2009

A Low-Power Low-Voltage Bandgap Reference in CMOS

Na Sun

A Low-Power Low-Voltage Bandgap Reference in CMOS

(Spine title: A Low-Power Low-Voltage Bandgap Reference in CMOS)

(Thesis format: Monograph)

by

Na Sun

Graduate Program
in
Engineering Science
Electrical and Computer Engineering

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Engineering Science

School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

A Low-Power Low-Voltage Bandgap Reference in CMOS

(Spine title: A Low-Power Low-Voltage Bandgap Reference in CMOS)

(Thesis format: Monograph)

by

Na Sun

Graduate Program
in
Engineering Science
Electrical and Computer Engineering

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Engineering Science

School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

A Low-Power Low-Voltage Bandgap Reference in CMOS

(Spine title: A Low-Power Low-Voltage Bandgap Reference in CMOS)

(Thesis format: Monograph)

by

Na Sun

Graduate Program
in
Engineering Science
Electrical and Computer Engineering

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Engineering Science

School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

Bandgap reference plays a substantial role in integrated circuit. Traditionally, it provides a constant reference voltage of $1.205V$ for other blocks in the circuit while itself is independent of temperature and power supply. However, the development of CMOS technology has brought us into a new era of high integration and ultra-low power consumption. As the gate length scales down, it is crucial to build circuits that are able to work under a very low voltage power supply, for instance, lower than the bandgap voltage of $1.205V$. Building bandgap circuits to generate the conventional bandgap voltage under a low voltage power supply such as $1.2V$ or $1V$ is no longer practical nor useful. Thus, bandgap references working under low-voltage and consuming low-power is becoming the trend of research and development nowadays.

In this thesis work, the potential structure of a low-voltage low-power bandgap reference is proposed, which is based on extracting a current that is a fraction of the traditional bandgap voltage. All the necessary blocks are designed to achieve the high accuracy bandgap reference, including bandgap core circuit, op-amp, start-up circuit and output stage. As a result, the designed bandgap reference is able to work under $1.2V$ power supply and provides an output reference voltage of $584.7mV$. It has a variation of only $244.38\mu V$ for the temperature range of $0℃ \sim 125℃$ and has a variation of only $1.1mV$ for a power supply range of $1.08V \sim 1.32V$. The layout design for the bandgap reference structure is also done carefully at the late stage, with an area of $100\mu m \times 85\mu m$.

Keywords: Bandgap Reference, Low-Voltage, Low-Power, Op-Amp, Start-Up, Output Stage, Layout.

Abstract

Bandgap reference plays a substantial role in integrated circuit. Traditionally, it provides a constant reference voltage of $1.205V$ for other blocks in the circuit while itself is independent of temperature and power supply. However, the development of CMOS technology has brought us into a new era of high integration and ultra-low power consumption. As the gate length scales down, it is crucial to build circuits that are able to work under a very low voltage power supply, for instance, lower than the bandgap voltage of $1.205V$. Building bandgap circuits to generate the conventional bandgap voltage under a low voltage power supply such as $1.2V$ or $1V$ is no longer practical nor useful. Thus, bandgap references working under low-voltage and consuming low-power is becoming the trend of research and development nowadays.

In this thesis work, the potential structure of a low-voltage low-power bandgap reference is proposed, which is based on extracting a current that is a fraction of the traditional bandgap voltage. All the necessary blocks are designed to achieve the high accuracy bandgap reference, including bandgap core circuit, op-amp, start-up circuit and output stage. As a result, the designed bandgap reference is able to work under $1.2V$ power supply and provides an output reference voltage of $584.7mV$. It has a variation of only $244.38\mu V$ for the temperature range of $0℃ \sim 125℃$ and has a variation of only $1.1mV$ for a power supply range of $1.08V \sim 1.32V$. The layout design for the bandgap reference structure is also done carefully at the late stage, with an area of $100\mu m \times 85\mu m$.

Keywords: Bandgap Reference, Low-Voltage, Low-Power, Op-Amp, Start-Up, Output Stage, Layout.

Acknowledgement

First I would like to thank The University of Western Ontario for granting me the opportunity to study and finish this work in such a nice environment.

I am extremely grateful for the help and support from my supervisor, Dr. Robert Sobot during the past two years, both for this thesis research work guidance and in every aspect of my studies at The University of Western Ontario.

I also would like to express my thanks to my present and former colleagues in our research group for their care and help. In particular, I would like to thank Sneha for proofreading this work.

My special appreciation goes to Tiam, for all his concerns for me and for all the memorable supports he gave me during this thesis work and the school days.

Last but not the least, I would like to express my deepest gratitude to my family– my parents for their unconditional love, constant encouragement and support for me to study abroad, my uncle and my aunt for their inspiration and insightful suggestions during my studies, and my other family members for their optimistic outlook.

Table of Contents

Table of Contents

Table of Contents

Table of Contents

List of Tables

List of Figures

List of Figures

List of Symbols and Abbreviations

ADC	Analog–to–Digital Converter
DAC	Digital–to–Analog Converter
PLL	Phase–locked–Loop
BJT	Bipolar Junction Transistor
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
DTMOST	Dynamic-Threshold MOS Transistor
BiCMOS	Bipolar Complementary Metal-Oxide-Semiconductor
CMOS	Complementary Metal-Oxide-Semiconductor
DC	Direct Current
IC	Integrated Circuit
TC	Temperature Coefficient
PTAT	Proportional To Absolute Temperature
OP-AMP	Operational Amplifier
CMRR	Common-Mode Rejection Ratio
VCCS	Voltage-Controlled Current Source
OTA	Operational-Transconductance Amplifier
PMOS	P-Channel MOSFET
NMOS	N-Channel MOSFET
DRC	Design Rule Checking
LVS	Layout Versus Schematic
ICMR	Input Common Mode range
CMRR	Common Mode Rejection Ratio
PSRR	Power Supply Rejection Ratio
SR	Slew Rate
MIM	Metal Insulator Metal

$V_{in,CM}$	Common Mode Input
V_{Dsat}	Saturation Voltage
v_{out}	AC Output Voltage
v_{in}	AC Input Voltage
v_{in}	Input Difference in Differential Input Stage
ω	Frequency in Radians
C_L	Load Capacitance
C_C	Compensation Capacitance
C_A	Capacitance at Point A
p	Pole in Frequency Response
z	Zero in Frequency Response
C_{load}	Load Capacitor
$G(A)$	Variation Gradient of A
V_{OS}	Offset Voltage
A_{CM}	Common Mode Gain
V_{dd-pp}	Power Supply Peak-to-Peak Voltage
V_{out-pp}	Output Peak-to-Peak Voltage
A_V	Gain
g_m	Transconductance
V_{GS}	Gate to Source Voltage
V_{DS}	Drain to Source Voltage
I_D	Drain Current
V_{DD}	Positive Power Supply Rail
V_{SS}	Negative Power Supply Rail
I_R	Current Through R
R	Resistor/Resistance
I_{REF}	Reference Current
V_{REF}	Reference Voltage
V_{BG}	Bandgap Voltage
V_{BE}	Base Emitter Voltage
V_T	Thermal Voltage
q	Electron Charge
I_C	Collector Current
I_S	Saturation Current

T	Absolute Temperature
μ	Mobility of Minority Carriers
n_i	Intrinsic Minority Carrier Concentration
E_g	Bandgap Energy
ΔV_{BE}	Base Emitter Voltage Difference
D	Current Density
k	Boltzmann Constant
W	Channel Width of Transistor
L	Channel Length of Transistor
λ	Channel Length Modulation Parameter
V_{th}	Threshold Voltage
V_{thn}	Threshold Voltage for NMOS
V_{thp}	Threshold Voltage for PMOS
(W/L)	Width to Length Ratio
m	Number of Elements in Parallel
V_{out}	Output Voltage
A_{diff}	Gain of Differential Stage
r_o	Output Impedance
A_{vcs}	Gain of Common Source Stage
g_d	Channel Conductance of Transistor
$A_{Vpush-pull}$	Gain of Push-Pull Stage
$+SR$	Slew Rate at Rising Edge
$-SR$	Slew Rate at Falling Edge
S_v	Thermal Noise
f	Frequency
V_{REF-pp}	Peak-to-Peak Reference Voltage

Chapter 1

Introduction

In integrated circuits nowadays, the bandgap voltage reference circuit, which has a very low dependence upon temperature and power supply, is used to provide a voltage reference for other blocks in the circuits. The requirement for a stable reference voltage is one of the most important issues in micro-electronic design as it is always a necessary part in many analog and mixed-signal integrated circuits such as ADC, DAC, PLL and memories.

As the technology has developed rapidly during the past several decades, the power supply voltage is required to scale down (for example, the power supply needs to be as low as $1.2V$ or even sub-$1V$) due to the shrinking gate lengths and the increased demand for low-power portable equipment in today's electronic applications. Therefore, it is critical to develop low-power consumption bandgap references working under low-voltage supply that are compatible to almost all the technologies at present.

In this chapter, a review of the bandgap reference is given and the research objective and the organization of this thesis are followed.

1.1 Review of The Bandgap Reference Circuits

The first bandgap reference circuit based on Bipolar Junction Transistor (BJT) was introduced in the early 1970s. It was in 1971 that the expensive Zener diode reference device was replaced by a bandgap reference circuit based on BJT in an integrated

circuit. By the concept introduced in [2], the bandgap voltage reference started a new era of development on chip. In 1973, a precision reference voltage source [3] was presented to replace the Zener diode as a discrete component, providing a $10V$ output voltage. A year later, A. Paul Brokaw illustrated a three-terminal bandgap reference with a much lower temperature coefficient, which is $5ppm/°C$ over a much wider temperature range [4]. Later on, the bandgap voltage reference continuously improved according to the temperature stability by innovative circuitry such as curvature compensation, laser trimming [3, 4, 5, 6] and op-amp offset cancellation [7].

In the 1990s, the dynamic-threshold MOS transistors (DTMOST) were implemented to replace the normal diodes in technologies. These DTMOSTs have an interconnected gate and backgate [8, 9] and the lateral bipolar structure [10, 11]. As a result of this architecture, the power supply for bandgap voltage reference circuit is reduced to $1.8V$ or lower and consumes only several micro-watts [9].

With the development of technology, the lower-voltage power supply bandgap circuits, which can even work under supply voltage of lower than $1.205V$ were required for application purposes. In [12], a $1V$ output was produced by a bandgap reference circuit working under $1.2V$, and the power-supply voltage was even lower (minimum $1V$) in a bandgap reference, which used standard bipolar process with base-diffused resistors [13]. The authors in [14] illustrated a very high accuracy bandgap reference with a temperature coefficient of $15ppm/°C$, in the absence of a low threshold voltage device, it was able to operate under $0.98V$ power supply. Since then, the bandgap voltage reference started another page of low power low voltage, sub-$1V$ bandgap references were developed in different technologies and with various compensation and trimming techniques [15, 16, 17, 18].

In the near past, the authors in [19] used the same technology as this thesis and by the time the thesis project was finished, a $1V$ bandgap reference without an op-amp

was developed by Lee, Edward K.F. [20]. The authors in [21] also demonstrated a novel bandgap reference circuit that is suitable for fabrication in any digital CMOS technology by eliminating resistors or operational amplifiers.

This work is based on the concept of introducing a fraction of original bandgap voltage, for example a temperature independent current, which was introduced by H.Banba et al. [22]. The concept was further applied and improved in the works of P. Malcovati et al. [1, 23]. The authors in [1] discussed a bandgap architecture capable of operating at $1V$ voltage-supply, while using a conventional BiCMOS technology to generate a reference voltage of $0.54V$, and this architecture consumes $92\mu W$ at room temperature. By correcting the curvature error, the voltage variation got $7.5ppm/°C$ in the temperature range from $0°C$ to $80°C$ and $212ppm/V$ of supply voltage dependence. Lower power consumption was achieved in [23] and the performance comparison of low-power bandgap voltage references was given.

1.2 Motivation and Research Objectives

The aim of this thesis is to develop a low-voltage bandgap voltage reference by using CMOS 90nm technology with optimized operational amplifier. The designed bandgap is also aimed for low power consumption, low noise and high temperature independence. The objective of this thesis are:

- To meet the demand of the low-voltage power supply in recent applications, the circuit is aimed to work under $1.2V$. The new technology may provide partial help, however, in order to achieve the goal, the main task should be focused on designing a bandgap circuit architecture that works under low-voltage power supply.

- To fully understand the available technology, especially the lateral BJT models, which will eventually determine the output results, characterization for the BJT transistors is required, including the characterization of their performance according to temperature change.

- To obtain a precise output voltage, the bandgap circuit architecture that contains an op-amp is implemented. Thus, an op-amp that is able to work under low voltage power supply with a moderate gain bandwidth product (as the circuit is mostly needed for DC voltage) is demanded.

- To get the right powering-up behavior in the circuit, a start-up circuit is necessary for breaking the zero operating point of the circuit and bringing it into the designed operating point.

- To achieve a voltage of high precision at the output node, an output stage whose behavior is independent of changes in temperature and power supply change should be developed.

As a result, the design includes an improved start-up block compared to [1], a bandgap core circuit, a high-quality two stage operational amplifier, and an output buffer sub-circuit.

1.3 Organization of The Thesis

In this thesis, a high accuracy bandgap voltage reference circuit is developed based on $90nm$ CMOS technology and its characterization is done by using IC design tools.

In chapter 2, the basic concept of building a bandgap reference is presented and several bandgap references are investigated as examples.

In chapter 3, the proposed bandgap voltage reference is described block by block in details. The schematic of each block is demonstrated and the further optimization is discussed.

In chapter 4, the layout for the proposed bandgap voltage reference is developed. As a result, the final bandgap voltage reference layout appears square and consumes considerably small area.

In chapter 5, a complete characterization is provided for the whole bandgap circuit as well as the composing blocks, and a comparison of the results is given at the end of the chapter.

In chapter 6, the conclusion and potential future work are presented.

Chapter 2

Bandgap Voltage Reference

In the previous chapter, the background of bandgap reference was introduced and the development of bandgap reference circuits was reviewed. The research objective, including the potential circuit blocks was also presented in the last chapter. In this chapter, the necessity of bandgap reference in applications is explained and the basic concept for building the bandgap circuit is discussed, followed with some case studies.

2.1 Introduction to Bandgap Reference

As it was demonstrated in Chapter 1, bandgap reference is a very important part in every single chip. A simple example can be found in Fig. 2.1(a). In this simple one-transistor Common Source amplifier configuration, the gain of the circuit is $A_V = R_s g_m$. To get the largest gain, one approach is to increase g_m and the other one is to increase R_s. The transconductance g_m can not be increased infinitely because the transistor must be guaranteed to stay in saturation region. To increase R_s, the best way is to use an ideal current source that has an infinite impedance.

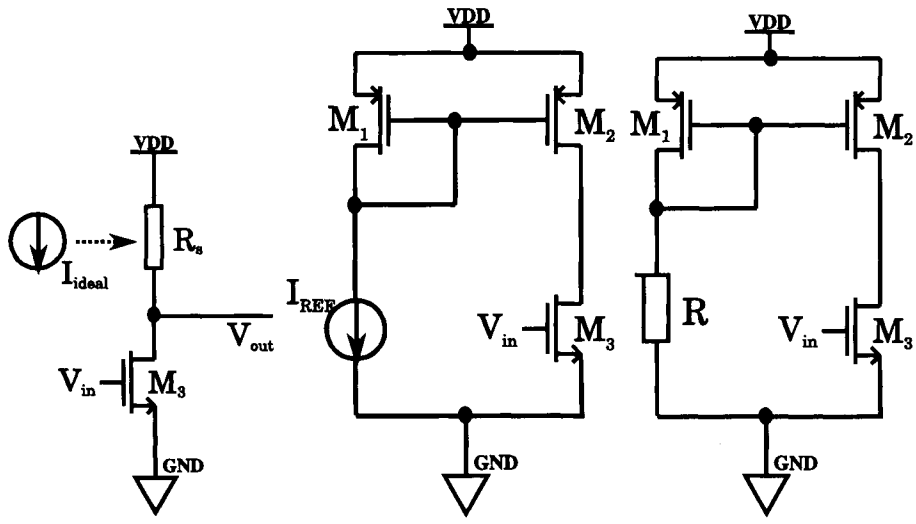
Fig. 2.1(b) gives an example of feeding the load of the amplifier using a current mirror. In this figure $V_{GS1} = V_{GS2}$, therefore, $I_{D1} = I_{D2}$. In other words, if transistor M_1 is identical to M_2, M_2 copies the current from M_1. In Fig. 2.1(b), a current source is used for generating and stablizing the current through M_1, so that both transistors of the current mirror can operate in saturation region. A resistor can be

used to realize the current source, as shown in Fig. 2.1(c). However, the resistor value varies a lot as the temperature changes. In addition, since $I_R = (V_{DD} - V_{GS1})/R$, the current it produces is related to the power supply. Thus, any disturbance of the power supply can influence its current. In the end, the resistor, as most of the electronic elements, is process dependent. All these limitations discussed above generates an unstable current in Fig. 2.1(c). This situation means that there is a chance that the saturation region lock will be lost.

In summary, it is essential to have a very stable current source (I_{REF}), which is independent of temperature, power supply and process, to make the circuit operate properly. This current can be obtained by either developing a current reference or by generating a current fraction from a voltage reference V_{REF}. Thus, voltage and current references are made with little dependence on temperature. What's more, less temperature dependence also reduces process dependence since most process parameters are related to temperature variations [24].

One of the simplest ways is to generate a voltage reference from the power supply by utilizing a voltage divider. However, the output voltage would be a fraction of the power supply voltage, which means it varies according to any variations on the power supply voltage if any disturbance takes place [25]. There are many other more complex approaches to realize current or voltage references that are independent of temperature variation and power supply disturbance, such as making use of a zener diode, applying the difference in the threshold voltage between an enhancement transistor and a depletion transistor, and trying to cancel the temperature dependence of bipolar transistors in a bandgap reference circuit [26]. The most commonly used method nowadays is the third one as the bandgap reference circuit provides high accuracy and is available in most technologies.

The term "bandgap" comes from bandgap energy, which represents for the energy



(a) Common source amplifier

(b) Feeding load with current mirror

(c) Current source realization by resister

Figure 2.1: Circuit diagram showing the demand for bandgap reference

difference between the top of the valence band and the bottom of the conduction band. In other words, an electron requires this amount of energy to jump from one covalent band to another. The bandgap voltage for reference circuits specifically refers to the bandgap voltage of silicon, which can be modeled as a function of temperature:

$$V_{BG} = 1.17 - \frac{4.73 \times 10^{-4} T^2}{T + 635}. \tag{2.1}$$

For the temperature concerned in most applications, this voltage can be further linearized as $1.20585 - 2.745 \times 10^{-4} T$. Obviously, at absolute temperature of $0K$, the bandgap voltage of silicon becomes a constant, $1.20585V$ [27].

2.2 Temperature Independent Reference

There are two kinds of temperature coefficients (TC), positive coefficient and negative coefficient. Summing the opposite TC voltage or current values with a suitable ratio, for example if

$$k_1 \oplus + k_2 \ominus = 0 \tag{2.2}$$

(here $\oplus$ represents a positive temperature coefficient value, and $\ominus$ represents a negative temperature coefficient value), then the final value is zero TC. For example, if there are two currents with the opposite TC direction, picking k_1 and k_2 will make

$$k_1 \frac{\partial I_1}{\partial T} + k_2 \frac{\partial I_2}{\partial T} = 0$$

(Here $\partial I_1/\partial T = \oplus$, and $\partial I_2/\partial T = \ominus$), so that the final value $I_{REF} = k_1 I_1 + k_2 I_2$ is independent of temperature variation.

2.2.1 Negative Temperature Coefficient

The base-emitter voltage (V_{BE}) of a bipolar-transistor has a negative TC. It is known that,

$$I_c = I_s e^{\frac{V_{BE}}{V_T}}$$

where V_T is the thermal voltage, I_s is the saturation current and V_{BE} is the base-emitter voltage of BJT. Voltage $V_T = kT/q$, in which k is the Boltzmann constant ($k = 1.3806503 \times 10^{-23} m^2 kgs^{-2} K^{-1}$), T is the absolute temperature in K (Kelvin), and q is the electron charge ($q = 1.602 \times 10^{-19} C$). Current $I_s \propto \mu k T n_i^2$, in which μ is the mobility of minority carriers and n_i is the intrinsic minority carrier concentration of silicon. The mobility of minority carriers $\mu \propto \mu_0 T^m (m \approx -\frac{3}{2})$,

and $n_i^2 \propto T^3 \exp[-E_g/(kT)]$. Symbol E_g is the bandgap energy of silicon, which approximately equals to $1.12eV$. Thus,

$$I_s = bT^{4+m} \exp[-E_g/(kT)] \tag{2.3}$$

where b is a proportionality factor [24]. Also,

$$V_{BE} = V_T \ln \frac{I_c}{I_s} \tag{2.4}$$

Under the condition that the dependence on temperature of I_s is much more significant than that of I_c, the derivative of V_{BE} with respect to T can be taken by assuming I_c is constant, in other words, computing the temperature coefficient of V_{BE} :

$$\frac{\partial V_{BE}}{\partial T} = \frac{\partial V_T}{\partial T} \ln \frac{I_c}{I_s} - \frac{V_T}{I_s} \frac{\partial I_s}{\partial T} \tag{2.5}$$

With the derivative of I_s with respect to T from (1.1) and the definition of V_{BE} in (1.2), we can get the final derivative of V_{BE} [24]:

$$\begin{aligned} \frac{\partial V_{BE}}{\partial T} &= \frac{\partial V_T}{\partial T} \ln \frac{I_c}{I_s} - (4+m)\frac{V_T}{T} - \frac{E_g}{kT^2} V_T \\ &= \frac{V_{BE} - (4+m)V_T - E_g/q}{T} \end{aligned} \tag{2.6}$$

It can be noticed that the V_{BE} temperature coefficient also depends on its magnitude. Assume $V_{BE} \approx 750mV$, at temperature of $300K$, $\frac{\partial V_{BE}}{\partial T} \approx -1.5mV/K$

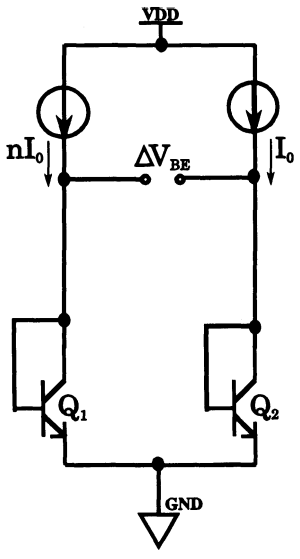


Figure 2.2: PTAT voltage generation circuit diagram

2.2.2 Positive Temperature Coefficient

In bandgap references, the positive temperature coefficient is realized by generating a Proportional to Absolute Temperature (PTAT) voltage. When different current densities are used, the voltage difference between two bipolar transistors' base-emitter voltages can be considered as a PTAT voltage, Fig. 2.2 [24]. Here two identical bipolar transistors are used, and the current through these two transistors has a ratio of n, which means, the ratio of current densities is n. The voltage difference between the two output nodes is ΔV_{BE}:

$$\begin{aligned} \Delta V_{BE} &= V_{BE1} - V_{BE2} \\ &= V_T \ln \frac{nI_0}{I_s} - V_T \ln \frac{I_0}{I_s} \\ &= V_T \ln n \end{aligned} \tag{2.7}$$

The temperature coefficient of ΔV_{BE} can be

$$\frac{\partial \Delta V_{BE}}{\partial T} = (\frac{kT}{q})' \ln n = \frac{k}{q} \ln n \tag{2.8}$$

As k, q and $\ln n$ are all positive, $\frac{\partial \Delta V_{BE}}{\partial T}$ will result in a positive number, and as shown in the equation, it is proportional to absolute temperature.

