Insert here your thesis' task.

Czech Technical University in Prague
Faculty of Electrical Engineering
Department of Microelectronics

Master's thesis

Ultra low quiescent current LDO regulator design

Bc. Lukáš Buryánec

Supervisor: doc. Ing. Jiří Jakovenko, Ph.D.

9th January 2018

Acknowledgements

Special thanks go to my consultants at STMicroelectronics company. Without them, this work could not be done. I would like to thank as well to my supervisor Doc. Ing. Jiří Jakovenko Ph.D for valuable advises and to my family for the support during my whole study.

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended. In accordance with Article 46(6) of the Act, I hereby grant a nonexclusive authorization (license) to utilize this thesis, including any and all computer programs incorporated therein or attached thereto and all corresponding documentation (hereinafter collectively referred to as the "Work"), to any and all persons that wish to utilize the Work. Such persons are entitled to use the Work in any way (including for-profit purposes) that does not detract from its value. This authorization is not limited in terms of time, location and quantity. However, all persons that makes use of the above license shall be obliged to grant a license at least in the same scope as defined above with respect to each and every work that is created (wholly or in part) based on the Work, by modifying the Work, by combining the Work with another work, by including the Work in a collection of works or by adapting the Work (including translation), and at the same time make available the source code of such work at least in a way and scope that are comparable to the way and scope in which the source code of the Work is made available.

In Prague on 9th January 2018 .....

Czech Technical University in Prague
Faculty of Electrical Engineering

Citation of this thesis

Buryánec, Lukáš. *Ultra low quiescent current LDO regulator design.* Master's thesis. Czech Technical University in Prague, Faculty of Electrical Engineering, 2018.

Abstrakt

Tato práce se zabývá návrhem LDO regulátoru napětí s ultra nízkým klidovým proudem v technologii BCD8 společnosti STMicroelectronics. Jedná se o návrh kompletního čipu včetně band-gap napěťové reference umožňující 4-bitový trimming, teplotní ochrany, nadproudové ochrany, enable logiky a proudové reference. V první kapitole jsou uvedeny definice důležitých parametrů týkajících se regulace napětí a funkce MOS tranzistorů. Druhá kapitola se věnuje principům napěťových regulátorů. Teorie bang-gap referencí je popsána v kapitole 3. Následující kapitola 4 je plně věnována samotnému návrhu kompletního čipu včetně popisu jednotlivých bloků. Výsledky simulací jsou uvedeny v kapitole 5.

Klíčová slova LDO, regulátor napětí, návrh integrovaných obvodů

Abstract

This thesis deals with design of a ultra-low quiescent current LDO voltage regulator in BCD8 technology of STMicroelectronics company. Design of a complete chip is described, including a band-gap voltage reference with 4-bit trimming circuit, thermal protection, current limiting circuit, enable control and a reference current generator as well.

The first chapter sums up definitions of voltage regulation parameters and important parameters of MOS transistors. The second chapter deals with principles of voltage regulators. Bang-gap voltage references theory and describing main types of band-gap references is done in chapter 3. Following chapter 4 is fully dedicated to the design of the complete chip with detailed description of each block. Simulation results are shown in chapter 5.

Keywords LDO, voltage regulator, integrated circuits design, ultra-low quiescent current

Contents

List of Figures

List of Tables

Introduction

Motivation and objectives

Power management blocks are parts of all electronic devices. Power management includes, among others, voltage regulators. The goal of voltage regulators is to assure stable output voltage regardless of input voltage variations, load conditions, transients etc.

Low drop-out regulators (LDOs) are linear voltage regulators capable of maintaining the input voltage even with its very low difference from the output voltage. This difference (drop-out voltage) can be as low as tens to hundreds of milivolts. Linear regulators function is based on a principle of dissipating energy across a pass transistor. Thanks to the very low drop-out voltage across the pass transistor, very high efficiency can be achieved. Nowadays, battery powered electronic devices such as cell phones are indivisible parts of everyday life. Thus, efficiency is one of the key features for long battery endurance of these devices. Besides the low drop-out voltage, there is another factor in terms of efficiency and it is a low quiescent current which is necessary especially when the LDO is operated at no or light small current conditions. These facts make micro-power integrated circuit design very important. Of course, everything in analog circuit design is a trade-off of something for something else. The lower quiescent current, the slower LDO is in general - worse transient responses, lower power supply rejection at higher frequencies etc. The challenge for a designer is to find optimal balance between these compromises and still fulfill desired specifications.

Problem statements

The main goal of this thesis is to design an ultra-low quiescent LDO voltage regulator in BCD8 technology provided by STMicroelectronics company. This designed LDO has to meet all requirements for fabrication in STMicroelectronics company. The emphasis has been put mainly on its efficiency. The quiescent current of the LDO core should be $I_q = 1$ μA or less at no load - $I_{load} = 0$ A at typical conditions (without the band-gap reference). The maximal drop-out voltage has to be $V_{drop} < 150$ mV at the maximal load current which is set to $I_{load} = 300$ mA. Because the application of this LDO could be as a block following a DC-DC buck converter, of which the output voltage could be less then one volt, the minimal output voltage of the LDO is set to be $V_{out} = 0.8$ V. The goal is to design a complete chip which includes a bad-gap voltage reference, enable control logic, a thermal protection, a current-limiting circuit and a reference current generator as well. Other parameters should be as follows:

- NMOS power transistor
- separate bias voltage and input voltage: V_{bias} = 2.75 to 5.5 V, V_{in} max 5.5 V
- output voltage: V_{out} = 0.8 to 4 V, by 50 mV step
- accuracy of the output voltage: ±2% at 27 °C
- power supply rejection ratio: PSRR = around 70 dB at 100 Hz
- optimal output capacitor C_{out} = 1 μF, stable with C_{out} = 500 nF to 10 μF with ESR in the range of 5 to 500 mΩ
- operating temperature range: -40 to 125 °C

The design of the band-gap voltage reference with the output voltage V_{out} = 0.8 V is and the maximum quiescent current I_q = 500 nA which will work in the range of the bias voltage V_{bias}. This design is supposed to include 4-bit trimming circuit.

Although, there are available ultra-low quiescent current LDOs on the market with a quiescent current in units of microamperes, some of their parameters are not usually sufficiently good, for example PSRR. The mentioned specification represents a quite unique LDO. Besides the ultra-low quiescent current, the ability of very low drop-out voltage even with non-ordinarily low output voltage and still achieving relatively high PSRR for a low-quiescent current category represents a potentially high demand on the market. A comparison of ultra-low quiescent current LDOs will be done at the end of this thesis.

Chapter 1

Definitions of parameters and theory of MOS transistors

Before explaining voltage regulation principles and theory, it is convenient to define and explain parameters used in further reading. Also, for better understanding, fundamental parameters of MOS transistors are explain in this chapter.

1.1 Voltage regulator parameters

The main goal of voltage regulators is to keep its output voltage ideally constant regardless of any circumstances which means that the ideal voltage regulator acts as a ideal voltage source. Of course, in reality, it cannot be possible to fulfill this definition. Hence, for classifying a quality of voltage regulators, some standard parameters are defined. Voltage regulators are closed-loop feedback systems. This means that static DC parameters and frequency dependent parameters have to be defined to satisfyingly describe their behavior.

1.1.1 Line regulation

Line regulation is a large-signal parameter which describes how the output voltage is affected by a change of the input voltage. It is a DC parameter. We can define load regulation as:

$$line\,regulation = \frac{\Delta V_{out}}{\Delta V_{in}} \tag{1.1}$$

where ΔV_{out} is a change in the output voltage caused by a change in the input voltage ΔV_{in}. Having been said that it is a DC parameter, load regulation does not consider frequency behavior of the regulator. As it will be described in a chapter 2, line regulation is mainly influenced by an open-open loop gain.

1.1.2 Load regulation

Load regulation is a DC large signal parameter as well. It describes the ability to keep the output voltage constant while there are changes of the load current. It is defined as:

$$load\,regulation = \frac{\Delta V_{out}}{\Delta I_{load}}. \tag{1.2}$$

Similarly as line regulation, the main factor in term of load regulation is an open-loop gain as well.

1.1.3 PSRR

Power supply rejection ratio is a parameter similar to line regulation. PSRR describes the ability of a regulator to keep the output voltage constant regardless of variation in the input voltage, but on contrary of line regulation, it considers frequency dependence of this ability. Another thing is that it is a small signal parameter, whereas line regulation is a large signal parameter. But in reality, we can say that PSRR for a very low frequency (almost DC) input variations it is approximately equal to line regulation. It is common to express PSRR in decibels. We can define PSRR as:

$$PSRR\,(f) = 20log\frac{v_{out}\,(f)}{v_{in}\,(f)}\ \text{[dB]}, \tag{1.3}$$

where v_{out} is a small signal change of the output voltage caused by a small signal change in the input voltage v_{in}. It is good to mention that sometimes PSRR is defined the same way as line regulation, but it is not that common. One of the most important aspect of PSRR is, again, an open-loop gain, but there are many factors which can influence PSRR.

1.1.4 Output noise

As in every electronic circuit, an unwanted noise signal appears at the output of the LDO. The amount of noise is usually represented by a noise spectral density - VSD_n for a chosen bandwidth - typically 10 Hz - 100 kHz. It is measured in units of $\frac{\mu V}{\sqrt{Hz}}$. There are three main types of voltage noise:

- thermal noise
- flicker noise - $1/f$ noise
- shot noise

Thermal noise has a characteristic of white noise - flat spectral density. On the other hand, flicker noise spectral density has a shape of 1/f which can be seen at lower frequencies where flicker noise is dominant. Output noise can be also classified by a single value - effective value - root mean square of the output noise voltage - V_{rms}. This can be calculated from the spectral density characteristic by integration across the bandwidth:

$$V_{n_{RMS}} = \sqrt{\int_{f_{min}}^{f_{max}} (VSD_n(f))^2 df}, \tag{1.4}$$

where f_{min} and f_{max} are the minimal and maximal frequency of the bandwidth.

Thermal noise is caused by thermally excited vibrations of carriers. Output noise level can be unacceptable high especially in low-power design. The need for low consumption forces to use large resistors and since the $V_{n_{rms}}$ of thermal noise caused by a resistor is given by [1]:

$$V_{n_{RMS}} = \sqrt{4kT \cdot B \cdot R}, \tag{1.5}$$

where k is a Boltzmann constant, T represents absolute temperature and B is the bandwidth, thermal voltage noise can be a problem in low power design.

MOS transistors are, besides others, sources of both thermal and flicker noise. For current thermal noise in MOS transistors working in strong inversion we can write [1]:

$$I_{n_{thermal}} = \sqrt{4kT\frac{2}{3} \cdot g_m \cdot B}, \tag{1.6}$$

where g_m is the transconductance of MOS.

Flicker noise in MOS structure can be significant at low frequencies. There are two theories of generation flicker noise in MOS structure. According to McWorther [1], [2], flicker noise is caused by random trapping carries near $Si - SiO_2$ interface. The charge fluctuation causes fluctuations in surface potential which affects channel carrier density. Another theory has been published by Hooge [2]. This theory describes flicker noise as consequence of bulk mobility fluctuations. For flicker drain current noise density of MOS in strong inversion using McWorther model it holds [3]:

$$I_{n_{flicker}} = \sqrt{\frac{K_F \cdot g_m^2}{C_{ox} \cdot W \cdot L_{eff} \cdot f^{E_F}}}, \tag{1.7}$$

where K_F represents a flicker noise coefficient and E_F is a flicker noise frequency exponent, W and L_{eff} represent the width and the effective length of a channel. C_{ox} means oxide capacitance and f represents frequency.

1.1.5 Quiescent current

Quiescent current is a current that is needed for proper functionality of the internal circuitry and it is not delivered from the input supply to the load. Sometimes it is called as ground current. Simple definition can be:

$$I_q = I_{in} - I_{load}, \tag{1.8}$$

where I_{in} represents the current supplied from the input (biasing voltage) and I_{load} is the load current. From this equation it is obvious that it has a major impact on efficiency of the regulator. The quiescent current can be either almost independent of the load current or it can be a function of the load current. Second option is suitable especially in low power design where, in terms of efficiency, very low quiescent current is needed at no or light loads. This quiescent current would be unnecessarily low at heavy loads and would unnecessarily degrade many parameters of given regulator.

1.1.6 Efficiency

Efficiency is simply given as a ratio between power delivered to the load $P_{delivered}$ and power consumed in the internal circuitry $P_{internal}$:

$$\eta = \frac{P_{load}}{P_{load} + P_{internal}}. \tag{1.9}$$

Efficiency is further discussed in section 2.4.

1.1.7 Dropout voltage

Series regulator are described by a parameter called drop-out voltage. Drop-out voltage is simply given by the difference between the input voltage and the output voltage when the power transistor of the regulator entered linear region of operation - so called *drop-out mode*. Thus, it represents the minimal difference between the input voltage and the output voltage for the given load current:

$$V_{drop} = V_{in} - V_{out} \quad at\ drop - out\ mode. \tag{1.10}$$

For proper function of the regulator, the power transistor needs a certain voltage drop to ensure that it works in active - saturation region (using MOS transistors), for the difference between the input and the output voltage it must hold:

$$V_{in} - V_{out} > V_{DS_{min}}, \tag{1.11}$$

where $V_{DS_{min}}$ is the minimal saturation voltage. If this condition is not fulfilled, the transistor enters linear region and starts to act as a simple resistor with a value of resistance called $R_{DS_{on}}$. The closed-loop system becomes an open-loop system and regulation capabilities do not apply. Generally the drop-out voltage can be as low as tens to hundreds of milivolts.

1.1.8 Line and load transient response

These characteristics describe behavior of the regulator in time domain in a response to the step change in the load current or the input voltage. As in every closed-loop feedback systems, the reaction of the loop is not immediate, so undesired undershoots, overshoots and ringing can occur in the step transient response. Every active element in the circuit has a certain time delay because parasitics capacitances need to be charged/discharged before the element can change something in the response. There are many factors that have an impact on the transient response. Besides others, the unity gain frequency - UGF, slew rate and phase margin have significant influence. Achieving high UGF and high slew rate is a trade-of between quiescent current. Therefore, especially in low quiescent current design, achieving satisfying transient response can cause difficulties. Transient behavior is disclosed in more details in other chapters.

1.2 MOS transistors - modes of operation

Metal-oxide-semiconductor transistors are field-effect transistors, which mean that they act similarly as a voltage controlled current source. The input voltage between their gate and source terminals controls the current flowing through their drain terminal. This chapter does not discuss transistor effect theory. It explains fundamental parameters and facts that are important for analog IC design. For further reading, for the sake of simplicity, the parameters are explained on NMOS transistors, equations for PMOS transistors are analogous. The active operation of MOS transistors can be divided into two main modes called *Strong inversion* and *Weak inversion* also known as *Subthreshold region.*

1.2.1 Strong inversion

Strong inversion mode can be divided into two regions - *Saturation region* and *Linear region.* If a transistor is in strong inversion, it enters saturation region when [4]:

$$V_{GS} - V_{th} > 4V_T;\ V_{DS} > V_{GS} - V_{th}, \tag{1.12}$$

where V_{th} is the threshold voltage and V_T is the thermal voltage $V_T = \frac{kT}{q}$. In this region, the relation between drain current I_D and V_{GS} is given by [5]:

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L} \cdot (V_{GS} - V_{th})^2 (1 + \lambda V_{DS}), \tag{1.13}$$

where μ_n represents the velocity of carries, C_{ox} is the capacitance of the oxide, W and L are physical dimensions of the transistor - width and length of the channel and λ is the channel

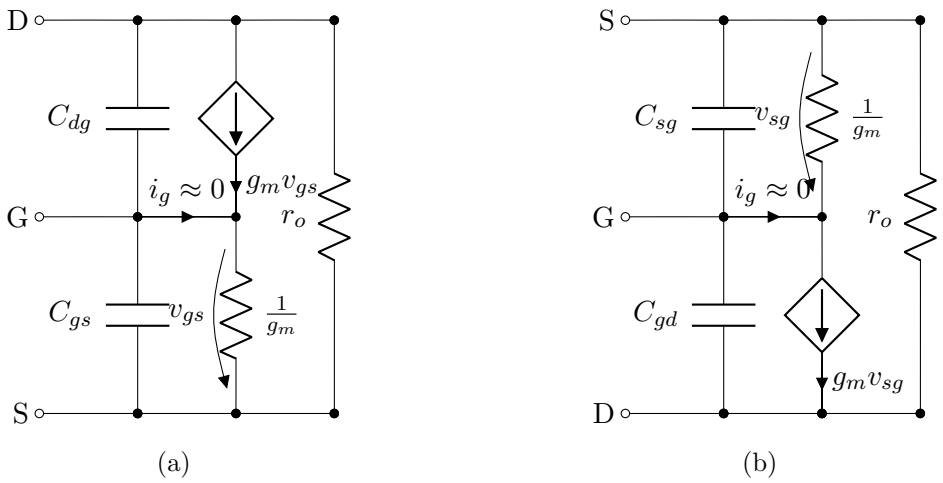


Figure 1.1: Ebbers-Moll small signal model - NMOS (a) and PMOS (b)

length modulation parameter which depends on physical dimensions and bias conditions, it is very complex and no such a simple model is available, but it is approximately inverse-proportional to:

$$\lambda \sim \frac{1}{V_E L}, \tag{1.14}$$

where V_E is a parameter similar to Early voltage in bipolar transistors. Typical values for parameter λ are usually in a range of 0.1 - 0.001.

If the second condition in equation 1.12 is not fulfilled so that: $V_{DS} < V_{GS} - V_{th}$, the transistor is in linear region where for the drain current it holds [5]:

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L} \cdot \left(V_{GS} - V_{th} - \frac{V_{DS}}{2}\right) \cdot V_{DS}, \tag{1.15}$$

in this region, the transistor behaves as a resistor controlled by the V_{GS} voltage.

For design, it is important to know a small-signal model. Small-signals models use parameters that are defined for a small signal variation around the operating point. Such a model is depicted in figure 1.1. The most important parameter in this Ebbers-Moll model is the transconductance g_m for which, in strong inversion saturation region, we can write:

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \mid_{OP} = \sqrt{2\mu C_{ox}\frac{W}{L} I_D}. \tag{1.16}$$

Another parameter r_o represents the small-signal output resistance defined as [4]:

$$r_o = \frac{\partial V_{DS}}{\partial I_D} \mid_{OP} = \frac{1}{\lambda I_D}. \tag{1.17}$$

Other parameters in the model represent the gate-source capacitance C_{gs} and drain-gate capacitance - called Miller capacitance - C_{dg}. These capacities are necessary to be bared in mind when analyzing higher frequency behavior and doing stability analysis.

1.2.2 Weak inversion - Subthreshold region

Even if $V_{GS} < V_{th}$, thanks to the thermal energy, some electrons at the source can still flow to the drain and there is still some drain current present which is called subthreshold current. If it holds [4]:

$$V_{GS} - V_{th} < 4V_T, \tag{1.18}$$

the transistor is said to be in weak-inversion region. In this region, for the drain current we can write [4], [6]:

$$I_D = I_s e^{\frac{V_{GS}-V_{th}}{nV_T}} \left(1 - e^{-\frac{V_{DS}}{V_T}}\right), \tag{1.19}$$

where I_s is the specific current $Is = \mu_n C_D \frac{W}{L} V_T^2$ and C_D represents the diffusion capacitance which depends on the doping density N_A and the bulk-source voltage V_{BS}:

$$C_D = \sqrt{\frac{q\varepsilon_{si} N_A}{2(\Phi_{si} - V_{BS})}}, \tag{1.20}$$

where ε_{si} is the permitivity of silicon and Φ_{si} represents the built-in potential of silicon. Finally, the n parameter is given by: $n = 1 + \frac{C_D}{C_{ox}}$. From equation 1.19, it can be seen that the drain current i_D is exponentially dependent on the V_{GS} voltage - similarly as in PN diodes and also on voltage V_{DS}. In weak inversion region, we can also say that there is a region similar to saturation in strong inversion - If the $V_{DS} \gg V_{th}$, the last term in the equation is negligible so the drain current I_D starts to be almost independent on V_{DS}. In reality, $v_{DS} >\approx 4V_{th}$ is sufficient [4], [6] and we can say that the transistor is in weak inversion saturation region - in weak inversion we can achieve the lowest minimal saturation voltage $V_{DS_{min}}$.

The small-signal transconductance in weak inversion saturation is defined as [4]:

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \mid_{OP} = \frac{I_D}{nV_T}. \tag{1.21}$$

From this equation, it is clear that weak inversion has a major advantage comparing to strong inversion - in weak inversion, the transconductance is directly proportional to the drain current, on the other hand, in strong inversion, it is dependent on the square root of the drain current. Therefore, for the maximal ratio $\frac{g_m}{I_D}$, the transistor should be operated in weak inversion region.

1.2.3 Strong vs Weak inversion in IC design

Operation of MOS in weak inversion is desired mainly for two reasons - the highest $\frac{g_m}{I_D}$ and the lowest minimal saturation voltage $V_{DS_{min}}$. Let us now compare other features of strong inversion and discuss its suitability on design of the two main blocks of analog IC design: differential pairs (differential amplifier) and current mirrors.

Differential pair: Usually, the goal is to achieve the maximal gain of the differential amplifier for the given biasing current - weak inversion is preferred for the maximal g_m which leads to larger gain. Larger g_m also reduces the input noise [7]. Also, minimal voltage offset is desired - in weak inversion, the V_{GS} of the differential pair transistor is the lowest for the given biasing current, so the mismatch between the drain currents causes minimal differences between V_{GS} of each transistor, thus the minimal offset voltage. Another parameter is bandwidth of the differential amplifier which is directly proportional to g_m, so again, the maximal $\frac{g_m}{I_D}$ is desired - thus weak inversion. On the other hand, operating the transistors in weak

Table 1.1: Weak inversion features

Pluses	Cons
• maximal g_m for given I_D - maximal $\frac{g_m}{I_D}$ ratio • minimal $V_{DS_{min}}$ • minimal V_{GS} • differential amplifier: – minimal voltage offset – maximal gain for given bias – minimal input voltage noise – maximal bandwidth	• larger area - larger capacitances • slower transient behavior • current mirror: – larger current offset – lower output impedance (for short channel lengths)

inversion leads to high $\frac{W}{L}$ ratios which, besides larger area on a chip, means larger C_{gs} which could show deterioration at transient response behavior.

Current mirror: In current mirrors, usually the goal is to achieve the minimal current offset between the mirroring current branches (difference in mirroring currents). To achieve this, transistors of the current mirror should be operated in strong inversion - V_{GS} is higher in strong inversion so any mismatch in the layout of the current mirror and, for instance, the difference between the threshold voltage V_{th} of each transistor has smaller impact on the overdrive voltage $V_{ov} = V_{GS} - V_{th}$ which then provides lower current offset. There is another thing, for achieving lower current offset - large output impedance is needed to minimize the dependence of the mirroring current on V_{DS}. For large output impedance r_o, the parameter λ should be minimal - the channel length should be long enough. If we wanted to operate current mirrors in weak inversion and still have large output impedance, it would yield to large $\frac{W}{L}$ ratio because of the long channel. These are the reasons why strong inversion is preferred in current mirrors design.

Overall summary of weak inversion features is depicted in table 1.1.

1.2.4 Threshold voltage

Threshold voltage V_{th} defines the voltage V_{GS} at which the channel between source and drain is made and a certain amount of current flows through the channel. This parameter is dependent on a variety of condition such as:

- so called body effect
- temperature
- technology parameters - such as doping densities, oxide thickness etc.
- physical dimensions

It is important to mention so called *body effect* because the threshold voltage can be influenced by the designer. The body effect describes dependence of the threshold voltage on the source-body voltage V_{SB}. The effect can be described by [8]:

$$V_{thB} = V_{th0} + \gamma \left(\sqrt{V_{SB} + 2\varphi_B} - \sqrt{2\varphi_B} \right), \tag{1.22}$$

where V_{th0} is the threshold voltage when $V_{SB} = 0$, φ_B represents surface-bulk potential when $V_{SB} = 0$ and γ is a body effect technology dependent parameter. This means that with higher V_{SB} the threshold voltage increases. The lowest V_{th} is for $V_{SB} = 0$. Parameter γ depends on the oxide thickness - the thinner the oxide thickness, the lower the this parameter, thus lower V_{th}.

Temperature dependence of threshold voltage is quite complex. Yet, it can be usually modeled as simple linear dependence on temperature - with increasing temperature, V_{th} decreases. Further reading can be found in [9], [10].

Dependence of V_{th} on physical dimensions is a really complex case. In classic CMOS technologies, two effect can occur. Due to so called short-channel effect, V_{th} decreases with decreasing channel length and V_{th} increases with decreasing channel width of the transistor - called narrow-channel effect. In modern technologies, the dependency of V_{th} on the dimensions can be different. The theory about that fact is complex and not a part of this thesis, further reading can be found in [11], [9], [10].

1.2.5 Subthreshold leakage current

Even when V_{GS}=0 there is still some current flowing through the channel - *leakage current*. This current shows exponential dependence on temperature. Approximately, it doubles with every 10 °C of rise in temperature. From equation 1.19 and for $V_{GS} = 0$ we can write for the leakage current:

$$I_D = I_s e^{\frac{-V_{th}}{nV_T}} \left(1 - e^{-\frac{V_{DS}}{V_T}} \right), \tag{1.23}$$

where I_s is the specific current $Is = \mu_n C_D \frac{W}{L} V_T^2$. From this equation, it is obvious that low threshold voltage devices are more leaky. It is useful to notice, that with higher V_{DS} voltage, the leakage current increases as well. One way to reduce the leakage current is to increase V_{th}, for instance by body effect. Also, longer channel length than minimal should be used to reduce the leakage current. It is also useful to mention that usually NMOS devices are more leaky than PMOS devices.

Chapter 2

Voltage Regulation Theory

This chapter discusses basic principles of voltage regulation theory. Division into main groups of regulators - parallel and series and their principles are disclosed. At the both groups of regulators, detailed scenarios of line and load regulation are described. Several materials regarding voltage regulators have been published, yet mathematical statements describing line and load regulation for every disclosed type have not been found. That is the reason why, in this chapter, explanation of load and line regulation is supported by mathematical expressions, these expressions are then plotted as a function of parameters with a significant influence.

2.1 Regulator division

Voltage regulators can be divided into two main groups - *parallel regulators* (also called *shunt regulators*) and *series regulators*. Choosing one type over the other depends on a purpose of application because each type has its advantages and disadvantages comparing to the other type, for instance, in power consumption, operating input voltage range, accuracy etc. Another division can be made according used regulating power transistors - *bipolar transistors - BJT* and *unipolar - MOS transistors*. Both types can be further divided into *NPN, PNP* type or *N-channel, P-channel* respectively. This thesis only deals with MOS transistors since their usage in integrated circuits is more common these days.

2.2 Parallel regulators

Basic principles of operation of parallel regulators can be compared to using zener diodes as voltage regulators. Parallel regulators have only two pins - an output pin with the regulated voltage V_{out} and a common ground pin. Application notes for parallel regulators can be found, for instance, in [12]. The input voltage is connected to the output pin via an input resistor. The parallel regulator provides a constant voltage at its output terminal, so the current which flows through the input resistor R_s is given by:

$$I_{R_s} = \frac{V_{in} - V_{out}}{R_s}. \tag{2.1}$$

From this situation it is obvious that this current does not change with a changing load current. This current is then divided to a current which flows to the load and to a current flowing to the regulator - *quiescent current*. This describes the key feature of parallel regulators -

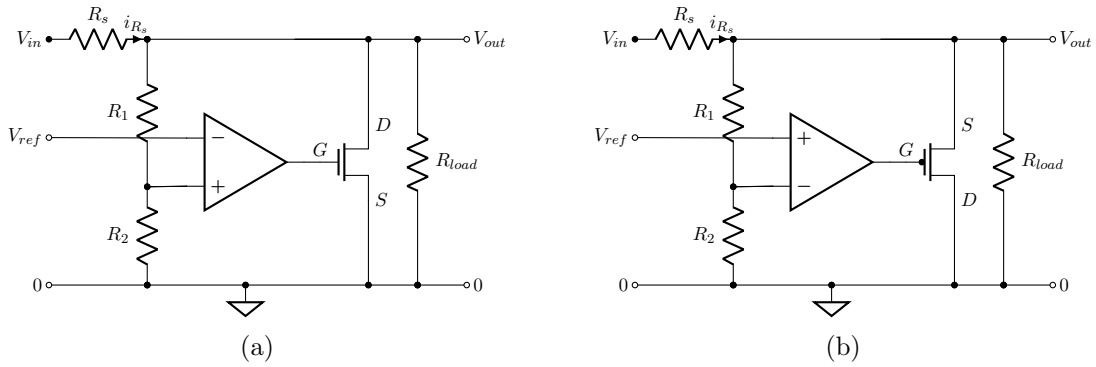


Figure 2.1: Parallel regulator using NMOS (a) and PMOS (b)

the sum of the quiescent current and the load current is always constant for the given input voltage. If the load current decreases, the quiescent current increases and vice versa so still the same voltage drop across the input resistor occurs for the given input voltage. If the input voltage changes, the current through the input resistor changes as well. Thus, for the given load current, the quiescent current changes regarding the input voltage.

The value of the input resistor has to be calculated regarding the operating input voltage range and the desired maximal load current:

$$R_s = \frac{V_{in_{MIN}} - V_{out}}{I_{load_{MAX}} + I_{q_{MIN}}}, \tag{2.2}$$

where $V_{in_{MIN}}$ is the minimal expected operating input voltage, $I_{load_{MAX}}$ is the maximum desired load current and $I_{q_{MIN}}$ is the minimal quiescent current needed to ensure proper functionality of the regulator. This value of the resistor has to determined by the minimal input voltage, if not, it would not be possible to assure the maximal load current under all conditions. On the other hand, the maximal input voltage $V_{in_{MAX}}$ determines the maximal power dissipation across the input resistor:

$$P_{R_{sMAX}} = \frac{(V_{in_{MAX}} - V_{out})^2}{R_s}. \tag{2.3}$$

Power dissipation in the regulator itself is maximal when the maximal input voltage ($V_{in_{MAX}}$ is applied - the current through the input resistor is maximal $I_{R_{s_{MAX}}}$ and there is no load current because the whole current through the input resistor flows to the regulator - the quiescent current is then maximal $I_{q_{MAX}}$. The maximal power dissipation in the regulator is then given by:

$$P_{reg_{MAX}} = V_{out} \cdot I_{q_{MAX}} = V_{out} \cdot I_{R_{s_{MAX}}}. \tag{2.4}$$

Thus, the maximal operating input voltage is restricted only by the maximal allowed power dissipation across the input resistor and the dissipation in the regulator. This shows another key feature of parallel regulators - they can be operated with very high input voltage.

Let us now explore more detailed operation of parallel regulators. A possible simplified realization of a parallel regulator is sketched in figure 2.1a for a NMOS power transistor and in figure 2.1b using a PMOS transistor. Considering a steady state scenario, the operational amplifier and a feedback network consisted of resistors R_1 and R_2 control the voltage between

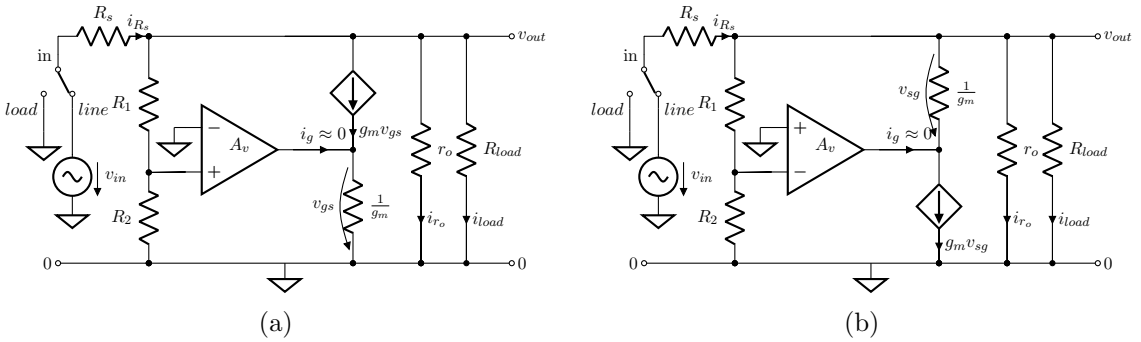


Figure 2.2: Small signal model of parallel regulator using NMOS (a) and PMOS (b)

the inputs of the opamp. This voltage is ideally zero considering infinite gain of the opamp. Thus, the output voltage is simply given as:

$$V_{out} = V_{ref} \cdot \left(1 + \frac{R_1}{R_2}\right), \tag{2.5}$$

where V_{ref} is the reference voltage provided, for instance, by a band-gap voltage reference. The power transistor conducts a proper amount of current so the sum of the quiescent current and the load current - the current through the input resistor R_s is still constant for the given input voltage.

2.2.1 Load and line regulation

In this section, both load and line regulation are described and are supported by mathematical expressions. Even though, as it has been said, the line and load regulation are large-signal parameters, for mathematical expressions they will be considered as small-signal parameters for better modeling. The inaccuracy of this decision is not too big to not have an overview what parameters have an impact on the line and load regulation. Therefore $\frac{v_{out}}{v_{in}} \approx \frac{\Delta V_{out}}{\Delta V_{in}}$, respectively $\frac{v_{out}}{i_{load}} \approx \frac{\Delta V_{out}}{\Delta I_{load}}$.

Now considering a load regulation scenario: Assuming the constant input voltage, when the load current starts to increase, the voltage drop across the input resistor R_s gets higher. Thus, the output voltage decreases and then the voltage at the plus terminal - in the case of NMOS power transistor or the minus terminal - in the case of PMOS of the opamp decreases as well. When this happens, the amplifier has a tendency to compensate the voltage difference between its terminals and, in the case of NMOS power transistor, pushes its output voltage lower or higher in the case of PMOS power transistor. The opamp drives the power transistor which stars to conduct less current. Thus, it equalizes the sum of the quiescent current and the load current. Therefore, it provides the desired regulated output voltage.

For mathematical description, a small signal model of a NMOS regulator is depicted in figure 2.2a and of a PMOS regulator in figure 2.2b. For the load regulation of the NMOS regulator we can write:

$$\begin{aligned} v_{out} = i_{load} \cdot R_{load} - \frac{v_{out}}{R_s} \cdot \frac{r_o \cdot R_{load}}{r_o + R_{load}} - \\ \left(v_{out} \cdot A_v \cdot \frac{R_2}{R_1 + R_2} \cdot g_m \cdot \frac{r_{o \cdot R_{load}}}{r_o + R_{load}}\right), \end{aligned} \tag{2.6}$$

Table 2.1: Model values for parallel regulators

parameter:	value:	parameter:	value:
A_v	1000 (-)	R_{load}	$\frac{1}{I_{load}}$
R_1	2.5 MΩ	g_m	$\sqrt{10 \cdot (I_{q_{MIN}} + I_{load_{MAX}}) - I_{load})}$
R_2	5 MΩ	r_o	$\frac{1}{\lambda \cdot ((I_{q_{MIN}} + I_{load_{MAX}}) - I_{load})}$
$I_{q_{MIN}}$	100 μA	R_s	10 Ω or sweep
$I_{load_{MAX}}$	100 mA	λ	0.1 V^{-1} or sweep

$$\frac{v_{out}}{i_{load}} = \frac{1}{1 + \frac{r_o}{R_s \cdot (r_o + R_{load})} + A_v \cdot \frac{R_2}{R_1 + R_2} \cdot g_m \cdot \frac{r_o}{r_o + R_{load}}}, \tag{2.7}$$

or equivalently for the PMOS regulator we can write:

$$\begin{aligned} v_{out} = i_{load} \cdot R_{load} - \frac{v_{out}}{R_s} \cdot \frac{r_o \cdot R_{load}}{r_o + R_{load}} - \\ \left(\left(v_{out} + v_{out} \cdot A_v \cdot \frac{R_2}{R_1 + R_2} \right) \cdot g_m \cdot \frac{r_{o \cdot R_{load}}}{r_o + R_{load}} \right), \end{aligned} \tag{2.8}$$

$$\frac{v_{out}}{i_{load}} = \frac{1}{1 + \frac{r_o}{R_s \cdot (r_o + R_{load})} + \left(1 + A_v \cdot \frac{R_2}{R_1 + R_2}\right) \cdot g_m \cdot \frac{r_o}{r_o + R_{load}}}, \tag{2.9}$$

where A_v is the DC voltage gain of the opamp. The transconductance g_m of the power transistor, assuming operation in strong inversion, is given as:

$$g_m = \sqrt{2\mu C_{ox} \frac{W}{L} I_d} \approx \sqrt{2\mu C_{ox} \frac{W}{L} ((I_{q_{MIN}} + I_{load_{MAX}}) - I_{load})}, \tag{2.10}$$

where μ. In parallel regulators, the drain current is not equal to the load current, but it is approximately the difference between the current flowing through the input resistor and the load current. Assuming that the quiescent current of the opamp does not change with the load current and is negligible we can use that $I_d \approx (I_{q_{MIN}} + I_{load_{MAX}}) - I_{load}$. The similar way we can write for the output resistance of the power transistor r_o:

$$r_o = \frac{1}{\lambda \cdot I_d} \approx \frac{1}{\lambda \cdot ((I_{q_{MIN}} + I_{load_{MAX}}) - I_{load})}. \tag{2.11}$$

For the sake of simplicity, for R_{load} simplyfing of $R_{load} \approx (R_1 + R_2) \parallel R_{load}$ is used further in this chapter since R_l and R_2 are usually in a order of MΩ.

We can see that the expressions for the load regulation are almost identical for the both cases. The dependency of the load regulation on the load current I_{load} is plotted if figure 2.3 for model values listed in table 2.1. Only one graph is presented since the difference between the NMOS and PMOS case cannot be observed. It is obvious that the best load regulation is at the medium load range. At the light loads, the regulation is worse due to the fact, that the drain current is maximal, thus the lowest r_o which leads to a lower value of expression $\frac{r_o}{r_o + R_{load}}$ in the equations. On the other hand, at the heavy loads, the load regulation also

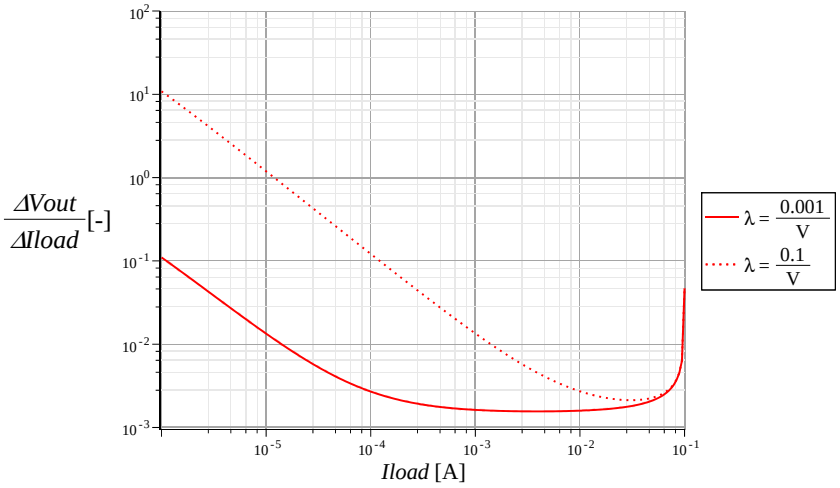


Figure 2.3: Load regulation of parallel regulators

get worse due the the fact, that the the drain current is the lowest, thus lowest g_m. We can see as well that the load regulation gets better for smaller channel length modulation λ. The load regulation is practically independent on the value of the input resistor, that is why the results are not shown for its different values.

Another situation is a line regulation scenario. Let us assume the constant load current. Suddenly, the input voltage changes, for instance, it increases. At this moment, the voltage drop across the input resistor is still the same thanks to the constant current situation, so the output voltage increases as well as the difference between the input terminals of the opamp. Therefore, it is necessary to increase the quiescent current to the proper value so the voltage drop across the input resistor increases to lower the output voltage back to the adjusted value. This is done by driving the gate of the power transistor higher by the opamp in the NMOS regulator or lower in the PMOS regulator. That forces the power transistor to conduct more current - the quiescent current increases and the output voltage decreases back to the adjusted value. We can describe this situation mathematically using, again, the models in figures 2.2a and 2.2b. For the NMOS regulator we cane write:

$$v_{out} = \frac{v_{in} - v_{out}}{R_s} \cdot \frac{r_o \cdot R_{load}}{r_o + R_{load}} - v_{out} \cdot A_v \cdot \frac{R_2}{R_1 + R_2} \cdot g_m \cdot \frac{r_o \cdot R_{load}}{r_o + R_{load}}, \tag{2.12}$$

$$\frac{v_{out}}{v_{in}} = \frac{1}{\frac{R_s \cdot (r_o + R_{load})}{r_o + R_{load}} + 1 + A_v \cdot \frac{R_2}{R_1 + R_2} \cdot g_m \cdot R_s} \tag{2.13}$$

and equivalently for the PMOS regulator we can express the line regulation as:

$$v_{out} = \frac{v_{in} - v_{out}}{R_s} \cdot \frac{r_o \cdot R_{load}}{r_o + R_{load}} - \left(v_{out} + v_{out} \cdot A_v \cdot \frac{R_2}{R_1 + R_2} \right) \cdot g_m \cdot \frac{r_o \cdot R_{load}}{r_o + R_{load}}, \tag{2.14}$$

$$\frac{v_{out}}{v_{in}} = \frac{1}{\frac{R_s \cdot (r_o + R_{load})}{r_o + R_{load}} + 1 + \left(1 + A_v \cdot \frac{R_2}{R_1 + R_2}\right) \cdot g_m \cdot R_s}. \tag{2.15}$$

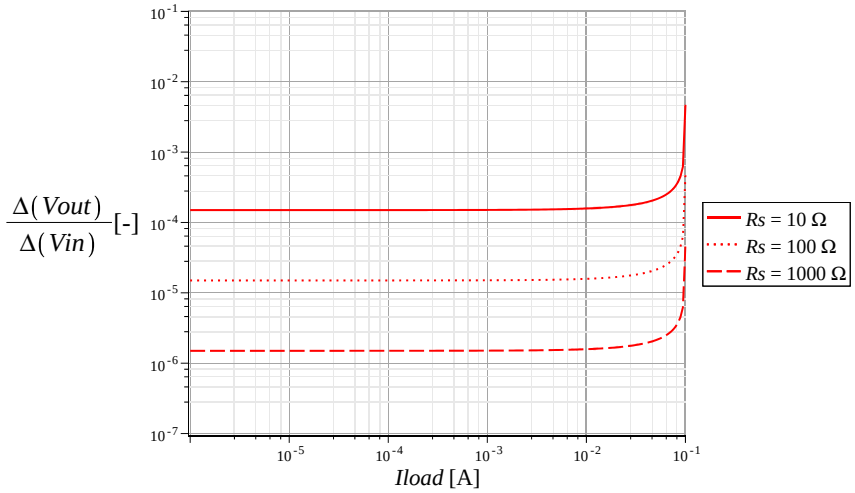


Figure 2.4: Line regulation of parallel regulators

The dependency of the line regulation on the load current I_{load} is plotted in figure 2.4 for model values listed in table 2.1. Since the expressions for the line regulation are almost identical for the both cases, only one graph is presented since the difference between the NMOS and PMOS case cannot be observed. The line regulation gets worse at heavy loads because of the lowest drain current, thus lowest g_m. We can also tell, that the highest value of the input resistor, the better the line regulation. This makes sense because we can imagine that the input resistor makes a simple voltage divider with the load resistor. But it is necessary to mention that the model is calculated with the constant current through the input resistor R_s as the sum of $I_{q_{MIN}} + I_{load_{MAX}}$, this means, for example, the value of the input resistor $R_s = 10^3\ \Omega$ and the maximal load current $I_{load} = 0.1$ A would make the input voltage $V_{in} \approx 100$ V, which may not be possible, of course.

