**PRECISION VOLTAGE REFERENCING
TECHNIQUES IN MOS TECHNOLOGY**

by

Bang-Sup Song

Memorandum No. UCB/ERL M83/17

14 March 1983

PRECISION VOLTAGE REFERENCING
TECHNIQUES IN MOS TECHNOLOGY

by

Bang-Sup Song

Memorandum No. UCB/ERL M83/17

14 March 1983

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Precision Voltage Referencing Techniques in MOS Technology

By

Bang-Sup Song

B.S. (Seoul National University, Korea) 1973
M.S. (Korea Advanced Institute of Science) 1975

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Engineering

in the

GRADUATE DIVISION

OF THE

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved: ....*Paul R. Gray*.......... *March 14/1983*
.................Chair................Date
....*David A. Hodges*.........................
....*Lucien Le Cam*..............................

....

PRECISION VOLTAGE REFERENCING TECHNIQUES
IN MOS TECHNOLOGY

Ph. D.   Bang-Sup Song   Department of
                                                            Electrical Engineering
                                                            and Computer Science

                              Signature : _____
                                          Chairman of Committee •

ABSTRACT

With the increasing complexity of functions on a single MOS chip, precision analog circuits implemented in the same technology are in great demand so as to be integrated together with digital circuits. The future development of MOS data acquisition systems will require precision on-chip MOS voltage references. This dissertation will probe two most promising configurations of on-chip voltage references both in NMOS and CMOS technologies.

In NMOS, an ion-implantation effect on the temperature behavior of MOS devices is investigated to identify the fundamental limiting factors of a threshold voltage difference as an NMOS voltage source. For this kind of voltage reference, the temperature stability on the order of $20 ppm/^{\circ}C$ is achievable with a shallow single-threshold implant and a low-current, high-body bias operation.

In CMOS, a monolithic prototype bandgap reference is designed, fabricated and tested which embodies a curvature compensation and exhibits a minimized sensitivity to the process parameter variation. Experimental results imply that an average temperature stability on the order of $10 ppm/^{\circ}C$ with a production spread of less than $10 ppm/^{\circ}C$ is feasible over the commercial temperature range.

ACKNOWLEDGEMENT

DEDICATION

I dedicate this dissertation to my wife Kyung-Sook and two love kids Juwan and Selin. Their love, endurance and sacrifice they have been filling me with every single day in the monotonicity of the Albany village life enabled me to concentrate on my work.

I also dedicate this dissertation to my late father who would be more delighted than anyone else if he is still in this world and my old mother in Korea who has been waiting so long for her youngest son to finish his Ph.D. work.

TABLE OF CONTENTS

LIST OF SYMBOLS

Subscripts U, E, D and B on the following symbols stand for unimplanted, enhancement, depletion and double-implant, respectively.

A	Emitter area ratio.
C_{OX}	Gate oxide capacitance per unit area, ϵ_{OX}/t_{OX}.
E_g	Si energy gap.
E_{g1}	Linearly-extrapolated Si energy gap at $0^o K$.
I_o	Temperature-independent bias current.
I_s	Reverse saturation current.
$I_{B,E,C}$	Base, emitter or collector current.
I_D	Bias current of V_{BE}.
I_T	PTAT(Proportional To Absolute Temperature) bias current.
kT/q	Thermal voltage, $\approx 26mV$ at $300^o K$.
KV_T	PTAT correction voltage.
FV_T^2	PTATS(PTAT Squared) correction voltage.
n	Exponent in the mobility temperature variation of T^{-n}.
n_i	Intrinsic concentration, $\approx 1.4 \times 10^{10} cm^{-3}$ at $300^o K$.
N_a	Bulk dopant concentration.
N_{ai}	Enhancement acceptor implant density.
N_{di}	Depletion donor implant density.
N_{peak}	Peak carrier density in the channel at threshold.
q	$1.6 \times 10^{-19} Coul.$
Q_d	Depleted charge density per unit area at threshold.
Q_i	Implanted charge density per unit area.
Q_{ss}	Fixed charge at the Si-Si dioxide interface.
Q_B	Total built-in base charge per unit area.
r_b	Base resistance of one unit transistor cell.
t_{OX}	Gate oxide thickness.
T	Temperature in $^o K$.
V_{BE}	Emitter-base or base-emitter potential.
V_{g1}	Linearly-extrapolated Si bandgap voltage at $0^o K$.
V_{g2}	Quadratically-extrapolated Si bandgap voltage at $0^o K$.
V_m	Gate-source voltage mismatch of a pair of IGFET's.
V_o	Bandgap voltage for the internal bias purpose.
V_{os}	Offset voltage of an op amp.

V_{ref}	Output reference voltage.
V_{th}	Gate threshold voltage.
V_{GS}	Gate to source voltage.
V_{SB}	Source to bulk body bias voltage.
V_T	Thermal voltage kT/q.
X_1	Depth of the maximum potential point.
X_d	Depth of the depletion edge at threshold.
X_i	Depth of the enhancement implant.
X_j	Depth of the depletion implant.
α	Exponent in the bias current temperature variation of T^α.
β	Current gain of bipolar transistors.
γ	$4-n$ where $n = 3/2\text{~}5/2$.
ϵ_s	Permittivity of Si, $\approx 10^{-12} F/cm$.
ϵ_{OX}	Permittivity of the gate oxide, $\approx 3.5 \times 10^{-13} F/cm$.
$\overline{\mu}$	Average channel mobility.
Φ_{bi}	Channel to bulk built-in potential at threshold.
Φ_E	Channel to bulk built-in potential in enhancement devices.
Φ_D	Channel to bulk built-in potential in depletion devices.
Φ_X	Potential drop from the maximum potential point to the surface.

Unless otherwise specified, the following data are employed for numerical calculations :

Q_{iE}	$2\times10^{11} cm^{-2}$,
Q_{iD}	$6\times10^{11} cm^{-2}$,
N_a	$5\times10^{14} cm^{-3}$,
N_{ai}	$5\times10^{15} cm^{-3}$,
N_{di}	$10^{17} cm^{-3}$,
X_i	$400 nm$,
X_j	$60 nm$,
t_{OX}	$100 nm$ and
V_{SB}	$5 V$.

CHAPTER 1

GENERAL INTRODUCTION

To meet the growing integration trend of analog and digital LSI functions, the development of precision analog components has been accelerated in MOS technology [1]. An essential element of the analog and digital interface function is a voltage reference to control the scale factor of conversion. The absolute accuracy of data conversion systems is limited by the accuracy of the voltage references used in the systems. The temperature stability of a reference source is a key factor in the accuracy of the overall data acquisition function. In reality, the performance of voltage references has not been able to catch up with those of other data acquisition functions. Therefore, the ability to integrate an entire data acquisition function within a single MOS VLSI chip is contingent upon the ability to realize an MOS-compatible voltage reference with a very low temperature drift.

This thesis explores the best-performing configurations of on-chip voltage references both in NMOS and CMOS technologies. In the first part (Chapter 2), an analysis of the effect of ion-implantation on the IGFET threshold temperature drift is presented, and simple design equations for predicting the threshold voltage temperature coefficients of implanted devices relative to those of unimplanted devices are developed. In the second part (Chapters 3, 4 and 5), a precision curvature-compensated switched-capacitor bandgap reference is described which employs a standard digital CMOS process and achieves a temperature stability significantly lower than has previously been reported for CMOS circuits without critical trim requirements. The theoretically achievable temperature stability of the reference output approaches $10ppm/^{o}C$ over the commercial temperature range utilizing a straight-forward room temperature trim procedure and the estimated production spread is less than $10ppm/^{o}C$. Experimental data from monolithic prototype samples are presented which fit the theoretical predictions reasonably well.

1.1. Historical Research Survey

Since its introduction by Widlar [2], the bandgap referencing technique has been applied for implementing a voltage source in bipolar integrated circuits. The temperature stability of a bandgap reference has been continuously improved via new circuits and technology innovation such as a curvature compensation and a laser trimming [3]-[6].

In NMOS technology, however, due to the lack of a forward-biased diode, the bandgap referencing scheme is not directly applicable. To date, the only technique actually utilized commercially for implementing NMOS reference sources has been based on developing the difference between the threshold voltages of two devices with different implants in the channel [7]-[8]. The temperature stability of such a voltage source depends on the details of the effect of the channel implant on the temperature coefficient of the device threshold. The temperature variation of the uniformly-doped device has been studied by many authors [9]-[11]. The temperature dependent behavior of the depletion device was analyzed by Gaensslen and Jaeger, numerically [12] and analytically [13]. Their work emphasizes the low temperature carrier freezeout phenomena and is not suitable for a circuit application in the commercial and military temperature ranges.

In CMOS technology, the bandgap referencing technique has been directly applied [14]-[15] and many versions of on-chip voltage references have been introduced during the past few years [16]-[17] using the well-known bipolar bandgap referencing technique. The flat-band voltage difference [18] resulting from different gate materials and the work-function difference [19] resulting from different gate dopings have also been proposed as reference sources. Unlike the bipolar implementation of a bandgap reference, however, the development of a high performance CMOS bandgap reference has been hindered by several factors such as the peculiarities of the bipolar devices available in a standard CMOS process, the high offset and drift of CMOS op amps that make up the circuit as well as the inherent curvature problem in the bandgap reference.

Table I summarizes the features of the several voltage reference works published in the chronical order. Note that the high performance in the bipolar reference is achievable through the use of thin film resistors and a laser trim technology. Due to the lack of these features and the existing problems unique to CMOS, a high-precision CMOS voltage source has been unattainable.

TABLE I

HISTORICAL SURVEY

Tech.	Author	V_{ref} (V)	T.C. (ppm/°C)	Feature	Correct	Noise (μV)
Bipolar	Widlar (1971)	5	200 (-55~125°C)	Thin film	1st order	40 (100kHz)
"	Kujik (1973)	9.118	4 (0~60°C)	Thin film Laser tr.	"	70 "
"	Brokaw (1974)	2.5	5-60 (-55~125°C)	"	"	-
"	Palmer (1981)	5	2.2 "	"	2nd order	-
"	Meijer (1982)	1.158	0.45 (-25~85°C)	Ni-Cr	"	-
NMOS	Blauschild (1978)	3.167	4.9 (-55~125°C)	-	-	200 (10kHz)
CMOS	Tsividis (1978)	1.105	100 "	Weak inv.	1st order	-
"	" (1979)	1.2	70 (0~60°C)	p well Weak inv.	"	-
"	Vittoz (1979)	1.3	110 (-50~100°C)	"	"	-
"	Gregorian (1981)	1.1~1.3	40 (0~70°C)	p well	"	-

CHAPTER 2
THRESHOLD VOLTAGE DIFFERENCE AS
AN NMOS VOLTAGE SOURCE

2.1. Introduction

Ion-implantation has been widely employed to adjust the threshold voltage of MOS transistors made on low conductivity substrates. The only commercially available NMOS voltage source relies on the threshold voltage difference of two devices with different threshold implants. However, the temperature variation of device characteristics associated with the channel implantation has not been fully understood. This chapter begins with an analysis of the temperature variations of the four types of implanted devices typical of those available in NMOS technology and the theoretical results are compared with experimental data. In this analysis, only the relative temperature variations of the implanted devices with respect to the unimplanted devices are considered.

In Section 2.2, the general threshold condition of IGFET's is described and, assuming the channel is gradual, the one-dimensional Poisson's equation based on a depletion approximation is solved for general non-uniform doping profiles. Also, the built-in potential difference and its effect on the threshold voltage temperature variation of the implanted devices will be discussed. In Section 2.3, the threshold voltage shifts resulting from various channel implants are calculated based on the threshold definition described in Section 2.2. In Section 2.4, using the analysis of Section 2.3, simplified explicit expressions for the temperature coefficients of the implanted devices relative to those of the unimplanted devices will be developed using a step profile approximation. In Section 2.5, the validity of the depletion approximation is discussed and, in Section 2.6, experimental results are compared with the theoretical results developed in Section 2.4.

Source   ○ Gate   Drain

Poly Si

t_{ox} SiO$_2$

n$^+$ Si   n$^+$

p$^-$ Substrate

X
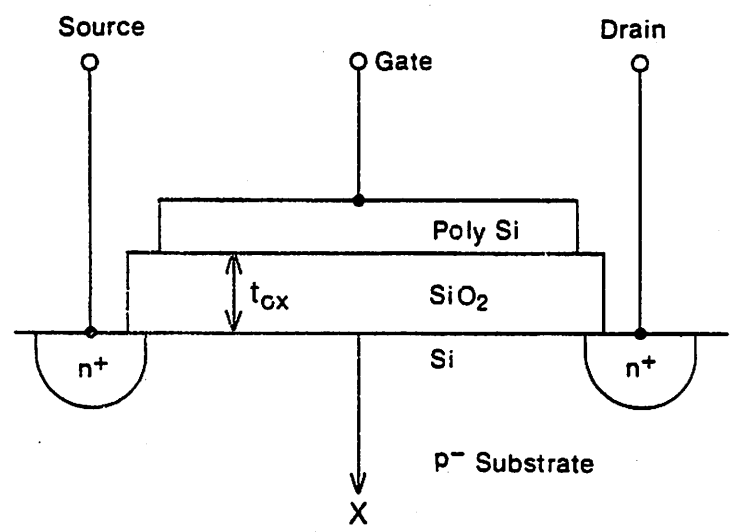
Fig. 2.1. N-channel Si-gate IGFET.

2.2. Threshold Voltage of IGFET's and its Temperature Dependence

The profile of an n-channel insulated gate field effect transistor (IGFET) is shown in Fig. 2.1. The heavily n^+-doped poly-Si gate is isolated from the substrate by the thermally-grown thin Si-oxide layer. The energy band diagram of the MOS system of Fig. 2.1 is illustrated in Fig. 2.2(a). As the gate voltage is increased above the flat-band voltage, the field at the surface of the Si increases and the energy band begins to bend. In the surface region, the majority carriers (holes) have been depleted and generation of carriers will therefore exceed recombination. The generated hole-electron pairs are separated by the field, the holes being swept into the bulk and the electrons moving to the oxide-Si interface where they are held because of the oxide energy barrier. The IGFET threshold voltage is defined as a potential drop from the gate to the substrate when a certain number of carriers exist in the channel. At this time, the potential drop from the point where potential is maximum to the substrate is defined as a built-in potential Φ_{bi} as shown in Fig. 2.2(b).

2.2.1. IGFET at Threshold

For a long-channel MOS transistor, the gate-substrate potential drop at threshold is the sum of the potential drop V_{OX} across the gate oxide and the potential $\Phi(x=0)$ at the surface. That is,

$$V_{th} = V_{OX} + \Phi(x=0) \quad . \tag{2.1}$$

The potential drop across the gate oxide is also given by

$$V_{OX} = V_{FB} - \frac{Q_d}{C_{OX}} = \Phi_{MS} - \frac{Q_{ss}}{C_{OX}} - \frac{Q_d}{C_{OX}} \tag{2.2}$$

where Q_d is the space-charge density per unit area, C_{OX} is the gate oxide capacitance per unit area, and V_{FB} is the flat-band voltage which is determined by the metal-semiconductor
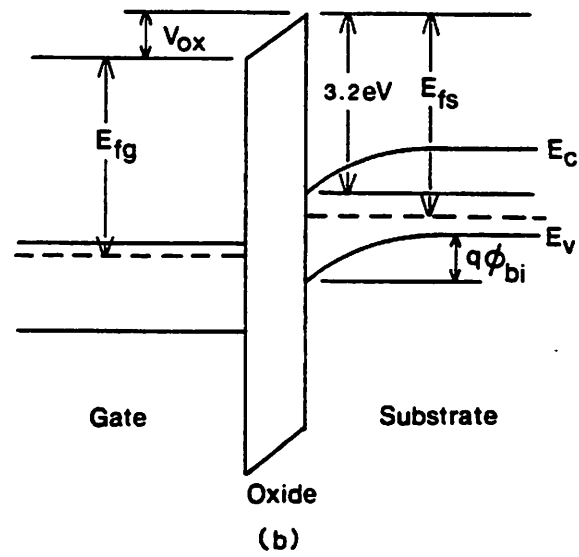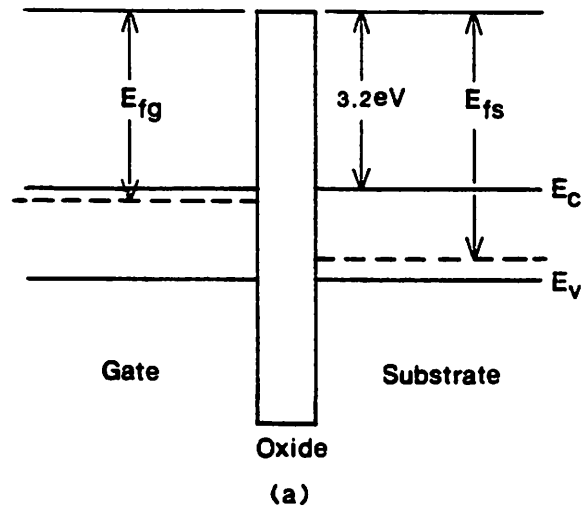
Fig. 2.2. Energy-band diagram : (a) at flat band and
(b) at threshold.

work function difference Φ_{MS} and the fixed interface charge per unit area Q_{ss}, etc.. From Fig. 2.2(a), the metal-semiconductor work function difference Φ_{MS} is simply

$$\Phi_{MS} = \Phi_M - \Phi_S = E_{fg} - E_{fs}$$

(2.3)

$$= -\frac{kT}{q} \ln \frac{N_{poly} N_a}{n_i^2}$$

where Φ_M and Φ_S are the work functions of the gate and the substrate, E_{fg} and E_{fs} are the potential difference from the oxide conduction-band edge to the Fermi potentials of the gate and the substrate, respectively, as shown in Fig. 2.2(a), and N_{poly} and N_a are the dopant concentrations of the gate and the substrate, respectively.

Assuming charge neutrality beyond the depletion edge and neglecting minority carriers, the one-dimensional Poisson's equation and the boundary conditions in the Si bulk are

$$\frac{d^2 \Phi(x)}{dx^2} = -\frac{q}{\epsilon_s} N(x) ,$$

(2.4)

$$\Phi(X_d) = 0 \quad \text{and} \quad \frac{d\Phi(x)}{dx}\bigg|_{x=X_d} = 0$$

where X_d is the depletion region edge and $N(x)$ is the total charge density as a function of depth from the Si surface. The solution of (2.4) is

$$\Phi(x) = -\frac{q}{\epsilon_s} [x \int_{X_d}^{x} N(t) dt - \int_{X_d}^{x} t N(t) dt] .$$

(2.5)

The depleted charge density per unit area at threshold is also given by

$$Q_d = -q \int_{0}^{X_d} N(x) dx$$

(2.6)

where the negative sign indicates the polarity of the space charge left over after the free carriers are depleted.

2.2.2. Channel-to-Bulk Built-In Potential

The channel-to-bulk built-in potential is defined as the potential difference between the peak potential of the conducting channel and the substrate potential. The potential with reference to the intrinsic Fermi level is a logarithmic function of the electron concentration so that

$$\Phi(x) = \frac{kT}{q} \ln \frac{n(x)}{n_i} \tag{2.7}$$

where $n(x)$ is the electron concentration at x. The definition of the built-in potential coincides with the definition of the peak-carrier density in the channel at threshold.

For the unimplanted (uniform channel doping) device without the body bias, the built-in potential between the maximum potential point in the channel and the bulk at threshold has traditionally been defined as the strong inversion value of

$$\Phi_{bi} = \frac{kT}{q} \ln \frac{N_a^2}{n_i^2} . \tag{2.8}$$

The same definition has been extended previously to the implanted devices [20]-[21]. This definition is equivalent to defining the threshold condition as that condition in which the concentration of electrons in the channel at the maximum potential point is the same as the concentration of holes in the bulk.

The built-in potential in the non-uniformly doped devices with channel implantations has been defined in two different ways. One definition is

$$\Phi_{bi} = \frac{kT}{q} \ln \frac{N_{doped} N_a}{n_i^2} \tag{2.9}$$

where N_{doped} is the actual dopant density at the potential maximum point [22]-[24]. The other definition is

$$\Phi_{bi} = \frac{kT}{q} \ln \frac{N(X-X_d)^2}{n_i^2} \tag{2.10}$$

where $N(X-X_d)$ is the dopant density at the edge of the depletion region [25]-[27]. Equation (2.10) is only valid when the implantation is deep and the body bias is low, and it coincides with (2.8) as far as the depletion region extend beyond the implantation depth. Equations (2.9) and (2.10) heavily depend on the depletion approximation and the strong inversion condition. Usually, it results in a high channel conductance at threshold. The generalization of the above definition leads to

$$\Phi_{bi} = \frac{kT}{q} \ln \frac{N_{peak} N_a}{n_i^2} \tag{2.11}$$

where N_{peak} is the peak electron density in the channel at threshold.

The above definitions are based on an arbitrary definition of the strong inversion. There is no way to define the threshold condition analytically other than the definitions mentioned. We can not say which one is better than the others absolutely. As a rule of thumb, however, the peak carrier density N_{peak} is defined as equal to the bulk doping concentration N_a. Whatever the definitions are, the calculated threshold voltages are close to each other due to the logarithmic relation as indicated by (2.7).

2.2.3. Threshold Voltage Temperature Dependence

The threshold voltage temperature dependence of the unimplanted IGFET has been investigated in many works [9]-[11], [28]-[30], especially at low temperature [31]-[33]. Assume the fixed interface charge is independent of temperature, the differentiation of (2.1) with respect to temperature yields

$$\frac{dV_{th}}{dT} = \frac{d\Phi_{MS}}{dT} - \frac{1}{C_{OX}} \frac{dQ_d}{dT} + \frac{d\Phi_{bi}}{dT} \tag{2.12}$$

for the unimplanted device.

The second term of (2.12) is adjustable by increasing the body bias and scaling down the gate oxide. However, the major contributions are from Φ_{MS} and Φ_{bi}. Therefore, the

direct dimension scaling of MOS devices does not scale the temperature drift of V_{th}. This may pose a problem in the low supply voltage operation of scaled MOS devices. From (2.11) and (AII.16), the built-in potential temperature dependence is given by

$$\frac{d\Phi_{bi}}{dT} = \frac{1}{T} [\Phi_{bi} - \frac{1}{q} (E_g - T \frac{dE_g}{dT}) - 3 \frac{kT}{q}]$$

(2.13)

$$\approx \frac{1}{T} (\Phi_{bi} - \frac{1}{q} E_{g1})$$

where the linearly-extrapolated Si energy gap E_{g1} at $0^o K$ is

$$E_{g1} = E_g - T \frac{dE_g}{dT}$$

$$\approx 1.205 \ eV \quad \text{for} \quad T = 300^o K \ .$$

For the substrate doping of $N_a = 5 \times 10^{14} cm^{-3}$, the temperature drift of the built-in potential is approximately

$$\frac{d\Phi_{bi}}{dT} \bigg|_{T_o} = \frac{1}{T_o} (\Phi_{bi} - 1.205)$$

(2.14)

$$\approx \frac{1}{300} (0.545 - 1.205)$$

$$\approx -2.2 \ mV/^o C \quad \text{for} \quad N_{peak} = N_a = 5 \times 10^{14} \ cm^{-3} \ .$$

The metal-semiconductor work function difference Φ_{MS} in (2.3) assumes the same form as Φ_{bi} in (2.11) and therefore its temperature drift is

$$\frac{d\Phi_{MS}}{dT} \bigg|_{T_o} = -\frac{1}{T_o} (0.862 - 1.205)$$

(2.15)

$$\approx 1.14 \ mV/^o C \quad \text{for} \quad N_{poly} \approx 10^{20} \ cm^{-3} \ .$$

The space-charge density Q_d is proportional to $(\Phi_{bi} + V_{SB})^{\frac{1}{2}}$ where V_{SB} is the source-to-bulk body bias. Therefore, we obtain

$$\frac{dQ_d}{dT} = -\frac{Q_d}{2C_{OX}(\Phi_{bi} + V_{SB})} \frac{d\Phi_{bi}}{dT}$$ (2.16)

$$\approx -\frac{[2\times1.6\times10^{-19}\times10^{-12}\times5\times10^{14}(0.545 + V_{SB})]^{1/2}}{2\times3.5\times10^{-8}(0.545 + V_{SB})} \times(-2.2 \quad mV/^{\circ}C)$$

$$\approx 0.52 \quad mV/^{\circ}C \quad \text{for} \quad V_{SB} = 0 \quad V$$

$$\approx 0.17 \quad mV/^{\circ}C \quad \text{for} \quad V_{SB} = 5 \quad V$$

where (AI.6) is used for the approximation. Also note that the temperature variation of Q_d decreases as V_{SB} increases.

Based on the above simple analysis from (2.12) to (2.16), the temperature coefficient of V_{th} is estimated by

$$\frac{dV_{th}}{dT} \approx 1.14 - 0.54 - 2.2 = -1.6 \quad mV/^{\circ}C \quad \text{for} \quad V_{SB} = 0 \quad V$$ (2.17)

$$\approx 1.14 - 0.17 - 2.2 = -1.23 \quad mV/^{\circ}C \quad \text{for} \quad V_{SB} = 5 \quad V \quad .$$

The above estimations are close to the experimentally observed temperature coefficients of the unimplanted devices. Specifically, the body bias effect on the threshold temperature variation fits well with the experiment.

2.3. Threshold Voltage Shifts by Ion-Implantations

In a typical enhancement/depletion NMOS technology, up to four types of devices are available depending on the dopant profiles as shown in Fig. 2.3. Using a depletion approximation, the dopant density, the space charge, the field distribution and the potential distribution at threshold for the enhancement device and the depletion device are illustrated in Fig. 2.4 and 2.5, respectively. At threshold, the surface of the enhancement-mode device is depleted of mobile free carriers, and minority carriers start to build up due to the polarity inversion at the surface where the potential peaks. In the depletion-mode device, a

Dopant Density

X

N_a

Unimplanted Device

(a)

Dopant Density

X

N_a

$N_{ai}(X)$

Enhancement Device

(b)

Dopant Density

$N_{di}(X)$

X

N_a

Depletion Device

(c)

Dopant Density

$N_{di}(X)$

X

N_a

$N_{ai}(X)$

Doubly-Implanted Device

(d)
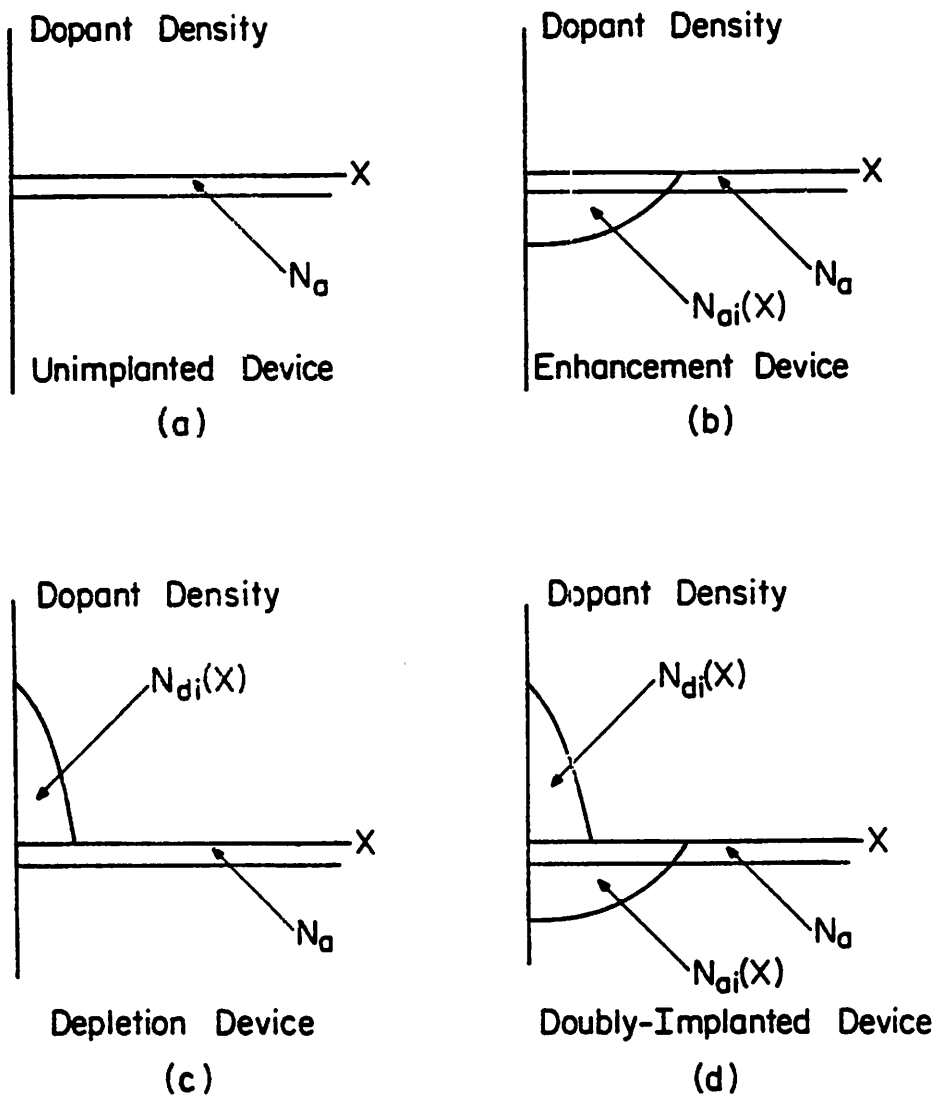
Fig. 2.3. Illustration of the four basic types of devices in NMOS technology. Dopant profiles of (a) unimplanted, (b) enhancement, (c) depletion and (d) doubly-implanted devices.

conducting channel is already formed by the depletion implant. Therefore, at flat band condition, there still exists a conducting channel under the surface. These free carriers must be cleared by pulling down the gate voltage. At threshold, the implanted conducting channel is depleted and the pinch-off condition occurs. The potential peaks at the subsurface as shown in Fig. 2.5(d) and free electrons accumulate at this peak potential point.

The threshold voltage shifts resulting from various channel implants are described for the enhancement device in numerical forms [25]-[26], [34] and in analytic forms [21], [27] and [35], for the depletion device in numerical forms [20], [36]-[37] and in analytic forms [22]-[24], respectively. In the following subsections, the equations describing the threshold voltage shifts resulting from the enhancement-type and the depletion-type channel implants are derived based on the depletion approximation discussed in Section 2.2.

2.3.1. Enhancement-Type Implantation

The threshold voltage of the unimplanted device is shifted due to the enhancement implant by

$$\Delta V_{th} = \Delta \Phi_{bi} - \Delta Q_d / C_{OX}$$ (2.18)

where $\Delta \Phi_{bi}$ is the built-in potential difference of the enhancement device and the unimplanted device at threshold and is given from (2.11) by

$$\Delta \Phi_{bi} = \Phi_E - \Phi_U = \frac{kT}{q} \ln \frac{N_{peak-E}}{N_{peak-U}}$$ (2.19)

and ΔQ_d, the depleted charge difference between them at threshold, is

$$\Delta Q_d = Q_{dE} - Q_{dU} = -q \int_0^{X_{dE}} \left[N_{ai}(x) + N_a \right] dx + q \int_0^{X_{dU}} N_a dx .$$ (2.20)

The depletion region edges of the enhancement device and the unimplanted device, X_{dE} and X_{dU}, respectively, can be obtained implicitly from the following relations ;

Dopant Density

Space Charge

N_a

$N_{ai}(X)$

(a)

X_{dE}

N_a

$N_a + N_{ai}(X)$

(b)

Field

ϕ_E

X_{dE}

(c)

Potential

ϕ_E

X_{dE}

(d)
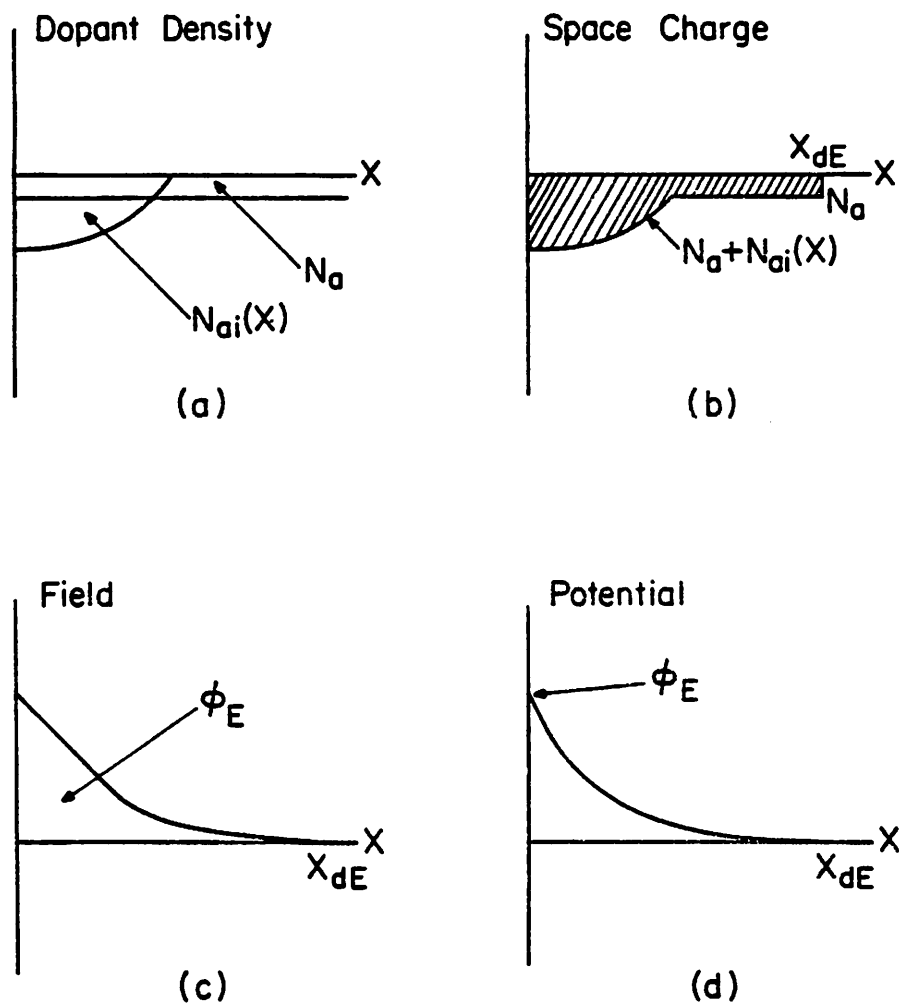
Fig. 2.4. Enhancement device at threshold. Sketches of (a) dopant density, (b) space charge, (c) field and (d) potential.

$$\Phi_E + V_{SB} = \frac{q}{\epsilon_s} \int_0^{X_{dE}} x \Big[N_{a_i}(x) + N_a \Big] dx \tag{2.21}$$

and

$$\Phi_U + V_{SB} = \frac{q}{\epsilon_s} \int_0^{X_{dU}} x N_a \, dx \tag{2.22}$$

where Φ_E and Φ_U are the built-in potentials of the enhancement device and the unimplanted device, respectively.