With positive temperature coefficient and negative temperature coefficient, a voltage value with zero temperature coefficient can be obtained, $\frac{\partial V_{BE}}{\partial T} + m\frac{\partial \Delta V_{BE}}{\partial T} = 0$. For example, if V_{BE}'s temperature coefficient is $-1.5mV/K$, in order for the total temperature coefficient to equal to zero, n needs to be 8 with m of approximately 9.

2.3 Bandgap Reference

As discussed previously, summing the negative and positive temperature coefficient (TC) voltages together, a voltage reference that is independent of temperature variation can be obtained. In other words, 0 TC voltage:

$$V_{REF} = \alpha_1 V_{BE} + \alpha_2 [V_T \ln n] \tag{2.9}$$

where $V_T \ln n$ is the difference between the base-emitter voltages of two bipolar transistors operating at different current densities. As mentioned in the last section, for example, if the environment is set to $300K$, $V_{BE} \approx 750mV$, the temperature coefficient of V_{BE} is approximately -1.5mV/K, and $\frac{\partial V_T}{\partial T} \approx 0.081mV/K$. By setting α_1=1,

$$\alpha_1 \frac{\partial V_{BE}}{\partial T} + 0.081\alpha_2 \ln n = 0 \tag{2.10}$$

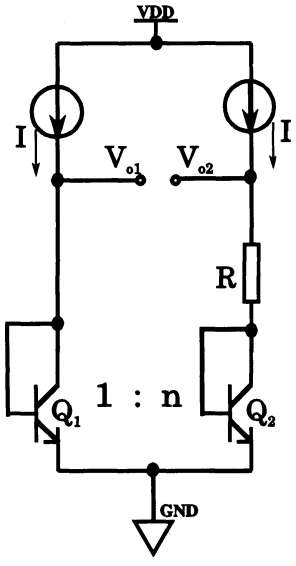


Figure 2.3: Bandgap-reference generating circuit diagram

it can be calculated that $\alpha_2 \ln n = 18.51$, so that

$$V_{REF} = V_{BE} + 18.51 V_T = 1.228V \tag{2.11}$$

where $V_T = 25.85mV$ at absolute temperature of $300K$.

The determination of the remaining variables α_2 and n is shown by the following example.

In Fig. 2.3, Q_2 has a junction area of n times as that of Q_1, so that the current density of Q_2 is n times less than that of Q_1. Suppose $V_{o1} = V_{o2}$ is forced true, then $V_{BE1} = V_{BE2} + RI$, where $RI = V_T \ln n$. If $\alpha_2 = 1$, then $\ln n = 18.51$.

V_{o2} can produce the exact voltage value demanded. However, there are two problems to be solved. First, $V_{o1} = V_{o2}$ must be forced true; second, $\ln n = 18.51$ is not practical as n should be $e^{18.51}$, a very high value in order to meet the demand of $\ln n = 18.51$. In this manner, an op-amp can solve the first problem while ratio circuit to increase α_2 is needed for solving the second problem.

2.4 Example of Bandgap Circuits

To further understand the bandgap reference circuits, two of the commonly used bandgap structures are shown in this section.

2.4.1 Simple Bandgap Circuits Without Op-amp

The simplest compensation for temperature dependence is first-order compensation, which happens when the difference between V_{BE} of the two bipolar transistors biased in different current densities is added to the base-emitter voltage of a bipolar transistor.

A simple implementation without an op-amp is shown in Fig. 2.4, [28]. Transistor Q_3 is n times larger than Q_4 in size and Q_2 is r times larger than Q_1. For Q_3 and Q_4, they form a current mirror that produces a current of $I_{Q_3}/I_{Q_4} = n$; while the density ratio between Q_1 and Q_2 is

$$\frac{D(Q_1)}{D(Q_2)} = rn \tag{2.12}$$

The voltage on R_2 is

$$\Delta V_{BE} = V_{BE1} - V_{BE2} = \frac{kT}{q}\ln(nr) \tag{2.13}$$

The output then will be

$$\begin{aligned} V_{REF} &= V_{BE1} + V_C \\ &= V_{BE1} + nI_{R_2}R_1 \\ &= V_{BE1} + n\frac{\Delta V_{BE}}{R_2}R_1 \\ &= V_{BE1} + n\frac{R_1}{R_2}\frac{kT}{q}\ln(nr) \end{aligned} \tag{2.14}$$

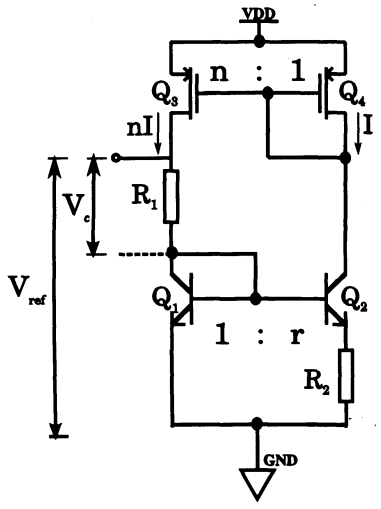


Figure 2.4: Example of an simple bandgap circuit diagram

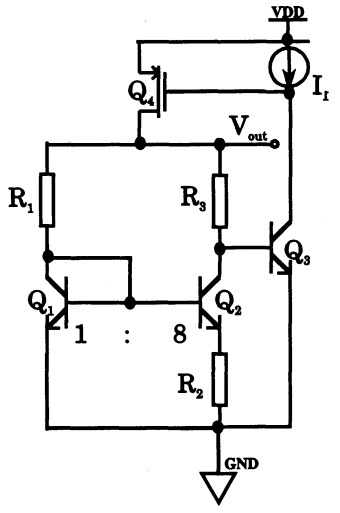


Figure 2.5: Simple bandgap with three Bipolar transistors circuit diagram

The simulation was done in Spice. With $R_1 = 40k\Omega$ and $R_2 = 102k\Omega$, the currents were set to $I_{Q_1} \approx 4\mu A$ and $I_{Q_2} \approx 1.03\mu A$ at $0°C$. Also because of this difference between the currents, the base-emitter voltages for Q_1 and Q_2 were different, forming ΔV_{BE} for this configuration. For example, in Fig. 2.4, $V_{BE2} \approx 0.74V$ and $V_{BE3} \approx 0.8V$ at $0°C$. In addition, the currents will vary according to the temperature variation, they are proportional to absolute temperature (PTAT) factors as can be seen in the simulation plot, Fig. 2.6.

Temperature dependance of reference voltage generated by circuit shown in Fig. 2.4 is shown in Fig. 2.7.

Another simple bandgap reference is shown in Fig. 2.5 [29]. The current source here produces a current for Q_3 so that V_{BE3} can be fixed as I_{C3} is fixed. The output voltage is the summation of V_{BE3} and the voltage across R_3. As the current through R_3 is a PTAT current, which has the positive temperature coefficient, the voltage across R_3 is a PTAT voltage, which can compensate for the negative temperature coefficient of V_{BE3}. The size ratio of the three bipolar transistors is $Q_1 : Q_2 : Q_3 =$

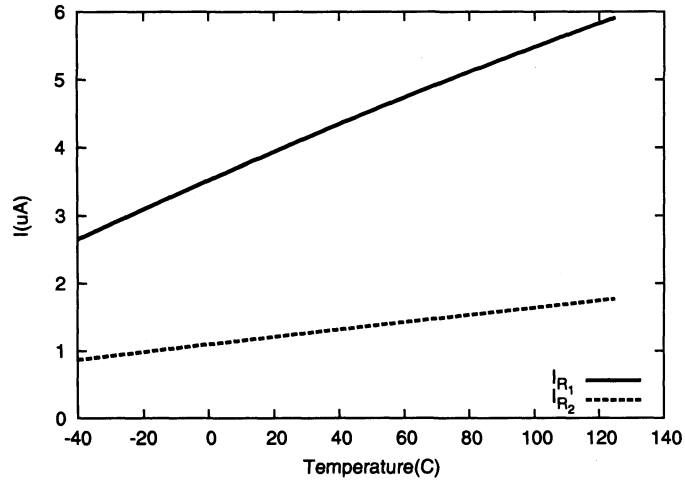


Figure 2.6: Simulation results of currents of the simple bandgap

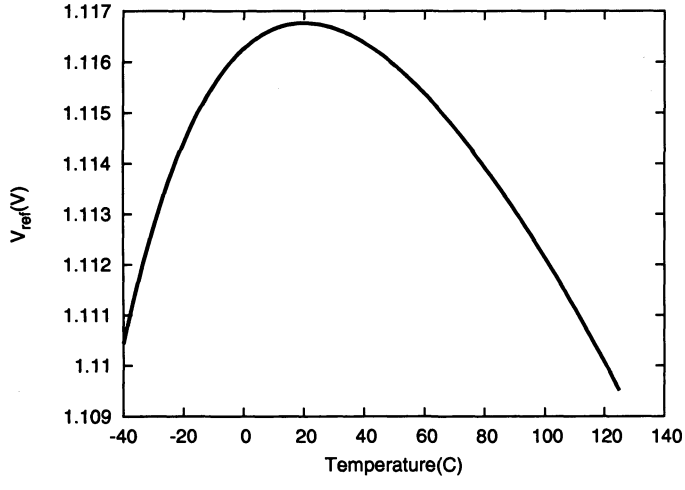


Figure 2.7: Simulation result of the simple bandgap output voltage

$1:8:1$. Then $\Delta V_{BE} = V_{R2} = \frac{kT}{q} \ln 8 = 46.06mV$ (at $0℃$).

As the *npn* bipolar model study showed, for a constant I_{CQ3}, for example $15\mu A$, V_{BE3} will have approximately a $-1.45389mV/K$ temperature dependence around room temperature. In Spice simulation, $R_1 = R_3 = 12k\Omega$ and $R_2 = 1.5k\Omega$ were applied, thus $R_3/R_2 = 12k\Omega/1.5k\Omega = 8$. As $\Delta V_{BE}/273K = 46.06mV/273K = 0.1797mV/K$ (at $0℃$), the PTAT part of the output has a positive temperature

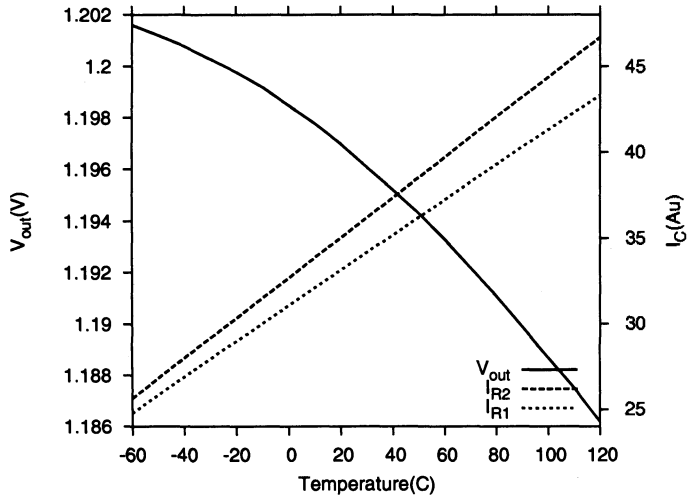


Figure 2.8: Simulation result of simple bandgap with three Bipolar transistors

dependence of $V_{R3}/273K = \frac{\Delta V_{BE}}{R_2} R_3/273K = 0.1797mV/K \times 8 = 1.4376mV/K$, which meets the demand for compensating the negative temperature dependence of V_{BE3} well. Figure 2.8 shows the output voltage of the circuit in Fig. 2.5 and the currents on R_1 and R_2. Voltage V_{out} varies from $1.205V$ to $1.186V$ with the temperature of $-60℃ \sim 125℃$, while the two currents are slightly different.

2.4.2 Bandgap Circuits With Op-amp

The most common configuration of a bandgap is composed of op-amp and ratio circuits, as Fig. 2.9 shows [26].

Points X, Y are at the same potential as the two inputs of the op-amp have the same DC voltage. In this way,

$$\begin{aligned} V_{REF} &= V_{BE2} + (R_2 + R_3) I_2 \\ I_2 &= \frac{V_{BE1} - V_{BE2}}{R_2} = \frac{V_T \ln n}{R_2} \\ V_{REF} &= V_{BE2} + (R_2 + R_3) \frac{V_T \ln n}{R_2} \end{aligned} \tag{2.15}$$

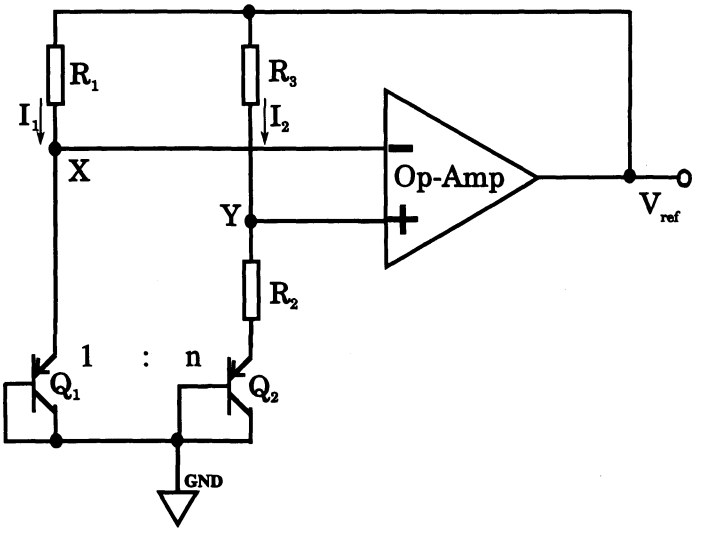


Figure 2.9: Schematic of bandgap reference with op-amp

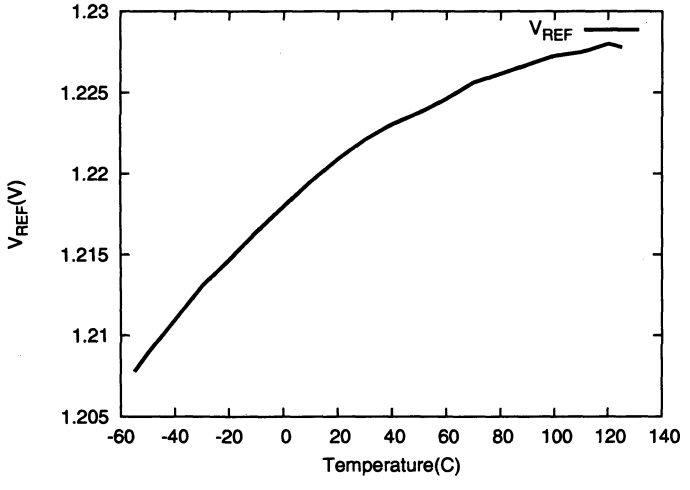


Figure 2.10: Simulation result of bandgap with op-amp output voltage

thus α_2 in (2.7) equals to $\frac{R_2+R_3}{R_2} = \frac{R3}{R2} + 1$ forms the ratio circuit. Figure 2.10 shows the simulation of the output of the circuit in Fig. 2.9

The *pnp* transistor model in the simulation for this circuit has a V_{BE} with a TC of $-2.0mV/K$ and $V_{BE} \approx 565mV$, $\alpha_2 \ln n = 24.69$ can be easily calculated. To minimize layout mismatch, the junction areas are chosen with a ratio of 1 : 8 so that when the same currents pass through the two BJTs, their current density ratio

is $8:1$. In this case, only when $\alpha_2 = 12$, $\alpha_2 ln8\frac{\partial \Delta V_{BE}}{\partial T} \approx 2.0mV/K$. This value can perfectly compensate for V_{BE}'s negative temperature coefficient. In order to meet the demand of $\alpha_2 = 12$ and to simplify the situation, the current was set to approximately $1mA$ by choosing $R_1 = R3 = 7.375k\Omega$ and $R_2 = 0.6676k\Omega$ respectively for ideal simulation in Spice. Here $R_1 = R3$ guarantees the currents through Q_1 and Q_2 are the same. However, in practical CMOS design, a limit of $0.1mA$ exists for the maximum current through the transistor, this limit is to minimize the errors due to the high base resistance caused by the large lateral dimension between the base contact and the effective emitter region [26].

In Fig. 2.10 as the temperature varies from $-55℃$ to $125℃$, the output voltage error is only $20mV$ $(1.2275V - 1.2075V = 0.02V)$, which is good for this simple configuration. The bandgap circuit with op-amp is expected to produce more accurate results. However, the simulation results in Fig. 2.7 and Fig. 2.8 had slightly better accuracy because the newer technology was used for the simulation in those bandgap circuits without op-amp while in the simulation for the circuit in Fig. 2.9, older technology and imprecise bipolar transistor models were used.

2.5 Start-Up Circuit

The common problem in all the references from the previous section is that there is a stable state at zero current point even if a power supply is provided. The bandgap circuit has two stable operating points. One is at zero current and the other one is at the desired current in the loop, Fig. 2.11. When there is zero current through the BJTs, the voltages at the input of the op-amp will also be at zero, plus there will not be any considerable difference between the two op-amp input voltages. This way the op-amp will not see any differential input and it will try to suppress any

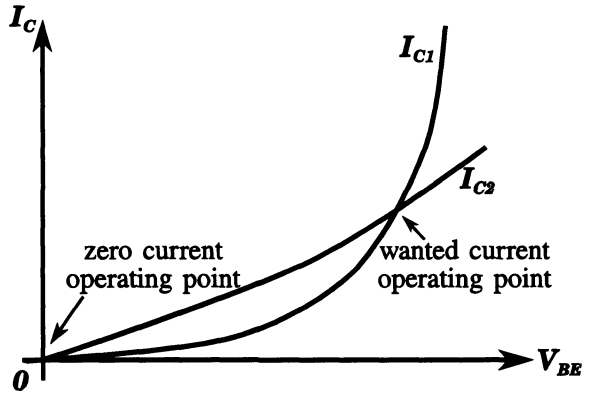


Figure 2.11: The two stable operating points in bandgap reference circuits

disturbance that may bring it out of the zero current state. That means the bandgap circuit can be locked in the zero-state forever if it is not forced out of it by an extra circuit. Although parasitic capacitance associated with every node may help to bring the circuit out of the zero current state,it is unpredictable. To avoid such unpredictability, a start-up circuit is used, which for sure prevents the circuit from getting locked in the zero current state [30]. In other words, an extra start-up circuit is needed for proper operation.

For further discussion on the start-up sub-circuit, please refer to Chapter 3.

2.6 Summary

In this chapter, an example is given for illustration of the bandgap reference's importance. The term of "bandgap" is explained and the concept of building a bandgap circuit is presented, which involves adding voltages that have opposite temperature coefficients. The case studies show the results of both the simple bandgap circuits and bandgap circuit architectures with op-amp. Also, the need of a start-up circuit in the design is presented. Next chapter gives further discussion on a specific bandgap circuit configuration and the details to build it are followed.

Chapter 3

Proposed Bandgap Circuit

In the previous chapter, bandgap voltage theory was presented and different bandgap voltage reference circuits were discussed. In this chapter, an accurate bandgap voltage reference working under $1.2V$ power supply is presented. The bandgap core circuit, the op-amp implementation, the start-up circuit and the output stage are discussed respectively.

3.1 Low-Voltage Bandgap Core Circuit

As it was presented in Chapter 2, the bandgap core circuit, nowadays, is usually composed by two or more bipolar-transistors. Thus, the analysis of BJT models in the available technology should be the first step to establish a whole circuit and the precise characterization and careful decision in choosing models guarantee the accuracy of the final output reference voltage. After choosing the model, two main concerns during the building of the circuit are the output's dependence on temperature change and power supply variation.

3.1.1 NPN Model Characterization

In CMOS90nm technology, there are two kinds of lateral *npn* bipolar-transistor available. One has an emitter area of $2\mu m \times 2\mu m$, and the other one has an emitter area

of $5\mu m \times 5\mu m$. Though the smaller one consumes much less area on chip, the bipolar-transistor, which is comparatively larger was chosen for its better characterization in accordance to temperature change. Figure 3.1 shows the base-emitter voltage V_{BE} variation with the temperature change when different collector currents are implemented. One can see that for higher temperature, the transistor has a lower V_{BE}. In other words, the bipolar transistor base-emitter voltage has a negative temperature coefficient, as it was previously mentioned. A family of curves with different current in Fig. 3.1 also shows that with higher I_C, V_{BE} is higher, 3.1.

$$V_{BE} = \frac{k}{q} T ln(\frac{I_c}{I_s}) \tag{3.1}$$

Also in this equation, it appears that V_{BE} should increase as temperature increases. However, V_{BE} acts oppositely because of the high temperature dependence of I_S.

It can be easily seen in Fig. 3.1, for different I_c, V_{BE} exhibits different temperature coefficient. For example, with a current of $1\mu A$, $V_{BEtc} = (786.214mV - 459.27mV)/(-40℃ - 125℃) = -1.98148mV/℃$; while with a current of $10\mu A$, $V_{BEtc} = (832.936mV - 538.652mV)/(-40℃ - 125℃) = -1.78354mV/℃$. In the available IC Design tool, the lowest temperature can be simulated is $-40℃$, however, if an extended temperature of $-273℃$ can be investigated, all those family curves with different I_C should converge to a single point, which has a V_{BE} of $1.205V$. This model, which is $5\mu m \times 5\mu m$ was chosen because with reference to the simulation, the manual calculation result of V_{BE} at temperature of $0K$ is closer to $1.205V$ than V_{BE} of the other model, which may provide more accurate results during the circuit development.