2.3 Series regulators

Besides parallel regulators, there is a group of series regulators. Comparing to parallel regulators, series regulators have three pins - an input voltage pin, an output voltage pin and a common ground pin. Series regulators act as a variable resistor connected in series with the load resistor. This makes a variable voltage divider. Assuming the constant input voltage, when the load current increases, the voltage drop across the variable resistor increases as well, so the output voltage decreases. To achieve the adjusted output voltage, the internal circuitry has to lower the resistance of the variable resistor to the proper value to decrease the voltage drop across the variable resistor. On the other hand, assuming the constant load current, when the input voltage increases, the voltage drop across the variable resistor is still the same due to the same load current. That fact makes the output voltage be higher than desired. The internal circuitry has to increase the resistance of the variable resistor to provide the desired output voltage.

This principle of operation determines substantial differences from parallel regulators. The quiescent current is not necessarily dependent on the load current, instead, it can be only given by the internal consumption of the internal circuitry. Another difference is that the whole excess of power is dissipated across the internal variable pass element. Thus, the

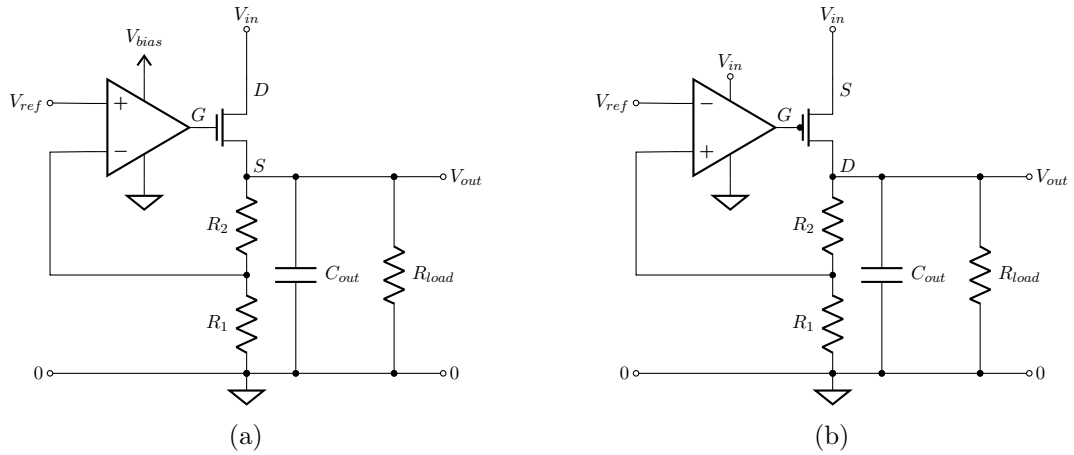


Figure 2.5: Series regulator using NMOS (a) and PMOS (b)

maximal dissipated power and operating temperature have to be considered and they limit the application of the circuit. The power dissipation in the regulator is given as the sum of the power dissipated on the variable pass element and the power consumption of the internal circuitry:

$$P_{reg} = V_{in} \cdot I_q + (V_{in} - V_{out}) \cdot I_{load}, \tag{2.16}$$

where I_q is the quiescent current needed for proper operation of the regulator. The fact that the input voltage is directly connected to the input pin also limits the allowed input voltage range, due to the maximal allowed voltage of the internal circuitry parts.

Now more detailed operation of series regulators can be disclosed. A simplified realization of a series regulator is sketched in figure 2.5a for a NMOS power transistor and in figure 2.5b using a PMOS transistor. The circuit is a closed-loop system. The output capacitor C_{out} is necessary for assuring stability of the circuit. The theory of stability will be discussed later in the LDO design chapter. Considering a steady state scenario, similarly to parallel regulators, the opamp and a feedback network of resistors R_1 and R_2 control the voltage between inputs of the opamp. Assuming infinite gain of the opamp, the output voltage is, again, simply given as:

$$V_{out} = V_{ref} \cdot \left(1 + \frac{R_1}{R_2}\right), \tag{2.17}$$

To ensure proper regulation, the power transistor has to work in saturation region. To keep the transistor in saturation, for the difference between its drain and source terminals (e.g. the difference between input and output voltage) it must hold:

$$V_{in} - V_{out} > V_{DS_{min}}, \tag{2.18}$$

where $V_{DS_{min}}$ is the minimal saturation voltage, when this condition is not fulfilled, the transistor enters linear region and starts to act as a simple resistor with a value of resistance called $R_{DS_{ON}}$ and the closed-loop system becomes an open-loop system and regulation capabilities do not apply. It can be said that the regulator entered a *drop-out mode*. The voltage drop across the power transistor in the drop-out mode - *drop-out voltage* V_{drop} depends on the $R_{DS_{ON}}$ which is given mainly by physical dimensions of the transistor. Generally it can be in the order of tens to hundreds of milivolts, thus the shortcut *LDO - low drop-out*.

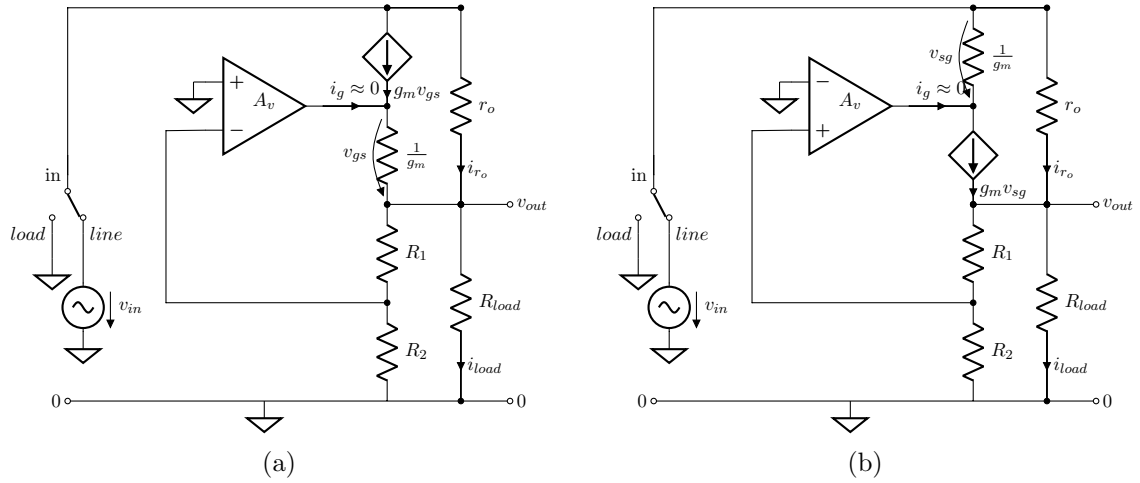


Figure 2.6: Small signal model of series regulator using NMOS (a) and PMOS (b) power transistor

In the case of the NMOS regulator, for achieving low drop-out voltage, it is necessary to assure higher biasing voltage V_{bias} of the opamp than the input voltage V_{in} which is connected to the power transistor. This is necessary for making a larger headroom for the V_{gs} of the power NMOS transistor. Larger biasing voltage can be either supplied using only V_{in} and a charge pump or using separate biasing voltage for V_{bias}.

2.3.1 Load and line regulation

Now the load regulation is depicted. Assuming the constant input voltage, when the load current starts to increase, the voltage drop across the power transistor gets higher. Therefore, the output voltage decreases and then the voltage at the minus terminal of the opamp - in the NMOS power transistor case or the plus terminal - in the case of PMOS decreases as well. When this happens, the opamp has a tendency to compensate the voltage difference between its input terminals and makes its output voltage increase - in the case of NMOS transistor or decrease - in the case of used PMOS transistor. The power transistor starts to conduct more current which appears similarly as lowering its resistance. Thus, it provides the adjusted output voltage. For mathematical expressions a small signal model in figures 2.6a and 2.6b can be used. Again, there is a small inaccuracy because the line and load regulation are large-signal parameters, as it has been already mentioned in the parallel regulators section.

For the load regulation of the NMOS regulator we can write:

$$v_{out} = i_{load} \cdot R_{load} + \left(-v_{out} \cdot A_v \cdot \frac{R_2}{R_1 + R_2} - v_{out} \right) \cdot g_m \cdot R_{load} - \frac{v_{out}}{r_o} \cdot R_{load}, \tag{2.19}$$

$$\frac{v_{out}}{i_{load}} = \frac{1}{\frac{1}{r_o} + \frac{1}{R_{load}} + \left(1 + A_v \cdot \frac{R_2}{R_1 + R_2}\right) \cdot g_m}, \tag{2.20}$$

and similarly for the PMOS regulator we can write:

$$v_{out} = i_{load} \cdot R_{load} - v_{out} \cdot A_v \cdot \frac{R_2}{R_1 + R_2} \cdot g_m \cdot R_{load} - \frac{v_{out}}{r_o} \cdot R_{load}, \tag{2.21}$$

Table 2.2: Model values for series regulators

parameter:	value:	parameter:	value:
A_v	1000 (-)	R_{load}	$\frac{1}{I_{load}}$
R_1	2.5 MΩ	g_m	$\sqrt{10 \cdot (I_{D_{MIN}} + I_{load})}$
R_2	5 MΩ	r_o	$\frac{1}{\lambda \cdot (I_{D_{MIN}} + I_{load})}$
$I_{D_{MIN}}$	100 nA	λ	sweep

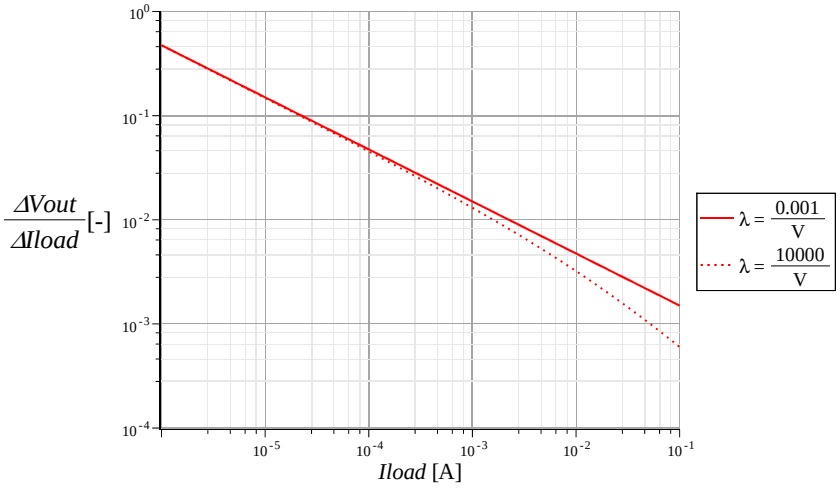


Figure 2.7: Load regulation of series regulators

$$\frac{v_{out}}{i_{load}} = \frac{1}{\frac{1}{r_o} + \frac{1}{R_{load}} + A_v \cdot \frac{R_2}{R_1+R_2} \cdot g_m}, \tag{2.22}$$

It can be seen that the expressions for the load regulation are almost identical for the both cases. The dependency of the load regulation on the load current is plotted on figure 2.7 for model values listed in table 2.2. Only one graph is present since the difference between the NMOS and PMOS case cannot be observed. It can be seen that the best load regulation is at heavy loads due the highest drain current which means highest g_m of the power transistor which is determined by the load current. The transconductance g_m is directly proportional to the load current which represents the main difference comparing to parallel regulators. For g_m we can write:

$$g_m = \sqrt{2\mu C_{ox} \frac{W}{L} I_d} \approx \sqrt{2\mu C_{ox} \frac{W}{L} (I_{D_{MIN}} + I_{load})}, \tag{2.23}$$

where $I_{D_{MIN}}$ represents the minimal current through the power transistor. This current may be given as the current through the voltage feedback divider. Surprisingly, at the first look, the load regulation gets better with higher channel length modulation - λ. Actually, it makes sense because when the load current gets higher, the output voltage drops and this makes higher V_{DS} of the power transistor which causes higher conduction of the transistor when λ is higher - this reduces the output voltage drop and improves the load regulation. It must

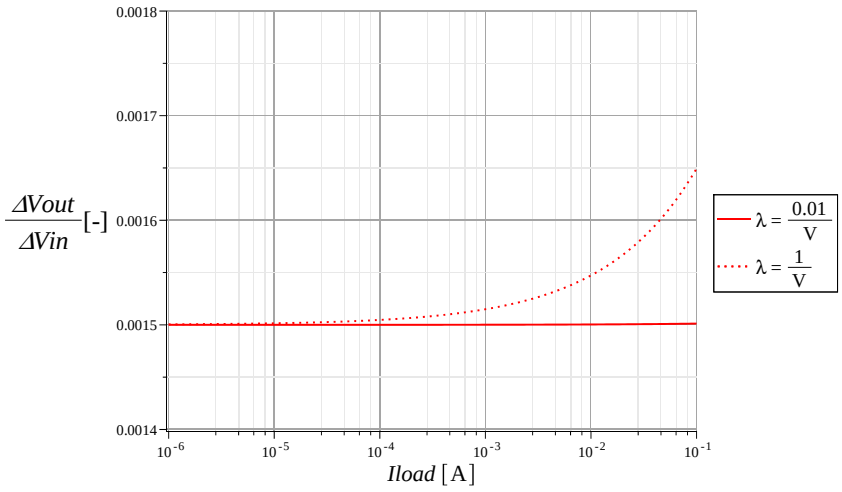


Figure 2.8: Line regulation of series PMOS regulators

be said, that the value of $\lambda = \frac{10000}{V}$ is extremely exaggerated for illustrative purposes and in reality it is never that high.

The load regulation works practically the same in the both NMOS or PMOS cases. On the other hand, the line regulation differs. Firstly, the situation using the PMOS is described. Assuming the constant load current, when the input voltage, for instance, increases, the output voltage also increases before the regulator starts to react and then the opamp drives the gate of the PMOS. The source of the PMOS is connected to the input voltage so before the opamp starts to react, its V_{sg} gets higher with the increasing input voltage. Thus, the opamp has to drive the gate of the PMOS to higher potential to assure the desired output voltage. Mathematically the line regulation of the PMOS regulator can be expressed as:

$$v_{out} = \frac{v_{in} - v_{out}}{r_o} \cdot R_{load} + \left(v_{in} - v_{out} \cdot A_v \cdot \frac{R_2}{R_1 + R_2} \right) \cdot g_m \cdot R_{load}, \tag{2.24}$$

$$\frac{v_{out}}{v_{in}} = \frac{1 + g_m \cdot r_o}{1 + A_v \cdot \frac{R_2}{R_1 + R_2} \cdot g_m \cdot r_o + \frac{r_o}{R_{load}}}. \tag{2.25}$$

The situation when using the NMOS regulator is different. The NMOS power transistor is connected as a voltage follower. The input voltage is connected to the drain terminal of the transistor. Assuming an ideal transistor which acts as a voltage controlled current source - its output resistance is infinite and the current is fully independent on the V_{DS} voltage. If we consider V_{GS} constant as well, we can say that the change of the potential of the drain terminal does not change the current situation. Therefore, the output voltage does not change at all. Of course, in reality this situation is not true, the NMOS output resistance is not infinite. Thus, increasing the input voltage makes the transistor conduct a little bit more which appears as increasing output voltage. This situation is compensated by the gain of the error amplifier and the closed-loop similar way as in the PMOS case. The line regulation of the NMOS regulator case can be expressed as:

$$v_{out} = \frac{v_{in} - v_{out}}{r_o} \cdot R_{load} + \left(-v_{out} \cdot A_v \cdot \frac{R_2}{R_1 + R_2} - v_{out} \right) \cdot g_m \cdot R_{load}, \tag{2.26}$$

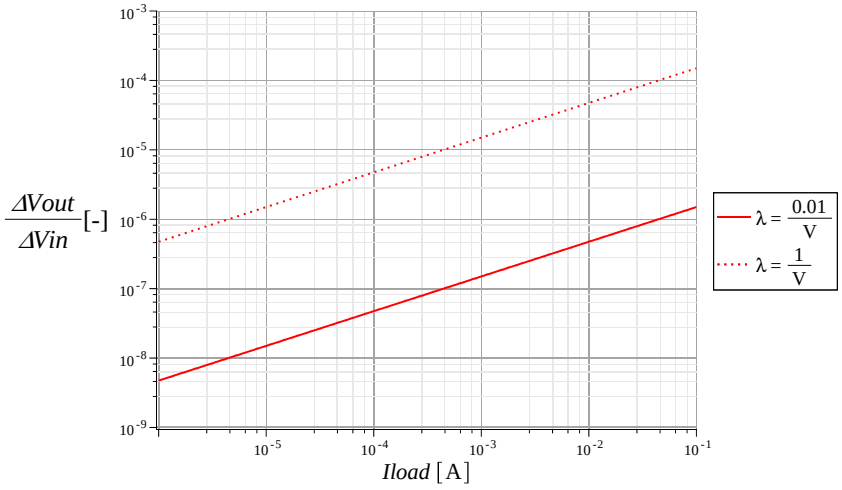


Figure 2.9: Line regulation of series NMOS regulators

$$\frac{v_{out}}{v_{in}} = \frac{1}{1 + \left(1 + A_v \cdot \frac{R_2}{R_1+R_2}\right) \cdot g_m \cdot r_o + \frac{r_o}{R_{load}}}. \tag{2.27}$$

The dependency of the line regulation on the load current I_{load} is plotted on figure 2.8 for the PMOS regulator and in figure 2.9 for the NMOS regulator for model values listed in table 2.2. It is obvious that the results are better for the NMOS regulator. In the NMOS case, the line regulation is improved approximately by the multiplication of the factor of $g_m \cdot r_o$. It can be also seen that it get worse with the load current. The transconductance g_m increases with the load current, but on the other hand, r_o decreases. The decreasing is more steep and dominant since $g_m \sim \sqrt{I_{load}}$, $r_o \sim \frac{1}{I_{load}}$. Also, with higher λ, the line regulation gets worse, the higher λ, the lower r_o. On the other hand, in the PMOS case, it is almost independent of the load current because the line regulation is mainly given by the gain of the operational amplifier and the ratio of the output voltage divider. For the same reason, the effect of λ is not that significant as in the NMOS case.

Needless to say, that the all results are mainly theoretical. In reality the results may differ substantially due to many other effects that have not been considered for the small signal models.

2.4 Efficiency: series vs parallel regulators

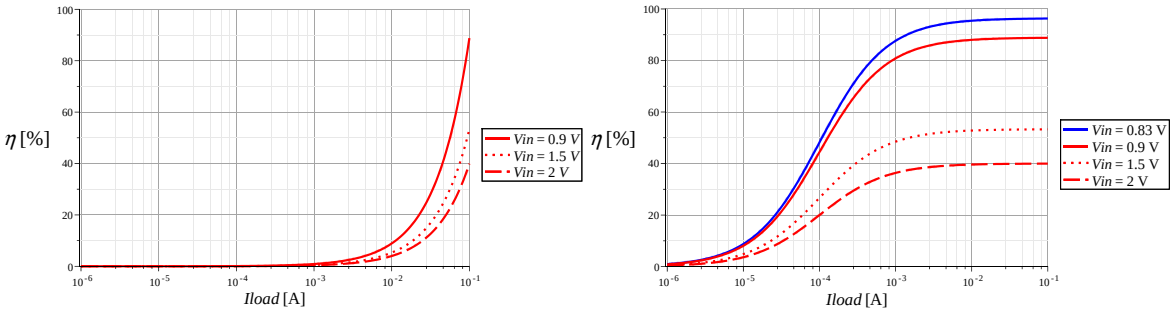
Efficiency could be the decision-making parameter when choosing between series and parallel regulators. Therefore, it is beneficial to explore the relation between efficiency and the load current. The definition of efficiency is stated in section 1.1.6. Firstly, for the efficiency of parallel regulators we can write:

$$\eta = \frac{P_{load}}{P_{load} + P_{internal}}, \tag{2.28}$$

$$\eta = \frac{1}{1 + \frac{(V_{in}-V_{out})\cdot\left(I_{q_{MIN}}+I_{load_{MAX}}\right)+V_{out}\cdot\left(\left(I_{q_{MIN}}+I_{load_{MAX}}\right)-I_{load}\right)}{I_{load}\cdot V_{out}}}. \tag{2.29}$$

Table 2.3: Parameters for computed efficiency

$I_{g_{MIN}}$	$I_{load_{MAX}}$	V_{in}	V_{out}
100 μA	1 A/100 μA	2/1.5/0.9/0.83 V	0.8 V



(a) (b)

Figure 2.10: Efficiency of parallel (a) and series (b) regulators

The dependency of efficiency on the load current for parameters listed in table 2.3 is depicted in figure 2.10a. As expected, the efficiency improves with the load current because at heavy loads, the quiescent current is minimal. The efficiency at light and medium loads is very poor. It is important to mention that the plotted results for the different input voltages are calculated for the constant current through the input resistor given as the sum of $I_{q_{MIN}} + I_{load_{MAX}}$. This means, that the value of the input resistor would have to change according the chosen input voltage. Although it is calculated with the fact that the quiescent current at maximal load current $I_{load_{MAX}}$ is equal to the minimal quiescent current $I_{q_{MIN}}$, in reality, the quiescent current at highest loads should be at least around 10% of $I_{load_{MAX}}$. Thus, the efficiency would be even worse, so the computed efficiency is only theoretical and practically not achievable. Finally, it is important to note that the efficiency of the parallel regulators depends on the value of $I_{load_{MAX}}$ - it determines the value of the input resistor. Thus, it determines the quiescent current as well - for higher desired maximal load currents, the quiescent current at lighter loads would be higher as well and the efficiency would be lower.

For comparison, the efficiency of series regulators is given by:

$$\eta = \frac{P_{load}}{P_{load} + P_{internal}} = \frac{1}{1 + \frac{(V_{in} - V_{out}) \cdot I_{load} + V_{in} \cdot I_q}{I_{load} \cdot V_{out}}}. \tag{2.30}$$

The dependency of efficiency on the load current for the same parameters is depicted in figure 2.10b. It is obvious that at light and medium loads it is much better than at parallel regulators. The main advantage of series regulators is the possibility of operation with very low drop-out voltage. This can lead up to achieving even higher efficiency at lighter loads than using DC-DC switching buck converters.

Chapter 3

Band-gap Voltage References

This chapter clears up fundamental theory of voltage references based on band-gap voltage principle. Main typologies and relation are discussed and shown.

3.1 Band-gap fundamentals

Voltage references are essential parts of almost any integrated circuit. The main role of voltage references is to provide nearly constant voltage regardless temperature variations and power supply fluctuations. The best performance is achieved by voltage references based on a so called band-gap principle. There is no parameter or value more precise than the band-gap voltage in integrated circuit design. The fundamental of these references is generation of a set of two voltages. One voltage has a positive temperature coefficient and is proportional to absolute temperature - *PTAT voltage*, the other voltage has a negative temperature coefficient and complements the PTAT voltage - so it is called complementary to absolute temperature - *CTAT voltage*. By suitable multiplication of this voltage by a constant and adding up with the PTAT voltage we can get nearly independent voltage on temperature variations.

The PTAT voltage can be generated using two PN junctions which are flown by two different current densities so that each forward-biased junction has different forward voltage V_F. It will be shown that the difference between these voltages ΔV_F is proportional to absolute temperature. The relation between current I_D flowing through a diode and its forward voltage V_F is given by Shockley equation [5]:

$$I_D = I_s \left(e^{\frac{V_F}{V_T}} - 1 \right) \approx I_s \left(e^{\frac{V_F}{V_T}} \right), \tag{3.1}$$

where I_s is a saturation current. So for the forward voltage we can write:

$$V_F = V_T \cdot ln \frac{I_D}{I_s}. \tag{3.2}$$

The difference voltage between two junctions biased with different current densities J_1 and J_2 is then given by:

$$\Delta V_F = V_T \cdot ln \frac{I_{D_1}}{I_{s_1}} - V_T \cdot ln \frac{I_{D_2}}{I_{s_2}} = V_T \cdot ln \left(\frac{I_{D_1} I_{s_2}}{I_{D_2} I_{s_1}} \right) = V_T \cdot ln \left(\frac{J_1}{J_2} \right), \tag{3.3}$$

$$\Delta V_F = V_T \cdot ln \left(\frac{I_{D_1} A_2}{I_{D_2} A_1} \right), \tag{3.4}$$

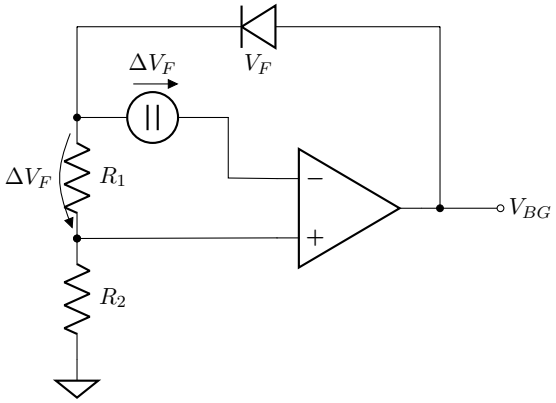


Figure 3.1: Band-gap reference principle model

this is valid because the saturation current I_s is proportional to the aperture A of the junction. Therefore, for making the voltage difference, the currents or the apertures of the junctions can be different. It can be seen that the ΔV_F is proportional to the thermal voltage $V_T = \frac{kT}{q}$. Thus, it is proportional to the absolute temperature - PTAT. ΔV_F has a positive temperature coefficient given by V_T:

$$\frac{\partial \Delta V_F}{\partial T} = \frac{k}{q} \cdot ln\left(\frac{I_{D_1} A_2}{I_{D_2} A_1}\right) = 0.086 \cdot ln\left(\frac{I_{D_1} A_2}{I_{D_2} A_1}\right) \text{ mV/°C}. \tag{3.5}$$

As the CTAT voltage, simple V_F can be used. Usually in many textbooks, the temperature dependence of V_F is approximately referred as:

$$\frac{\partial V_F}{\partial T} \approx -2 \text{ mV/°C}. \tag{3.6}$$

Thus, to ensure temperature Independence of the reference voltage, multiplying ΔV_F by a suitable constant M and adding up with V_F we get for the band-gap voltage:

$$V_{BG} = V_F + M \cdot \Delta V_F = V_F + M' \cdot V_T, \tag{3.7}$$

where $M = \frac{2}{0.086 \cdot ln\left(\frac{I_{D_1} A_2}{I_{D_2} A_1}\right)}$ and $M' = \frac{2}{0.086}$.

Till now, the linear temperature dependency of V_F has been considered, in reality, this is not true. From equation 3.2, this temperature dependency is given by temperature dependency of the saturation current. The relation of this dependency is very complex. In [5] there is a stated relation for V_F:

$$V_F = V_{G_0} - V_T \cdot (\beta - \alpha)\, ln(T) - ln(\gamma\sigma), \tag{3.8}$$

where V_{G_0} is the band-gap voltage of silicon at 0 K, α, σ represent circuit parameters and β, γ are device parameters independent on temperature. Similarly, as in equation 3.7, for the band-gap voltage we can write:

$$V_{BG} = V_{G_0} - V_T \cdot (\beta - \alpha)\, ln(T) + V_T(M'' + ln\,(\gamma\sigma)), \tag{3.9}$$

where M'' is another temperature independent constant. From this it can be seen that the band-gap voltage is not entirely independent on temperature. In fact, the dependency has a shape of a curve around the V_{G_0} which is approximately $V_{G_0} \approx 1.205$ V, thus the name

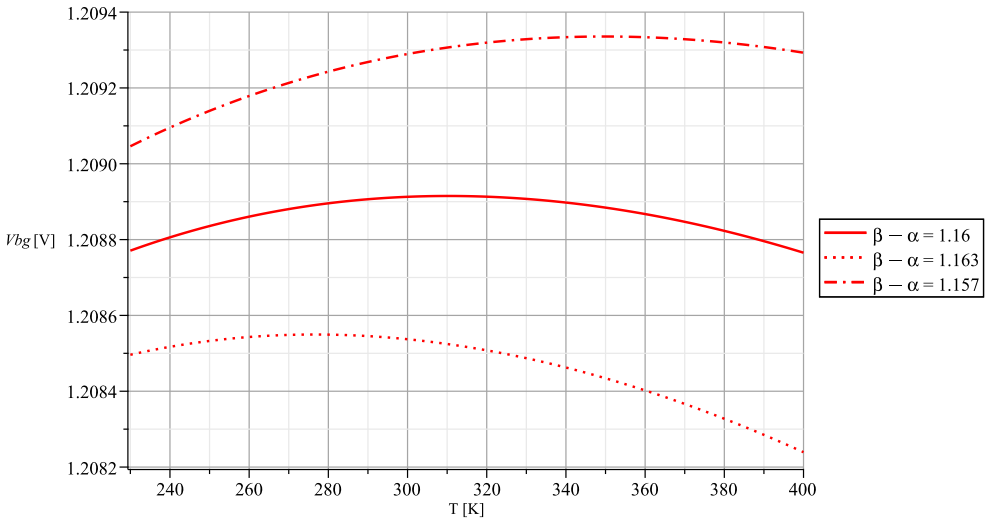


Figure 3.2: Bang-gap curve, different tuning

band-gap reference. The challenge is to find suitable values of the constants to tune the curve the way that its peak is at the desired temperature - usually around the room temperature of 300 K. The equation 3.9 is plotted in figure 3.2 for different values of constants $(\beta - \alpha)$ and for $(M'' + ln\,(\gamma\sigma)) = 1$. V_{G_0} is considered to be $V_{G_0} = 1.205$ V. It can be seen that each curve is tuned to have a peak at different temperature, so when designing a band-gap reference, the constant M in equation 3.7 has to be tuned somewhere around its stated value.

A principal implementation of a band-gap voltage reference is depicted in figure 3.1. The operational amplifier keeps voltage ΔV_F across the resistor R_1 which generates PTAT current $I_{PTAT} = \frac{\Delta V_F}{R_1}$. This current generates a voltage-drop across R_1. R_1 and R_2 make a voltage divider. This acts as a multiplication of ΔV_F by a constant. For the output voltage we can then write:

$$V_{out} = V_F + M\Delta V_F = V_F + \left(1 + \frac{R_2}{R_1}\right)\Delta V_F. \tag{3.10}$$

3.2 Brokaw band-gap reference

Practical realizations of band-gap references use bipolar junction transistors - BJTs instead of simple diodes. Firstly, let us sum up voltage-current relations of BJTs [5].

$$I_C = I_s\left(e^{\frac{V_{BE}}{V_T}} - 1\right) \tag{3.11}$$

$$I_B = \frac{I_C}{\beta_F} = \frac{I_s}{\beta_F}\left(e^{\frac{V_{BE}}{V_T}} - 1\right) \tag{3.12}$$

$$I_E = I_C + I_B = \frac{\beta_F + 1}{\beta_F}\left(e^{\frac{V_{BE}}{V_T}} - 1\right) \tag{3.13}$$

where I_s is a specific current similar to a saturation current as in a simple diode, so it can be seen that the relation between collector current I_C and base-emitter voltage V_{BE} is almost identical as I_D - V_D relation of a diode. One of the most used typologies of band-gap references is a topology invented by Paul Brokaw. The topology is sketched in figure 3.3.

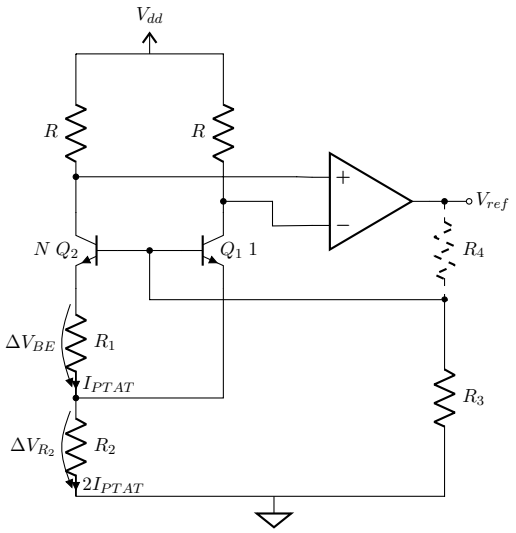


Figure 3.3: Brokaw band-gap reference

For the sake of simplicity of analysis, let us assume that the base current of the BJTs are negligible - $\beta \gg 1$ so that $I_C \approx I_E$. Although, making ΔV_{BE} is possibly either by different emitter currents or by different emitter areas, the second option is more accurate and more common. So let us assume emitter-base junction areas of the BJTs in a ratio of $1 : N$. The operational amplifier in a closed loop drives the base voltage of the BJTs so that the difference between the opamp input terminals is ideally zero and since resistors R are ideally identical, the currents flowing through the collectors of the BJTs are the same as well. Thanks to the different base-emitter areas, for the voltage drop across R_1 it holds:

$$V_{R_1} = \Delta V_{BE} = V_{BE_1} - V_{BE_2} = V_T \cdot ln\frac{I_{E1}}{I_{s_1}} - V_T \cdot ln\frac{I_{E2}}{I_{s_2}} = V_T \cdot ln\frac{I_{s_2}}{I_{s_1}} = V_T \cdot lnN, \quad (3.14)$$

this voltage drop across R_1 generates PTAT current:

$$I_{PTAT} = \frac{\Delta V_{BE}}{R_1} \approx I. \quad (3.15)$$

Since the currents through the BJTs are assumed to be identical, the current through R_2 is doubled and the voltage drop across R_2 is:

$$V_{R_2} = 2 \cdot I_{PTAT} \cdot R_2 = 2\frac{R_2}{R_1}\Delta V_{BE} = 2\frac{R_2}{R_1}V_T lnN. \quad (3.16)$$

Now the output voltage can be calculated. If resistor R_4 is shorted the output voltage is given as:

$$V_{ref} = V_{R_2} + V_{BE_1} = 2\frac{R_2}{R_1}V_T lnN + V_{BE_1}, \quad (3.17)$$

if we consider resistor R_4 which makes a voltage divider with R_3, the output voltage is then given by:

$$V'_{ref} = V_{ref} \cdot \frac{R_3}{R_3 + R_4}. \quad (3.18)$$

To achieve a zero temperature coefficient - assuming the linear temperature dependence of V_{BE} from equation 3.6, the ratio of R_1 and R_2 has to fulfill:

$$\frac{\partial V_{ref}}{\partial T} = 0 \rightarrow \frac{R_2}{R_1} = \frac{2}{2 \cdot 0.086 \cdot lnN} \quad (3.19)$$

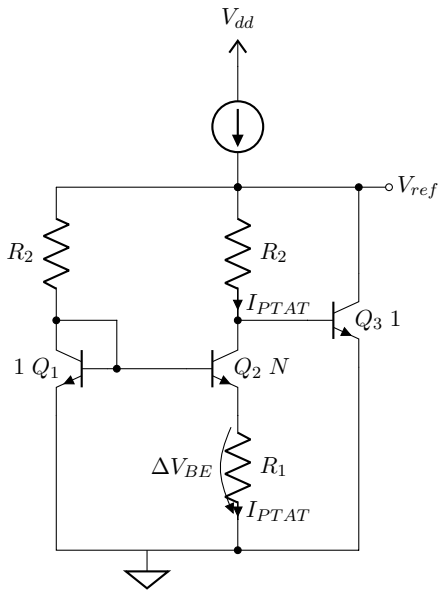


Figure 3.4: Widlar band-gap reference

3.3 Widlar band-gap reference

The first invented bad-gap reference was Bob Widlar's structure which is principally depicted in figure 3.4. Assuming again different emitter-base areas of the BJTs in a ratio of $1 : N$. The transistor Q_3 provides that the collector voltage of Q_2 is approximately equal to the collector voltage of Q_1. This is only valid if currents through Q_3 and Q_1 are identical - in reality the circuit will be more complex for assuring this. This depicted structure is only for principle describing. The collector resistors R_2 are set to be identical. Therefore, almost identical currents flow through each BJT. The voltage drop across R_1 is equal to ΔV_{BE}:

$$V_{R_1} = \Delta V_{BE} = V_{BE_1} - V_{BE_2} = V_T \cdot ln\frac{I_{E1}}{I_{s_1}} - V_T \cdot ln\frac{I_{E2}}{I_{s_2}} = V_T \cdot ln\frac{I_{s_2}}{I_{s_1}} = V_T \cdot lnN \tag{3.20}$$

this voltage drop across R_1 generates PTAT current:

$$I_{PTAT} = \frac{\Delta V_{BE}}{R_1}. \tag{3.21}$$

Again, assuming negligible base currents, I_{PTAT} makes a voltage drop across the collector resistor:

$$V_{R_2} = I_{PTAT} \cdot R_2 = \frac{R_2}{R_1} \cdot \Delta V_{BE}. \tag{3.22}$$

The output voltage is given by the sum of V_{BE_3} and V_{R_2}:

$$V_{ref} = V_{R_2} + V_{BE_3} = \frac{R_2}{R_1} V_T lnN + V_{BE_3}. \tag{3.23}$$

Now again, to achieve a zero temperature coefficient - again assuming the linear temperature dependence of V_{BE} from equation 3.6:

$$\frac{\partial V_{ref}}{\partial T} = 0 \rightarrow \frac{R_2}{R_1} = \frac{2}{0.086 \cdot lnN}. \tag{3.24}$$

Widlar band-gap references have a major advantage in terms of high PSRR. The whole core of the band-gap is self-biased from the band-gap voltage. This makes this structure very immune against power supply fluctuations. On the other hand, this structure is only suitable for technology processes where large values of β can be achieved. Otherwise there are errors in the reference output voltage, as it is discussed in section 3.5.1.

3.4 Summing-current low voltage band-gap reference

In many applications, a band-gap reference working with low supply voltage even around 1 V could be necessary. One of the possible ways how to achieve that is depicted in figure 3.5. This structure is based on summing CTAT and PTAT currents, not voltages as in previous cases. For analysis, again, negligible base currents are assumed. The opamp in a closed loop ensures that the difference voltage between its input terminals is almost zero. Thus, the voltage drop across resistor R_1 is given by ΔV_{BE}:

$$V_{R_1} = \Delta V_{BE} = V_{BE_1} - V_{BE_2} = V_T \cdot ln\frac{I_{E1}}{I_{s_1}} - V_T \cdot ln\frac{I_{E2}}{I_{s_2}} = V_T \cdot ln\frac{I_{s_2}}{I_{s_1}} = V_T \cdot lnN, \quad (3.25)$$

this voltage drop across R_1 generates PTAT current:

$$I_{PTAT} = \frac{\Delta V_{BE}}{R_1} = \frac{V_T lnN}{R_1}. \quad (3.26)$$

CTAT current is generated by the voltage drop across resistor R_2, the voltage drop across this resistor is equal to V_{EB_1}:

$$I_{CTAT} = \frac{V_{EB_1}}{R_2}. \quad (3.27)$$

I_{PTAT} and I_{CTAT} are then summed up at node 1. This current is then mirrored by transistors T_2 and T_3 so that the output voltage is given by the voltage drop caused by the the sum of the currents flowing through resistor R_4:

$$V_{ref} = R_4 \left(I_{PTAT} + I_{CTAT}\right) = R_4 \left(\frac{V_T lnN}{R_1} + \frac{V_{EB_1}}{R_2}\right). \quad (3.28)$$

Finally, for a zero temperature coefficient - again assuming the linear temperature dependence of V_{BE} from equation 3.6:

$$\frac{\partial V_{ref}}{\partial T} = 0 \rightarrow \frac{R_2}{R_1} = \frac{2}{0.086 \cdot lnN}. \quad (3.29)$$

For a proper operation, this type of band-gap reference theoretically needs minimal operating voltage only given by the sum of V_{BE} and $V_{DS_{min}}$, but that depends on design of the opamp which may probably need higher voltage. The low voltage is the major advantage of this structure. One the other hand, precision is its weakness. There are high requirements on precision of the current mirrors, so the overall precision is not that good as in Brokaw references, for instance.

3.5 Errors in reference voltage

Until now, ideal scenarios have been considered in calculation of the band-gap voltage. In real design, there will occur some errors. Among significant sources of errors we can put:

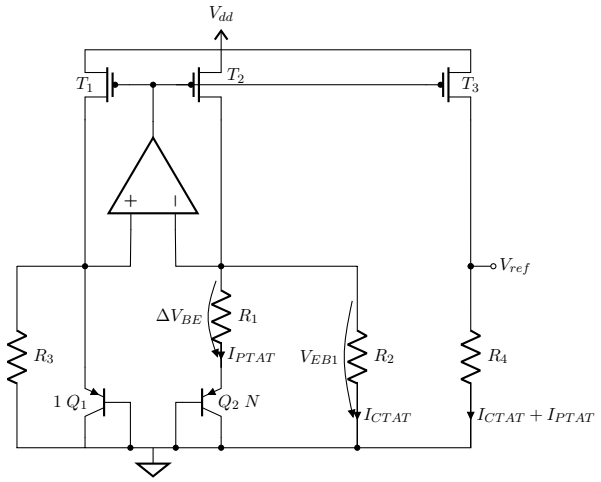


Figure 3.5: Summing current band-gap reference

- mismatch of resistors and transistors
- temperature coefficient of components, mainly resistors
- technology corners
- finite parameter β of BJTs - non-negligible base currents
- errors in current mirroring
- finite gain of used opamp and its offset

Some sources of errors cannot be avoided. On the other hand, effect of some sources of errors can be reduced by a proper design technique.

3.5.1 Finite β effects

Bipolar transistors in CMOS technology are made using parasitic structure. Either lateral or vertical BJT transistors can be used. In some technology processes, parameter β cannot achieve values large enough to neglect its effects on precision of the reference voltage. To explore effects of β, large-signal model of BJT can be used. Such a model is depicted in figure 3.6. This model considers "parasitics" base resistance of BJT. It can be seen that the real voltage V'_{BE} does not equal to the ideal voltage V_{BE} given by the standard exponential relation in equation 3.2, to get real V'_{BE}, V_{BE} has to be summed up with the voltage drop across the base resistance flown by base current I_B:

$$V'_{BE} = V_{BE} + R_B \frac{I_C}{\beta}. \tag{3.30}$$

Another problem is the fact that V_{BE} in this equation is not simply given by equation 3.2 because the collector current has to be summed up with the base current as well to know the emitter current which gives V_{BE} :

$$V_{BE} = V_T \cdot ln\left(\frac{I_C + I_B}{I_S}\right) = V_T \cdot ln\left(\frac{I_C\left(1 + \frac{1}{\beta}\right)}{I_S}\right). \tag{3.31}$$

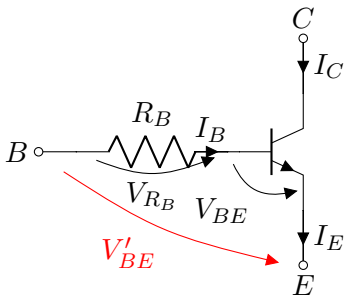


Figure 3.6: BJT model for finite β effects

There is a question suggesting itself: Is there the same influence of finite β in the mentioned typologies of band-gap references? Let us compare the Brokaw and Widlar topology and use figure 3.7a. Firstly, let us take a look on the Brokaw topology. For the whole generation of the band-gap voltage, only emitter currents are used. The base currents have impact on generation of both V_{BE} and ΔV_{BE} so the influence of the base currents is suppressed. If there is a need for higher output voltage, then resistor R_4 is added. In this situation, the voltage divider consisting of R_3 and R_4 is loaded by the base currents. In this situation, the output voltage gets slightly higher because of the loaded divider. Especially in low quiescent current design, the voltage divider needs to consume a low quiescent current and then the base currents are not negligible and cause errors in the output reference voltage. This errors can be theoretically suppressed by adding resistor R_5 between the base terminals of the BJTs. This resistor is supposed to be equal to [13]:

$$R_5 = \frac{R_1}{R_2} \cdot \frac{R_3 R_4}{R_3 + R_4}. \tag{3.32}$$

This resistor flown by the base current I_{B_2} makes a little voltage drop. This voltage drop influences generation of ΔV_{BE}. Thus, ΔV_{BE} is now slightly lower and this compensates the effect of the loaded voltage divider. But there is a drawback of this solution and it is increased level of output noise because R_5 value can be in the order of MΩ. So the real usage of R_5 is debatable.