2.3.2. Depletion-Type Implantation

Under the same set of assumptions, the threshold voltage shift due to the depletion implant is

$$\Delta V_{th} = \Delta \Phi_{b_i} + \Phi_X - \Delta Q_d / C_{ox} \tag{2.23}$$

where $\Delta \Phi_{b_i}$ is the built-in potential difference of the unimplanted device and the depletion device at threshold and is given from (2.11) by

$$\Delta \Phi_{b_i} = \Phi_U - \Phi_D = \frac{kT}{q} \ln \frac{N_{peak-U}}{N_{peak-D}} \quad , \tag{2.24}$$

the potential drop from the maximum potential point to the surface, Φ_X, in the depletion device is

$$\Phi_X = \frac{q}{\epsilon_s} \int_0^{X_1} x \Big[N_{d_i}(x) - N_a \Big] dx \tag{2.25}$$

and ΔQ_d, the depleted charge difference of the unimplanted device and the depletion device at threshold, is

$$\Delta Q_d = Q_{dU} - Q_{dD} = - q \int_0^{X_{dU}} N_a \, dx - q \int_0^{X_1} \Big[N_{d_i}(x) - N_a \Big] dx \ . \tag{2.26}$$

Unlike the enhancement device, there is an additional potential drop Φ_X from the max-
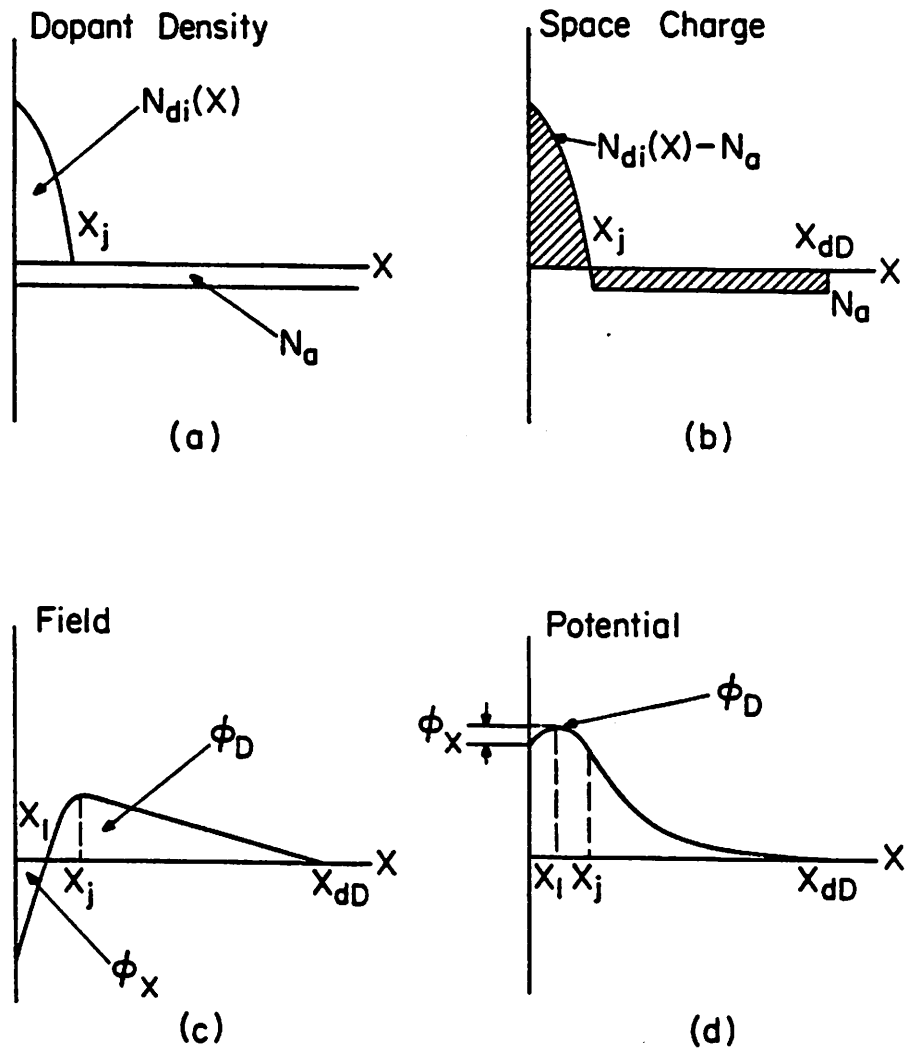
Fig. 2.5. Depletion device at threshold. Sketches of (a)
dopant density, (b) space charge, (c) field and (d)
potential.

imum potential point X_1 to the surface as shown in Fig. 2.5(d). This occurs because the same type of impurity as the channel is implanted and there exists a p-n junction at $x = X_j$ where $N_{di}(X_j) = N_a$. The maximum potential point, X_1, and the depletion region edge of the depletion device, X_{dD}, can be obtained implicitly from the following relations ;

$$\int_{X_1}^{X_j} \left[N_{di}(x) - N_a \right] dx = \int_{X_j}^{X_{dD}} N_a \, dx \qquad (2.27)$$

and

$$\Phi_D + V_{SB} = \frac{q}{\epsilon_s} \int_{X_1}^{X_{dD}} x \left[N_{di}(x) - N_a \right] dx \qquad (2.28)$$

where Φ_D is the built-in potential of the depletion device. For the doubly-implanted device, the equations (2.23) to (2.28) hold after substituting $N_{ai}(x) + N_a$ for N_a.

As shown in (2.18) and (2.23), the threshold voltage shifts resulting from the channel implantations consist of the built-in potential difference and the depleted charge difference. To the first order, the built-in potential differences (2.19) and (2.24) exhibit no dependence on other parameters except the peak carrier densities in the channels of two devices at threshold while the depleted charge differences (2.20), (2.25) and (2.26) depend on the body bias as well as process parameters. The equations in this section are also summarized in Appendix I employing a step profile approximation.

2.4. Temperature Drift of the Threshold Voltage Difference

If a difference in the built-in potentials of two devices exists, the built-in potential difference will contribute a constant temperature drift which is determined by the ratio of the peak carrier densities in the channels of two devices at threshold as shown in (2.19) and (2.24). For example, if the peak carrier densities of two devices at threshold differ from each other by a factor of 2, the temperature drift due to this difference is

$(K/q)\ln 2 = 0.06 mV/^{\circ}C$ at $300^{\circ}K$. The above temperature drift is introduced by the built-in potential difference of only $(kT/q)\ln 2 = 18 mV$ at $300^{\circ}K$. A small difference in the definition of the built-in potentials is negligible in estimating the threshold voltage. However, in the temperature analysis, even a very small difference in the definition of the built-in potentials is likely to defeat the validity of the whole analysis, especially in analyzing the differential temperature coefficient of two threshold voltages. In this work, the built-in potentials are defined as the same for all types of devices, and the temperature drift of the threshold voltage difference of two devices is assumed to be dominated by the variation of the depleted charge difference of two devices. This assumption is justified by the following argument.

The temperature coefficient of the depleted charge difference of two devices is the temperature coefficient difference of the depleted charges of each device at threshold. The temperature coefficient of the depleted charge of one device is proportional to the temperature coefficient of the built-in potential of the device at threshold because the depleted charge depends on the built-in potential. That is, if Φ_{bi1} and Φ_{bi2} are the built-in potentials of two devices, the temperature coefficient of the depleted charge difference of two devices is approximated by

$$T.C. = C_a \frac{d\Phi_{bi1}}{dT} - C_b \frac{d\Phi_{bi2}}{dT} \tag{2.29}$$

$$\approx \left[(C_a - C_b) + \eta \frac{(C_a + C_b)}{2} \right] \frac{d\Phi_{bi}}{dT} \approx (C_a - C_b) \frac{d\Phi_{bi}}{dT}$$

where C_a and C_b are assumed to be constant. The built-in potential difference $\Delta\Phi_{bi}$ is

$$\Delta\Phi_{bi} = \Phi_{bi1} - \Phi_{bi2} , \tag{2.30}$$

and the following relation holds from (2.13) ;

$$\eta = \left| \frac{d\Delta\Phi_{bi}}{dT} \middle/ \frac{d\Phi_{bi}}{dT} \right| = \left| \frac{\Delta\Phi_{bi}}{(\Phi_{bi} - 1.205)} \right| \tag{2.31}$$

$$\approx \ 3.3 \ \% \quad \text{for} \quad \Delta\Phi_{bi} = 18 \ mV \ .$$

For example, from (2.31), if the built-in potential of one device is different from that of the other device by $18mV$, $d\Delta\Phi_{bi}/dT$ is only 3.3% of $d\Phi_{bi}/dT$. In (2.29), the $\Delta\Phi_{bi}$ contribution is neglected since C_a-C_b is actually much bigger than the other term and the $18mV$ difference corresponds to the difference of the peak carrier densities of two devices by a factor of 2. Therefore, even though the built-in potentials are defined to be the same for all types of devices at threshold, it makes little difference in the calculation of the relative temperature coefficient contributed by the depleted charge difference of two devices. By defining $\Phi_{bi}=\Phi_E=\Phi_D=2(kT/q)\ln(N_a/n_i)$, the analyses of the individual channel implants based on a step profile approximation will be discussed in the following subsections.

2.4.1. Enhancement Implant

In this subsection, the temperature coefficient of the threshold voltage shift which results from the enhancement implant is calculated. Assuming a step enhancement implant as shown in Fig. 2.6(a) with an implant depth of X_i and a constant dopant density of N_{ai} and using (2.18)-(2.22), the threshold voltage shift by the enhancement implant is

$$\Delta V_{th} \ = \ \frac{Q_{iE}}{C_{OX}} + \frac{[2q\epsilon_s N_a(\Phi_{bi} + V_{SB})]^{\frac{1}{2}}}{C_{OX}} \left\{ \left[1 - \frac{qN_{ai}X_i^2}{2\epsilon_s(\Phi_{bi} + V_{SB})} \right]^{\frac{1}{2}} - 1\right\} \qquad (2.32)$$

where the enhancement-implant charge density per unit area Q_{iE} is given by

$$Q_{iE} \ = \ qN_{ai}X_i \qquad (2.33)$$

and is independent of temperature. The implant dose $N_{ai}X_j$ only shifts up the threshold voltage by approximately 1 V. The enhancement implant also contributes a higher field in the bulk due to the ionization of the implanted charge. Therefore, the depletion depth of the enhancement device is shorter than that of the unimplanted device with the same bulk concentration.
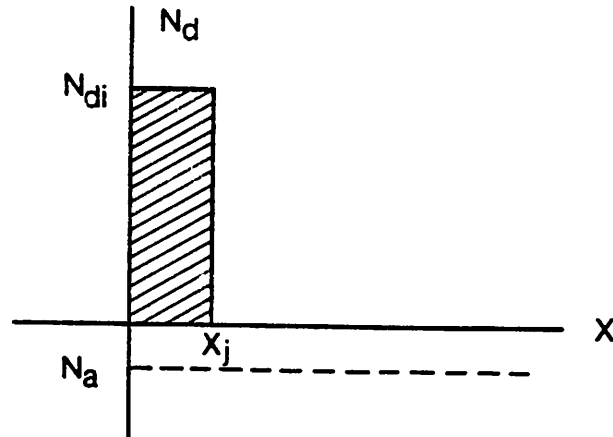
Fig. 2.6. Step profile approximation : (a) enhancement implant
and (b) depletion implant.

If a derivative of (2.32) is taken with respect to temperature, the temperature coefficient of the threshold voltage shift due to the enhancement implant is given by

$$T.C. = \frac{\epsilon_s}{C_{OX}} \left[\frac{qN_a}{2\epsilon_s(\Phi_{bi}+V_{SB})} \right]^{\frac{1}{2}} \left\{ \frac{1}{\left[1 - \frac{qN_{ai}X_i^2}{2\epsilon_s(\Phi_{bi}+V_{SB})}\right]^{\frac{1}{2}}} - 1 \right\} \frac{d\Phi_{bi}}{dT} . \qquad (2.34)$$

Since ΔV_{th} in (2.32) and T.C. in (2.34) all depend on $1/C_{OX}$ or t_{OX}, the temperature coefficient in $ppm/^oC$ unit is thus insensitive to the oxide thickness t_{OX}. Note that the temperature drift is mainly a function of process parameters such as C_{OX}, N_a and N_{ai}, etc.. The only non-process parameter is the body bias V_{SB}. The depleted charge difference is reduced by increasing the body bias and the temperature coefficient of the threshold voltage shift due to the enhancement implant decreases correspondingly as shown in Fig. 2.7(a). As the implant depth is made shallower, the temperature coefficient of the enhancement device relative to that of the unimplanted device approaches zero.

In the figure, the incompleted part at a low body bias is when the depletion edge falls within the implant depth. Therefore, in applying (2.34), care must be taken for the depletion edge not to fall within the implant depth. If $X_{dE}<X_i$, the enhancement device is merely an unimplanted device with a higher bulk density and (2.34) no longer holds. For this case, (2.32) must be modified to

$$\Delta V_{th} = \frac{kT}{q} \ln\frac{N_{ai} + N_a}{N_a} + \frac{[2q\epsilon_sN_a(\Phi_{bi} + V_{SB})]^{\frac{1}{2}}}{C_{OX}} [(1 + \frac{N_{ai}}{N_a}) - 1] . \qquad (2.35)$$

There is a discontinuity in using (2.32) and (2.35) around $X_{dE} \approx X_i$. In case the depleted edge is comparable to the implanted depth, the depletion approximation fails inherently. In this analysis, this limiting case is not treated.
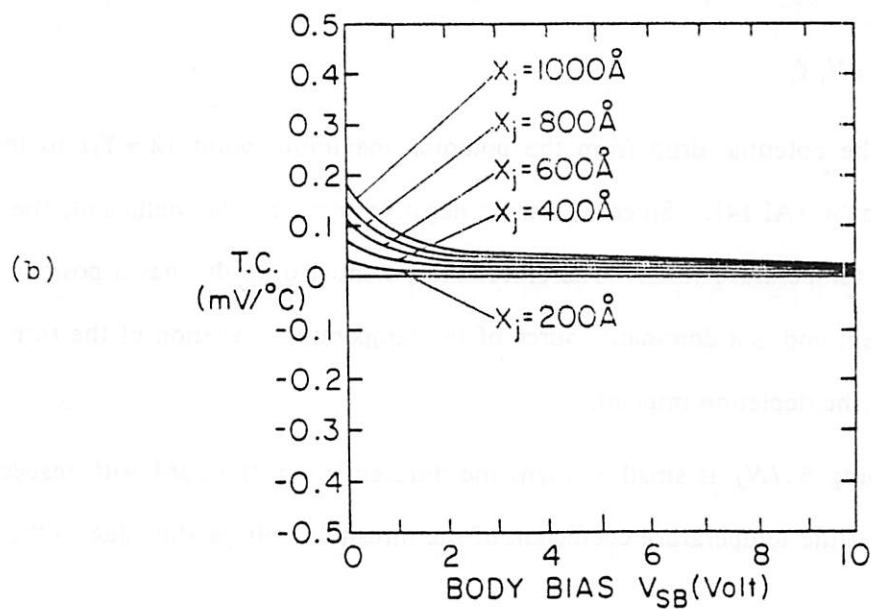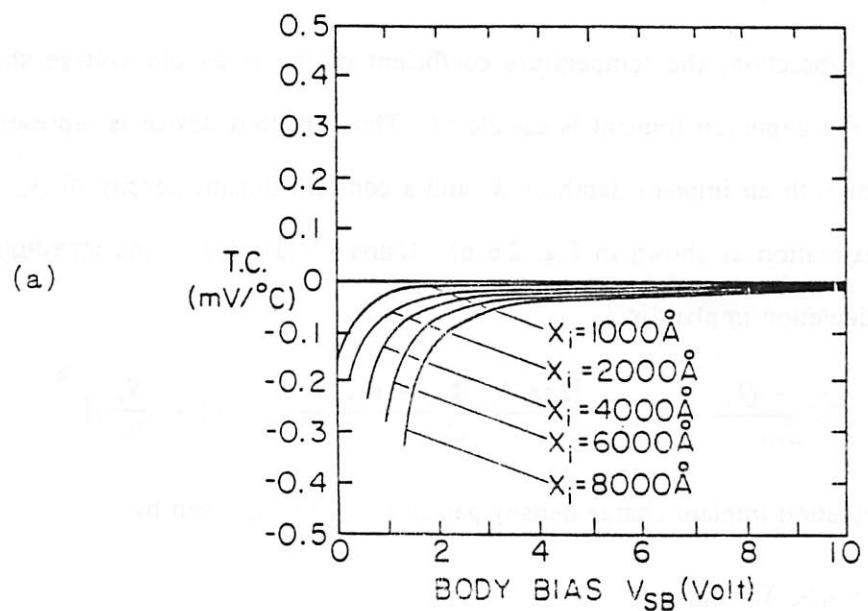
Fig. 2.7. Temperature coefficients of the threshold voltages of
implanted devices relative to those of the unimplanted
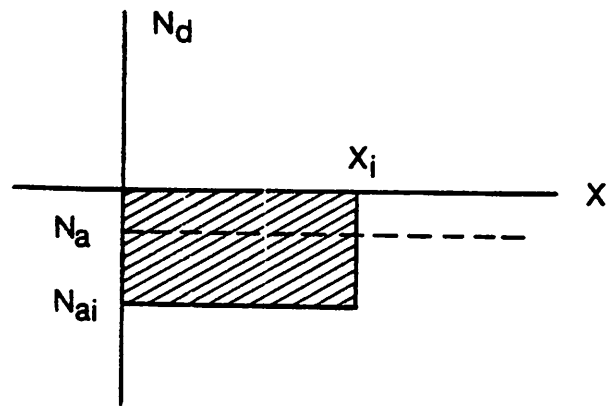device : (a) enhancement implant and (b) depletion
implant.

2.4.2. Depletion Implant

In this subsection, the temperature coefficient of the threshold voltage shift which results from the depletion implant is calculated. The depletion device is represented by a donor implant with an implant depth of X_j and a constant dopant density of N_{di} in a step profile approximation as shown in Fig. 2.6(b). Using (2.23)-(2.28), the threshold voltage shift by the depletion implant is

$$\Delta V_{th} = \frac{Q_{iD} + Q_{iL}}{C_{OX}} + \Phi_X + \frac{[2q\epsilon_s N_a (\Phi_{bi} + V_{SB})]^{1/2}}{C_{OX}} [1 - (1 - \frac{N_a}{N_{di}})] \qquad (2.36)$$

where the depletion-implant charge density per unit area Q_{iD} is given by

$$Q_{iD} = - q N_{di} X_j \qquad (2.37)$$

which is independent of temperature, and only shifts down the threshold voltage by -3 V after compensating the equivalent bulk charge density of

$$Q_{iU} = q N_a X_j \quad , \qquad (2.38)$$

and Φ_X is the potential drop from the potential maximum point $(X = X_1)$ to the surface $(X = 0)$ given by (A1.14). Since Φ_{bi} has a negative temperature coefficient, the depth X_1 increases as temperature rises. Therefore, the potential drop Φ_X has a positive temperature coefficient and is a dominant source of the temperature variation of the threshold voltage shift by the depletion implant.

Assuming N_a/N_{di} is small enough, the differentiation of (2.36) with respect to temperature gives the temperature coefficient of the threshold voltage shift due to the depletion implant of

$$T.C. = \left\{ \frac{N_a}{N_{di}} - X_j \left[\frac{q N_a}{2\epsilon_s (\Phi_{bi} + V_{SB})} \right]^{1/2} \right\} \frac{d\Phi_{bi}}{dT} . \qquad (2.39)$$

The temperature coefficient calculated from (2.39) is shown in Fig. 2.7(b). As in the enhancement implant case, the temperature coefficient of the depletion device relative to

that of the unimplanted device approaches zero as the body bias increases and the implant depth is made shallower. Equation (2.39) consists of two temperature variations which are proportional to N_a and $N_a^{1/2}$, respectively. These variations cancel each other. Unlike the enhancement implant case, C_{OX} does not affect T.C. given by (2.39) while ΔV_{th} in (2.36) depends on $1/C_{OX}$. Therefore, the temperature coefficient in $ppm/^\circ C$ increases as the oxide thickness t_{OX} decreases. If all other parameters are fixed, a shallow single implant and a high body bias help reduce the temperature coefficient given by (2.39).

Let us further examine the body bias sensitivity of the threshold voltage shift due to the depletion implant. The differentiation of (2.36) with respect to the body bias yields

$$\frac{d\Delta V_{th}}{dV_{SB}} = \frac{N_a}{N_{di}} - X_j \left[\frac{qN_a}{2\epsilon_s(\Phi_{bi} + V_{SB})} \right]^{1/2} \tag{2.40}$$

$$\approx -0.02, \quad \text{for} \quad V_{SB} = 2V$$

$$\approx -0.01, \quad \text{for} \quad V_{SB} = 5V.$$

For example, if the body bias changes from $2V$ to $2.1V$, ΔV_{th} changes by $-2mV$ which may be intolerable in a precision circuit. If the body bias is increased to $5V$, the ΔV_{th} variation is reduced by half compared to the $2V$ body bias case for the same amount of the body bias change. Hence, the high body bias helps reduce the body bias sensitivity of the threshold voltage difference. This argument applies to all types of devices.

In applying (2.39), we have to consider the non-pinchoff problem. Equation (2.39) does not hold when X_1 reaches $X_{1,max}$ [22]-[23] which is approximately

$$X_{1,max} = \left[\frac{2\epsilon_s kT}{q^2(N_{di} - N_a)} \ln \frac{(N_{di} - N_a)N_a}{n_i^2} \right]^{1/2} \tag{2.41}$$

$$\approx 92.6 \ nm \quad \text{for} \quad N_{di} = 10^{17} \ cm^{-3} \quad \text{at} \quad 300^\circ K.$$

For the shallow implant of less than $100nm$, the non-pinchoff problem is not present. For

the typical values of $X_j = 100nm$, $N_{d_i} = 10^{17} cm^{-3}$ and $N_a = 5 \times 10^{14} cm^{-3}$, the following equation from (2.39) will estimate the temperature drift of the threshold voltage shift due to the depletion implant.

$$T.C. = -0.011 + 0.14(0.545 + V_{SB})^{1/2} \quad mV/^\circ C \qquad (2.42)$$

$$\approx 0.077 \quad mV/^\circ C \quad \text{for} \quad V_{SB} = 2 \ V$$

$$\approx 0.046 \quad mV/^\circ C \quad \text{for} \quad V_{SB} = 5 \ V \quad \text{at} \quad 300^\circ K \ .$$

2.4.3. Double Implant (Enhancement and Depletion)

In this subsection, the temperature coefficient of the threshold voltage shift which results from the double implant is calculated. The doubly-implanted device has a combination of acceptor and donor implants. That is, the four parameters, X_i and N_{a_i}, X_j and N_{d_i}, are involved in a step profile approximation. Generally, what has been done in the depletion implant case is applied in this case. However, due to the combination of two implementations, the equations are more complex. Using (2.23)-(2.28) and assuming $X_j < X_i$, the threshold voltage shift by the double implant is

$$\Delta V_{th} = \frac{Q_{iD} + Q_{iE} + Q_{iU}}{C_{OX}} + \Phi_X \qquad (2.43)$$

$$+ \frac{[2q\epsilon_s N_a(\Phi_{b_i} + V_{SB})]^{1/2}}{C_{OX}} \left\{ 1 - (1 - \frac{N_a}{N_{d_i} - N_{a_i}}) \left[1 - \frac{qN_{a_i}N_{d_i}(X_i - X_j)^2}{2\epsilon_s(\Phi_{b_i} + V_{SB})(N_{d_i} - N_{a_i})} \right]^{1/2} \right\}$$

where Φ_X is given by (AI.11).

Assuming $N_a/(N_{d_i} - N_{a_i})$ is small enough, the differentiation of (2.43) with respect to temperature gives the temperature coefficient of the threshold voltage shift due to the double implant of

$$T.C. = \left[\frac{N_a}{N_{d_i} - N_{a_i}} + \left[\frac{qN_a}{2\epsilon_s(\Phi_{b_i} + V_{SB})} \right]^{1/2} \right] \left\{ \frac{\epsilon_s}{C_{OX}} \right. \qquad (2.44)$$

$$- \frac{\left[\frac{\epsilon_s}{C_{OX}} + X_j - \frac{N_{a_l}(X_l - X_j)}{N_{d_l} - N_{a_l}}\right]}{\left[1 - \frac{qN_{a_l}N_{d_l}(X_l - X_j)^2}{2\epsilon_s(\Phi_{b_l} + V_{SB})(N_{d_l} - N_{a_l})}\right]^{1/2}} \right\} \right] \frac{d\Phi_{b_l}}{dT} .$$

Equation (2.44) is basically similar to the equation for the depletion implant (2.39) except that the enhancement implant is involved. Thus, the general discussion on the depletion implant holds in the double implant case.

2.5. Validity of a Depletion Approximation

For the convenience of analysis, semiconductor potential barriers are usually represented by two discrete regions such as a depletion region where no free carrier exists and a quasi-neutral region where the charge neutrality is assumed. Generally, the depletion approximation based on the above assumption is not valid within a few Debye lengths from the depletion edge since a free carrier density cannot change abruptly. The free carrier distribution is characterized by the Debye length which is a function of the doping density so that

$$L_D = \left[\frac{2kT\epsilon_s}{q^2 N_l}\right]^{1/2} \tag{2.45}$$

where N_l is the impurity concentration. Therefore, the application of the depletion approximation to the shallow implant case needs considerations.

The potential difference between two points where the carrier concentrations are n_1 and n_2, respectively, is a logarithmic function of the ratio, n_1/n_2,:

$$\Delta\Phi = \frac{kT}{q} \ln \frac{n_1}{n_2} . \tag{2.46}$$

Table II lists the calculated parameter values in the shallow depletion implant case. For example, if $X_j = 28 nm$, the carrier density difference between the field zero point X_1 and the

TABLE II

DEPLETION IMPLANT CALCULATION DATA BASED ON A DEPLETION APPROXIMATION

N_a (cm^{-3})	X_j (nm)	X_i (nm)	L_D (nm)	Φ_X (V)	$\exp\dfrac{q\Phi_X}{kT}$
5×10^{17}	14	11.3	5.7	0.051	7
2.5×10^{17}	28	22.6	8	0.102	52
1×10^{17}	70	56.8	12.7	0.258	2.2×10^4
7×10^{16}	100	81.3	15.2	0.372	1.7×10^6
5×10^{16}	140	114.2	18	0.521	6×10^8
4×10^{16}	170	138.9	19.8	0.653	5×10^{10}
3×10^{16}	220	180.2	22.5	0.826	8.2×10^{13}

surface is only a factor of S2. That is,

$$\frac{n(X=X_1)}{n(X=0)} = \exp\frac{q\Phi_x}{kT} = S2 \; . \tag{2.47}$$

This means free carriers cannot be accumulated like a delta function and the distribution of free carriers is characterized by the Debye length. Therefore, it is reasonable to define the threshold condition of the depletion device in the same way as in the enhancement device.

2.6. Temperature Measurement of IGFET Characteristics

The $I-V$ characteristics of five samples of the four basic types of devices are measured with drain-source voltage kept constant by $4V$ from -30 to $110°C$. The samples are prepared employing a typical Si-gate NMOS process on a 25-30 Ω-cm Boron doped p-type <100> substrate with an oxide thickness of $100nm$, a gate width of $50\mu m$ and a gate length of $7\mu m$. The five sets of measured $I-V$ data are averaged and fitted employing the linear regression method to give the typical temperature variation of the four threshold

Fig. 2.8. Typical measured temperature variations of the four
devices in NMOS technology.

voltages as shown in Fig. 2.8. At room temperature, the threshold voltages of the unimplanted device, the enhancement device, the depletion device and the doubly-implanted device are about 0, 1, −3 and −1V, respectively. Table III summarizes typical process parameters simulated by the process simulation program SUPREM [38]. The implant depth in a step profile approximation of a Gaussian profile is approximated by $R+2\Delta R$ as in [27] where R is a range and ΔR is a straggle.

2.6.1. Threshold Voltage Difference

The measured temperature coefficients of each threshold voltage are processed to give the temperature coefficients of the implanted devices relative to those of the unimplanted device. The dashed lines in Figs. 2.9(a), (b) and (c) show the theoretical temperature coefficients of the enhancement device, the depletion device and the doubly-implanted device relative to those of the unimplanted device, respectively. The dots with the error bars represent the measured values of the relative temperature coefficients. The theoretical

TABLE III

PROCESS PARAMETERS (SUPREM SIMULATION)

Parameters	Enhancement Boron	Depletion Arsenic	Double-implant Boron	Arsenic
Dose ($10^{11}cm^{-2}$)	2.6	14	2.6	9
Energy (KeV)	50	180	50	180
t_{ox} (nm)	103	103	103	103
Range (nm)	160	94.5	160	94.5
Straggle (nm)	55.1	33.2	55.1	33.2
% in Si*	94	57	94	57
Depth in step Approximation (nm)*	300	80	300	80

*Note : Values after drive-in of 30 Min. at $900^{\circ}C$ and 30 Min. at $1000^{\circ}C$.

Fig. 2.9. Measured temperature coefficients of the threshold
voltages of implanted devices relative to those of the
unimplanted device : (a) boron-implanted device and (b)
arsenic-implanted device.
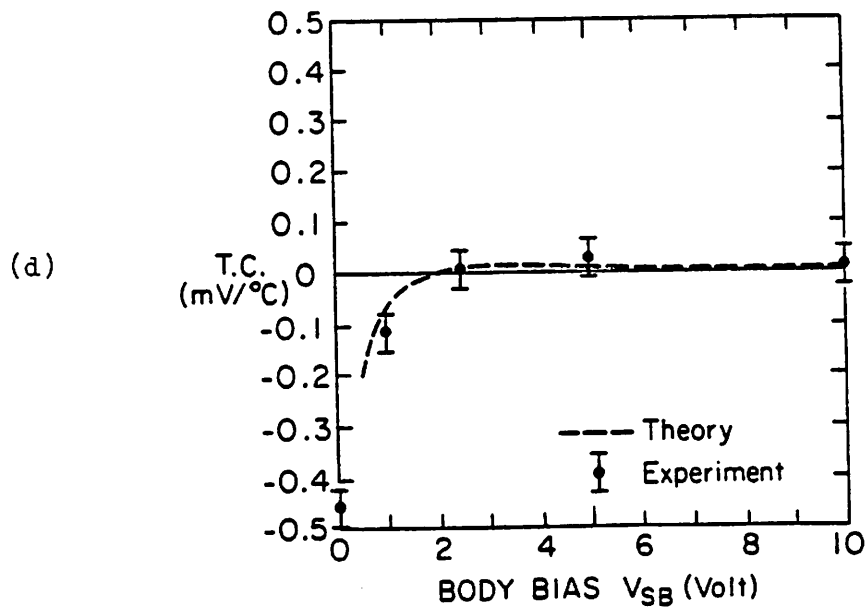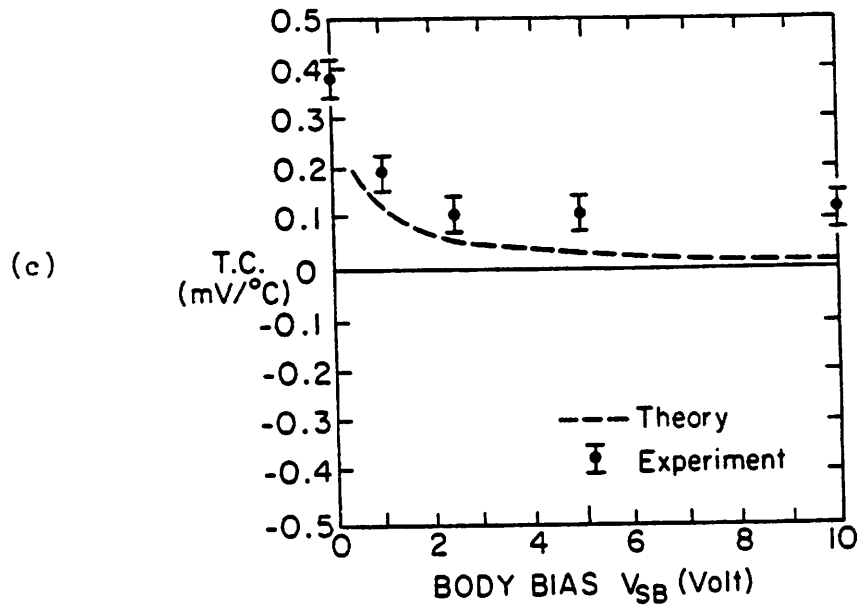
Fig. 2.9. Continued : (c) boron- and arsenic-implanted device and (d) boron-implanted device relative to the arsenic-implanted device.

curves are calculated at $300^{\circ}K$ using (2.34), (2.39) and (2.44) based on the process data listed in Table III. The error bar indicates an equivalent temperature variation of an estimated $1mV$ measurement error.