For a higher I_c, V_{BE} has a lower absolute temperature coeffient. While the current can not be infinitely high as low power consumption is demanded, $I_C = 10\mu A$ is

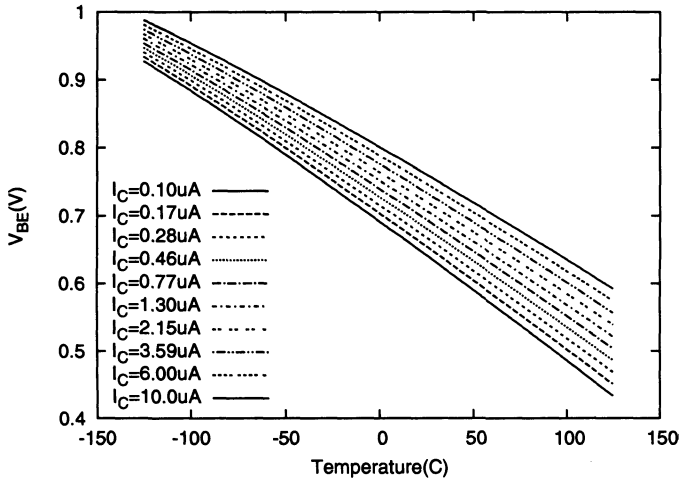


Figure 3.1: Simulation results showing the dependence of V_{BE} on temperature

chosen as the operating point of the bipolar transistors in the bandgap voltage reference circuit. Accordingly, the negative temperature coefficient to be compensated is fixed at $-1.78354mV/°C$.

3.1.2 Bandgap Voltage Generating Circuit

Given the intention that the nominal power supply should be at $1.2V$ in this low-voltage circuit design, it would be impractical to design a traditional bandgap whose output voltage is around $1.205V$ (Refer to Chapter 2.1). However, the temperature-independent output voltage can be realized by taking a fraction of the original bandgap voltage, specifically, using current terms proportional to it. In Fig. 3.2, V_A and V_B are forced equal by the op-amp. The voltage across R_0 is a PTAT voltage as $V_{R0} = \Delta V_{BE} = V_{BE1} - V_{BE2}$. Thus, the current through R_0 is PTAT (assuming R_0 is temperature independent). The voltage value across R_1 and R_2 equals to the base-emitter voltage of the diode connected BJT transistor Q_1, which has a negative temperature coefficient. The current through M_2 (also M_1, because $I_1 = I_2$ as M_1 and M_2 are identical and have the same V_{GS} and V_{DS}) can be temperature

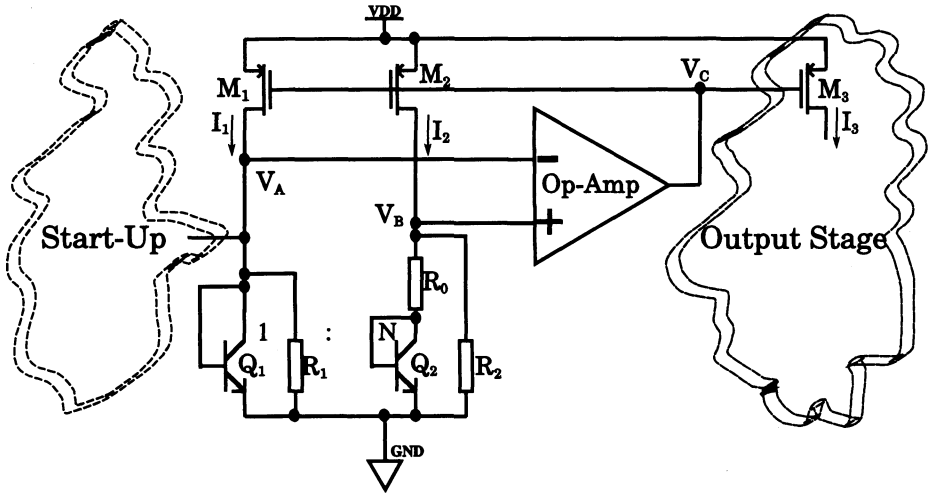


Figure 3.2: Bandgap voltage generating circuit diagram

independent, as follows:

$$I_1 = I_2 = \frac{V_{BE}}{R_2} + \frac{V_T \ln N}{R_0} = \frac{1}{R_2}(V_{BE} + \frac{R_2}{R_0} V_T \ln N) \tag{3.2}$$

I_3 can be slightly different, depends on the size of M_3 and the load. (Refer to Chapter 1.4) However, I_3 should be proportional to I_1 and I_2, since the gate of M_3 is controlled by the same op-amp output. Assuming the load is a temperature independent resistor at this time, the output voltage is as follows,

$$V_{REF} = I_3 R_{load} = k I_1 R_{load} = \frac{k R_{load}}{R_2}[V_{BE} + \frac{R_2}{R_0} V_T \ln N] \tag{3.3}$$

If $V_T = kT/q$ is used to obtain a PTAT voltage, it is well known that V_T has to be multiplied by approximately 22 to compensate for the temperature dependence of the diode voltage [31]. Thus,

$$\frac{R_2}{R_0} \ln N = 22 \tag{3.4}$$

For the sake of BJT transistor layout matching issues, $N = 8$ is used. (Refer to Chapter 4) R_2/R_0, therefore, should be around 10. According to the current needed in the two bandgap core branches, R_0 and $R_2(R_1)$ can be determined. The decision of current is also related to the power change tolerance of the circuit, which is discussed in the next section.

3.1.3 Sensitivity to Power Supply Variation

In Fig. 3.2, the bandgap voltage generating circuit, M_1, M_2 should be large enough to tolerate the power change. The saturation current equation shows:

$$I_{D1,2} = \frac{1}{2}K_p'\frac{W}{L}(V_{GS1,2} - V_{th})^2(1 + \lambda V_{DS1,2}) \tag{3.5}$$

when the power supply disturbance is $\pm 10\%$, $1.2V \times \pm 10\% = \pm 0.12V$, $V_{G1,2}$, which are controlled by the op-amp should be changed in order to hold the constant current $I_{D1,2}$. The op-amp is working in a feedback network to track the changes in $V_{S1,2}(V_{DD})$ to make sure that the range in which $V_{GS1,2}$ changes is small enough. As the op-amp in this design can not serve an extremely high gain, $V_{G1,2}$ can not be tracked perfectly. The purpose is to keep $I_{D1,2}$ as stable as possible. Even with the limitation of a low gain op-amp, the stablization can still be achieved by increasing the width to length ratio of M_1, M_2 to make $I_{D1,2}$ less dependent on $V_{GS1,2}$. In other words, $I_{D1,2}$ can be made more dependent on $(W/L)_{1,2}$, which does not change with the power supply. However, when the size (W/L) of a transistor is very large, while the drain current remains very low, the transistor may even enter into subthreshold conduction region. This situation is because the constant current is more dependent on the size other than the gate-to-source voltage, which means less gate-to-source voltage is needed to maintain the same current, even less than the threshold voltage. The proper way is to increase the current of each transistor, M_1, M_2 to ensure

these large devices are actually open. Transistors M_1, M_2 are guaranteed to work in saturation region as the drain voltage is approximately equal to V_{BE}, which leaves around $500mV$ for $V_{DS1,2}$ if the power supply is $1.2V$. (Also see section on 'Output Dynamic Range and the Load' in 'Op-Amp Design Considerations', Chapter 3.2)

To only make sure V_{GS} is large enough to open M_1, M_2 transistors, for instance $V_{GS} - V_{th} \leq 50mV$ is not sufficient to meet the tolerance demands for the high power supply change. The saturation current of M_1, M_2 is a function of $(V_{GS} - V_{th})^2$, if $V_{GS} - V_{th} \approx 0V$ or $V_{GS} - V_{th} \leq 50mV$, a slight disturbance on V_{GS} can cause a comparatively significant change on the current. For example, assume the nominal power supply is $1V$, a transistor with $V_{th} = 220mV$ works at $V_{GS} = 230mV$ when no disturbance occurs, then $(V_{GS} - V_{th})^2 = 0.0001V^2$, which is proportional to the drain current, and $I_D = h(V_{GS} - V_{th})^2 = 0.1h \times 10^{-3}$. If there is a disturbance on V_{DD} so that it changes to $1.1V$, then the instantaneous $V_{GS} = 330mV$ forces $(V'_{GS} - V_{th})^2 = 0.0169V^2$. The drain current $I'_D = 0.0169h$, and $I'_D/I_D = 169$. On the other hand, if V_{GS} is $320mV$ for nominal power supply, a $0.1V$ increase in the power supply can cause less disturbance at the drain current:

$$\begin{aligned} I'_D/I_D &= (420mV - 220mV)^2/(320mV - 220mV)^2 \\ &= 0.04/0.01 = 4 \ll 169 \end{aligned} \tag{3.6}$$

Obviously, as V_{GS} increases, the drain current will be enlarged if other parameters do not change. The current through the bandgap paths can not be enlarged infinitely as to keep the system within a certain power consumption. The only other parameter that can be changed is (W/L). If (W/L) is reduced, with the same drain current, V_{GS} can be made larger. It is a triangular trade-off between power consumption, V_{GS} and (W/L) ratio. A lot of work can be done to optimize this trade-off.

Two cases are studied with the same diagram and similar path current, in order

Table 3.1: The effect of different V_{GS} on power supply variation

Parameters	Case 1	Case 2
W/L	$1.6\mu m/0.8\mu m = 2$	$2\mu m/1\mu m = 2$
m	50	10
$V_{GS} - V_{th}$	$-10.7mV$	$-106.9mV$
I_D	$27.73\mu A$	$21.75\mu A$
I_D Variation with $\pm 10\%$ V_{DD} Change	$\pm 20nA$	$-44nA \sim +24nA$
Output Voltage	$812mV$	$517.5mV$
V_{out} Variation with $\pm 10\%$ V_{DD} Change	$-17.74mV \sim 7.97mV$	$0.3mV \sim 0.9mV$

to prove the effect of V_{GS} on power supply variation tolerance, Table. 3.1. Case 2 apparently produces less output variation with power change, the transistor size of which is 1/5 of that in Case 1.

3.2 The Design of Op-Amp for Use in Bandgap Reference

The op-amp is used to provide a more precise output voltage for the bandgap reference circuit. The specific application needs in bandgap circuit defines the performance demands for the op-amp.

3.2.1 Op-amp Blocks and Two Stage Op-Amp Circuits

Ideally, the operational amplifier is a voltage-controlled voltage source with infinite voltage gain and zero input admittance as well as zero output impedance [32].

In order to obtain the nearest ideal performance for a practical op-amp, the general block diagram is shown in Fig. 3.3.

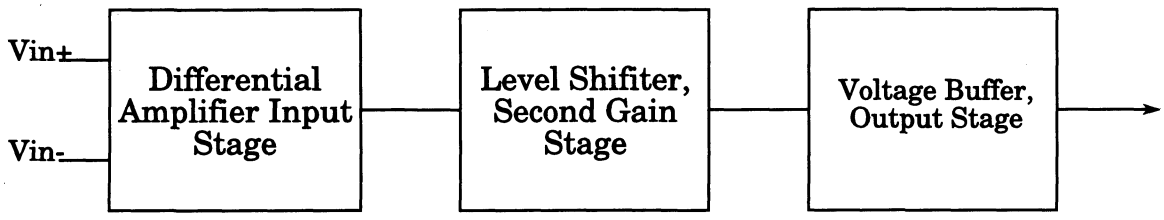


Figure 3.3: Blocks of a general op-amp

The differential input stage's aim is to achieve high input impedance, large common-mode rejection ratio (CMRR) and high gain. Its output should usually be single-ended in order to drive the next stage. The middle stage plays two roles in the op-amp circuit. First, it adds more gain to the input stage. Even though the differential amplifier (input stage) provides good gain it is still not enough when it comes to an op-amp circuit. The second stage amplifies the gain even further into several hundred times. For example, if the first stage gives a gain of 30 times, and the second stage gives a gain of 20 times, then the total gain should be $30 \times 20 = 600$ times. Second, the second stage serves as a level shifter, which compensates for the DC voltage change from the input stage. The last stage is usually an output voltage buffer, which provides a very low output impedance and strong current drive capability.

3.2.1.1 Conventional Two-Stage Operational Amplifiers

According to the three blocks of an op-amp, a commonly used op-amp architecture is shown in Fig. 3.4

The role of the differential stage, as the first stage, is to amplify the difference

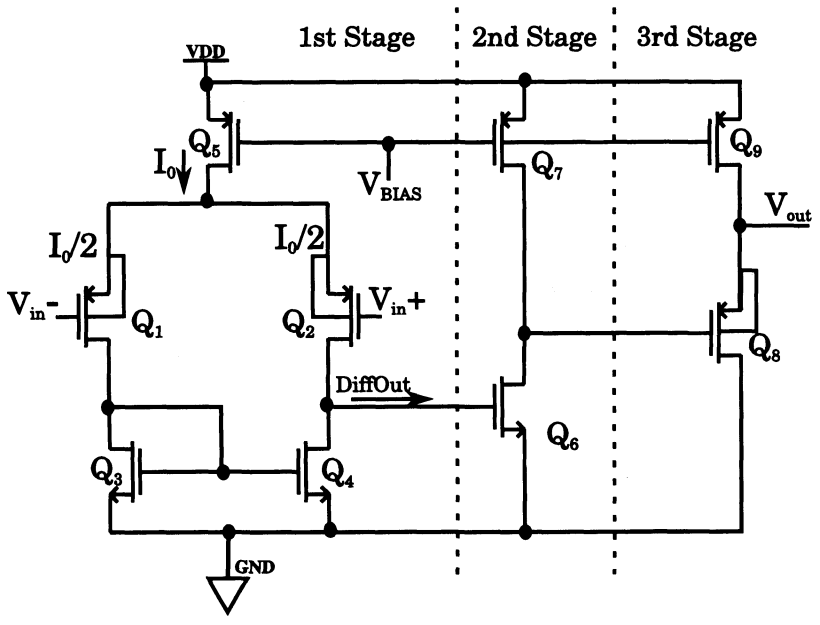


Figure 3.4: A conventional op-amp circuit diagram

between the two input voltages. It provides a voltage differential gain of

$$\begin{aligned} A_{diff} &= V_{diff-out}/(v_{in}^{+} - v_{in}^{-}) \\ &\approx g_{m1} r_0 \\ &= g_{m1}(r_2 \| r_4) \\ &= \frac{g_{m1}}{g_{d2} + g_{d4}} \end{aligned} \tag{3.7}$$

As can be seen from the equation: $g_{m1} = 2\sqrt{k'(W/L)I_o/2}$ the transconductance of the input transistor can be increased by increasing its ratio of width to length or the current through the path. The current cannot be increased too much as it consumes much power. Therefore, the two input transistors should be made extremely large in width while it is also acceptable to reduce the length of the two input transistors. It results in the uniqueness of these two transistors compared to other transistors in the circuit (especially the length, it can be different from the other transistors, as

the layout of these two transistors are separately developed rather than together with the other transistors in the op-amp circuit, and also the direct connection between their bodies to sources is allowed, which eliminates the body effect from these two transistors.)

The second stage is a common source amplifier, which acts as a level shifter and a second gain stage. Transistor Q_6 produces the gain while Q_7 acts as a load. The gain of the second stage is

$$A_{vcs} = \frac{-g_{m6}}{g_{d6} + g_{d7}} \tag{3.8}$$

where $\frac{1}{g_6+g_7}$ is the output impedance of the second stage. Up to the end of the second stage, the overall voltage gain is $A_v = A_{diff} \cdot A_{vcs}$, which can be as high as 80dB according to the simulation results in CMOS 90nm technology.

The third stage is composed of Q_8 as a voltage buffer and Q_9 as the load. This stage is the output buffer whose gain is less than 1. It may cause some loss in the overall gain, however it eliminates the output impedance.

In the bandgap circuit, the op-amp output is used to drive gates of the other PMOS transistors whose capacitance is small, typically less than $2pF$. In this situation, the 3^{rd} stage summation can be omitted and the op-amp will then function as a voltage-controlled current source (VCCS), in other words, an operational-transconductance amplifier (OTA) whose output impedance is not close to zero. As the op-amp has a high voltage gain and is in a stable feedback network, its output impedance is reduced to a very low value, and the difference between the performance of an op-amp and an OTA can be neglected [32].

The two stage op-amp without the voltage buffer can be designed as in Fig. 3.5.

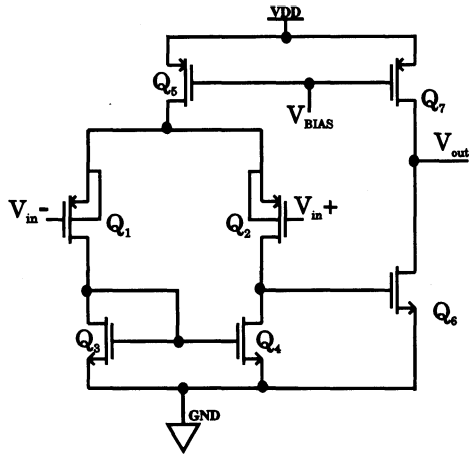


Figure 3.5: A two stage op-amp circuit diagram

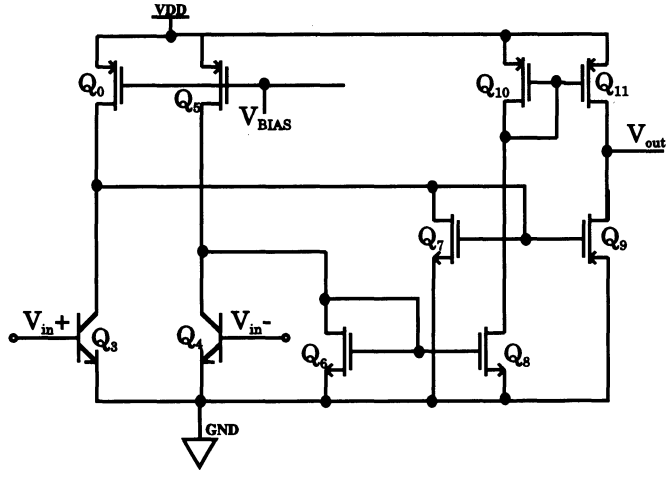


Figure 3.6: Circuit diagram of the op-amp proposed in [1]

3.2.1.2 Operation Analysis at Low Temperature of The Op-Amp in The Referred Work

The op-amp circuit proposed in [1] is as shown in Fig. 3.6. This circuit is aimed for low voltage power supply, around $1V$ in [1]. For example, the folded structure is used frequently and as a result, there are no more than two tansistors in every path between the power supply rails. The combination of the diode connected BJT in the bandgap core and the BJT of the input stage constitutes a current mirror [1]. The

current generated by Q_3 and Q_4 is then folded and collected by two diode connected MOS transistors Q_6 and Q_7. The output would be a current difference between Q_6 and Q_7 while the currents on them are mirrored into Q_8 and Q_9, the next stage. As explained in [1], the gain of the differential stage is most likely to be around 8. This gain is quite low compared to traditional CMOS input differential stage even with the addition of the next stage.

When implementing the op-amp into the bandgap circuit, with the different sizes of $Q_{10}(Q_{11})$ and $Q_8(Q_9)$, the bandgap reference voltage output varies differently with the temperature change. The result is interesting for when the size ratio between $Q_8(Q_9)$ and $Q_{10}(Q_{11})$ is less than 8/3, the bandgap behaves worse within the low temperature region (negative temperature) with a decrease of the width to length ratio, Fig. 3.7. There must be transistors not working properly when the temperature is low. Moreover, the inputs of the differential stage are currents fed into $Q_3(Q_4)$'s base nodes from the collector current of $Q_1(Q_2)$. The currents on $Q_1(Q_2)$ are so substantial in the whole bandgap circuit that seemingly very trivial disturbance (having very little currents fed into or driven from them) can influence the temperature independence of the currents through the two core paths, thus reduces the accuracy of the output voltage in response to temperature variation.

3.2.1.3 The Op-Amp Used in This Bandgap Circuit Design

The op-amp proposed in [1] provides less gain compared to the conventional two stage op-amp. The most attractive part of this op-amp is that it works under $1V$ power supply, which is not necessary in this design with a nominal power supply of $1.2V$. With a better gain, the circuit can perform much better. Figure 3.8 shows a reliable op-amp implementation based on the two stage op-amp discussed previously. This configuration is an improved version of the one shown in Fig. 3.4, which provides

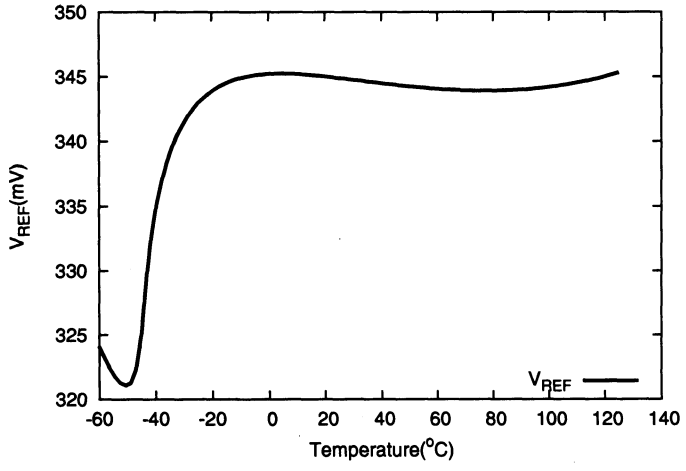


Figure 3.7: V_{REF} vs. temperature using the op-amp proposed in [1]

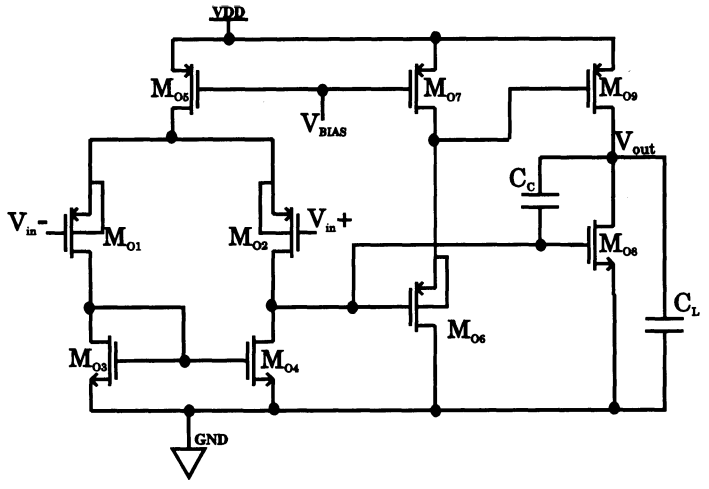


Figure 3.8: Proposed two stage op-amp circuit diagram

increased output range and current drive capacity. The first stage is the same differential stage and the second stage, which is also the output stage, is composed by transistors M_{o6} to M_{o9}. Transistors M_{o6} and M_{o7} act as the level shifter while M_{o8} and M_{o9} act as a class AB push-pull output [32].