On the other hand, in Widlar topology depicted in figure 3.7b. V_{BE} is generated by the emitter currents, ΔV_{BE} generates the PTAT current, but the PTAT voltage is not generated by the emitter current but by the collector current which generates an error. From this, it is evident that Widlar references are suitable for using in technologies where β is sufficiently large enough to fulfill accuracy requirements.

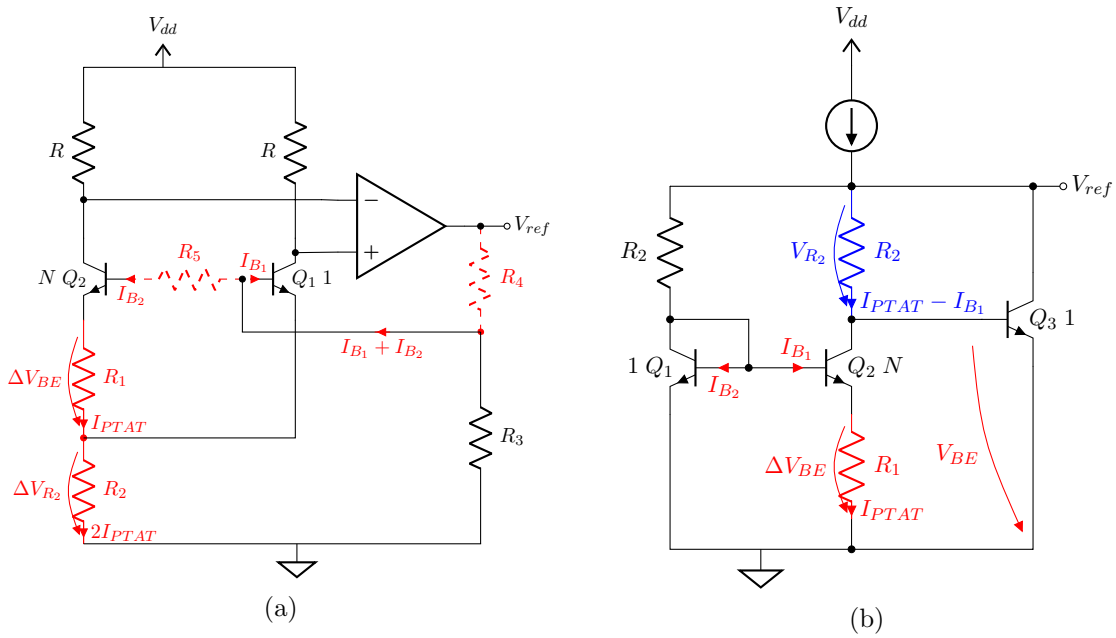


Figure 3.7: Effect of base currents for Brokaw (a) and Widlar (b) band-gap reference

Chapter 4

Design of LDO Regulator

This chapter deals with design of the LDO ultra-low quiescent current regulator. Design of each used block is explained. Difficulties and trade-offs during the design are mentioned as well.

4.1 Specification of requirements

The regulator has been designed for purposes of STMicroelectronics company in their technology BCD8. All standard requirements for fabrication have had to be met. The proposed regulator can be put into an ultra-low quiescent current category since the emphasis has been put on minimizing the quiescent current at no load condition. Also very high efficiency can be achieved since the drop-out voltage is extraordinary low for the ultra-low quiescent current category. This makes the LDO ideal for battery powered applications.

The required specification is following:

- quiescent current $I_q = 1$ μA @ $I_{load} = 0$ mA (without a band-gap reference)
- load current $I_{load} = 0$ to 300 mA
- drop-out voltage $V_{drop} < 150$ mV @ $I_{load} = 300$ mA
- N-channel power transistor
- separate bias voltage from a battery and input voltage: $V_{bias} = 2.75$ to 5.5 V, V_{in} max 5.5 V
- output voltage: $V_{out} = 0.8$ to 4 V, by 50 mV step
- accuracy of the output voltage: $\pm 2\%$ at 27 °C
- power supply rejection ratio: PSRR = around 70 dB @ 100 Hz
- optimal output capacitor $C_{out} = 1$ μF, stable with $C_{out} = 500$ nF to 10 μF with the ESR in the range of 5 to 500 mΩ
- temperature range: -40 to 125 °C
- technology BCD8
- design also includes:

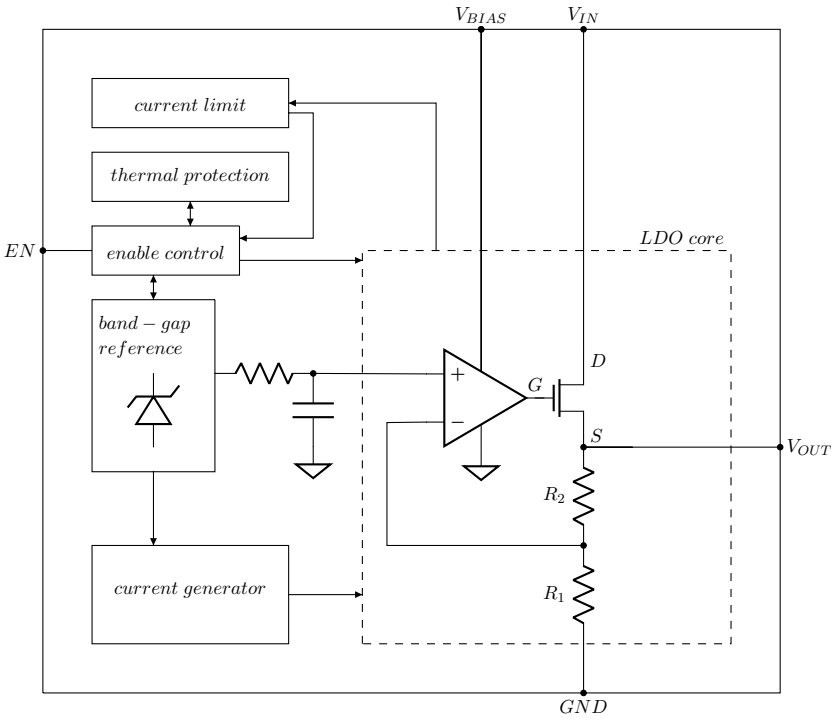


Figure 4.1: Simplified block schematic of the designed LDO regulator

- band-gap voltage reference with the maximal quiescent current $I_q = 500$ nA
- current generator
- enable control
- thermal protection
- current limiting circuit

The need for the minimal output voltage $V_{out} = 0.8$ V and the capability of very low drop-out voltage even at this output voltage level determines the usage of NMOS power transistor. Using PMOS transistor and fulfilling these requirements is not possible since in this case, the minimal biasing voltage would need to be less than 1 V (not considering separate input and biasing voltage) and there may not be enough headroom for V_{sg} of the power PMOS. It has been decided to use separate biasing voltage V_{bias} and the input signal V_{in}. The simplified block schematic of the proposed regulator is depicted in figure 4.1.

4.2 Obstacles caused by the requirements

Before doing firsts steps and sketching first attempts of the opamp structure, several things have to be considered. From the specification of requirements it can be seen that the main determinant in the design is the low quiescent current and the very low drop-out voltage. These two factors bring difficulties with assuring stability under all possible conditions and achieving sufficient transient responses.

The low quiescent current makes difficulties mainly at following parameters and scenarios:

- Less current branches - lower "freedom" of design.

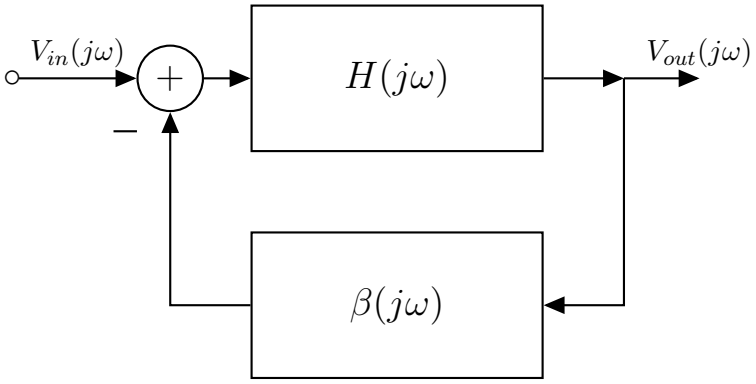


Figure 4.2: Block schematic of a feedback system

- Requirement for the low drop-out means using a large power transistor which is difficult to drive with the low current.
- The large power transistor together with low current makes assuring stability at every condition very difficult.
- For high PSRR, large bandwidth of the open-loop transfer function of the LDO is needed, to achieve the large bandwidth, the large current is needed as well.
- For sufficient transient response, the large bandwidth is needed and capacitances have to be charged/discharged fast enough - the current is needed again.

4.3 Stability

Assuring stability under any condition is necessary. Unfortunately, assuring stability is one of the most difficult things in voltage regulators design, mainly due the fact that there are many factors that can change and vary a lot and the regulator has to show sufficient universality and immunity against these variables.

For examining stability, a closed-loop feedback system can be generally depicted as in figure 4.2, where $H(\omega)$ is a transfer function of a system without a feedback and $\beta(\omega)$ represents a transfer function of a feedback. An overall transfer function of a closed-loop system, also called closed-loop gain, can be expresses as:

$$H_{CL}(j\omega) = \frac{V_{out}(j\omega)}{V_{in}(j\omega)} = \frac{H(j\omega)}{1 + \beta(j\omega)H(j\omega)}, \tag{4.1}$$

from this expression it can be seen that for the case where:

$$|H(j\omega)\beta(j\omega)| = 1, \tag{4.2}$$

$$arg(H(j\omega)\beta(j\omega)) = 180^{\circ}, \tag{4.3}$$

the denominator of equation 4.1 becomes equal zero and the closed-loop gain becomes equal infinity, which means that the system is not stable and starts to oscillate. Investigation of

Table 4.1: Normalized transfer function of zeros and poles

LHP zero	RHP zero
$H(j\omega) = \left(\frac{j\omega}{-z} + 1\right); z < 0$	$H(j\omega) = \left(\frac{j\omega}{-z} + 1\right); z > 0$
LHP pole	RHP pole
$H(j\omega) = \frac{1}{\left(\frac{j\omega}{-p}+1\right)}; p < 0$	$H(j\omega) = \frac{1}{\left(\frac{j\omega}{-p}+1\right)}; p > 0$

stability is provided by exploring $H_{OL}(j\omega) = H(j\omega)\beta(j\omega)$, this expression is called open-loop transfer function or open-loop gain. The open loop transfer function can be also express as:

$$H_{OL}(j\omega) = A\frac{(j\omega - z_1)(j\omega - z_2)...(j\omega - z_n)}{(j\omega - p_1)(j\omega - p_2)...(j\omega - p_m)}, \tag{4.4}$$

or in the normalized form:

$$H_{OL}(j\omega) = G \cdot \frac{(\frac{j\omega}{-z_1} + 1)(\frac{j\omega}{-z_2} + 1)...(\frac{j\omega}{-z_n} + 1)}{(\frac{j\omega}{-p_1} + 1)(\frac{j\omega}{-p_2} + 1)...(\frac{j\omega}{-p_m} + 1)}, \tag{4.5}$$

where z represents the frequency for which the numerator of the function becomes zero - this frequency is called *zero*, p represents the frequency for which the denominator becomes zero - this frequency is called *pole*, finally A represents a constant, G is a constant as well and it acts as a DC gain of the transfer function.

It is obvious that zeros and poles have a significant impact on the phase and the gain of the transfer function. Both zeroes and poles can be either negative - left half plane (it is located in the left side of a complex plane) - LHP zeros/poles or positive - right half plane - RHP zeros, poles. If a system has a RHP pole, it is not stable. But in analog circuit design, RHP poles are not a concern. They are a concern when it comes to discrete systems. The transfer functions of LHP/RHP zeros and LHP/RHP poles are summed up in table 4.1 and are plotted in figure 4.3. It can be seen that each pole or zero adds up to 90 degrees positive/negative shift of phase and the decrease/increase of magnitude is 20 dB/decade using Bode approximation.

4.3.1 Stability analysis

Stability of a closed-loop system can be analyzed using various methods. The most used method in analog circuits design is by examining its open-loop transfer function. The system is stable if the phase shift of the open-loop transfer function is less than -180 degrees at the frequency at which the magnitude of open-loop transfer function is unity (or equals zero in decibels). This method is called *Bode plots*. For further reading, it it suitable to define frequently used parameters - *phase margin*, *gain margin* and *unity gain frequency*.

Unity gain frequency (UGF) is the frequency at which the magnitude of the open-loop transfer function reaches unity gain (or zero level when using decibels):

$$UGF = f_{UGF}\ [Hz]\ where\ |H_{OL}(j2\pi f_{UGF})| = 1\ [-], \tag{4.6}$$

Phase margin is given by the difference of the value of the phase of the open-loop transfer function at the unity gain frequency and 180 degree shift:

$$Phase\ margin = \angle H_{OL}(j2\pi f_{UGF}) - (-180)\ [^\circ], \tag{4.7}$$

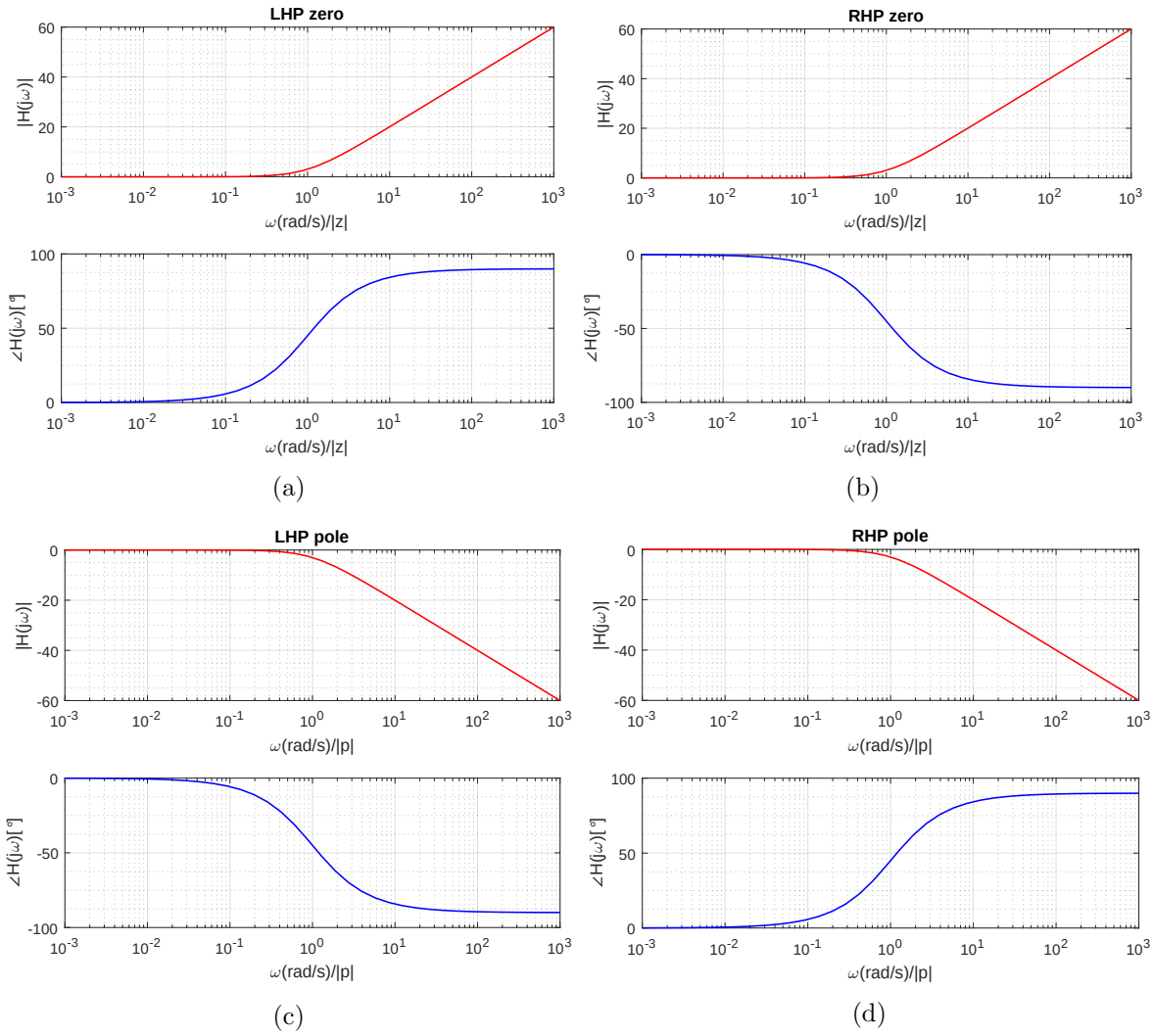


Figure 4.3: Normalized transfer functions of LHP zero (a), RHP zero (b), LHP pole (c) and RHP pole (d)

Gain margin is defined as the difference of the magnitude of the open-loop transfer function from the zero magnitude (in decibels) at the frequency f_{180} at which the phase crosses 180 degree shift:

$$Gain\ margin = -\left|H_{OL}(j2\pi f_{180})\ [dB]\right| \ \ where\ \angle\left|H_{OL}(j2\pi f_{180})\right| = -180°. \tag{4.8}$$

When plotting open-loop transfer functions, several shapes of waveforms are possible to be seen. For better illustration, how the position of poles and zeros affects stability of a closed-loop system, Matlab script for plotting open-loop transfer functions has been created. In figure 4.4 several different open-loop transfer functions of systems are depicted. Chosen positions of zeros and poles for each system and their stability parameters are summed up in table 4.2. Let us now examine each system:

- ***System 1*** is a scenario where there are three poles located very closed to each other. These poles can add up the phase shift up to 270 degrees. The phase reaches 180 degree shift before the magnitude can reach unity (zero in decibels). This system, of

Table 4.2: Position of poles and zeros for various systems and their stability parameters

$G = 1000$ [−]	p_1 [Hz]	p_2 [Hz]	p_3 [Hz]	p_4 [Hz]	z_1 [Hz]	z_2 [Hz]
System 1	-70	-70	-70	$-5 \cdot 10^3$	$-5 \cdot 10^2$	-10^3
System 2	-10^2	-10^3	-10^8	-10^9	$-8 \cdot 10^3$	$-\infty$
System 3	-10^2	-10^6	-10^8	-10^9	$-\infty$	$-\infty$
System 4	-10^3	-10^4	$-5 \cdot 10^5$	-10^7	$-\infty$	$-\infty$

	UGF [Hz]	PM [°]	GM [dB]	stability
System 1	$1.1 \cdot 10^3$	22	–	*yes*
System 2	$1.4 \cdot 10^4$	75	98	*yes*
System 3	$9.95 \cdot 10^4$	85	60	*yes*
System 4	$9.87 \cdot 10^4$	-5	–	*no*

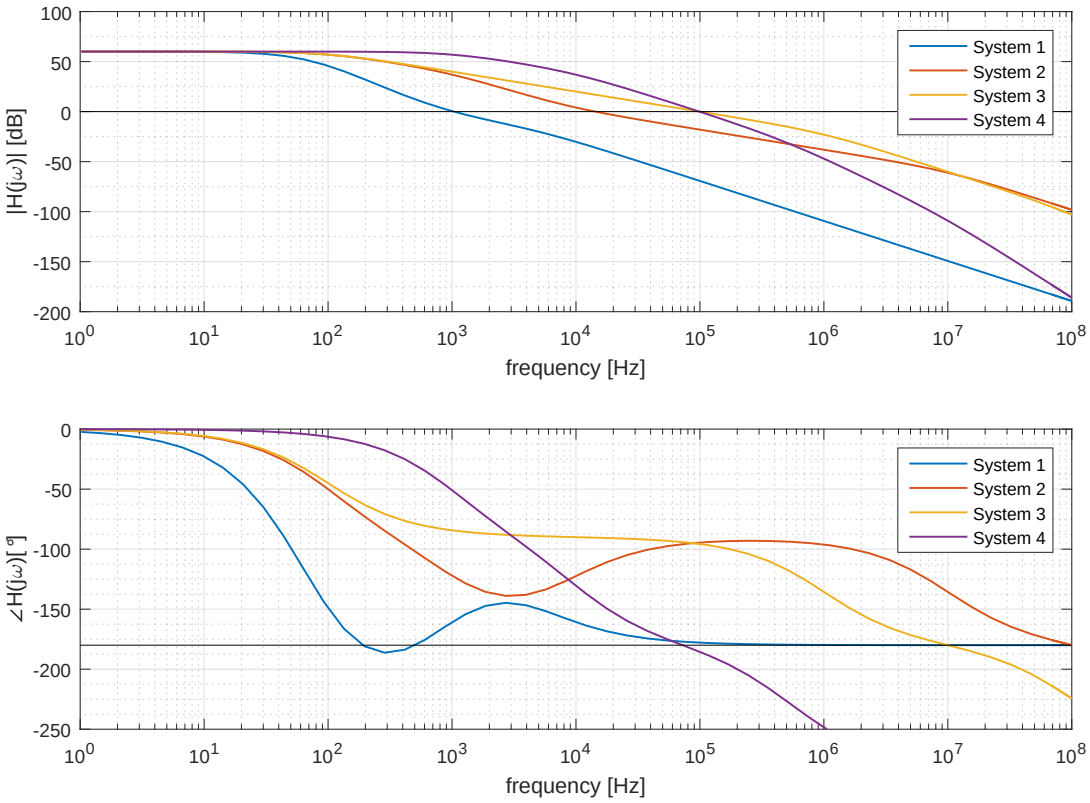


Figure 4.4: Open-loop transfer functions for various systems

course, would be unstable. But there are also located two zeros closed to each other at frequencies slightly higher than the frequency of the poles. The zeros compensate the steep phase shift of the poles so that the phase shift is less than 180 degrees at UGF - phase margin is positive and the system is stable. This system is a good example that the system can be stable even if the phase reaches 180 degree shift before UGF if there are zeros to compensate the phase shift.

- ***System 2*** is a scenario where there are two poles at relatively close frequencies. If there were only these two pole the system may have been stable if the poles were at frequencies far enough from each other to allow the magnitude reach the unity gain before 180 degree shift. The phase margin would probably be not so large. But there are also other higher frequency poles that can add up another phase shift so the system can become unstable. To compensate this, there is one zero located at the frequency a little behind the second pole. This zero compensates the phase shift of second pole so that the phase margin is quite large and the system is stable. It is important to mention that the zero should not be located very close to the second pole: Because we want to compensate the phase shift of the second pole, so the first thought would probably be to locate the zero at the frequency of the second pole to fully compensate its phase shift. But on the other hand, we need this second pole to allow the magnitude to reach unity fast enough to eliminate the effect of the other high frequency poles, which could otherwise cause instability again. The same thing relates to *System 1* where fully compensation of the two poles by two zeros is not desired as well. In this *System 2* the first and second pole are located within the UGF, other poles are outside the UGF.

- ***System 3*** is a situation where there is no zero in the system. But the first and the second pole are very separated so that the magnitude can decrease and cross unity before the effect of the phase shift of the second pole and other higher frequency poles can have significant impact on the phase margin. In this system, only the first pole is located within the UGF.

- ***System 4*** is an example of an unstable system where two poles are located close to each other at frequencies close to other high frequency poles. There are no zeros in the system for compensation so that the phase shift reaches 180 degrees before the magnitude can decrease to unity and the system is then unstable.

Nyquist diagram represents another way how to analyze stability instead of using *Bode plots*. Such a diagram for the same systems is depicted in figure 4.5. Nyquist diagram uses plotting the open-loop transfer function in a complex plane. Axis x represents the real part of the transfer function and axis y is the imaginary part of the function. The waveform starts at the point at the real axis for which the frequency equals zero. At this point, there is no phase shift so the imaginary part is zero and this point represents the DC gain of the open-loop transfer function. Let us now increase the frequency. Until the significant phase shift starts to show, we are still very close to the starting point. When the frequency is approaching the first pole (or zero) the phase starts to change rapidly, so does the the imaginary part of the function and the starting point starts to move as well. The distance from the origin of coordinates to the actual position gives the magnitude of the open-loop transfer function and the phase of the function is given by the pitch of the actual point from the x axis. Increasing the frequency, the actual point is moving to the left side and is drawing an "ellipse like" curve. For frequency equal infinity, the actual point gets to the origin of coordinates.

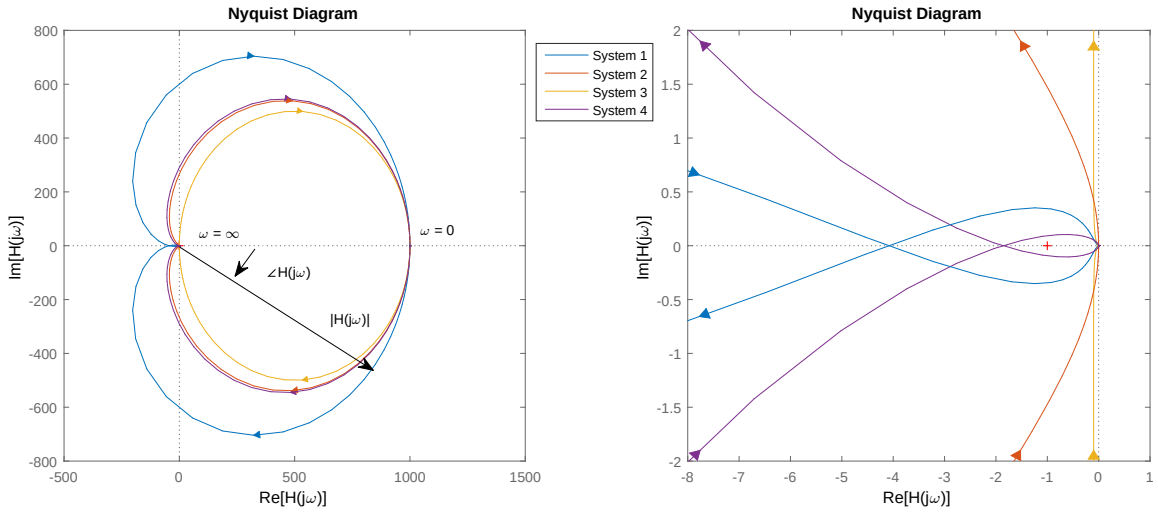


Figure 4.5: Nyquist diagram for various systems

The critical point in the Nyquist diagram is the point with coordinates (-1,0). If the actual point is at this point, the magnitude of the function is unity and the phase shift is 180 degrees. For a stable system, the waveform must not encircle the point from "the outside". This situation would mean that there is a frequency for which the phase shift is equal or higher than 180 degrees while the magnitude is unity - this is a case of *System 4*. An interesting case is for *System 1* where the waveform crosses x axis "outside" of the point (-1,0) - its phase shift crosses 180 degree while the magnitude is still higher then unity, and then the phase shift starts to decrease and then it encircles the point (-1,0) from the "inside" so the system is stable. This situation is fully equivalent to the representation using Bode plots. The upper half of the diagram is identical for analog systems - it is just a representation for negative frequencies. The drawback of Nyquist diagrams and the reason why its usage is not frequent in analog circuit design is the fact that there is no information about the frequency on the contrary of Bode plots.

4.3.2 Unity step response

Unity step response describes how a closed-loop system response to a unity step signal. This behavior in time domain is really important in the case of line and load transient response. It will be shown that the phase margin and the unity gain frequency have major impact on this behavior. The step response can be defined as convolution of the impulse response of the closed-loop system $h_{CL}(t)$ and the Heaviside step function $u(t)$:

$$y_{out}(t) = h_{CL}(t) * u(t), \tag{4.9}$$

it can be also computed using the transfer function of the closed-loop system $H_{CL}(j\omega)$:

$$y_{out}(t) = F^{-1}\left\{H_{CL}(j\omega)\right\} * u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H_{CL}(j\omega)e^{j\omega t}d\omega * u(t), \tag{4.10}$$

also we can use:

$$y_{out}(t) = F^{-1}\left\{H_{CL}(j\omega) \cdot F\left\{u(t)\right\}\right\}, \tag{4.11}$$

where F^{-1} represents Inverse Fourier transform and F is Fourier transform.

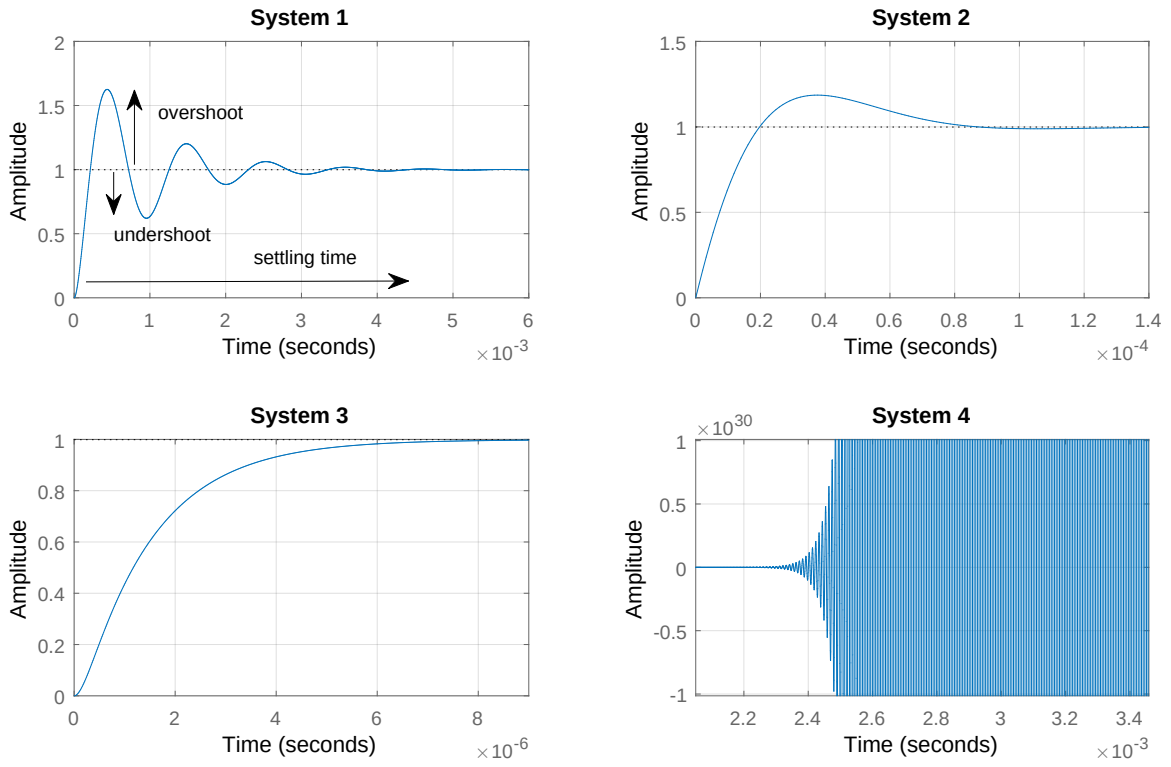


Figure 4.6: Unity step response for various systems

Step response for the four systems being analyzed is plotted in figure 4.6. It is seen that there are significant differences in each characteristic:

- ***System 1*** shows largest ringing and slowest settling time. Ringing is caused by relatively small phase margin equals $PM_1 = 22°$. Settling time is mainly determined by UGF. This system has the lowest UGF, thus the longest settling time.
- ***System 2*** shows substantial better step response behavior, only one small overshoot occurs - the phase margin is quite large - $PM_2 = 75°$ and the settling time is shorter due to the larger UGF.
- ***System 3*** shows the best unity step response - there is neither overshoot nor undershoot nor ringing due to the highest phase margin $PM_3 = 85°$ and its UGF is highest as well.
- ***System 4*** is not stable, so its unity step response is unbounded.

4.4 Poles and zeros of LDO regulator

For assuring good stability and designing good frequency compensation it is important to know where the most significant poles and zeros are located in the LDO structure. For this purpose, the simplified small signal model of the typical NMOS LDO structure for determining its open-loop transfer function is shown in figure 4.7. This model consists of three main blocks - differential amplifier stage, driver of power the transistor stage which is considered to be based on a voltage follower and finally the power NMOS transistor. Now

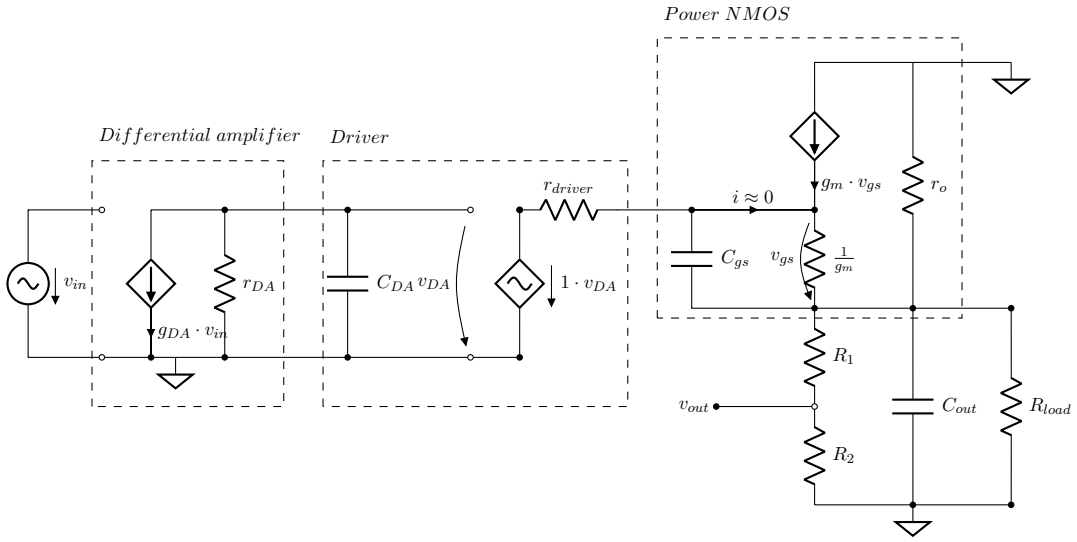


Figure 4.7: Simplified block schematic of the designed LDO regulator for determining open-loop gain

we can determine the poles. The first significant poles is given by the output resistance of the differential amplifier r_{DA} and the input capacitance of the driver C_{DA}:

$$p_1 = -\frac{1}{2\pi \cdot r_{DA} C_{DA}} \qquad \text{[Hz]}. \tag{4.12}$$

Another poles is located on the interface of the driver and the NMOS - it is given by the output resistance of the driver r_{driver} and the input capacitance of the power NMOS C_{gs}:

$$p_2 = -\frac{1}{2\pi \cdot r_{driver} C_{gs}} \qquad \text{[Hz]}. \tag{4.13}$$

Another pole of the LDO is located at its output, it is given by the output capacitor C_{out} and the resistance r_{out} seen at the node of the capacitor which is given by:

$$r_{out} = \frac{1}{g_m} \parallel r_o \parallel (R_1 + R_2) \parallel R_{load} \approx \frac{1}{g_m}, \tag{4.14}$$

this simplifying can be done considering that the resistors of the feedback divider are quite large, usually in a order of MΩ, the output resistance r_o of the NMOS is large as well and the transconductance g_m of the NMOS is high enough due to its high $\frac{W}{L}$ ratio, g_m then practically determines the r_{out}. Finally, for the output pole we can write:

$$p_3 = -\frac{1}{2\pi \cdot r_{out} C_{out}} \qquad \text{[Hz]}. \tag{4.15}$$

It has to be mentioned that, of course, in a real circuit, there will be more poles and zeros located at higher frequencies which may or may not have an impact on the phase margin. If the most significant poles are known, then the open-loop transfer function can be expressed as:

$$H_{OL}(j\omega) = \frac{\frac{g_{DA} r_{DA} R_2}{R_1 + R_2}}{(j\omega r_{DA} C_{DA} + 1)(j\omega r_{driver} C_{gs} + 1)(j\omega r_{out} C_{out} + 1)}, \tag{4.16}$$

where g_{DA} represents the transconductance of the differential amplifier.

4.4.1 Problems and complications

Since the open loop transfer function shows three significant poles, the system can be unstable. It is worth mentioning that the output pole is determined by $\frac{1}{g_m}$ so it is dependent on the load current, with higher loads, it is pushed to higher frequencies. On the other hand, it can be located at very low frequencies when light or no loads. If the LDO is adaptively biased - its quiescent current is dependent on the load current, the first pole p_1 and the second pole p_2 can be dependent on the load as well. There is not the only one strategy how to compensate this system. Theoretical options are:

- The first, really theoretical, option is to make the driver pole p_2 pushed to as high frequency as possible (pushed outside UGF), p_1 located at sufficiently low frequencies, to be separated from the output pole p_3, so that the output pole can have less impact on the phase shift. Also p_1 allows the magnitude of the open-loop transfer function reach unity before the phase reaches 180 degree shift, thus it ensures the sufficient phase margin - similarly as in *System 3* above. This is sometimes called "pole splitting". This is almost impossible in real design since there are lot of variable conditions, especially dependency of the poles on the load. Separation of the poles can occur especially at heavy loads where the output pole is pushed to highest frequencies but at light loads this situation may not occur.

- Another option is to have the pole p_1 located at low frequencies, the driver pole p_2 pushed to as high frequency as possible again and the effect of the output pole p_3 compensated by a suitably inserted zero into the system - this is described in similar scenario of *System 2* above. Again, the problem with the dependence on the load is still here. Therefore, some kind of technique of the "output pole tracking" may be needed especially for the application where the load can vary in a large range.

- There is also theoretical possibility of making the output pole p_3 located at lowest frequencies and pushing the pole p_1 to higher frequencies - to have similar situation as in the first option but vice versa. This situation may be suitable mainly for PMOS power transistor LDOs since their output pole is located at lower frequencies than in the NMOS case because the output resistance which determines the output pole is given approximately by the output resistance r_{out} of the PMOS power transistor, on contrary of g_m in case of NMOS LDOs. This scenario may work at lighter loads. If the application requires a large range of allowed loads, at high loads it would not be possible to make pole p_1 located at higher frequencies than the output pole. These poles can even switch positions which brings another complication. Also pushing pole p_1 to higher frequencies leads to lowering the output resistance r_{DA} of the differential amplifier which means lowering the overall gain which could be unacceptable.

These option are mainly theoretical, in real design, other higher frequencies poles can affect stability as well. Also requirements for universality and robustness of the LDO are high, for instance the need for allowed value of the output capacitor can be typically from less then 1 μF to tens of μF which means variability of the output pole, so assuring stability under every condition can be a really tough task and more complex approaches with compensation such as inserting more zeros into the system etc. might be necessary.

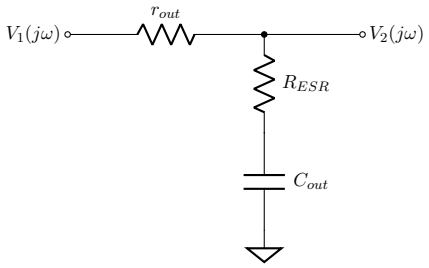


Figure 4.8: Compensation using ESR of the output capacitor

4.5 Compensation using ESR of the output capacitor

One possibility of inserting a zero into the system is simply using the equivalent series resistance - ESR of the output capacitor, the circuit for locating the frequency of the zero is depicted in figure 4.8. The transfer function of this circuit is following:

$$H(j\omega) = \frac{V_1(j\omega)}{V_2(j\omega)} = \frac{1 + j\omega C_{out} R_{ESR}}{1 + j\omega C_{out}(r_{out} + R_{ESR})}, \tag{4.17}$$

it is obvious that the zero is located at:

$$z_{ESR} = -\frac{1}{2\pi \cdot C_{out} R_{ESR}} \qquad [\text{Hz}], \tag{4.18}$$

on the other hand the output pole is given by:

$$p_{out} = -\frac{1}{2\pi C_{out}(r_{out} + R_{ESR})} \approx -\frac{1}{2\pi C_{out} r_{out}} \qquad [\text{Hz}] \tag{4.19}$$

and since $R_{ESR} << r_{out}$, the output pole is practically at the same frequency as without considering ESR.

The usage of ESR compensation is mainly for the second scenario of compensation method mentioned above - compensating the phase shift of the output pole. The problem is to specify the value of ESR. Too low value of ESR may not have any significant effect because the zero would be on too high frequency. On the other hand, too high ESR can cause the zero be located at undesirably low frequencies - of course, it would fully compensate the effect of the phase shift of the output pole but on the other hand we need a certain phase shift of the output pole to allow the magnitude of the open-loop transfer function to decrease fast enough - without that the effect of other higher frequency poles would occur - especially the pole created at the output of the driver. So the ideal position of the zero is at frequencies slightly higher then the output pole, but not too much. Similar situation is described in *System 2* above. From this, it can be seen that the drawback is again the dependency of the output pole on the load. Thus, choosing the one value of the ESR which is suitable for every condition is nearly impossible.

The ESR compensation had important usage in the past days when large tantalum output capacitor with ESR in a range even of units of Ω were used. These days, there is a large need for universality and ceramic capacitors are mainly used and their ESR can reach very low values of units of $m\Omega$. Therefore, the compensation technique cannot fully depend on using the ESR compensation, it has only a supporting role. On the other hand, there is a possibility of using a special "ESR enhancement" circuit, such a circuit is used in the design

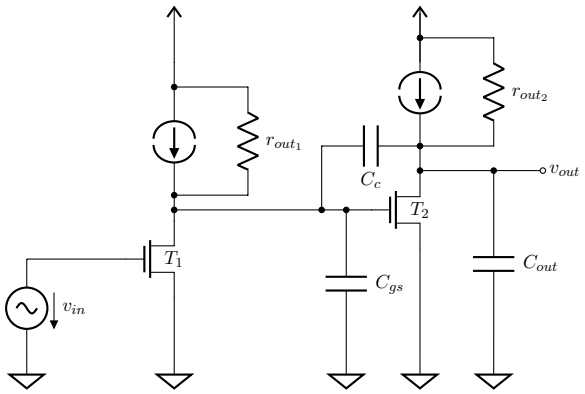


Figure 4.9: Miller compensation

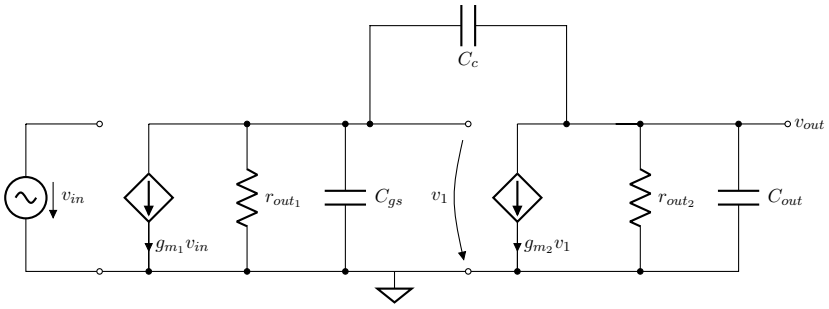


Figure 4.10: Small signal model of two stage amplifier with Miller compensation

of LDO and will be discussed latter. But still, the compensation strategy cannot be fully based on this type of compensation.

Finally, it is worth mentioning that larger value of ESR means worsening transient response behavior because the output capacitor then acts as a "weaker" voltage source. Also, the capacitor is charging more slowly - larger time constant, which is not desirable - fast charging is needed to compensate undershoots at the output.

4.6 Miller compensation

Miller compensation is a suitable method for "pole splitting" - separating two poles at close frequencies. The principle of Miller compensation can be well described on a two stage amplifier in figure 4.9. In this figure, C_{gs} represents the gate capacitance of the second stage NMOS T_2 and C_{out} is the load capacitance, finally, C_c represents a compensation capacitor. The effect of this compensation can be described by a small signal model of the two stage amplifier depicted in figure 4.10. Each stage is an inverting amplifier with DC gain of $G_1 = g_{m1}R_{out_1}$ and $G_2 = g_{m2}R_{out_2}$ respectively.