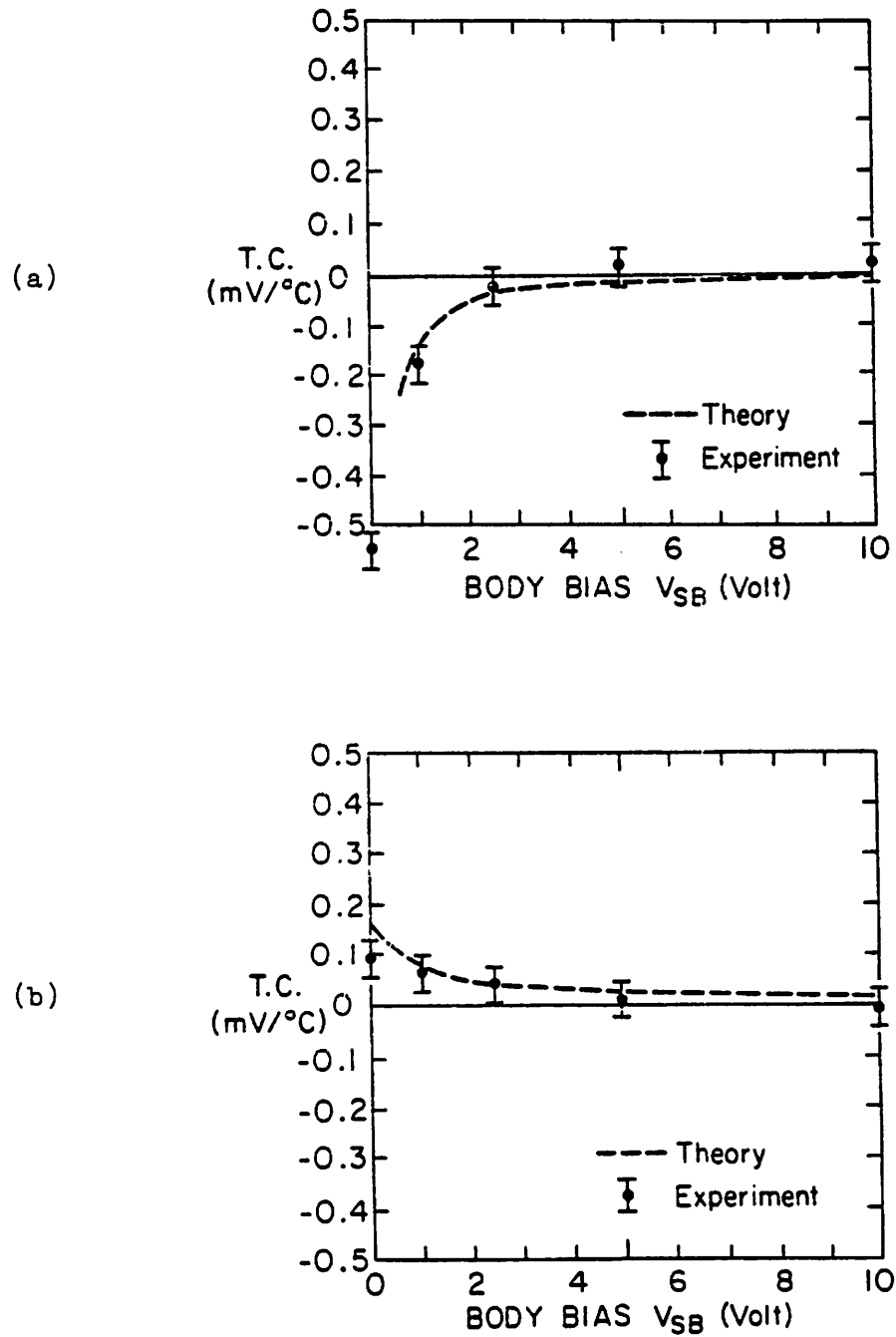In the experimental results, the effect of the body bias on the relative temperature coefficients is as expected. However, the body bias independent constant temperature drifts are observed in all types of devices, especially in the doubly-implanted device (Fig. 2.9(c)). In both the enhancement device and the depletion device, the temperature coefficient deviations from the theoretical values (Figs. 2.9(a) and (b)) are small compared to the measurement error. In the double implant case (Fig. 2.9(c)), however, the measured deviation is on the order of $0.1mV/^{\circ}C$ and is much bigger than expected. This deviation corresponds to the temperature coefficient contributed by the $30mV$ built-in potential difference of the unimplanted device and the doubly-implanted device at threshold. In the enhancement and depletion implant cases, the temperature coefficient deviations from the theoretical values are around $0.02mV/^{\circ}C$ with one in the positive direction and the other in the negative direction. These deviations cancel each other out if two temperature coefficients are added as shown in Fig. 2.9(d).

2.6.2. Gate-Source Voltage Difference

The gate-source voltage of IGFET's is the sum of the threshold voltage and a voltage which depends on the bias current and the channel mobility. Therefore, the channel mobility temperature variation [39] will effect the temperature characteristics of IGFET's. Figure 2.10(a) shows a differential pair voltage reference using the devices with different channel implants. The gate-source voltage difference of two devices with the equal bias current I is

$$\Delta V_{GS} = \Delta V_{th} + [\Delta(\frac{2}{K})^{\frac{1}{2}}]I^{\frac{1}{2}} \tag{2.48}$$

where $K = \bar{\mu}C_{OX}(W/L)$ and (W/L) is the ratio of the channel width to the channel length. The temperature variation at a high bias current level is mainly due to the mismatch in the
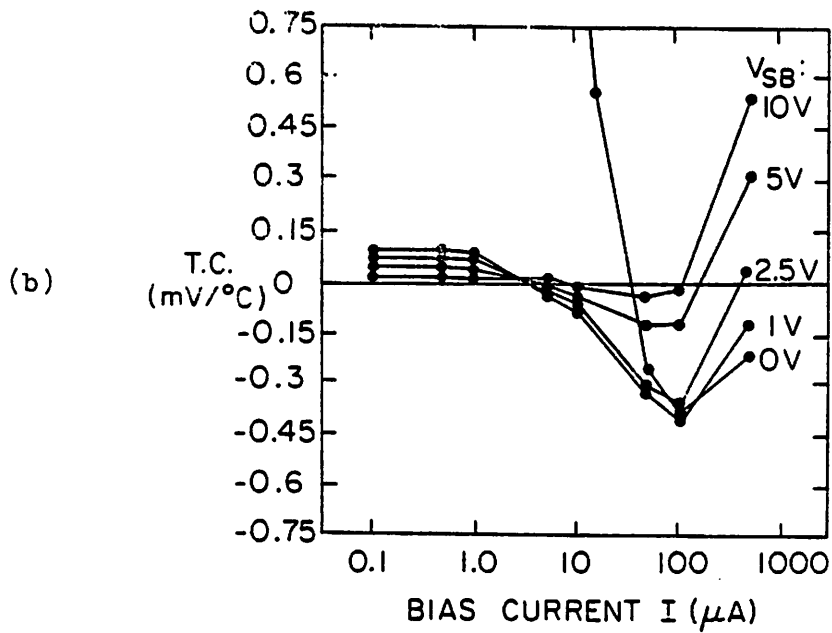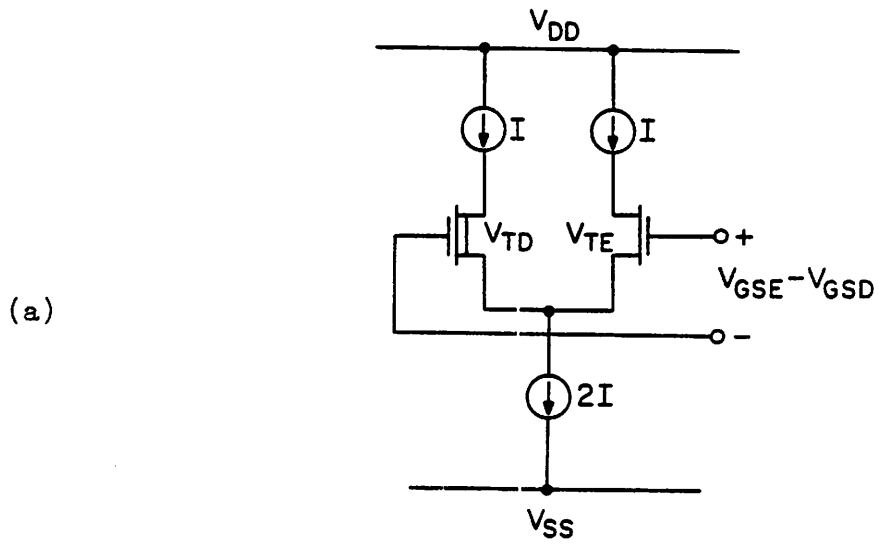
Fig. 2.10. (a) Differential pair composed of two devices with different channel implants. (b) Measured temperature coefficients of the gate-source voltage difference.

channel mobilities of two devices while the temperature variation at a low bias current level is dominated by the threshold voltage difference of two devices.

As an example, Figure 2.10(b) shows the measured temperature coefficient of the gate-source voltage difference for the differential pair composed of the unimplanted device and the depletion device. At a low bias current level, the temperature variation observed is close to that of the threshold voltage difference. On the other hand, at a high bias current level, the temperature variation is quite steep and unpredictable because the mobility mismatch of two devices is involved. As the bias current is lowered, the temperature variation is reduced. Notice that the gate-source voltage difference also changes due to the bias current as well as the body bias as shown in Fig. 2.11. A low bias current operation helps reduce the bias current sensitivity and the temperature drift of the gate-source voltage difference. However, if a device is biased at too low a current level, the device enters the subthreshold region and difficulties are expected in an actual circuit design. The bias current and body bias sensitivities of the gate-source voltage difference may be critical in a precision voltage reference circuit such as proposed by Blauschild et al. [8]. The steep variation at zero body bias (Figs. 2.10(b) and 2.11) in the depletion device is due to the non-pinchoff problem discussed in [22]-[23].

2.6.3. Device Test Results

From the theory developed in Section 2.4 and the discussions of two previous subsections (2.6.1 and 2.6.2) based on the temperature measurements of I-V characteristics in saturation region, the following conclusions can be drawn.

1. The temperature coefficient introduced by a double implant is much bigger than that by a single implant.

2. The body bias helps reduce the temperature drift of both the gate-source voltage difference and the threshold voltage difference of two devices.
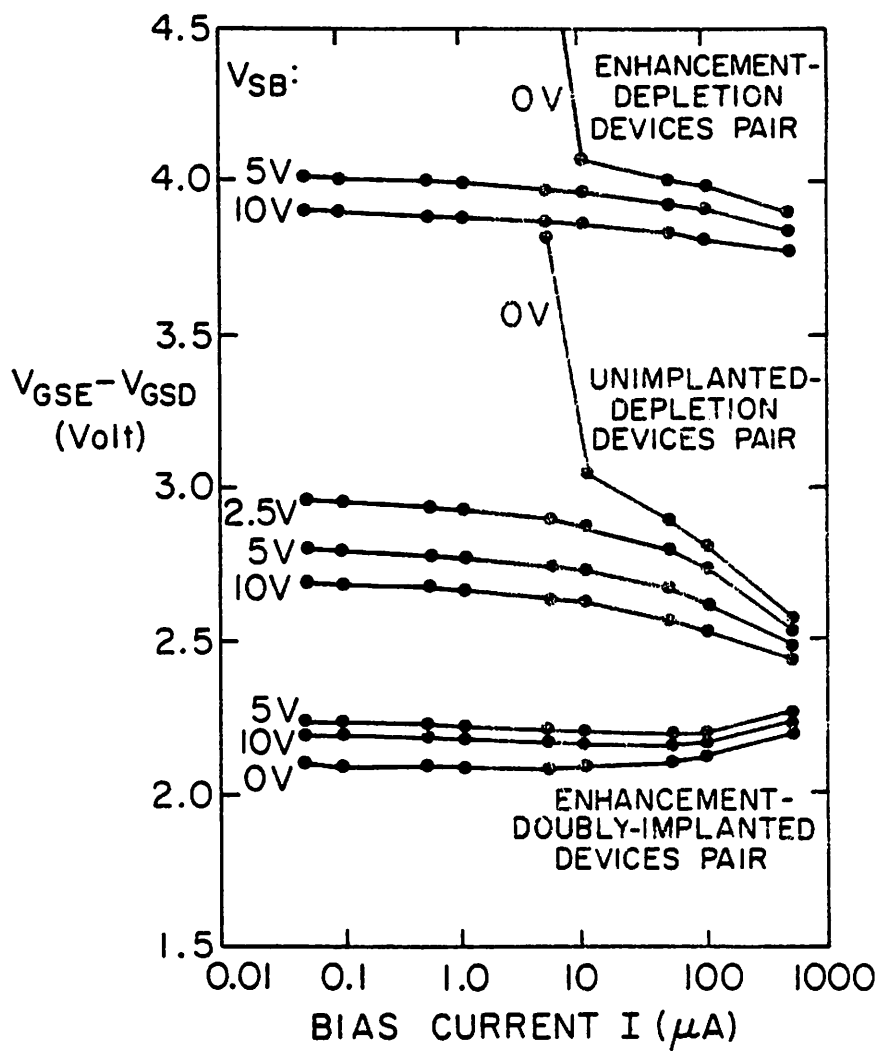
Fig. 2.11. Typical measured gate-source voltage differences as
a function of bias current and body bias.

3. The temperature drift of the threshold voltage shift resulting from the channel implantation is noticeably lowered in a single implant case when the body bias is increased.

4. Below the bias current level of $10\mu A$, the temperature drift of the gate-source voltage difference is very low in a single implant case.

Considering the analysis and the above experimental observations, the threshold voltage difference temperature drift as well as the instability in the bias current and in the body bias will limit the overall performance of the voltage reference circuit based on the threshold voltage difference. So as to achieve a performance predicted, we must

1. make the bias current low and constant,

2. apply a high body bias and keep it constant, and

3. use a single channel implant for the threshold shift.

CHAPTER 3

Si BANDGAP AS A CMOS VOLTAGE SOURCE

3.1. Introduction

Precision bandgap references have been successfully implemented in bipolar technology. However, the performance achievable in bipolar has never been achieved in CMOS. In Chapters 3, 4 and 5, one circuit approach to implement a precision curvature-compensated switched-capacitor CMOS bandgap reference which is compatible with a standard digital CMOS process and achieves a temperature stability on the order of $10ppm/°C$ over the commercial temperature range will be described. In the reference, a temperature-stable voltage is developed by adding a linear and a quadratic temperature correction voltages to the forward-biased diode which is obtained from the substrate pnp transistors available in CMOS process. The linear temperature correction voltage is proportional to the absolute temperature (commonly called PTAT) while the quadratic temperature correction voltage is proportional to the absolute temperature squared (PTATS). They are independently adjustable to set the reference output voltage for a minimum temperature drift.

The offset voltage of a CMOS op amp is eliminated using the correlated-double sampling (CDS) technique [40]. The finite current gain and the base spreading resistance of the native substrate pnp transistors in a standard CMOS process are canceled to the first order and the amplification ratio is set by a capacitor ratio rather than by a resistor ratio. Due to the cyclic behavior of the offset cancellation in this technique, the output reference voltage is not valid at all times. However, the reference can be operated synchronously with other elements of the systems. Since the periodic offset sample and subtraction cycle effectively removes the low frequency $1/f$ noise of a CMOS op amp along with its offset, the dominant noise source is the thermal noise of a CMOS op amp which is designed to be

on the order of $100\mu V$ (rms) at the output in the $500kHz$ bandwidth.

In Section 3.2, the Si bandgap is treated qualitatively to understand its temperature dependence. In Sections 3.3, 3.4 and 3.5, the temperature characteristics of the substrate pnp transistor and other components are investigated theoretically and the primary limitations in a conventional CMOS bandgap reference implementation are discussed. In Section 3.6, a temperature compensation technique suitable for a CMOS bandgap reference is introduced.

3.2. Silicon Energy Gap and its Temperature Dependence

For any semiconductor, there is a forbidden energy region in which no allowed states can exist. Above and below this forbidden energy gap are permitted energy bands. The upper bands are called the conduction bands, and the lower bands the valence bands. The separation between the energy of the lowest conduction band and that of the highest valence band is called the bandgap, E_g ($1.12eV$ for Si at room temperature). The bandgap varies noticeably with temperature. There are two reasons for its temperature dependence. First, the crystal lattice expansion (dilation) due to external stress, pressure or temperature accounts for the change of the energy gap since the energy band depends on the distance between neighboring ions. Second, the electron-lattice interaction energy which results from the indirect optical transition of Si reflects the change of the energy gap with temperature. Although there exist qualitative theoretical analyses describing the temperature dependence of the energy gap, theoretical analyses do not exhibit the accuracy that many empirical equations have, and heavily depend on the parameters obtained from direct measurements such as elastic constant, mobility, compressibility and volume expansion coefficient, etc.. Especially, the quantitative analysis of the second-order temperature dependence is not available.

The concepts of effective mass, density of states, intrinsic carrier concentration and mobility are summarized in Appendix II and, in the following subsections, the qualitative treatment of the Si energy gap is given based on Appendix II for the understanding of the temperature dependence of the Si bandgap.

3.2.1. Energy Gap Narrowing with Impurity Concentration

As impurity concentration increases, the Fermi level approaches the allowed bands, impurity states broaden, and the intrinsic bandgap is narrowed (so called band-edge trailing). The extreme case is the degenerate semiconductor when the impurity concentration is greater than the effective density of states (N_c or N_v). In this case, the Fermi level is located within the conduction or valence band. The conduction or valence band edge is not sharp, and it tails off gradually into the forbidden energy gap. This is due to the fact that the distribution of impurity changes from a delta function to an impurity band. This broadened band joins the conduction or valence band. This makes the band edges trail into the bandgap.

In Si, the energy gap narrowing becomes important at impurity concentrations greater than about $10^{17} cm^{-3}$. The experimentally derived Si bandgap narrowing as a function of the impurity concentration N_i was proposed by Slotboom and De Graaff [41] :

$$\Delta E_{g1}(N_i) = 9 \times 10^{-3} [\ln\frac{N_i}{10^{17}} + (\ln^2\frac{N_i}{10^{17}} + 0.5)^{1/2}]$$

(3.1)

where E_{g1} represents the linearly-extrapolated Si bandgap at $0°K$. For example, the Si bandgap narrows by approximately $68 eV$ for the impurity concentration of $9 \times 10^{19} cm^{-3}$.

3.2.2. Shift of Energy Gap with Lattice Dilation

The band structure varies with a lattice constant for a crystal such as Si or Ge. In these crystals, a decrease in a lattice constant increases the separation between the valence

band and the conduction band. The deformations of lattice such as compression and dilation which are caused by external strain, pressure or temperature, etc. result in the change of energy gaps in semiconductors. Bardeen and Shockley [42] treated the effect of a lattice expansion on the energy gap for the non-polar crystals such as Si and Ge.

When a cubic crystal is subject to a homogeneous strain, the energy gap can be expressed in the form of

$$E_g = E_{og} + E_{1g} \Delta \tag{3.2}$$

where Δ is the lattice dilation and E_{1g} is dependent on the lattice deformation while E_{og} is not. The energy gap change per unit dilation E_{1g} is given approximately by [42]

$$E_{1g} = V \frac{dE_g}{dV} \tag{3.3}$$

$$= -(|E_{1c}| + |E_{1v}|) \approx -17.8 \quad eV/\text{unit dilation} \quad \text{for} \quad \text{Si}$$

where E_{1c} and E_{1v} are the changes of the conduction and valence bands per unit dilation of lattice, with the values of 6.5 and 11.3 per unit dilation, respectively. From the above equation, we can estimate the thermal-expansion effect, or what is often called the dilation contribution, which accounts for the effects of the change of a lattice constant on the energy gap. That is, if it is assumed that the temperature variation is entirely due to the thermal expansion, the temperature variation of the energy gap with dilation can be obtained by multiplying the volume expansion coefficient of $10 \times 10^{-6}/^\circ C$ for Si, so that

$$\frac{dE_g}{dT} = E_{1g} \frac{1}{V} \frac{dV}{dT} \approx -1.78 \times 10^{-4} \quad eV/^\circ K \quad . \tag{3.4}$$

However, this dilation effect on the temperature dependence of the energy gap does not explain the typically measured temperature drift of $-2.4 \times 10^{-4} eV/^\circ K$ of the Si bandgap. The other factor for the extra drift is due to the electron-lattice interaction which will be discussed in the next subsection.

The temperature dependence of conductivity will be dominated by the exponential dependence $\exp(-E_g/2kT)$ since the conductivity is proportional to the product of n, in (AII.16) and μ in (AII.20), with μ being proportional to $T^{-3/2}$. Therefore, the conductivity σ is given by

$$\sigma = \sigma_\infty \exp(-\frac{E_g}{2kT})$$ (3.5)

where the conductivity σ_∞ is independent of temperature. If σ_∞ is assumed to be independent of pressure, the differentiation of (3.5) with respect to pressure gives the pressure dependence of the energy gap so that

$$\frac{dE_g}{dp} = E_{1g} \frac{1}{V} \frac{dV}{dp} = -2kT \frac{1}{\sigma} \frac{d\sigma}{dp} \ .$$ (3.6)

The energy gap can be estimated from two conductivities σ_1 and σ_2 measured at temperatures, T_1 and T_2 $(T_2 > T_1)$:

$$E_g = 2K \ln(\frac{\sigma_2}{\sigma_1}) \ (\frac{1}{T_1} - \frac{1}{T_2})$$ (3.7)

where E_g is assumed to remain constant for T_1 and T_2. The value obtained in this way can be combined with the compressibility data (the change of volume with pressure) to estimate the change in the energy gap with dilation, E_{1g}. For Si, the bandgap decreases with pressure typically at a rate of $2.4\times10^{-6} eV-cm^2/Kg$.

3.2.3. Shift of Energy Gap with Electron-Phonon Interaction

The bottom of the conduction band in Si appears along <100> axes while the valence bands are degenerate at the zone center (000). The direct transition between the valence band and the conduction band occurs in most semiconductors as illustrated in Fig 3.1(a) and (b). In Si, electrons at the lowest energy in the conduction band have a nonzero momentum. Since the holes at the valence-band edge do have a zero momentum, an indirect transition that conserves both energy and momentum is impossible without

Direct Gap Crystal

(a)

(b)
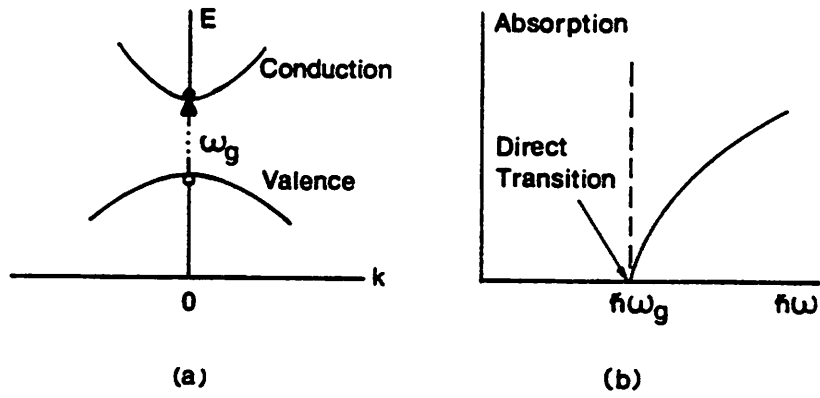
Indirect Gap Crystal

(c)

(d)
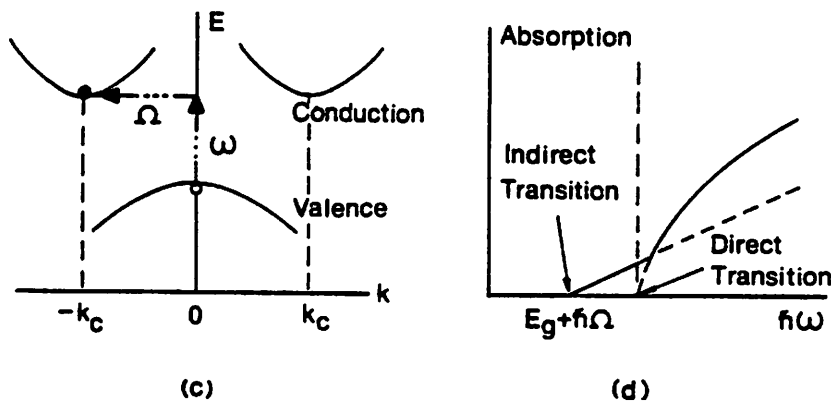
Fig. 3.1. Comparison between the direct optical transition (a)
-(b) and the indirect optical transition (c)-(d).

a lattice interaction occurring simultaneously. Thus, the indirect transition across the forbidden energy gap corresponds to a simultaneous interaction of three particles; the electron, the hole and a phonon that represents the lattice interaction as illustrated in Fig. 3.1(c) and (d). The threshold of continuous optical absorption at the frequency ω_g determines the bandgap $E_g = \hbar\omega_g$ in the direct absorption process because the absorbed photon has a very small wavevectors. However, in the indirect absorption process, generated electrons and holes (exitons) are separated by a substantial wavevector $\bar{k}_c$. For the energy and momentum conservation, the following relations hold depending on the absorption (-) and the emission (+) of phonons.

$$\hbar\omega = E_g - E_{ex} \pm \hbar\Omega \tag{3.8}$$

and

$$\bar{k}(photon) = \bar{k}_c + \bar{K} \approx 0 \tag{3.9}$$

where E_{ex} ($\approx 14 meV$) is the exiton dissociation energy, $\hbar\Omega = 57.3 meV$, and $\bar{K}$ is the wave vector of a phonon [43].

The effect of a lattice dilation on the energy gap accounts only a part of the commonly observed temperature dependence. For example, Bardeen and Shockley [42] estimated E_{1g} to be $-17.8 eV$ from the mobility measurement given by (3.3) while the observed values of E_{1g} are between -30 and -50 eV. Fan [44] explained this discrepancy with the lattice vibration. The reason is that a crystal with a vibrating lattice not only has a different electron-lattice interaction energy, but also has an additional vibrational energy of the ions as compared with a stationary lattice. The energy gap in a vibrating lattice is

$$E_g = E_{og} + E_{1g} + \Delta E_g \tag{3.10}$$

where ΔE_g is the change in the interaction energy when an electron is shifted from a state near the top of the valence band to a state near the bottom of the conduction band.

$E_{uK} + E_{1g}$ denotes the change in energy for the lattice undistorted by the vibration. The temperature variation of the interaction energy ΔE_g of Si is given by [44]

$$\frac{d\Delta E_g}{dT} = -\left(\frac{3}{4\pi^4}\right)^{1/3}\frac{K}{\hbar^2 c_{11} a_0} \ (m^*_e E^2_{1c} + m^*_h E^2_{1v}) \tag{3.11}$$

$$\approx -1.8 \times 10^{-4} \ eV/^\circ K \ .$$

where a_0 is the Si lattice constant, c_{11} is the Si elastic constant, and $m^*_{e,h}$ are the conductivity effective masses of electrons and holes, respectively. Therefore, the summing of (3.4) and (3.11) leads to the value of $-3.58 \times 10^{-4} eV/^\circ K$, which explains the first-order temperature variation of the Si bandgap properly.

3.2.4. Empirical Temperature Dependence of the Si Energy Gap

The direct measurement data of the Si bandgap by Macfarlane et al [45] were fitted by a simple empirical equation from $0^\circ K$ to $400^\circ K$ by Varshni [46].

$$E_g(T) = E_{g1} - \frac{cT^2}{T+d} \tag{3.12}$$

where $E_{g1} = 1.1557 eV$ in case we neglect the exiton dissociation energy E_{ex}, $c = 7.021 \times 10^{-4} eV/^\circ K$ and $d = 1108^\circ K$. These constants were modified in the later works [43], [47] for better fitting to measurement data.

Another purely empirical equation was proposed by Bludau et al [48] which fits measurement data within $0.2 meV$ below $300^\circ K$. Especially from $150^\circ K$ to $300^\prime K$, the Si bandgap varies with temperature so that

$$E_g(T) = E_{g1} - aT - bT^2 \tag{3.13}$$

where

$$E_{g1} = 1.1785 \ eV \ ,$$

$$a = 9.025 \times 10^{-5} \ eV/^\circ K \quad \text{and}$$

$$b = 3.05 \times 10^{-7} \ eV/{}^{\circ}K^2 \qquad \text{for} \quad 150^{\circ}K < T < 300^{\circ}K \ .$$

Assuming a linear temperature dependence above $300^{\circ}K$, Tsividis [49] calculated the coefficients from $300^{\circ}K$ to $400^{\circ}K$ as

$$E_{g1} = 1.20595 \ eV \ , \tag{3.14}$$

$$a = 2.7325 \times 10^{-4} \ eV/{}^{\circ}K \quad \text{and}$$

$$b = 0 \qquad \text{for} \quad 300^{\circ}K \leqslant T < 400^{\circ}K \ .$$

When estimating from (3.13) and (3.14), $E_g(T)$ exhibits about $3mV$ $(12ppm/{}^{\circ}C)$ maximum deviation from the ideal linear temperature dependence from $200^{\circ}K$ to $400^{\circ}K$. This temperature nonlinearity appears as an inherent curvature in the voltage reference based on the Si bandgap.

3.3. Temperature Dependence of Substrate PNP Transistors

3.3.1. Temperature Variation of the Emitter-Base Potential

The collector current I_C is an exponential function of the emitter-base potential V_{BE} :

$$I_C = I_s \exp(\frac{V_{BE}}{V_T}) \ . \tag{3.15}$$

The current I_s is the reverse saturation current and is related to the device structure by

$$I_s = \frac{q^2 A_e^2 n_i^2 \overline{D}_B}{Q_B} \tag{3.16}$$

where A_e is the effective emitter area, n_i is the intrinsic carrier concentration in the base, Q_B is the total built-in base charge per unit area, and $\overline{D}_B$ is the average minority carrier diffusion constant in the base. The number of base dopant atoms per unit area (the Gummel number) in the quasi-neutral base region is

$$\frac{Q_B}{qA_e} = \int_0^{X_B} N_B(x)\, dx \tag{3.17}$$

where N_B is the dopant density in the base and X_B is the depth of the base region. The diffusion constant $\overline{D}_B$ is not a direct spatial average, but rather

$$\overline{D}_B = \frac{\displaystyle\int_0^{X_B} N_B(x)\, dx}{\displaystyle\int_0^{X_B} \frac{N(x)}{D_B(x)}\, dx} . \tag{3.18}$$

Using (3.16) and the Einstein relation, $D=(kT/q)\mu$, I_s is written as

$$I_s = BT\overline{\mu}n_i^2 \tag{3.17}$$

where B is a temperature-independent constant. If we assume the temperature variation of the minority carrier mobility is the same as that of the majority carrier mobility [41], the mobility is a function of temperature, T^{-n} ($n\approx 3/2-5/2$) from (AII.20). Using (3.15), (3.19) and (AII.16), the emitter-base potential drop is

$$V_{BE} = V_g - V_T(\gamma\ln T - \ln I_C - \ln E) \tag{3.20}$$

where E is a temperature-independent quantity whose exact value is not important in the temperature analysis, $\gamma = 4-n$, and V_g is the Si bandgap voltage.

In an actual circuit, I_C is a temperature-dependent current as

$$I_C = GT^\alpha \tag{3.21}$$

where G is another temperature-independent coefficient. Combining (3.20) and (3.21), we obtain the following relation which explains the temperature dependence of the emitter-base potential V_{BE} [50] :

$$V_{BE} = V_g - V_T[(\gamma-\alpha)\ln T - \ln EG] . \tag{3.22}$$

In (3.22), the Si bandgap voltage V_g is generally a nonlinear function of temperature which is described in the previous section.

The derivative of (3.22) with respect to temperature gives the temperature drift of the emitter base potential V_{BE} at T so that

$$\frac{dV_{BE}}{dT} = \frac{dV_R}{dT} - \frac{V_T}{T} [(\gamma-\alpha)(\ln T - 1) - \ln EG] \quad . \tag{3.23}$$

Assuming $\gamma=2.623$ and $EG=0.00556^o K^{(\gamma-\alpha)}$, the first-order temperature dependence of V_{BE} is approximated by

$$\frac{dV_{BE}}{dT} \approx -0.273 mV/^oC - \frac{0.026}{300} [2.623\times(\ln 300 - 1) - \ln 5.56\times 10^{-3}] \tag{3.24}$$

$$\approx -1.79 \quad mV/^oC \quad \text{at} \quad 300^o K$$

where (3.14) is used for the first-order temperature drift of the Si bandgap. Experimentally observed temperature drift of V_{BE} is about $-2mV/^oC$.