This implementation achieves the elimination of crossover distortion of class B output stage and reduces the quiescent power dissipation of class A output stage

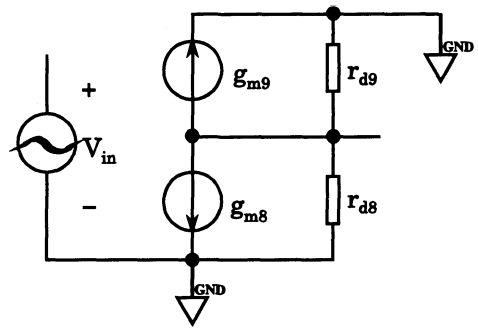


Figure 3.9: Low-frequency small-signal analysis of the output stage in Fig. 3.8

while serves a better gain than the circuit in Fig. 3.4.

When the output signal of the first stage is above M_{o8}'s V_{thn}, the drain current of M_{o8} is larger than the drain current of M_{o9}, thus the output stage pulls a current from the load, or vice versa. The crossover distortion is eliminated by the biasing of M_{o6}, who acts as a level shifter, having M_{o8} and M_{o9} biased at a small, nonzero current continuously. Therefore, M_{o8}, M_{o9} allow for the rail-to-rail output range.

Compared to the architecture in Fig. 3.4, this push-pull output op-amp also gives better gain. Consider, for example, the low-frequency small-signal gain of M_{o8} and M_{o9}, Fig. 3.9

$$(g_{m8} + g_{m9})v_{in} + (g_{d8} + g_{d9})v_{out} = 0$$

$$A_{V push-pull} = \frac{v_{out}}{v_{in}} = -\frac{g_{m8} + g_{m9}}{g_{d8} + g_{d9}} \tag{3.9}$$

Since $g_m \gg g_d$, the gain A_{V2} is high. It can be higher than A_{VCS} in 3.8.

Thus, $A_V = A_{diff} \cdot A_{V push-pull}$, and it can be higher than the gain in Fig. 3.4. Also given that this op-amp is only required to drive three PMOS transistors' gates as the load (only very small capacitance no more than $10fF$), once it is implemented

in the whole circuit, the gain remains relatively high without any extra output buffer.

The frequency response, the frequency compensation, and consideration of other aspects are discussed in the following sections.

3.2.2 Op-Amp Design Considerations

3.2.2.1 The Input Stage

Before considering the open-loop gain, frequency response, input-common mode range, output range, etc., first the input common mode voltage must be figured out, as this op-amp is for special use in bandgap. The two input nodes of the op-amp are connected directly to the collector node of the bipolar-transistor and a resistor whose node voltage is very close to the collector voltage of the bi-polar transistor, Fig. 3.2. At room temperature, V_A and V_B are equal to $694.9mV$, approximately.

That means, in prospective of optimization for those characters mentioned above, the input voltage should be fixed in a very small range. In the bandgap circuit, the chosen *npn* transistor model's base-emitter voltage is around $695mV$ when operating under room temperature. So the input common-mode range in consideration should be $650mV \sim 750mV$ (the base-emitter voltage varies in this range according to the temperature change from $-20℃$ to $50℃$).

The input common-mode range of a *p-channel* input differential stage and a *n-channel* input differential stage are different. For example, in the *p-channel* input differential stage of Fig. 3.5, the drain-to-source DC voltage of Q_1 is

$$V_{DS1} = V_{SS} + V_{GS3} - (V_{in,CM} - V_{GS1}) \tag{3.10}$$

To ensure that transistor Q_1 works in saturation:

$$\begin{aligned} |V_{DS1}| \geq & |V_{GS1}| - |V_{thp}| \\ -V_{DS1} \geq & V_{thp} - V_{GS1} \\ V_{in,CM} - V_{GS1} - V_{SS} - V_{GS3} \geq & V_{thp} - V_{GS1} \\ V_{in,CM} \geq & V_{thp} + V_{SS} + V_{GS3} \\ V_{in,CM} \geq & V_{thp} + V_{SS} + V_{Dsat3} + V_{thn} \end{aligned} \tag{3.11}$$

$V_{in,CM}$ has a minimum voltage of V_{SS} plus the drain-to-source saturation voltage of Q_3, as V_{thp} is negative and has a similar value to V_{thn}, this value results in $V_{thp} + V_{thn} = 0$. The drain-to-source DC voltage of Q_5 is

$$V_{DS5} = V_{in,CM} - V_{GS1} - V_{DD} \tag{3.12}$$

For Q_5 to also work in the saturation region,

$$\begin{aligned} |V_{DS5}| \geq & |V_{GS5}| - |V_{thp}| \\ -V_{DS5} \geq & |V_{GS5}| - |V_{thp}| \\ V_{GS1} + V_{DD} - V_{in,CM} \geq & -V_{Dsat5} \\ V_{in,CM} \leq & V_{Dsat1} + V_{thp} + V_{DD} + V_{Dsat5} \end{aligned} \tag{3.13}$$

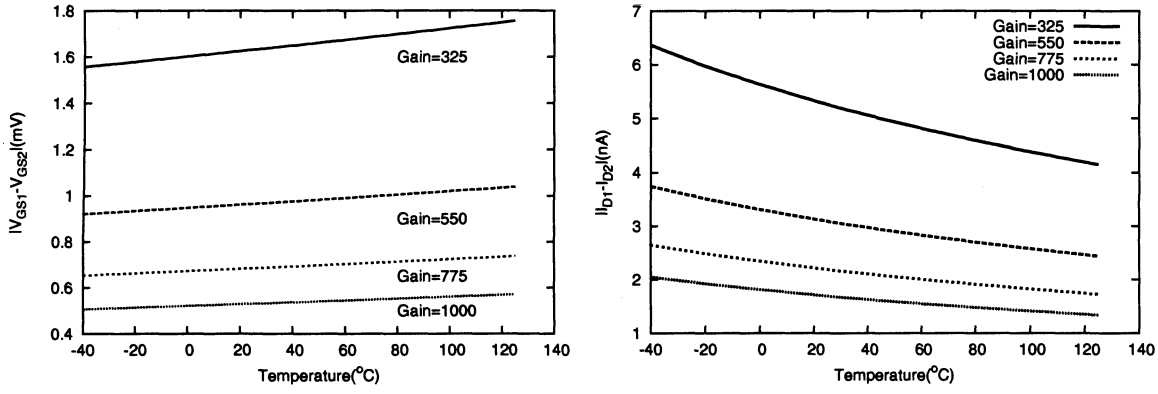
$V_{in,CM}$ has a maximum value of $V_{DD} - |V_{Dsat5}| - |V_{thp}| - |V_{Dsat}|$. If V_{DD} is $1.2V$ and $|V_{thp}| > 0.3V$, then $V_{in,CM}$ should be lower than $0.8V$. From this point, it is clear that the *p-channel* input differential stage demonstrates a good negative common-mode swing while its positive common-mode swing is poor. Similarly, one can prove that *n-channel* input differential stage offers a good positive common-mode swing but poor negative common-mode swing. As stated above, this op-amp only needs a

very small input common mode range, which is($650mV \sim 750mV$), which is in the middle of the rail-to-rail voltage range (given that power supply is $1.2V$), either the *p-channel* input differential stage or the *n-channel* input differential stage is suitable here for building the op-amp. The *p-channel* input differential stage is chosen for this design.

3.2.2.2 The Output Dynamic Range and The Load Capacitance

The purpose for the op-amp is to drive the PMOS transistors ($M_{1,2}$) in bandgap paths as a feedback, controlling the currents in these two paths by controlling the gate voltage of the two PMOS transistors, the output voltage of the op-amp can be determined by the demands of the gate voltage of the PMOS transistors, Fig. 3.2. Firstly, to ensure that the PMOS transistors are turned on, the output voltage of the op-amp should be lower than $V_{DD} - |V_{thp}|$. Secondly, it is easy to make these two transistors work in saturation region as V_{DS} is approximately as large as $V_{DS} = 1.2V - 0.7V = 0.5V$. In other words, as long as the output voltage can reach the range of $V_{DD} - |V_{thp}| - |V_{DS}| \leq V_{Opamp-out} \leq V_{DD} - |V_{thp}|$, the op-amp output stage is doing a satisfied job. In this circuit, the output voltage of the op-amp can be within $[0.3V, 0.8V]$

The load capacitance of this particular op-amp should be very small, for instance $100fF$, as it only drives the gates of three PMOS transistors. This small load capacitance is important when characterizing the op-amp because it influences the frequency and gain, etc. (Refer to the section of 'Frequency Response and Compensation', Chapter 3.2)



(a) Drain voltage difference varies with gain

(b) Drain current difference varies with gain

Figure 3.10: Simulation results for op-amp gain determination

3.2.2.3 The Op-Amp Gain

The op-amp is particularly designed for the bandgap circuit, its output voltage swing matters a lot as it supplies the bias voltage for the bandgap core circuit. First, it should be made low enough ($|V_{GS}| > |V_{thp}|$) to drive the PMOS load transistors. Second, it can not be too low because in that case a larger gain is needed for the op-amp to compensate for the bias voltage in PMOS load transistors. By simulating using the ideal op-amp (Voltage-Controlled Voltage Source), with a close enough PMOS driving voltage, such as $430mV$, the gain of the op-amp can be determined by the demands to compensate the biasing circuit. If the output common mode voltage of an op-amp is close enough to the biasing demands, then the op-amp does not have to supply a high gain, which is easier in this case.

The result showed that if the op-amp has a gain of $60dB$, it's already very good and if it's $40dB$, it's still working well under both temperature variation and power supply disturbance. Upon research and simulation, the common-mode output voltage can be set to $427mV$ when simulating with ideal voltage-controlled voltage source.

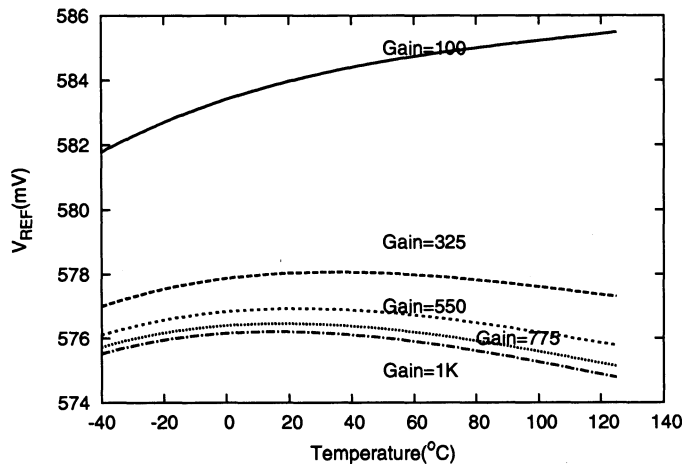


Figure 3.11: V_{REF} varies with gain

In the built op-amp, the common-mode output voltage is at $485mV$. With this common mode output voltage to serve as the load's gate voltage in the two bandgap paths, a sweep upon different gain of the ideal-op-amp is conducted to show the determination of gain. As Fig. 3.10 shows, the higher the gain is, the less the drain current difference and also the gate-source voltage difference between the two transistors (M_1, M_2) in the circuit shown in Fig. 3.2 can be obtained. This slight difference in drain currents with the same V_{GS} is caused mainly by the channel-length modulation as the transistors are small in scale. To eliminate this current difference, one way is to increase the gain of the op-amp and the other, is to increase the length of the two transistors M_1, M_2. One of the limitations for $90nm$ technology under $1.2V$ power supply is that the op-amp's gain cannot be made too large, for this reason M_1 and M_2's channel lengths should be increased. The trade-off between the op-amp gain and channel length of M_1 and M_2 should be considered too. After all of the investigation and simualtions based on the considerations discussed above, the results in Fig. 3.11 shows that for the chosen transistor $(M_{1,2})$ length, an op-amp gain of 100 can give a close enough result.

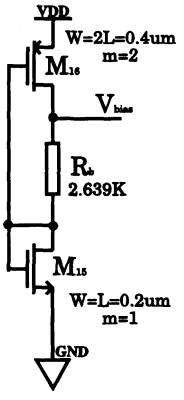


Figure 3.12: Biasing circuit diagram for the proposed op-amp

3.2.2.4 The Biasing Circuit for The Proposed Op-amp

The biasing voltage required by the current source in the differential pair is proportional to the power supply voltage. It should be $0.7V$ for $1.2V$ supply, $0.63V$ for a $1.08V$ supply and $0.77V$ for a $1.32V$ supply. According to the characteristics of the biasing voltage, the simplest configuration is the resistor-composed voltage-divider between the two power rails (between $1.2V$ and the ground voltage potential), which however, is almost impossible to build for a low power consumption circuit. On one hand, if the biasing path current is limited to $20\mu A$, then the total resistance needed is $60k\Omega$, which takes up a very large area when realizing the layout. On the other hand, less resistance reduces the area while consumes more current and power. It is again desirable to use the transistor instead of a resistor to build the biasing circuit.

The circuit shown in Fig. 3.12 provides a proportional biasing voltage to the power supply voltage. Transistor M_{15} and M_{16} are diode connected to act as resistors (if R_b is negligible as it is quite small). The biasing voltage is derived from the drain node voltage of M_{16} while the small resistor R_b is used for raising the voltage slightly. As the transistors' widths and lengths can not be configured arbitrarily, the slight adjustments can be made by using a resistor R_b with a small value. The biasing

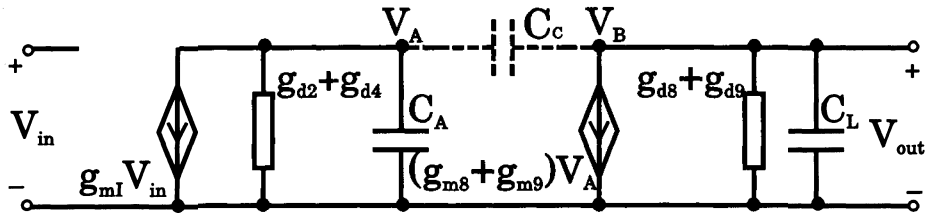


Figure 3.13: Small signal analysis of Fig. 3.8

voltage generated here are $683mV$ for $1.2V$ power supply (the nominal), $600mV$ for $1.08V$ and $770mV$ for $1.32V$ power supply. The slight variation of $25mV$ at the biasing voltage node when the temperature changes between $-40℃ \sim 125℃$ is negligible according to the simulation results.

3.2.3 Frequency Response and Compensation for the Proposed Op-Amp

Appendix A explains the general concept of frequency response and compensation for op-amp, and especially the Miller compensation for a two-stage op-amp. Having introduced these knowledge and techniques, the proposed op-amp is analysed and the compensated.

The small signal analysis of the circuit in Fig. 3.8 is as shown in Fig. 3.13 showing [32]:

The transfer function is:

$$\begin{aligned} A_V(s) &= \frac{v_{out(s)}}{\Delta v_{in}} \\ &\approx \frac{g_{m1}}{g_{d2}+g_{d4}} \frac{g_{m8}+g_{m9}}{g_{d8}+g_{d9}} \frac{1}{(1-s/s_A)(1-s/s_B)} \\ &= \frac{A_V(0)}{(1-s/s_A)(1-s/s_B)} \end{aligned} \tag{3.14}$$

where $s_A = -(g_{d2}+g_{d4})/C_A$ and $s_B = -(g_{d8}+g_{d9})/C_L$.

When frequency is high,

$$A_V(j\omega) = \frac{-A_V(0)}{\omega^2/s_A s_B} = \frac{-g_{m1}(g_{m8}+g_{m9})}{\omega^2 C_A C_L} \tag{3.15}$$

The amplifier inverts its input voltage under high frequency, which may cause an instability once plugged into a feedback system. However, the amplifier no longer inverts its input voltage for low frequencies as $A_V(j\omega) > 0$, and thus stability is achieved. This is because in a closed loop system, if at ω_1,

$$KH(s) = KH(j\omega_1) = -1 \tag{3.16}$$

the system will enter oscillation, App. A.

For bandgap circuit, the op-amp works close to DC signal, thus makes the compensation task easier. As high gain is needed and working under high frequency is not desirable in this implementation, some bandwidth can be sacrificed in order to get a high gain.

To compensate this op-amp, firstly, Miller compensation should be applied. By connecting node A and node B in Fig. 3.13, the two poles are splitted from each other. However, there is a zero rising. See the node equations for node A and B below:

$$g_{m1}V_{in} + (g_{d2}+g_{d4}+sC_A)V_A + sC_C(V_A - V_{out}) = 0 \tag{3.17}$$

$$sC_C(V_{out} - V_A) + (g_{m8}+g_{m9})V_A + (g_{d8}+g_{d9}+sC_L)V_{out} = 0 \tag{3.18}$$

From 3.18, V_A is achieved:

$$V_A = \frac{-sC_C V_{out} - (g_{d8}+g_{d9}+sC_L)V_{out}}{9_{m8}+g_{m9}-sC_C} \tag{3.19}$$

Using this result in 3.17, the transfer function can be obtained:

$$\frac{v_{out(s)}}{v_{in(s)}} = \frac{g_{m1}(g_{m8}+g_{m9}-sC_C)}{\frac{s^2[C_LC_A+C_CC_A+C_CC_L]}{(g_{d2}+g_{d4})(g_{d8}+g_{d9})} + \frac{s[C_C(g_{m8}+g_{m9})+(C_C+C_L)(g_{d2}+g_{d4})+(C_A+C_C)(g_{d8}+g_{d9})]}{(g_{d2}+g_{d4})(g_{d8}+g_{d9})} + 1} \tag{3.20}$$

If the two poles are far away from each other and if $C_L \gg C_C > C_A$, the poles and zero are as below:

$$p_1 = \frac{-(g_{d2}+g_{d4})(g_{d8}+g_{d9})}{C_C(g_{m8}+g_{m9})+(C_C+C_L)(g_{d2}+g_{d4})+(C_A+C_C)(g_{d8}+g_{d9})} \tag{3.21}$$

$$\approx \frac{-(g_{d8}+g_{d9})}{C_L}$$

$$p_2 = \frac{-[C_C(g_{m8}+g_{m9})+(C_C+C_L)(g_{d2}+g_{d4})+(C_A+C_C)(g_{d8}+g_{d9})]}{C_LC_A+C_CC_A+C_CC_L} \tag{3.22}$$

$$\approx \frac{-(g_{d2}+g_{d4})}{C_A+C_C}$$

$$z_1 = \frac{g_{m8}+g_{m9}}{C_C} \tag{3.23}$$

If $p_2 = z_1$,

$$\begin{aligned} \frac{-(g_{d2}+g_{d4})}{C_A+C_C} &= \frac{g_{m8}+g_{m9}}{C_C} \\ C_C &= \frac{-C_A(g_{m8}+g_{m9})}{g_{d2}+g_{d4}+g_{m8}+g_{m9}} \end{aligned} \tag{3.24}$$

Thus, by letting $p_2 = z_1$, z_1 can be cancelled and if the capacitive load is large enough, the dominant pole is determined by the output pole. The approximation in the second pole's function may have omitted its dependence on output capacitance too, which may result in the zero not being cancelled completely. Therefore, simulations had been done to adjust the position of pole 2. And also, by enlarging the load capacitance, the first pole is pushed close to low frequency, which means that $0dB$

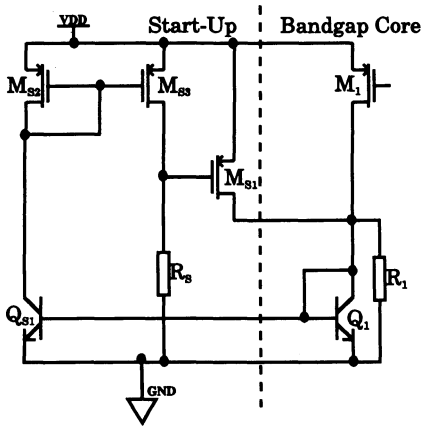


Figure 3.14: Start-up circuit diagram in design [1]

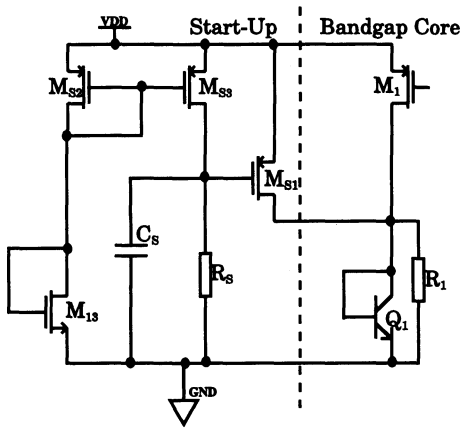


Figure 3.15: Improved start-up circuit diagram

bandwidth may not include the second pole and the zero, making them less important to consider. This result is obtained from a sacrifice of bandwidth for stability and high gain.

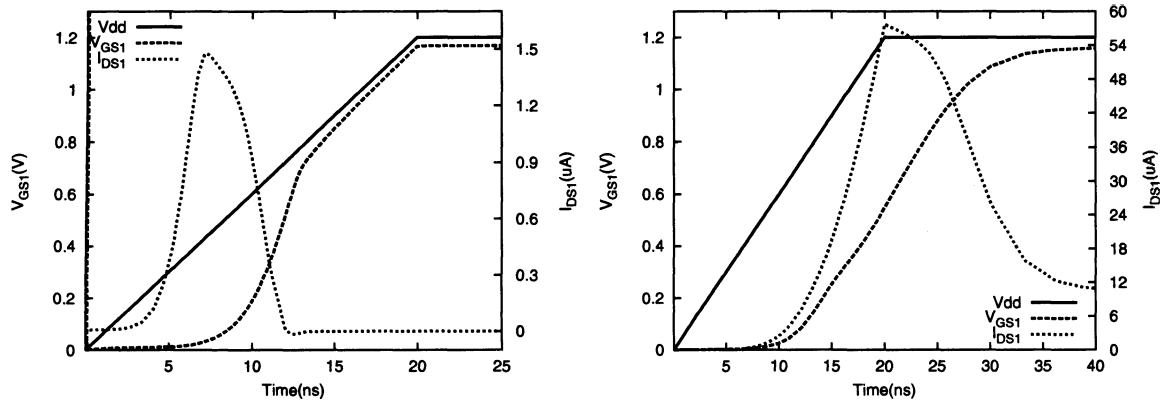
With the IC Design tool's *AC* analysis assistance, the bode plot of the compensated op-amp is shown in Fig. 5.1, and also some discussion on the results are given in Chapter 5.