Without considering the compensation capacitor, there are two poles, first pole located at:

$$p_1 = -\frac{1}{2\pi R_{out_1} C_{gs}} \quad [\text{Hz}], \tag{4.20}$$

and the second pole:

$$p_2 = -\frac{1}{2\pi R_{out_2} C_{out}} \quad [\text{Hz}]. \tag{4.21}$$

Since r_{out1} and r_{out2} can have similar values, so do C_{gs} and C_{out}, the poles can be located at very close frequencies. The effect of Miller compensation is separating these two poles. The Overall transfer function of the small-signal model using C_c is quite complicated and can be found in many textbooks such as [14]:

$$H(j\omega) = \frac{g_{m_1} r_{out_1} r_{out_2}(g_{m_2} - j\omega C_c)}{1 + j\omega r_{out_1} r_{out_2} g_{m_2} C_c + (j\omega)^2 r_{out_1} r_{out_2}(C_{gs}C_{out} + (C_{gs} + C_{out})C_c)}, \tag{4.22}$$

from this function, the poles and zeros can be determined:

$$p_1 \approx -\frac{1}{2\pi g_{m_1} R_{out_1} R_{out_2} C_c} \quad [\text{Hz}], \tag{4.23}$$

and the second pole:

$$p_2 \approx -\frac{g_{m_2} C_c}{2\pi(C_{gs}C_{out} + (C_{gs} + C_{out})C_c)} \quad [\text{Hz}], \tag{4.24}$$

there is also one zero:

$$z = \frac{g_{m_2}}{2\pi C_c} \quad [\text{Hz}]. \tag{4.25}$$

The first pole p_1 is now moved to significantly lower frequencies - due to the Miller effect of inverting amplifier, the compensating capacitance is multiplied by the gain of the second stage and transformed to the input of the second stage. C_c is also transformed to the output of the second stage so the second pole p_2 is moved to slightly higher frequencies then it was originally without compensation - pole splitting occurs. But there is also an undesired RHP zero which can potentially make compensation ineffective or can even cause instability - RHP zeros makes magnitude increase while phase decrease - shown in figure 4.3b. The effect of the RHP zero can be compensated by adding a resistor R_c in series with C_c. It can be shown that the original zero is then replaced by:

$$\frac{1}{2\pi C_c(\frac{1}{g_{m_2}} - R_c)} \quad [\text{Hz}], \tag{4.26}$$

so by appropriate choosing of R_c we can either eliminate its effect by pushing it to infinite frequencies when:

$$R_c = \frac{1}{g_{m_2}}, \tag{4.27}$$

or we can even change the RHP zero to the desired LHP zero and reach another phase compensation when:

$$R_c > \frac{1}{g_{m_2}}. \tag{4.28}$$

Having been said, the pole at the output of differential amplifier and the output pole of the LDO can be located at very close frequencies or they can even switch position at some load conditions. This may cause Miller compensation be ineffective - in the scenario where the output pole is located at a lower frequency than the pole at the output of the differential amplifier - because of Miller compensation - the pole a the the output of differential amplifier would be pushed closer to frequencies where the output pole is located and the situation would get even worse. Another fact is that for Miller compensation we need an inverting amplifying stage. When it comes to NMOS based LDOs, the NMOS power transistor is connected as a voltage follower with a non inverting gain approximately equal unity. Although, it is possible to use Miller compensation in the NMOS LDO, but not in a "classical way".

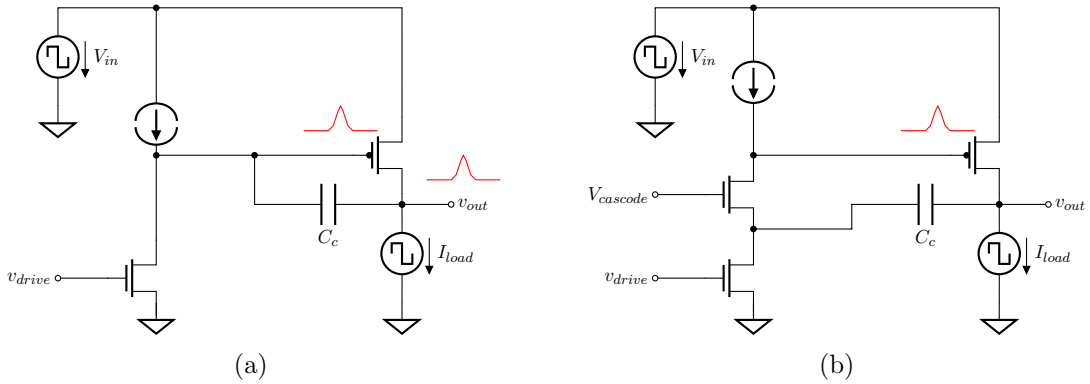


Figure 4.11: Principle of Miller compensation and transient response of LDO (a) and improved topology (b)

4.6.1 Miller compensation and transient response

It is good to mention that Miller compensation also affects the transient behavior of the LDO. A scenario with the line and the load transient response using Miller compensation can be explain using figure 4.11a. The PMOS power transistor is driven by transistor T_1.

Firstly, let us explain the line transient response (for that purpose, let us assume the restive load instead of the depicted current source). Assuming a steady state scenario, suddenly the input voltage, for instance, starts to increase. The gate voltage of the PMOS cannot react immediately, so its V_{gs} increases, which causes more current to flow to the load. Thus, the overshoot at the output voltage occurs. The Miller compensation capacitor directly couples the gate of the power transistor to the output. For frequencies high enough we can consider this path with the very low impedance. As the gate voltage starts to increase during the step change of the input voltage, this high frequency change is then transfer to the output and the overshoot gets bigger - Miller compensation causes worsening of the line transient response.

On the other hand, when assuming the load transient response, when there is a sudden increase of the load current, the gate voltage cannot be pushed immediately lower to assure higher V_{gs} so that the PMOS could conduct more. Hence, the undershoot at the output voltage happens. With the Miller compensation capacitor, the undershoot is directly transferred to the gate of the power PMOS and pushes the gate lower faster, which reduces the undershoot - Miller compensation improves the load transient response.

The disadvantage of worsening the line transient response can be reduced by cascoding driver transistor T_1 by transistor T_2 and connecting the compensation capacitor between these transistors as it is in figure 4.11b. When the gate voltage suddenly starts to increase during the step increase of the input voltage the source voltage of T_2 is not almost affected by the change of the input voltage, so the high frequency coupling of the gate of the power transistor is almost eliminated.

The cascode will not much affect improving the load transient response - when the undershoot at the output voltage is transferred to the source of cascoding transistor T_2, its V_{gs} increases so its drain is pushed lower - still improving the load transient response.

4.7 Parallel compensation

Another way how to compensate the LDO is using parallel compensation. The goal of this compensation is to make the pole at the output of the differential amplifier located at lower

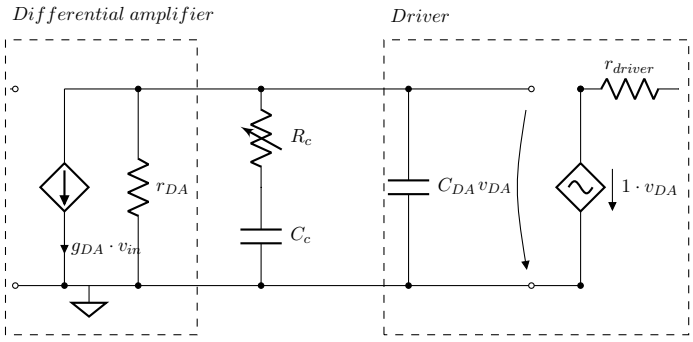


Figure 4.12: Parallel compensation

frequencies and the phase shift caused be the output pole of the LDO compensate by adding the zero into the system - described in the second situation above. The principle of the parallel compensation is depicted in figure 4.12. It is done by adding a capacitor in a series with a resistor between the differential amplifier and the driver. For proper function, it must hold $C_c >> C_{DA}$, then this compensation makes the pole at the output of the differential amplifier located at:

$$p_1 \approx -\frac{1}{2\pi C_c(r_{DA} + R_c)} \qquad [\text{Hz}], \tag{4.29}$$

also a zero is formed by:

$$z_1 = -\frac{1}{2\pi C_c R_c} \qquad [\text{Hz}]. \tag{4.30}$$

In this kind of compensation, choosing an appropriate value of the zero-making resistor R_2 may possibly be difficult. As it has been said, the position of this zero has to track the output pole of the LDO. In the application where a quite small range of loads is allowed, it might be sufficient to use only a fixed value resistor, but in most cases, some kind of the output pole tracking technique is needed. This leads to using an transistor working in linear region instead of a resistor. This is also suitable for another reason, at light loads, the zero has to be located at low frequencies which yields to having quite large resistance. Using a resistor with very large resistance would cost a large area on a chip. Thus, using the transistor is a better choice. This compensation technique is used for the proposed LDO, the design of active compensation - output pole tracking - will be discussed latter.

Disadvantage of this compensation technique is the fact that it leads to usage of relative large capacitance which may take a quite large area on a chip.

4.8 Feed forward compensation

Feed forward compensation represents another way how to insert a zero into the system. It consist of a capacitor connected in parallel with the output voltage divider as depicted in figure 4.13. The transfer function of this capacitor in a combination with the voltage divider can be expressed as:

$$H(j\omega) = \frac{V_1(j\omega)}{V_2(j\omega)} = \frac{R_2}{R_1 + R_2} \cdot \frac{1 + j\omega C_{ff} R_1}{1 + j\omega C_{ff} \frac{R_1 R_2}{R_1 + R_2}}, \tag{4.31}$$

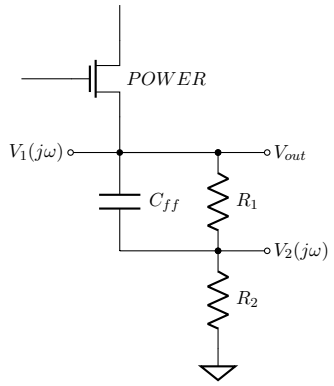


Figure 4.13: Feed forward compensation

it can be seen that the zero is located at:

$$z_{ff} = -\frac{1}{2\pi \cdot C_{ff} R_1} \qquad [\text{Hz}], \tag{4.32}$$

on the other hand the output pole is given by:

$$p_{ff} = -\frac{1}{2\pi C_{ff}(\frac{R_1 R_2}{R_1+R_2})} \qquad [\text{Hz}]. \tag{4.33}$$

Due to the fact that the pole is given by a parallel combination of the voltage divider resistor, the pole is located always at higher frequency than the zero so for a certain frequency range starting at the frequency of the zero, there is a phase shit compensation. For frequencies higher that the frequency of the pole, there is no effect of the feed forward compensation since the effect of the zero and the pole are eliminated.

A certain complication can occur when designing the output voltage variable LDO. In this situation, the ratio of the voltage divider resistors varies, so the zero and the pole vary as well. This may cause difficulties for a proper compensation under all possible conditions.

4.9 Structure of the designed LDO core

The design of the LDO core - the operational amplifier plus the power transistor, besides the band-gap voltage reference, is a bedrock of the whole design. The block schematic of the designed LDO core is depicted in figure 4.14. Each block will be later discussed separately. The most important thing is choosing a sufficient topology of a combination of the differential amplifier and the driver together with the power transistor. Each of these blocks cannot be designed separately without considering other. Thus, the first step is to figure out a simplified structure of the whole, which may be possible to meet the requirements with and after that to modify and adjust each block.

When figuring out the first simplified structure, besides others, main factors and determinants may be:

- Stability and possibility of compensation
- Quiescent current consumption, amount of branches
- Possibility of adaptive biasing
- PSRR considerations

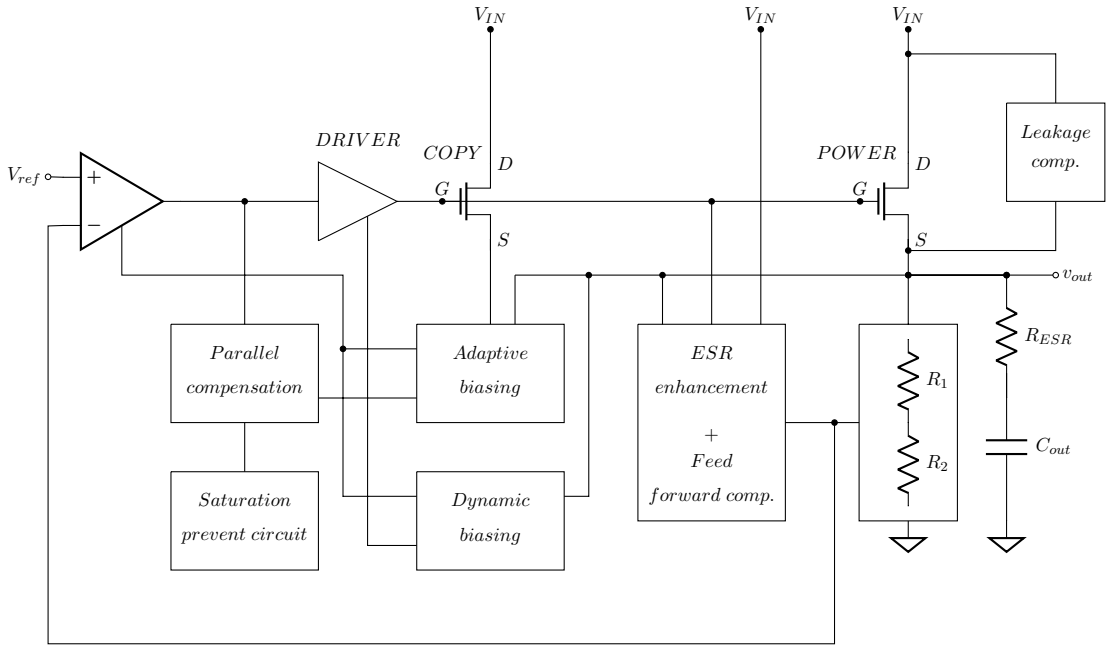


Figure 4.14: Simplified block schematic of the designed LDO core

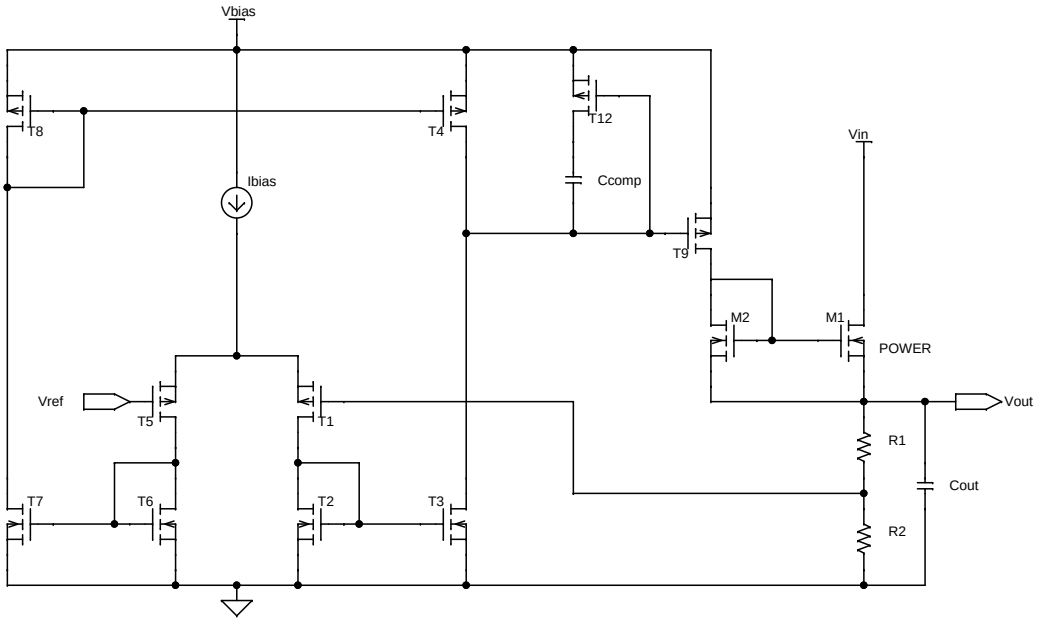


Figure 4.15: Simplified first approach structure

4.9.1 First approach structure

The firstly considered LDO structure is depicted in figure 4.15. The differential amplifier is consisted of so called symmetrical operational transconductance amplifier. The output of the differential amplifier is followed by the driver which consist of transistor T_9 and transistor M_2 which forms a current mirror with the power NMOS. For the frequency compensation, parallel compensation technique was intended.

This structure was firstly considered mainly because of the following features:

a. The power NMOS has to be driven by a low impedance source to move the pole which is formed by the gate capacitance of the power NMOS and the output resistance of the driver to frequencies high enough to have low impact on the phase margin. The feeding side of the current mirror is diode-connected so its output impedance is approximately given as:

$$r_{driver} \approx \frac{1}{g_{m_{M2}}}, \tag{4.34}$$

which can be considered as low impedance for sufficiently large $g_{m_{M2}}$. But the problem is that $g_{m_{M2}}$ is determined by the biasing current of M_2, which may bring complications in ultra-low quiescent current design.

b. It has been decided to use adaptive biasing. Since there is the requirement for ultra-low quiescent current at no load, without using the adaptive biasing, parameters like PSRR, line and load transients response and others would not be satisfying, because one of the main determinants in these parameters is UGF which is dependant on the biasing current of the differential amplifier. UGF of the differential amplifier is approximately given as:

$$UGF_{DA} \approx \frac{g_{m_{DA}}}{C_{out_{DA}}} \sim \frac{\sqrt{I_D}}{C_{out_{DA}}}, \tag{4.35}$$

where $g_{m_{DA}}$ is the transconductance of differential pair and $C_{out_{DA}}$ is the capacitance seen at the output node of differential amplifier and I_D is the biasing current of the input differential pair. Another factor when it comes to the transient response behavior is the slew rate. Before a system can react to a change, capacitances in the circuit need to be recharged, the speed of recharging is only determined by the current that can be sourced or sunken. Thus, the adaptive biasing where the quiescent current is a function of the load current is desired. The overall efficiency of the LDO may not be much affected since at higher loads, there is no need for extremely low quiescent current in terms of efficiency.

The desired feature of this structure is that the adaptive biasing is "automatically" ensured by the current-mirror driver. Adaptive biasing of the differential amplifier can be provided, again, by another simple current mirror.

c. For sufficient PSRR, the structure has to be designed the way that any capacitance coupling between high-side and low-side node has to be minimized. With this structure, it is possible to achieve quite high PSRR since there is no direct coupling between high and low side nodes.

d. Parallel frequency compensation was considered. The structure has to be compensated at the output of the differential amplifier. The compensation capacitor has to be coupled to the high-side because the gate of the driver transistor is tied to the high-side, coupling of the compensation capacitor to the ground would cause worsening of PSRR. Since a relatively large range of the load current is needed, the compensation zero has to track the output pole - active compensation. The output pole tracking is implemented using the adaptive biasing. The output pole shifts with the load current so the adaptive biasing includes the information about the output pole position. With the higher load, there is larger V_{GS} of T_9. This voltage controls the resistance of liner-region operated zero-making transistor T_{12}.

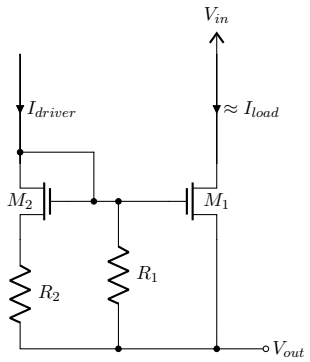


Figure 4.16: Source degenerated current mirror of the first approach structure driver

Unfortunately there are several complications which has caused that this has not been used finally. The main problem was the fact that stability for under all conditions could not be accomplished. This is mainly related to the output resistance of the driver. As it has been mentioned - quite large current is needed to achieve sufficiently low resistance of the driver to push the driver pole to sufficiently high frequencies.

The current mirroring ratio of M_2 and the power NMOS has to be very large, in order of thousands not to have too high current consumption through the driver at high loads. Let us say, that the maximum allowed current through the driver is set to be 100 μA. Taking the maximal load as 300 mA, the ratio has to be:

$$\frac{W_{power}}{W_{M_2}} \geq 3000, \tag{4.36}$$

considering the same channel lengths. This means that almost no current flows through the driver at light loads. So for that reason, a resistor connected between the gate and sources of the current mirror as in figure 4.16 has to be added to assure some minimal current through the driver. But this current has to be set not to cross the maximal quiescent current at no load. But on the other hand, this amount of current may not be enough to assure that the transconductance $g_{m_{M2}}$ is large enough for the driver pole to have a low impact on the phase margin.

At light loads, the pole at the output of the differential amplifier and the output pole of the LDO are located at very low frequencies, so that the UGF of the LDO is quite low, so even the smallest current through the driver is enough to push its pole out of the UGF. On the other hand, the UGF is maximal at heavy loads but the current through the driver is maximal as well, so its pole is not a concern. The worst situation is at medium loads where there is a still quite small current through the driver but the UGF is at a medium frequency range.

To suppress this complication, some way ho to add more current through the driver whilst keeping the requirement for the quiescent current at no load and keeping the maximal allowed quiescent current at maximal loads at some "reasonable" values has to be implemented. One possibility is to implement a source-degenerated current mirror at the driver as it is depicted in figure 4.16. This kind of current mirroring has non-linear characteristic between I_{load} and I_{driver}. This is caused by the fact that the resistor connected to the source acts as a local negative feedback. When the reference current is low, the voltage drop across the resistor is very small, so it almost acts like there is no resistor connected. On the other hand, when the reference current gets higher, the voltage drop gets higher as well and the feedback gets

stronger and causes lowering of V_{GS} of the degenerated transistor. This causes the non-liner characteristic. To achieve high non-linearity, large $\frac{W}{L}$ of the degenerated transistor has to be chosen as well as the resistance of the resistor. This may lead to undesirable large values of both, which yields to a large occupied area on a chip.

There are other complications with source degenerating. If there is a need for precision of the non-liner mirroring function, the technology corners and the temperature coefficients of the resistor may cause problems. Usually the resistors with a lower sheet resistance have the lowest technology corner spread. Using these kinds of resistor may not be desirable if high non-linearity is needed, because of the occupied area on a chip. Unfortunately, after a lot of experimentation, sufficiently large non-linearity has not been achieved.

All that mentioned compilations and inability to achieve good stability under all conditions and corners has caused not using this structure.

4.9.2 Final structure

For the final structure of the LDO, it has been finally decided to implement the driver based on a voltage follower block. Pros and cons of this decision will be discussed latter. Overall complexity of this structure is bigger than the first approach structure, nevertheless, it was necessary to fulfill all the requirements. Further in this chapter, each block will be discussed separately.

4.10 Differential amplifier

With regard to the fact that the driver is based on a voltage follower and the power NMOS is connected as a voltage follower as well, the differential amplifier represents the only one gain stage of the LDO. Main requirements on the differential amplifier have been put on:

- Sufficiently large gain - the only one gain stage
- Large output swing - since the driver is based on a voltage follower and the fact that the LDO has to supply the output voltage in a large range of $V_{out} = 0.8 - 4\,\mathrm{V}$.
- Ability of the differential pair to work with quite low input voltage, which is set by the band-gap voltage reference to be $V_{ref} = 0.8\,\mathrm{V}$
- Sufficiently large PSRR

Basic simplified structure of used differential amplifier is depicted in figure 4.17. For further reading, the parameters of the transistors in this structure are symmetrical and the current mirroring ratio is set to be 1:1 for all current mirrors. The virtue of this structure is its large output swing which is only limited by the minimal saturation voltage $V_{DS_{min}}$ of transistors T_3 and T_4 so it is in a range of:

$$output\ swing = from\ V_{DS_{3min}}\ to\ V_{bias} - V_{DS_{4min}}. \tag{4.37}$$

The relatively low input voltage determines usage of the PMOS differential pair. All transistors in the structure have to be kept in saturation region. Thus, the input common voltage range for PMOS differential pair is given by:

$$V_{in_{min}} = V_{GS_2} + V_{DS_{1min}} - V_{GS_1} = V_{GS_2} - V_{TH_1}, \tag{4.38}$$

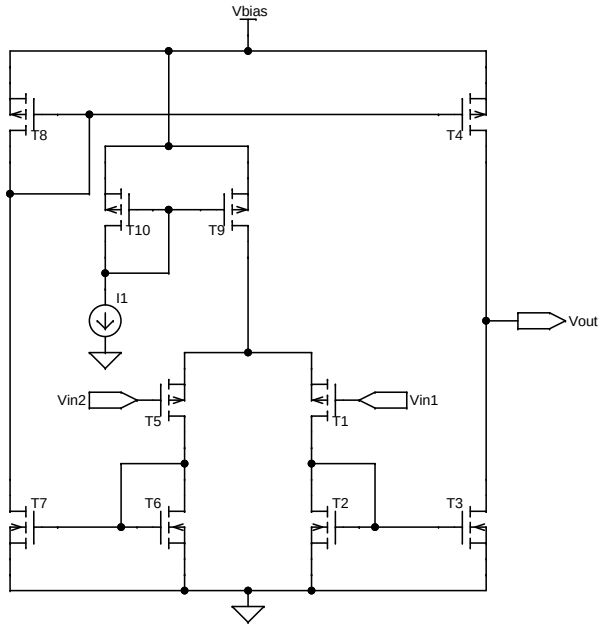


Figure 4.17: Simplified basic structure of differential amplifier

$$V_{in_{max}} = V_{bias} - (V_{GS_1} + V_{DS_{9min}}), \tag{4.39}$$

on the other hand, for the NMOS differential pair, the input common voltage range is given as (labeling of transistors is assumed the same but for "inverted" structure):

$$V_{in_{min}} = V_{DS_{9min}} + V_{GS_1}, \tag{4.40}$$

$$V_{in_{max}} = V_{bias} - (V_{GS_2} + V_{DS_{1min}} - V_{GS_1}) = V_{bias} - (V_{GS_2} - V_{TH_1}). \tag{4.41}$$

From these equations it is obvious that the PMOS differential pair is more suitable in a case when there is a need for the low minimal input voltage. On the other hand, the NMOS differential pair should be used when the high maximal input voltage range is necessary. So for the case when $V_{ref} = 0.8$ V the PMOS is the only option since 0.8 V may not be enough to cover one V_{GS} and one V_{DS}.

4.10.1 Small signal behavior

Now, let us explore the behavior of this structure. Just for now, ideal transistors with the same transfer characteristic for PMOS and NMOS are considered. If the input voltages are identical, the same current flows through each transistor of the differential pair. The current through T_1 is mirrored to the output branch by T_2 and T_3.

The current current through T_5 is mirrored by current mirrors formed by transistors T_6, T_7, T_8 and T_4 to the output branch. In this branch the currents of each transistor of differential pair are summed up. If the NMOS and PMOS transistors are assumed to have the same transfer characteristic, the output voltage should be in the middle of biasing voltage. If the input voltages differ, differential pair transistor are flown by different currents. In this situation, one of the output transistors is forced to enter linear region, so the output voltage

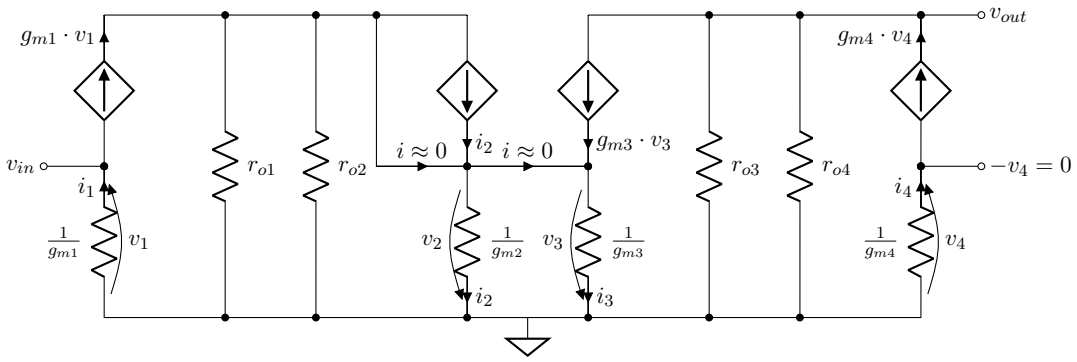


Figure 4.18: Small-signal model of differential amplifier

swings to one or another rail. At the output branch the "lowest current wins" so the transistor which should be flown by higher current enters the linear region. Of course, this situation is related to ideal situation where the transistors have the infinite output resistance and this situation would mean infinite gain. For figuring out the real gain, small signal model such as in figure 4.18 can be used. Assuming ideal current mirroring by T_6 and T_7, the gain can be determined using only one half of the circuit and transforming the differential signal to the single ended signal. According to the model we can write:

$$i_1 = -v_{in} \cdot g_{m_1}, \tag{4.42}$$

$$i_2 = i_1 \cdot \frac{r_{o_1} \parallel r_{o_2}}{r_{o_1} \parallel r_{o_2} + \frac{1}{g_{m_2}}}, \tag{4.43}$$

$$i_3 = i_2, \tag{4.44}$$

$$v_{out} = -i_3 \cdot r_{o_3} \parallel r_{o_4} = v_{in} \cdot g_{m_1} \cdot \frac{r_{o_1} \parallel r_{o_2}}{r_{o_1} \parallel r_{o_2} + \frac{1}{g_{m_2}}} \cdot r_{o_3} \parallel r_{o_4} \approx v_{in} \cdot g_{m_1} \cdot r_{o_3} \parallel r_{o_4}, \tag{4.45}$$

this simplifying can be done taking in account that that output resistance of a MOS transistor is usually significantly larger than $\frac{1}{g_m}$, finally, the gain can be expressed as:

$$G \approx g_{m_1} \cdot r_{o_3} \parallel r_{o_4} = g_{m_1} \cdot r_{out}, \tag{4.46}$$

where $r_{o_3} \parallel r_{o_4}$ represents the output resistance of the differential amplifier r_{out}. For achieving as large gain as possible, the differential pair transistor should be operated in weak-inversion to achieve the maximal g_m for the given biasing current. Thus, the $\frac{W}{L}$ ratio of the differential pair should be large enough.

Attention should be paid on connecting bulks of the differential pair. If the bulks are connected to the biasing voltage, the body effect occurs and the threshold voltage of the differential pair transistors increases. This can be advantage when there is a need of larger headroom at the low-side. Increasing the threshold voltage yields at larger V_{gs} of the differential pair while its minimal saturation voltage $V_{DS_{min}}$ remains the same. This increases the headroom at the low-side. Thus, according to equation 4.41, it makes the minimal common input voltage lower. But there is a drawback in terms of PSRR. In that situation, every fluctuation of the bias voltage is transfer to the bulks and the fluctuation then affects the body effect as well, which can show as decreasing PSRR. For that reason, the bulks are connected with the sources which eliminates worsening of PSRR, since there is no body effect.

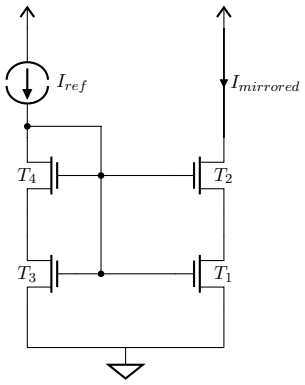


Figure 4.19: Modified cascoded current mirror

4.10.2 Cascoded current mirrors

With this structure of the differential amplifier, usually achievable values of the gain are around 50 dB. This may not be sufficiently large for the good line and load regulation, especially for our case where the differential amplifier is the only gain stage. To overcome this complication, some kind of a cascoded mirror to increase the output resistance and thus increase the gain has to be implemented. The drawback of using cascoding is its need for larger voltage headroom. For our case where $V_{ref} = 0.8$ V, there is not enough headroom, especially at low side, for using a classical cascode mirror which needs two V_{gs}. Thus, improved cascoded mirrors depicted in figure 4.19 have been used. This cascoding mirror requires only one V_{gs}. On the other hand, of course, every transistor needs to be in saturation region for proper function. The difficulties can occur especially when assuring that the lower current mirroring transistor T_1 is in saturation. Assuming that this transistor is operated in strong inversion, for its drain-source voltage V_{DS} to be in saturation it must hold:

$$V_{DS_1} = V_{GS_1} - V_{GS_2}, \tag{4.47}$$

$$V_{DS_1} \geq V_{DS_{1min}} = V_{GS_1} - V_{th_1}, \tag{4.48}$$

$$V_{GS_2} < V_{th_1}, \tag{4.49}$$

this condition can be achieved two ways:

a. using depletion type of upper transistors, if the technology allows this

b. operating upper transistors in weak inversion

BCD8 technology does not allow using depletion transistors so option **b.** must have been used. Requirements for operating the lower transistor in strong inversion for better accuracy of current mirroring and lower current offset due to the mismatch (discussed in chapter 1) lead to small $\frac{W}{L}$ ratio. On the other hand, operating upper transistors in weak inversion leads to large $\frac{W}{L}$ ratio. It is important to mention that the bulks of upper transistors must be connected with their sources. Connecting them with ground would cause increasing of the threshold voltage of the upper transistors by body effect and the condition from equation 4.49 could not be fulfilled since V_{gs} of upper transistor would be increased.

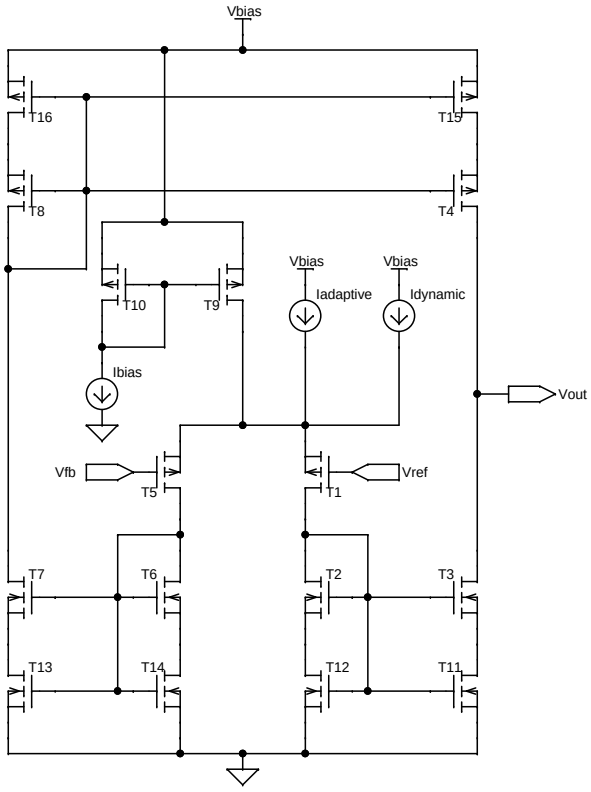


Figure 4.20: Simplified structure of the used differential amplifier

If all transistors are in saturation region, then the output resistance R_{out} of this cascoded current mirror is given by the same relation as in a classical cascoded mirror [5]:

$$R_{out} = r_{o2}(1 + g_{m2}r_{o1}) + r_{o1}, \tag{4.50}$$

from this equation is is obvious that operating upper transistor in weak inversion has also advantage in maximal transconductance g_{m2} for given current. Thus, the output resistance is maximal as well.

In real design, it can be very difficult to assure that all transistor are always in saturation region at every condition and corner, especially when adaptive biasing of differential amplifier is used, so achieving maximal theoretical output resistance may not be fulfilled because the transistors may work at a "border" between linear and saturation region, anyway, a certain improvement of the output resistance still occurs in this situation.

4.10.3 PSRR considerations

For good PSRR it is necessary to use cascoded mirrors in this structure as well. The driver is based on a voltage follower which yields in a large variance of the output voltage of the differential amplifier. This means that the large difference between V_{DS} of transistors T_4 and T_8 or T_3 and T_7 back in figure 4.17 can occur since T_8 is diode-connected. This large difference of V_{DS} causes errors in current mirroring which occurs as an offset voltage at the output. The value of this offset depends on the biasing voltage, thus not sufficient PSRR. When using cascoded mirrors, their output resistance is much higher so influence of the biasing voltage

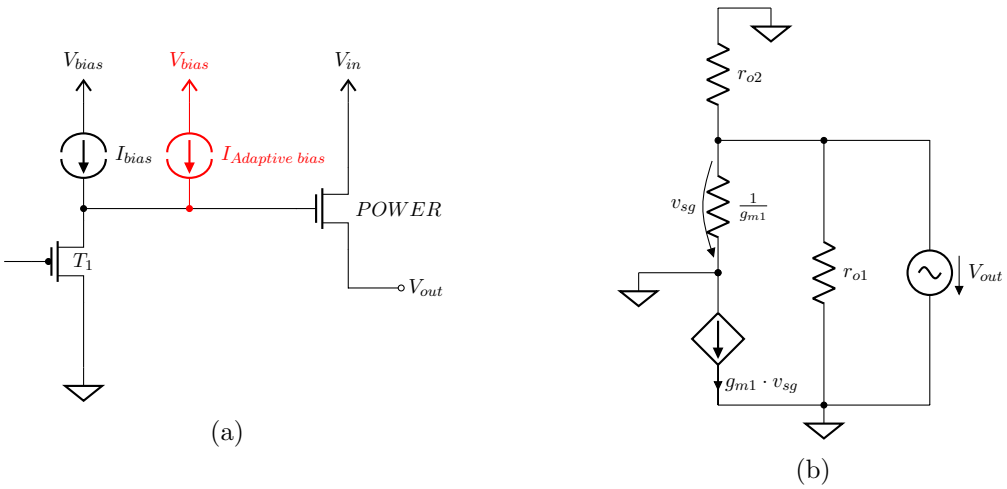


Figure 4.21: Principle of the driver (a) and a small-signal model of a simple voltage follower for determining its output resistance (b)

variation is suppressed. In other words, the large difference of V_{DS} is mainly "absorbed" by the cascoding transistors, while the current-mirroring transistors have relatively similar V_{DS}.

This situation represents one the disadvantages of using a driver based on a voltage follower. In the first approach structure back in figure 4.15, there is no need for cascoded mirrors because the output voltage of the differential amplifier is still in the range one V_{gs}.

The final simplified structure of the differential amplifier with depicted adaptive and dynamic biasing is represented in figure 4.20.

4.11 Driver

The driver of the power transistor represents one of the most critical blocks of the LDO regrading the transient response behavior. The main goal of the diver is to supply a low impedance source for driving the large power transistor. Before the LDO can react to changes of the load current or the input voltage, capacitances of the circuit have to be recharged. The large area of the power NMOS represents a large gate capacitance. This capacitance, besides the output capacitor, may stand for the largest capacitance in the circuit. So how fast this capacitance is able to recharge determines the transient behavior significantly. There are two main challenges that need to be assured:

- Pushing the pole which is made of the gate capacitance of the power NMOS and the output resistance of the driver to frequencies as high as possible to suppress its impact on stability.

- Assuring sourcing and sinking of a current large enough, which is necessary for sufficiently large slew rate for the fast recharging of the gate capacitance of the power NMOS.

The simplified structure of the proposed driver is depicted in figure 4.21a The core of the driver is transistor T_1 connected as a voltage follower which is biased from fixed current

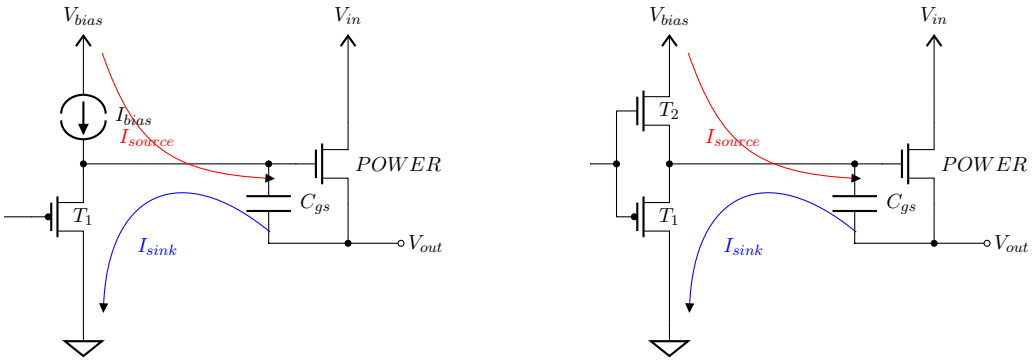


Figure 4.22: Principle of sourcing and sinking current to the load of a simple follower (a) and a push-pull follower (b)

source I_{bias}, adaptive biasing block $I_{adaptive\ bias}$ will be disclosed later. The output resistance of a simple voltage follower can expressed by using a small-signal model in figure 4.21b as:

$$r_{out} = \frac{1}{g_{m1}} \parallel r_{o1} \parallel r_{o2} \approx \frac{1}{g_{m1}}, \tag{4.51}$$

where r_2 represents the output resistance of the biasing current source. This simplifying in the equation can be done because in most cases $\frac{1}{g_{m1}} << r_o$. From this relation, it can be seen that the output resistance of the voltage follower is the same as for the driver based on a current mirror used in the first approach structure in figure 4.15. So the same problems remain:

- The requirement for the ultra-low quiescent current at no load does not allow using sufficiently large biasing current to assure that g_{m1} is large enough to push the pole to frequencies high enough. Thus, again, there must be some technique how to make the output resistance lower.
- Large biasing current is needed for the high slew rate which is again in breach of the ultra-low quiescent current requirement.

Regarding the slew rate, a voltage follower is not an ideal solution for sourcing the current to the load. As depicted in figure 4.22a, the sourcing capability of a current to charge the gate capacitance is restricted by the biasing current source, which is very limiting in ultra-low quiescent current design. On the other hand, sinking capability of a current for discharging the gate capacitance is high, restricted only by the $\frac{W}{L}$ ratio of the voltage follower transistor and its maximal applied V_{gs} voltage. This situation is depicted in figure 4.22b. An ideal solution would be using a push-pull topology but this structure cannot be used due to the fact that there is a need for two V_{gs} voltages above the output voltage of the LDO, this would yield to unacceptably high minimal biasing voltage V_{bias}.

4.11.1 Adaptive biasing of the driver

From these mentioned facts, it is obvious that, as in the first approach structure, some way how to to add more current to the driver, whilst keeping the requirement for the quiescent

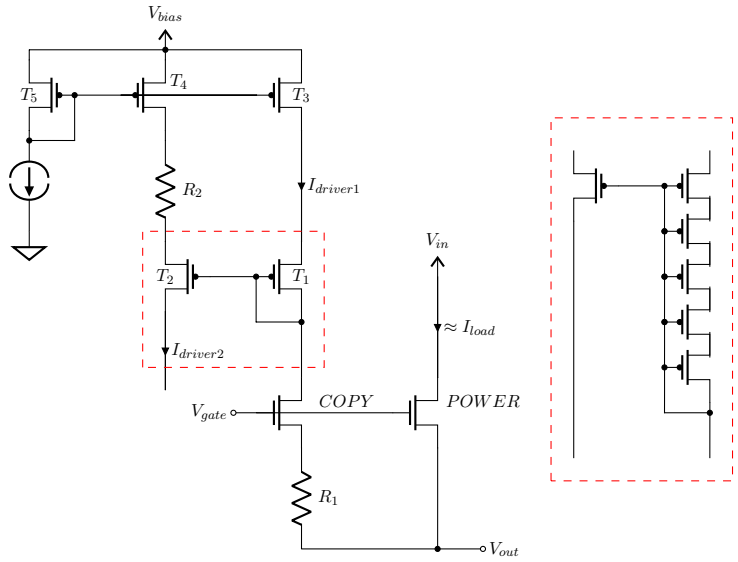


Figure 4.23: Simplified schematic of the adaptive biasing circuit for the driver

current at no load and keeping the maximal allowed quiescent current at the maximal load has be to introduced. It has been decided to use non-linear adaptive biasing.