3.3.2. Temperature Variation of the Transistor Current Gain

The temperature dependence of the bipolar transistor current gain has been accounted for many factors. Among them, the bandgap narrowing due to the high emitter doping [51]-[53] and the non-ideality of the base current [54] have been suspected for the variation of β with temperature. Since a rigorous theoretical analysis is not feasible, an order-of-magnitude analysis is given here to understand the current gain temperature dependence.

First assume a transistor is ideal and has the uniform emitter and base doping concentrations of N_E and N_B with the corresponding depths of X_E and X_B. Also neglect the space charge recombination in the emitter-base junction. Then, the current gain is

$$\beta = \frac{I_C}{I_B} = \frac{\alpha_F}{1-\alpha_F} \tag{3.25}$$

where α_F is the ratio of the collector current to the emitter current, I_C/I_E. α_F is generally represented by a product of two terms so that

$$\alpha_F = \alpha_T \alpha_E \qquad (3.26)$$

where α_T is the base transport factor and α_E is the emitter efficiency. If the emitter and base regions are assumed to be short compared with the diffusion lengths, α_T and α_E are given by

$$\alpha_T = 1 - \frac{X_B^2}{2L_B^2} \qquad (3.27)$$

and

$$\alpha_E = \frac{1}{1 + \dfrac{D_E X_B N_B n_{ie}^2}{D_B X_E N_E n_{ib}^2}} \qquad (3.28)$$

where D is the diffusion constant, L_B is the diffusion length, n_i is the intrinsic carrier concentration, and the subscripts e, E, b and B represent emitter and base, respectively. For high β transistors, usually $X_B/L_B << 1$ and α_F is mainly determined by the emitter efficiency. Neglecting α_T, we have

$$\beta = \frac{\alpha_E}{1 - \alpha_E} = \frac{D_B X_E N_E n_{ib}^2}{D_E X_B N_B n_{ie}^2} . \qquad (3.29)$$

Usually, n_{ie}^2 and n_{ib}^2 are assumed to be equal. Due to the bandgap narrowing in the emitter region where doping concentration is relatively high, however, the intrinsic carrier concentration in the emitter is different from that in the base so that

$$n_{ie}^2 = C_E T^3 \exp(-\frac{E_g - \Delta E_g}{kT}) \quad \text{and} \qquad (3.30)$$

$$n_{ie}^2 = C_B T^3 \exp(-\frac{E_g}{kT})$$

where C_E and C_B are constants. Therefore, the current gain β is

$$\beta = C \frac{\mu_B}{\mu_E} \exp(-\frac{\Delta E_g}{kT}) \qquad (3.31)$$

where the Einstein relation $D = (kT/q)\mu$ is used, and the constant

$C = (C_B X_E N_E)/(C_E X_B N_B)$ is assumed to be temperature independent. In the base region, the impurity concentration is low and the mobility μ_B decreases with temperature due to the lattice scattering while, in the emitter region, the mobility μ_E increases with temperature due to the scattering by the ionized impurities resulting from the high emitter dopant concentration. Therefore, the mobility ratio μ_B/μ_E should decrease with temperature while the exponential dependence of ΔE_g increases with temperature. Experimentally observed temperature coefficients range from $6000 ppm/^oC$ to $8000 ppm/^oC$. Thus, the exponential variation attributable to the bandgap narrowing is accounted for the variation of β with temperature. For the emitter doping density of $2.2 \times 10^{19} cm^{-3}$ in the substrate pnp structure, ΔE_g is approximately $0.1 eV$. This ΔE_g alone gives

$$\frac{1}{\exp(-\frac{\Delta E_g}{kT})} \frac{d\exp(-\frac{\Delta E_g}{kT})}{dT} = \frac{\Delta E_g}{kT} \frac{1}{T} \tag{3.32}$$

$$\approx \frac{0.1 eV}{0.026 eV} \times \frac{1}{300^o K} = 12820 \quad ppm/^oC$$

which is approximately twice the typically observed temperature coefficient of β.

3.4. Temperature Dependence of Components

3.4.1. Temperature Variation of Diffused Resistors

Resistors available in CMOS process are n^+ diffusion resistor (25 $\Omega/square$), p^+ diffusion resistor (75 $\Omega/square$), polysilicon resistor (30 $\Omega/square$) and n^- well diffusion resistor ($10K$ $\Omega/square$). The polysilicon and well diffusion resistors are harder to control compared with the n^+ or p^+ source/drain diffusion resistors because of a poly undercut ($0.75 \mu m$) and a well lateral diffusion ($3 \mu m$). Although the n^+ diffusion has less side diffusion than the p^+ diffusion, the lower resistance of the n^+ diffusion (a third of the

p^+ diffusion) occupies more area than the p^+ diffusion. Therefore, a $6\mu m$-wide p^+ diffused resistor in the n^- well is chosen as a resistor in the bias circuit. In the current process, the depth of the p^+ diffusion is $0.9\mu m$ and it has a $0.7\mu m$ side diffusion. Thus, the effective width is $7.4\mu m$. The boron implant dose of the p^+ diffusion is $2\times10^{15}cm^{-2}$ at the energy of $60KeV$ while the phosphorus implant dose of the n^- well is $1.5\times10^{12}cm^{-2}$ at the energy of $100KeV$.

If step profiles are assumed, the dopant density of the p^+ diffusion is

$$N_A = \frac{2\times10^{15}cm^{-2}}{0.9\mu m} \approx 2.2\times10^{19} \ cm^{-3} \ , \tag{3.33}$$

while that of the n^- well is

$$N_D = \frac{1.5\times10^{12}cm^{-2}}{5.5\mu m} \approx 2.7\times10^{15} \ cm^{-3} \ . \tag{3.34}$$

The depleted region depth at the p^+ diffusion side of the p^+ diffusion and n_- well junction is approximately

$$X_{dp} = \left[\frac{2\epsilon_s(\Phi_{bj} + V_R)}{qN_A(1 + N_A/N_D)} \right]^{1/2} \tag{3.35}$$

$$= \left[\frac{2\times10^{-12}(0.866 + 5)}{1.6\times10^{-19}\times2.2\times10^{19}(1 + 2.2\times10^{19}/2.7\times10^{15})} \right]^{1/2} \approx 2\times10^{-4} \ \mu m$$

where the built-in potential across the junction is

$$\Phi_{bj} = \frac{kT}{q} \ln \frac{N_A N_D}{n_i^2} = 0.866 \ V \ , \tag{4.36}$$

and V_R is the junction reverse bias voltage.

The diffused resistor with the charge density $N(x)$ as a function of the depth x from the surface has the resistance per square of

$$R_{sq} = \frac{1}{\int_0^t q\mu(x)N(x)dx}$$

(3.37)

where t is the depth from the surface to the junction. If we assume the step dopant density extending to the depth t and assume the constant mobility, R_{sq} becomes

$$R_{sq} = \frac{1}{q\mu N_A t}$$

(3.38)

Since all the dopant impurities are assumed to be ionized, N_A is independent of temperature. Therefore, most of the temperature variation is from the variation of the mobility and the depth with temperature. If we differentiate (3.38) with temperature, we obtain

$$\frac{1}{R_{sq}} \frac{dR_{sq}}{dT} = -\frac{1}{\mu} \frac{d\mu}{dt} - \frac{1}{t} \frac{dt}{dT} .$$

(3.39)

If the two scattering mechanisms discussed in Appendix II are considered, we have

$$\frac{1}{\mu} \frac{d\mu}{dT} = \frac{1}{\mu_l} \frac{d\mu_l}{dT} + \frac{1}{\mu_i} \frac{d\mu_i}{dT}$$

(3.40)

where μ_l and μ_i are the mobilities due to the lattice scattering and to the ionized impurities, respectively. Theoretically, μ_l changes with temperature approximately by $T^{-3/2}$ and μ_i by $T^{3/2}$. In reality, other scattering mechanisms are involved and μ_l varies with temperature approximately by $T^{-5/2}$. The following order-of-magnitude calculation explains the mobility temperature variations.

$$\frac{1}{\mu_l} \frac{d\mu_l}{dT} \approx -\frac{5}{2} \frac{1}{T} \approx -8300 \quad ppm/^oC \quad \text{at} \quad 300^oK$$

(3.41)

and also

$$\frac{1}{\mu_i} \frac{d\mu i}{dT} \approx \frac{3}{2} \frac{1}{T} \approx 5000 \quad ppm/^oC \quad \text{at} \quad 300^oK .$$

(3.42)

The junction depth t must be the the real junction depth t_1 minus the depletion region. From (3.35), we have

$$t = t_1 - X_{dp} \qquad (3.43)$$

where t_1 ($\approx 0.9\mu$) is the depth where the dopant density equals the background density, $N(t_1)=N_D$. Therefore, we obtain

$$\frac{1}{t}\frac{dt}{dT} = \frac{X_{dp}}{2(t_1 - X_{dp})}\frac{1}{\Phi_{bj}}\frac{d\Phi_{bj}}{dT} \qquad (3.44)$$

Since $X_{dp} \ll 2(t_1 - X_{dp})$, the thickness variation is negligible compared with the mobility variation given by (3.40). the temperature variation of Φ_{bj} can be estimated from (2.13) as

$$\frac{1}{\Phi_{bj}}\frac{d\Phi_{bj}}{dT} = \frac{1}{\Phi_{bj}}\frac{1}{T}(\Phi_{bj} - 1.21) \qquad (3.45)$$

$$= \frac{1}{0.866}\times\frac{1}{300}(0.866 - 1.21) \approx -1324 \; ppm/^{\circ}C \quad \text{at} \quad 300^{\circ}K \quad .$$

We can therefore conclude the temperature variation of p^+ diffused resistors is mainly determined by the mobility variation with temperature. The typically observed temperature coefficient of p^+ diffused resistors is about $900 ppm/^{\circ}C$.

3.4.2. Temperature Variation of Capacitors

The poly-diffusion capacitance is the series sum of the oxide capacitance and the space charge capacitances as

$$\frac{1}{C} = \frac{1}{A_c}\left[\frac{1}{C_{OX}} + \frac{1}{C_s}\right] = \frac{1}{A_c}\left[\frac{1}{C_{OX}} + \frac{1}{C_{s1}} + \frac{1}{C_{s2}}\right] \qquad (3.46)$$

where C_{s1} and C_{s2} are the space charge capacitances per unit area of the top-plate polysilicon and the bottom n^+ diffusion, respectively, and A_c is the capacitor top plate area. The space charge capacitances are originated from the spatial distribution of the majority carriers at the poly-oxide and the oxide-silicon interfaces. The temperature coefficient of a capacitance is [55]

$$\frac{1}{C}\frac{dC}{dT} = \left[\frac{1}{A_c}\frac{dA_c}{dT} - \frac{1}{t_{OX}}\frac{dt_{OX}}{dT}\right] + \frac{C_{OX}}{C_s^2}\frac{dC_s}{dT} + \frac{1}{\epsilon_{OX}}\frac{d\epsilon_{OX}}{dT} \tag{3.47}$$

where the first term represents the thermal expansion effect of the physical dimensions, the second the change of the space charge regions, and the last the oxide dielectric constant change.

Since the polysilicon and the diffusion layer are heavily doped, C_s is much greater than C_{OX}. Then the space charge capacitance contribution is negligible compared to the other variations. The thermal expansion accounts approximately for $6 ppm/^\circ C$ and ϵ_{OX} for $20 ppm/^\circ C$. The typically observed temperature coefficient of capacitors is about $25 ppm/^\circ C$. In this work, what is important is the differential temperature coefficient of a capacitor ratio which is supposed to be much smaller than the absolute temperature coefficient on the order of $25 ppm/^\circ C$.

3.4.3. Temperature Variation of MOS Transistor Mismatches

The temperature variation of the mismatch voltage (the gate-source voltage difference of two identical devices) due to the channel ion-implantation is discussed extensively in Chapter 2. However, the temperature behavior of the mismatch voltage of two identical devices which can also be called the offset temperature drift is important in the design of the temperature circuit such as a voltage reference because MOS transistors are used in pair to form an input stage of a differential pair and a current mirror which depends on the identical device characteristics. Judging from the results from Chapter 2, the mismatch voltage is a limiting case of the gate-source voltage when the ion-implantation depth approaches zero. That is, the delta profile of the implanted ions does not affect the temperature behavior of the devices, but rather does shift only the device threshold voltage. For this reason, the mismatch voltages of two identical MOS devices under the same bias conditions should not drift with temperature. This results contradict with the offset drift of two identical bipolar devices which exhibits a linear temperature dependence. The

TABLE IV

MISMATCH VOLTAGE (mV) TEMPERATURE DATA

	$Bias(\mu A)$	$-30^\circ C$	$25^\circ C$	$75^\circ C$	$110^\circ C$
Pair #1	5	-27	-25	-29	-26
	10	-29	-28	-30	-29
	20	-35	-34	-36	-38
	30	-39	-38	-39	-38
pair #2	5	-22	-16	-20	-18
	10	-20	-17	-21	-20
	20	-21	-20	-25	-25
	30	-24	-23	-23	-23
Pair #3	5	-46	-40	-43	-43
	10	-48	-45	-47	-46
	20	-53	-51	-52	-54
	30	-58	-58	-56	-55
pair #4	5	21	24	24	26
	10	23	26	25	27
	20	24	24	25	23
	30	24	25	25	26

measured mismatch voltage temperature variations of four MOS transistor pairs with the gate width of $50\mu m$ and the length of $7\mu m$ are listed in Table IV. For the measurement, V_{DS} and V_{SB} were kept constant at $4V$ and $5V$, respectively. Note the magnitudes of the mismatch voltages are big (20 to $60mV$) while their temperature drifts are randomly distributed within a few mV which is about the same order as the measurement error corresponding to the temperature setting error.

The figure shows a conventional CMOS bandgap reference circuit with resistors R_2, R_3, R_1, an op amp with offset voltage V_{os}, transistors Q_1 and Q_2, producing output V_{ref}, with currents I_1 and I_2. The relevant equations are:

$$V_{ref} = V_{BE} + \left(1 + \frac{R_2}{R_1}\right)(\Delta V_{BE} + V_{os})$$
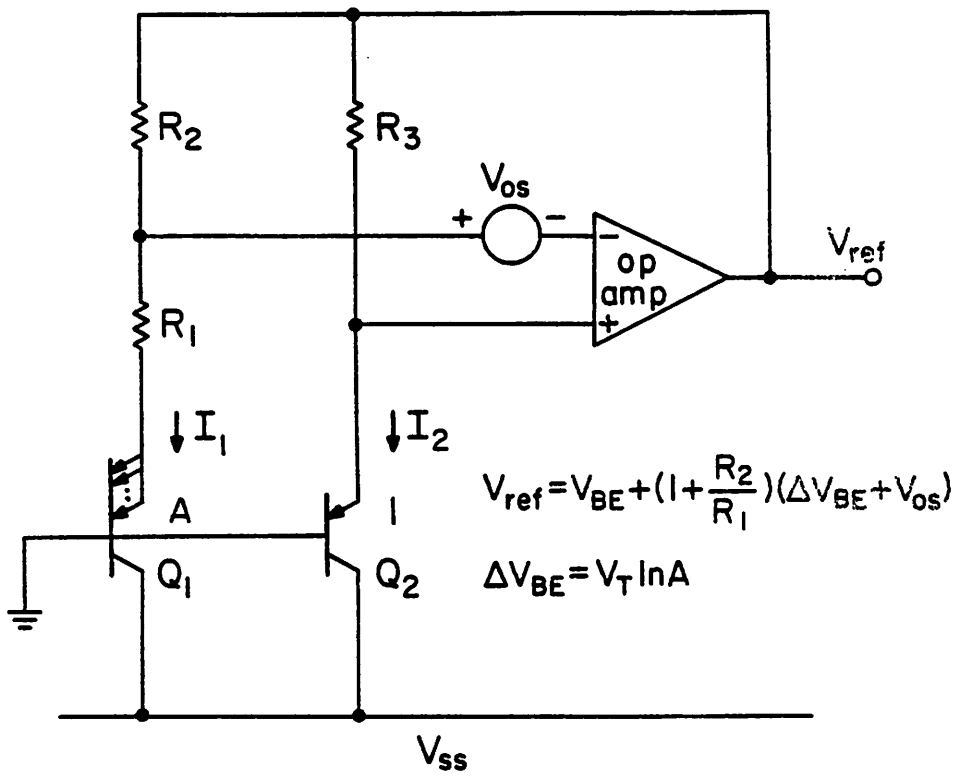
$$\Delta V_{BE} = V_T \ln A$$

Fig. 3.2. A conventional CMOS bandgap reference implementation.

3.5. Conventional CMOS Bandgap Reference and Its Error Sources

A conventional CMOS BGR implementation is shown in Fig. 3.2 [56]. The transistors Q_1 and Q_2 are substrate pnp transistors whose collectors are always tied to the most negative power supply because, in an n^- well CMOS process, the p^+ diffusion in the n^- well, the n^- well itself and the p^- substrate form a vertical pnp structure as shown in Fig. 3.3(a). Therefore, it is not possible to sense the collector current directly as in the bipolar bandgap reference so as to reduce the error due to the finite current gain [4], [6]. In a p^- well process, a dual circuit incorporating npn transistors would be used. While many other circuit implementations are possible, this circuit appears to be as good as any, and the contributions to the non-ideal V_{ref} will be analyzed one by one. All resistors are the p^+ diffusion resistor in the n^- well and the op amp is assumed to have an infinite gain with the offset voltage of V_{os}. This assumption is justified because CMOS op amps usually have enough gains.

Assume that Q_1 has A times larger emitter area than Q_2 and they are biased in the forward active region. Then, including the nonidealities shown in Fig. 3.3(b), V_{ref} is given by

$$V_{ref} = V_{BE} + (1 + \frac{R_2}{R_1})(\Delta V_{BE} + V_{os}) \tag{3.48}$$

where

$$V_{BE} = V_T \ln \frac{I_1}{I_{s1}} + V_T \ln \frac{1}{1 + \frac{1}{\beta_1}} + \frac{r_b I_1}{A \beta_1} \quad \text{and} \tag{3.49}$$

$$\Delta V_{BE} = V_T \ln A + V_T \ln \frac{I_2}{I_1} + V_T \ln \frac{1 + \frac{1}{\beta_1}}{1 + \frac{1}{\beta_2}} + r_b (\frac{I_2}{\beta_2} - \frac{I_1}{A \beta_1}) \quad . \tag{3.50}$$

V_{BE} is the emitter-to-base voltage and all others have their usual meanings. The remainings except the first terms of (3.49) and (3.50) represent the error sources we may
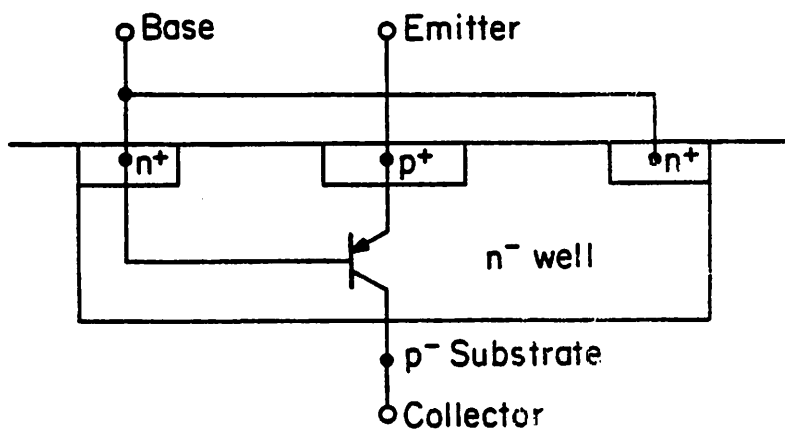
(a)
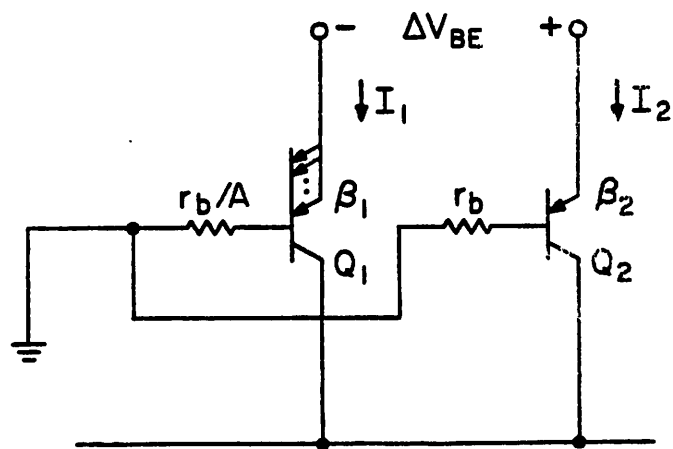
(b)

Fig. 3.3. (a) Profile of the substrate vertical pnp transistor
in an n-well CMOS process. (b) A transistor pair with
non-ideal parameters.

encounter in the existing designs such as shown in Fig. 3.2. Among these, the four important ones we may not overlook are the offset voltage V_{os}, the bias current variation of V_{BE}, the finite current gain β and the base resistance r_b. What is worse in V_{os} and the error in ΔV_{BE} is that they are amplified by a factor of $(1+R_2/R_1)$ which is about 10.

In the following subsections, these error sources are considered one by one to see an order-of-magnitude contribution of the individual error sources.

3.5.1. OP Amp Offset

The offset of the op amp is the biggest error source that causes the spread of the temperature stability. Normally, a bandgap reference is trimmed to an output voltage which is predetermined to give a near-zero temperature coefficient of the output. The large, non-PTAT component in the output due to the op amp offset causes the trimming operation to give an erroneous result. If we assume V_{os} is independent of temperature, the resulting temperature coefficient error due to V_{os} is approximately

$$T.C. \ Error \ = \ \frac{(1+\frac{R_2}{R_1})\,V_{os}}{V_{ref}\,T_o} \ \approx \ \frac{10\times 5mV}{1.26\,V\times 300^o K} \ \approx \ 132 \ ppm/^oC \ . \tag{3.51}$$

That is, the spread on the order of $132 ppm/^oC$ in the reference output temperature coefficient is expected from the $5mV$ temperature-independent offset voltage.

3.5.2. Bias Current Variation

The bias current of Q_1 is not exactly PTAT. This is mainly caused by the temperature variation of the diffused resistors R_1, R_2 and R_3. The diode voltage V_{BE} is expanded in a Taylor series around T_o so that

$$V_{BE} \ = \ V_{BE}\Big|_{ideal} + HV_T - LV_T^2 + \ldots\ldots \tag{3.52}$$

where

$$H = T_o \frac{1}{R}\frac{dR}{dT}\Big|_{T_o} \quad \text{and} \quad L = \frac{T_o}{V_{T_o}}\frac{1}{R}\frac{dR}{dT}\Big|_{T_o} \ .$$

Therefore, neglecting the higher orders, the temperature coefficient error due to the bias current variation is

$$T.C. \ Error = -\frac{\frac{1}{R}\frac{dR}{dT}\Big|_{T_o} V_{T_o}}{V_{ref}} \approx -\frac{1000ppm/^{\circ}C \times 26mV}{1.26V} \tag{3.53}$$

$$\approx -21 \ ppm/^{\circ}C \ .$$

Unlike (3.51), this error is unidirectional and can be partly compensated by setting V_{ref} several mV higher than a theoretical value. However, a complete cancellation is impossible unless a curvature-compensation technique is employed.

3.5.3. Current Gain and Base Resistance

The last term of (3.50) is the most significant compared to the other contributions of β and r_b. The temperature coefficient error due to β and r_b is from (3.48) and (3.50)

$$T.C. \ Error = (1+\frac{R_2}{R_1})\frac{r_b I_2}{V_{ref}\beta_2}\left[\frac{1}{r_b}\frac{dr_b}{dT} + \frac{1}{I_2}\frac{dI_2}{dT} - \frac{1}{\beta_2}\frac{d\beta_2}{dT}\right] \tag{3.54}$$

$$\approx \frac{10\times 2k\Omega \times 30uA}{1.26V\times 150} \times (1000+3300-7000)ppm/^{\circ}C$$

$$\approx -8.6 \ ppm/^{\circ}C \ .$$

This error can also be compensated partly as in the case of (3.53).

3.5.4. Finite Op Amp Gain

If the op amp gain A_{op} is finite, the gain factor becomes

$$(1 + \frac{R_2}{R_1}) \frac{1}{1 + \frac{1}{A_{op}} (1 + \frac{R_2}{R_1})} \approx (1 + \frac{R_2}{R_1})[1 - \frac{1}{A_{op}} (1 + \frac{R_2}{R_1})] \tag{3.55}$$

for $A_{op} >> 1$. Therefore, the gain error due to the op amp finite gain is

$$G_\epsilon = - \frac{1}{A_{op}} (1 + \frac{R_2}{R_1}) \approx -0.022 \ \% \quad \text{for} \quad A_{op} \approx 50000 \ . \tag{3.56}$$

The effect of this gain error is hard to predict since the op amp gain A_{op} is also heavily dependent on temperature. However, the error attributable to the gain is negligible as far as the gain is big enough. For example, if A_{op} is independent of temperature, the temperature coefficient error due to G_ϵ is

$$T.C. \ Error = \frac{G_\epsilon (1 + \frac{R_2}{R_1}) \Delta V_{BE}}{V_{ref} T_o} \approx - \frac{0.022\% \times 10 \times 60 mV}{1.26 V \times 300^o K} \tag{3.57}$$

$$\approx -0.35 \ ppm/^oC \quad \text{at} \quad 300^o K \ .$$

3.5.5. Curvature Problem in the Bandgap Reference

As shown in Section 3.3, the emitter-base potential V_{BE} varies linearly with temperature approximately at a rate of $-1.79 mV/^oC$. This linear temperature variation can be completely compensated by a PTAT correction voltage. However, the higher-order temperature variation still exists. The deviation from the linear temperature dependence is the sum of the higher-order terms when V_{BE} is expanded in a Taylor series. That is, using (3.23), the temperature curvature $Cur[V_{BE}(T)]$ of V_{BE} is defined by

$$Cur[V_{BE}(T)] = V_{BE}(T) - [V_{BE}(T_o) + \frac{dV_{BE}}{dT}\bigg|_{T_o} (T-T_o)] \tag{3.58}$$

$$= Cur[V_g(T)] + V_T(\gamma-\alpha)(1 - \frac{T_o}{T} + \ln\frac{T_o}{T})$$

where the temperature curvature $Cur[V_g(T)]$ of the Si bandgap is from (3.13) and (3.14)

$$Cur[V_g(T)] = V_g - [V_g(T_o) + \frac{dV_g}{dT}\Big|_{T_o} (T-T_o)]$$ (3.59)

$$= -0.02745 + 1.83 \times 10^{-4} T - 3.05 \times 10^{-7} T^2 \quad \text{for} \quad 200^o K \leqslant T \leqslant 300^o K$$

$$= 0 \quad \text{for} \quad 300^o K \leqslant T \leqslant 400^o K \quad .$$

The maximum deviation occurs at $200^o K$. Estimating from this maximum deviation from the linear temperature dependence, the best-achievable temperature coefficient of the ideally first-order compensated bandgap reference is limited by

$$T.C. \quad Error = \frac{7.3mV}{1.26V \times 200^o K} \approx 29 \quad ppm/^o C \quad .$$ (3.60)

When the temperature variation of V_{BE} is under-compensated, the temperature curvature of the bandgap reference can be made symmetric around T_o. In this case, the temperature coefficient is minimized below the value predicted by (3.60). However, for the further reduction of the temperature coefficients, the higher-order temperature variation is to be compensated.

3.5.6. Others

Other errors are the differential temperature coefficient of the resistors R_1 and R_2, the finite β and the current mismatch, etc.. Typically, the temperature coefficients of resistors match each other within 1.2%, and the pnp transistors can be biased in the current range where β stays relatively constant and greater than 100. Therefore, these errors are negligible compared to the errors explained in the previous subsections.

3.6. Temperature Compensation Technique

Employing (3.22) and (3.52), the forward-biased emitter-base potential is given by [3], [50], [57]

$$V_{BE} = V_g - V_T[(\gamma-\alpha)\ln T - \ln EG] + HV_T - LV_T^2 \quad . \tag{3.61}$$

V_g is the Si bandgap which is a function of base doping [41] and temperature [48]. γ is the parameter which depends on the process and α represents the temperature dependence of the bias current. E and G are the parameters whose magnitudes are insignificant in the temperature analysis. H and L represent the bias current variation.

3.6.1. First-Order Temperature Compensation

After compensating the first-order temperature variation by adding a linear temperature correction voltage KV_T to V_{BE}, the V_{ref} is from (3.61)

$$\begin{aligned} V_{ref} &= V_{BE} + KV_T \\ &= V_g - V_T(\gamma-\alpha)\ln T + (K+H+\ln EG)V_T - LV_T^2 \quad . \end{aligned} \tag{3.62}$$

By equating the derivative of V_{ref} at T_o to zero and eliminating the unknown constants, we have

$$V_{ref} = V_g - \frac{dV_g}{dT}\bigg|_{T_o} T + V_T(\gamma-\alpha)(1+\ln\frac{T_o}{T}) - LV_T^2(1-2\frac{T_o}{T}) \quad . \tag{3.63}$$

The last term is due to the bias current variation and it will not disappear by the first-order temperature compensation. The first two terms are the nonlinearity of the Si bandgap. The voltage V_g can be expanded as follows.

$$V_g = V_g(T_o) + \frac{dV_g}{dT}\bigg|_{T_o}(T-T_o) + \frac{1}{2}\frac{d^2V_g}{dT^2}\bigg|_{T_o}(T-T_o)^2 + \dots\dots \quad . \tag{3.64}$$

Using (3.63) and (3.64), the voltage at T_o is

$$V_{ref}\big|_{T_o} = V_{g1} + V_{T_o}(\gamma-\alpha) + LV_{T_o}^2 \qquad (3.65)$$

where

$$V_{g1} = V_g(T_o) - \frac{dV_g}{dT}\bigg|_{T_o} T_o \approx 1.205 \quad V \quad . \qquad (3.66)$$

As commonly assumed, V_{g1} is the linearly-extrapolated Si bandgap voltage at $T=0^o K$. (3.65) shows that the bias current variation resulting from the temperature coefficient of resistors causes the nominal voltage different from the theoretical value.

3.6.2. Second-Order Temperature Compensation

The second-order temperature compensation requires two temperature correction voltages. The first one is a PTAT correction voltage KV_T compensates the linear temperature variation of V_{BE} and the second one is a PTATS correction voltage FV_T^2 which compensates the quadratic temperature variation of V_{BE}.

After adding two correction voltages, V_{ref} is from (3.61)

$$V_{ref} = V_{BE} + KV_T + FV_T^2$$
$$= V_g - V_T(\gamma-\alpha)\ln T + (K+H+\ln EG)V_T + (F-L)V_T^2 \quad . \qquad (3.67)$$

Similarly, by equating the first-order and the second-order derivatives of V_{ref} at T_o to zero, we obtain

$$K+H+\ln EG = (\gamma-\alpha)(1+\ln T_o) - \frac{T_o}{V_{T_o}} \frac{dV_g}{dT}\bigg|_{T_o} - 2(F-L)V_{T_o} \quad \text{and} \qquad (3.68)$$

$$F-L = \frac{1}{2}\frac{(\gamma-\alpha)}{V_{T_o}} - \frac{1}{2}\frac{T_o^2}{V_{T_o}^2}\frac{d^2V_g}{dT^2}\bigg|_{T_o} \quad . \qquad (3.69)$$

Therefore, from (3.67), (3.68) and (3.69),

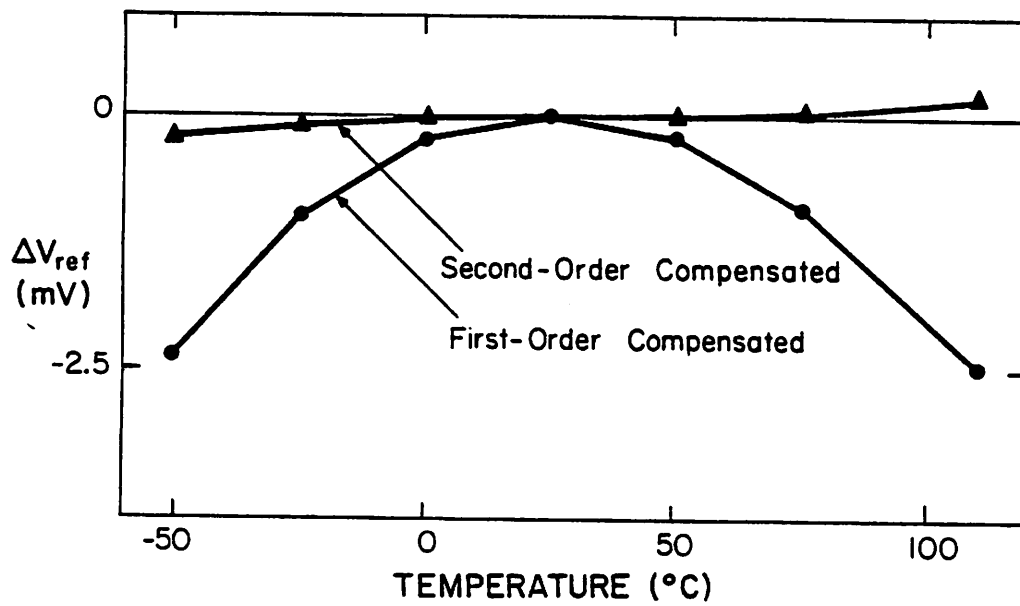$$V_{ref} = V_g - \frac{dV_g}{dT}\bigg|_{T_o}T - \frac{1}{2}\frac{d^2V_g}{dT^2}\bigg|_{T_o}T^2(1-2\frac{T_o}{T})$$

Fig. 3.4. Temperature variations of two ideally compensated bandgap
references when the Si bandgap curvature is neglected.

$$+ V_T(\gamma-\alpha)(\ln\frac{T_o}{T} +\frac{1}{2} \frac{T}{T_o}) \quad . \tag{3.70}$$

The nominal voltage at T_o is

$$V_{ref}|_{T_o} = V_{g2} + \frac{1}{2} V_{T_o}(\gamma-\alpha) \tag{3.71}$$

where

$$V_{g2} = V_g(T_o) - \frac{dV_g}{dT}\bigg|_{T_o} T_o + \frac{1}{2} \frac{d^2V_g}{dT^2}\bigg|_{T_o} T_o^2 \approx 1.179 \; V \quad . \tag{3.72}$$

Now V_{g2} is the quadratically-extrapolated Si bandgap voltage at $T=0^oK$. The voltage V_{g2} is more close to the Si bandgap at T_o than V_{g1} in (3.66). Approximately, $1.179V$ is the theoretical value of $V_g(0^oK)$ [48]. The typical temperature variations of (3.63) and (3.70) are compared in Fig. 3.4 neglecting the temperature curvature of V_g. Also note the difference of the nominal voltages at T_o in (3.65) and (3.71). The bias temperature variation no longer appears in (3.71) because it is compensated by the PTATS voltage.