3.3 Start-Up Circuit

As the circuit has two stable operating points at $I_{C0} = 0$ and $I_{C0} = 8.8\mu A$, a start-up circuit is needed to guarantee the desired mode of operation [1].

In the design in [1], the start-up circuit is as shown in Fig. 3.14. Once the power supply is on, the gate of M_{S1} is pulled down to the ground, thus injecting a significant current into R_1 and Q_1. At the end of the start-up phase, when the circuit reaches the nominal operating point, the current in M_{S3} makes the voltage across R_S high enough (close to V_{DD}) to turn off M_{S1}, thus the start-up circuit is turned off.

There are two problems about this start-up circuit.



(a) Setting-up of power and M_5 before adding the capacitor

(b) Setting-up of power and M_5 after adding the capacitor

Figure 3.16: The difference in setting-up of M_5 caused by the capacitor

Firstly, the slight base current injected into Q_{S1} is from a portion of I_{C1}, which significantly disturbs the balance in the bandgap core circuit, thus influences the output reference voltage, because it's not totally disconnected from the bandgap core circuit when M_{S1} is turned off. That means, there is a slight amount of current drawn from bandgap core circuit and then is fed into the start-up circuit continuously even when the circuit gets to the desired operating point. This current seriously influences the bandgap output voltage as the output voltage is highly dependent on the currents on M_1, M_2 and M_3, which are stable and temperature-independent.

To solve this problem, Q_{S1} can be replaced by a diode connected NMOS transistor M_{13}, as shown in Fig. 3.15. This replacement guarantees that the start-up circuit is completely disconnected from the bandgap circuit when it finishes the start-up task. The other problem is the drain current of M_{S1} appears quickly when the power is turned on but also disappears fast even before the power is totally set up. See Fig. 3.16(a).

To solve this problem, a transistor gate composed capacitor C_s is used between the gate of M_{S1} and ground to delay the rise of the gate voltage of M_{S1}. The plot

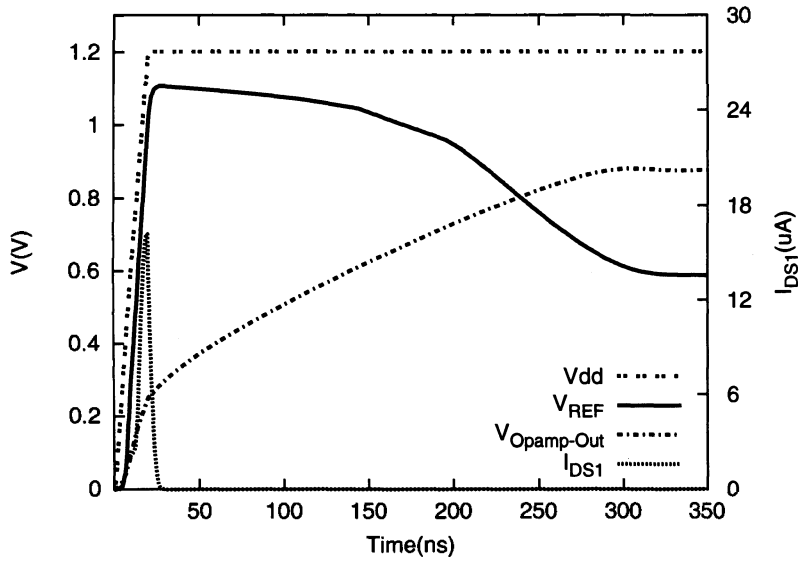


Figure 3.17: The response of V_{REF} and op-amp output when powering up

obtained after adding the capacitor is shown in Fig. 3.16(b). The other benefit this capacitor brings is that because of the delay, M_{S1} is given more time to turn on thoroughly, at the same time supplying considerably larger instant current whose maximum value is $16.5\mu A$, compared to the current in Fig. 3.16(a), which is only $1.5\mu A$.

Fig. 3.17 shows the response of the whole circuit when powered up. The op-amp starts powering up due to the disturbance introduced by the current from M_{S1} (the rising of I_{DS} is followed by the rising of the op-amp output voltage). Until op-amp finishes the power up, alternatively, when the op-amp becomes steady, and is able to provide a right DC voltage, the output voltage of the bandgap reference thus becomes steady, which is around $580mV$.

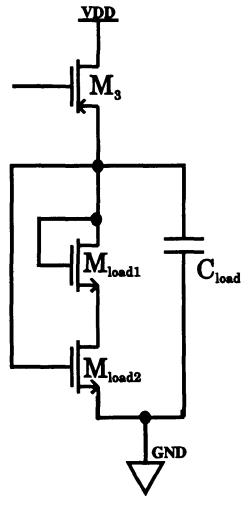


Figure 3.18: The output stage of bandgap circuit diagram

3.4 The Output Stage

As shown in Fig. 3.18, M_3 copies the bandgap current proportionally by being biased by an op-amp. For this application, the output current is fixed around $20\mu A$. In order to get a voltage output, a resistor should be added. As a resistor here needs very high value, for example $30k\Omega$ to generate an output voltage of $0.6V$, it is easier to have biased transistors to work as the resistor. In this way, it saves some area of the whole circuit and moreover, if the transistors are biased properly, it can get rid of the temperature dependence, which the real resistors in $90nm$ technology library have.

M_{load2} is biased in linear region while M_{load1} is diode connected to raise the V_{REF} voltage for a V_{thn}. The biasing of M_{load2} is done by the output voltage itself as this node is the most stable voltage node in the whole circuit according to the variation of temperature and the power supply voltage. This biasing concept makes the transistor-resistor's value independent of temperature and power supply voltage as the resistance of a transistor working in the linear region is only controlled by V_{GS}, which equivalently is V_{REF} for this circuit. A $0.75pF$ C_{load} capacitor composed by

three gate to body transistors, is added at the output for two reasons. First, it acts as a capacitive load to the bandgap voltage reference output; second, as the op-amp should be frequency-compensated to eliminate oscillation, the whole bandgap circuit needs to be compensated by a capacitor at the output too, according to practical demands.

3.5 Summary

In this chapter, a proposed bandgap circuit is presented. The *npn* transistor models are first investigated and the bandgap voltage generating circuit is described, with consideration in both temperature independence and power supply tolerance. Then an optimized op-amp is built as an important part in the bandgap reference. The start-up circuit with some improvement and a novel stable output stage circuit are discussed. The schematic of each block is given respectively through this chapter. The layout of each block is developed in the next chapter.

Chapter 4

Layout Design

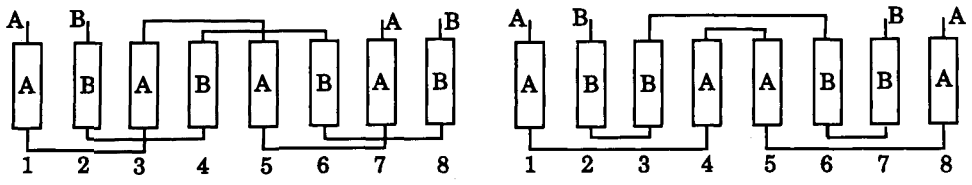
In the previous chapter, an accurate bandgap voltage reference circuit was presented, including bandgap core circuit, op-amp, start-up and output stage. In this chapter, some layout issues are first discussed with implementation illustrations for this work, followed by the layout realizations of each block. In the end, certain concerns are given for the whole bandgap voltage reference circuit layout and the finished layout figure is shown.

4.1 Layout Issues

For the purpose of good matching, there are some layout issues to be considered specially for analog circuit, in contrast of the digital circuit layout.

4.1.1 Using Unit Elements

In practical layout design, it is common to have long devices, such as long resistors, which have a width/length ratio of 1/200. These long devices are almost infeasible if one does not break it into the same units and connect them with wires. Usually, a unit resistor is laid out with some nominal resistance [33]. To find out a unit resistor's value, one should first check to see if all of the values have a greatest common factor [34]. For example, in the bandgap design, 2 p-type diffusion resistors with length=$56\mu m$, width=$1\mu m$ and another resistor of the same type with a length



(a) Interdigitated (b) Common centroid

Figure 4.1: The layout types illustration diagram

of $6\mu m$ and a width of $1\mu m$ are needed. In order to find the greatest common factor, the resistors with length of $56\mu m$ should be tuned down a little, to $54\mu m$ for instance, as long as the system can tolerate the change and work properly. Then in this case, the greatest common factor is $6\mu m$ and in total it is $1+9+9 = 19$ strips in the layout design.

The matched resistor segments should never contain less than five squares, and preferably they should contain at least ten [34]. Although partial resistor segments with a value larger than 70% of a complete segment is allowed in the array [34], it should be avoided as much as possible. When the changed parameter is brought back into the schematic, the changes in the result is acceptable, thus, the example above is doing a good job in both layout design and schematic change toleration.

4.1.2 Common-Centroid Layout

Common-Centroid (common centre) layout helps improve the matching between two resistors [33]. Consider the example in Fig. 4.1.

In Fig. 4.1(a), if there is a linear gradient variation increasing from left to right, demonstrated by a number from 1 to 8. The gradient influence on resistor A is $G(A) = 1+3+5+7 = 16$, and for B, is $G(B) = 2+4+6+8 = 20$. Gradients $G(A)$ and $G(B)$ have a large difference compared to the structure in Fig. 4.1(b) in which $G(A) = 1+4+5+8 = 18$ and $G(B) = 2+3+6+7 = 18 = G(A)$.

The common-centroid layout has four rules [34]: 1. Coincidence; 2. Symmetry; 3. Dispersion; 4. Compactness

The coincidence rule states that the centroids of the matched devices should at least coincide approximately. The symmetry rule suggests that the array should be symmetric to both the X_ and Y_ axes, and it should arise from the whole array instead of the individual segments (see example followed). Dispersion, apparently, means that the segments of each device should be distributed throughout the array as uniformly as possible. It helps to reduce the sensitivity of a common-centroid array to higher order gradients, which is nonlinear. And at last, compactness states that the array should be as compact as possible, ideally it should appear square.

Another resistor array in this layout is also designed according to these rules. In the design, three resistors that have the same material and same width are laid out in one island. Two of the resistors have the same length of $256\mu m$ and the other one has $24\mu m$. Since 24 has a factor of 3, it is much easier to change the lengths of the bigger resistors so that they have a factor of 3 as well, thus making the strips uniform. By tuning the length from $256\mu m$ to $264\mu m$ for the bigger resistors, the strips' lengths can be determined without much change of the bandgap output voltage. In the final island design, a length of $8\mu m$ as a unit was used, generating three strips for the small resistor and 33 strips for each of the bigger resistors. In order to make the strips coincide fairly and also symmetric, three rows of array each containing 23 strips is developed as shown in Fig. 4.2. Each strip in Fig. 4.2 demonstrates a resistor whose length is $8\mu m$ for each. Three **A**s are dispersed as well as possible among the three rows using common-centroid rule, so are all the **B** strips and **C** strips. Figure 4.3 is showing the final layout of this resistor island, also with the complementary p-type diffusion resistor island. As we can see, this three-row structure makes the whole island appear almost square.

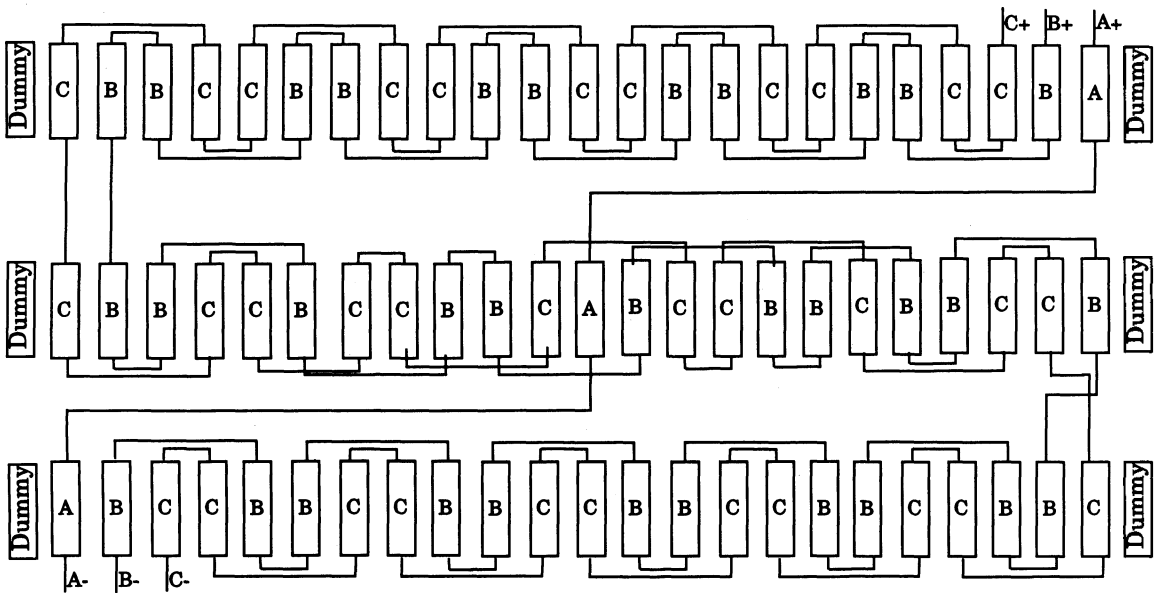


Figure 4.2: The n-type diffusion resistor array diagram

4.1.3 Dummy Elements

To compensate for the effect of mismatch caused by the different process between the outer and the inner unit cells, dummy elements can be added [33]. For example, in Fig. 4.2 the final amount of diffusion under the resistors, on the edges, is different between the inner and outer unit cells because of the different dopant concentrations. Thus, it is better to add dummy elements on both of the edges, left and right, Fig. 4.3.

The dummy elements are not connected into the circuit, so they do nothing electrically. They are introduced to guarantee that the matched units see the same adjacent structures. Usually, these dummy elements are covered by metal and tied off to either of the power-rails instead of leaving them floating.

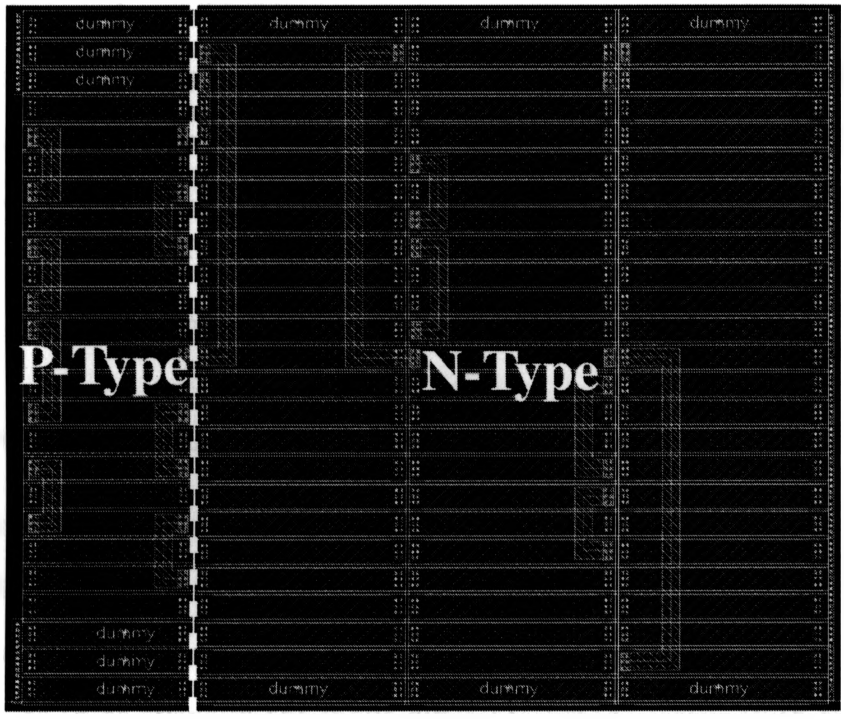


Figure 4.3: The resistor array layout including both diffusion types

4.2 Block Design

At the verification stage, it is very frustrating if the layout is made by building a whole circuit in one cellview. In order to easily track the errors in DRC and LVS, and fix them with the high efficiency, the whole circuit can be broken into blocks that are as small as possible. In this layout, the bandgap circuit is divided into several blocks including the op-amp block with the differential input stage sub-block, the bandgap core resistor array, the BJT array, the bandgap core PMOS current sources array, the gate-capacitor coin, the start-up circuit, and the output stage. DRC and LVS can be done separately at the block level before the blocks are put together.

4.2.1 Op-Amp Layout

The matching issue in op-amp is extremely substantial, because it is the op-amp, which affects the precision at the output and determines the bandwidth of the circuit in frequency domain. The most important part in an op-amp is the differential input stage, thus, the layout of these two transistors is done alone as a block in order to easily check for errors and to ensure good performance.

4.2.1.1 Differential Input Stage Layout

In the differential input stage of the op-amp, there are two PMOS transistors with a width of $10\mu m$ and a length of $0.2\mu m$. The tansistors are divided into twenty-five fingers in parallel, whose widths are $0.4\mu m$ for each finger. In other words, there are fifty fingers (gate strips)in the input differential stage array. With the fingering method, the active capacitance is less as the drain region is surrounded with gate poly instead of field [35], and also a considerable amount of area is saved by drain and source region sharing. According to the common centroid rule in "layout issues", the array is developed, Fig. 4.4. Label **A** and **B** stand for the two input transistors. These two PMOS transistors have their sources connected to the same voltage potential (the drain of the PMOS current source M_{o5} in Fig. 3.8). Thus, it is possible to share the diffusion area between the element transistors to save area [36]. Dual gates structure, for example, **AABB** instead of interdigited structure (for instance, **ABAB**) is also applied due to this reason.

Note that the dummy transistors at the edge of every row for both sides are also added for matching issues and a row of dummy transistors on both top and bottom of the array are added as well, Fig. 4.5. CMOS designs are more prone to latchup than standard bipolar [34]. Thus, in order to avoid the latchup effect, a guard ring is added around the PMOS differential input array. The guard ring is composed of the

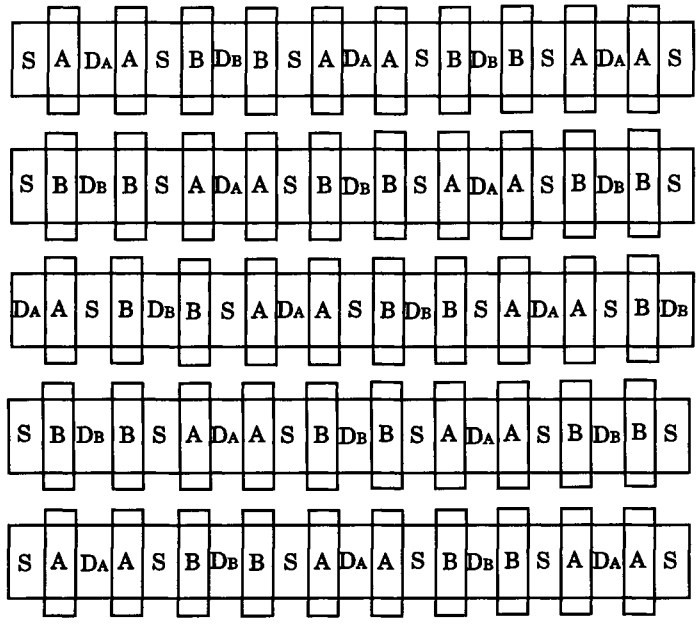


Figure 4.4: The differential input transistor array diagram

N-Plus layer, the diffusion layer and the metal contacts, which are to be connected to the power supply. As in the schematic of Fig. 3.8, the two input PMOS transistors' body terminals are connected to their source instead of directly being connected to the power supply. This connection is for the purpose of eliminating the body effect. In layout design, accordingly, the two input PMOS transistors should be built in an N-Well, which is tied to their source terminal S_{Mo1} and S_{Mo2}.

4.2.1.2 Layout of The Other Parts of Op-Amp

Similarly, in the op-amp design, M_{o6}'s body terminal is connected to its source terminal (Fig. 3.8), so this transistor needs a separate N-Well, which is tied to its source only, while other PMOS transistors that are sharing the same N-Well are tied up to the power supply. There is a trade off between body effect elimination and the area consumption.

The NMOS transistors in the op-amp can be built in a same N-Plus array and all

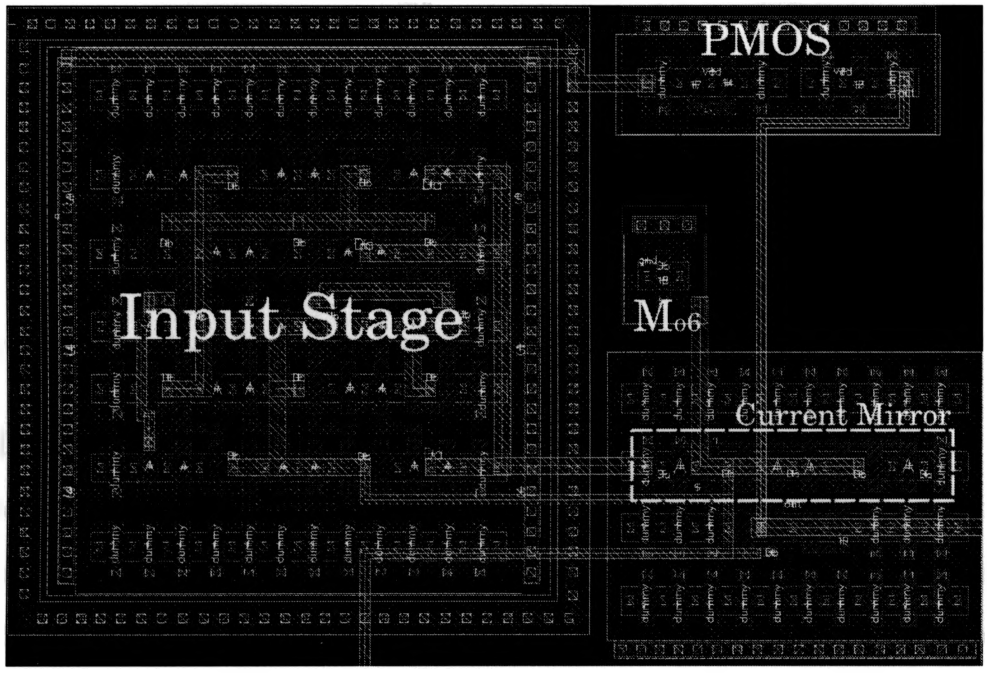


Figure 4.5: Op-amp circuit layout without capacitors

the matching and dummy rules mentioned previously apply to this part as well, especially for the current mirror. All the widths and lengths of the fingers are unified to $0.4\mu m \times 0.2\mu m$ at the schematic design stage for layout convenience. The transistors part of the op-amp is shown in Fig. 4.5.