It is necessary that the adaptive biasing would be as simple as possible to assure a low time delay. For instance, every current mirror has a certain time delay due to the parasitic capacitances which have to be recharged. Too large time delay would cause unacceptably high undershoots at the step transient response. Simple and effective way is using source degenerating current mirrors. The simplified structure of the adaptive biasing is depicted in figure 4.23. Current mirroring is provided by the copy NMOS with a suitable $\frac{W_{power}}{W_{copy}}$ ratio with the power NMOS and by the current mirror of T_1 and T_2. Non-linearity is caused by the effect of two source degenerating resistors R_1 and R_2. Let us for now consider transistors T_3 and T_4 shorted.

On contrary of the first approach structure, where source degenerating has not been satisfying due to the low precision and not high enough non-linearity to assure the low output resistance of the driver for pushing the driver pole far enough at the worst case of a load current range, in this structure, there is no such a need for very precise current mirroring function. Also, much higher non-linearity can be achieved because there can be used two source degenerating resistors and their effect for the non-linearity is then multiplied.

It has been simulated that the critical factor for the sufficiently good load transient response and the phase margin is to supply a sufficient current to the driver at the mid-range loads. In this range, in our case around from $I_{load} = 100\mu$A to $I_{load} = 1$mA, the UGF of the open-loop transfer function of the LDO is at medium values, the pole at the output of differential amplifier and the output pole of the LDO are not too separated and the pole of the driver is difficult to be pushed far enough because the adaptive biasing current may not be sufficiently large. On the other hand, the value of the biasing current of the driver at high loads has not shown so big impact on improving the slew rate and reducing undershoots.

When designing the adaptive biasing there are several complications:

a. There is a need for low current $I_{driver1}$ through the branch of the copy NMOS for the following reason. Every transistor needs to stay in saturation region so the biasing voltage V_{bias} needs to be at least one V_{gs} and one $V_{DS_{min}}$ plus the voltage drop across the

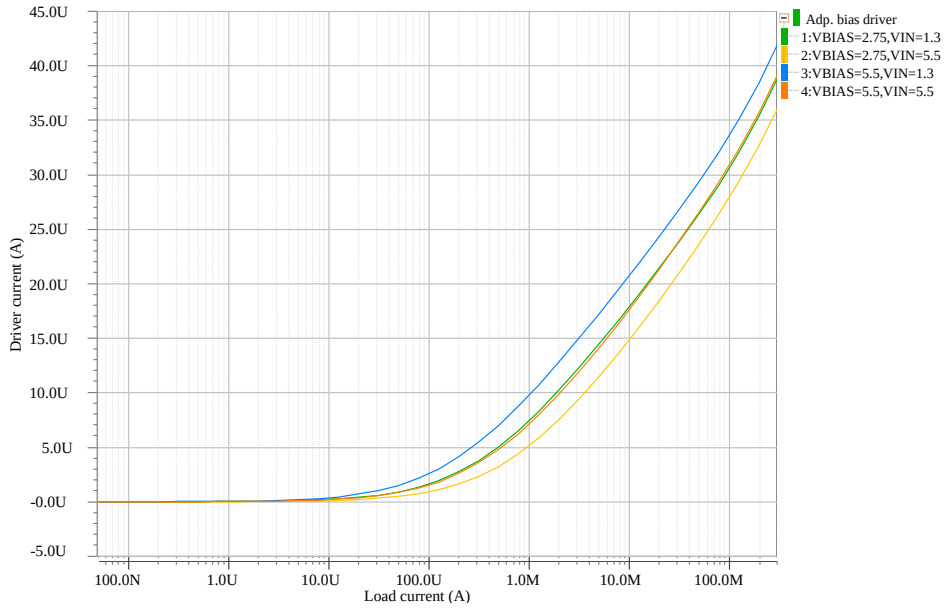


Figure 4.24: Adaptive biasing of the driver - simulation results

resistor R_1 higher than the output voltage. Due to the fact that there are tendencies for lowering V_{bias} as possible, it may not be easy to assure saturation. This yields to using large $\frac{W}{L}$ of T_1 to achieve its low V_{gs}. This is not ideal for current mirrors (reasons are mentioned in chapter 1). Making $\frac{W}{L}$ of the copy NMOS larger does not help reducing $V_{DS_{min}}$ because its larger $\frac{W}{L}$ ratio also means larger current through this branch because of the higher current mirroring ratio with the power NMOS. Thus, a lower current through this branch reduces both $V_{DS_{min}}$ and low V_{gs}.

Also, there is a source of an error in the current mirroring - V_{DS} of the copy NMOS can be significantly different than V_{DS} of the power NMOS, and due to the finite output resistance of the transistors, there may be higher dependency of the current mirroring on V_{bias} and V_{in} ratio. On the other hand, source degeneration resistor acts as a local negative feedback which suppresses this dependency.

b. As it has been already mentioned, achieving large non-linearity of the current mirroring yields to using large resistors and high $\frac{W}{L}$ ratio of the source degenerated transistors. Also $\frac{W_2}{W_1}$ ratio has to be quite large as well (assuming the same channel lengths). The problems is that $\frac{W_1}{L_1}$ of T_1 may already be not so small. This would mean larger area on a chip. Choosing different channel lengths is not an option - different lengths may cause different threshold voltage and not precise current mirroring due to different transfer function. To reduce the area, a series connection of transistor is used. N Series connected transistors with the same channel length act in a current mirror similarly as a single transistor with N-times longer channel length than originally but the threshold voltage and the transfer function is nearly the same as originally. Thus, a reduction of the occupied area is achieved.

Simulation of dependency of the adaptive biasing current of the driver on the load current

for certain combinations of V_{bias} and V_{in} is depicted in figure 4.24. Still, for the high slew rate during large transients, the adaptive biasing may not be sufficient. To overcome this problem, dynamic biasing circuit boosting a large amount of current to the driver only during the transients, when it is needed, is implemented. This block will be discussed separately later in this chapter.

4.11.2 Drop-out behavior

When the LDO is entering the drop-out region, it stops working as a closed-loop system and starts to behave as an open-loop system, this causes significant differences in the behavior of the internal circuitry. When that happens, the output voltage starts to be below the nominal value. This force the differential amplifier to push its output voltage to the high rail and since the driver is a voltage follower, the gate of the power NMOS goes to the high rail as well. V_{DS} voltage of the copy NMOS is still high to remain the copy NMOS in saturation, on contrary of the power NMOS. Proper copying current stops working and a large amount of current starts to flow through the branch of the copy NMOS. Although, the requirement for quiescent current is not usually defined for the drop-out mode and larger consumption in the drop-out is a common thing in LDOs, this might not be desirable.

Proposed solution is to clamp the current through the branch of copy NMOS by adding transistor T_3 which makes a current mirror with T_5, again using figure 4.23. The principle of function is following:

- In the normal mode of operation of the LDO, the current through the copy NMOS is lower than a current set by current source I_1 and multiplying factor of the current mirror consisted of T_5 and T_3 - let us call it I_{clamp}, this forces T_3 to enter the linear region and act as a switch with a certain negligible $R_{DS_{on}}$. Thus, it practically does not effect the function of the driver.

- When the LDO is entering the drop-out, the current through the branch increases, when the current reaches the value of I_{clamp}, T_3 enters the saturation region and starts to act as current source which forces the copy NMOS to enter linear region. Thus, the current through the branch is clamped at the value of I_{clamp} .

The value of I_{clamp} must be set higher than the maximal current through the driver in the normal mode, not to effect the proper function of the driver.

Analogous clamping must be done at the other branch of the adaptive biasing. Although, in the "full" drop-out, the gate voltage of the power NMOS and copy NMOS is at high rail, so V_{DS} voltage of T_2 is close to zero and almost no current flows through its branch. But, on the "boundary" between the normal mode and the drop-out, as the gate voltage approaches the high rail, there is still some headroom to allow T_2 to stay almost in saturation and an undesirably high current may flow through the branch. Hence, similar clamping using transistor T_4 should be done as well.

The best solution would be designing the circuit that assures lower quiescent current when the LDO is in the drop-out. Such a circuit is in [15]. This circuit works only for PMOS regulators. Hence, some modification would have to be done. On the other hand, these modification may mean increasing the quiescent current at no load.

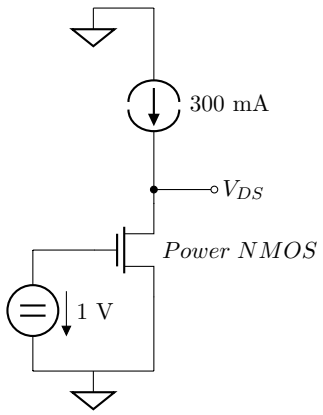


Figure 4.25: Test-bench for determining $R_{DS_{on}}$ of the power NMOS

4.12 Sizing of the power NMOS

The size - $\frac{W}{L}$ ratio of the power NMOS is determined by a value of required drop-out voltage V_{drop} at the maximal load current $I_{load_{max}}$. The higher the maximal current and the lower the drop-out voltage, the larger size. Besides the occupied area on a chip, the drawback of the large size is a large gate capacitance which brings complication regarding stability and the slew rate. Also, larger size means larger leakage current. In this case, the leakage current can be significant since it is generally higher at NMOS than at PMOS transistors. On the other hand, the size of NMOS can be smaller due to higher mobility of charge carriers.

When the LDO enters drop-out mode, the power transistors acts as a resistor with a value of $R_{DS_{on}}$ which has to be smaller than:

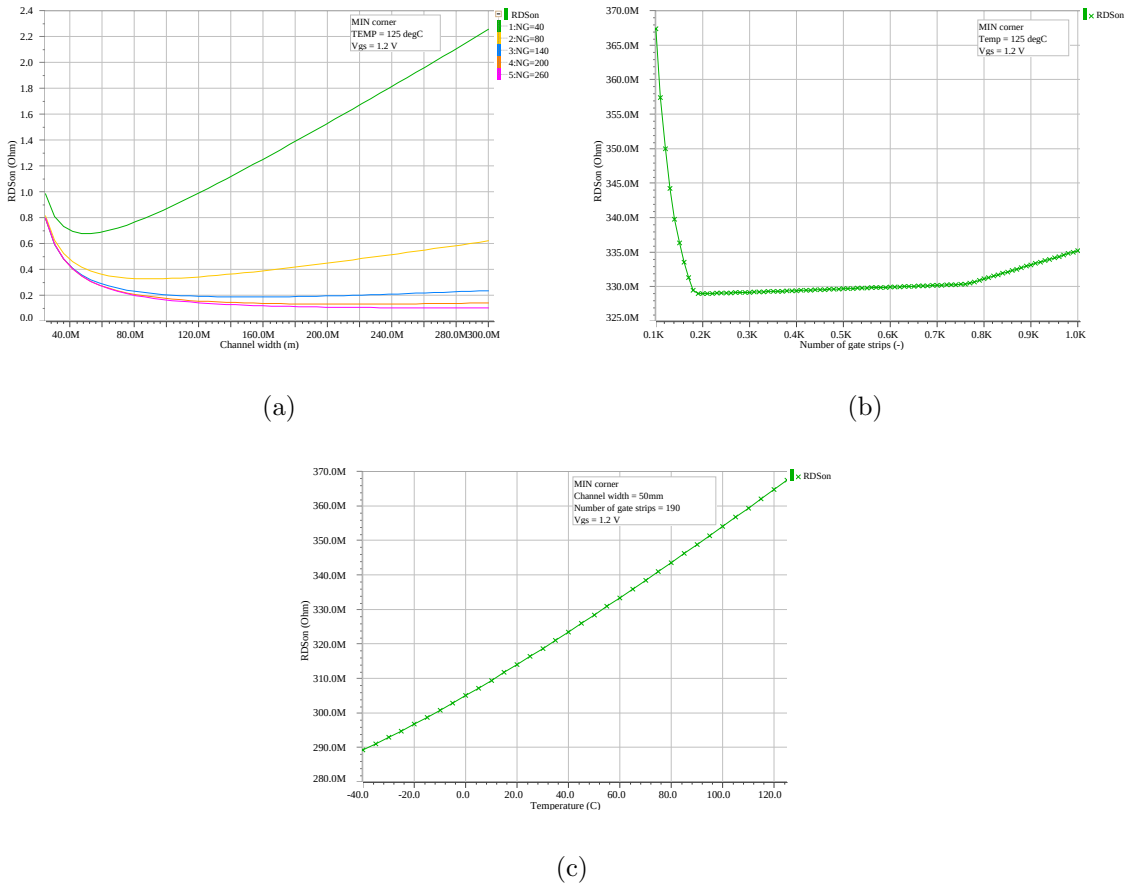
$$R_{DS_{on}} \leq \frac{V_{drop}}{I_{load_{max}}} = \frac{150\ mV}{300\ mA} = 0.5\ \Omega. \tag{4.52}$$

$R_{DS_{on}}$ can be determined using a characteristic equation for NMOS operated in linear region:

$$R_{DS_{on}} = \frac{V_{DS}}{I_D} \approx \frac{1}{\frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{th})}, \tag{4.53}$$

from this, it is obvious that low $R_{DS_{on}}$ leads to larger $\frac{W}{L}$ ratio. A good way how simulate $R_{DS_{on}}$ is using a test-bench depicted in figure 4.25. Firstly, a value of available V_{GS} has to be chosen, the higher V_{GS} the lower $R_{DS_{on}}$. But higher available V_{GS} leads to higher minimal V_{bias}. For the simulation, it has been chosen $V_{GS} = 1.2$V. In the final design, there may be a headroom for larger V_{GS} but for covering worst cases, it it suitable to chose a smaller value. The power NMOS in the test-bench is biased by the ideal current source representing the maximal load current, in this case 300 mA. The power NMOS is then forced to linear region and for determining $R_{DS_{on}}$ it does just to measure V_{DS} and divide it by the current.

The channel length has been set to the minimum, in this case $L = 600$ nm. It must be said that there are many other factors that affect $R_{DS_{on}}$. The most important are temperature and technology corners. The worst case is for MIN technology corner where the threshold voltage is maximal and according to equation 4.53 it means larger $R_{DS_{on}}$. Regarding the temperature, $R_{DS_{on}}$ increases with increasing temperature. Although the threshold voltage decreases with temperature, the increasing resistance of the metalization of the power NMOS has a more significant effect.



(a) (b)

(c)

Figure 4.26: Dependency of $R_{DS_{on}}$ on the channel width for various numbers of gates strip (a), dependency on the number of gate strips (b) and on temperature (c)

Table 4.3: Parameters of the power NMOS

W :	L :	number of gate strips:
50 mm	600 nm	190

The equation 4.53 is simplified, in reality, this relation is by far more complex. For now the goal has been to only determine $\frac{W}{L}$ ratio. But there is another factor that has to be figured out - the number of gate strips of the power NMOS. It can be observed that the lowest $R_{DS_{on}}$ is for the number of gate which provides an approximately square shape of the power NMOS in its layout.

Final parameters of the power NMOS are summed up in table 4.3. Simulation results are shown in figure 4.26. Figure 4.26a shows the dependency of $R_{DS_{on}}$ on the channel width for various number of gate strip. Figure 4.26b represents the dependency on the number of gate strips for chosen channel width and figure 4.26c shows the temperature dependency for chosen channel width.

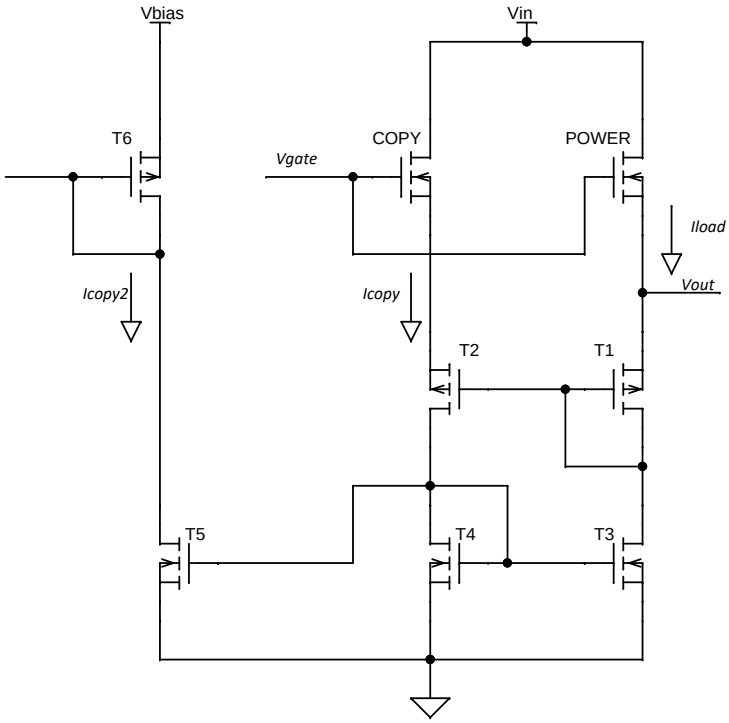


Figure 4.27: First approach of adaptive biasing circuit

4.13 Adaptive biasing

It has been already mentioned that it is suitable to assure adaptive biasing for the differential pair as well as for the active frequency compensation. Of course, first thought would be implementing the adaptive biasing similar way as it has been designed for the driver. Having been said, the drawback of this type of adaptive biasing is its imprecision. For the case of the differential amplifier and the frequency compensation it is not acceptable mainly for stability reasons:

- The biasing current is the main factor for determining the UGF of the differential amplifier and of the whole LDO as well. Too large dependency of the adaptive biasing on, for instance, the biasing voltage V_{bias}, input voltage V_{in} and temperature as well as technology corners, could potentially make assuring sufficient phase margin under all conditions unattainable due to large changes of the UGF.

- Imprecision of mirroring the load current leads to imprecision of tracking the output pole by the active parallel compensation which may cause achieving good stability under all conditions not possible.

4.13.1 First approach structure

The important thing for improving precision of the adaptive biasing is ensuring the same V_{DS} voltage for both the power NMOS and for the copy NMOS. The structure in figure 4.27 was chosen as the first approach namely for its simplicity and low current consumption.

Firstly, let us assume a certain current through the copy NMOS, the same current flows into T_4, this current is mirrored by T_3. The mirrored current flows through diode-connected T_1 with a certain value of voltage V_{GS1}. If transistors T_1 and T_2 are ideally identical and assuming ideal current mirroring of T_4 and T_3 in 1:1 ratio, then $V_{GS1} = V_{GS2}$, thus the same V_{DS} and V_{GS} for the power NMOS and the copy NMOS is assured. Transistors T_5 and T_6 assure mirroring of the copy current to other circuitry.

For proper operation, it is necessary that all transistors are in saturation region. To assure that, there is a need for one V_{GS} and one $V_{DS_{min}}$ below the output voltage of the LDO. This brings a drawback of this structure. In the situation when the output voltage is set to be $V_{out} = 0.8$ V, there may not be enough headroom. Firstly, this situation yields to large $\frac{W}{L}$ ratios of T_1 - T_5 transistors and operating them in weak inversion to lower their V_{GS} and $V_{DS_{min}}$. Again, this is not ideal for current mirrors (as mentioned in chapter 1).

Several following issued has caused not using this topology:

a. The need for the large voltage headroom has been the main issue with this structure. Even for the lowest allowed channel width of the copy NMOS, the copying current has been too high at heavy loads to fulfill the condition for assuring saturation region for all transistors under every corner condition, especially for MIN technology corner and the lowest operating temperature.

b. The overall precision has not been satisfying, mainly at light loads when the copy current is too low for proper biasing of the current mirrors. Also higher dependency on technology corners and temperature have occurred.

c. Using all transistors operating in weak-inversion leads to large $\frac{W}{L}$ ratios. This means that non-negligible leakage currents at high temperatures may occur. This represents a significant problem regarding the frequency compensation at light loads - output pole tracking by the parallel active compensation is highly dependent on temperature at light loads, which causes assuring a sufficient phase margin unattainable. Although several leakage current compensation circuits have been implemented, the results have not been a hundred percent reliable and satisfying.

4.13.2 Final structure

The simplified structure of the final adaptive biasing circuit is depicted in figure 4.28. This structure is based on an operational amplifier working in a closed-loop. Sources of the both copy and power NMOS are connected to the inputs of the PMOS differential pair made of transistors T_1 and T_2. Transistors T_3 and T_4 represent an active load of the differential stage. The second stage consists of transistor T_5 connected as a common source amplifier loaded by the copy NMOS. The closed-loop provides ideally identical voltages (assuming infinite open-loop gain and zero voltage offset) at the inputs of differential pair.

Resistor R_1 makes the characteristic between load and copy current slightly nonlinear. At light loads, the copy current is small so is the voltage drop across the resistor and the effect of R_1 is negligible and V_{DS} and V_{GS} of the copy NMOS and the power NMOS is practically the same. At high loads, the voltage drop starts to increase and the effect of the resistor gets bigger - acts as a local negative feedback and slightly decreases V_{GS} of the copy NMOS, thus decreases slightly the copy current I_{copy} - the principle is basically the same as for source degenerating current mirrors. It has been figured out that slightly non-linear function is suitable for slightly lower quiescent current at high loads without affecting stability of the LDO - the "active" zero of the active parallel compensation is already pushed to very high

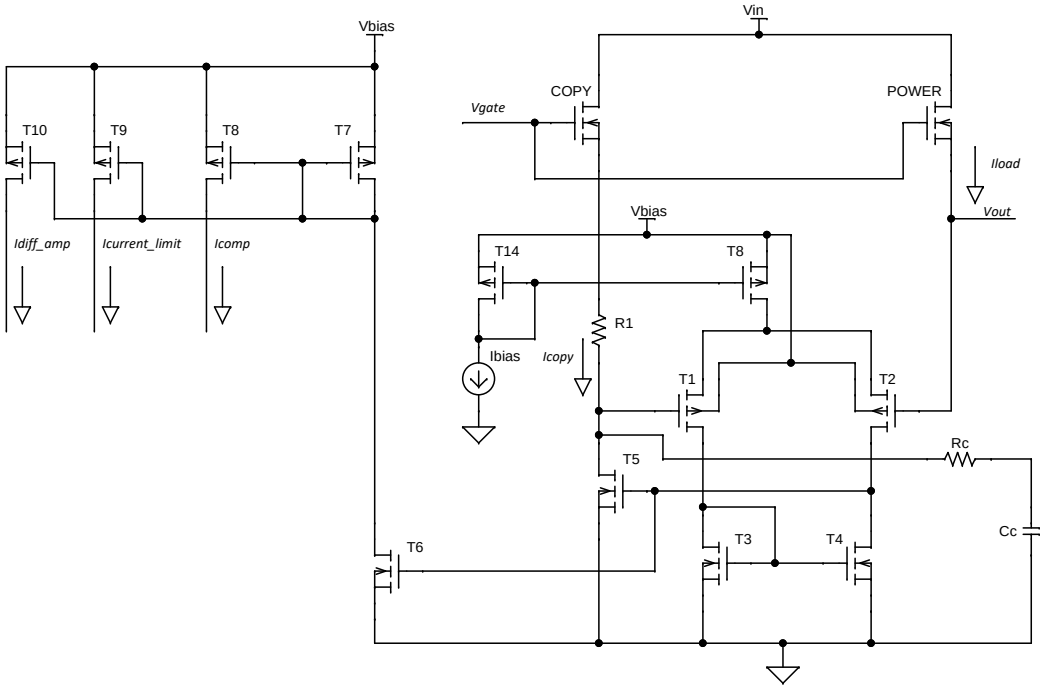


Figure 4.28: Simplified schematic of adaptive biasing circuit

frequencies that more copy current has almost no effect on stability (only a small difference of the UGF caused by the adaptive biasing of the differential amplifier can be observed). This is also the reason why the dependency of the resistance value on temperature and technology corners is not an issue in this case.

As a two stage amplifier, some kind of frequency compensation is needed for a sufficient phase margin. Compensation capacitor C_c together with compensation resistor R_c acts similar way as mentioned compensation technique using ESR together with the output capacitor in section 4.5. C_c ensures that the open-loop gain crosses unity at lower frequencies while R_c inserts a zero into the system.

Finally, transistors $T_6 - T_{10}$ provide current mirroring of I_{copy} and supplying of the adaptive biasing current to other circuit blocks.

Special attention should be paid on:

a. Connecting the bulks of the PMOS differential pair to highest potential to use body effect for increasing the threshold voltage V_{th} of the differential pair. Increased threshold voltage helps providing larger voltage headroom for transistors T_3 and T_4 to be in saturation region. This situation is related to the situation when the lowest output voltage of the LDO $V_{out} = 0.8$ V is set.

b. Sizing of T_5 is important. Its $\frac{W_5}{L_5}$ ratio has to be small enough to assure that V_{GS5} is high enough that T_4 stays in saturation at light loads where there is only a small current through T_5. On the other hand, too low $\frac{W_5}{L_5}$ may lead to too large V_{GS5} at high loads that T_2 enters liner region since there may not be a voltage headroom for its minimal saturation voltage V_{DS2min}. Again, this situation may occur at $V_{out} = 0.8$ V.

Simulated dependency of the adaptive biasing of differential amplifier on the load current for various combinations of V_{bias} and V_{in} is depicted in figure 4.29.

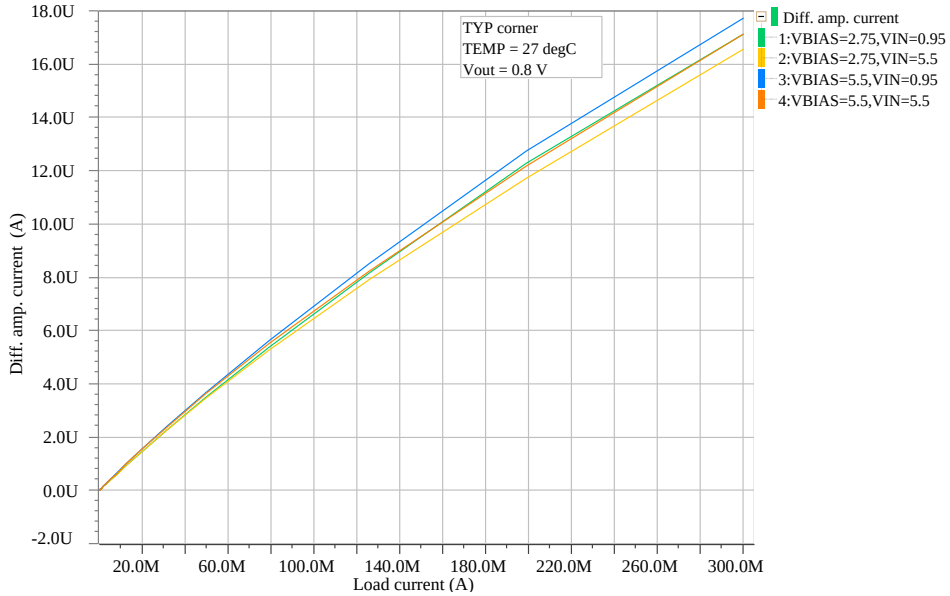


Figure 4.29: Simulated dependency of the adaptive biasing current of the differential amplifier on the load current

4.14 Dynamic biasing

Transient response behavior, especially the load transient response is a general weakness of ultra-low quiescent current LDOs. Having been said already, the load transient response is mainly determined by the speed of recharging capacitances in the circuit. The are two largest capacitances in the proposed LDO structure:

a. Gate capacitance of the power NMOS of which recharging is supplied by the driver

b. Compensation capacitance of the active parallel compensation of which recharging is provided by biasing of the differential amplifier

Both the differential amplifier and the driver are adaptively biased. At great transient steps, the adaptive biasing current will not be sufficiently large to assure a good slew rate - fast recharging of the capacitances and providing small undershoots at the output voltage.

Dynamic biasing boosts the current to the differential pair and to the driver only when there is an undershoot at the output voltage during the transients steps. At steady state, the dynamic biasing current is nearly zero. The proposed dynamic biasing is based on sensing the output voltage (the feedback voltage) and comparing this voltage with the reference voltage. The voltage between the feedback and the reference controls the boosting current sources for the differential amplifier and the driver. The boosting current ratio (K, L) for each block is different as it is depicted in a block schematic in figure 4.30.

Several issues had to be solved during designing the dynamic biasing:

a. Assuring as fast reaction of the dynamic biasing circuit as possible, again, this is mainly determined by the biasing current of this block.

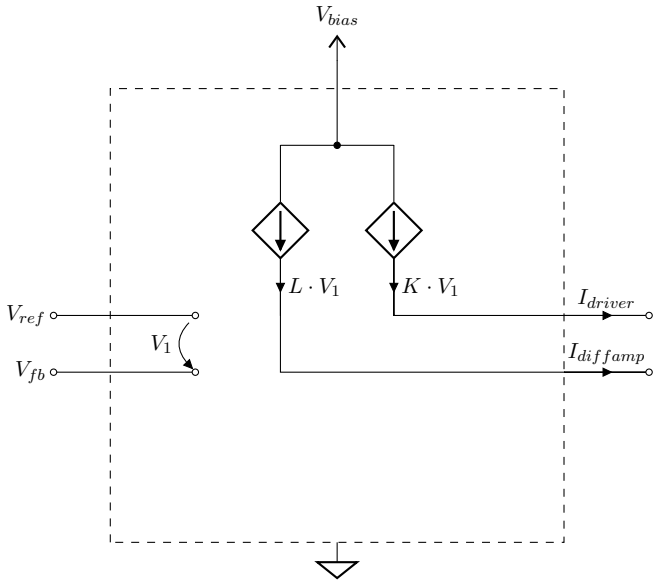


Figure 4.30: Simplified block schematic of dynamic biasing

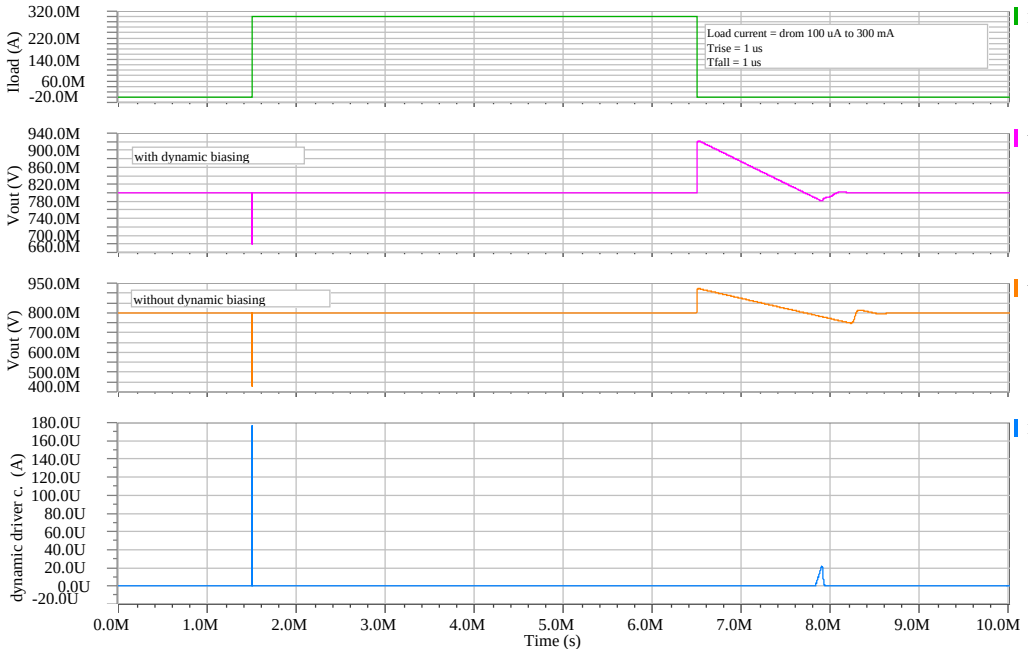


Figure 4.31: The effect of the dynamic biasing during large transient load steps

b. Low current consumption, which is in breach of **a.**, again.

c. Assuring almost zero current in the steady state under all conditions. Special attention had to be paid on Monte Carlo simulations.

d. Leakage current complications had occurred, compensation circuitry had to be implemented.

e. Choosing optimal current boost ratio. Too low ratio may not be enough for fast transients, on the other hand, too high ratio may cause some overshoots at the output voltage, this is related mainly for the biasing of the differential amplifier.

Example of boosting a current to the driver during load transients is depicted in figure 4.31. It can be seen, as expected, that the boosting current only occurs when there is about an undershoot at the output voltage. The effect of the dynamic biasing on reducing undershoots is significant as it can be seen in the figure.

4.15 Leakage current compensation

The size of the power NMOS transistor is significant, its $\frac{W}{L} \approx 83300$ while the channel length is minimal $L = 600$ nm. Transistors with minimal channel lengths are the most leaky and very large $\frac{W}{L}$ ratio of the power NMOS, according to equation 1.23, means large leakage current as well. Besides the current consumption, the leakage current of the power transistor might cause inability of voltage regulation at light loads. In a case of the NMOS power transistor, the operational amplifier of the LDO forces its gate to be pushed to a lower potential to reduce the conduction of the power NMOS. If a significant leakage current occurs, the gate of the power transistor can be forced below the potential of the source and even if the gate voltage is located in the low rail, it still does not have to be sufficient for lowering the conduction of the power NMOS enough to maintain the desired output voltage - the power NMOS cannot be turned off sufficiently - the output voltage is now higher than the desired value and the regulation stops working. For these reasons, a compensation of the leakage current is necessary.

The compensation of the leakage current can be implemented based on a generation of proportional leakage current and subtracting it from the leaky power transistor branch. Proposed simplified schematic of the leakage current compensation is depicted in figure 4.32. Firstly, let us consider the situation when the leakage current causes problems. Assuming the compensation circuit disconnected and the situation when the expected load current given by the desired output voltage V_{out} and the load resistance R_{load} as:

$$I_{load} = \frac{V_{out}}{R_{load}}, \tag{4.54}$$

and the current through the feedback voltage divider is smaller than the leakage current I_{leak} - the leakage current cannot be fully absorbed by the divider. If $I_{leak} > I_{load}$. Neglecting the small current through the divider, the current through the power NMOS is the actual load current I'_{load} expressed as:

$$I'_{load} = I_{leak} > I_{load}. \tag{4.55}$$

The leakage current then causes the actual output voltage V'_{out} to be:

$$V'_{out} > V_{out}, \tag{4.56}$$

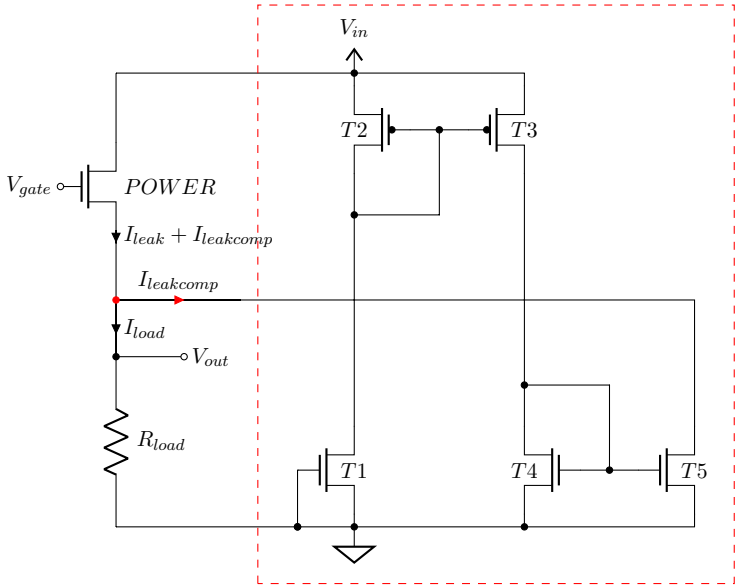


Figure 4.32: Simplified structure of leakage current compensation

so the voltage regulation does not work properly. Now let us describe the compensation connected. Transistor T_1 has the gate terminal connected with its source, so it represents a proportional leakage generator. For better accuracy, the channel lengths must be the same $L_{power} = L_1$. For the size of T_1 it then should hold:

$$W_1 \ll W_{power}, \tag{4.57}$$

$$\frac{W_{power}}{W_1} = M \tag{4.58}$$

for not too large area occupation. Leakage current is directly proportional to the $\frac{W}{L}$ ratio. Ideally T_1 generates M-times smaller leakage current than the power NMOS. Transistors $T_2 - T_5$ have to provide current mirroring and multiplying the proportional leakage current so that:

$$I_{leak} = I_{leakcomp}. \tag{4.59}$$

In this situation, the whole leakage current I_{leak} is now absorbed be the compensation and there is no unwanted excess of current to the load which causes $V'_{out} > V_{out}$ without the compensation. Thus, now we can write:

$$I'_{load} = I_{load} \tag{4.60}$$

and then for the output voltage with the compensation it holds:

$$V'_{out} = V_{out}. \tag{4.61}$$

- If $I_{leakcomp} < I_{leak}$, then the gate potential of the power NMOS can still be bellow its source and the NMOS cannot be sufficiently turned-off and the voltage regulation can still be corrupted.
- If $I_{leakcomp} > I_{leak}$, the regulation will, of course, work normally because the opamp of the LDO is capable of regulation the conduction of the power NMOS, but there is unnecessary current consumption.

There are several complications achieving $I_{leakcomp} = I_{leak}$:

a. Transistors $T_2 - T_5$ have to supply multiplying factor by M. Too large M can lead to large $\frac{W}{L}$ of these transistors and then their own significant leakage current can potentially occur. To reduce their own leakage, it is necessary to use longer channel lengths than minimal.

b. The leakage current of a transistor is dependent on its V_{DS} as well. In our case, The power NMOS and the proportional leakage generator transistor T_1 do not have the same V_{DS} so imprecision of generation $I_{leakcomp}$ can occur as well.

4.16 Feedback voltage divider

The LDO has to allow the adjustable output voltage. The customer's desired output voltage will be provided by a laser beam burning metal connections.

The voltage divider requirements are:

- $V_{out} = 0.8 - 4$ V
- adjustable by 50 mV step

The ultra-low quiescent current requirement forces us to design the diver using large resistors to provide the low current consumption which is a trade-off with an occupied area on a chip. To reduce the area, using a type of resistors with the largest sheet resistance in given technology may be needed. On the other hand, resistors with the largest sheet resistance do not usually provide the best matching, this can lead to worse precision of the output voltage which can be shown using Monte Carlo simulations. Another drawback of using large resistors is increased output noise caused by thermal noise according to equation 1.5.

For better matching and more precise output voltage, attention should be paid on:

a. Using resistors with bigger width than minimal, again, there is a trade-off between precision and occupied area.

b. Layout matching technique should be considered. For improved matching, it is suitable to consist the whole divider using approximately the same size parts to allow using layout techniques such as a common centroid etc.

It has been decided to use high resistance poly-silicon (HIPO) resistors. At typical technology corners and temperature, the current through the divider has been set to $I_{div} = 100$ nA. Now we can calculate the resistance needed for the step of 50 mV voltage drop:

$$R_{50m} = \frac{V_{50m}}{I_{div}} = \frac{50 \text{ mV}}{100 \text{ nA}} = 500 \text{ k}\Omega. \tag{4.62}$$

This value can be achieved with a reasonable size so it is suitable to consist the whole divider with resistors approximately of this value of resistance, as it is depicted in figure 4.33. The number of 500 kΩ pieces for the whole divider then is:

$$N_{R_{50m}} = \frac{V_{outmax}}{V_{50m}} = 80 \ (-). \tag{4.63}$$

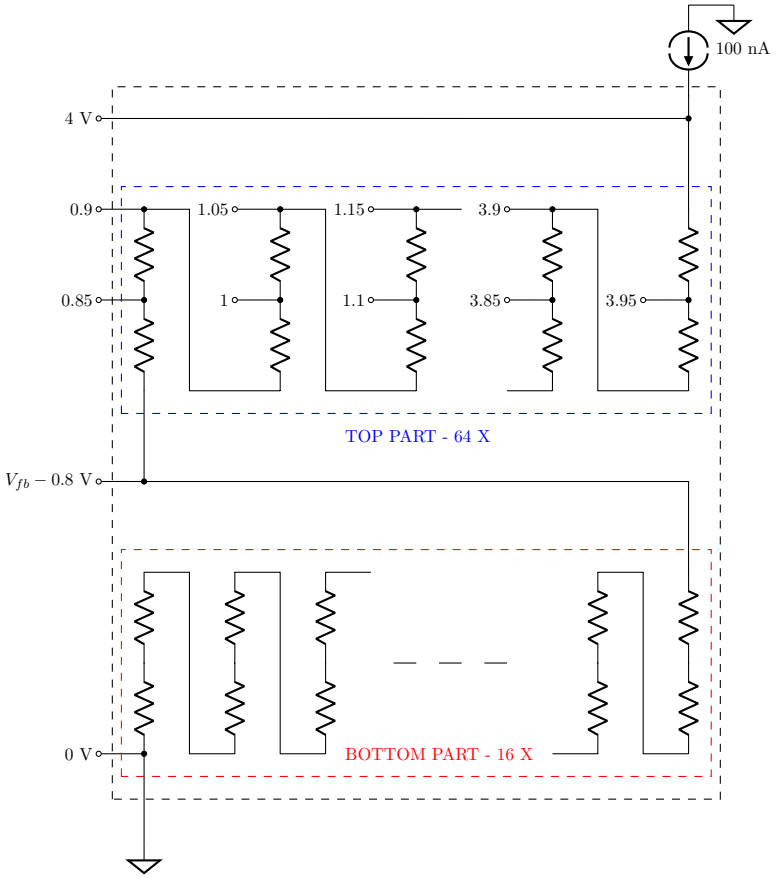


Figure 4.33: Simplified schematic of voltage feedback devider and the test-bench for its design

Besides temperature dependency and technology corners of resistors, there is another fact that must be taken into account. Real HIPO resistors have three terminals - an additional terminal is a connection to the substrate. The voltage potential between poly-Si contacts and the substrate has also an impact on the resistance - there is a voltage dependency. To compensate this effect, each resistor has to have slightly different dimensions and the goal is to tune each resistor piece to just right value to have the desired output voltage.

4.17 Frequency compensation

An active parallel frequency compensation has been used as the "main" compensation. For proper parallel frequency compensation, the output pole tracking has to be usually provided to set the proper frequency of the compensating zero - as it is mentioned in section 4.7. In this proposed compensation technique, the output pole tracking is based on an adaptive biasing current. At heavy loads - the output pole is located at highest frequencies and the adaptive biasing current is highest, on the other hand, at light loads, the output pole is located at low frequencies and the adaptive biasing current is low as well - the adaptive biasing current contains information about the output pole. Simplified structure of the designed compensation is depicted in figure 4.34.

The Theory of parallel compensation is explained in section 4.7. Capacitor C_c represents a compensation capacitor. Resistor R_c together with transistor T_2 and T_3 form a zero making resistor. Because the transistors work in liner region, they act as a voltage controlled resistor.

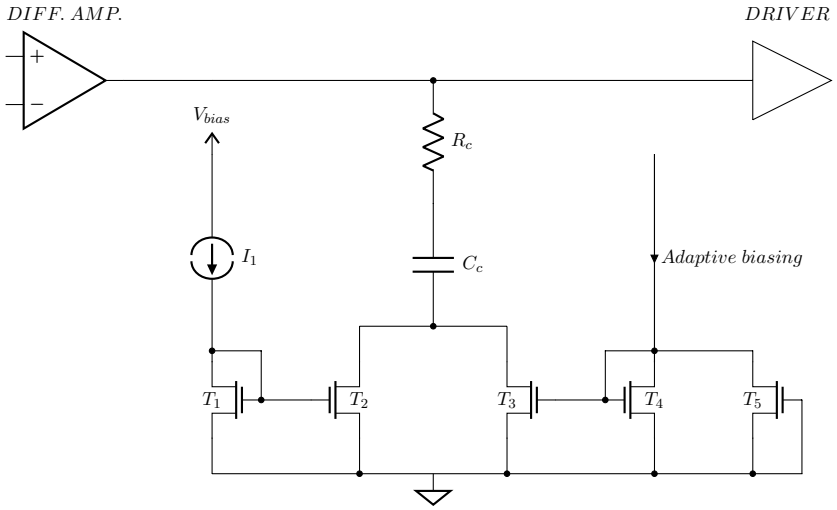


Figure 4.34: Simplified structure of active parallel compensation

- Transistor T_2 acts as a fixed value resistor, its resistance is set by the standard equation for linear region of NMOS:

$$R_{T_2} = \frac{V_{DS_2}}{I_{D_2}} \approx \frac{1}{\frac{1}{2}\mu_n C_{ox} \frac{W_2}{L_2}(V_{GS_2} - V_{th_2})}, \tag{4.64}$$

  so the resistance is determined by its $\frac{W_2}{L_2}$ ratio and V_{GS2} voltage which is determined by the current I_1 and $\frac{W_1}{L_1}$ of the diode-connected transistor T_1.