After the PTAT voltage is added, the nominal voltage at T_o is

$$(V_{BE}+KV_T)\bigg|_{T_o} = V_{g2} + \frac{1}{2} \frac{d^2V_g}{dT^2}\bigg|_{T_o} T_o^2 - LV_{T_o}^2 \quad . \tag{3.73}$$

The nominal value of (3.73) after the first-order temperature compensation is quite different from that of the conventional approach given by (3.65) because (3.73) is the voltage at the intermediate step. The magnitude of the PTATS voltage FV_T^2 at T_o is obtained by the subtraction of (3.73) from (3.71) as

$$FV_T^2\bigg|_{T_o} = \frac{1}{2} V_{T_o}(\gamma-\alpha) - \frac{1}{2} \frac{d^2V_g}{dT^2}\bigg|_{T_o} T_o^2 + LV_{T_o}^2 \quad . \tag{3.74}$$

As expected, the correction voltages include the nonlinearity in the inherent Si bandgap reference curvature as well as the bias current variation. Other error sources neglected can be included in (3.74) easily as in the same manner as in the bias current variation.

CHAPTER 4

DESIGN OF A PRECISION CMOS BANDGAP REFERENCE

As discussed in Section 3.5, existing CMOS bandgap reference implementations suffer from many error sources unique to CMOS. In Sections 4.1 and 4.2, a systematic approach suitable for a CMOS bandgap reference is described which reduces the effects of those error sources. In Sections 4.3 and 4.4, the error and the noise analyses are given to predict the performance of the proposed CMOS bandgap reference.

4.1. A Precision CMOS Bandgap Reference

4.1.1. Curvature Compensation

The curvature compensation procedure is divided into two steps as illustrated in Fig. 4.1. The first step is to add a PTAT correction voltage KV_T to V_{BE}. After a PTAT correction voltage is added, V_{ref} will exhibit the second-order temperature variation as in Fig. 4.1(c). After a PTATS correction voltage FV_T^2 is added to that, A zero temperature coefficient at room temperature is achieved as shown in Fig. 4.1(d). Figure 4.2 is the overall schematic of the curvature-compensated CMOS bandgap reference which performs a curvature compensation as well as an offset-free amplification. The A_o is the fixed gain of the gain block determined by the capacitor ratio C_2/C_1. The current I_o is temperature-independent while I_T is PTAT. The current I_D is the bias current whose temperature variation should be well defined as a function of temperature.

If β and r_b effects are neglected, V_{ref} is given by

$$V_{ref} = V_{BE} + \frac{C_2}{C_1} \Delta V_{BE} \tag{4.1}$$
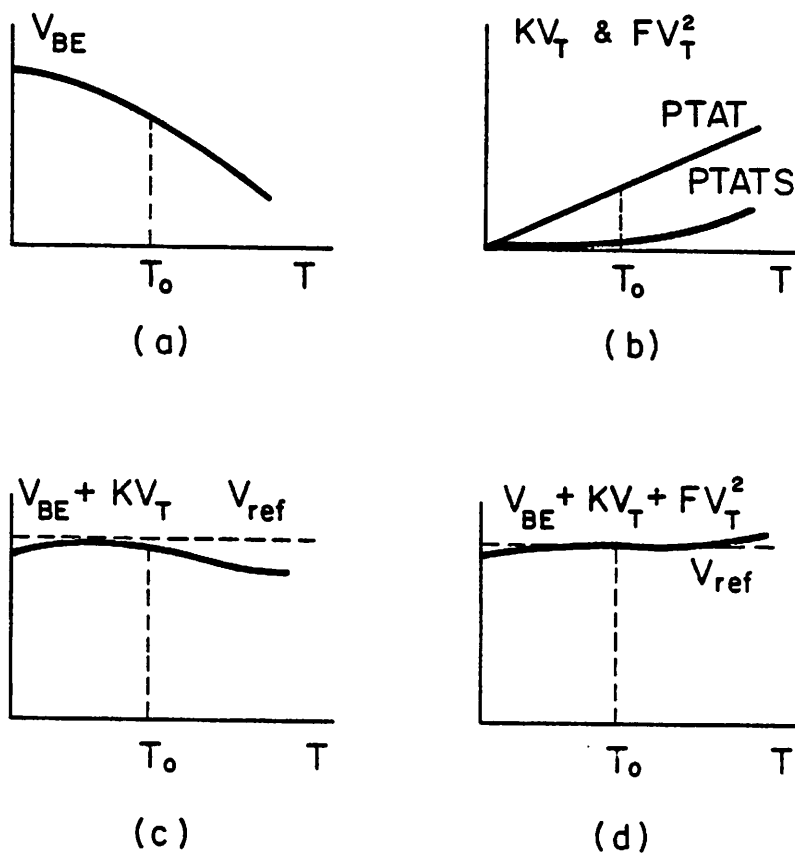
where

Fig. 4.1. Curvature compensation concept : (a) diode temperature variation, (b) a linear and a quadratic temperature correction voltages, (c) first-order compensated output voltage at the intermediate step and (d) the output achieving a zero temperature coefficient at room temperature.
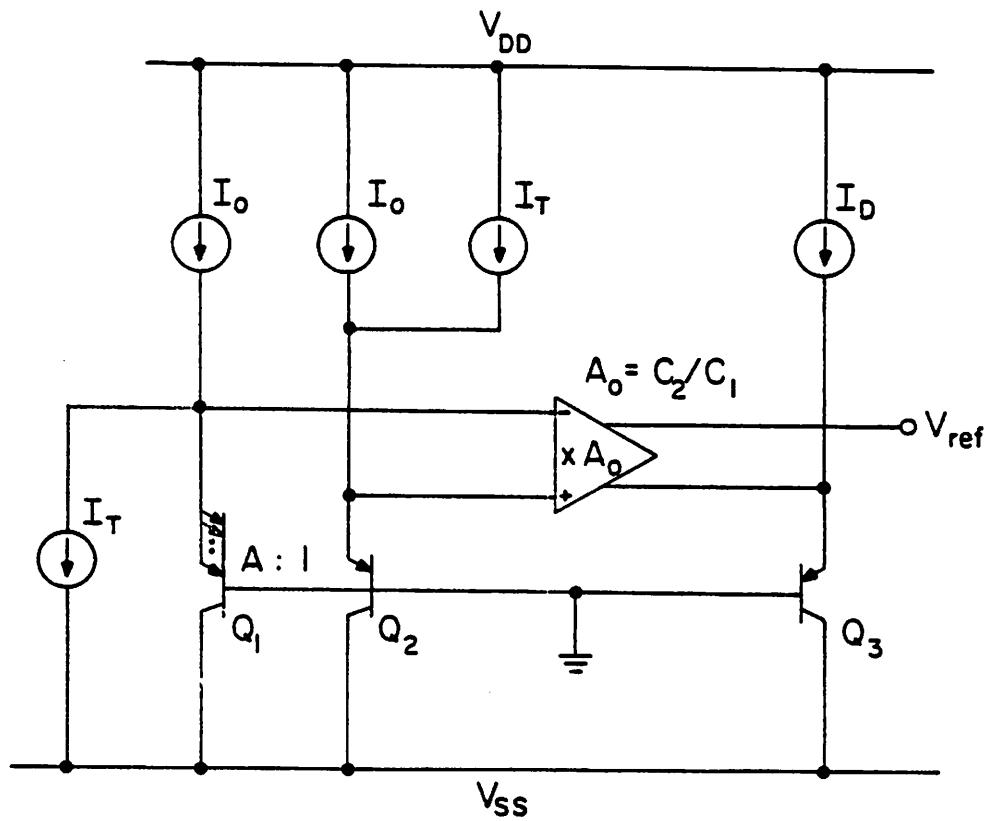
Fig. 4.2. Overall schematic of the curvature-compensated CMOS bandgap reference.

$$\Delta V_{BE} = V_T \ln A \frac{I_o + I_T}{I_o - I_T} = V_T \ln A + 2V_T(\frac{I_T}{I_o}) + \frac{2}{3} V_T(\frac{I_T}{I_o})^3 + \cdots\cdots\cdots . \qquad (4.2)$$

By trimming the ratio of two currents I_T/I_o, we can adjust the PTATS correction voltage independently of the PTAT correction voltage. The first-order compensated voltage $V_{BE} + KV_T$ is set by trimming C_2/C_1 and I_D with I_T disconnected at room temperature and the final V_{ref} is set by trimming I_T/I_o with I_T connected also at room temperature. The combined trims of the 6-bit binary-weighted capacitor array and the 6-bit resistor string insure a total of 12-bit trim accuracy in setting V_{ref} with comparison to a single 12-bit resistor trim in the existing designs.

4.1.2. Bias Current Generation

The circuit for generating the bias currents I_T and I_o is illustrated in Fig. 4.3. The current I_D is generated in the same manner in I_o. The stacked-cascode connection formed by M_1 to M_8 forces the emitter currents of Q_6 and Q_7 to be equal. The same scheme is also employed for the matching of the emitter currents of Q_9 and Q_{10}. The transistors M_9 and M_{10} start this self-bias circuit. As indicated, Q_4 and Q_6 have A times larger emitter areas than the remaining transistors. Therefore, the voltage developed across the resistor R_1 is PTAT and the current I_T through R_1 is also PTAT. The voltage V_o formed by Q_8 and R_2 is a first-order temperature compensated bandgap voltage. The temperature stability of V_o is not so critical because it only affects the PTATS voltage (not PTAT). The voltage V_o is developed across R_3 and the current I_o through R_3 is also independent of temperature. Therefore, the ratio I_T/I_o is easily trimmed by R_3. However, due to the mismatches of the $M_1 - M_2$ and $M_{11} - M_{12}$ transistor pairs, the voltage across R_1 and R_3 are not what they are supposed to be. Let the mismatches of V_{GS}'s of those MOS transistor pairs V_{m1}, V_{m2} and V_{m3}, etc.. Assuming current mirrors and pnp transistors are ideal, the currents I_T, I_o and I_D are given by
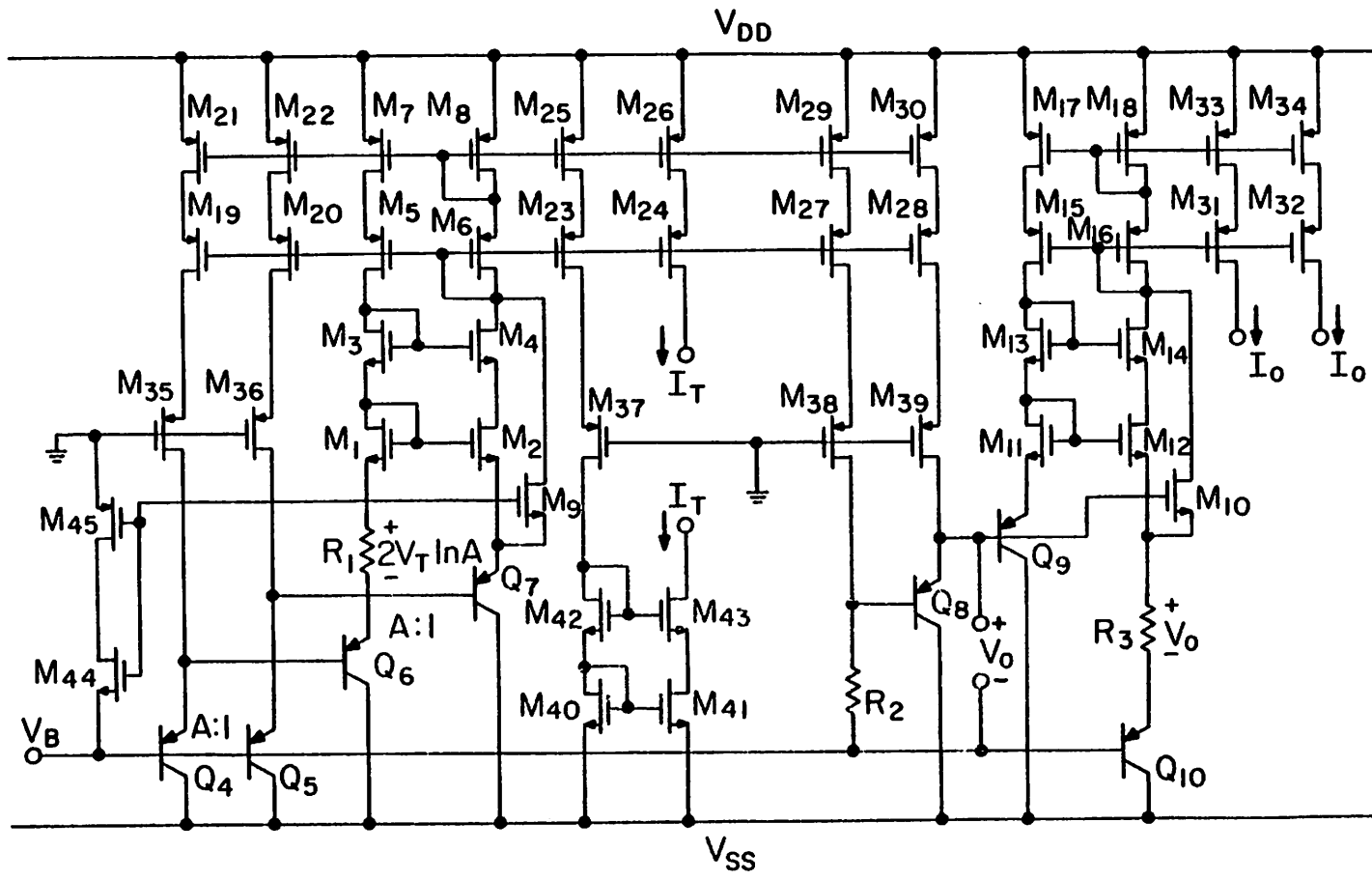
Fig. 4.3. Circuit for generating the FTAT current and the temperature-independent current.

$$I_T = \frac{1}{R_1}(2V_T \ln A + V_{m1}) \quad ,$$ (4.3)

$$I_o = \frac{1}{R_3}(V_o + V_{m2}) \quad \text{and}$$ (4.4)

$$I_D = \frac{1}{R_4}(V_o + V_{m3}) \quad .$$ (4.5)

Substituting (4.3), (4.4) and (4.5) into (4.1) and (4.2), and neglecting the higher orders, we obtain

$$V_{ref} \approx V_{BE}\Big|_{ideal} + \frac{C_2}{C_1}V_T \ln A + 4\frac{C_2}{C_1}\frac{R_3}{R_1}\ln\frac{A}{V_o}V_T^2 + V_\epsilon$$ (4.6)

$$= V_{BE}\Big|_{ideal} + KV_T + FV_T^2 + V_\epsilon$$

where

$$V_\epsilon = \frac{\rho}{100}\frac{1}{R}\frac{dR}{dT}\Big|_{T_o}(V_{T_o}-V_1)T + 2\frac{C_2R_3V_TV_{BE}}{C_1R_1V_o^2}V_{m1}$$ (4.7)

$$-4\frac{C_2R_3V_T^2\ln A}{C_1R_1V_o^2}V_{m2} + \frac{V_T}{V_o}V_{m3} \quad .$$

In (4.7), ρ represents the standard deviation (%) of the temperature coefficient of R_4 which is used for biasing V_{BE}.

To see the effects of the mismatch voltage, V_ϵ is approximated as in the following. Estimating from (3.73) and (3.74), we have

$$K \approx 24 \quad \text{and} \quad F \approx 72 \quad .$$ (4.8)

If $A \approx 10$, from the relation of

$$V_o = V_{BE} + 2\frac{R_2}{R_1}V_T \ln A \approx 1.256 \quad V,$$ (4.9)

we have

$$\frac{R_2}{R_1} \approx 5.48 \quad \text{for} \quad V_{BE} = 0.6 \ V \ . \tag{4.10}$$

Also, from (4.6), we obtain

$$\frac{C_2}{C_1} = \frac{KV_T}{V_T \ln A} = \frac{24}{\ln 10} \approx 10.4 \quad \text{and} \tag{4.11}$$

$$\frac{R_3}{R_1} = \frac{FV_T^2}{4\frac{C_2}{C_1}\frac{\ln A}{V_o}V_T^2} \approx 0.9 \ . \tag{4.12}$$

Therefore, from (4.7),

$$V_\epsilon = 0.19 V_{m1} - 0.04 V_{m2} + 0.02 V_{m3} \approx 1 \ mV \quad \text{at} \quad 300^o K \ . \tag{4.13}$$

The above equation indicates that the temperature coefficient uncertainty due to the mismatch voltages is 50 times smaller than the uncertainty resulting from the offset of the conventional bandgap reference given by (3.51). That is, in (3.51), the offset voltage is amplified by a factor of 10 while, in (4.13), it is reduced by a factor of 5. Thus, the proposed approach is 50 times less sensitive to the MOS transistor mismatch than the existing designs.

Table V summarizes the drawn dimensions and the bias conditions of the MOS transistors used in the bias circuit when the nominal values of the resistors are used. The current generation circuit of I_D is exactly the same as that of I_o except that the sizes of the transistors M_{11-14}, M_{15-18} and M_{31-34} are modified to $30\mu m/6\mu m$ to give a bias level of $30\mu m$ for the nominal value of R_4.

4.1.3. Offset-Free Amplification

For an offset-free amplification, the amplifier is divided into 2 stages as shown in Fig. 4.4. In the first offset storage mode, all the MOS switches are closed to sample the second offset voltages of the individual op amps. In the process of opening the MOS switches S_1 and S_2, the channel charges are injected into the op amp summing nodes to load the
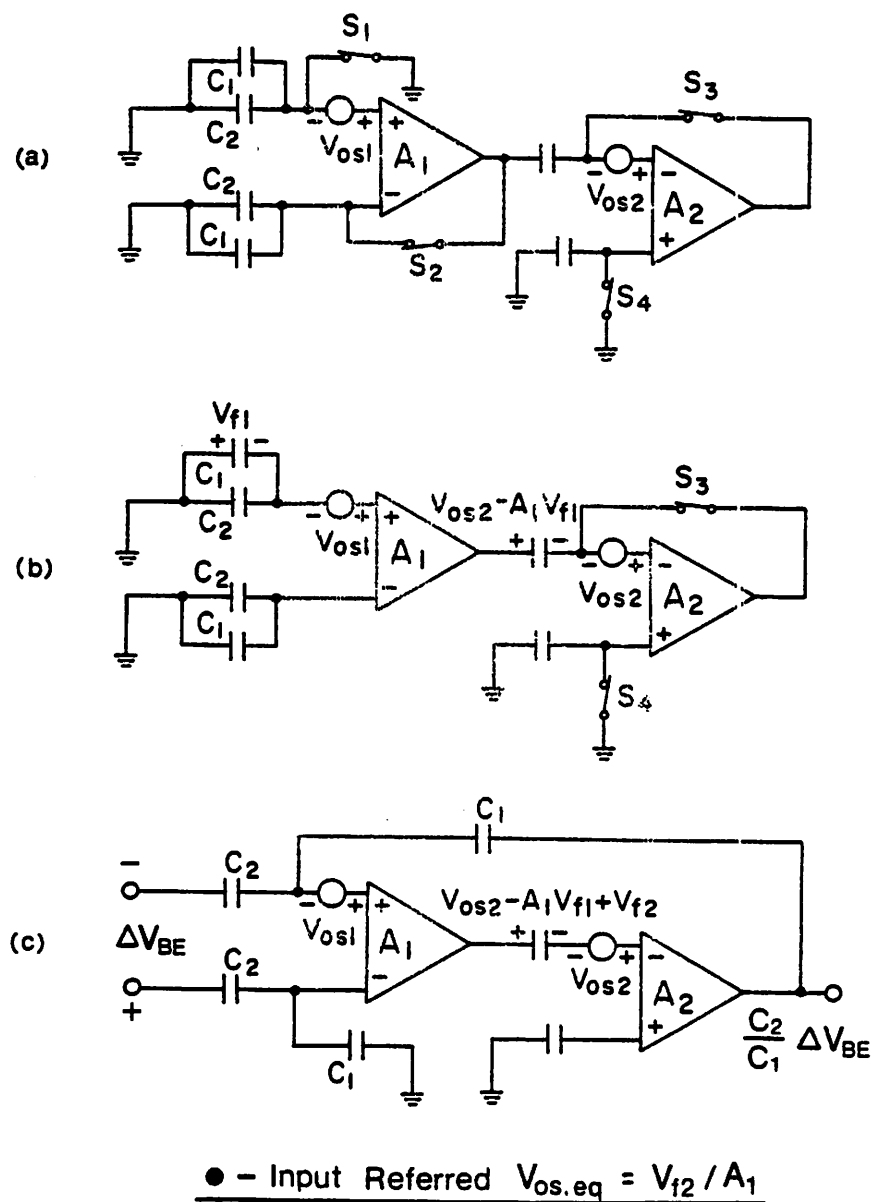
Fig. 4.4. Offset-free amplification : (a) first offset storage mode, (b) second offset storage mode and (c) final amplification mode.

TABLE V

DRAWN DIMENSIONS AND BIAS CONDITIONS
OF MOS TRANSISTORS IN THE BIAS CIRCUIT

Transistors	Widths(μm)	Length(μm)	Bias(μA)
M_{1-4}	18	6	9.12
M_{5-8}	24	6	9.12
M_{9-10}	6	6	0
M_{11-14}	78	6	89.5
M_{15-18}	123	6	89.5
M_{19-22}	24	6	9.11
M_{23-26}	24	6	9.11
M_{27-30}	33	6	12.5
M_{31-34}	123	6	89.5
M_{35-36}	24	6	9.11
M_{37}	24	6	9.11
M_{38-39}	33	6	12.5
M_{40-43}	18	6	9.11
M_{44}	72	6	20.1
M_{45}	6	78	20.1

capacitors C_1 and C_2. The charge injection differential voltage V_{f1} of the switches S_1 and S_2 is sampled across C_1 and C_2 along with the offset voltage V_{os1}. In the offset storage mode, the first gain stage charges the coupling capacitor C_c to compensate for the input differential voltage V_{f1}. After the switches S_3 and S_4 are opened, two stages are connected in a feedback amplification mode and the amplification of ΔV_{BE} takes place by the capacitor ratio C_2/C_1. When referred to input, the feed-thru difference of the switches S_3 and S_4 is reduced by the open-loop gain of the first stage.
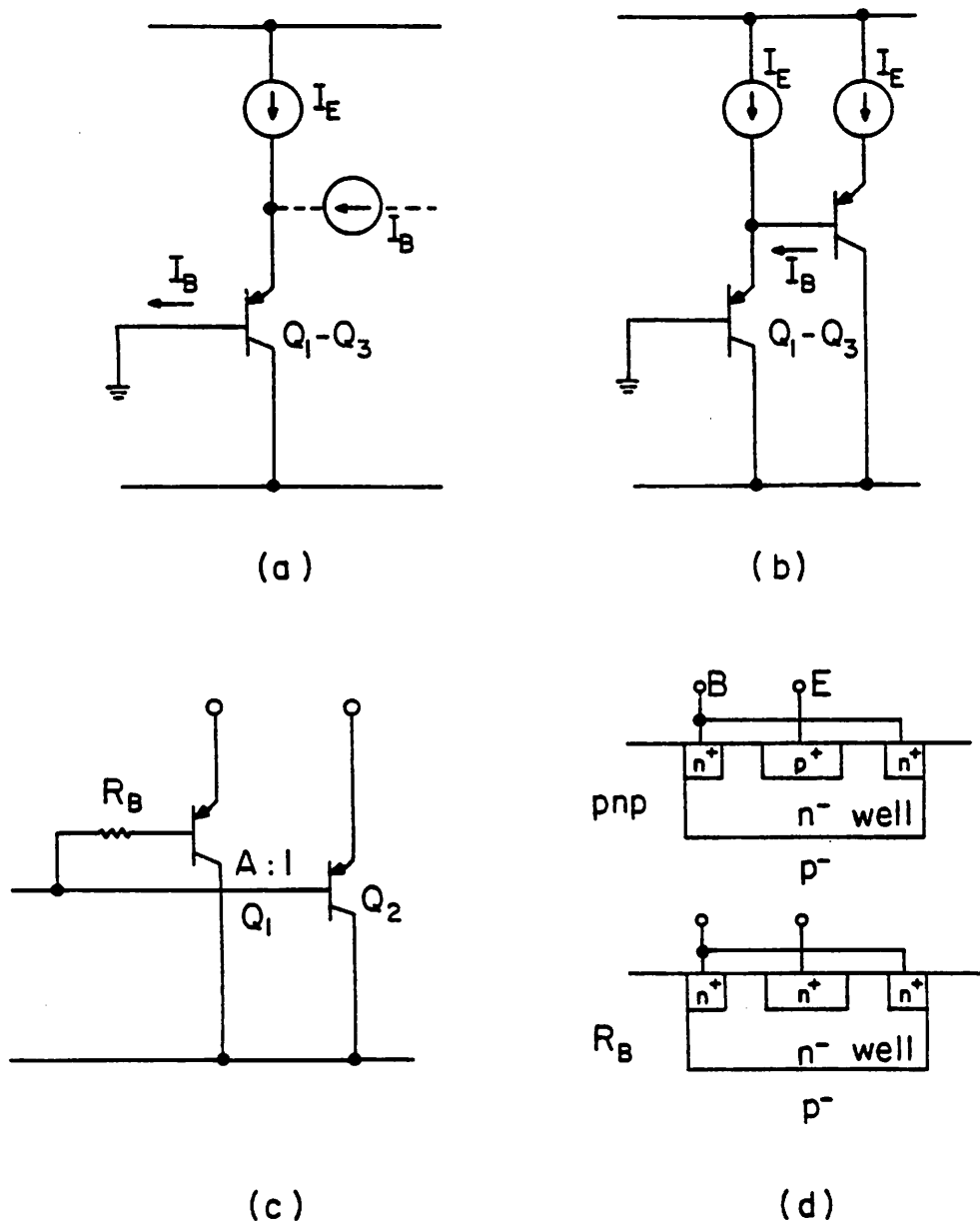
Fig. 4.5. Base current and base resistance cancellation : (a) base current returning, (b) replication, (c) extrinsic compensation resistance and (d) comparison between intrinsic base resistance and the extrinsic compensation resistance.

4.1.4. Base Current and Base Resistance Cancellation

To compensate for the difference between the collector current and the emitter current, the base current has to be returned to the emitter as shown in Fig. 4.5(a). However, a simpler approach is to replicate the base current and allow it to flow into the emitter as shown in Fig. 4.5(b). The base current is canceled up to 90% because the base currents can match each other within 10% for the adjacent transistors on a single chip.

Although it is difficult to cancel the intrinsic base resistance r_b exactly by the extrinsic compensation resistance, assume the extrinsic compensation resistance R_B is required for the compensation of the intrinsic base resistance r_b as shown in Fig. 4.5(c). Then, R_B should be

$$ R_B = (\frac{1+\beta_1}{1+\beta_2} - \frac{1}{A}) r_b . \tag{4.14} $$

The extrinsic compensation resistance R_B is made of the resistors which are formed using the same n^- well as the pnp transistor base with a different n^+ plug instead of the emitter p^+ plug as illustrated in Fig. 4.5(d). This extrinsic compensation resistance is supposed to track the temperature variation of the intrinsic base resistance. For the same geometry, the magnitude of the extrinsic compensation resistance is about 500Ω which is approximately one quarter of the measured intrinsic base resistance of pnp transistors.

4.2. OP Amp Design

The following requirements were considered in the design of two CMOS operational amplifiers. Two amplifiers are identical.

1. Moderate gain for each stage (100 to 300).

2. Single dominant pole per stage.
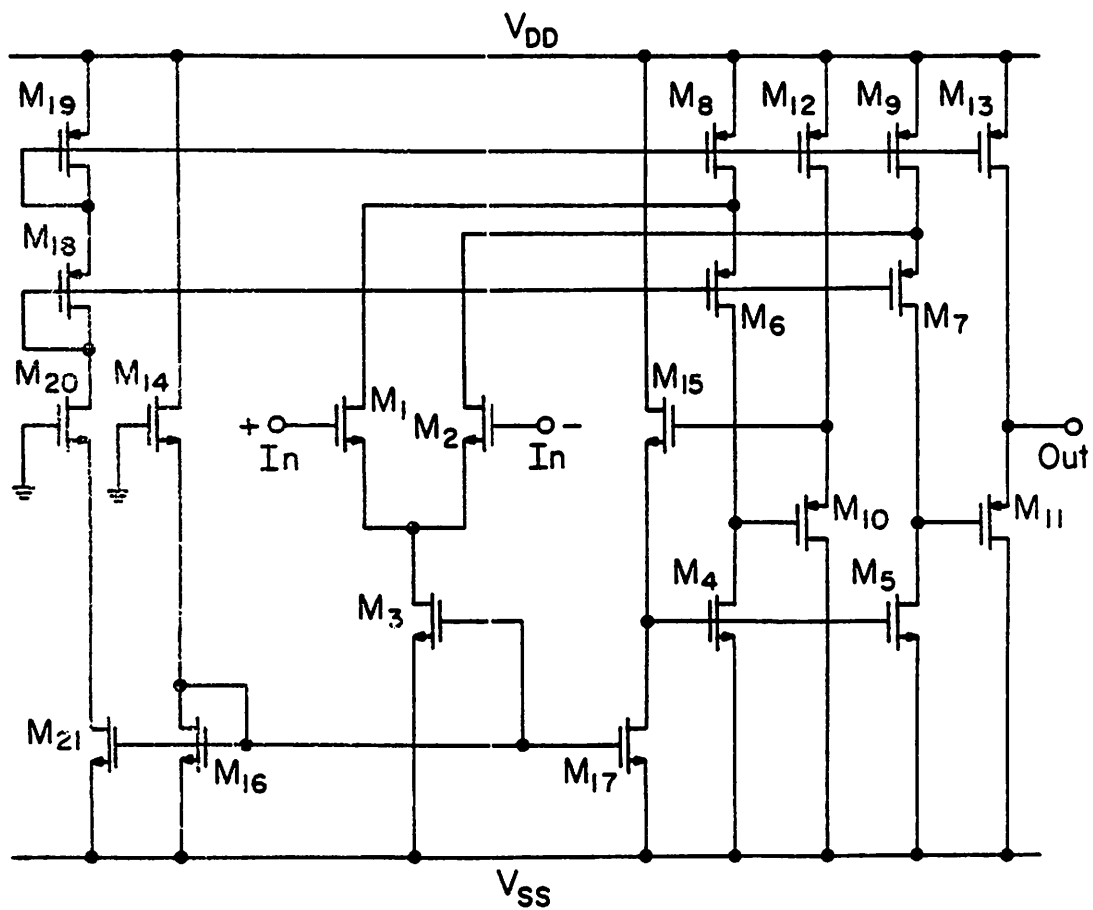
3. Inherent zero systematic offset voltage.

Fig. 4.6. Folded-cascode single-pole CMOS operational amplifier schematic.

4. Capacitor driving capability ($15pF$ for one stage and $100pF$ for two stages)

The configuration for two op amps is shown in Fig. 4.6. The transistors $M_{19}-M_{21}$ form a bias string for the amplifier. The replica bias circuit composed of $M_{14}-M_{17}$ performs a level shift and a differential to single-ended conversion, and reduces the inherent systematic offset. In order to limit the gain of each stage, the lower part composed of M_4-M_5 is not cascoded while the upper part M_6-M_9 is. The class-A source follower stages formed by $M_{10}-M_{13}$ are added to meet the capacitor driving capability.