4.2.1.3 Capacitors Layout

There are two capacitors in the op-amp for frequency compensation. As the schematic in Fig. 3.8 shows, one of the capacitors should be connected from the output node to ground, which is more suitable for gate-capacitor implementation for a higher square capacitance, and also all the metal layers can be overlapped on the gate poly to add an extra 10% capacitance to it. Figure 4.6 shows a cross-sectional view of the

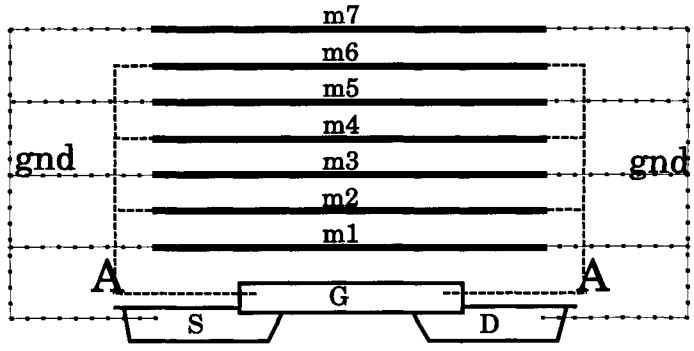


Figure 4.6: Cross section diagram of a unity coin gate capacitor

interdigited connection between the transistor layers and the metal layers.

The gate-capacitor coins layout is also shown in Fig. 4.7. Each coin consumes an area of $49\mu m^2$ ($7\mu m \times 7\mu m$) and forms approximately a $0.25pF$ capacitance. In the design, a $2.5pF$ gate-capacitor is needed, alternatively, an array of ten gate-capacitor coins is required.

The other capacitor is connected between the output node of the op-amp and the output node of the differential stage, which means that the gate-capacitor is unavailable for this connection, so a metal-metal capacitor is implemented for this purpose. There are seven metal layers in this technology, as also shown in Fig. 4.6, the connection is also similar to the one shown in Fig. 4.6. The area is the same as one gate-capacitor coin, $7\mu m \times 7\mu m$ as it is intentionally designed to be aligned with the gate-capacitor at the schematic design stage.

The overall layout view of the op-amp is shown in Fig. 4.7. It is made rectangular so its length will fit together with the BJT array.

4.2.2 Bandgap Core Resistor Array Layout

The resistor array for the bandgap core circuit is shown in Fig. 4.3. As discussed previously, by following the rules and the layout issues, the resistor array is developed

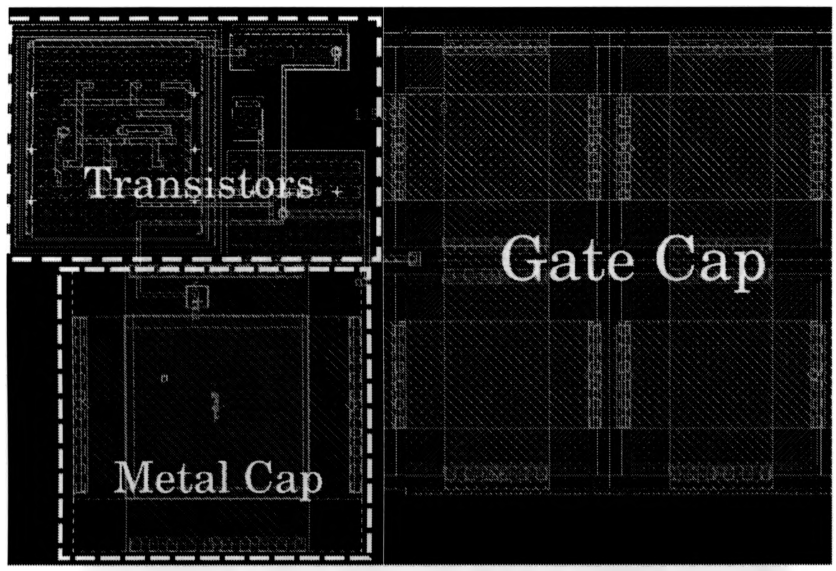


Figure 4.7: Op-amp layout *(with four coins of gate capacitor visuable)*

tightly in area appears square-like. Note that different types of contacts are added because of the diffusion type difference (p-type diffusion resistor and n-type diffusion resistor).

4.2.3 BJT Array Layout

In this design, the larger BJT model is used for realizing the diodes. It is a laterally structured BJT with an emitter area of $25\mu m^2$ in the middle. The different current densities of the bandgap circuit is formed by applying two bipolar transistors whose emitter area ratio is $1:8$, Fig. 3.2. Thus, in the layout design, one coin of this BJT is placed in the middle of the array to form the smaller bipolar transistor and another eight coins are laid arround it to compose the other bigger bipolar transistor. An array of 3×3 is develped for the bipolar transistors and a dummy ring composed by the same bipolar transistor model is added around it, forming a 5×5 array, Fig 4.8.

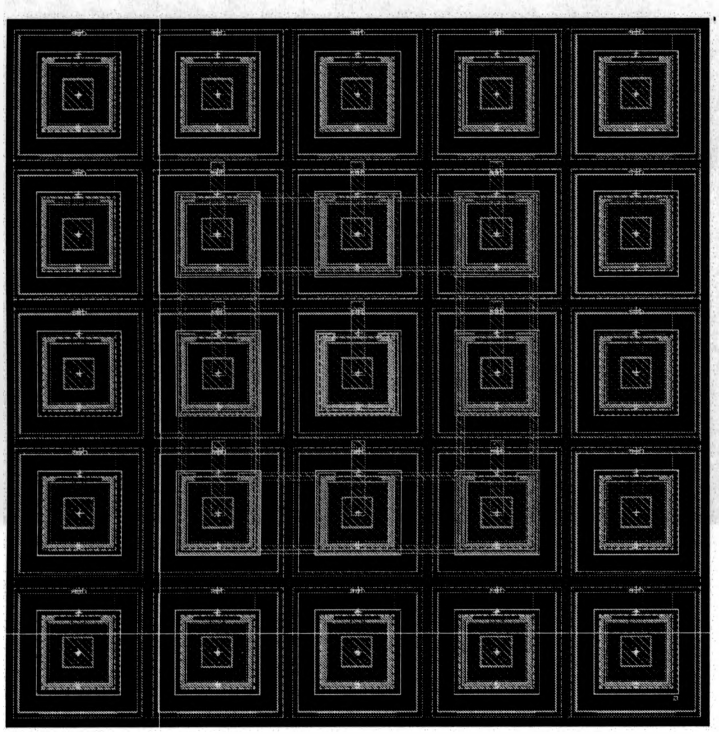


Figure 4.8: BJT array layout

4.2.4 Bandgap Core PMOS Current Sources Array Layout

In the bandgap core circuit, the PMOS current source transistors on top of the BJTs are also very important for the whole circuit's accuracy. Thus, to reduce the sensitivity of the process variation, the length of the transistors is chosen as large as $1\mu m$, five times as the other CMOS transistors in the desgin. The matching between the two transistors is also critical, so the development of these two transistors calls upon all aspects of matching techniques, similar to the current mirror in the op-amp. Again, drain and source sharing is considered and dummy transistors are added around the array, Fig. 4.9. The width of both transistors is $20\mu m$ so that in the array, there are twenty fingers of transistors whose size is $1\mu m \times 2\mu m$, for the sake of matching.

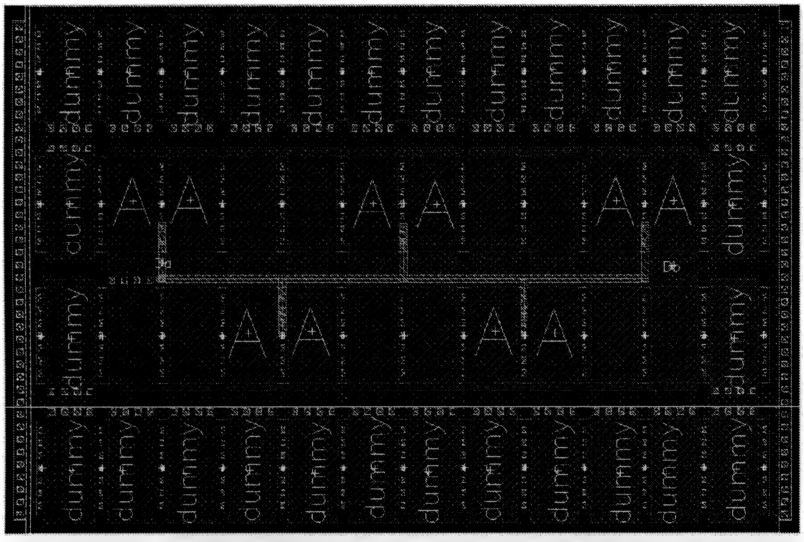


Figure 4.9: PMOS current sources array layout

4.2.5 Output Stage and Start-Up Circuit Layouts

In the output stage, three gate-capacitors coins are needed for frequency compensation and one coin of this kind is utilized for charging and discharging purposes in the start-up circuit. No extra attention need to be paid to these two circuits' layout except for there is one current mirror demanding good matching mechanism in the start-up circuit. Figure 4.10.

4.3 The Top Level Layout

At the top level, it becomes a higher level management of blocks. The placement of the blocks, the empty space utilization, and the power lines are all the issues to be concerned.

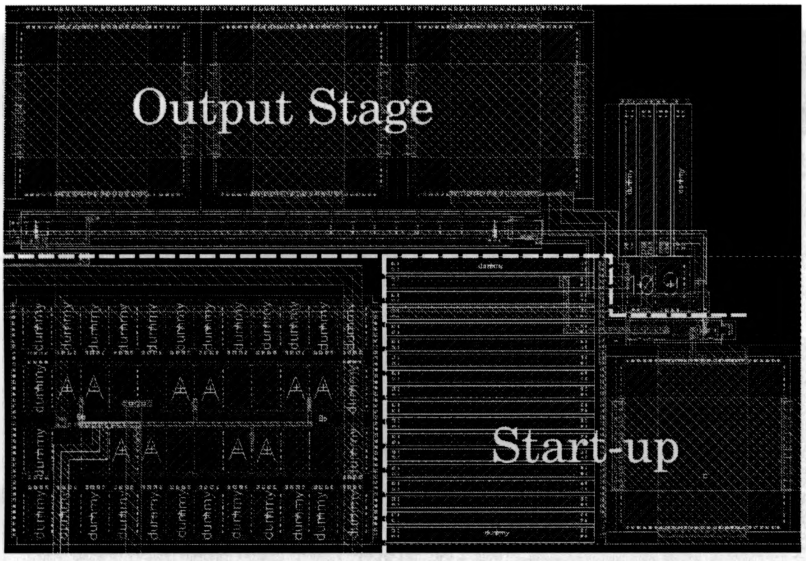


Figure 4.10: Output stage and start-up circuit layout

4.3.1 Placement of The Blocks

The finished bandgap reference layout is shown in Fig. 4.11, $100.92\mu m \times 84.74\mu m$, square alike. Just as when designing the blocks, it is also very important to make sure that the blocks are close enough to each other to save space. In the meantime, the blocks that are related to each other closely, in other words, connected to each other by direct wires, need to be put close to each other for the convenience of wire routing and for reducing the parasitic resistance and parasitic capacitance that may be caused by long wires.

The whole layout should appear as square-like as possible in order to be inserted into other blocks as implementation takes place later on. Figure 4.11 shows the square-like finished layout.

When trying to achieve the square-like appearance for the whole circuit, there is always some space left blank within the square as it is impossible to unify all of the

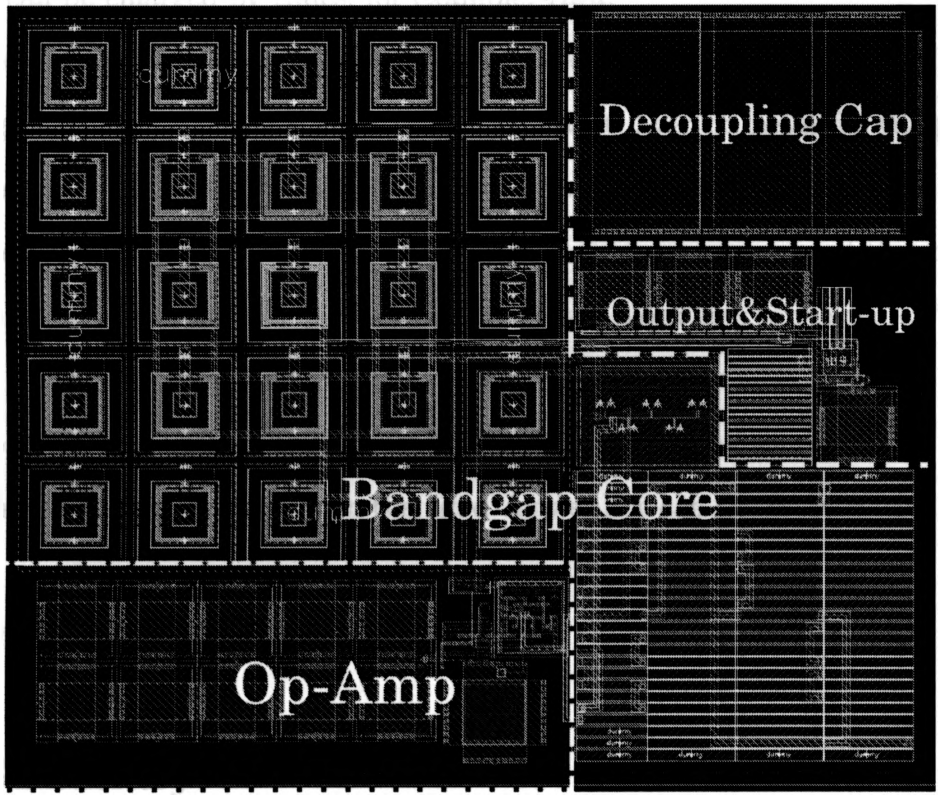


Figure 4.11: The finished layout

analog blocks in shape. For the empty space, a decoupling capacitor including every metal layer should be added to fill it, Fig. 4.11. This manner ensures the highest function efficiency for the area that is taken up by the whole circuit.

4.3.2 The Power Lines Layout

In determining the power line width, the current of the whole circuit should be considered. It is known from practice that a $1\mu m$ metal wire can pass $0.5mA$ of current. Thus, wider metal is able to carry higher current. In this design, approximately a total current of $100\mu A$ is consumed, which means the width of the power line can be

$0.2\mu m$, very close to the minimum width. To ensure that everything works decently, the width can be enlarged by times, for example, $1\mu m$.

Considering that using a higher layer of metal for power requires vias and other connecting elements [35], it is better to use the lowest layer of metal for the power line. As shown in Fig. 4.11, *metal1* is used for the wide power line.

In the circuit, "vdd" and "gnd" are both laid to occupy one corner of the square. This diagonal structure ensures that both sides of every block gets connected to "vdd" and "gnd". As the dummy around the BJT array is connected to "gnd", and it takes up the most space of the whole circuit, the other blocks can be connected to "gnd" through the BJT array dummy when the "gnd" power line can not be seen from them directly.

Note there is no power routing over any cell, only signal wires.

4.4 Summary

In this chapter, the layout of the proposed bandgap voltage reference is developed by applying the layout techniques , such as using unit elements, implementing common-centroid structure, and adding dummy elements. The finished layout is DRC, LVS clean, and square-like, consuming an area of only $100.92\mu m \times 84.74\mu m$.

Chapter 5

Circuit Characterization

In the last chapter, the layout of the bandgap circuit was given with the discussion of some layout issues. In this chapter, each block is characterized and the results after replacing the ideal elements with practical elements are given. At the end, Monte-Carlo statistic analysis is provided.

5.1 Characterization of Op-amp

By applying the simulation and measurement techniques discussed in [37], the designed op-amp is characterized in aspects of open-loop gain and frequency response, common mode input range and output voltage range, common mode gain, power supply rejection ratio, open-loop output resistance and slew rate in transient response.

5.1.1 The Open-Loop Gain and Frequency Response

In open-loop configuration, the open-loop gain can be determined by both AC analysis and transient analysis. Figure 5.1 shows the frequency compensated op-amp open-loop gain and frequency response. For frequencies lower than $60KHz$, the open-loop gain is $58.74dB$, at which the op-amp is supposed to be functional because the bandgap voltage reference circuit is mainly working under DC condition. The unity-gain point happens at a frequency of $77.6MHz$, which provides approximately a $30°$ phase margin. The phase margin is not extremely good compared to other op-amps,

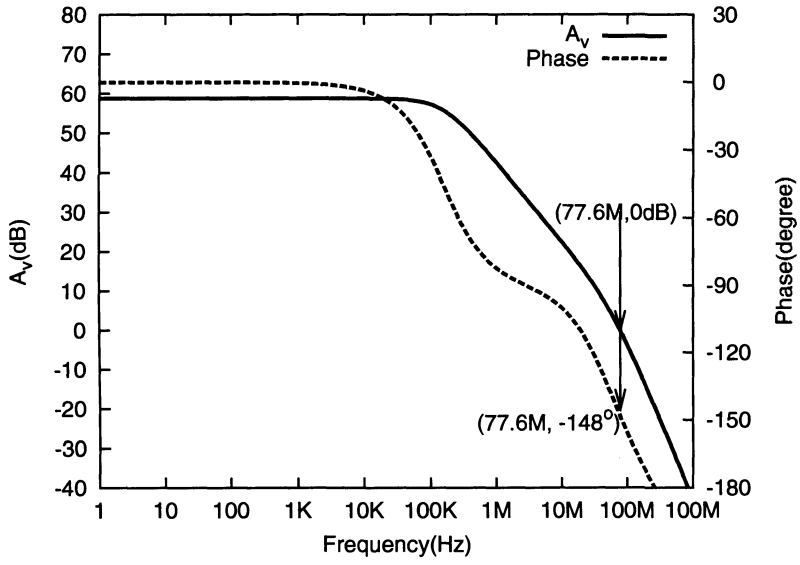


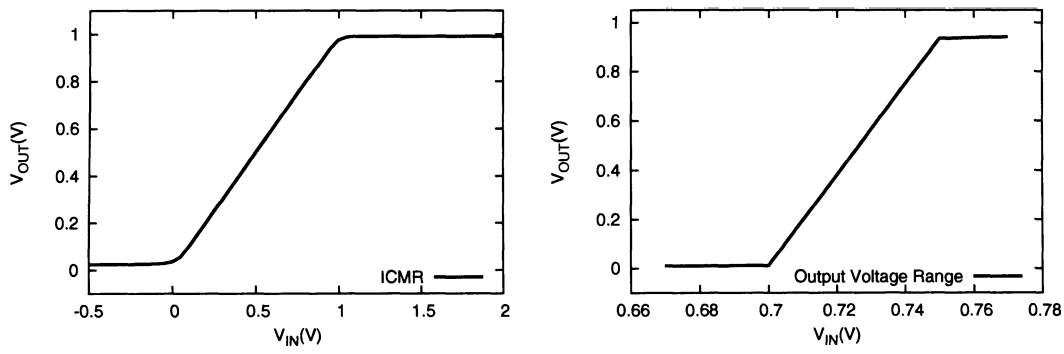
Figure 5.1: Frequency response bode plot of the op-amp

however, for a bandgap voltage reference circuit working under DC voltage that has very low demands for high frequency, it is more than enough.

5.1.2 ICMR and Output Voltage Range

The input common mode range in this op-amp is not that demanding, for it is fixed to around $0.7V$ in this circuit. In the unity-gain configuration, by sweeping V_{IN} at the available node, the input common mode range can be measured. Figure 5.2(a) shows that the input common mode range is from 0 to $1V$. This common mode range is more than enough for the op-amp's implementation as part of the bandgap voltage reference circuit.

As the voltage gain is very large, to measure the output voltage range by the open-loop configuration directly may not be accurate enough. From Fig. 5.2(a), it is shown roughly that the output voltage range would be $0 \sim 1V$, however, this range is measured by hypothetically sweeping the input common mode voltage from $-0.5V$ to $2V$, which may not be true in practice. The moderate gain configuration



(a) Input common mode range

(b) Output voltage range

Figure 5.2: Simulation results of input common mode range and output swing

for the op-amp is used for measuring the output voltage range whose gain is 20(with one $2K\Omega$ resistor connected to one of the input nodes of the op-amp and another $40K\Omega$ resistor connected between this input node and the output node, while the other one input node is connected to a DC voltage of $694.9mV$). By sweeping the voltage at the input node, which is connected to the resistor, the output voltage can be measured, Fig. 5.2(b). The output voltage range is $23.5mV \sim 925.6mV$ according to an input voltage of $0.7V \sim 0.75V$, which is slightly shifted to the right because of the input-offset voltage.

5.1.3 Input-Offset Voltage

One input (V_{in-}) is set as $694.9mV$(the nominal input DC voltage of the op-amp in the bandgap circuit, otherwise, the op-amp would not work because of the bias demands), while the other input node voltage (V_{in+}) is sweeped. Ideally, the output voltage should be zero when the two DC input voltages are the same. However, it's not the case in the practical world. When the differential input voltage equals to zero, the voltage at the output node is $V_{OS}(out)$. Knowing that $V_{out} = V_{in}A_v$, then

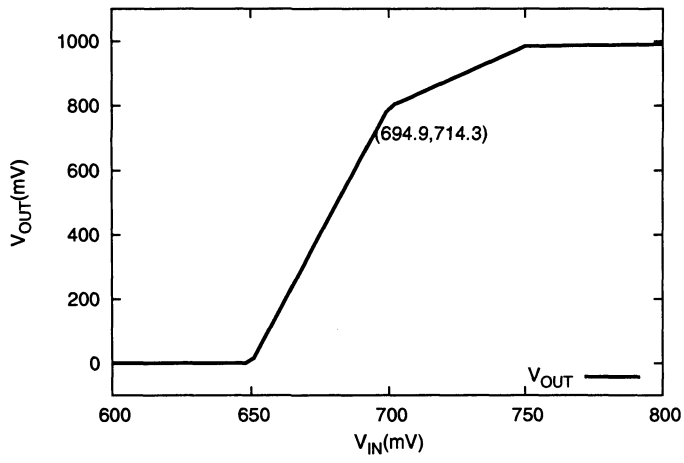


Figure 5.3: Simulation result to determine the input-offset voltage

$V_{OS}(in) = V_{OS}(out)/A_v$. Thus the input-offset voltage can be determined. Figure 5.3 shows when $V_{in+} = V_{in-} = 694.9mV$, $V_{out} = 714.3mV$. The input-offset voltage $V_{OS}(in) = V_{OS}(out)/865.1 = 0.825mV$.