- The resistance of T_3 is given the same way, the difference is only in the variable - adaptive biasing current through T_4. Thus, higher the adaptive biasing current, the higher V_{GS4}, thus lower resistance of T_3.

- The purpose of T_2 is to set the maximal resistance. At the light or no load, the adaptive biasing current can be in an order of picoamperes, so the resistance of T_3 can reach too high values so the created compensation zero is at too low frequencies. But, since T_2 and T_3 are in parallel with each other, the maximal resistance can be adjusted by T_2.

- At heavy loads, the adaptive biasing current is quite large so the resistance of T_3 is too low and the zero would than be located at too high frequencies and there would be no effect of it. In that situation, the effect of R_c shows - the minimal resistance is determined by this resistance so the zero can be located at right frequencies.

There are several complications:

a. Setting the optimal value of the adaptive biasing current and $\frac{W}{L}$ ratios of all transistors is really difficult for accomplishing a sufficiently good phase margin under all conditions, such as temperature, technology corners and the output capacitors values.

b. At light loads, the effect of leakage current shows. Due to the leakage of transistors of the adaptive biasing circuit, the unacceptable rise of adaptive biasing current can occur. For that reason, the emphasis on minimizing the leakage current of adaptive biasing has been put, as it has been already mentioned in section of adaptive biasing 4.13. Also, for

a small compensation of the leakage current, there is transistor T_5. This transistor has the gate connected with its source terminal so it acts as a leakage generator. At high temperatures, this generated leakage current is subtracted from the adaptive biasing branch so the current through T_4 is lowered and the leakage of adaptive biasing circuit is then compensated. Although more complex compensation circuit can be implemented, their accuracy and reliability may be their weakness. Either way, the minimizing of the leakage current is a priority, compensation has only a "supporting" role.

c. In reality, the compensation capacitor does not have only two terminals but, in our case where junction capacitor based on a MOS structure is used, it has four terminals - additional connections to isolating n-well and p-well. The leakage current between the junctions may cause problems - transistors T_2 and T_3 might be bypassed by the leakage path. Thus, achieving a very large resistance needed at light loads may be difficult when the leakage current is not negligible. A special attention should be paid on the right connections of the n-well and p-well terminals to minimize the leakage current.

4.18 Saturation preventing circuit

For proper function of the active frequency compensation, the zero-making transistor has to work in linear region no matter what. During very large transient steps, under some specific conditions, because the compensation capacitor represents very low impedance for high frequencies, so the V_{DS} of compensation transistors T_2 and T_3 may potentially be sufficiently large enough for allowing the transistors to enter the saturation region and the transistors start to act as current sources which is not desirable. This situation can occur as ringing and at the output voltage during the transients.

The proposed saturation preventing circuit assures that the V_{DS} of the compensating transistors does not rise too much to cause saturation. It is based on monitoring the V_{DS}. If this voltages is about to cross a certain level, the circuit activates and clamps the V_{DS} at this value. A block schematic of the circuit is depicted in figure 4.35.

The difficulties are:

a. Setting the optimal voltage level at which the circuit activates - too high level would mean no effect of the circuit because the compensation transistor may already be in saturation region. On the other hand, too low level would cause that there would be no enough voltage headroom for a proper function of the compensation - not enough voltage headroom for a normal voltage drop across the resistor made by the liner region operating transistors.

b. Setting the right level under all conditions, especially technology corners.

c. Minimizing the leakage current from the output of the circuit to ground - leakage current may cause achieving a large resistance, needed at light loads, difficult because the compensating transistor would be bypassed by the leakage path the same way as it has been mentioned above.

This circuit also improves a recovering time at the load transient response at the step from high to low loads as plotted in figure 4.36.

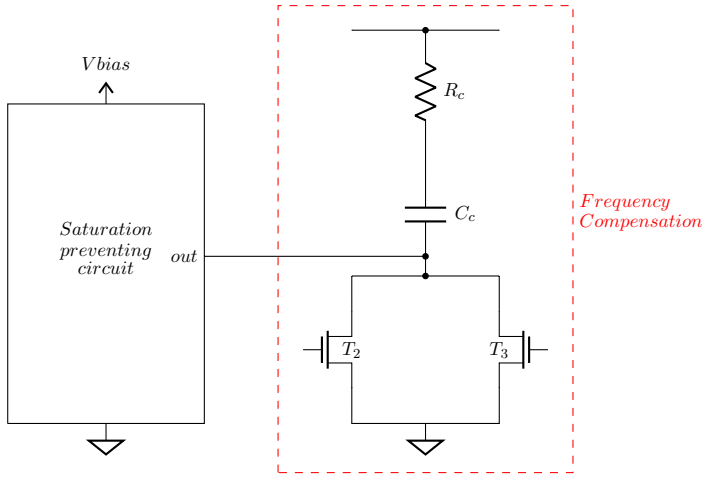


Figure 4.35: Block schematic of the saturation preventing circuit

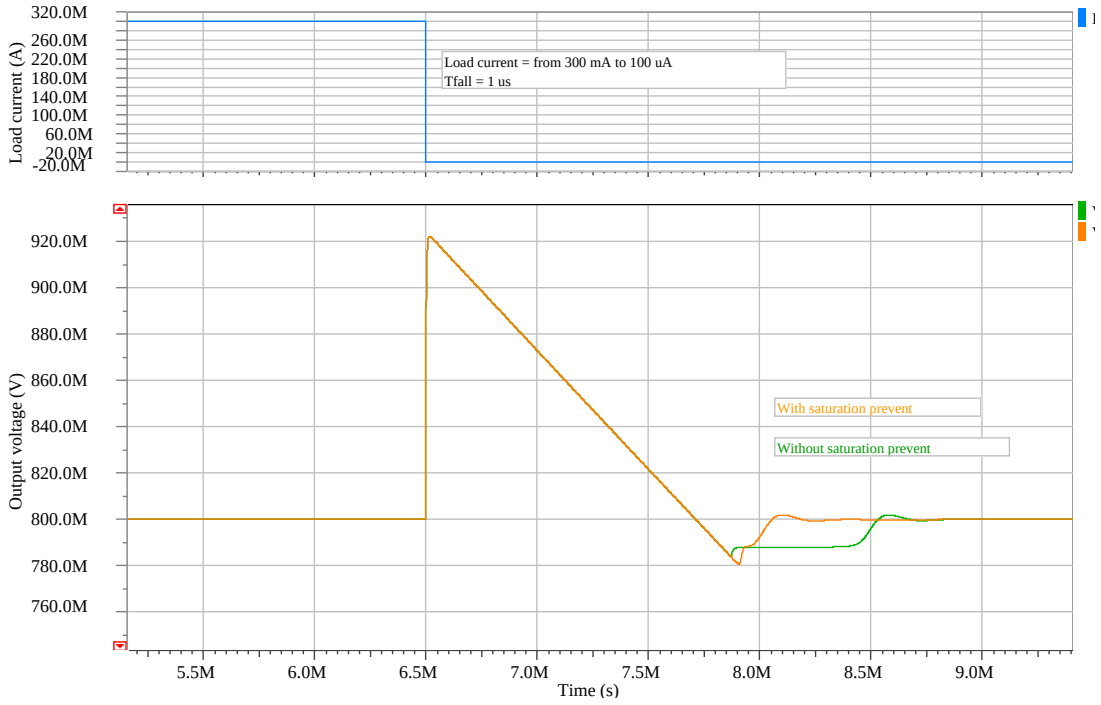


Figure 4.36: Effect of saturation preventing circuit at the step from high to low loads

4.19 ESR enhancement and feed forward compensation

The two goals of the ESR enhancement and feed forward compensation circuit are:

a. If the output voltage of the LDO is set to be $V_{out} = 0.8$ V and the band-gap reference output voltage is also at this level, the feedback voltage divider is not applicable for feed forward compensation - as it follows from the feed forward compensation theory in section 4.8 since the division ratio is 1:1. This circuit provides a mechanisms how to overcome this disadvantage and still allow using feed-forward compensation.

b. There is a requirement for allowing using output capacitors with very low ESR. The minimal allowed ESR of the output capacitor is 5 mΩ. This value is too low to have a significant effect on improving the phase margin - the theory of ESR compensation is explained in section 4.5. This circuit virtually enhances the ESR. Thus, it improves the phase margin.

The circuit is based on providing a high frequency feedback path. This path has to be separated from the DC feedback path, which assures DC voltage regulation, by a high pass filter. The high frequency path virtually enhances ESR as well as provides feed forwarding for the phase margin improvement.

4.20 Band-gap voltage reference design

A core of the proposed band-gap reference is based on a Brokaw topology. This topology has been chosen for its better precision and higher finite β error resistance than other mentioned structures. Also, as few as possible current branches for a low current consumption are necessary and emphasis has been put on simplicity of the structure. The disadvantage of a Brokaw topology of higher needed biasing voltage does not represent a significant problem since the requirement of minimal biasing voltage of the LDO is $V_{biasmin} = 2.75$ V. A simplified schematic of the proposed band-gap core is depicted in figure 4.37.

The theory of a Brokaw topology is disclosed in chapter 3.2. For determining approximate values of resistors R_1 and R_2, neglecting finite β effects, following equations can be used. Firstly, we have to know the ratio between these two resistors for independence of the band-gap voltage V_{BG} on temperature (in ideal scenario):

$$\frac{\partial V_{BG}}{\partial T} = 0 \rightarrow \frac{R_2}{R_1} = \frac{2}{2 \cdot 0.086 \cdot lnN} = \frac{2}{2 \cdot 0.086 \cdot ln8} = 5.59\ (-), \tag{4.65}$$

where N represents a ratio of emitter areas of the BJTs. In this situation, $N = 8$ has been chosen. The ratio of 1:8 is suitable for layout purposes for better matching, because a topology such as common centroid can be used. Now we have to determine the value of R_1, this values determines the PTAT current I_{PTAT} through each BJT. It has been decided to use I_{PTAT} approximately equal 50 nA at the typical technology corner and typical temperature $T = 27$ °C. The value of the PTAT current represents a trade-off between quiescent current and the output noise because low PTAT current means using large resistors, thus increased output noise. For R_1 it holds:

$$R_1 = \frac{\Delta V_{BE}}{I_{PTAT}} = \frac{V_T \cdot lnN}{I_{PTAT}} \approx \frac{0.026 \cdot ln8}{50 \cdot 10^{-}9} \approx 1\ \text{MΩ}, \tag{4.66}$$

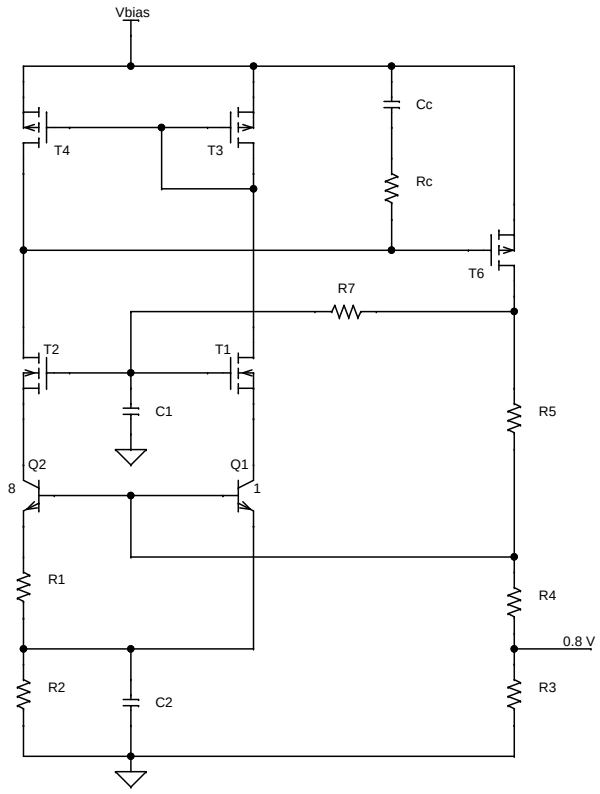


Figure 4.37: Simplified schematic of proposed band-gap core

and finally for R_2:

$$R_2 = 5.59 \cdot R_1 = 5.59 \text{ M}\Omega, \tag{4.67}$$

these values are, of course, only theoretical and in practice they need some tuning to achieve an optimal-tuned band-gap curve due to finite β and other effects.

The differences between the proposed structure and the standard Brokaw structure depicted in chapter 3.2 are following.

- In the proposed structure, the currents through each BJT are assured to be approximately equal by the current mirror consisted of transistors T_4 and T_3. The requirements for precision of this current mirror are high. This means using large channel lengths and low $\frac{W}{L}$ ratio for better layout matching and low current offset (mentioned in chapter 1).

- The base terminals of the BJTs are driven by transistor T_6 which forms a closed-loop system and regulates the bang-gap voltage. The requirement for the output voltage of the band-gap reference to be $V_{out} = 0.8$ V causes using an output voltage divider to scale-down the standard band-gap voltage $V_{BG} \approx 1.205$ V.

- Transistors T_1 and T_2 form a cascode of the BJTs. Without cascoding, the static line regulation and PSRR of the band-gap reference are poor due to not satisfyingly high Early voltage of the BJTs and thus quite low output resistance. The cascode assures low changes of V_{CE} of the BJTs with changes of the biasing voltage. Hence, it enhances their output resistance - it is multiplied by the transconductance g_m of T_1 and T_2. To

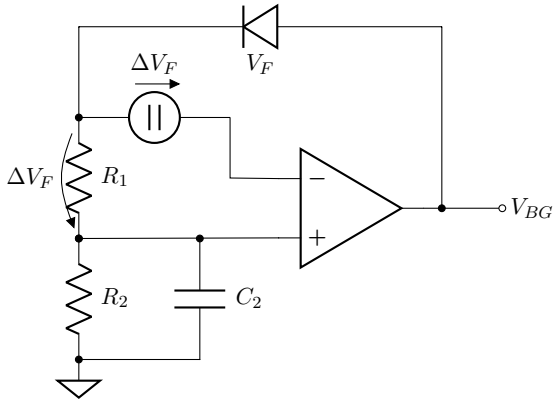


Figure 4.38: Band-gap reference model

achieve highest g_m for a given current, T_1 and T_2 have to work in weak inversion which yields in large $\frac{W}{L}$ ratio. The potential for biasing the cascode is provided by resistor R_5 which forms a voltage divider and for the cascode biasing potential $V_{cascode}$ we can then write:

$$V_{cascode} = \left(\frac{R_5}{R_3 + R_4} + 1\right) \cdot V_{BG}, \tag{4.68}$$

this voltage has to be high enough to allow V_{CE} of the BJTs to be high enough to operate the BJTs in active region, on the other hand, larger $V_{cascode}$ means higher minimal biasing voltage V_{bias}.

- Frequency compensation is needed because the band-gap core acts similarly as a two stage amplifier. Capacitor C_c and resistor R_c represent parallel frequency compensation. The theory of parallel frequency compensation is disclosed in section 4.7. There are other frequency compensation parts. The cascode of T_1 and T_2 forms a local feedback loop and it has been simulated that this local closed-loop has a negative impact on the phase margin of the whole band-gap reference loop. Hence, some kind of high frequency separating of these two loops has to done. For this purpose, there is resistor R_7 and capacitor C_1 which form a low pass filter - the DC path for biasing the cascode is not disturbed while the AC path is suppressed. For another frequency compensation, there is capacitor C_2. For easier understanding, a model of bang-gap reference is presented in figure 4.38. It can be seen that there are two feedback loops presented - negative and positive. For assuring stability, the open-loop gain of the negative feedback loop must be higher than the open-loop gain of the positive feedback loop. This is fulfilled since the open-loop gain of the positive feedback is $\frac{R_2}{R_1+R_2}$ times lower (not considering C_2). The capacitor C_2 forms a low pass filter with R_1 and R_2 with a transfer function:

$$H(j\omega) = \frac{R_2}{R_1 + R_2} \cdot \frac{1}{1 + j\omega C_2 \cdot R_1 \parallel R_2}, \tag{4.69}$$

it is obvious that the open-loop gain of the positive feedback loop gets lower with frequency. Thus, it improves the phase margin of the whole band-gap reference.

4.20.1 Output filter

The band-gap reference has to be connected to the opamp of the LDO via a low-pass filter. The purpose of this filter is:

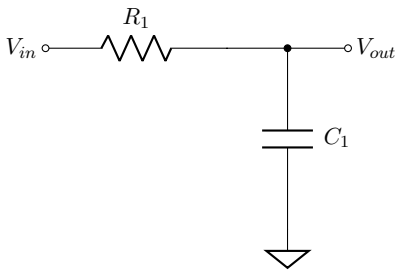


Figure 4.39: Output filter of the band-gap reference

a. Reducing the output noise of the bang-gap reference.

b. Significant improvement of PSRR of the band-gap reference.

c. Reducing undershoots and overshoots during line transient response and during start-up sequence of the band-gap reference.

d. Providing a soft-start of the LDO.

The filter has been designed as a simple 1st order RC filter, depicted in figure 4.39. The cut-off frequency for $C = 100$ pF and $R = 5$ MΩ is given as:

$$f_{cut} = \frac{1}{2\pi RC} = \frac{1}{2\pi \cdot 5 \cdot 10^6 \cdot 100 \cdot 10^{-12}} = 318 \text{ Hz}. \tag{4.70}$$

The price for the mentioned advantages of this filter is a larger area of the capacitor.

4.21 Start-up circuit of the band-gap reference

Every band-gap reference needs some kind of a start-up circuit because there are more stable operating points. Explaining can be done using again figure 4.37. Let us consider raising biasing voltage V_{bias} from zero. The output of the bang-gap reference is zero as well at this point. With raising V_{bias}, the V_{gs} of the top current mirror of T_3 and T_4 does not have to get higher, neither does the V_{BE} voltage of BJTs. In this situation, almost no current flows through the branches of the BJTs and the circuit is in a stable operating point and the output voltage is close to zero. For proper starting, there must be some way how to make either V_{GS} of the current mirror or V_{BE} high enough during the start-up sequence to get the circuit to the desired operating point - this is the goal of the start-up circuit. On the other hand, the start-up circuit must not affect the behavior of the band-gap reference at the normal steady mode.

Several start-up circuits have been simulated, the best result have been achieved with a circuit depicted in figure 4.40. The function of this circuit is following:

- Transistors T_5 and T_8 act as a current comparator. If the current set by the current source I_1 and a current mirroring ratio of transistors T_7 and T_8 (let us call it I_a) would be lower than the current given by a mirroring ratio of T_5 and T_4 and I_{PTAT} current through the branches of the BJTs (let us call it I_b), transistor T_5 would be forced into linear region and the potential V_{comp} would be close to V_{bias} and the current I_a would flow through T_8 and T_5. In the other situation, if I_b would be lower than I_a, T_8 would be forced into liner region and V_{comp} would be close to zero and the current I_b would flow through T_5 and T_8 - "lower current wins".

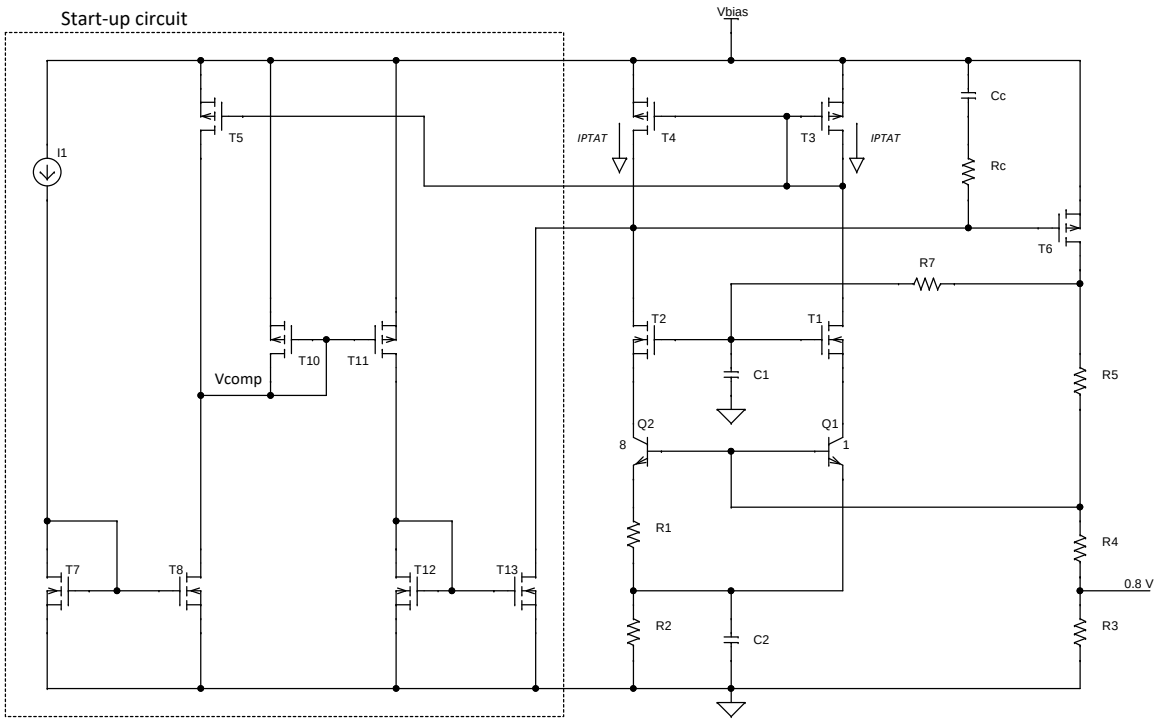


Figure 4.40: Simplified schematic of the stat-up circuit

- At the very first moment of the starting sequence, there is almost no current flowing through the band-gap core, so the V_{comp} reaches zero because $I_a > I_b$. V_{comp} close to zero forces V_{gs} of diode connected transistor T_8 to increase and a current starts to flow through current mirrors of T_8 - T_{11}. This current charges the capacitor C_c so that V_{gs} of regulation transistor T_6 increases and this transistor pushes the base voltage of the BJTs to increase as well and the circuit enters the normal operating point.
- As the circuit is reaching the normal operating point, the current through the BJT branches increases, when this current crosses a certain level when $I_b > I_a$, V_{comp} gets close to V_{bias} and only a leakage current flows through T_8 - T_{11} and the start-up circuit is then deactivated.

Several issues are presented:

a. Choosing values of comparing currents I_a and I_b brings difficulties. For the sake of simplicity, let us assume that the current mirroring ratio of T_7 and T_8 is 1:1 as well as T_5 and T_4. For a proper start-up, I_1 must be always lower that I_{PTAT} in the normal operating point. If this was not fulfilled, the start-up circuit would never be deactivated. On the other hand, choosing I_1 too much lower, the start-up circuit may be deactivated too early and even though the band-gap reference would potentially enter the normal operating point, the output voltage during the start-up sequence may not look "nice".

b. The band gap core is sensitive to a leakage current of T_{11} which can cause worsening of PSRR and untuning of the band-gap curve. For that reason, a dummy leakage compensation transistor should be connected to the other branch of the band-gap core to make it symmetrical as much as possible.

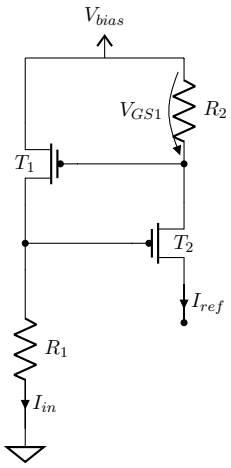


Figure 4.41: Current generator for the start-up circuit

c. The most important thing is to provide a sufficient current generator for the start-up circuit, Which is difficult when it comes to low quiescent current.

4.21.1 Current generator for start-up circuit

The low quiescent current requirement brings complications when designing a current generator for the stat-up circuit. Generating a current in the order of nanoamperes may lead to using large resistances in order of hundreds of megaohms which may occupy a significant area on a chip. A good way how to generate a current only in an order of nanoamperes would be using depletion region transistors, unfortunately, they are not available in BCD8 technology. Other approach has been using a leakage current generator, but the generated current has shown a high spread under different technology corners, temperature range and biasing voltage. Finally, it has been decided using a resistor for a generation of the current at the expense of larger occupied area. On the other hand, using high-ohmic polysilicon (HIPO) resistors with large sheet resistence does not require so large area. The final solution is depicted in figure 4.41. The input current I_{in} is given as:

$$I_{in} = \frac{V_{bias} - V_{SG1} - V_{SG2}}{R_1}, \tag{4.71}$$

and for the output reference current I_{ref} it holds:

$$I_{ref} = \frac{V_{SG1}}{R_2}, \tag{4.72}$$

from this, assuming temperature independence of R_2, the temperature dependence of I_{ref} is given by temperature dependence of V_{GS} which is mostly given by temperature dependence of the threshold voltage V_{th}. To reduce the dependence of I_{ref} on biasing voltage V_{bias}, transistor T_1 should be operated in weak inversion - the changes of V_{GS1} are then lowest when I_{in} changes, so the ratio $\frac{W}{L}$ of T_1 should be high enough. Operating T_1 in weak inversion also means lowest V_{GS} for given I_{in}, this allows using smaller values of R_2 which is suitable since R_2 can reach high values of resistance when generation of I_{ref} in order of nanoamperes is needed.

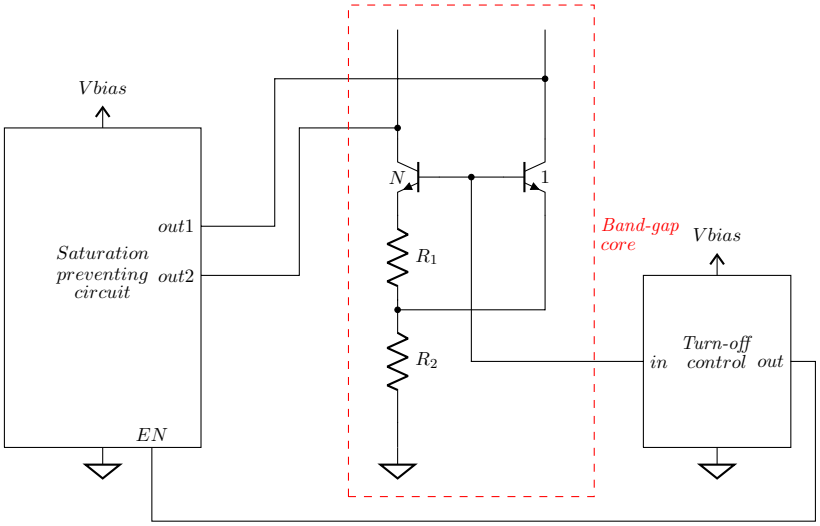


Figure 4.42: Block schematic of saturation preventing circuit for BJT

4.22 Saturation preventing circuit for BJT of the band-gap reference

Especially during longer rising of V_{bias} during the start-up sequence, there is an issue with BJT transistor Q_2 in figure 4.40 entering saturation. This represents a very common problem with BJTs. The drain potential of T_4 is pushed lower, so is the collector potential of Q_2. It may happen that the collector potential gets lower than the base potential and Q_2 is forced to enter saturation. This occurs as an undesirable glitch of the output voltage during the start-up sequence - the output voltage does not follow rising V_{bias}.

The proposed circuit prevents Q_2 from entering saturation. The circuit uses a fact that Q_1 does not enter saturation since T_3 is diode-connected and does not allow the collector potential of Q_1 to get lower than the base potential. The proposed circuit tries to keep the collector potential of Q_2 at the level of Q_1. A block schematic of the circuit is depicted in figure 4.42. Because the fact that the circuit is connected to collectors of the BJTs, PSRR of the band-gap reference may be potentially affected and a little worsening of PSRR may occur. For that reason, the circuit is made inactive after the start-up sequence. The complication is that the circuit needs to be disconnected somewhat later than the start-up circuit for preventing the saturation during the whole start-up sequence. Therefore, another mechanisms how to turn off the circuit have been implemented. The turn-off logic is based on monitoring the output voltage.

The effect of the saturation preventing circuit is shown in figure 4.43 where a slow start-up sequence is presented.

4.23 Voltage dividers and trimming circuit for the band-gap reference

For designing the voltage divider for the band-gap core and the output divider of the bang-gap reference, the same rules as have been mentioned in section 4.16 regarding the output feedback divider of the LDO have to be taken into account. The divider of the band-gap core should be consisted of a type of resistors with better matching. Therefore, diffused resistors

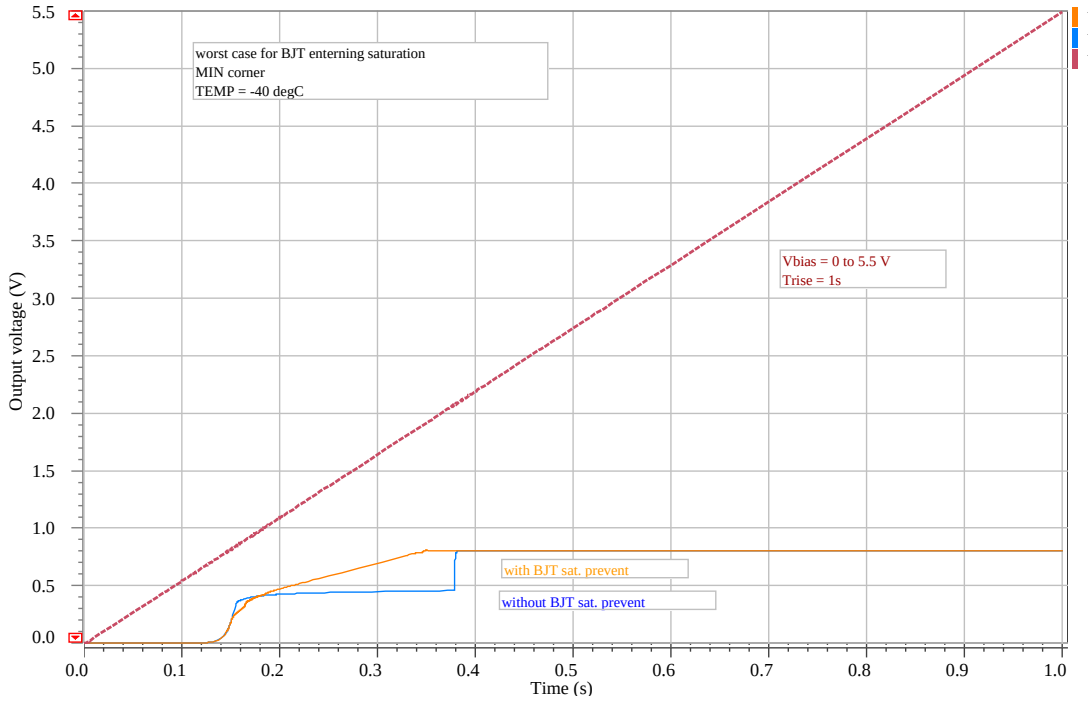


Figure 4.43: Effect of BJT saturation preventing circuit at slow start-up

which show good matching in BCD8 technology have been used, at the expense of grater occupied area comparing to HIPO resistors which have been used for the output divider of the band-gap reference.

Unfortunately, technology corners and non-ideal matching cause that a bang-gap reference with better precision of the output voltage than ±2% cannot be usually achieved without using trimming circuits. For that reason, a 4-bit trimming circuit of the band-gap core divider have been designed. The lower part of the divider (R_2 in figure 4.37) has to be trimmed because the upper part determines the PTAT current. A simplified structure of the trimming circuit is depicted in figure 4.44. The logic is based on two's complement. The band-gap curve in simulations is tuned when resistors R_2 - R_4 are shorted. Resistor R_1 can cause the output voltage to move by -4%, at typical temperature $T = 27$ °C, if shorted. If R_1 is shorted and R_2 - R_4 are not shorted, the difference of the output voltage is -0.5%. Similar way, we can change the output voltage in a range of ±4% by a 0.5% step. Transistors T_1 - T_4 act as switches and together with inverters they provide the trimming logic. A truth table of this logic can be found in table 4.4.

4.24 Current generator

A current generator with a reference current $I_{ref} \approx 50$ nA is needed for biasing all blocks. It has to show sufficiently large independence on biasing voltage and temperature. On contrary of band-gap voltage references, generation of a precise current is not possible. The maximal achievable precision of a reference current generator is usually approximately ±20% to ±30%.

Table 4.4: Truth table of the trimming circuit

D	C	B	A	ΔV_{BG}
0	0	0	0	0 %
0	0	0	1	+0.5%
0	0	1	0	+1%
0	0	1	1	+1.5%
0	1	0	0	+2%
0	1	0	1	+2.5%
0	1	1	0	+3%
0	1	1	1	+3.5%
1	0	0	0	-4%
1	0	0	1	-3.5%
1	0	1	0	-3%
1	0	1	1	-2.5%
1	1	0	0	-2%
1	1	0	1	-1.5%
1	1	1	0	-1%
1	1	1	1	-0.5%

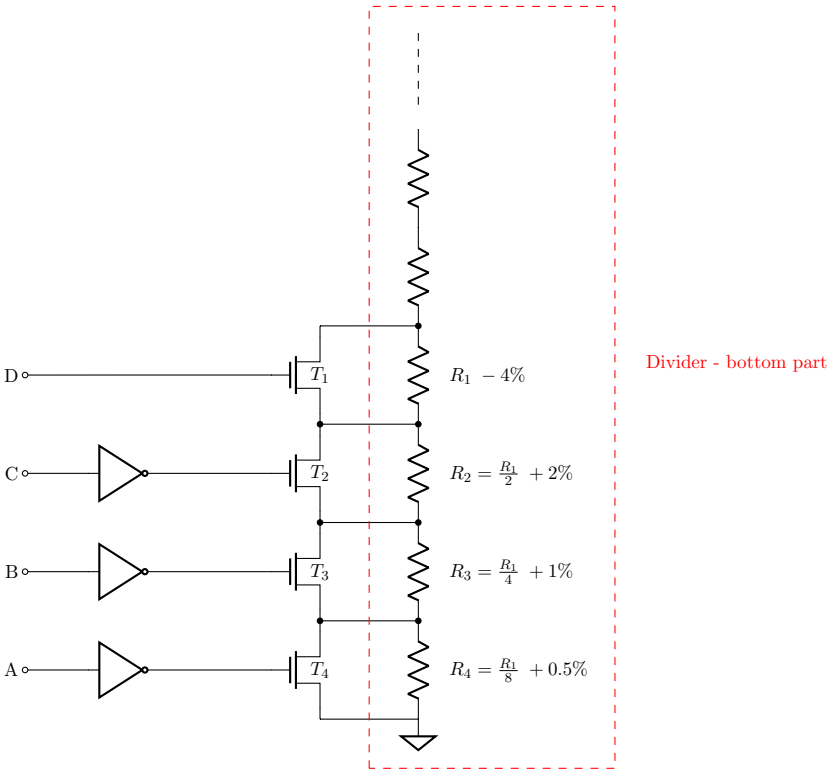


Figure 4.44: Simplified schematic of the trimming circuit

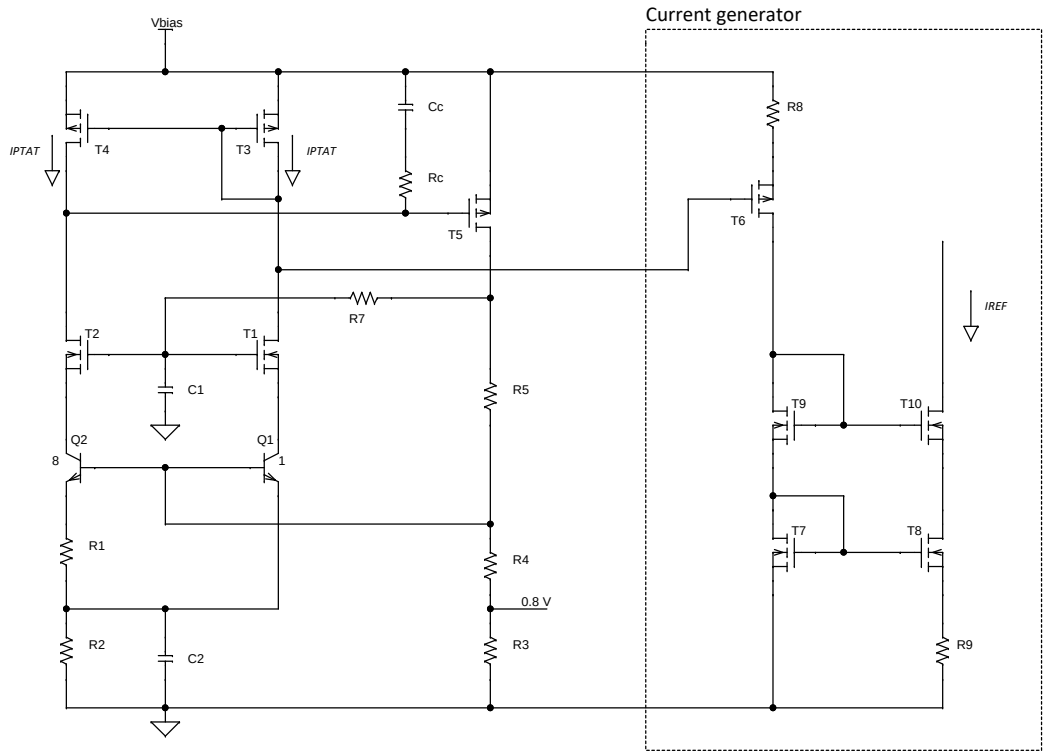


Figure 4.45: Simplified schematic of the current generator

Also the low quiescent current requirement restricts freedom of design.

Several typologies have been designed and simulated, the best results have been obtained with a structure shown in figure 4.45. It is based on the PTAT current which is already generated by the band-gap reference core. The PTAT current is then mirrored by source degenerated current mirrors. The purpose of source degenerating is for using temperature dependency of used resistors for compensation of the positive temperature coefficient of the PTAT current. The resistors have to have positive temperature coefficient as well. With increasing temperature, the resistance increases as well and the V_{GS} of source degenerated transistors decreases. Thus, the mirrored current is reduced and the overall temperature dependency of I_{ref} is then reduced as well. Resistors with highest positive temperature coefficient in BCD8 technology are n-well resistors. The disadvantage of this type of resistors is its lower sheet resistance and wider minimal width. The cascode of transistors T_9 and T_{10} increases the output resistance of the current mirror of T_6 and T_7 and reduces dependency of I_{ref} on biasing voltage V_{bias}. The simulation results of the proposed current generator, dependency of I_{ref} on temperature and biasing voltage V_{bias} for different technology corners, are shown in figure 4.46.

4.25 Thermal protection

The LDO needs to be protected against high temperature of the environment and against high temperature of the power stage caused by dissipating enormous power. If the temperature crosses a certain level, the opamp of the LDO needs to be disabled and if the temperature decreases under another level, automatic enabling has to be ensured.

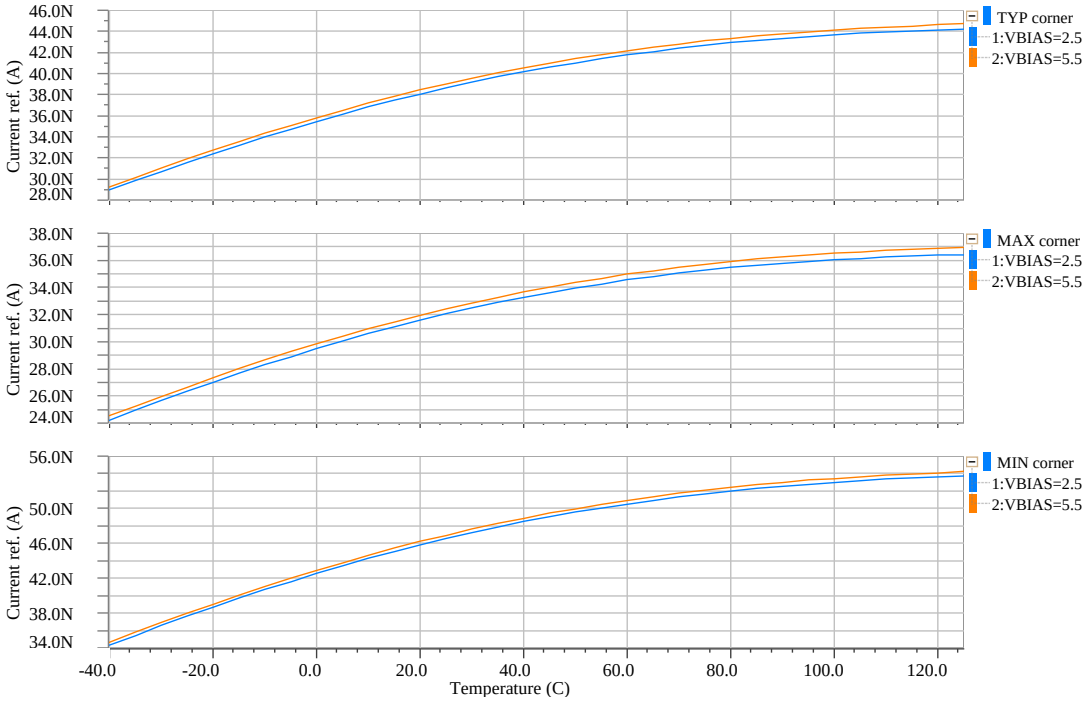


Figure 4.46: Simulation results of the current generator

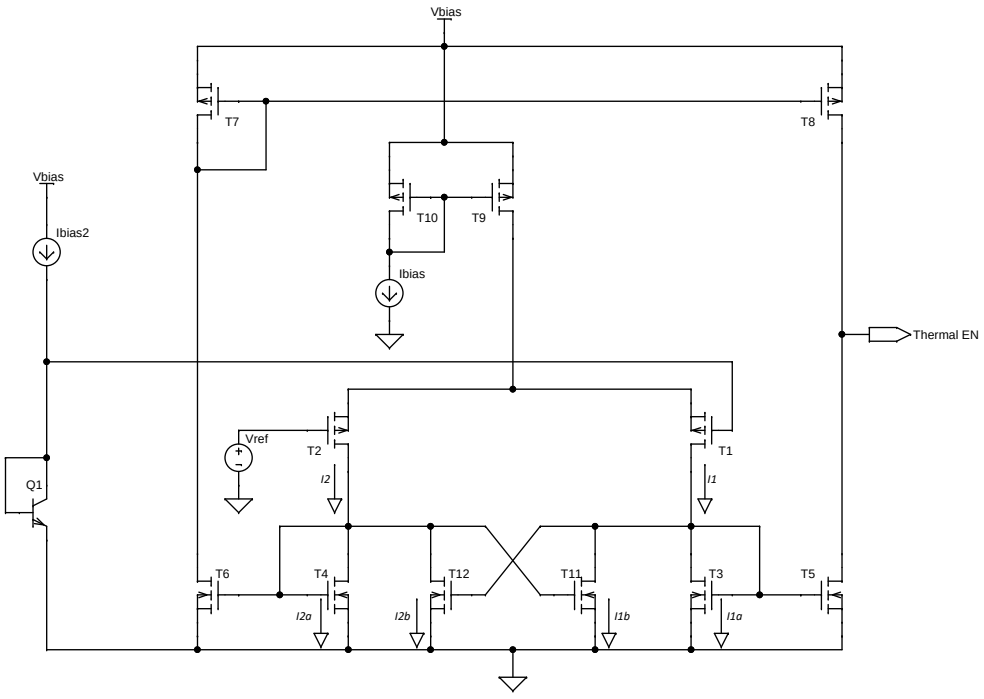


Figure 4.47: Simplified schematic of the thermal protection circuit

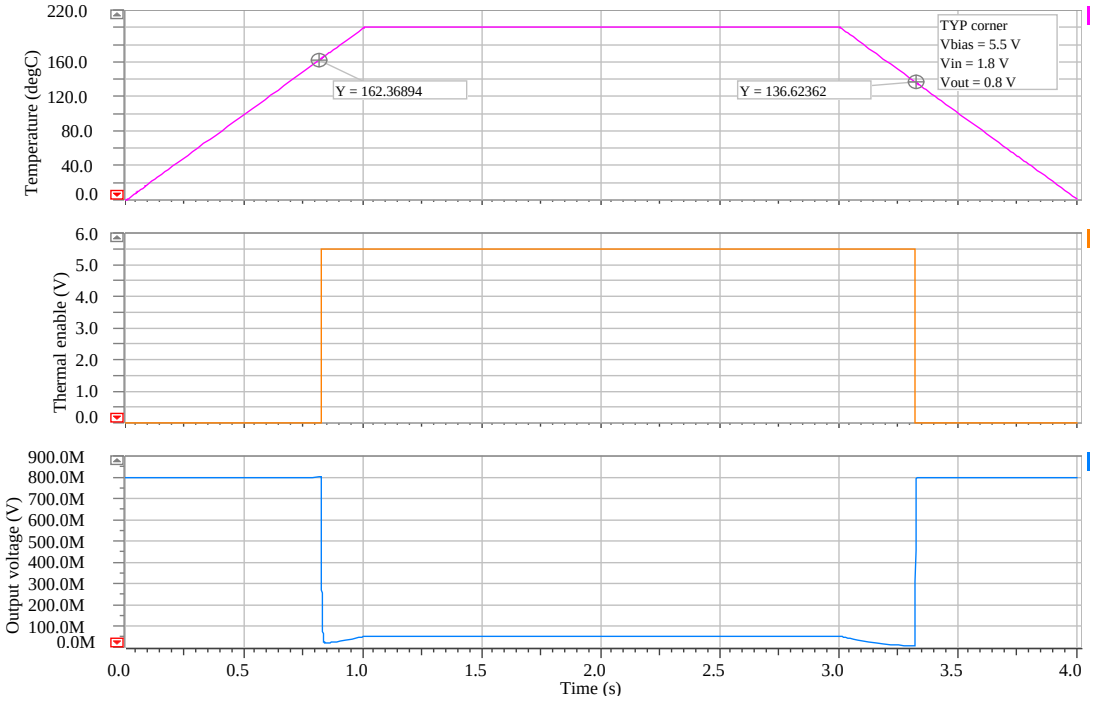


Figure 4.48: Simulation results of the thermal protection circuit

The first step of designing a thermal protection is choosing a suitable thermal detector. Almost every device has its temperate dependence and potentially can be used as a thermal detector, but the problem is its reliability. Using thermal dependence of forward-biased PN junction seems as the best choice. Forward-biased voltage V_{FB} of a PN junction decreases approximately with -2 mV/°C for a constant biasing current. A thermal protection can be based on sensing V_{FB} and comparing it with a certain voltage level, if the temperature crosses its upper limit value, V_{FB} decreases under the comparing voltage level and then other circuitry provides a thermal enable logic signal to switch its logical level. This signal then controls the enable logic of the opamp. The proposed thermal protection is depicted in figure 4.47. Beside the thermal detector of the diode-connected BJT Q_1, the thermal protection circuit includes a comparator which detects if the level of V_{FB} is below the reference voltage V_{ref}. Some hysteresis of the comparator is necessary to prevent unwanted enabling/disabling if the temperature oscillates around the threshold level.