Each amplifier has only one single dominant pole determined by the time constant by the output impedance of M_5 and the stray capacitance connected to the drain of M_5. Therefore, the frequency compensations of individual op amps are done by connecting the frequency compensation capacitors C_{c1} and C_{c2} between this high impedance nodes and the ac ground V_{SS} as shown in Fig. 4.7. For the frequency compensation of two stages, the

TABLE VI

SIMULATED AMPLIFIER PERFORMANCE
(-55 to 125°C)

A1 and A2 with $20pF$ load;

Gain	$48dB$
Freq (Gain = 1)	$3.9MHz$
Phase Margin	$65°$
Open-Loop Pole	$~500kHz$

Two Stages with $100pF$ load;

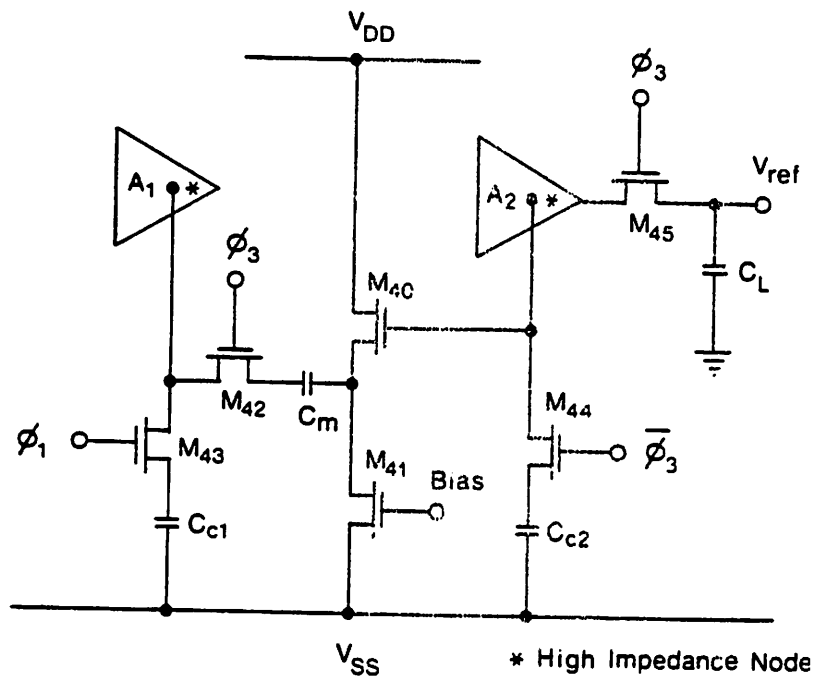|	
Gain	$97dB$
Freq (Gain = 10)	$650kHz$
Phase Margin	$65°$
Cyclic Time	$5\mu s$
Output Noise	$100\mu V$ ($500kHz$)

Fig. 4.7. Three-phase switched frequency compensation. The high impedance node is an internal node of each amplifier.

Miller capacitance C_m is returned from the high impedance node of the second stage to the same node of the first stage to achieve a pole-splitting compensation. To avoid the right-half plane zero, the source follower is used as shown in Fig. 4.7. SPICE simulated performances are summarized in Table VI. Table VII lists the drawn dimensions of the transistors and their bias conditions, and Table VIII lists the nominal capacitance values used.

4.3. Error Analysis

The error voltages in setting V_{ref} as expressed in (3.71) and (3.73) contribute a random spread of the temperature coefficient of the reference output V_{ref} since they are not deterministic and their magnitudes depend on the process. In the following subsections, the effects of the individual error sources on the temperature coefficients of V_{ref} will be discussed one by one. To make the analysis meaningful, we assume that the PTAT correction voltage KV_T and the PTATS correction voltage FV_T^2 are the ideal correction voltages which compensate the temperature variation of V_{ref} completely up to the second-order, and that another PTAT and PTATS correction voltages such as $K'V_T$ and $F'V_T^2$ are the actual correction voltages which result from the random errors in setting V_{ref} such as the temperature coefficient spread of resistors, the mismatch voltages and so on.

4.3.1. Spread of Resistor Temperature Coefficients

As explained in Section 3.6.2, the bias current variation due to the temperature coefficients of the diffused resistors used in the bias circuit can be compensated up to the second-order by the curvature compensation as far as the magnitude of the temperature coefficient is known. However, the temperature coefficient of the diffused resistors is not considered fixed, but rather has a production spread of

$$\frac{1}{R}\frac{dR}{dT}\bigg|_{T_o} = X\,(1\pm\frac{\rho}{100})\;\;ppm/^{\circ}C \tag{4.15}$$

TABLE VII

DRAWN DIMENSIONS AND BIAS CONDITIONS OF MOS TRANSISTORS IN THE OP AMP

Transistors	Widths(μm)	Length(μm)	Bias(μA)
M_{1-2}	252	6	26.1
M_{4-5}	48	12	26.2
M_{8-9}	24	6	52.3
M_{10}	48	6	59.6
M_{11}	328	6	408
M_{12}	24	6	59.6
M_{13}	164	6	408
M_{14-15}	6	48	12.1
M_{16-17}	24	12	12.1
M_{18}	24	6	26.1
M_{19}	12	6	26.1
M_{20}	252	6	26.1
M_{21}	48	12	26.1
M_{40}	224	6	287
M_{41}	224	6	287
M_{42}	336	6	0
M_{43-44}	63	6	0
M_{45}	366	6	0

TABLE VIII

CAPACITANCE VALUES

Capacitors	pF
C_1	0.65
C_2	5.2 - 10.4
C_c	4.3
C_{c1}	7.5
C_{c2}	5.5
C_m	5.5

where X is the average temperature coefficient of R and ρ is the standard deviation (%). Therefore, including the spread of resistor temperature coefficients, the emitter base potential V'_{BE} becomes

$$V'_{BE} = V_{BE} + \frac{\rho}{100}(HV_T - LV_T^2) + \ldots \quad . \tag{4.16}$$

where H and L defined in (3.52) represent the error occurred due to the spread of resistor temperature coefficients.

We correct V'_{BE} in the same way as V_{BE} because we are not affordable to perform an individual temperature trim of V_{ref} on every chip. Thus, in a room temperature trim, we make V_{ref} at T_o by adding a PTAT correction voltage such that

$$(V'_{BE} + K'V_T)\big|_{T_o} = (V_{BE} + KV_T)\big|_{T_o} \tag{4.17}$$

where the right side is the predetermined setting voltage after the first-order correction voltage is added as given by (3.73). After neglecting the higher orders, we obtain from (4.16) and (4.17)

$$K' = K - \frac{\rho}{100}H + \frac{\rho}{100}LV_{T_o} \quad . \tag{4.18}$$

For the optimum correction of V'_{BE}, however, $K'V_T$ should have been chosen such that

$$K' = K - \frac{\sigma}{100}H \quad . \tag{4.19}$$

Therefore, the error in setting the PTAT voltage is obtained by subtracting (4.19) from (4.18) so that

$$[K' - (K - \frac{\rho}{100}H)]V_T = \frac{\rho}{100}LV_{T_o}V_T \quad . \tag{4.20}$$

Similarly, for the second-order compensation, we make V_{ref} at T_o such that

$$(V'_{BE} + K'V_T + F'V_T^2)\Big|_{T_o} = (V_{BE} + KV_T + FV_T^2)\Big|_{T_o} \quad . \tag{4.21}$$

Then,

$$F' = F \quad . \tag{4.22}$$

However, F' should have been

$$F' = F + \frac{\rho}{100} L \tag{4.23}$$

Therefore, the error in setting the PTATS voltage is from (4.22) and (4.23)

$$[F' - (F + \frac{\rho}{100} L)] = -\frac{\rho}{100} LV_T^2 \quad . \tag{4.24}$$

By adding (4.20) and (4.24), the total correction voltage error $V_{\epsilon 1}$ is given by

$$V_{\epsilon 1} = \frac{\rho}{100} L(V_{T_o} - V_T)V_T \quad . \tag{4.25}$$

Therefore, the temperature coefficient error due to the production spread of the resistor temperature coefficient is

$$\frac{dV_{\epsilon 1}}{dT}\Big|_{T_o} = -\frac{\rho}{100} L\frac{V_{T_o}^2}{T_o} = -\frac{\rho}{100} V_{T_o} \frac{1}{R} \frac{dR}{dT}\Big|_{T_o} \tag{4.26}$$

$$\approx \pm\frac{10\%}{100} \times 26mV \times 1000ppm/^{\circ}C \approx \pm 2.6 \ \mu V/^{\circ}C \quad .$$

That is, for the 10% variation of the resistor temperature coefficients, we have only $2.6\mu V/^{\circ}C$ which approximately corresponds to $2ppm/^{\circ}C$.

4.3.2. Gate-Source Voltage Mismatches

The gate-source voltage mismatches V_{m1} and V_{m2} affect the output only when the second-order correction voltage is applied since they are included in the PTATS correction voltage, while V_{m3} appears all the time since V_{m3} appears in the bias current of V_{BE}. The same analysis as in the previous section applies. If a PTAT correction voltage is added to

satisfy (4.17), we have

$$K' = K - P \qquad (4.27)$$

where

$$P = \frac{1}{V_o} V_{m3} \quad .$$

If a PTATS correction voltage is added to satisfy (4.21), we have

$$F' = F - \frac{M}{V_{T_o}} + N \qquad (4.28)$$

where

$$M = 2\frac{C_2 R_3}{C_1 R_1 V_o} V_{m1} \quad \text{and}$$

$$N = 4\frac{C_2 R_2 R_3 \ln A}{C_1 R_1^2 V_o^2} V_{m1} + 4\frac{C_2 R_3 \ln A}{C_1 R_1 V_o^2} V_{m2} \quad .$$

For the optimum compensation, K' and F' should have been chosen as

$$K' = K - M - P \quad \text{and} \qquad (4.29)$$

$$F' = F + N \quad .$$

From (4.27), (4.28) and (4.29), the correction voltage error becomes

$$V_{e2} = [K' - (K - M - P)]V_T + [F' - (F + N)]V_T^2 \qquad (4.30)$$

$$= MV_{T_o}(V_{T_o} - V_T)V_T \quad .$$

Therefore, the temperature coefficient error due to the random gate-source voltage mismatches is

$$\left.\frac{dV_{e2}}{dT}\right|_{T_o} = M\frac{V_{T_o}}{T_o} = 2\frac{C_2 R_3 V_{T_o}}{C_1 R_1 V_o T_o} V_{m1} \qquad (4.31)$$

$$= \pm 2 \times \frac{10.4 \times 0.9 \times 26mV \times 5mV}{1.2 \times 300^o K}$$

$$\approx \pm 6.8 \ \mu V/^{\circ}C \quad .$$

where (4.11) and (4.12) are employed in the approximation.

From (4.26) and (4.31), the statistical sum of two independent temperature coefficient errors is given by

$$\frac{dV_{\epsilon}}{dT}\bigg|_{T_o} = \left(\left| \frac{dV_{\epsilon 1}}{dT} \right|^2_{T_o} + \left| \frac{dV_{\epsilon 2}}{dT} \right|^2_{T_o} \right)^{1/2} \tag{4.32}$$

$$\approx (2.6^2 + 6.8^2)^{1/2} \approx 7.3 \ \mu V/^{\circ}C \quad .$$

This temperature coefficient error results from the 10% random spread of resistor temperature coefficients and the $5mV$ random gate-source voltage mismatch of two identical adjacent MOS transistors. $7.3 \mu V/^{\circ}C$ approximately corresponds to $6.1 ppm/^{\circ}C$. This level of the production spread in the reference output temperature coefficients is to be expected.

4.3.3. Leakage from the Op Amp Summing Node

The source or drain of MOS switchs is to be connected to the op amp summing node so as to charge the capacitive node periodically. A leakage current I_l through a reverse-biased source or drain junction connected to the summing node contributes V_{out} to drift up as illustrated in Fig. 4.8. This leakage current heavily depends on process and exhibits an exponential temperature drift. In the circuit shown in Fig. 4.8, the following relation holds :

$$I_l = - C_2 \frac{\Delta V_s}{\Delta T} + C_1 (\frac{\Delta V_{out}}{\Delta t} - \frac{\Delta V_s}{\Delta t}) \tag{4.33}$$

where the summing node voltage change ΔV_s is related to the change of V_{out} by

$$\Delta V_s = - \frac{C_1}{(C_1 + C_2)} \Delta V_{out} \quad . \tag{4.34}$$

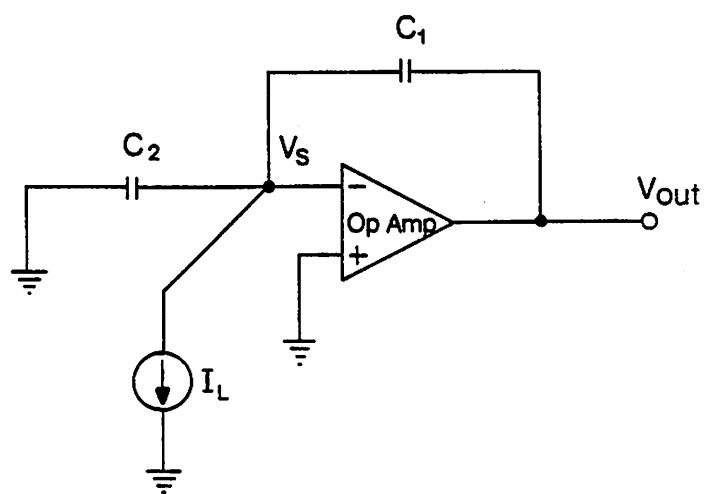Therefore, the time rate of the output voltage change is

Fig. 4.8. Junction leakage current from the op amp summing node.

$$\frac{\Delta V_{out}}{\Delta t} = \frac{I_l}{2C_1} \quad .$$

(4.35)

The leakage current in the reverse biased junction is mainly due to the space-charge generation current although the injected minority carrier still exists. The generation current in the space charge region of width, $X_s = 0.5\mu m$, and area, $A_j = 664\mu m^2$, is approximately [58]

$$I_l(300^\circ K) = \frac{qA_j X_s n_i}{2\tau_o} \approx \frac{1.6\times10^{-19}\times664\mu m^2\times0.5\mu m\times1.4\times10^{10}}{2\times10^{-6}}$$

(4.36)

$$\approx 3.7\times10^{-13} \, A \quad .$$

where τ_o is the excess carrier life time. Assuming X_s and τ_l are constant with temperature, the junction leakage current $I_l(T)$ at T with respect to $I_l(T_o)$ at the reference temperature T_o is from (AII.16)

$$I_l(T) = I_l(T_o)\left[\frac{T}{T_o}\right]^{3/2} \exp[\frac{1.21}{2kT_o}(1 + \frac{T_o}{T})] \quad .$$

(4.37)

$$\approx I_l(T_o)\left[\frac{T}{T_o}\right]^{3/2} \exp[23(1- \frac{T_o}{T})]$$

For example, at $400^\circ K$, the leakage current in the area of $664\mu m^2$ is from (4.36) and (4.37)

$$I_l(400^\circ K) \approx 484 \, I_l(300^\circ K) \approx 0.18 \, nA \quad .$$

(4.38)

Then, the time rate change of V_{out} expressed by (4.35) is

$$\frac{I_l(400^\circ K)}{2C_1} = \frac{0.18\times10^{-9}}{2\times0.65\times10^{-12}} \approx 138 \, V/sec \quad \text{at} \quad 400^\circ K \quad .$$

(4.39)

For $10\mu sec$ cyclic time, the maximum departure of V_{ref} during one full cycle time is

$$\frac{I_l(400^\circ K)}{2C_1} \times 10\mu sec = 138V/sec\times10\mu sec = 1.38 \, mV \quad .$$

(4.40)

The temperature error due to leakage current is

$$T.C. \quad Error \quad = \quad \frac{1.38mV}{1.2V \times 200^{\circ}K} \quad = \quad 5.3 \quad ppm/^{\circ}K \quad .$$ (4.41)

In the actual circuit, however, part of this error is canceled because the amplifier is in a differential configuration. Even though I_l is heavily dependent upon the process, the typical leakage current is on the order of $10^{-4}A/m^2$ level which is much lower than the estimation of (4.36). Therefore, the temperature coefficient error due to this level of leakage is virtually negligible.

4.3.4. Bias Current Instability

The bias current I_D of V_{BE} which is generated in the same manner as I_o in Fig.4.3 is from (4.15)

$$I_D \quad = \quad \frac{1}{R_4} \quad (V_o + V_{m3}) \quad .$$ (4.42)

The process spread of R_4 temperature coefficient makes the spread of the output V_{ref} as analyzed in Section 4.3.1. Since the temperature variation of V_o is random, the drift of V_o including V_{m3} will affect the reference output V_{ref} directly through the relation of

$$V_{BE} \quad = \quad V_T \ln \frac{(V_o + V_{m3})}{I_s R_4} \quad .$$ (4.43)

Assuming V_{m3} is constant as discussed in Section 3.4.3, the temperature coefficient error of V_{ref} due to the temperature variation of V_o is from (4.43)

$$T.C. \quad Error \quad = \quad \frac{V_T}{V_o + V_{m3}} \frac{1}{V_o} \frac{dV_o}{dT}$$ (4.44)

$$= \quad \frac{26mV}{1.256V} \frac{1}{V_o} \frac{dV_o}{dT}\Big|_{T_o} \quad \approx \quad 48.3 \quad \frac{1}{V_o} \frac{dV_o}{dT}\Big|_{T_o} \quad at \quad 300^{\circ}K \quad .$$

The instability of V_o appears at the reference output reduced by a factor of 48.3. For example, the $100ppm/^{\circ}C$ temperature drift of V_o makes V_{ref} drift at a rate of

approximately $2ppm/^{\circ}C$.

4.3.5. Others

Besides the errors discussed in the previous subsections, errors may come from the finite op amp gain, the differential temperature coefficient of capacitors and the uncanceled β and r_b, etc.. As far as these error sources are constant, they are actually canceled out because the temperature variation can be compensated up to the second order. The production spreads of these error sources directly appear as a spread of the reference output temperature coefficient.

The error involved in the trim procedure is the absolute error in setting the reference output voltage. The output trim can be performed with $\pm0.3mV$ accuracy. For example, $0.3mV$ PTAT missetting gives rise to approximately $1.2ppm/^{\circ}C$ in the worst case. When statistically summing all the error sources discussed in Section 4.3, the production spread of the reference output temperature coefficient is less than $10ppm/^{\circ}C$ theoretically.

4.4. NOISE ANALYSIS

The primary merit of a correlated-double sampling is to remove the switching transient and the switch reset noise such as found in charge sensing circuits. This feature is desirable in a low frequency signal processing because $1/f$ noise density at low frequency is effectively eliminated in the correlated-double sampling process. However, the application of this technique in high-speed data-acquisition systems results in the increase of a noise spectral density. The reason is that the cut-off frequency of the chopper-stabilized amplifier or the comparator in data-acquisition systems is usually much higher than the noise sample-subtraction cycle. As a consequence of aliasing, the white thermal noise is doubled and the $1/f$ noise remains at the same level for the typical ratio of the clock frequency to the cut-off frequency.
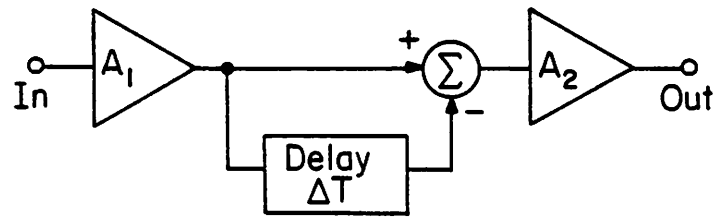
(a)
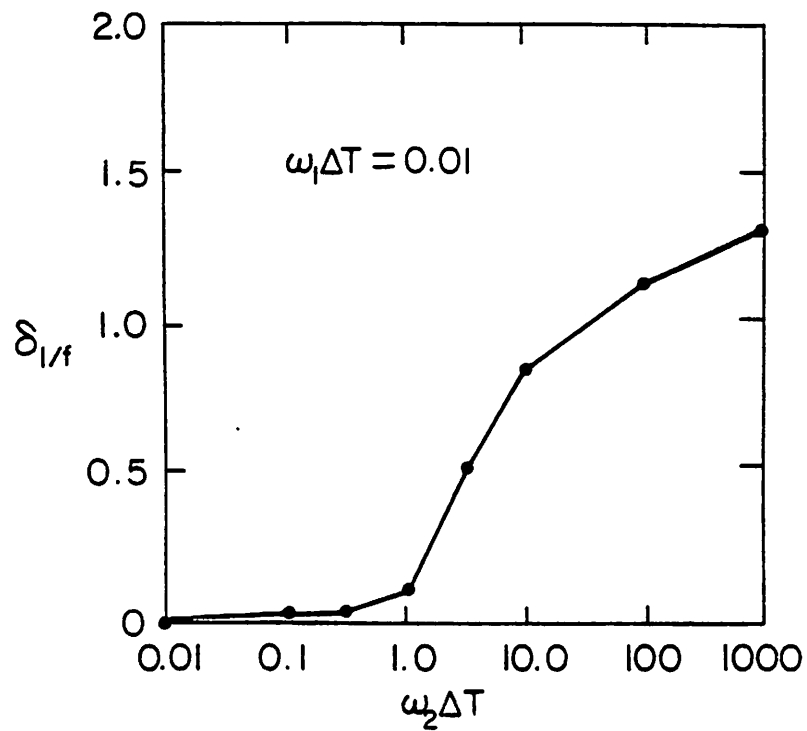
(b)

Fig. 4.9. (a) Correlated double sampling (CDS). (b) The scale
factor of 1/f noise after CDS.

4.4.1. Effects of a Correlated-Double Sampling

Fig. 4.9(a) shows the circuit equivalence of the correlated-double sampling scheme. Assuming the extremely fast unity-gain input sense stage A_1, the output spectral density [40], [59]-[60]

$$S_o(\omega) = 2(1 - \cos\omega\Delta T)S_i(\omega) \tag{4.45}$$

where ΔT is the time delay between the offset sample and the output sample. The output spectral density at the output is nulled at every multiple frequency of $1/\Delta T$.

The white noise variance at the output in the effective bandwidth ω_2 of A_2 is

$$\overline{v_o^2} = \frac{\delta_{white}}{2\pi} \int_0^{\omega_2} S_i(\omega)\,d\omega \tag{4.46}$$

where δ_{white} is the white noise scale factor associated with the correlated-double sampling and is given by

$$\delta_{white} = \frac{1}{\omega_2} \int_0^{\omega_2} 2(1 - \cos\omega\Delta T)\,d\omega \approx 2 \quad \text{for} \quad \omega_2\Delta T >> 1 \quad . \tag{4.47}$$

The doubled white thermal noise is due to the subtraction of two uncorrelated noises. Similarly, the $1/f$ noise variance at the output is

$$\overline{v_o^2} = \frac{\delta_{1/f}}{2\pi} \int_{\omega_1}^{\omega_2} S_i(\omega)\,d\omega \tag{4.48}$$

where the low frequency limit is taken as ω_1 instead of zero and the $1/f$ noise scale factor associated with the correlated-double sampling is

$$\delta_{1/f} = \frac{\displaystyle\int_{\omega_1}^{\omega_2} \frac{2(1 - \cos\omega\Delta T)}{\omega}\,d\omega}{\displaystyle\int_{\omega_1}^{\omega_2} \frac{1}{\omega}\,d\omega} \quad . \tag{4.49}$$

The scale factor $\delta_{1/f}$ versus ω_2 is illustrated in Fig. 4.9(b). If A_2 bandlimits the noise up
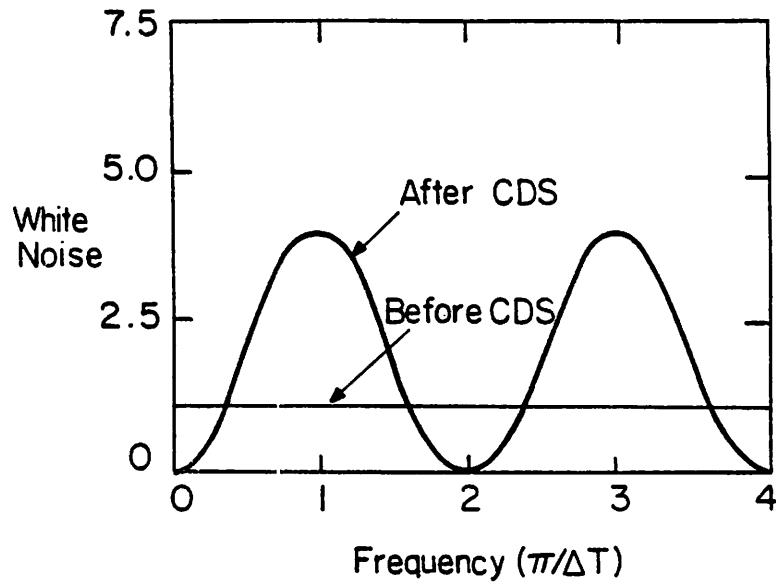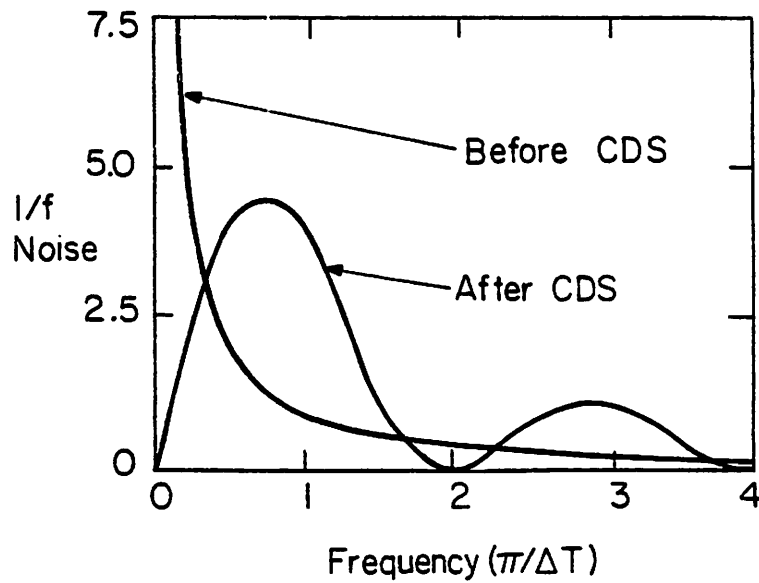
Fig. 4.10. Effect of correlated double sampling (CDS) on the op amp white thermal noise (a) and 1/f noise (b).

to $\omega_2 \Delta T = 1.0$, 90% of the $1/f$ noise is canceled while the increased bandwidth, for example $\omega_2 \Delta T = 10$, reduces the $1/f$ noise only by 15%. If the $1/f$ noise corner frequency where the $1/f$ noise equals the white thermal noise is less than $1/4\Delta T$, the $1/f$ noise is negligible compared to the thermal noise after the correlated-double sampling. This corner frequency is heavily dependent upon the process. Fig. 4.10 illustrates the effects of the correlated-double sampling on the white noise and the $1/f$ noise of the op amp as indicated by (4.45).

4.4.2. Op Amp Thermal Noise and $1/f$ Noise

Noises associated with MOS transistors can be divided into the following two. They are :

1. Thermal Noise ; Thermal noise of MOS devices arises because of the finite channel resistance. The input-referred spectral density is

$$\frac{\overline{v_{eq}^2}}{\Delta f} = 4kT(\frac{2}{3} \frac{1}{g_m})$$

(4.50)

where g_m is the transconductance of the device. This noise is white.

2. Flicker Noise (1/f Noise) ; Flicker Noise is associated with the presence of extra electron energy states at the Si-Si dioxide interface. The slow charging and discharging of these surface states give rise to the noise concentrated at low frequency. The input-referred spectral density is

$$\frac{\overline{v_{eq}^2}}{\Delta f} = \frac{K_1}{C_{OX} WL} \frac{1}{f}$$

(4.51)

where K_1 is a constant representing the particular process.

When the noise contributions of each transistors in op amps are considered (Fig. 4.6), the equivalent input noise of op amps A_1 and A_2 is

$$\overline{v^2_{vq1-2}} = 2\overline{v^2_{m1-2}} + 2\left(\frac{g_{m4-5}}{g_{m1-2}}\right)^2 \overline{v^2_{m4-5}} + 2\left(\frac{g_{m6-7}}{g_{m1-2}}\right)^2 \overline{v^2_{m6-7}} + 2\left(\frac{g_{m8-9}}{g_{m1-2}}\right)^2 \overline{v^2_{m6-7}}$$

$$\approx 2\overline{v^2_{m1-2}} \qquad\qquad (4.52)$$

where $\overline{v^2_{mi}}$, ($i=1,2,\bar{},7$) is the equivalent noise of the transistor M_i and g_{mi} is its transconductance. To reduce the thermal noise of M_{1-2}, relatively big devices ($252\mu m/6\mu m$) were used for M_1 and M_2. Since the g_{m1-2} are much bigger than the rest of the transistors, other contributions can be neglected in the noise estimation. For this geometry, the total equivalent input-referred noise resistance of M_{1-2} is

$$R_{m1-2} = \frac{2}{3}\frac{1}{g_{m1-2}} = \frac{2}{3}\frac{1}{(2\mu C_{ox}\frac{W}{L}I_{D1-2})^{\frac{1}{2}}}$$

$$\approx \frac{2}{3}\times\frac{1}{(2\times35\times10^{-6}\times\frac{256}{6}\times26\times10^{-6})^{\frac{1}{2}}} \approx 1.9 \ k\Omega \ . \qquad (4.53)$$

4.4.3. Noise of Substrate PNP Transistors

When the input ΔV_{BE} is amplified in Fig. 4.2 and 4.4, the capacitor C_2 is connected to the emitter of the substrate pnp transistor, and thus the equivalent noise voltage at the emitter is also amplified by a gain factor of C_2/C_1 and band-limited by the gain stage. There are four kinds of noises in bipolar transistors except the avalanche noise which accompanies Zener breakdown. They are

1. Shot Noise : Shot noise is associated with a direct current across the junctions of diodes or transistors. The spectral density is

$$\frac{\overline{i^2}}{\Delta f} = 2qI \qquad\qquad (4.54)$$

where I is a direct current through a junction. The corner frequency of this noise is so high that it can be assumed white in the frequency range of interest.

2. Thermal Noise : The spectral density of the thermal noise in series with a resistor R is

$$\frac{\overline{v^2}}{\Delta f} = 4kTR \quad . \tag{4.55}$$

3. Flicker Noise $(1/f)$: Flicker noise is mainly caused by traps associated with contamination and crystal defects in the emitter-base depletion layer. It is also associated with a flow of direct current. The spectral density is

$$\frac{\overline{i^2}}{\Delta f} = K_2 \frac{I^b}{f} \Delta f \tag{4.56}$$

where K_2 is a constant for a given process, $b \approx 0.5 \tilde{} 2$.

4. Burst Noise : Burst noise is associated with heavy metal contamination and also related to a direct current. The spectral density is

$$\frac{\overline{i^2}}{\Delta f} = K_3 \frac{I^c}{1 + (f/f_c)^2} \tag{4.57}$$

where K_3 is a constant, I is a direct current, $c \approx 0.5 \tilde{} 2$ and f_c is a particular frequency for a given process.