5.1.4 Common Mode Gain and CMRR

With the input-offset voltage cancellation by adding an extra DC voltage source $V_{OS}(in)$ at one input node to force the output voltage to be zero when differential input voltage is zero, the common mode gain can be measured. When fed in a $2mV$ peak-to-peak sinusoidal signal, the output has a peak-to-peak voltage of $132.8\mu V$, Fig. 5.4. Thus the common mode gain A_{CM} is 0.0664. Also, Common Mode Rejection Ratio is $20log(A_{diff}/A_{CM}) \simeq 82.3dB$.

5.1.5 Power Supply Rejection Ratio

A sinusoidal voltage source whose peak-to-peak voltage is $2mV$ is inserted in series with power supply V_{dd} in a unity-gain configuration. The output voltage is at DC

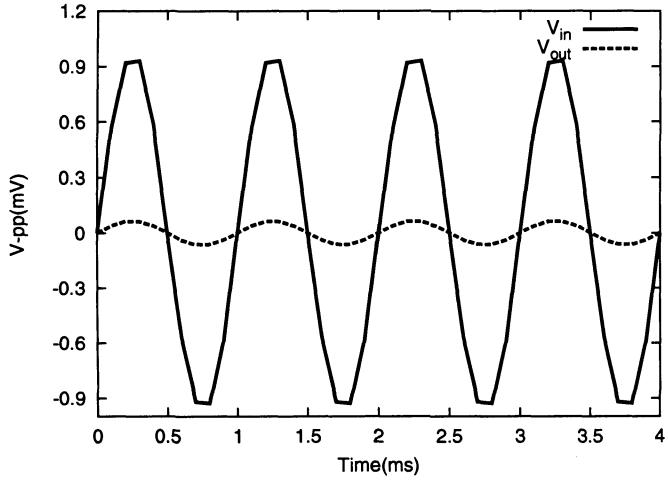


Figure 5.4: Common mode gain measurement through transient analysis

$699.359mV$ with a sinusoidal small signal whose peak-to-peak voltage is $17\mu V$ added on it. We can determine $PSRR^+$ as

$$PSRR^+ = \frac{V_{dd-pp}}{V_{out-pp}} = 117.6 = 41.408d \tag{5.1}$$

As V_{ss} of the circuit is connected to ground, there is no need to investigate $PSRR^-$.

5.1.6 Slew Rate and Settling Time

The large signal step response in the unity-gain configuration is measured, Fig. 5.5 shows the result.

The rising slew rate($+SR$), also the slope of the ramp part in the output response, is $886.3K$, but the falling ramp is not quite obvious, having a large $-SR$ of $-52.95M$. With the same configuration, but small signal, the settling time can be determined. In Fig. 5.6, the settling time is measured as $200ns$.

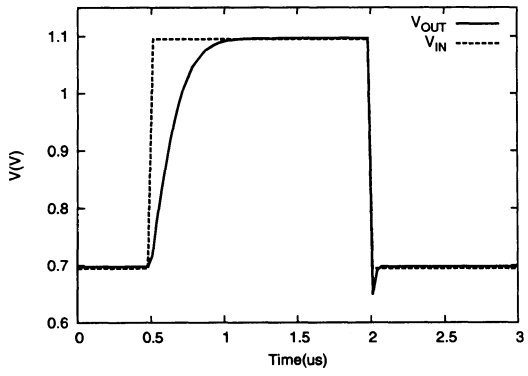


Figure 5.5: Simulation result of slew rate measurement

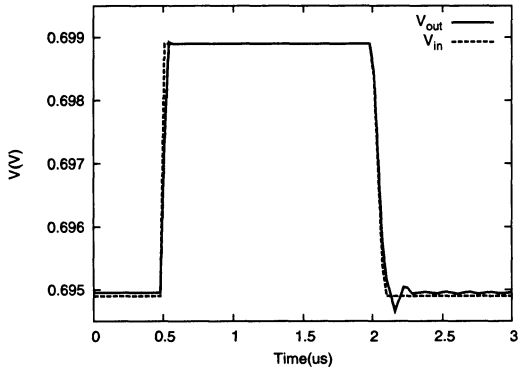


Figure 5.6: Simulation result of the settling time measurement

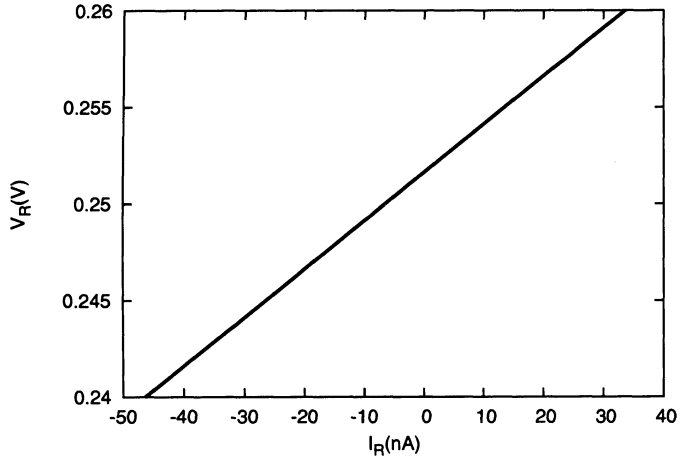
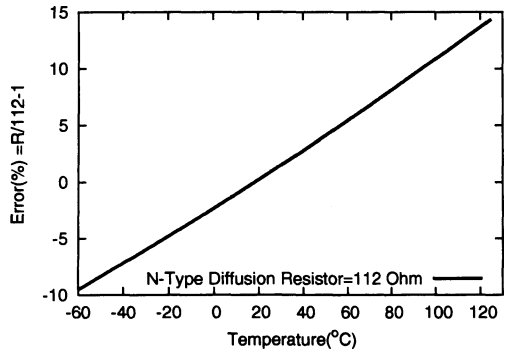


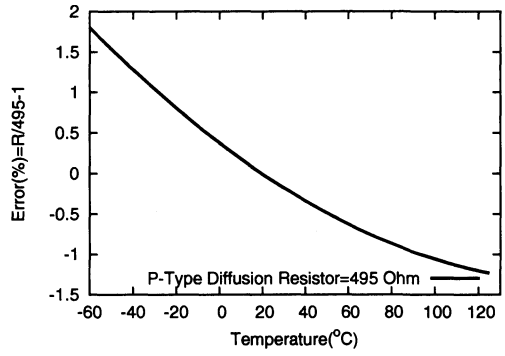
Figure 5.7: The slope of voltage vs. current determines the output impedance

5.1.7 Output Impedance

By sweeping a voltage source connected between the output node and the ground in the open-loop configuration within a small range of the supposed output, the output resistance can be determined by measuring the slope of voltage v.s. current, which is $250K\Omega$, Fig. 5.7.



(a) Variation of the square resistance of the n-type diffusion resistor whose norminal value is 112Ω vs. temperature



(b) Variation of the square resistance of the p-type diffusion resistor whose norminal value is 495Ω vs. temperature

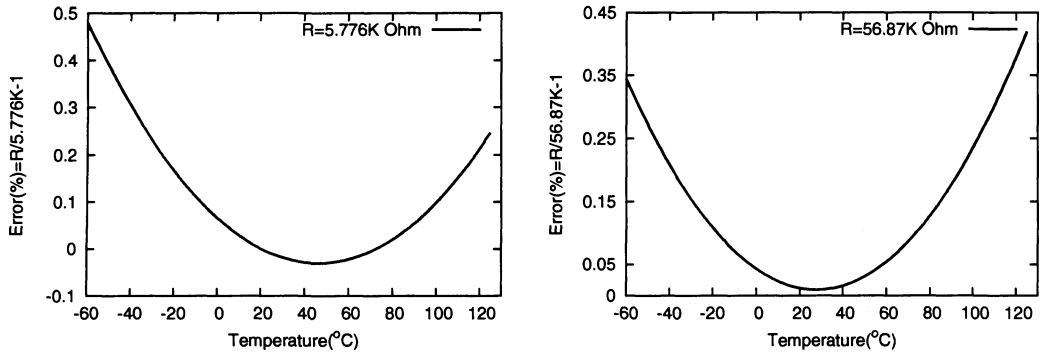
Figure 5.8: Characterization results of the poly resistors in the technology

5.2 Characterization of The Bandgap Voltage Reference

5.2.1 Temperature Independent Resistor in Bandgap Core

Each type of resistor in CMOS 90nm technology responds differently to changes in temperature. As the "n-type diffusion+p-type diffusion" resistors in series have larger value per square and less dependence on temperature among the available resistor models in this technology, they were chosen to compose the resistors in the bandgap core circuit. The characterization result of these two sheet resistors ($2\mu m \times 2\mu m$) is shown in Fig. 5.8. It is clear that both of their values vary with temperature, in an opposite ways however. (p-type diffusion resistor has a sheet resistance temperature coefficient of $0.14471\Omega/\square \cdot °C$, while p-type diffusion resistor sheet resistance has a temperature coefficient of $-0.0833\Omega/\square \cdot °C$.)

By combining two resistors with an appropriate ratio in series, the resistance achieved has little dependence on temperature. Figure 5.9 shows the characteristics



(a) The characteristics of 6K resistor against temperature

(b) The characteristics of 57K resistor against temperature

Figure 5.9: Characterization results of the temperature independent resistors

of the target resistors' realization character against temperature. For the realized resistor, the absolute value of sheet resistance TC is $0.00396\Omega/\square \cdot °C$. Note this coefficient is for square resistance, which means it's derived by dividing the error by the number of squares included in the resistors.

The comparison in Fig. 5.10 is done between the output voltages of the bandgap voltage reference before and after replacing the ideal resistors. Only a slight shift is observed in the average output voltage, but the reduction in the output voltage error is clearly visible with the real resistors. This error may be caused by the slight temperature dependence of the real resistors, which has a concave characteristics as Fig. 5.9 shows. Considering that the bandgap voltage has a convex characteristic, the resistors have a slight complementary effect on the output voltage.

5.2.2 The Replacement of Ideal Capacitors

The capacitors used in the early stage design were chosen from the ideal library in the IC Design tool. In order to realize those capacitors, gate-capacitors composed of transistors available for the current technology are applied. When the source, drain

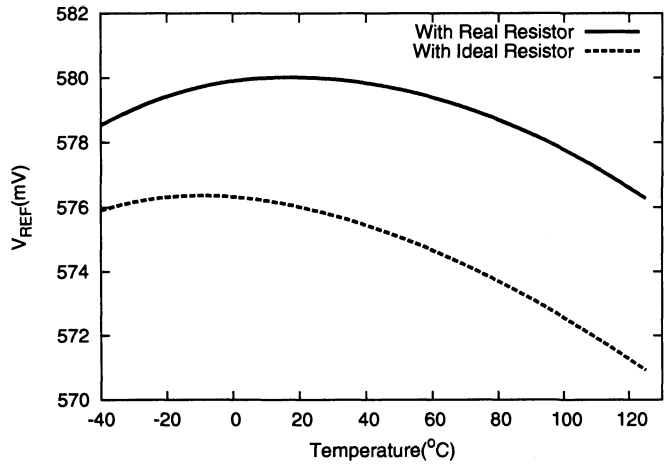


Figure 5.10: Dependence of V_{REF} on the replacement of the ideal resistors

and body of a transistor are connected to ground while the gate is connected to a non-0-voltage node in the circuit, the transistor acts as a capacitor. This kind of capacitor has its merit over other kinds of capacitors because the metal capacitor in the technology(90nm) has very limited square capacitance, and "MIM" capacitors are unavailable in this technology. In other words, to save some space on the circuit, the large capacitors($> 0.5pF$) are realized by transistor gate-capacitor.

After replacing the ideal capacitors by transistor composed capacitors, the bandgap curve appears to rise from around $20℃$ to higher temperature, Fig. 5.11.

The plot of a family of output curves with different power supplies is shown in Fig. 5.12. From the plot, one can see that the bandgap circuit works very well at the nominal voltage, $1.2V$, which produces only $244\mu V$ of error among with a temperature variation of $0℃ \sim 125℃$, which is better than [1]. (In this paper, even with curvature compensation, the voltage error is $300\mu V$ with a temperature variation of $0℃ \sim 80℃$, while the proposed bandgap circuit is operating without curvature compensation).

At lower voltages, the bandgap circuit can still survive, even if the voltage is

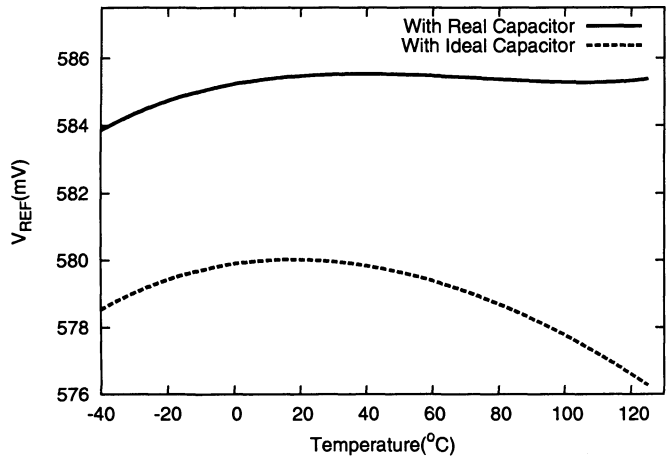


Figure 5.11: Dependence of V_{REF} on the replacement of the ideal capacitor

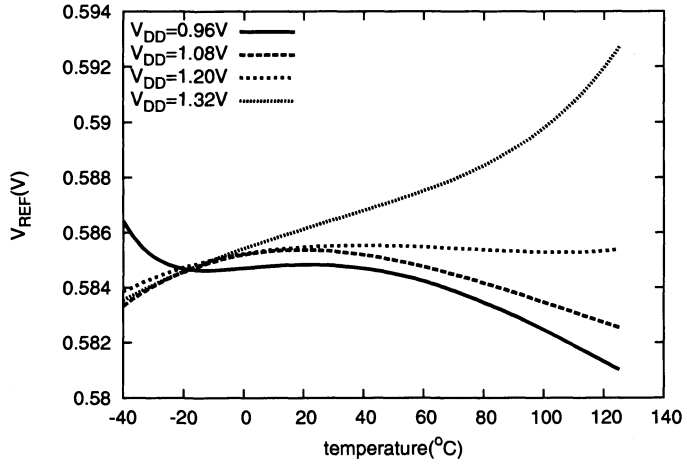


Figure 5.12: The simulation result of output voltage with different V_{dd}

as low as $0.96V$ and at $1V$, it still gives a good curve. The slight rises at the low temperature for lower power supply and at the higher temperature for higher power supply are caused by the real capacitors mentioned above, which are composed by the gate-capacitors. This change is due to the temperature dependence of the gate-capacitors and is actually good for adjusting the output voltage under the nominal voltage supply. It introduces a negative effect for a non-nominal power supply, which

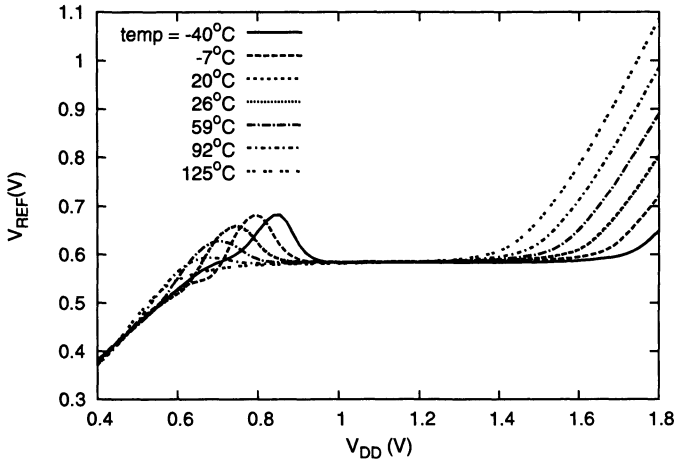


Figure 5.13: Simulation result of the output voltage against v_{dd} variation

is obvious from the plot and the effect is more pronounced for a higher power supply, for example $1.32V$. The error goes from $1.67mV$ to $9mV$ across the whole temperature range with different power supplies. Considering the moderate temperature range, ($0°\ C \sim 80℃$), the error is at most $3mV$, which is acceptable for the purpose of this design.

With the transient analysis, it is also proved that the bandgap circuit schematic works well. With a power supply variation of $\pm 10\%$, the variation of the output voltage is only $1.1mV$.

To better investigate the power supply dependence, the family plot of the output voltage against the supply voltage sweep is shown in Fig. 5.13.

The ripples at $600mV$ to $1V$ are caused by the current difference between the two bipolar transistors in the bandgap core circuit. When **Vdd** is low(around $600mV$ but high enough to turn the bipolar transistors on), for the bipolar transistors with lower current density(larger emitter area, $m = 8$), it is easier to carry more current under the same condition of low V_{BE}. However, when **Vdd** is around the nominal voltage value, the current difference is negligible. For a high temperature like $125°\ C$,

the ripple disappears as the high temperature makes the bipolar transistor capable to work under higher current with the same low V_{BE}, which eliminates the current difference between the two bipolar transistors. [38]

$$I_{BE}(T_2) = I_{BE}(T_1)(\frac{T_2}{T_1})^{X_{T1}/n_F} exp[-\frac{qE_g(300)}{n_F k T_2}(1 - \frac{T_2}{T_1})] \tag{5.2}$$

5.2.3 The Noise Analysis

The thermal noise of a resistor can be presented by

$$S_v(f) = 4kTR, f \geq 0 \tag{5.3}$$

where $k = 1.38 \times 10^{-23} J/K$ is the Boltzmann constant, [24], Chapter 7.2. It is proportional to the absolute temperature while it is independent of frequency from 0 to as high as $100THz$.

By the same method of measuring the output impedance of the op-amp, the output impedance of the bandgap voltage reference at the output node is measured as $8.819K\Omega$. With this resistance, the thermal noise can be determined by the equation above. For example, at room temperature, the circuit exhibits $1.4604 \times 10^{-16} V^2/Hz$ of thermal noise.

When fed in a $2mV$ peak-to-peak sinusoidal signal in series with power supply V_{dd}, there is a peak-to-peak sinusoidal small signal at $184.8\mu V$, which is a fraction in the output voltage.

We can determine $PSRR^+$ as

$$PSRR^+ = \frac{V_{dd-pp}}{V_{REF-pp}} = 10.82 = 20.685dB \tag{5.4}$$

This PSRR is satisfactory as it only influences the output voltage with a μm scale.

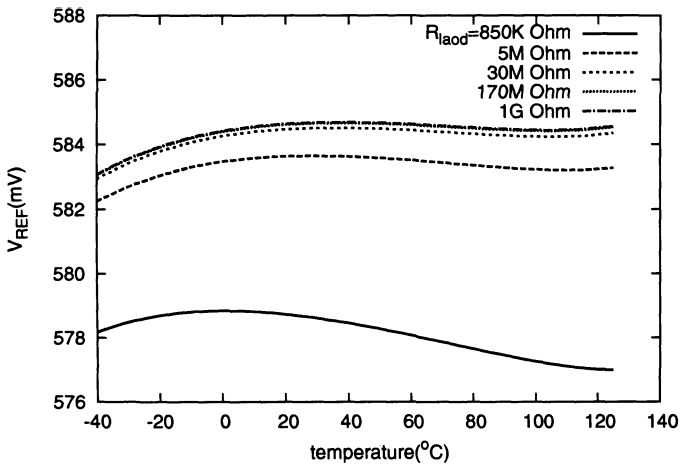


Figure 5.14: Dependence of V_{REF} vs. load resistance

5.2.4 The Resistive Load Limitation

The bandgap voltage reference always acts as a function to drive very high input impedance blocks. Therefore, the resistive load at the output node should be high enough to ensure minimum influence to the reference output voltage. Other impedance such as capacitive and inductive load is not of concern here as the bandgap voltage reference works on DC and frequency related components are not involved.

In order to specify the lower limit of the resistive load, the bandgap output node is intentionally connected with a load resistor whose resistance can be swept as a parameter. The nominal output voltage with an infinite resistive load is $584.7mV$ at room temperature. Assume that the highest tolerable error in the system is within 1% of the nominal value at the same temperature. With this assumption, the result shows that the resistive load lower limit is around $850K\Omega$, Fig. 5.14. The output voltage at room temperature is $578.74mV$ while it has a variation of $1.849mV$ across the temperature range of $-40℃ \sim 125℃$, which means the error according to temperature variation is 0.3%.

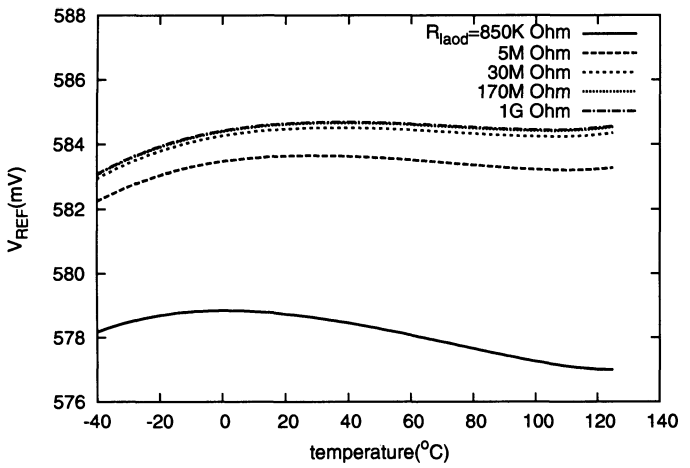


Figure 5.14: Dependence of V_{REF} vs. load resistance

5.2.4 The Resistive Load Limitation

The bandgap voltage reference always acts as a function to drive very high input impedance blocks. Therefore, the resistive load at the output node should be high enough to ensure minimum influence to the reference output voltage. Other impedance such as capacitive and inductive load is not of concern here as the bandgap voltage reference works on DC and frequency related components are not involved.

In order to specify the lower limit of the resistive load, the bandgap output node is intentionally connected with a load resistor whose resistance can be swept as a parameter. The nominal output voltage with an infinite resistive load is $584.7mV$ at room temperature. Assume that the highest tolerable error in the system is within 1% of the nominal value at the same temperature. With this assumption, the result shows that the resistive load lower limit is around $850K\Omega$, Fig. 5.14. The output voltage at room temperature is $578.74mV$ while it has a variation of $1.849mV$ across the temperature range of $-40℃ \sim 125℃$, which means the error according to temperature variation is 0.3%.