- The comparator is based on a similar structure as the structure used for the differential amplifier of the LDO. The difference is that the comparator works in an open-loop. Not considering transistors T_{12} and T_{11} and assuming a mirroring ratio of all current mirrors 1:1 and symmetrz of the structure, transistors $T8$ and $T5$ compare the currents I_1 and I_2. Assuming infinite gain, if $I_1 > I_2$, the output of the comparator is at the low rail and vice versa, this means that there is no hysteresis and the comparing currents are only given by the voltage at the inputs. The needed hysteresis is provided by transistors T_{11} and T_{12}. For proper hysteresis, it must hold $\frac{W_{11}}{L_{11}} > \frac{W_3}{L_3}$ respectively $\frac{W_{12}}{L_{12}} > \frac{W_4}{L_4}$. The comparing currents are now $I_{1a} = I_1 - I_{1b}$ and $I_{2a} = I_2 - I_{2b}$. Before the comparator

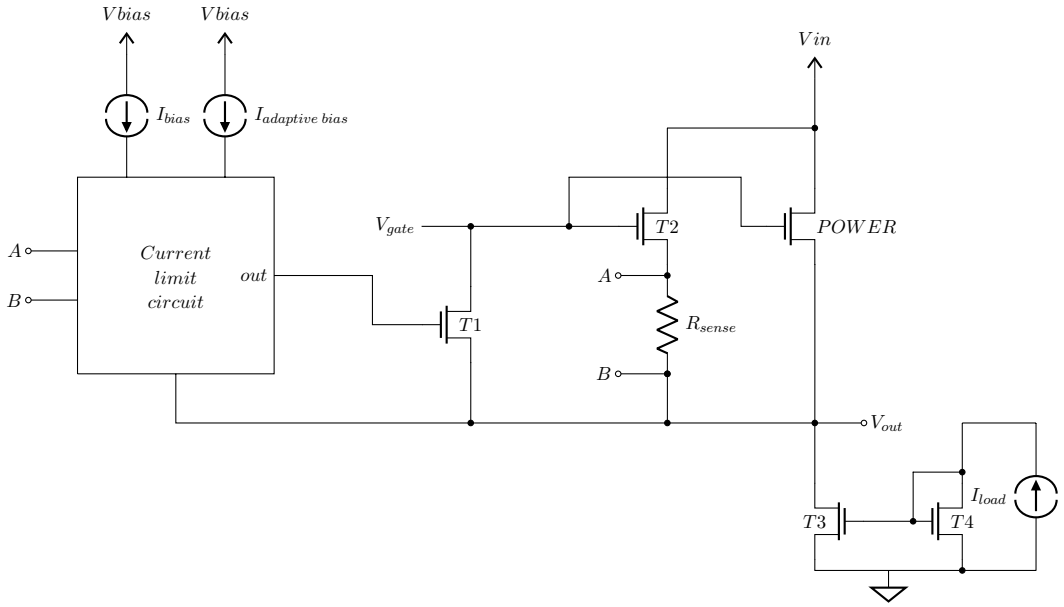


Figure 4.49: Block schematic of the current limiting circuit

can switch the output level when I_{1a} starts to be greater than I_{2a} or vice versa, the current I_1 respectively I_2 must firstly "cross the level" of I_{1b}, I_{2b} respectively. This assure the hysteresis.

- The reference voltage V_{ref} is supplied from the output band-gap reference voltage divider.
- The output of the comparator is connected to the enable control circuitry of the opamp.

Simulation results are shown in figure 4.48. First plot shows the temperature waveform, the second waveform represents the output signal of the thermal protection (In fact, this signal is inverted comparing to figure 4.47 for other purposes of the enable control logic.). The third waveform represents the output voltage of the LDO. It can be seen that the thresholds levels of the thermal protection have been set to approximately 162 °C when increasing and 137 °C when decreasing temperature. The non-zero level of the LDO output voltage is caused by a large leakage current at the temperatures around 200 °C.

4.26 Current limiting circuit

The LDO has to be protected against crossing the maximal allowed load current. In a normal situation, LDO regulators act as a voltage source. When the maximal load current is reached, the operation of the LDO has to switch to acting as a current source - sourcing still the maximal allowed current even if the load resistance still decreases. In this situation, the output voltage of the LDO has to decrease to keep the maximal current with the decreasing load resistance.

A block schematic of the proposed current limiting circuit is depicted in figure 4.49. The circuit is based on sensing a voltage-drop across resistor R_{sense}. This resistor is a part of a special current branch through the copy transistor T_2 which is flown by a part of the load current. So the voltage drop across R_{sense} is proportional to the load current. If the load

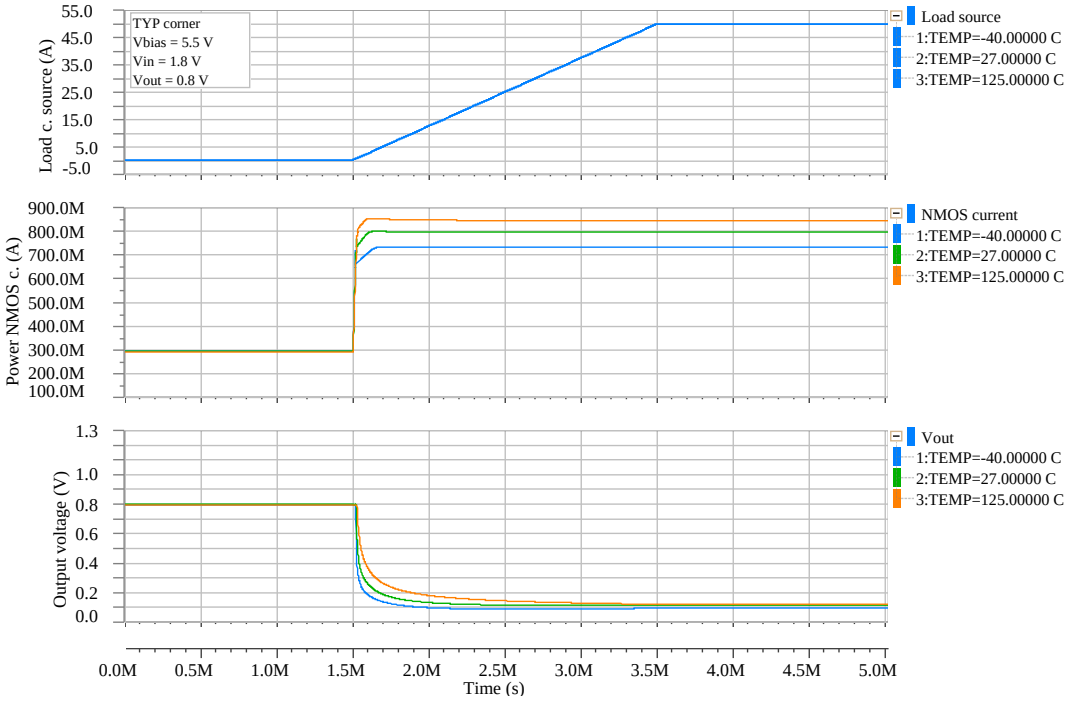


Figure 4.50: Simulation results of the current limiting circuit

current reaches the maximal allowed level - the voltage drop across R_{sense} reaches its threshold level and the current limiting circuit activates and drives transistor T_1 which regulates the gate voltage of the power NMOS, so that the current through the power transistor is kept on the maximal level until the load resistance increases back, so the load current is lower than the threshold. Therefore, if the circuit is activated, it forms another regulating loop which "fights" with the main regulation loop. It has been simulated, that higher biasing current improves the transient response of the current limiting circuit. For that reason, adaptive biasing has been used since there is the ultra-low quiescent current requirement at no load condition.

The simulation results of the current limiting circuit are shown in figure 4.50. For the load of the LDO, a current mirror has been used as it is depicted in figure 4.49. The first plot represents the waveform of the loading source I_{load}. Second plot represents the current through the power NMOS and the last plot is the output voltage of the LDO. The maximal current level have been set around $I_{max} \approx 800$mA, but there is some spread presented. It is caused mainly by the temperature dependence and the technology corner spread of the sensing resistor R_{sense}.

4.27 Enable control

If the operation of the LDO is not needed in a system. Its current consumption should be almost zero ideally. This is the purpose of the enable control logic. If the function of LDO is not enabled, the consumption of the LDO should be only given by leakage currents plus

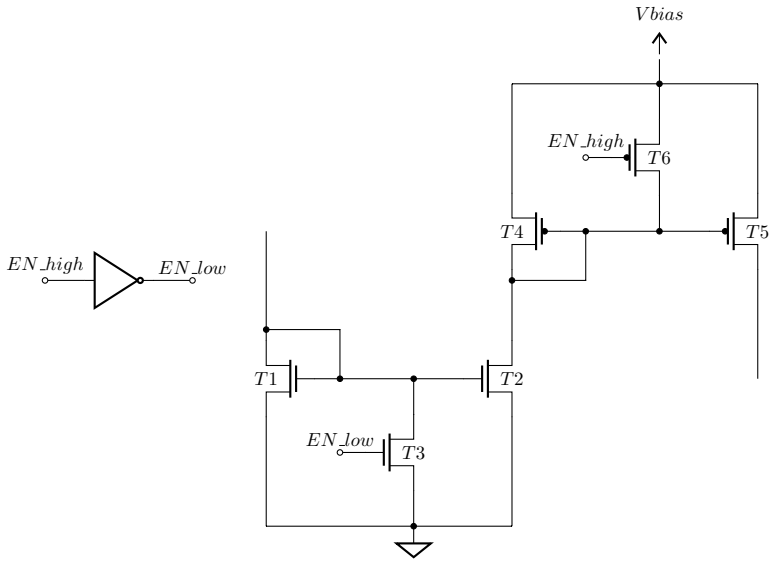


Figure 4.51: Principle of enable control switching nodes

potentially consumption of a circuit which defines a voltage threshold of the enable signal logic. To assure that, every node, of which the voltage potential may potentially generate some current, needs to be switch to ground or to biasing voltage - as it is shown in figure 4.51 for a case of NMOS and PMOS current mirrors. Transistors T_3 and T_6 work as switches controlled by the enable logical signal. For not disturbing the normal operation of the LDO, the channel lengths of the switching transistors should be longer than minimal and their $\frac{W}{L}$ ratio should be small enough to reduce their leakage current in their off-state which may cause problems especially to the nodes highly sensitive to the leakage.

There are several issues:

a. The opamp of of the LDO can be disabled either by the total enable logic or by the thermal protection. If the temperatures rises above the threshold and the thermal protection switches its output signal, only the opamp has to be disabled because the thermal protection requires the band-gap reference operation for generating the comparing voltage reference. The easiest logic can be implemented without a need of logical gates. The total enable signal controls the band-gap reference and the thermal protection while the opamp is controlled by the output signal of the thermal protection.

b. To assure a soft-start after the thermal logic enables the the opamp, the capacitor of the band-gap output filter has to be discharged before the the thermal enable signal activates the opamp. Discharging mechanism have been implemented to assure that.

c. Some nodes of the opamp are tied to V_{in} while the enable logic is tied to V_{bias}. Therefore, some level-shifters to cover this issue have been designed.

The simulation results of the current consumption of the LDO while disabled are shown in figure 4.52. The first plot represents a current consumption which is supplied from V_{bias} while the other plot represents the total consumption from both V_{bias} and V_{in}. It is obvious that at high temperatures, the leakage current occurs. The high leakage current in the second plot is caused by the leakage current of the power NMOS which is tied to V_{in}.

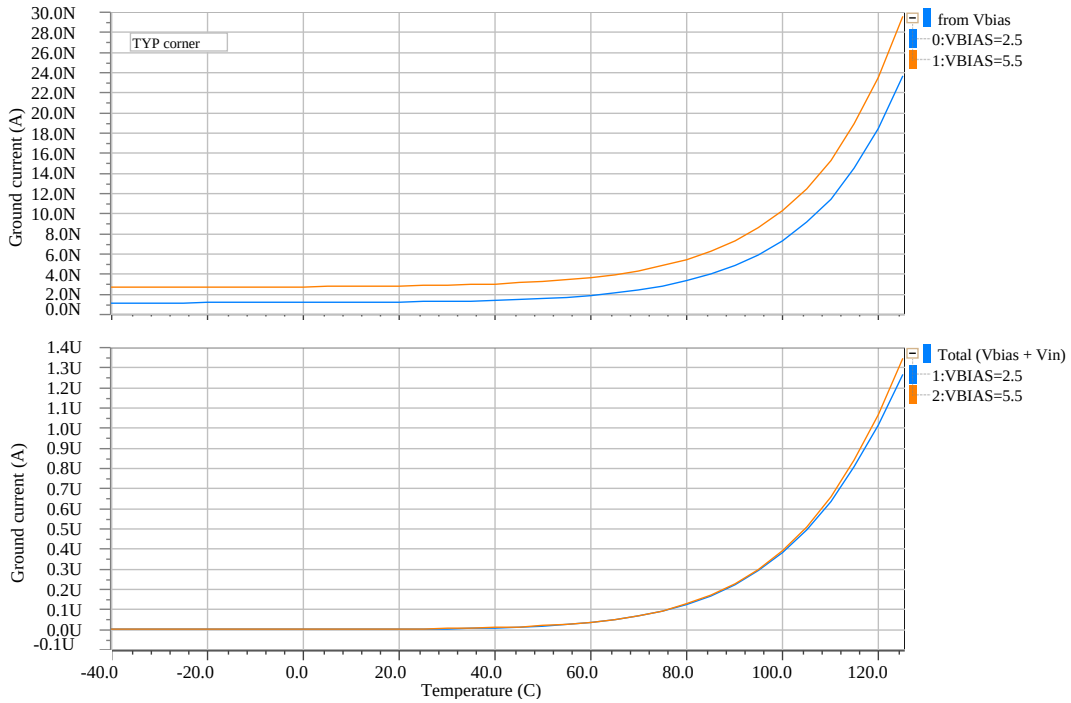


Figure 4.52: Simulation results of the current consumption while the LDO is disabled

Chapter 5

Simulation results

In this chapter, the most important simulation results of the designed LDO are presented. A large amount of simulations has been done under a lot of different conditions. For the sake of illustrative matter, not all results are stated here. For further reading, technology corner MAX represents the case where transistors have the lowest threshold voltage and resistances have the largest values. On the other hand, technology corner MIN means the highest threshold voltage and the lowest resistances. Typical values are represented by TYP technology corner. Operating range values of the LDO are shown in table 5.1.

5.1 Output voltage accuracy

Monte Carlo simulation of the output voltage has been done. Number of Monte Carlo runs has been set to 350. Biasing voltage V_{bias}, input voltage V_{in} and temperature have been set to change according the uniform distribution - ranges are stated in the figures. Four figures are presented 5.1 - 5.4, for the output voltage of $V_{out} = 0.8$ V and $V_{out} = 0.4$ V. For each value of the output voltage, there is the option of the complete chip with the band-gap reference connected, the other figure represents the situation when the ideal voltage source is connected instead of the band-gap reference (BGR). This assures better distinguishing of the output voltage inaccuracy sources. It can be seen that the accuracy of the LDO without the BGR is around ±0.5% for 4-sigma (standard deviations). The accuracy with the connected BGR is less than ±2% without using the trimming circuit of the BGR - this represents a common precision of ban-gap references without trimming.

Table 5.1: Operating range values

	min	max	unit	note
V_{bias}	2.5	5.5	V	$V_{bias} \geq V_{out} + 1.5$, for full load range
V_{in}		5.5	V	$V_{in_{min}} = V_{out} + drop$
V_{out}	0.8	4	V	50 mV step
I_{load}	0	300	mA	
$I_{load_{lim}}$	≈ 800		mA	
T	-40	125	°C	
T_{lim}	160		°C	hysteresis, up 160 °C, down 135 °C

5.2 Temperature dependence of the output voltage

The temperature dependence of the output voltage is shown in figure 5.5 for the situation with and without the BGR. It is obvious that the dependency is given by the temperature dependency of the bad-gap reference - standard band-gap curve. Theoretical possibility of compensation of the band-gap curve is adding up a resistor with a different temperature coefficient into the band-gap core divider. Unfortunately, this option works online theoretically, because in practice, adding a different type of resistor causes unacceptably higher spread of the output voltage in Monte Carlo simulations. The output voltage of the LDO without the BGR is nearly constant regardless of temperature.

5.3 Quiescent current

The temperature dependence of the quiescent current at no load is depicted in figure 5.6 for TYP, MAX and MIN technology corner. In each plot there are three waveforms. The green waveform represents the total quiescent current of the whole chip. The orange waveform shows the quiescent current which is supplied from V_{bias} source and not considering the consumption of the BGR. The blue waveform represent the quiescent current of the band-gap reference. Step increase of the green line at high temperatures is caused by the leakage current of the power NMOS transistor which is supplied from V_{in}. Normally, the quiescent current from V_{in} is only about 100 nA which flows through the feedback voltage divider. Overall, it can be said, that the total quiescent current of the LDO without the BGR is around $I_q \approx 900$ nA at the room temperature and typical technology corner, the total quiescent current of the whole chip including the band-gap reference is about $I_q \approx 1.2$ μA. Thus, the requirements have been fulfilled.

The dependency of the total quiescent current of the whole chip on the load current is shown in figure 5.7. It can be seen that the maximal quiescent current at the maximal load current is $I_q \approx 100$ μA.

5.4 Line and Load regulation

Line regulation - versus changes in V_{in} for various load currents is depicted in figure 5.8. The static line regulation - versus changes in V_{bias} is shown in figure 5.9.

Static load regulation for various combinations of V_{bias} and V_{in} is depicted in figure 5.10 for $V_{out} = 0.8$ V, and in figure 5.11 for $V_{out} = 4$ V. The little "wave" at higher loads in figure 5.10 is related to the adaptive biasing of the differential amplifier and used cascodes for the gain improvement.

5.5 Power supply rejection ratio - PSRR

PSRR versus V_{in} and V_{bias} for various load current is depicted in figure 5.12 for $V_{out} = 0.8$ V and in figure 5.13 for $V_{out} = 4$ V. With higher load current, the UGF of the LDO is higher as well, thus higher PSRR at higher frequencies. The frequency of the peak of waveforms is approximately determined by the UGF. Improvement of PSRR at higher frequencies than the peak is caused by the output capacitor of the LDO which acts as a low pass filter and its influence is dominant. For very low load currents, the UGF is low, however, the output impedance which is given by the load resistance and the output capacitor is higher. Thus,

the effect of the output capacitor is dominant even at low frequencies - as it can be seen on the green waveform in the figures.

PSRR for $V_{out} = 4$ V is a little worse than for $V_{out} = 0.8$ V. This behavior is normal and makes sense - the open-loop gain of the LDO is lower because of the feedback voltage divider.

At very high frequencies, the effect of Miller capacitance of power transistors of NMOS based LDOs can show. The drain (input voltage) of a NMOS power transistor is coupled with its gate terminal via Miller capacitance, this may cause some worsening of PSRR at very high frequencies. Improvement of PSRR at high frequencies can be based, for instance, on [16]. This principle have been tried, but finding an optimal value of the compensation capacitor was difficult in this particular case and possible worsening of PSRR under some conditions occurred.

5.6 Transient load and line regulation

Transient load regulation for the fast large step in the load current - from 100 μA to 300 mA and back for $V_{out} = 0.8$ V is shown in figure 5.14 and for $V_{out} = 4$ V in figure 5.15. The slow getting back to the nominal value of the output voltage after the step from high to low current is normal - after the overshoot happens, the output capacitor starts to discharge. The load resistance is now very high - the load current is only 100 μA so the time constant of discharging is high as well. With higher load current, the discharging is faster.

Transient line regulation versus V_{bias} for various load currents is depicted in figure 5.16 and versus V_{in} in figure 5.17.

5.7 Drop-out voltage

The dependency of the drop-out voltage on the load current is shown in figure 5.18 for $V_{out} = 0.8$ V for TYP and MIN technology corner and various combination of V_{bias} and temperature. For $V_{out} = 4$ V, the results are in in figure 5.19. The worst case is for MIN corner and for the maximal temperature as it has been mentioned in section 4.12 and also for $V_{out} = 4$ V, because in this situation, the headroom for V_{GS} of the power NMOS is the lowest. It can be seen that even at the worst case, the drop-out voltage is around 60 mV. Under typical conditions, the drop out voltage is less than 35 mV even for $V_{out} = 4$ V. Thus, the requirement of 150 mV has been overcome by far. That low value puts the LDO into the ultra-low drop-out category. The measurement of the drop-out voltage is done under condition when the output voltage is 100 mV below the nominal level, thus 700 mV, respectively 3.9 V.

5.8 Output noise

The output noise spectral density is shown in figure 5.20 for various load currents and for $V_{out} = 0.8$ V and $V_{out} = 4$ V. It can be seen that with higher output voltage, the noise increases. This behavior is standard, it is caused by the voltage feedback divider resistors and their thermal noise. With higher output voltage, the resistance of the divider increases, thus increased noise. Effective values of the output noise are shown in table 5.2.

Table 5.2: Output noise

conditions:	$T = 27°C$, $I_{load} = 10$ mA, 10 Hz - 100 kHz
$V_{out} = 0.8$ V	51 μV
$V_{out} = 4$ V	188 μV

5.9 Frequency characteristics

Frequency characteristics - gain and phase of the open-loop transfer function for various load current are depicted if figure 5.21. The dependency of the phase and gain margin on the load current for different temperatures is depicted in figure 5.22. The LDO is stable under all conditions using the output capacitor with the recommended minimal value of 500 nF, the maximal value is not restricted. The ESR should be in the range of 5 mΩ - 1 Ω.

5.10 Start-up

Start-up sequence for two scenarios of the enable signal, V_{bias} and V_{in} is shown in figures 5.23 and 5.24. The start-up sequence of the reference voltage - start up sequence of the band-gap reference is depicted in figure 5.25 for various temperatures.