Including all the above noise sources, a pnp transistor is shown with three noise sources in Fig. 4.11(a). Their spectral densities are

$$\frac{\overline{v_b^2}}{\Delta f} = 4kTr_b \quad , \tag{4.58}$$

$$\frac{\overline{i_c^2}}{\Delta f} = 2qI_C \quad \text{and} \tag{4.59}$$

$$\frac{\overline{i_b^2}}{\Delta f} = 2qI_B + K_2 \frac{I_B^b}{f} + K_3 \frac{I_B^c}{1 + (f/f_c)^2} \quad . \tag{4.60}$$

These three noise sources can be replaced with two input-referred noise sources $\overline{v_i^2}$ and $\overline{i_i^2}$ as shown in Fig. 4.11(b).
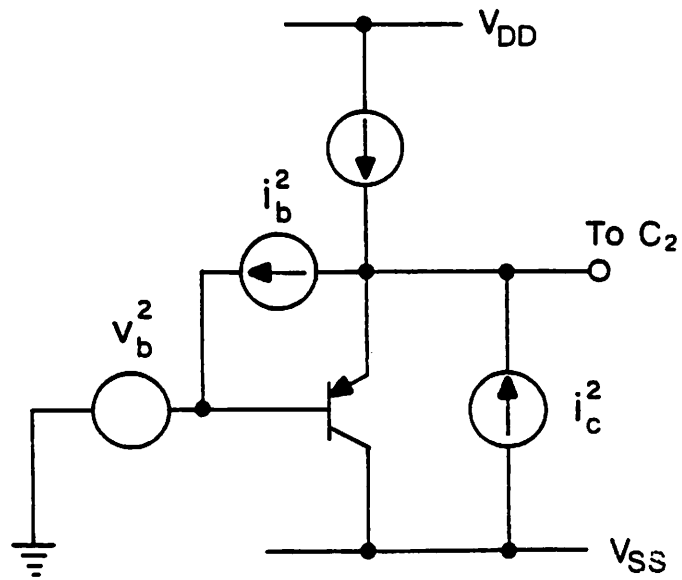
(a)
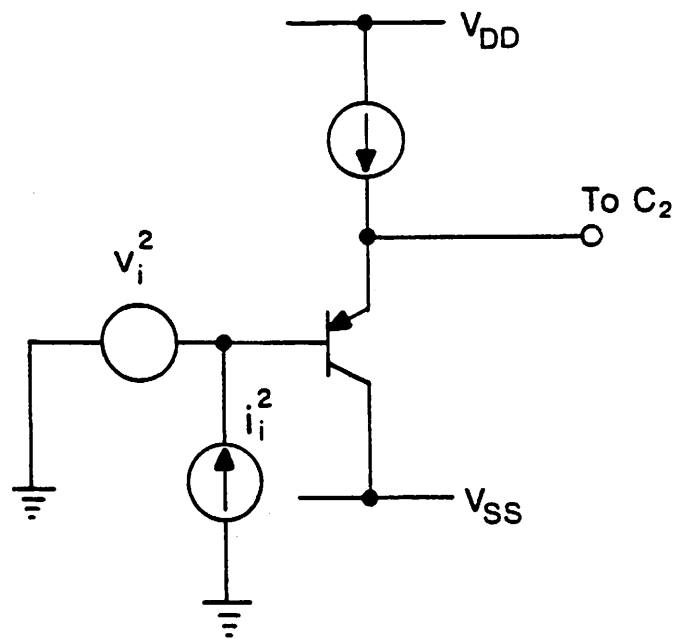
(b)

Fig. 4.11. (a) Noise sources of pnp transistors. (b) Input-
referred equivalent noises.

$$\overline{v_i^2} = \overline{v_b^2} + \frac{\overline{i_c^2}}{g_m^2} \qquad \text{and} \tag{4.61}$$

$$\overline{i_i^2} = \overline{i_b^2} + \frac{\overline{i_c^2}}{|\beta(j\omega)|^2} \quad . \tag{4.62}$$

For the circuit of Fig. 4.11(b), the dominant source is $\overline{v_i^2}$ because the base terminal is grounded. From (4.58), (4.59) and (4.61), the equivalent input noise spectral density is given by

$$\frac{\overline{v_i^2}}{\Delta f} = 4kTR_{eq} \tag{4.63}$$

where the equivalent input noise resistance R_{eq} is

$$R_{eq} = r_b + \frac{1}{2g_m} \quad . \tag{4.64}$$

This noise source is directly connected to the input of the gain stage through the emitter follower. The input-equivalent noise resistance R_{eq} is estimated to be

$$R_{eq} \approx 1.5k\Omega + \frac{26mV}{2\times30\mu A} \approx 2 \ k\Omega \quad . \tag{4.65}$$

4.4.4. Output Noise and its Spectral Density

At the output of the proposed reference, there appear two kinds of noises. One is the direct noise while the output sync pulse stays high and the other is the sampled-held noise at the load capacitor while the output sync pulse is low. When the sync pulse is high, the reference is represented equivalently as in Fig. 4.12(a) after replacing all the noise sources with the ideal noise generators. the resistors R_{c1}, R_{c2} and R_{c3} represent the on resistances of the MOS switches for the capacitors C_1, C_2 and C_3, respectively. Their thermal noise densities are $\overline{v_{n1}^2}/\Delta f$, $\overline{v_{n2}^2}/\Delta f$ and $\overline{v_{n3}^2}/\Delta f$, respectively, and $\overline{v_{eq}^2}/\Delta f$ is the equivalent input noise of the op amp which has a gain of A_{op} and a dominant pole at p_o. Among these, the op amp input noise and the pnp transistor noise are dominant in magni-
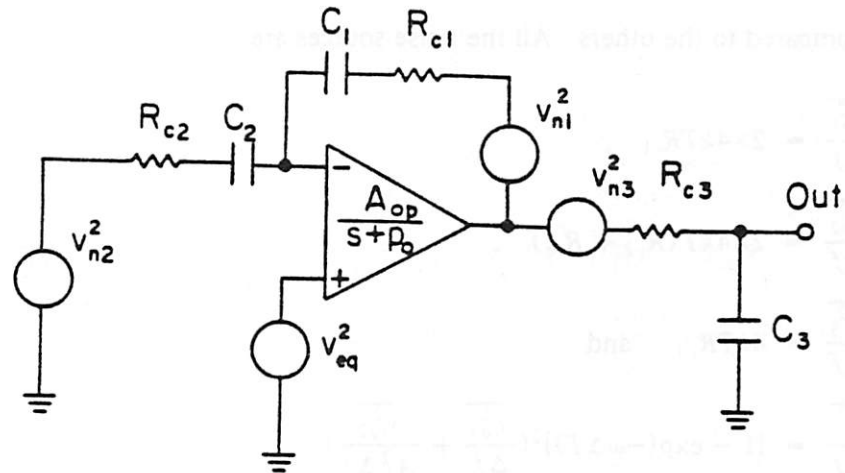
(a)
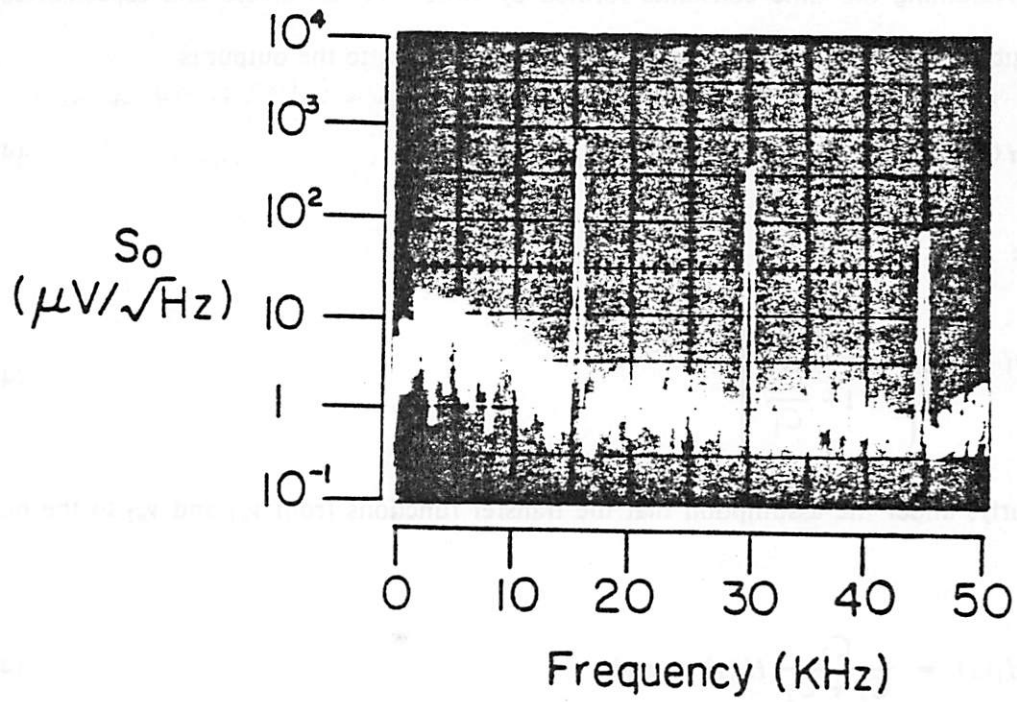
(b)

Fig. 4.12. Simplified noise equivalent circuit of the reference
(a) and measured noise spectral density (b).

tude compared to the others. All the noise sources are

$$\frac{\overline{v_{n1}^2}}{\Delta f} = 2 \times 4kTR_{c1} \quad ,$$

(4.66)

$$\frac{\overline{v_{n2}^2}}{\Delta f} = 2 \times 4kT(R_{c2} + R_{eq}) \quad ,$$

(4.67)

$$\frac{\overline{v_{n3}^2}}{\Delta f} = 4kTR_{c3} \quad \text{and}$$

(4.68)

$$\frac{\overline{v_{eq}^2}}{\Delta f} = [1 - \exp(-\omega \Delta T)]^2 \left(\frac{\overline{v_{eq1}^2}}{\Delta f} + \frac{\overline{v_{eq2}^2}}{A_1^2 \Delta f} \right)$$

(4.69)

where the factor of 2 results from the differential configuration of the actual circuit and R_{eq} is the equivalent noise resistance of the substrate pnp transistor.

Assuming the time constants formed by switch on resistances and capacitances are negligible compared to $1/p_o$. The transfer function of v_{eq} to the output is

$$H(s) = \frac{A_{op} p_o}{s + p_1}$$

(4.70)

where

$$p_1 = p_o \left[1 + \frac{A_{op}}{1 + \frac{C_2}{C_1}} \right] \quad .$$

(4.71)

Similarly, under the assumption that the transfer functions from v_{n1} and v_{n2} to the output are

$$H_1(s) = \frac{C_2}{C_1 + C_2} H(s) \quad \text{and}$$

(4.72)

$$H_2(s) = \frac{C_2}{C_1 + C_2} H(s) \quad ,$$

(4.73)

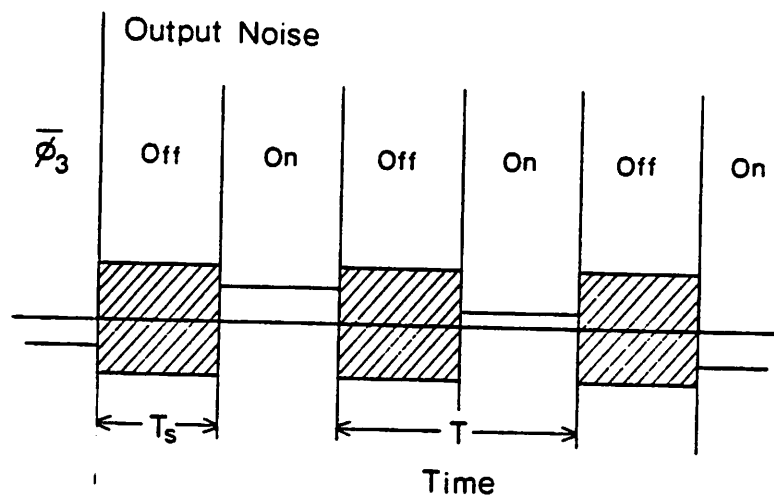respectively. The equivalent input referred noise spectral density is thus

Fig. 4.13. Output noise in time domain. T_s is the time during which op amp noise appears at the output.

$$S_i(\omega) = \frac{\overline{v_{vq}^2}}{\Delta\omega} + \frac{\overline{v_{n1}^2}}{\Delta\omega}\left[\frac{C_1}{C_1+C_2}\right]^2 + \frac{\overline{v_{n2}^2}}{\Delta\omega}\left[\frac{C_2}{C_1+C_2}\right]^2 + \frac{\overline{v_{n3}^2}}{\Delta\omega}\left|\frac{1}{H(\omega)}\right|^2 . \tag{4.74}$$

The output V_{ref} is the sum of two processes as shown in Fig. 4.13. A stationary process is applied periodically (on-off) to a linear system with an impulse response $h(t)$ and the output is sampled during on cycle. From (4.70), the impulse response of $H(\omega)$ is the form of

$$h(t) = A_{op}p_0 e^{-p_1 t} u(t) \tag{4.75}$$

where $u(t)$ is an unit step function. As in Appendix III, the autocorrelation function of the output $R_0(\tau)$ is related to the input autocorrelation function $R_i(\tau)$ so that

$$R_0(t_1, t_2) = R_i(t_1, t_2) * h(t_1) * h(t_2) \tag{4.76}$$

where

$$R_i(\tau) = \frac{1}{2\pi}\int_0^\infty S_i(\omega)e^{j\omega\tau}d\tau . \tag{4.77}$$

Therefore, we can get the autocorrelation of the output $R_0(\tau)$ from (4.75), (4.76) and (4.77).

For an order-of-magnitude calculation, let's assume that $S_i(\omega)$ is white. Then,

$$R_i(t_1, t_2) = R_i(0)\delta(t_1 - t_2) \qquad t_1 \geqslant 0 \text{ and } t_2 \geqslant 0 \tag{4.78}$$

$$= 0 \qquad \text{otherwise}$$

where $R_i(0)$ is the equivalent input white noise variance. From (4.76) and (4.78), we have

$$R_i(t_1, t_2) * h(t_1) = \int_0^\infty R_i(t_1-\tau-t_2)h(\tau)d\tau \tag{4.79}$$

$$= A_{op}p_0 R_i(0)e^{-p_1(t_1-t_2)}u(t_1-t_2) \qquad \text{for } t_1, t_2 \geqslant 0 ,$$

and finally

$$R_o(t_1, t_2) = R_i(t_1, t_2) * h(t_1) * h(t_2) \tag{4.80}$$

$$= A_{op}^2 p_o^2 R_i(0) \int_0^\infty e^{-p_1(t_1-t_2+\tau)} u(t_1-t_2+\tau)$$

$$= \frac{A_{op}^2 p_o^2 R_i(0)}{2p_1} e^{-p_1(t_1-t_2)}(1 - e^{-2p_1 t_2}) \quad \text{for} \quad t_1 \geqslant t_2 .$$

For $t_1 \leqslant t_2$, the role of t_1 and t_2 is to be reversed. Note that, as t is getting bigger, (4.80) becomes

$$R_o(\tau) = \frac{A_{op}^2 p_o^2 R_i(0)}{2p_1} e^{-p_1|\tau|} . \tag{4.81}$$

If we limit the time to the interval of $-T_s \leqslant \tau \leqslant T_s$, we have

$$R_o(\tau) = \frac{T_s}{T}(1 - \frac{|\tau|}{T_s}) \frac{A_{op}^2 p_o^2 R_i(0)}{2p_1} e^{-p_1|\tau|} \quad \text{for} \quad |\tau| \leqslant T_s \tag{4.82}$$

$$= 0 \quad \text{otherwise} .$$

Taking the Fourier transform of (4.82), the output spectral density is

$$S_o(\omega) = \int_{-T_s}^{T_s} R_o(\tau) e^{-j\omega\tau} d\tau \tag{4.83}$$

$$= \frac{T_s}{T} \frac{A_{op}^2 p_o^2 R_i(0)}{p_1^2} \frac{1}{1 + (\omega/p_1)^2} .$$

Therefore, the output noise variance during T_s is from (4.83)

$$R_o(0) = \frac{T_s}{T} \frac{A_{op}^2 p_o^2}{2p_1} R_i(0) \tag{4.84}$$

where $R_i(0)$ is the magnitude of input noise spectral density when $S_i(\omega)$ is assumed to be white. The noise variance of (4.84) is sampled when the sync pulse is low. During the other time slot of $T-T_s$, we have

$$R_o(\tau) = \frac{T-T_s}{T} \frac{T_s}{T} \left(1 - \frac{|\tau|}{T-T_s}\right) \frac{A_{op}^2 p_o^2 R_i(0)}{2p_1} \quad \text{for} \quad |\tau| \leqslant T-T_s \tag{4.85}$$

$$= 0 \quad \text{otherwise} \ .$$

The spectral density of the output during $T-T_s$ is by taking the Fourier transform of (4.85)

$$S_o(\omega) = (T-T_s)^2 \frac{T_s}{T^2} \, sinc^2 \frac{\omega(T-T_s)}{2} \frac{A_{op}^2 p_o^2}{2p_1} R_i(0) \ . \tag{4.86}$$

where $sinc\,x = \sin x/x$. The total output noise variance is from (4.83) and (4.86)

$$\overline{v_o^2} = \frac{1}{2\pi} \int_0^\infty S_o(\omega) d\omega = R_i(0) \frac{A_{op}^2 p_o^2}{2p_1} \frac{T_s}{T} \left(1 + \frac{T-T_s}{T}\right) \ . \tag{4.87}$$

Also by adding (4.83) and (4.86), the total output noise spectral density is given by

$$S_o(\omega) = R_i(0) \frac{A_{op}^2 p_o^2}{p_1^2} \frac{T_s}{T} \left[\frac{1}{1+\frac{\omega^2}{p_1^2}} + \frac{(T-T_s)^2}{T} \frac{p_1}{2} sinc^2 \frac{\omega(T-T_s)}{2} \right] \tag{4.88}$$

where T and T_s are the clock period and the sync pulse on time, respectively.

The output sampling introduces the under-sampled noise spectrum as in (4.88). The observed spectrum is shown in Fig. 4.12(b). The output noise of the reference differs from chip to chip. Typically, $400\mu V$ at the output in the $500kHz$ bandwidth is observed. This value is a little higher than the theoretical noise. The discrepancy is believed to be due to the high $1/f$ noise and not to the limitation of this kind of reference. When only white thermal noise is considered based on (4.47), (4.53), (4.65) and (4.69), the theoretical noise should be less than $100\mu V$ in the $500kHz$ bandwidth.

CHAPTER 5

EXPERIMENTAL RESULTS FOR A CMOS BANDGAP REFERENCE

5.1. Design Considerations

5.1.1. Trimming Accuracy

To set the reference output voltage, we need some kind of trim networks which are usually composed of passive components such as resistor strings and binary-weighted capacitor arrays. The trim networks of a resistor string and a binary-weighted capacitor array are illustrated in Fig. 5.1 and 5.2, respectively. Note that the capacitor trim is performed electronically while the resistor trim is done by blowing the fuses between the taps. Table IX summarizes the trim ranges of the individual resistor strings and the capacitor array if the nominal resistance of 6μ-wide p^- diffused resistor is 70Ω. There are four trims of the passive components.

1. R_2 Trim : A 6-bit resistor string is necessary to set the internal bias voltage V_o which is shown in Fig. 4.3. the voltage V_o is trimmed with $\pm 5.2mV$ accuracy. If everything is ideal, this $5.2mV$ setting error corresponds to the V_o temperature drift of

TABLE IX

OUTPUT VOLTAGE TRIM RANGE

	Bits	Min.	Max.	LSB	Range (mV)
R_2	6	$35k\Omega$	$63k\Omega$	441Ω	343 ± 5.2
R_3	6	$10.5k\Omega$	$24.5k\Omega$	221Ω	46.7 ± 0.7
R_4	6	$31.5k\Omega$	$59.5k\Omega$	441Ω	16.5 ± 0.3
C_2	6	$5.2pF$	$10.4pF$	$0.081pF$	480 ± 7.5
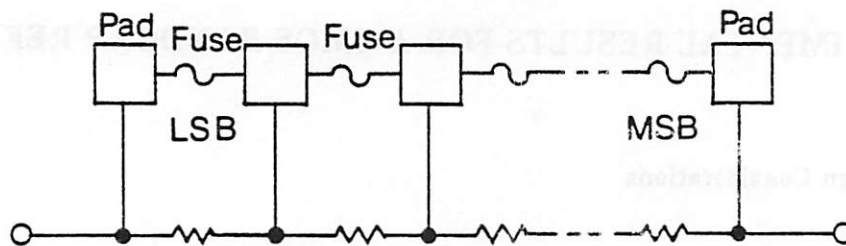
6-Bit Binary-Weighted R String

(a)
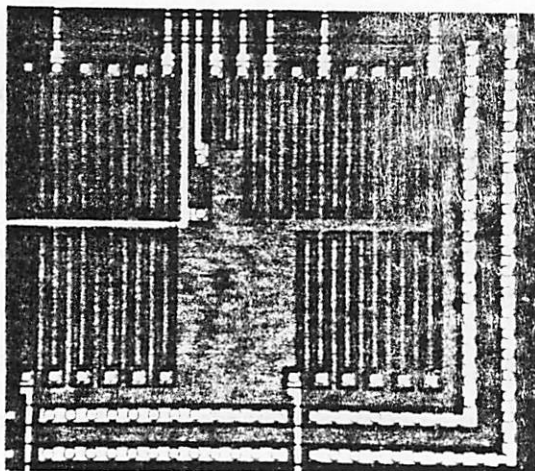
(b)

Fig. 5.1.   Blowing fuse type resistor trim : (a) a resistor
string and (b) its photo.

$$\frac{5.2mV}{300^\circ K} = 17.3 \ ppm/^\circ C \quad . \tag{5.1}$$

2. R_3 Trim : The resistor R_3 is trimmed to give the second-order correction voltage determined by the ratio of I_T/I_o. The ratio of the PTAT current I_T and the temperature independent current I_o is related to the ratio of R_3/R_1 from (4.3) and (4.4). If this ratio is adjusted by R_1 instead of R_3, we have to trim R_2 again to set V_o because V_o depends on R_1 through I_T. For this reason, R_3 is chosen for the R_3/R_1 ratio trim. The R_1 is fixed at $14k\Omega$ approximately.

3. R_4 Trim : The resistor R_4 is to generate the bias current I_D which is given by (4.5). I_D is generated exactly in the same way as I_o as shown in Fig. 4.3. Due to the logarithmic character of V_{BE} as a function of I_D, we can achieve a fine trim accuracy. The diode voltage V_{BE} is not a linear function of R_4. However, if the incremental variation of R_4 is small enough, the emitter-base potential V_{BE} changes linearly by

$$INCREMENT \ of \ V_{BE} = V_T \ln \frac{I_D + \Delta I_D}{I_s} - V_T \ln \frac{I_D}{I_s} \tag{5.2}$$

$$\approx V_T \frac{\Delta I_D}{I_D} \approx -V_T \frac{\Delta R}{R}$$

4. C_2 Trim : The capacitor C_2 trim is combined with R_4 trim to set the output V_{ref} with the total trim range of $480mV \pm 0.3mV$. The LSB of C_2 is designed to be equal to the second MSB of R_4 to insure no discontinuity in the trim of the output.

To sum up, the room temperature trim procedures of the reference is as follows

First Step : Adjust R_2 to set the predetermined V_o (4.9).

Second Step : Adjust R_4 and C_2 to set the predetermined V_{ref} with I_T disconnected. At this step, the reference output V_{ref} is the intermediate voltage without the PTATS correction voltage $FV_{T_o}^2$ (3.73).

6-Bit Binary-Weighted C Array
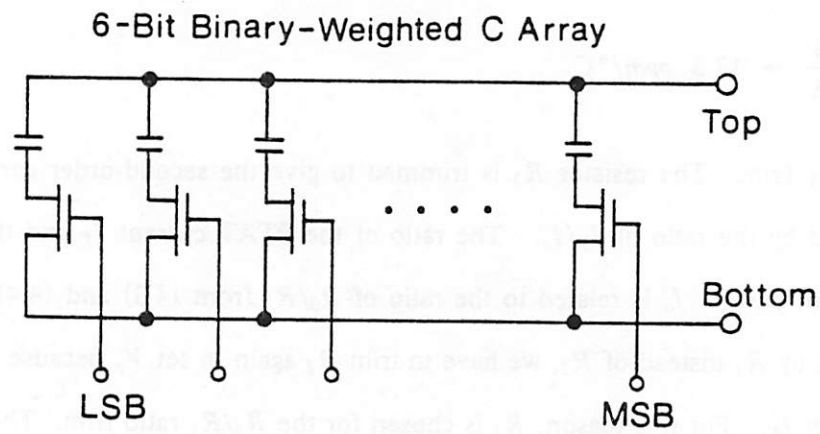
(a)
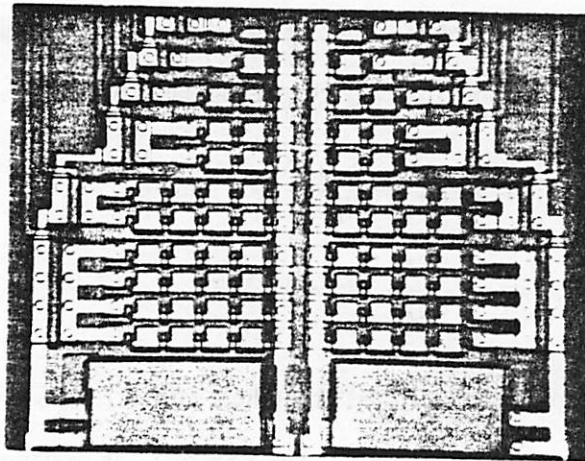
(b)

Fig. 5.2. Electronic capacitor trim : (a) a capacitor array and
(b) its photo.

Third Step : Adjust R_3 to set the predetermined V_{ref} with I_T connected. After this step, the final reference output V_{ref} contains both the PTAT and PTATS correction voltages (3.71).

5.1.2 Latch-Up Prevention

The regenerative mechanism of the four-layer pnpn structures has long been known to explain a latch-up failure in CMOS. Fig. 5.3 shows a typical CMOS cross section showing a latch-up path. For example, one pnpn path runs from the p^+ source of the p-channel device, through the n^- well, continuing into the p^- substrate, and ending at the n^+ source of the n-channel device. This pnpn structures consists of a vertical pnp transistor coupled with a lateral npn transistor so that the pnp collector sources the base current of the npn transistor while the npn collector sinks the base current of the pnp transistor. This pnpn structure resembles a semiconductor controlled rectifier (SCR). Under the normal bias condition, the pnpn structure remains in off condition. Only a small leakage current will flow across the n^- well and the substrate junction under normal bias. However, if certain conditions are met, the parasitic bipolar transistor may become active. Generally, for a latch-up to occur, the following three conditions must be met:

Condition 1 : Both npn and pnp transistors must be biased into the forward active region for appreciable minority carrier injection to occur.

Condition 2 : To allow regeneration, the following sufficient loop gain is required ;

$$\beta_{npn} \cdot \beta_{pnp} > 1 \ . \tag{5.3}$$

Condition 3 : Sufficient current source or sink external to the pnpn path is required to sustain the high-current condition.

The above condition 1 is easily met if the lateral currents I_{LS} in the substrate and I_{LW} in the n^- well are present as shown in Fig. 5.3. The n^+ source can be shorted to the p^+ substrate contact by the external metal line. Likewise, the p^+ source can also be shorted to
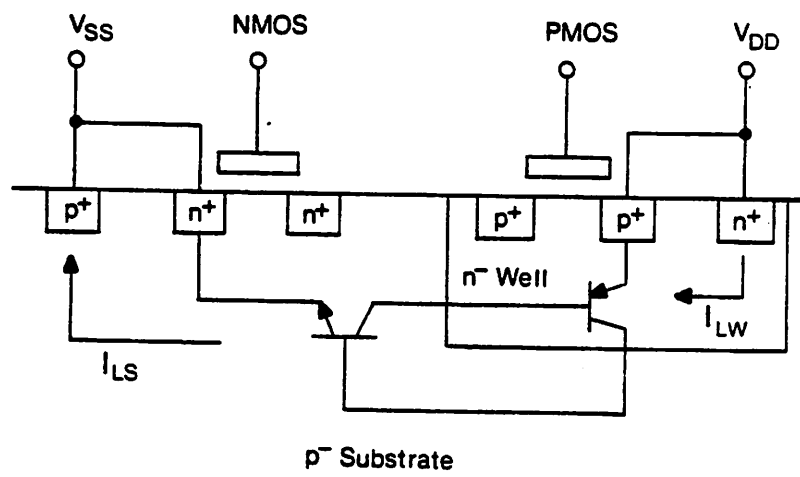
Fig. 5.3. Cross-sectional view of one CMOS latch-up path.

the n^+ well contact by the external metal line. When a lateral current is induced by some means between V_{DD} and V_{SS} terminals, lateral ohmic voltage drop appears in the substrate and the well due to their high resistivities. This lateral ohmic voltage drop forward-biases the emitter-base junctions of the parasitic npn and pnp transistors, thus to initiate latch-up. There are many ways in which the lateral current can arise. They are

1. Excess Supply : If an applied terminal voltage $V_{DD}-V_{SS}$ exceeds the breakdown voltage of the n^- well and p^- substrate junction will produce a lateral current.

2. Voltage Transient : The voltage transient or spike on the supply line produces a displacement current from V_{DD} to V_{SS}. The increase of $V_{DD}-V_{SS}$ widens the n^- well and p^- substrate depletion layer to support the additional voltage between the supply terminals. The resulting displacement current from V_{DD} to V_{SS} is proportional to both the junction capacitance and the time rate of the voltage change across the junction so that

$$i_d = C_{jw}\frac{\Delta(V_{DD} - V_{SS})}{\Delta t} \tag{5.4}$$

where C_{jw} is the well-substrate junction capacitance.

3. Radiation : The ionizing radiation such as X-ray or γ-ray generates electron-hole pairs. Generated minority carriers in the depletion region and within a diffusion length are swept across the depletion layer. Upon crossing the depletion layer, they participate in the lateral current flow as majority carriers.

4. Input Protection Circuit Overdrive ; The gate protection diode at the input terminal will be a source for the lateral current by injecting minority carriers into the substrate when the input terminal voltages are lower or higher than the supply voltage.

Although it is unusual, however, in a CMOS bandgap reference, the n^- well is intentionally used as a base of the vertical pnp structure. Thus the forward-biased p^+ emitter in the n^- well inject holes into the base and these injected holes are swept across the well-substrate junction and collected at the substrate. Therefore, in the layout of the vertical

pnp transistors, special care was taken to avoid latch-up conditions. Without changing the process, the general guide lines of good CMOS layout for avoiding latch-up are to:

1. Minimize the lateral current which can flow along a potential latch-up path.

2. Reduce the magnitude of the voltage-producing shunt resistance components.

3. Avoid the close spacing between the n^+ diffusion and the n^- well to suppress the parasitic lateral npn transistor action.

Specifically, the n^+ diffusion tied to the negative supply V_{SS} should not be placed close to the n^- well to reduce the current gain of the lateral npn transistors. Although the extensive use of guard bands is not compatible with the layout compaction for a high density CMOS, every n^- well should be surrounded with a n^+ guard band so as to effectively lower the lateral resistance. When possible, a pseudo collector should be placed around the gate-protection diode. For the substrate pnp transistor, the p^+ diffusion should surround the base n^- well to collect the majority holes injected from the base. Even though the holes injected from the well can still drift and diffuse under the surrounding collector, the order of two immunity to latch-up

5.1.3. MOS Switches

Depending on the device size and the applied voltage, an MOS switch has a finite resistance of

$$R_{on} = \frac{1}{\mu C_{OX} \frac{W}{L} (V_{GS} - V_{th})}$$

(5.5)

where V_{th} is the threshold voltage. As shown in Fig. 5.4(a), the MOS switches for M_{s1} and M_{s2} for capacitors C_1 and C_2 introduce a delay in the feedback. The loop gain Af becomes

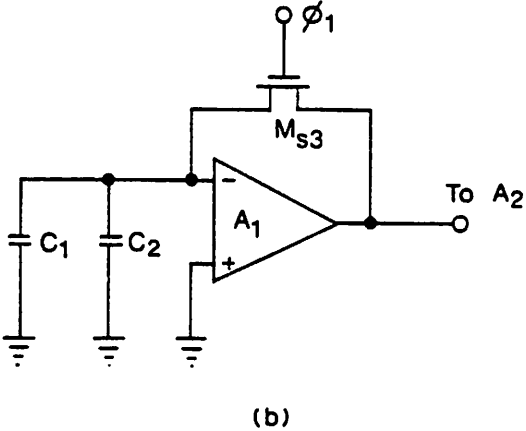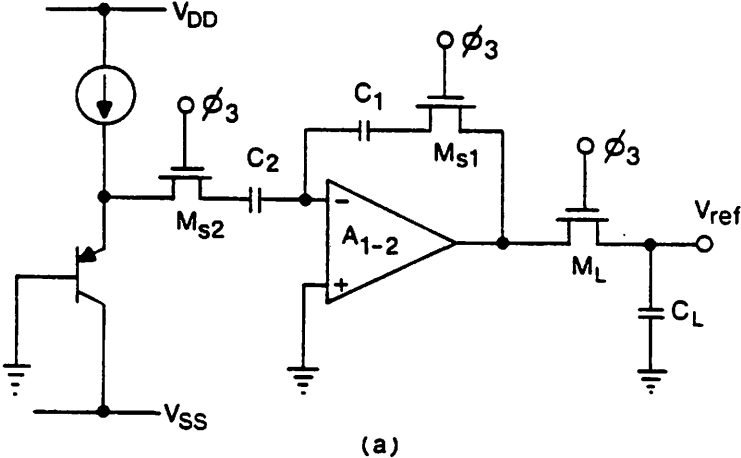Fig. 5.4. Switch on-resistance effects : (a) amplification
mode and (b) first offset storage mode.

$$Af = A \left| \frac{R_{on2}^2 + (\frac{1}{\omega C_2})^2}{(R_{on1} + R_{on2})^2 + (\frac{1}{\omega C_1} + \frac{1}{\omega C_2})^2} \right|^{1/2} \exp^{-\Delta\Theta} \qquad (5.6)$$

where R_{on1} and R_{on2} are the on resistances of the MOS switches M_{s1} and M_{s2}, and $\Delta\Theta$ is the phase delay of

$$\Delta\Theta = \tan^{-1} \frac{1}{R_{on1} + R_{on2}} (\frac{1}{\omega C_1} + \frac{1}{\omega C_2}) - \tan^{-1} \frac{1}{\omega R_{on2} C_2} \qquad (5.7)$$

$$\approx \tan^{-1} \frac{1}{\omega R_{on1} C_1} - \tan^{-1} \frac{1}{\omega R_{on2} C_2}$$

Thus, if we make the time constant $R_{on1}C_1$ equal to the time constant $R_{on1}C_2$, the phase delay $\Delta\Theta$ due to the on resistances of the MOS transistors approaches zero. In the actual circuit, however, the capacitor C_2 is variable from $5.2pF$ to $10.4pF$ to trim the output voltage V_{ref}, and the matching of the two time constants is not economical because the sizes of the MOS switches is to be adjusted along with the capacitors. We can make the size of M_{s1} and M_{s2} arbitrarily big enough for the external phase delay to be negligible. For the geometry of $W/L = 36\mu m/6\mu m$, R_{on1} is from (5.5)

$$R_{on1} \approx \frac{1}{35 \times 10^{-6} \times \frac{36}{6}(5-1)} \approx 1.2 k\Omega \qquad (5.8)$$

Even though 50% change of C_2 is allowed, the phase delay amounts to

$$\Delta\Theta = \tan^{-1} \frac{1}{2\pi 600 kHz \times 1.2 k\Omega \times 0.65 pF} - \tan^{-1} \frac{1}{2\pi 600 kHz \times 1.2 k\Omega \times 10.4 pF}$$

$$= 0.1^{\circ} \qquad (5.9)$$

which is virtually negligible.