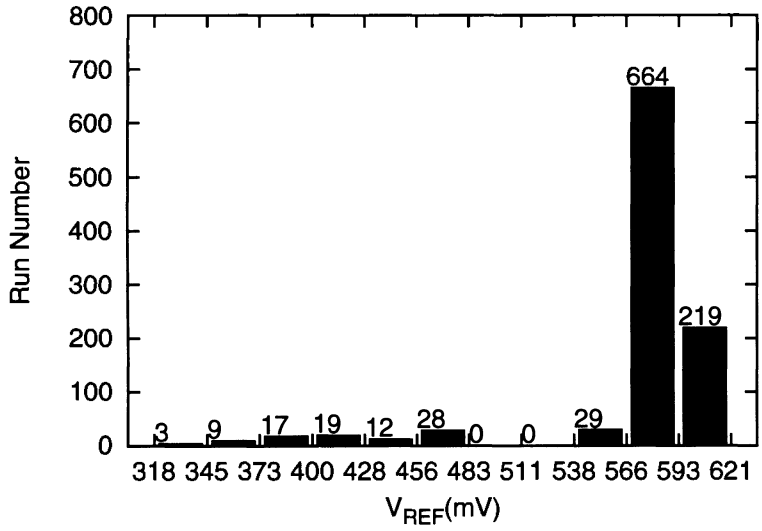


Figure 5.15: Output voltage histogram on 1000 Monte Carlo simulation runs

5.3 Monte Carlo Analysis of Process Variation and Component Mismatch

In order to characterize the behaviour of the circuit when process and environment variation are involved, Monte Carlo analysis can be applied. Over 1000 runs, there are only 50 runs that totally failed(the output voltage of the run has an error of 20% $\sim$ 40%). 125 runs display a more than 1% higher output voltage($590mV \sim 617mV$, whose error is 1% $\sim$ 5%) than $584.7mV$ and 82 runs display a more than 1% lower output voltage($555mV \sim 578mV$, whose error is $-5\% \sim -1\%$) than $584.7mV$. If all of these runs(including the runs with $|error| > 1\%$) are counted as failed, then the successful rate for process variation and mismatch is around 85% and the yield is over 90%. Figure 5.15.

Table 5.1: Comparison between this work and the referred paper work [1]

Parameters	Malcovati [1]	This Work
Power Supply Voltage	1V	1.2V
Technology	$0.8\mu m$ BiCMOS	$90nm$ CMOS
Bandgap Cell Area	$0.25mm^2$	$0.008484mm^2$
Power Consumption @ $T = 25℃$	$92\mu W$	$160\mu W$
Reference Voltage @ $T = 25℃$	$536mV$	$584.7mV$
Operation Temperature Range	$0℃ \sim 80℃$	$-40℃ \sim 125℃$
Temperature Dependence (Without Curvature Compensation)	$800uV$,$20ppm/℃$ ($0℃ \sim 80℃$)	$244.37\mu V$, $3.343ppm/℃$ ($0℃ \sim 125℃$) or $1.6119mV$, $16.7ppm/℃$ ($-40℃ \sim 125℃$)
Temperature Dependence (With Curvature Compensation)	$300uV$,$7.5ppm/℃$ ($0℃ \sim 80℃$)	NA
Dependence on the Supply Voltage	$114\mu V/V$	$1.1mV$($1.08V \sim 1.32V$)

5.4 Summary

The results from the characterization process are given in this chapter. One can see that with the different op-amp configuration and the improved output stage and start-up circuit compared to [1], the bandgap voltage reference provides an ultimate better accuracy without an extra compensation circuit and is able to operate within a much wider temperature range. Table.5.1

The relatively high dependence on the supply voltage of this work may be caused by the high demand of the pure CMOS op-amp, while the op-amp proposed in [1], which includes bipolar transistors may be more suitable for low voltage operation. However, the performance in other aspects of this work demonstrates advantages such as high accuracy, wide operation temperature range and much smaller area consumption.

Chapter 6

Conclusion

The aim of this thesis is to design a low-voltage power supply and high temperature-independence bandgap voltage reference. In correspondence to this goal, the bandgap circuit is developed block by block in CMOS 90nm technology. This chapter gives a summary of the contribution made by this thesis, plus suggestions on future work.

6.1 The Contribution of The Thesis

- In order to build a bandgap voltage reference to work under $1.2V$, the implementation makes use of the concept introduced in [1], which is to utilize a fraction of the current from the conventional bandgap voltage. The final bandgap circuit is able to work under a power supply range of $0.96V \sim 1.4V$, where the error of the output voltage is only $1.1mV$ with a power supply variation of $\pm 10\%$ $(1.08V \sim 1.32V)$.

- With the careful characterization of the resistor models in the available technology, the temperature independent resistors are build for the bandgap core circuit. Other works regarding how to increase the temperature independence of the output are done, such as BJT model characterization and the novel output design. Thus the output of the bandgap circuit has a very high accuracy even with temperature variation, which gives only $244\mu V$ of error over the temperature variation range of $0℃ \sim 125℃$.

- In order to avoid the bandgap circuit being locked in the zero current operating point, the start-up circuit is designed with high efficiency and little influence on the bandgap circuit.

- In order to obtain a precise output voltage, a two stage op-amp working under $1.2V$ power supply is developed. The characterization shows that the op-amp has an open-loop gain of $58.74dB$ and the unity-gain point is at $77.6MHz$. The phase margin of the op-amp is $30°C$, however, this phase margin is good enough for the bandgap circuit mainly works under DC voltage. Both the input common-mode range and the output voltage range cover almost the entire rail-to-rail voltage range, which shows a very good performance. Also, according to the results shown in Chapter 5, the CMRR and PSRR of the op-amp are satisfactory, and it achieved an output impedance of $250K\Omega$.

- The layout of the whole bandgap circuit is designed by referring to the rules of matching issues to eliminate the process and other mismatch effects. The final layout is a square-like chip, which takes up an area of $100.92\mu m \times 84.74\mu m$, with DRC and LVS clean. Over the 1000 runs in the Monte Carlo analysis, around 85% of the runs are satisfactory with an error of less than 1%.

6.2 Future Work

Based on the achievements of this thesis, there are some future work to be done to improve the performance of this work.

- Nowadays the low-voltage power supply demands continue to require lower and lower power supply, thus, bandgap circuit working under lower power supply, such as $1V$ or even lower is the direction of the current circuit development.

- This thesis presented a bandgap voltage reference (without any curvature compensation circuit) whose output voltage has a $3.343 ppm/^{\circ}C$ error across a wide temperature range, which is much better than the referenced work, [1] (with curvature compensation circuit). However, it is believed that with an additional curvature compensation circuit, less error can be achieved and even better performance of the circuit can be obtained.

- The results of this thesis work are based on schematic simulation, although all the elements used in the circuit are real elements from the technology, there is still a possibility that the experimental results would be different. To obtain the absolute practical results, post-layout simulation, or even chip verification should be done.

Bibliography

[1] P. Malcovati, F. Maloberti, C. Fiocchi, and M. Pruzzi, "Curvature-compensated bicmos bandgap with 1-v supply voltage," *Solid-State Circuits, IEEE Journal of*, vol. 36, no. 7, pp. 1076–1081, Jul 2001.

[2] R. Widlar, "New developments in ic voltage regulators," *Solid-State Circuits, IEEE Journal of*, vol. 6, no. 1, pp. 2–7, Feb 1971.

[3] K. Kuijk, "A precision reference voltage source," *Solid-State Circuits, IEEE Journal of*, vol. 8, no. 3, pp. 222–226, Jun 1973.

[4] A. Brokaw, "A simple three-terminal ic bandgap reference," *Solid-State Circuits, IEEE Journal of*, vol. 9, no. 6, pp. 388–393, Dec 1974.

[5] C. Palmer and R. Dobkin, "A curvature corrected micropower voltage reference," in *Solid-State Circuits Conference. Digest of Technical Papers. 1981 IEEE International*, vol. XXIV, Feb 1981, pp. 58–59.

[6] G. Meijer, P. Schmale, and K. Van Zalinge, "A new curvature-corrected bandgap reference," *Solid-State Circuits, IEEE Journal of*, vol. 17, no. 6, pp. 1139–1143, Dec 1982.

[7] B. Song and P. Gray, "A precision curvature-compensated cmos bandgap reference," *Solid-State Circuits, IEEE Journal of*, vol. 18, no. 6, pp. 634–643, Dec 1983.

[8] F. Assaderaghi, D. Sinitsky, S. Parke, J. Bokor, P. Ko, and C. Hu, "A dynamic threshold voltage mosfet (dtmos) for ultra-low voltage operation," in *Electron Devices Meeting, 1994. IEDM '94. Technical Digest., International*, Dec 1994, pp. 809–812.

[9] A.-J. Annema, "Low-power bandgap references featuring dtmosts," *Solid-State Circuits, IEEE Journal of*, vol. 34, no. 7, pp. 949–955, Jul 1999.

[10] S. Verdonckt-Vandebroek, S. Wong, J. Woo, and P. Ko, "High-gain lateral bipolar action in a mosfet structure," *Electron Devices, IEEE Transactions on*, vol. 38, no. 11, pp. 2487–2496, Nov 1991.

[11] S. Verdonckt-Vandebroek, J. You, J. Woo, and S. Wong, "High-gain lateral p-n-p bipolar action in a p-mosfet structure," *Electron Device Letters, IEEE*, vol. 13, no. 6, pp. 312–313, Jun 1992.

[12] Y. Jiang and E. Lee, "Design of low-voltage bandgap reference using transimpedance amplifier," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 47, no. 6, pp. 552–555, Jun 2000.

[13] M. Gunawan, G. Meijer, J. Fonderie, and J. Huijsing, "A curvature-corrected low-voltage bandgap reference," *Solid-State Circuits, IEEE Journal of*, vol. 28, no. 6, pp. 667–670, Jun 1993.

[14] K. N. Leung and P. Mok, "A sub-1-v 15-ppm/c cmos bandgap voltage reference without requiring low threshold voltage device," *Solid-State Circuits, IEEE Journal of*, vol. 37, no. 4, pp. 526–530, Apr 2002.

[15] S. Mehrmanesh, M. Vahidfar, H. Aslanzadeh, and M. Atarodi, "A 1-volt, high psrr, cmos bandgap voltage reference," in *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, vol. 1, May 2003, pp. I–381–I–384 vol.1.

[16] Y.-T. Lin, W.-Y. Chung, D.-S. Wu, H.-C. Lin, and R. Lin, "A low voltage cmos bandgap reference," in *IEEE-NEWCAS Conference, 2005. The 3rd International*, June 2005, pp. 227–230.

[17] A. Bendali and Y. Audet, "A 1-v cmos current reference with temperature and process compensation," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 54, no. 7, pp. 1424–1429, July 2007.

[18] K. Sanborn, D. Ma, and V. Ivanov, "A sub-1-v low-noise bandgap voltage reference," *Solid-State Circuits, IEEE Journal of*, vol. 42, no. 11, pp. 2466–2481, Nov. 2007.

[19] S. Wadhwa, "A low voltage cmos bandgap reference circuit," in *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, May 2008, pp. 2693–2696.

[20] E. K. Lee, "A low voltage cmos bandgap reference without using an opamp," in *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, 24 2009-Yearly 27 2009, pp. 2533–2536.

[21] A. Becker-Gomez, T. Lakshmi Viswanathan, and T. Viswanathan, "A low-supply-voltage cmos sub-bandgap reference," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 55, no. 7, pp. 609–613, July 2008.

[22] H. Banba, H. Shiga, A. Umezawa, T. Miyaba, T. Tanzawa, S. Atsumi, and K. Sakui, "A cmos bandgap reference circuit with sub-1-v operation," *Solid-State Circuits, IEEE Journal of*, vol. 34, no. 5, pp. 670–674, May 1999.

[23] A. Cabrini, G. De Sandre, L. Gobbi, P. Malcovati, M. Pasotti, M. Poles, F. Rigoni, and G. Torelli, "A 1 v, 26 w extended temperature range band-gap

reference in 130-nm cmos technology," in *Solid-State Circuits Conference, 2005. ESSCIRC 2005. Proceedings of the 31st European*, Sept. 2005, pp. 503–506.

[24] B. Razavi, *Design of Analog CMOS Integrated Circuits.* Mc Graw Hill Education, 2001.

[25] X. Dai, "Explicit characterization of bandgap references," Master of Science Thesis, Iowa State University, 2006.

[26] D. A.Johns and K. Martin, *Analog Integrated Circuit Design.* John Wiley and Sons,Inc, 1997.

[27] L. Ge, "Cmos bandgap reference circuit," Master of Applied Science Thesis, Dalhousie University, 2004.

[28] J. H.Huijsing, R. J. de Plassche, and W. M.C.Sansen, *Analog Circuit Design, Low-Noise, Low-Power, Low-Voltage; Mixed-Mode Design with CAD Tools; Voltage, Current and Time References.* Kluwer Academic Publishers, 1996.

[29] W. M. Sansen, *Analog Design Essentials.* Springer, 2006.

[30] Absjoshi, *http://www.edaboard.com/ftopic331729.html.* EDAboard.com Forum Index, 2007.

[31] A.Cabrini, G. Sandre, L.Gobbi, P.Malcovati, M.Pasotti, M.Poles, F.Rigoni, and G.Torelli, "A 1v, 26uw extended temperature range band-gap reference in 130-nm cmos technology," 2005.

[32] R. Gregorian, *Introduction to CMOS Op-Amps and Comparators.* John Wiley and Sons, Inc., 1999.

[33] R. Baker, *CMOS Circuit Design, Layout, and Simulation.* British Library Cataloguing-in-Publication Data, 2005.

[34] A. Hastings, *The Art of Analog Layout.* Pearson Education, Inc., publishing as Prentice Hall, 2006.

[35] D. Clein, *CMOS IC LAYOUT Concepts, Methodologies, and Tools.* British Library Cataloguing-in-Publication Data, 1999.

[36] K. Lampaert, G. Gielen, and W. Sansen, *Analog Layout Generation For Performance And Manufacturability.* Kluwer Academic Publishers, 1999.

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

[38] G. Massabrio and P. Antognetti, *Semiconductor Device Modeling with Spice.* McGraw-Hill Professional, 1998.

[39] B. Razavi, *Fundamentals of Microelectronics.* Jonh Wiley and Sons, Inc., 2008.

Appendix A

Frequency Response and Compensation of Op-Amp

Although the bandgap circuit works mostly under DC voltage, it is demanded that the op-amp in this circuit operates well in the frequency domain and possess a reasonable phase margin. This section refreshes the general knowledge of freqency response in a closed-loop system, and the frequency compensation techniques for op-amps.

A.0.1 Frequency Response and The Problem of Instability

A closed loop system is shown in Fig. A.1(a), [39]. The transfer function is:

$$\frac{Y}{X}(s) = \frac{H(s)}{1 + KH(s)} \tag{A.1}$$

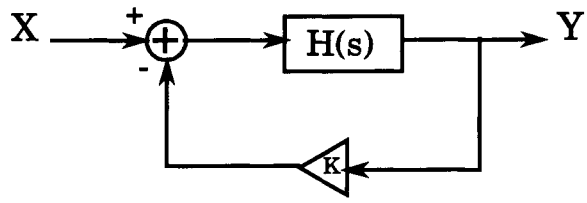
If at ω_1

$$KH(s) = KH(j\omega_1) = -1 \tag{A.2}$$

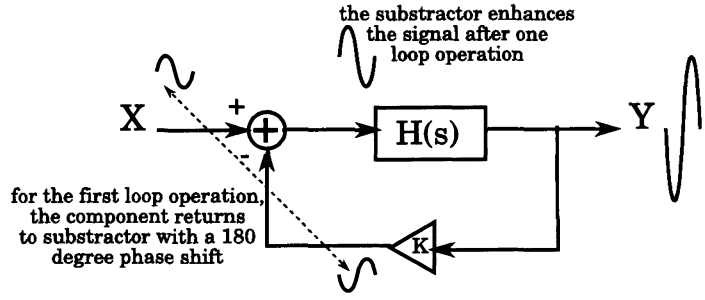
the system will enter oscillation.

The two conditions that cause oscillation (also called "Barkhausen's criteria") are: **first**, $\angle KH(j\omega_1) = -180°$, which guarantees the positive feedback and **second**, $|KH| = 1$ ensures sufficient loop gian for the circulating signal to grow.

Thus, $Y/X \approx \infty$ and the feedback becomes positve at the frequency ω_1, thereby producing a feedback signal that enhances the input. In other words, if $|KH| < 1$,



(a) The closed loop system



(b) At frequency ω_1

Figure A.1: Closed loop system diagram

then the output cannot grow indefinitely while for $|KH| \geq 1$ and $\angle H(j\omega_1) = -180°$, the negative feedback system oscillates.

To determine the stability of the closed loop system, the freqency response should be analysed. The bode plot is a commonly used tool for this purpose, Fig. A.2.

There are three definitions here.

Gain Crossover Frequency (ω_{GX}) : the frequency when the loop gain falls to unity (0dB).

Phase Crossover Frequency (ω_{PX}) : the frequency when the phase shift reaches $-180°$).

Phase Margin : $\angle H(\omega_G X) + 180°$.

The stability requires that $\omega_{GX} < \omega_{PX}$, and the greater the phase margin is, the more stable a system is. Typically, a phase margin of 60° is enough to provide stability.

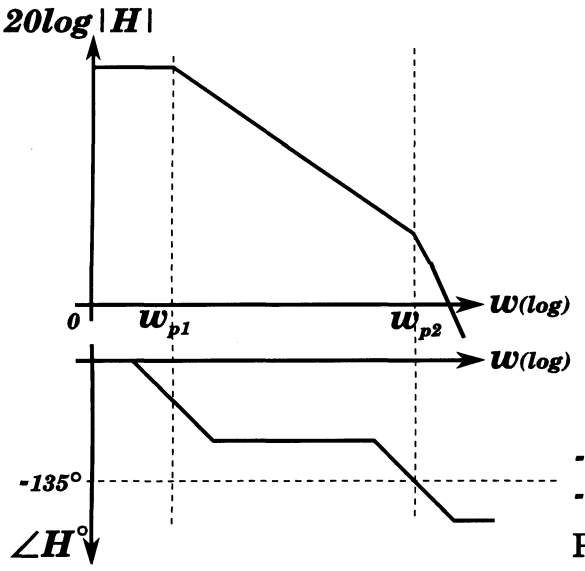


Figure A.2: Bode plot of a closed loop system

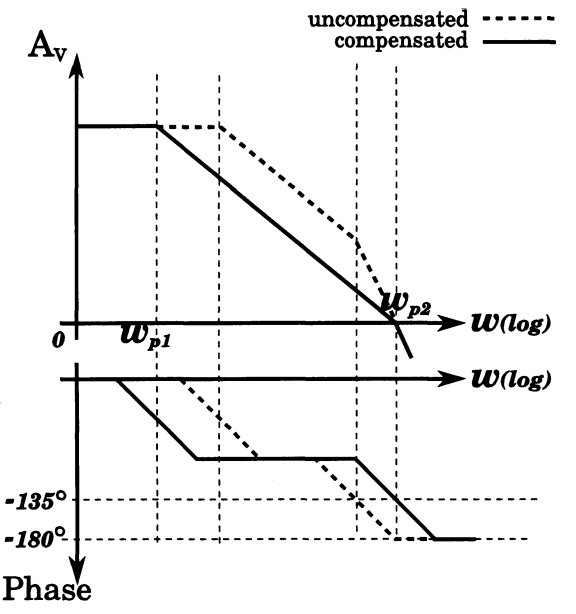


Figure A.3: Bode plot of Miller compensation

A.0.2 Miller Compensation of the Two-Stage Op-Amp

To prevent oscillation in feedback amplifiers, and to ensure a good transient performance, an additional step, frequency compensation, is needed, [32] Chapter 4.3.

The task in compensating an amplifier for closed-loop applications is to move all the poles and zeros, except for the dominant pole (p_1), sufficiently away from the origin of the complex frequency plane (beyond the unity-gain bandwidth frequency) to result in a phase shift [37].

Miller compensation is commonly used in two-stage op-amps. It is applied by connecting a capacitor between the output node and the input node of the second transconductance stage, Fig. A.4.

By adding C_{MC}, two results are accomplished. Firstly, the capacitance at node A is increased, thus, by adding a capacitor of $g_{m\text{II}}R_{\text{II}}C_{MC}$, the first pole is moved closer to the origin of complex freqency plane. Secondly, the second pole is moved away from the origin of the complex frequency plane, see the transfer functions below:

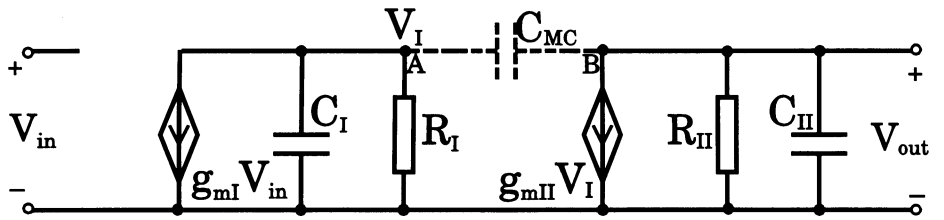


Figure A.4: Small signal analysis of Miller compensation for two-stage op-amp

$$\frac{v_{out(s)}}{v_{in(s)}} = \frac{a}{1 + sb + s^2c} \tag{A.3}$$

where, $a = (g_{mI})(g_{m\text{II}})(R_I)(R_{\text{II}})(1 - sC_{MC}/g_{m\text{II}})$,

$b = R_I(C_I + C_{MC}) + R_{\text{II}}(C_{\text{II}} + C_{MC}) + g_{m\text{II}}R_IR_{\text{II}}C_{MC}$, and

$c = R_IR_{\text{II}}[C_IC_{\text{II}} + C_{MC}C_I + C_CC_{\text{II}}]$.

The first pole is changed to $p_1 \approx \frac{-1}{g_{m\text{II}}R_IR_{\text{II}}C_{MC}}$;

The second pole is changed to $p_2 = \frac{-g_{m\text{II}}C_{MC}}{C_IC_{\text{II}}+C_{\text{II}}C_C+C_IC_C} \approx \frac{-g_{m\text{II}}}{C_{\text{II}}}$. (Fig. A.3).

The magnitude of the second pole p_2 must at least be equal to GB (Gain-Bandwidth) and is resulted from the capacitance at the output of the op-amp. Pole p_2 is also called the ouput pole, it strongly depends on the load capacitance.