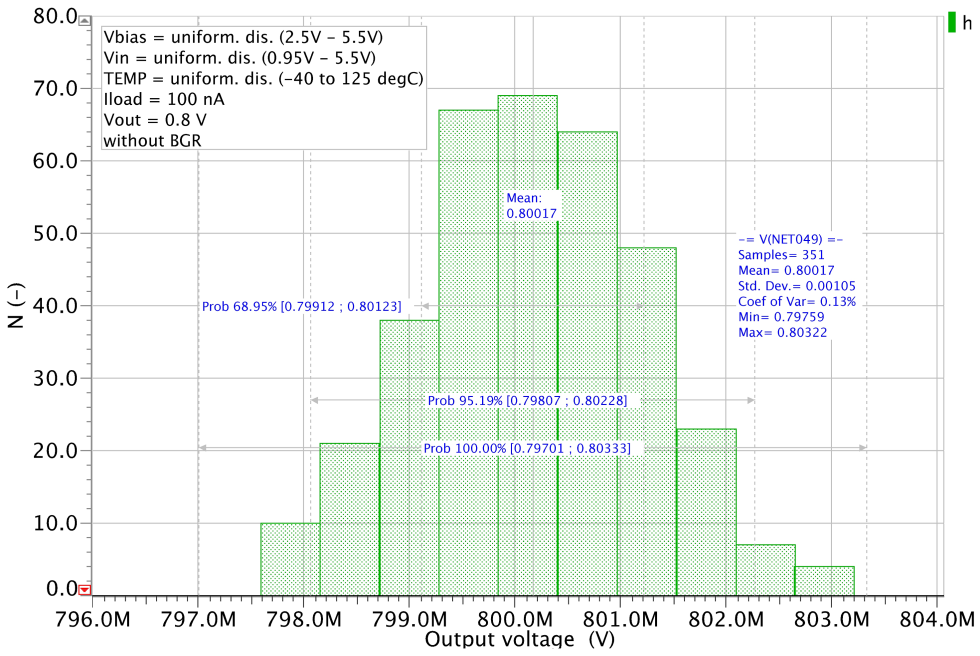


Figure 5.1: MC simulation of the output voltage, $V_{out} = 0.8$ V, without BGR

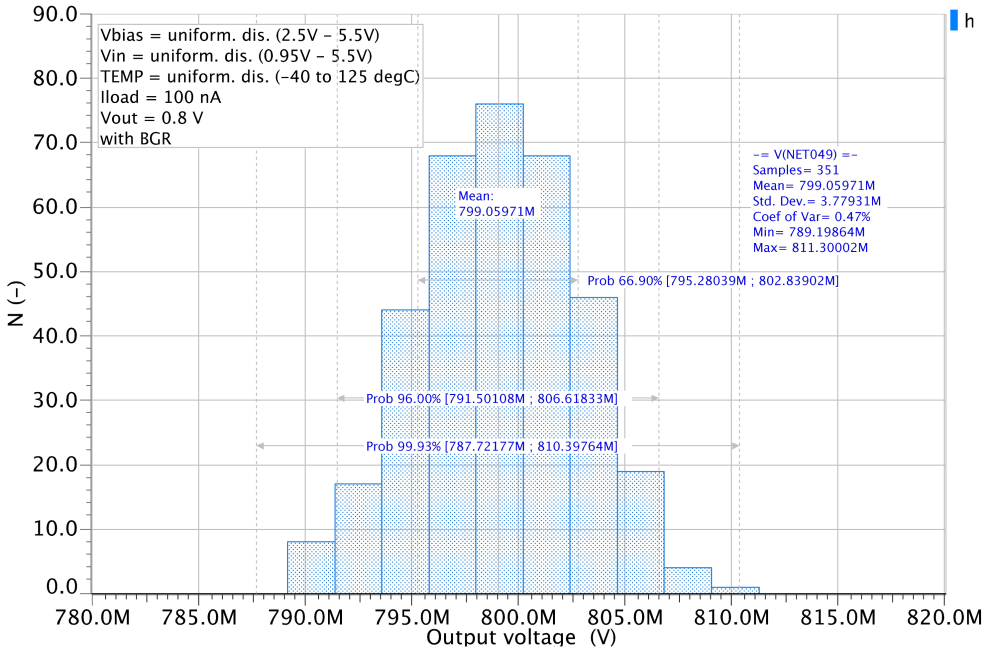


Figure 5.2: MC simulation of the output voltage, $V_{out} = 0.8$ V, with BGR

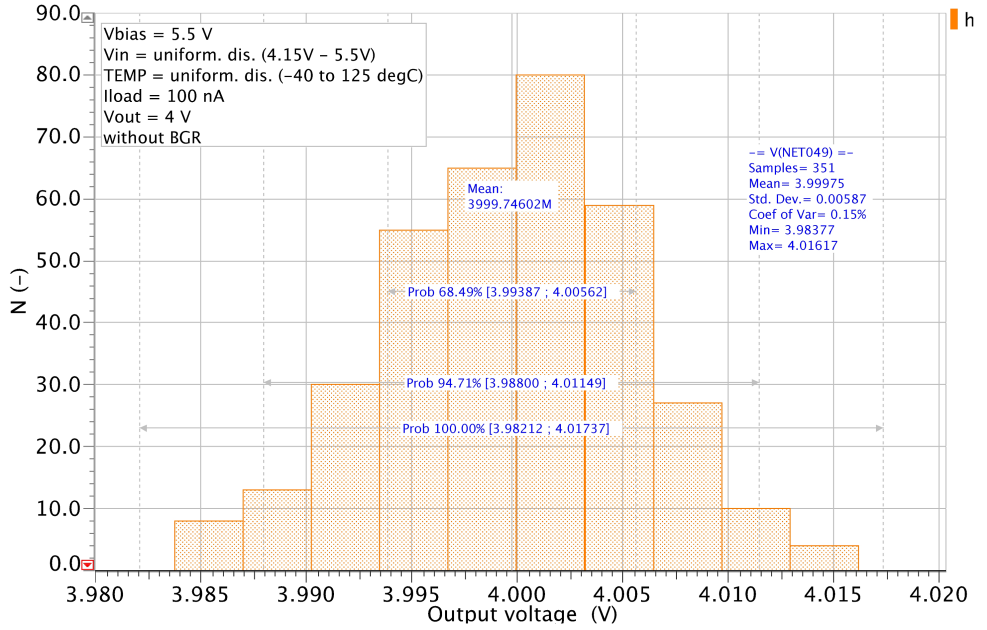


Figure 5.3: MC simulation of the output voltage, $V_{out} = 4$ V, without BGR

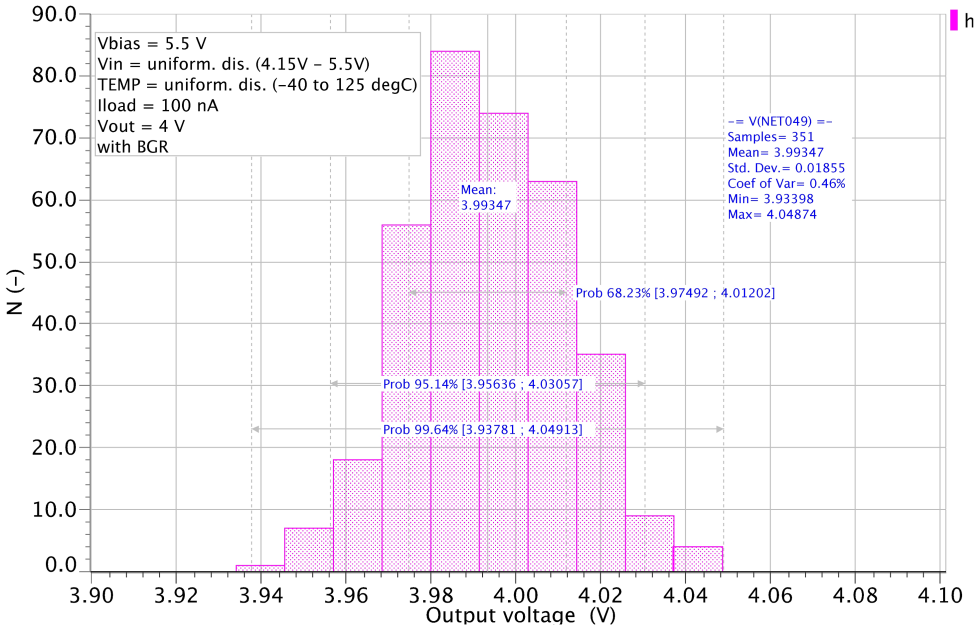


Figure 5.4: MC simulation of the output voltage, $V_{out} = 4$ V, with BGR

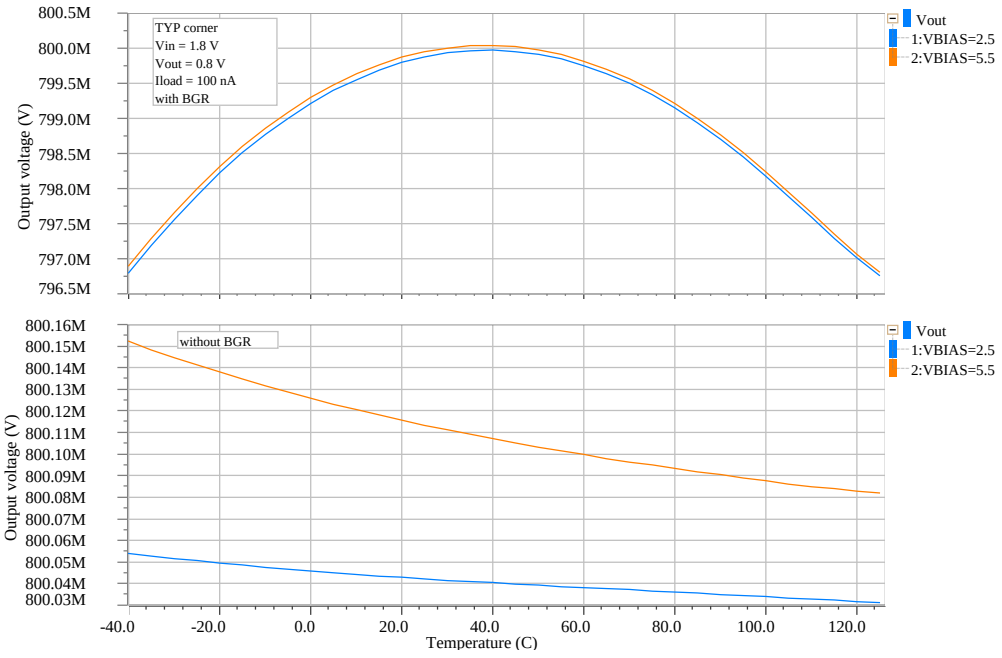


Figure 5.5: Temperature dependence of the output voltage

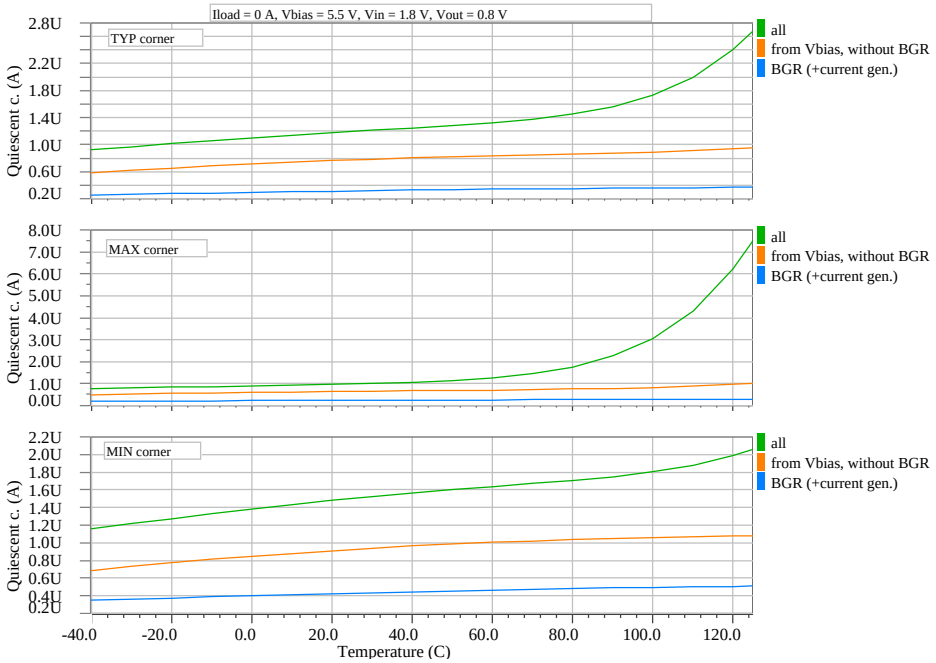


Figure 5.6: Temperature dependence of the quiescent current at no load

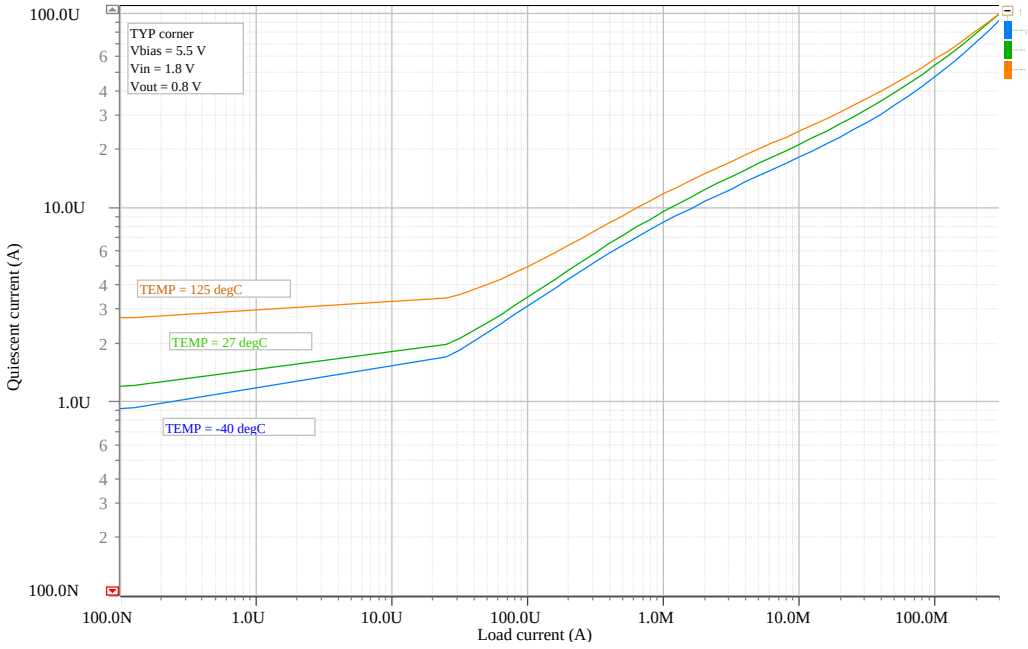


Figure 5.7: Dependence of the quiescent current on the load current

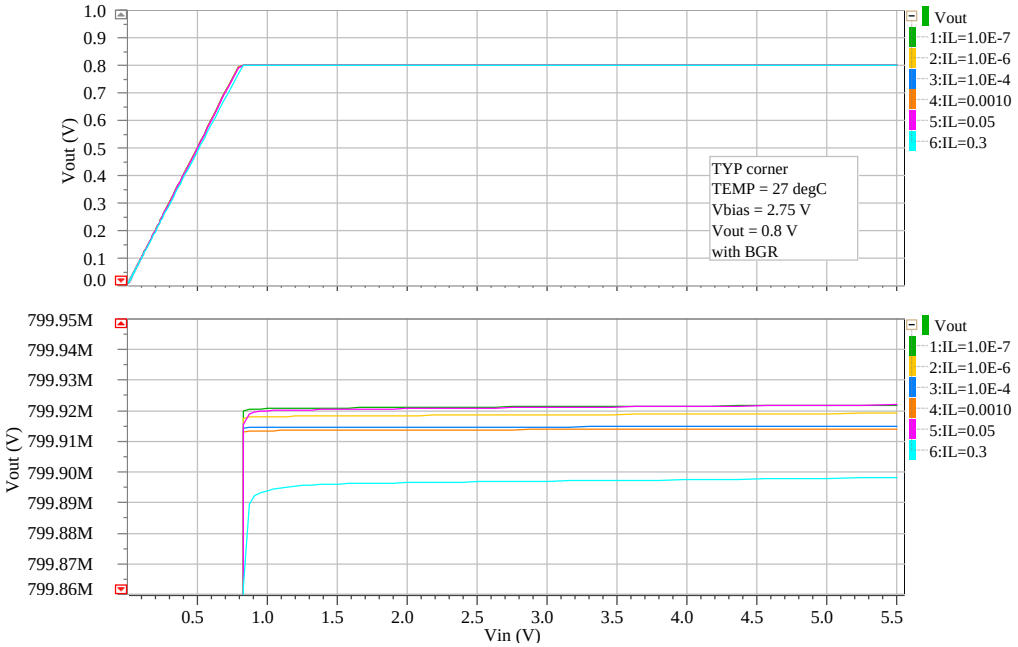


Figure 5.8: Line regulation - vs V_{in}

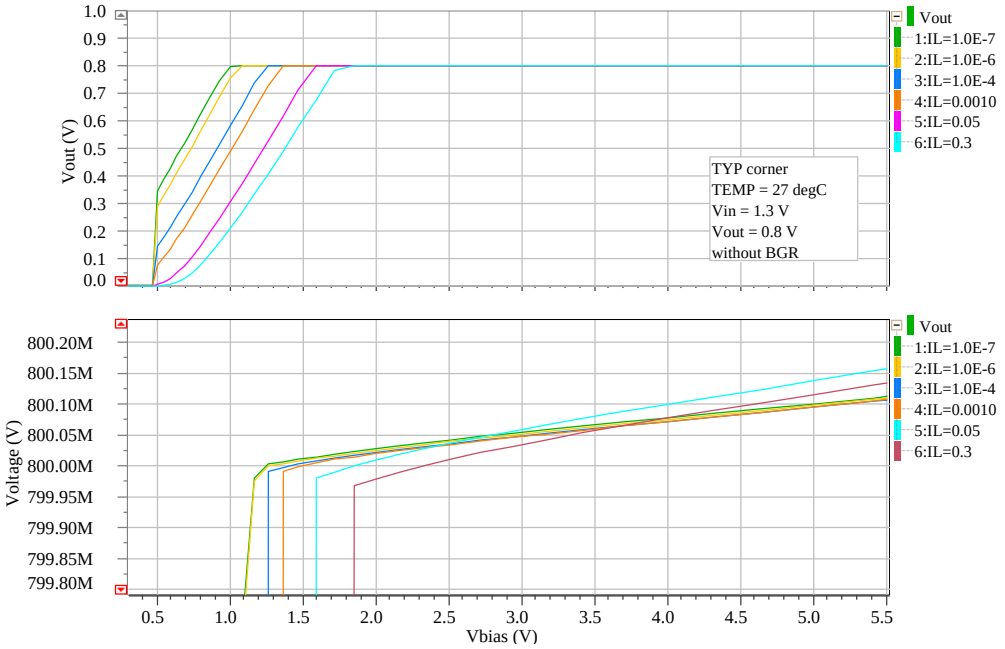


Figure 5.9: Line regulation - vs V_{bias}

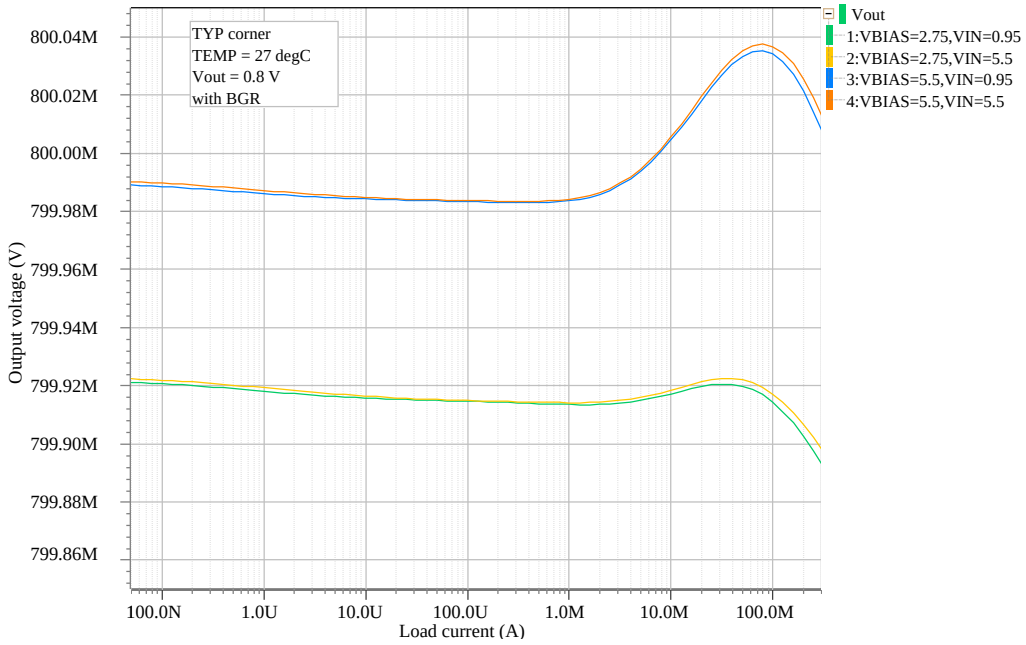


Figure 5.10: Load regulation, $V_{out} = 0.8$ V

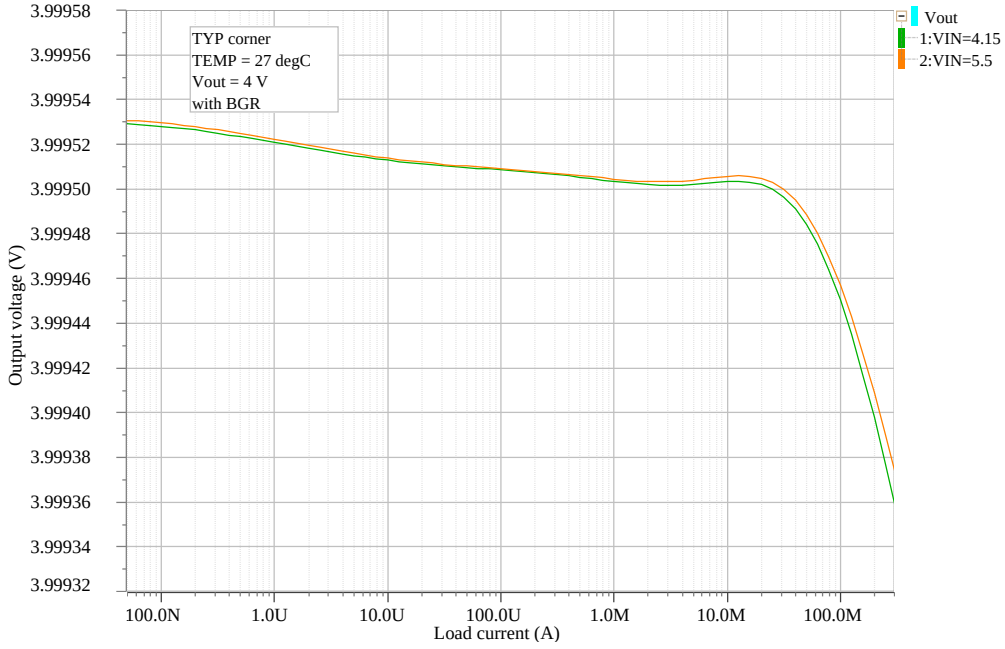


Figure 5.11: Load regulation, $V_{out} = 4$ V

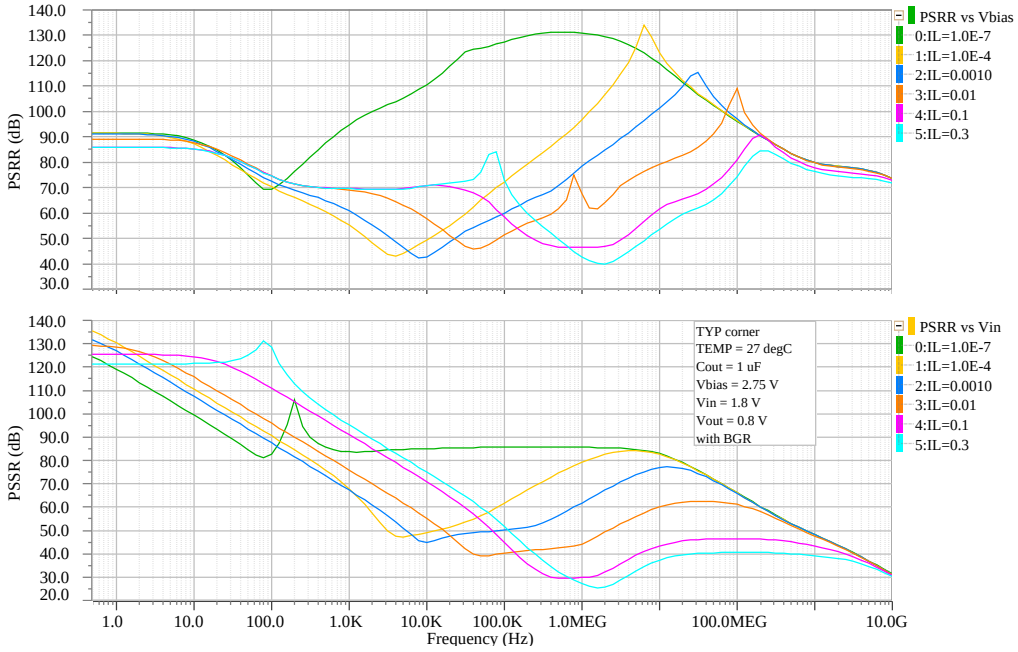


Figure 5.12: PSRR, $V_{out} = 0.8$ V

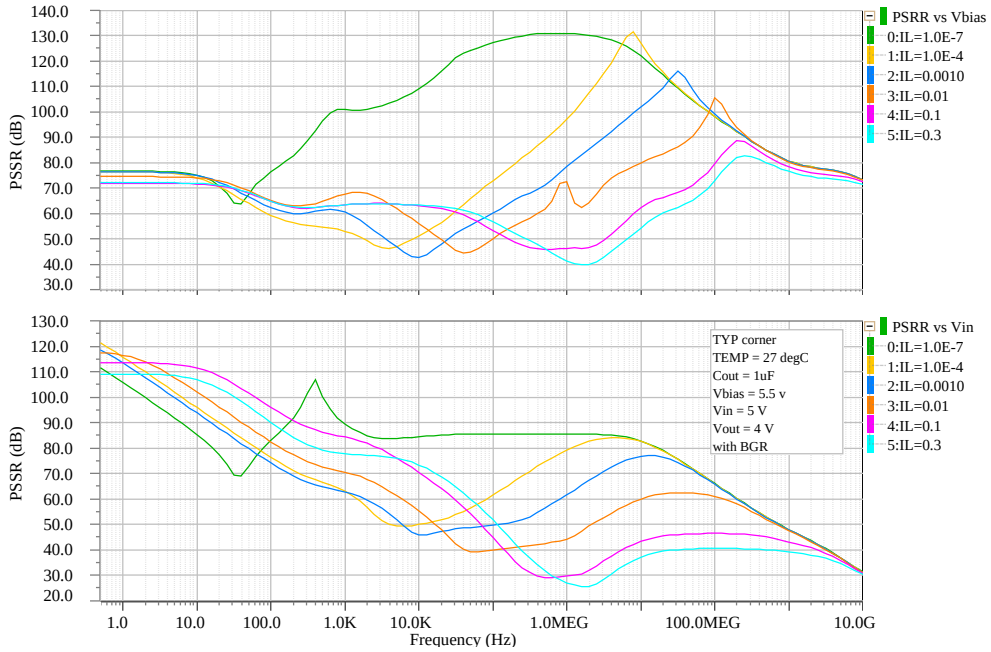


Figure 5.13: PSRR, $V_{out} = 4$ V

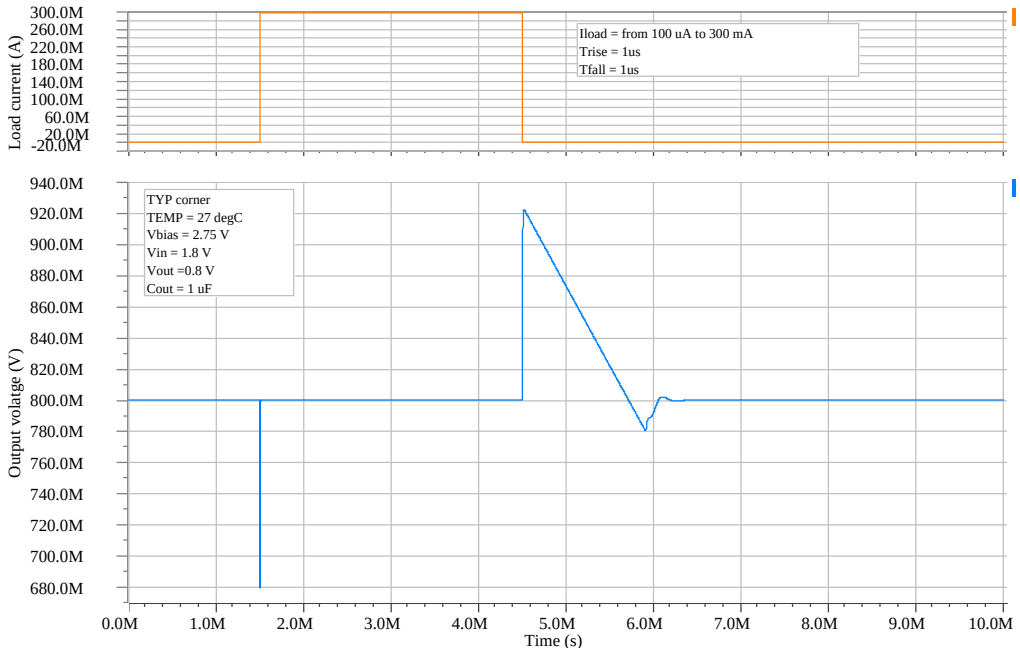


Figure 5.14: Transient load regulation, $V_{out} = 0.8$ V

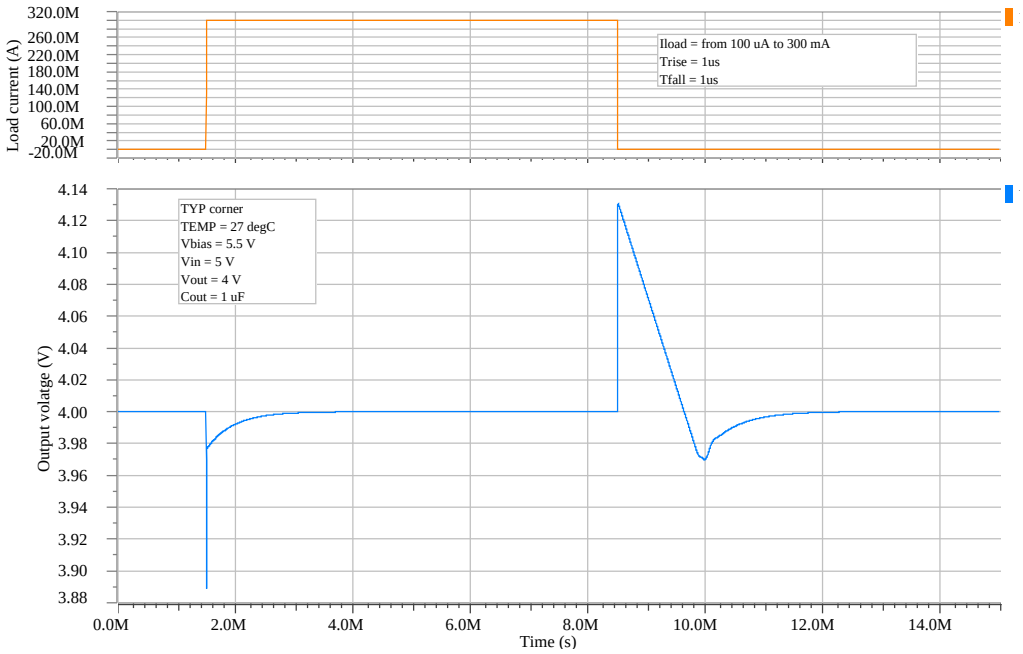


Figure 5.15: Transient load regulation, $V_{out} = 4$ V

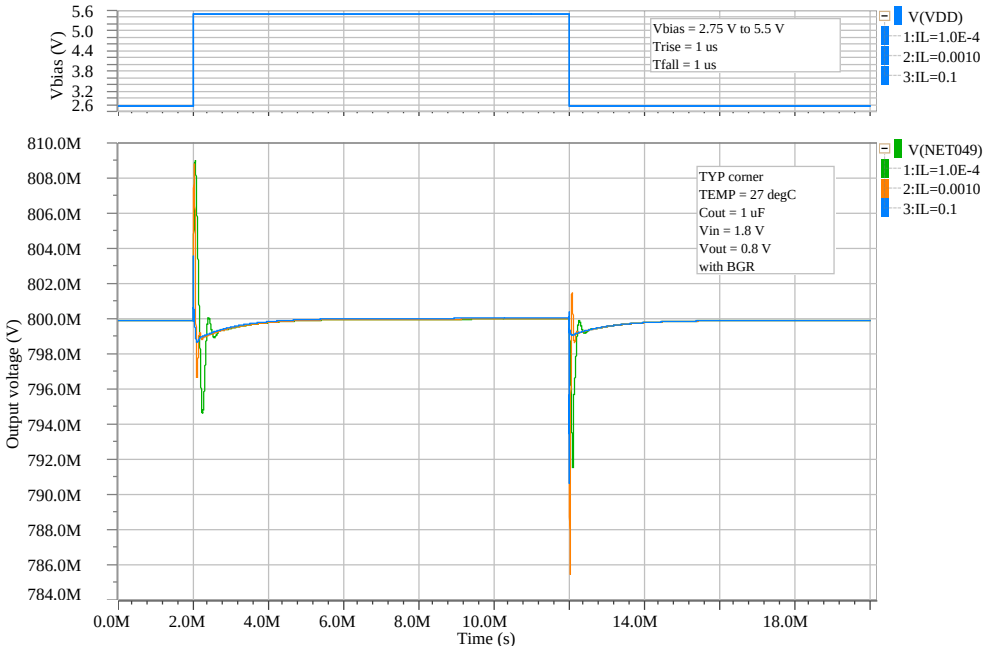


Figure 5.16: Transient line regulation - vs V_{bias}

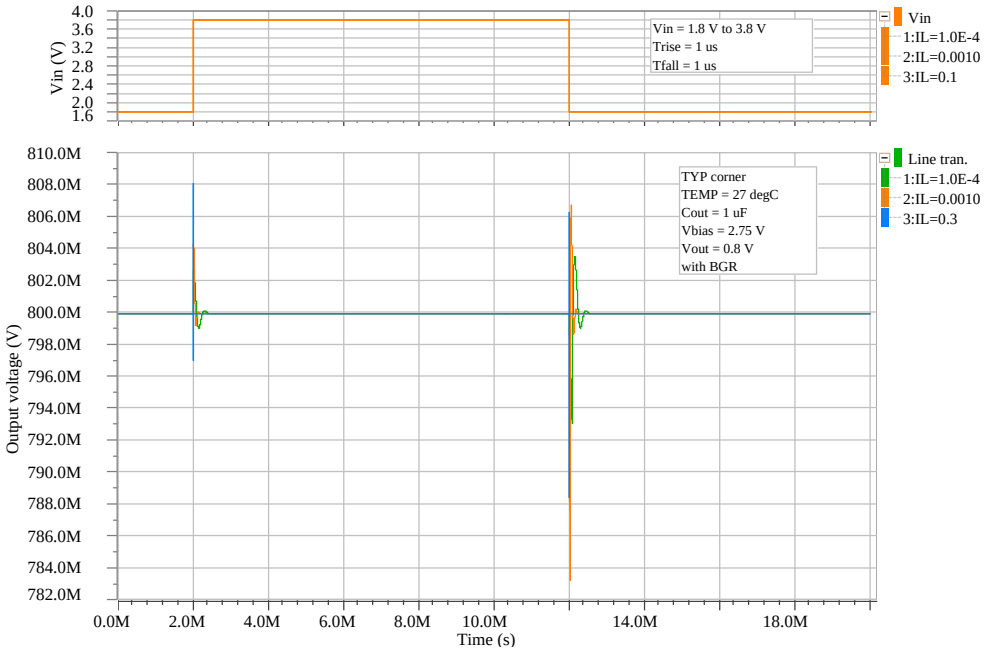


Figure 5.17: Transient line regulation - vs V_{in}

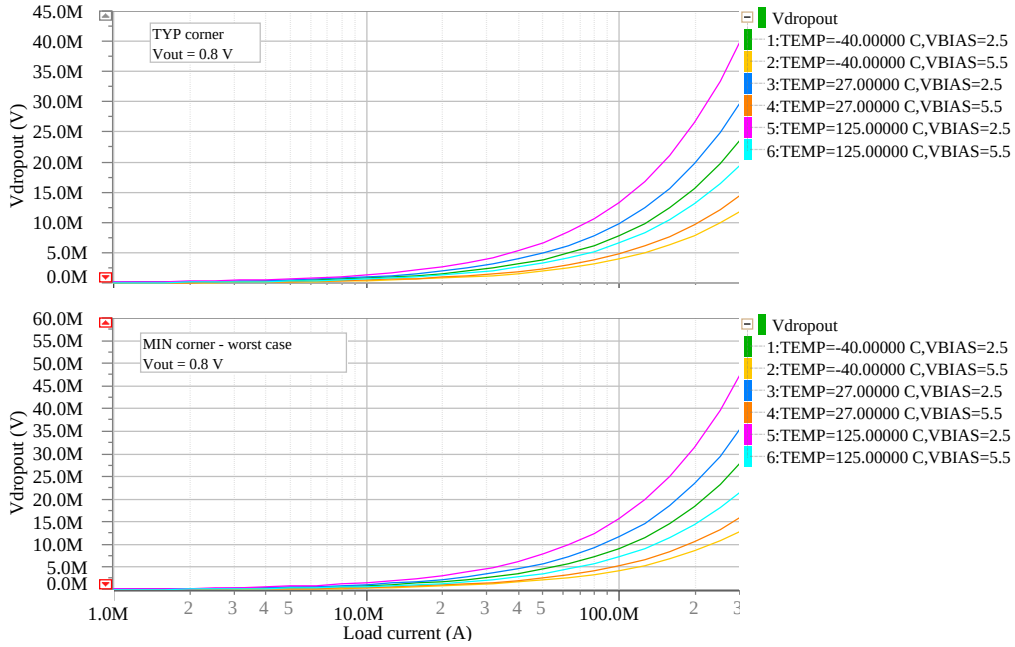


Figure 5.18: Drop-out voltage, $V_{out} = 0.8$ V

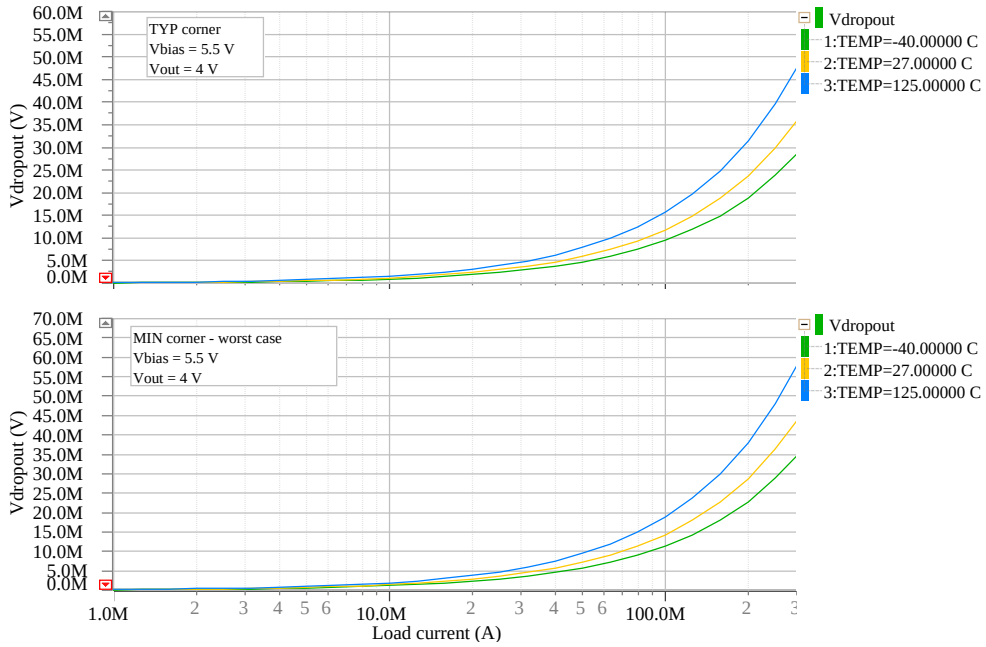


Figure 5.19: Drop-out voltage, $V_{out} = 4$ V

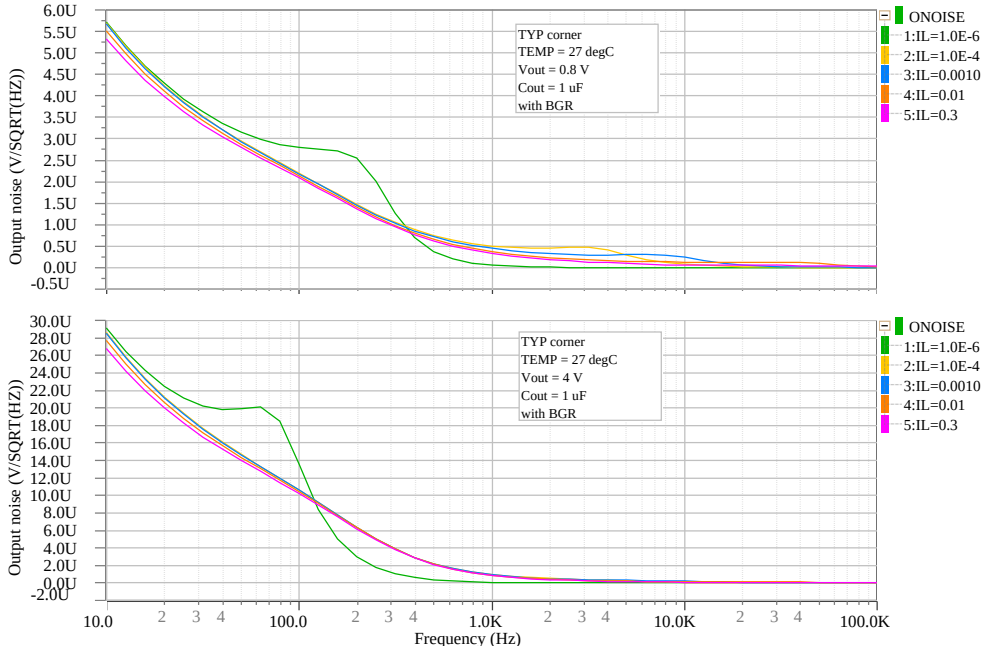


Figure 5.20: Output noise spectral density

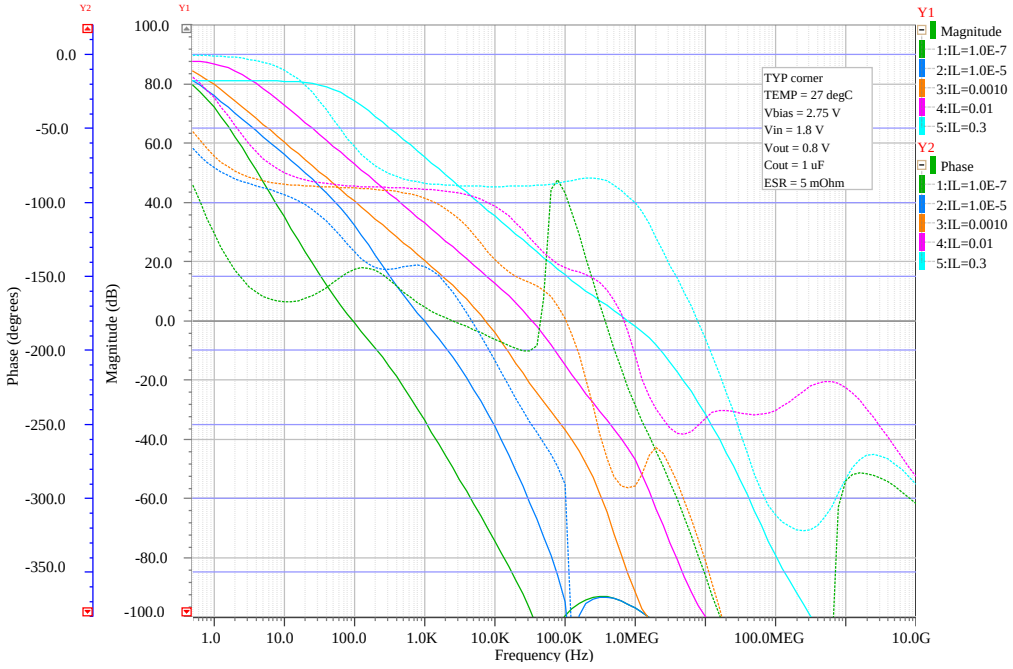


Figure 5.21: Frequency characteristics, $V_{out} = 0.8$ V

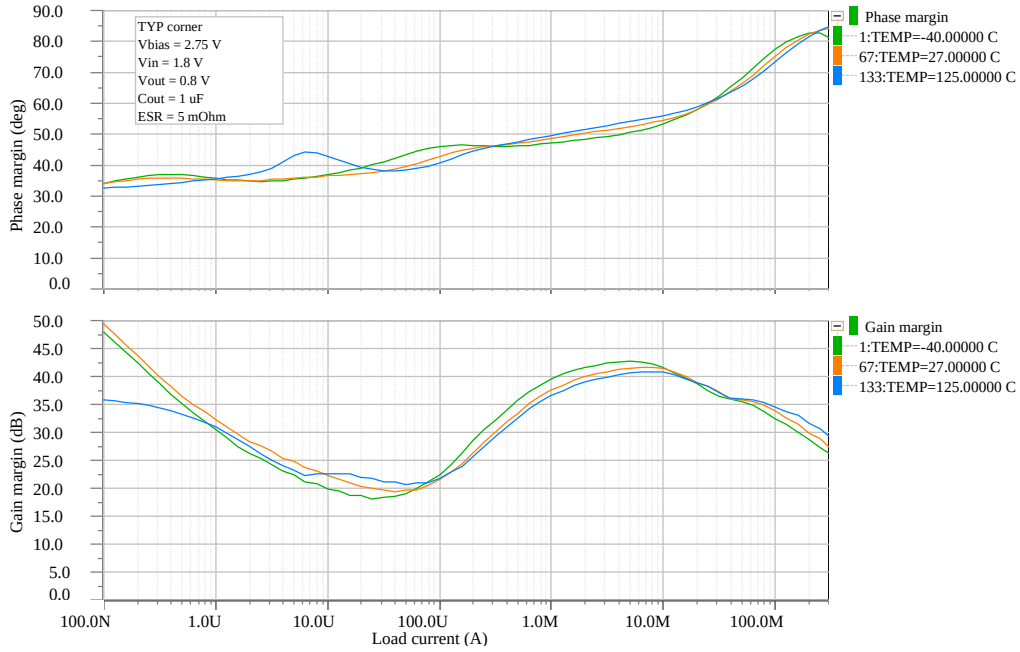


Figure 5.22: Phase and gain margin vs load current, $V_{out} = 0.8$ V

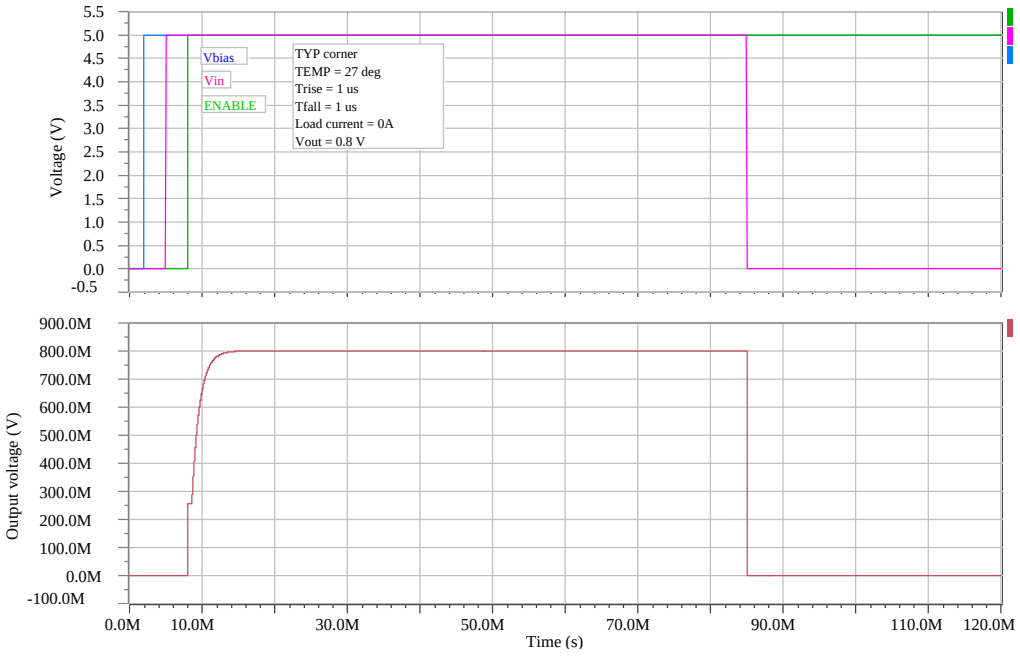


Figure 5.23: Start-up sequence, option A

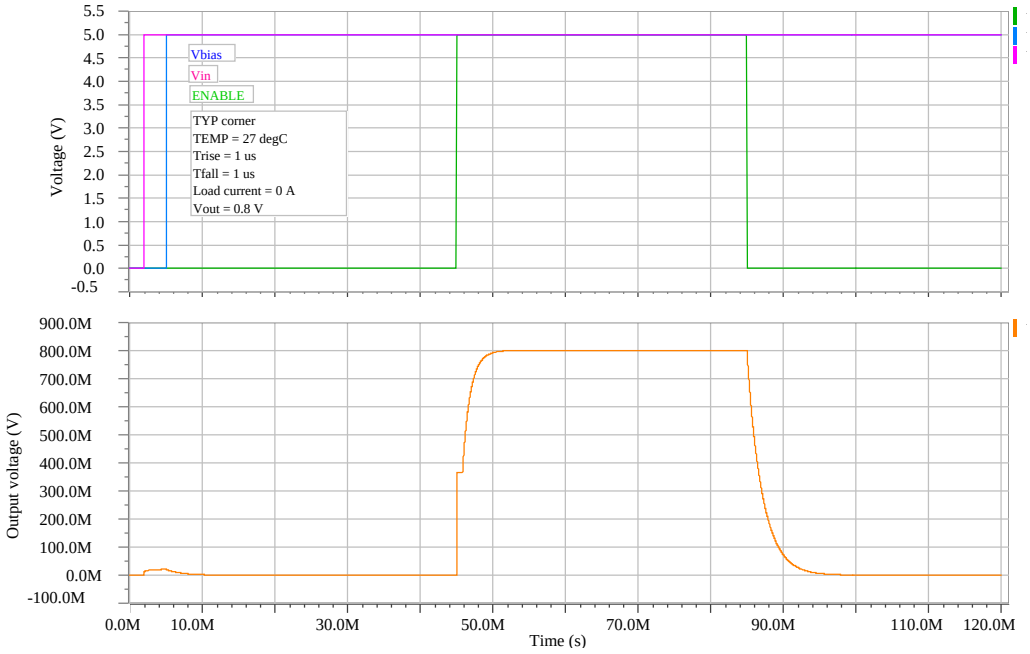


Figure 5.24: Start-up sequence, option B

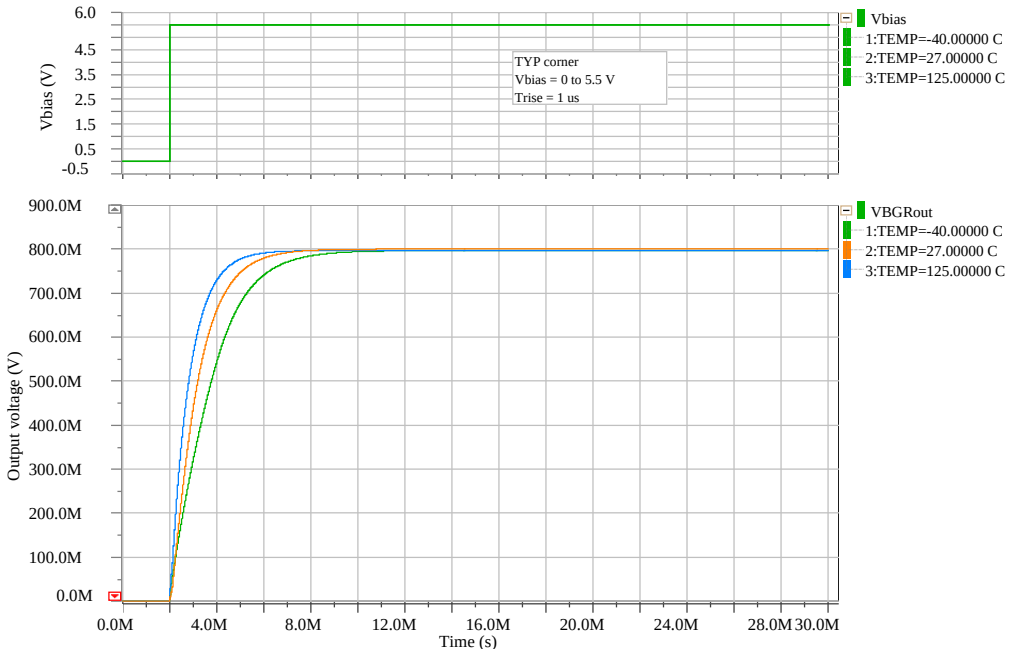


Figure 5.25: Start-up sequence - reference voltage - BGR start-up reference

Chapter 6

Conclusion

In this thesis, design of an ultra-low quiescent current LDO regulator in BCD8 technology of STMicroelectronics company has been performed. Design of a complete chip has been described. Besides the LDO core, it includes the design of a bang-gap voltage reference with a 4-bit trimming circuit, thermal protection circuit, current limiting circuit, enable control and a reference current generator as well.

The main emphasis has been put on the ultra-low quiescent current and efficiency requirements. All the specification requirements have been fulfilled, some of them have been even improved. The quiescent current of the complete chip under typical conditions and no load is only $I_q = 1.2$ μA and the quiescent current without the ban-gap reference is only $I_q = 0.9$ μA. The chip is ideal for battery-powered applications, its efficiency can reach unusually high values. This is achieved, besides the quiescent current, by the drop-out voltage which is about $V_{drop} = 35$ mV (specification was 150 mV) at the maximal load current $I_{load_{max}} = 300$ mA, which puts this LDO to an ultra-low drop-out category as well. The minimal output voltage is $V_{out_{min}} = 0.8$ V and thanks to the used NMOS power transistor, the input voltage can be as low as $V_{in_{min}} = 0.8$ V plus the voltage drop. The mentioned ultra low drop-out voltage can be reached even for the minimal output voltage. Power supply rejection ratio is about $PSRR = 80$ dB versus the input voltage V_{in} and about $PSRR = 70$ dB versus the biasing voltage V_{bias} at 100 Hz. This value is unusually high for an ultra-low quiescent category. Other parameters are on a very good level as well. Improved transient response thanks to implemented dynamic biasing technique is worth mentioning. Precision of the LDO is about $\pm 0.5\%$ across all conditions without considering the band-gap reference and $\pm 2\%$ with connected band-gap reference without using trimming circuit.

The offer of LDOs with a quiescent current on the level of the designed LDO is very restricted in the marker. Comparison of available LDOs produced by competitors of STMicroelectronics have been done. Products of companies such as Texas Instruments, Analog Devices, Linear Technology, OnSemiconductor and others have been compared. Products with a quiescent currents less than 5 μA have been found only by companies Texas Instruments and OnSemiconductor. The comparison of the designed LDO with ultra-low quiescent current LDOs of these two companies is shown in table 6.1. It has to be said that comparing of product its very difficult for maybe not completely the same measurement conditions for each product. Therefore, the results shown in the table are rather suitable for a first-look comparison. For more details it is necessary to look up the information in datasheets.

Values of parameters which make significant differences in an ultra-low quiescent category are shown. It can be said that the parameters of the designed LDO are quite unique, especially

Table 6.1: Comparison of ultra-low quiescent current LDOs

TYPE	this LDO	TPS7B82Q1	TLV704	TPS783
MANUFACTURER	STM	TI [17]	TI [18]	TI [19]
I_q (μA) $I_{load} = 0$	1.2	1.9	3.2	0.5
$V_{in_{min}}$ (V)	0.8 + drop	3 + drop	2.5	2.2
$V_{out_{min}}$ (V)	0.8	3	1.2	1.8
$I_{load_{max}}$ (mA)	300	300	150	150
V_{drop} (mV) @ $I_{load_{max}}$	30	600	1600	130
$PSRR$ (dB) @ 100 Hz, around $I_{load} = 10$ mA	80 (vs V_{in}) 70 (vs V_{bias})	60	50	20
V_{noise} (μV) @ 10 Hz - 100 kHz, $V_{out_{min}}$	51	N/A	N/A	86

TYPE	NCP718	NCP583	NCP170	NCP4624
MANUFACTURER	OnSemi [20]	OnSemi [21]	OnSemi [22]	OnSemi [23]
I_q (μA) @ $I_{load} = 0$	4	1	0.5	2
$V_{in_{min}}$ (V)	2.5	1.7	2.2	2.5
$V_{out_{min}}$ (V)	1.2	1.5	1.2	1.2
$I_{load_{max}}$ (mA)	300	150	150	150
V_{drop},(mV) @ $I_{load_{max}}$	480	600	350	1680
$PSRR$ (dB) @ 100 Hz, around $I_{load} = 10$ mA	70	50	65	40
V_{noise} (μV) @ 10 Hz - 100 kHz, $V_{out_{min}}$	36	N/A	86[(1)]	105

values are for $T \approx 27°C$
[(1)] @ 100 Hz - 1 MHz, $I_{load} = 1$ mA

the combination of ultra-low quiescent current, ultra-low drop-out voltage and high PSRR comparing to others. The ability of operation with very low minimal input voltage and still achieving the mentioned drop-out voltage is not usual either because the vast majority of available LDOs uses PMOS power transistors which do not allow that and in the compared ultra-low quiescent current category no such a variant with very low input voltage has been found.

Bibliography

[1] Lundberg, K. H. *Noise Sources in Bulk CMOS.* Massachusetts Institute of Technology, 2002.

[2] Hung, K. K.; Ko, P. K.; Cheng, Y. C. *A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors.* IEEE Transaction on electron devices, Vol. 37 No. 3, 1999.

[3] Silvaco. *Noise Modeling in MOSFET and Bipolar Devices [online].* [cit. 2018-01-01]. Available from: `https://www.silvaco.com/content/kbase/noise_modeling.pdf`

[4] Bruschi, P. *Notes on Mixed Signal Design [online].* University of Pisa, [cit. 2018-01-01]. Available from: `http://docenti.ing.unipi.it/~a008309/mat_stud/MIXED/2014/Dispense/Chap_3/App_3_1_device_models.pdf`

[5] Gray, P. R.; Hurst, P. K.; Lewis, S. H.; et al. *Analysis and design of analog integrated circuits, fourth edition.* New York, USA: John Wiley & Sons, Inc.

[6] Witoz, E. A. *Weak inversion in analog and digital circuits.* Centre Suisse d'Electronique et de Microtechnique SA, 2003.

[7] Sedra, A. S.; Smith, K. C. *Microelectronic circuits, fourth edition.* New York, USA: Oxford University Press, Inc.

[8] Zeghbroeck, B. V. *Principles of semiconductor devices [online].* University of Colorado, 2011. Available from: `https://ecee.colorado.edu/~bart/book/book/chapter7/ch7_4.htm`

[9] Liu, W.; Jin, X.; Xi, X.; et al. *BSIM3v3.3 MOSFET Model Users' Manual.* University of California, Berkeley, 2005.

[10] Wolpert, D.; Ampadu, P. *Managing Temperature Effects in Nanoscale Adaptive Systems.* New York, USA: Springer.

[11] Zabeli, M.; Caka, N.; Limani, M.; et al. *The impact of MOSFET's physical parameters on its threshold voltage [online].* University of Prishtina, 2007. Available from: `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.636.9728&rep=rep1&type=pdf`

[12] Maxim Integrated. *Applicaation note 4003 Series or Shunt Voltage Reference? [online].* [cit. 2018-01-01]. Available from: https://www.maximintegrated.com/en/app-notes/index.mvp/id/4003

[13] Brokaw, P. A. *How to make a bad gap voltage reference in one easy lesson.* Integrated Device Technology, 2011.

[14] Becvar, D.; Stehlik, J. *Navrh analogovych integrovanych obvodu.* Brno University of Technology, 2006.

[15] Petenyi, S. *Voltage regulator with dropout detector and bias current limiter and associated methods.* US Patent US20170102724A1, 2017.

[16] Petenyi, S. *Voltage regulator circuits, systems and methods for having improved supply to voltage rejection (svr).* US Patent US2017019485511, 2017.

[17] Texas Instruments. *TP37B82-Q1 [online].* 2017. Available from: http://www.ti.com/lit/ds/symlink/tps7b82-q1.pdf

[18] Texas Instruments. *TLV704 [online].* 2015. Available from: http://www.ti.com/lit/ds/symlink/tlv704.pdf

[19] Texas Instruments. *TPS783 [online].* 2014. Available from: http://www.ti.com/lit/ds/symlink/tps783.pdf

[20] OnSemiconductor. *NCP718 [online].* 2017. Available from: http://www.onsemi.com/pub/Collateral/NCP718-D.PDF

[21] OnSemiconductor. *NCP583 [online].* 2016. Available from: http://www.onsemi.com/pub/Collateral/NCP583-D.PDF

[22] OnSemiconductor. *NCP170 [online].* 2017. Available from: http://www.onsemi.com/pub/Collateral/NCP170-D.PDF

[23] OnSemiconductor. *NCP4624 [online].* 2017. Available from: http://www.onsemi.com/pub/Collateral/NCP4624-D.PDF