Another case is shown in Fig. 5.4(b) where M_{s3} hold the amplifier A_1 to the unit gain configuration. The time constant formed by R_{on3} and $C_1 + C_2$ is about $8.6nsec$. Therefore, the phase delay due to M_{s3} is

$$\Delta\Theta = 90^o - \tan^{-1}\frac{1}{\omega R_{on3}(C_1 + C_2)} \tag{5.10}$$

$$\approx 90^o - \tan^{-1}\frac{1}{2\pi\times3.9MHz\times8.6nsec} \approx 12^o \quad .$$

To reduce this extra phase shift due to M_{s3}, the size of transistor must be increased. As the size is getting bigger, the clock feedthru and the leakage current to the substrate at the summing nodes will complicate the situation. Fortunately, this phase shift is not critical in the operation because as far as the amplifier is compensated to be stable, the speed of the unit gain settling is only a small part of the offset sample and subtraction cycle.

5.2. Substrate PNP Transistor Design

Since the aim of this work is to build a precision MOS voltage reference without modifying an existing standard CMOS process, no arbitrary change of the process to optimize component performance is permitted. Therefore, the worst case component performance must be taken into account. The most critical limitation is the base spreading resistance and the finite current gain of the substrate pnp transistors. The substrate pnp transistor whose base is the n^-well has an inappropriately high resistivity as a transistor base. For the further reduction of the base spreading resistance, the base contact surrounds the emitter junction as shown in Fig. 5.5. The emitter junction and the base contact plug are separated by $9\mu m$ even though the separation may be reduced further. The finite current gain of less than 50 must be considered in the design although β of 100 to 250 was observed experimentally. In the following two subsections, the design considerations upon the base spreading resistance and the current gain will be discussed.

5.2.1. Base Spreading Resistance Limitation

The base of the substrate pnp transistor shown in Fig. 5.6(a) can be divided into two regions. The active base region is directly under the emitter diffusion and transistor action
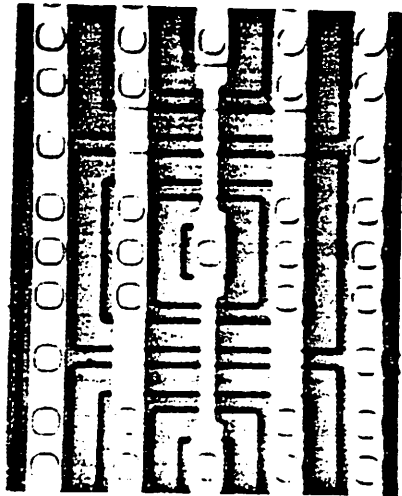
(a)
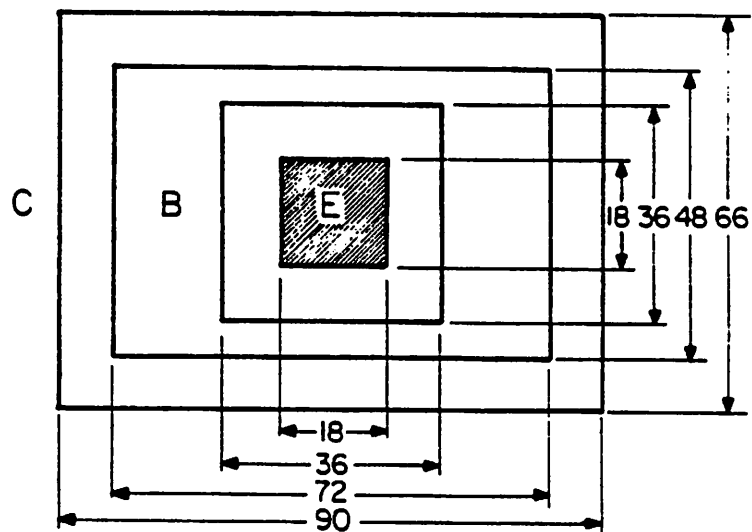
(b)

Fig. 5.5. (a) Photograph of one substrate pnp transistor
         unit cell and (b) its actual dimension (drawn).

takes place in this region. The remainder of the base is the inactive neutral region. Assuming low-level injection and neglecting the depletion region under the emitter-base junction, the lumped equivalent r_{b1} and r_{b2} represent the active and inactive base regions, respectively. T_1 and T_2 are the depths of the corresponding base regions.

Assume the differential emitter area of thickness dx as shown in Fig. 5.6(a) to be equipotential. The currents I_1 to I_n and the resistances R_1 to R_n are the currents and the resistances of the differential elements. Then, the active base area is modeled as shown in Fig. 5.6(b). Since the current density remains constant all over the emitter area, the current through the i—th differential element is

$$I_i = \frac{8}{a^2} x\,dx\; I_B \quad . \tag{5.11}$$

Symmetry of the emitter and the resistance calculating method [50] give the resistance of the differential element of

$$R_i = \frac{\rho_s}{8T_1} \ln\frac{2x+dx}{2x-dx} \approx \frac{\rho_s}{8T_1}\frac{1}{x}\,dx \tag{5.12}$$

where ρ_s is the resistivity of the base region. Therefore, the transverse ohmic voltage drop across the active base area is from (5.11) and (5.12)

$$\begin{aligned}
r_{b1}I_B &= \lim_{n\to\infty}\sum_{i=1}^{n} I_i(\sum_{j=i+1}^{n} R_j)\\
&\approx \frac{\rho_s I_B}{a^2 T_1}\int_0^{\frac{a}{2}} x\int_x^{\frac{a}{2}} (\ln\frac{a}{2}-\ln y)\,dy\,dx \approx \frac{\rho_s}{16T_1}\,I_B
\end{aligned} \tag{5.13}$$

where ρ_s/T_1 is the resistance per square of the active base region. From (5.13), we obtain

$$r_{b1} = \frac{\rho_s}{16T_1} \quad . \tag{5.14}$$

In the same manner, r_{b2} is given by
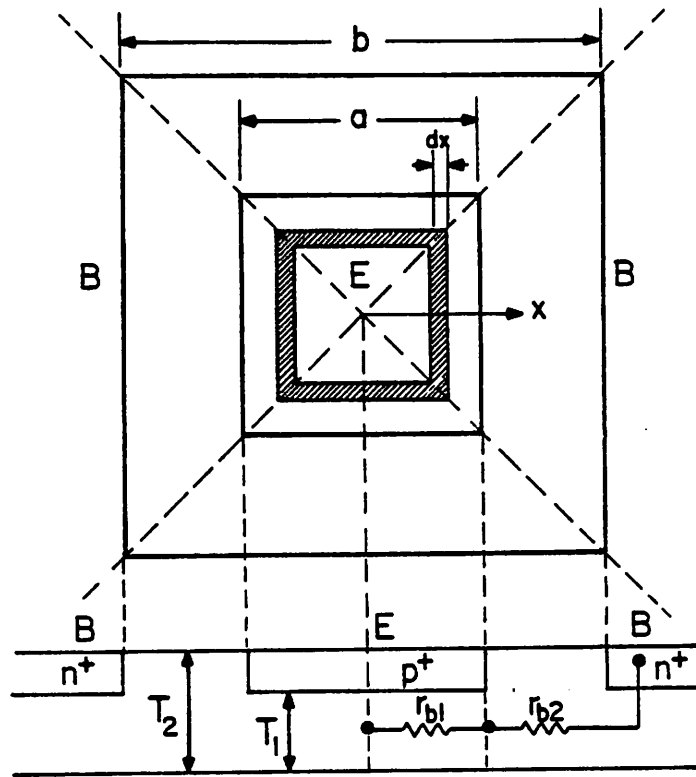
(a)

Emitter-Base Junction
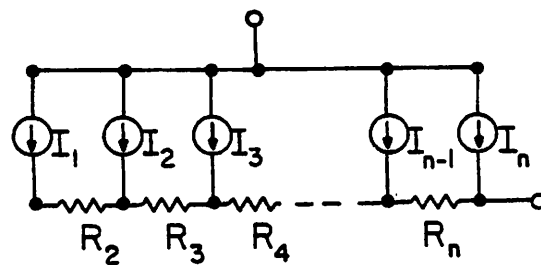
(b)

Fig. 5.6. Base resistance analysis model of pnp transistors (a) and equivalent circuit representation of the pinched base region (b).

$$r_{b2} = \frac{\rho_s}{8T_2} \ln\frac{b}{a} \quad . \tag{5.15}$$

Adding (5.14) and (5.15) yields the total intrinsic base resistance of

$$r_b = r_{b1} + r_{b2} = \frac{\rho_s}{16T_1} + \frac{\rho_s}{8T_2} \ln\frac{b}{a} \quad . \tag{5.16}$$

In the current process, the r_b is estimated to be $1.5k\Omega$. For $1.5k\Omega$, the ohmic voltage drop is $r_b I_B = 0.27mV$ which is much smaller than the thermal voltage $26mV$ at $25^\circ C$. Therefore, the effective emitter area reduction [62] due to the base crowding is negligible.

5.2.2. Current Gain Limitation

At a low bias current level, the current gain β is limited by the space charge recombination current. Therefore, the minimum base current level should be much greater than the space charge recombination current in the emitter-base junction to avoid the β reduction due to this mechanism. The space charge recombination current in the emitter-base space charge region is a weak exponential function of V_{BE} [58] :

$$I_r = \frac{qA_e X_s n_i}{2\tau_o} \exp\frac{qV_{BE}}{2kT} \tag{5.17}$$

where A_e is the emitter area, X_s is the width of the space-charge region, and $\tau_o = 1/N_t \delta v_{th}$ is the lifetime associated with the recombination of excess carriers in a region with a density N_t of recombination centers, a cross section δ of recombination centers and a thermal velocity v_{th} of carriers. For typical values, the recombination current is estimated to be

$$I_r = \frac{1.6\times10^{-19}\times324\mu m^2\times0.5\mu\times1.4\times10^{10}}{2\times10^{-6}} \exp\frac{0.6V}{2\times26mV} \tag{5.18}$$

$$\approx 19 \ nA \quad \text{at} \ 300^\circ K \quad .$$

Assuming a current gain of 100, emitter current should be greater than $1.9\mu A$. In an actual circuit, however, the current gain β does not decrease down to the emitter current of
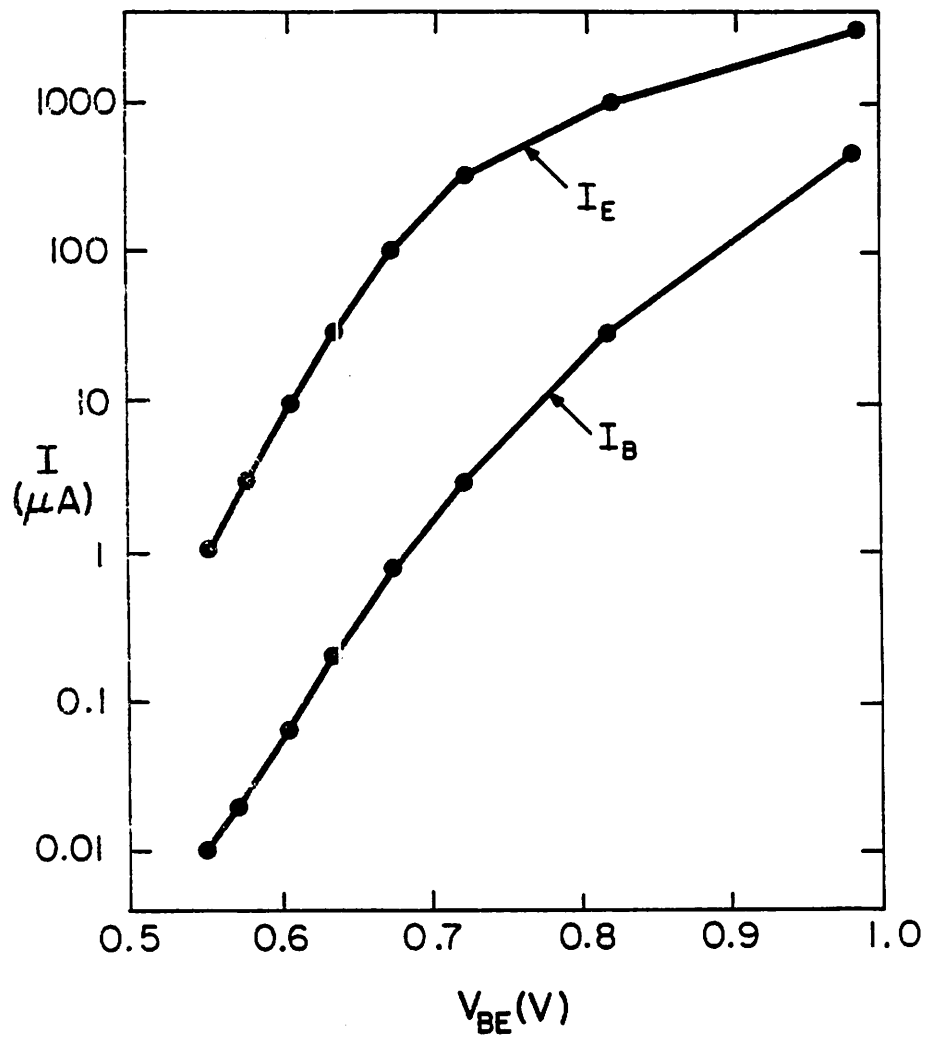
Fig. 5.7. Typical measured performance of the substrate pnp
transistors in Berkeley CMOS process.

$1\mu A$. The minimum emitter bias current is designed to be more than $5\mu A$.

5.3. Component Measurement Results

The minimum feature size in the layout is $6\mu m$. As jump wires, n-diffusion resistors $(25\Omega/Sq.)$ and polysilicon resistors $(30\Omega/Sq.)$ are used outside and inside the n^- well, respectively, to reduce a connection resistance. The p^+ diffusion resistor $(75\Omega/Sq.)$ in the n^- well is used as a resistor component. As a capacitor component, the polysilicon-n^+ diffusion capacitor $(5\times10^{-9}F/cm^2)$ separated by $70nm$ thin oxide is used. The problem in the use of the substrate pnp transistor is the direct carrier injection into the substrate which may initiate the latch-up plaguing CMOS circuits. No latch-up is observed in the measurement even under the power supply transient.

The typical measured characteristics of pnp transistors are shown in Fig. 5.7. No base crowding effect is observed over the $1^-100\mu A$ emitter current range. The current gain β is also relatively constant within this range. However, in the emitter current range over $100\mu A$, the base crowding effect is detrimental. In the range lower than $1\mu A$, β is reduced by the space charge recombination current. The unit cells of Fig. 5.5(a) are connected in parallel to make the multiple emitter devices. The emitter area ratio of $A=9$ is chosen in an actual layout. The typical measured temperature variations of β and diffused resistors are shown in Figs. 5.8(a) and (b), respectively. The current gain β is minimum at $-50^\circ C$ and the average value at room temperature is 155. Table X summarizes pnp transistor parameters derived from the experimental data. The reverse saturation current temperature data is listed in Table XI and Table XII summarizes the β and V_{BE} measurement data including temperature coefficients. The temperature data of the p^+ diffused resistor in the n^-well is listed in the Table XIII. The temperature coefficient of the diffused resistors has a standard deviation of 3% and can match each other within 1.2%. Finally, the measured device parameters for the typical sizes of NMOS and PMOS transistors are
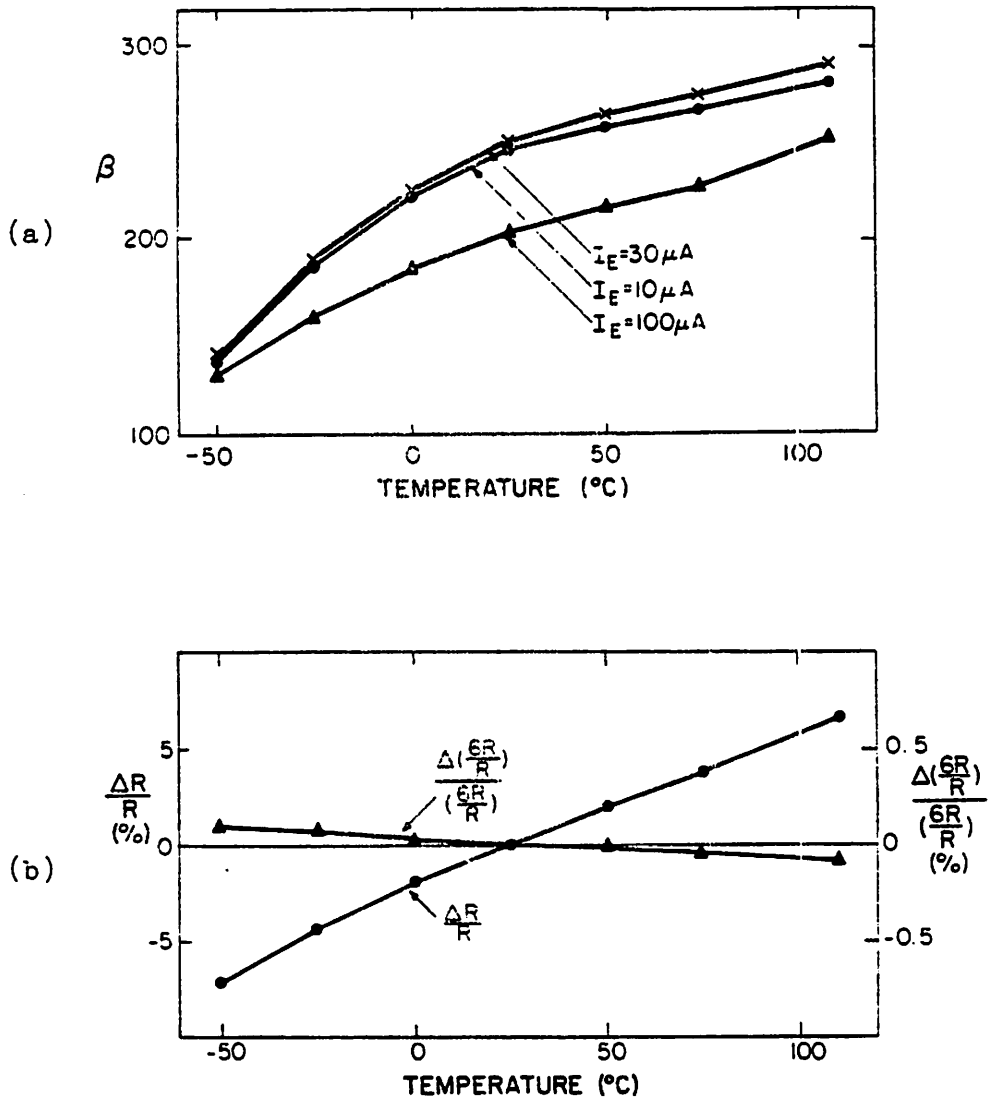
Fig. 5.8. Typical measured temperature characteristics of (a) the current gain of pnp transistors and (b) diffused resistors.

summarized in Table XIV.

TABLE X

PNP TRANSISTOR PARAMETERS*

β	I_s	V_A	η
155	$9.7×10^{-16}$ A	42 V	$6.2×10^{-4}$

*$I_e = 30 μA$ at 25°C.

TABLE XI

REVERSE SATURATION CURRENT (Amp)

	Saturation Current I_s				
Tem.	Unit	Mean	Dev.	Min.	Max.
−50°C	$×10^{-23}$	8.9	6.6	3.2	19.3
−25°C	$×10^{-20}$	6.4	1.5	5.1	9.1
0°C	$×10^{-18}$	8.6	2.2	5.9	1.2
25°C	$×10^{-16}$	9.7	1.9	7.5	12.4
50°C	$×10^{-14}$	2.7	0.75	2	3.9
75°C	$×10^{-13}$	7.2	0.92	6.1	8.4
110°C	$×10^{-11}$	2.5	0.35	2.1	3

TABLE XII

CURRENT GAIN AND EMITTER-BASE POTENTIAL

	Mean	Dev.	Min.	Max.
Current Gain β				
β at $300^\circ K$	155	89	91	273
T.C. of β*	6503	872	6471	7792
Emitter-Base Potential V_{BE}				
$V_{BE}(V)$ at $300^\circ K$	0.6309	0.005	0.6268	0.6360
T.C. of V_{BE}**	1.985	0.095	-1.86	-2.07

*$ppm/^\circ C$.

**First-order T.C., $mV/^\circ C$.

TABLE XIII

TEMPERATURE DATA OF DIFFUSED RESISTORS

$(-55$ to $125^\circ C)$

|	Mean	Dev.	Min.	Max.
p^+ Diffused Resistor;				
Sheet Resistance (Ω/square)	68.9	3.1	64.6	72.1
T.C. of R*	886	27.9	849	917
Ratio of $6R/R$**	5.952	0.02	5.9304	5.9794
T.C. of $6R/R$*	10.6	2.2	8.9	13.6

*T.C. unit is $ppm/^\circ C$.

**R and $6R$ are composed of 40 and 240 squares of $6\mu m$-wide p^+ diffusion.
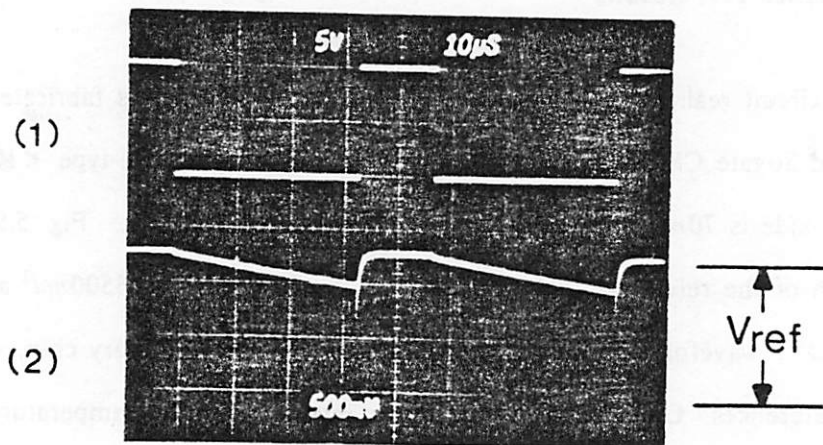
TABLE XIV

MOS TRANSISTOR PARAMETERS

$\dfrac{W}{L}$	$\dfrac{123}{6}$	$\dfrac{123}{12}$	$\dfrac{123}{6}$	$\dfrac{123}{12}$
Type	NMOS	NMOS	PMOS	PMOS
V_{TO}	$0.7V$	$0.8V$	$-0.6V$	$-0.7V$
Lambda	$0.037V^{-1}$	$0.025V^{-1}$	$0.035V^{-1}$	$0.01V^{-1}$
Field V_{th}	$>13V$		$>12V$	
Gamma	0.167		0.312	
$\bar{\mu}C_{OX}$	$35\mu A/V^2$		$20\mu A/V^2$	

5.4. Reference Test Results

The circuit realizing a precision CMOS bandgap reference is fabricated employing a self-aligned Si-gate CMOS process on a $20\text{-}30\,\Omega\,cm$ Boron-doped p-type $<100>$ substrate. The gate oxide is $70nm$ and the drawn minimum feature is $6\mu m$. Fig. 5.9(a) shows the photograph of the reference chip whose active area is occupying $3500mil^2$ and Fig. 5.9(b) shows its V_{ref} waveform as well as the output sync pulse. On every chip, we made three types of references. One sample is adjusted to give a minimum temperature drift. For a minimum temperature drift, the optimum values of the second-order temperature compensated V_{ref} at $25^{\circ}C$ and $FV_{f_0}^2$ are found to be $1.192V$ and $61mV$, respectively. The nominal value of the first-order temperature compensated V_{ref} is around $1.256V$ which is a little lower than the theoretical value predicted by (3.65). Estimating from the measured data based on (3.65), (3.71) and (3.74), the parameters V_{go1}, V_{go2} and $\gamma - \alpha$ necessary for specifying the prototype bandgap reference are approximately $1.181V$, $1.158V$ and 2.623, respectively. No effort has been made to adjust the temperature coefficients of the other 6 samples except setting the nominal voltages which are obtained from the first sample. The measured temperature coefficients of 7 samples from one wafer are listed in Table XV.
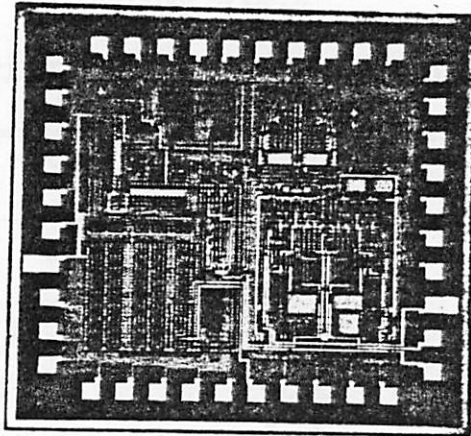
(a)

(b)

Fig. 5.9. (a) Experimental prototype chip and (b) output wave-
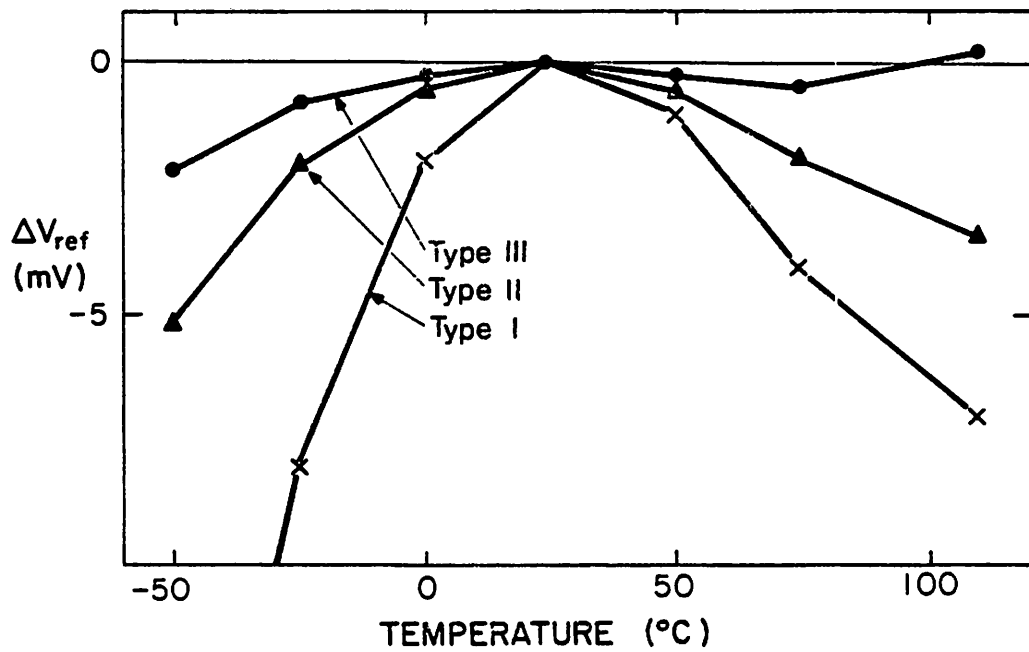forms : (1) output synchronization clock and (2) the
reference output.

Fig. 5.10. Temperature variations of three types of reference out-
put voltages.

TABLE XV

STATISTICS OF MEASURED TEMPERATURE COEFFICIENTS
OF 7 SAMPLES ($ppm^{\circ}C$)

	Mean	Dev.	Min.	Max.
0 to $70^{\circ}C$;				
Type I	105	43	42	167
Type II	22.3	10.8	11.1	42
Type III	13.1	7.1	5.6	25.7
-55 to $125^{\circ}C$;				
Type I	185	56.5	107	273
Type II	35.1	18.8	17.6	66.7
Type III	25.6	10.5	12.1	39.9

The Type I reference is the internal voltage V_o in Fig. 4.3 which is used to generate the temperature-independent current I_o. The voltage V_o represents a conventional CMOS bandgap reference without any consideration for the second-order effects. The Type II reference is the offset-canceled first-order temperature compensated bandgap reference which has the base current and the base resistance cancellations. The Type III reference is the curvature-compensated bandgap reference which has all the cancellations discussed so far. Note that the offset cancellation and the first-order compensation of r_b and β effects have already brought a factor of 5 improvement over the conventional approach in the achieved temperature stability. By the curvature compensation, we get a factor of 2 further improvement. In fact, the curvature compensation does not improve a temperature stability significantly as predicted in Fig. 3.4. This implies that that the uncertainty due to the process variation amounts to $10 ppm/^{\circ}C$ as discussed in Section 4.3 and the best achievable temperature coefficient in CMOS technology is mainly limited by the process variation not by the temperature compensation technique. Individual temperature
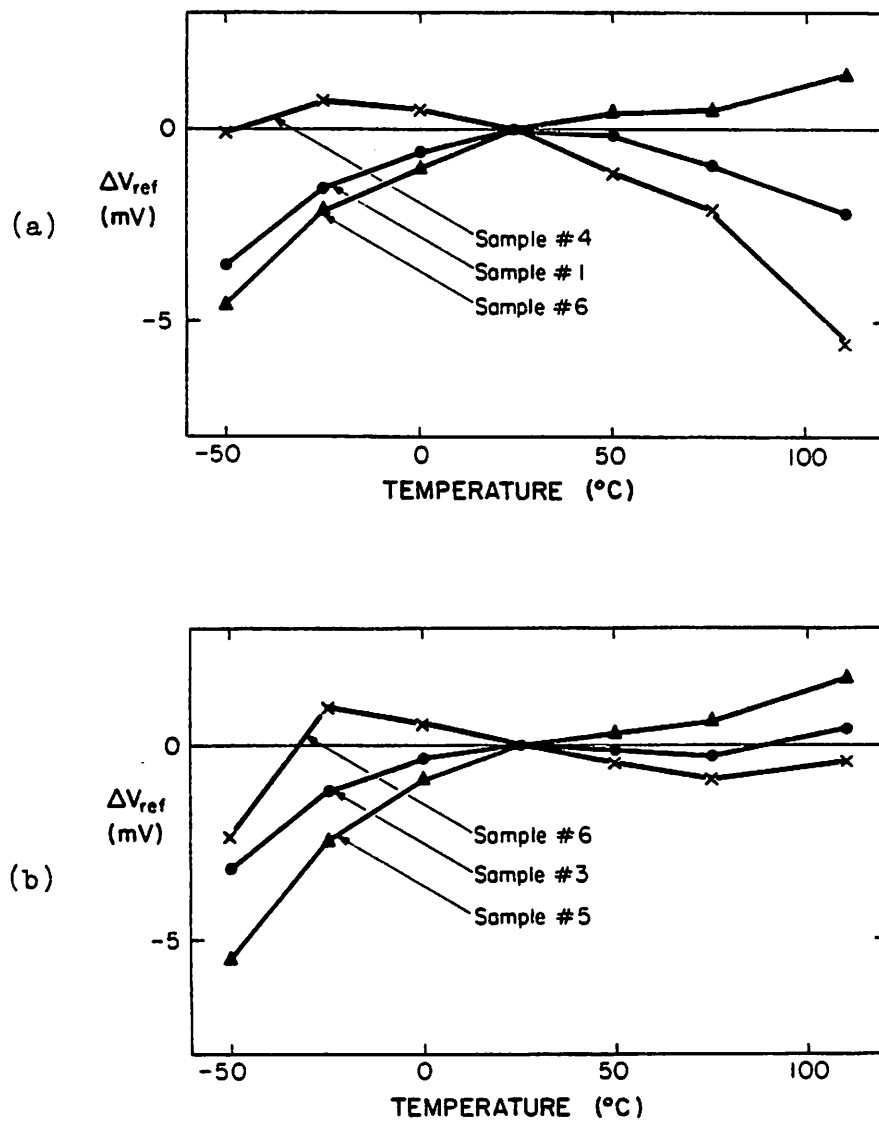
Fig. 5.11. Chip-to-chip process variations : (a) Type II reference and (b) Type III reference.

TABLE XVI

PERFORMANCE SUMMARY

V_{ref}	$1.192V \pm 1mV$ at $25°C$
T.C.	See TABLE XV
Power	12^-15mW with $\pm 5V$ Supply
Load	$100pF$ Capacitor
Cycle Time	$5\mu s$
+PSRR	$>50dB$ (DC)
-PSRR	$>60dB$ (DC)
Clock RR	$>75dB$
Output Noise	$400\mu V$ $(500kHz)$

coefficient can be trimmed and minimized below $10ppm/°C$ through the adjustment of the ratio of I_T/I_o. However, reliable and consistent process development has to precede because individual temperature trimmings are expensive. Fig. 5.10 compares graphically those three optimally compensated bandgap references. The typical drifts of the last two references due to the process variation are illustrated in Figs. 5.11(a) and (b), respectively. The measured performance of the prototype reference is listed in the Table XVI. To improve the power supply rejection ratio (PSRR), the base of all the pnp transistors should be biased constant relative to the negative supply line. Otherwise, the base width modulation (Early effect) will limit the PSRR of the reference.

CHAPTER 6

CONCLUSIONS

This dissertation consists of two works. One is the theoretical analysis of the ion-implantation effect on the device threshold voltage so as to locate the problems in implementing an NMOS voltage source based on the threshold voltage difference. The other is the implementation of a curvature-compensated switched-capacitor CMOS bandgap reference whose performance approaches the theoretical limit with a lower production spread than the existing CMOS BGR implementations.

In the first part of the work, the analytically simple equations are developed for the analysis of the temperature drift of the IGFET threshold voltage tailored by the ion-implantation. The discovered facts from the theoretical analysis and the experimental results are that each ion-implantation gives rise to two kinds of temperature variation. The temperature drift associated with the potential drop across the gate oxide is controlled by the body bias. However, the temperature drift associated with the channel to bulk built-in potential is independent of the body bias. So as to minimize the temperature drift resulting from the channel implantation, a shallow single implant is required, and IGFETs are recommended to be operated with a low bias current and a high body bias, and if feasible, the changes in the bias current, the body bias and the drain-source voltage are to be kept constant.

In the second part of the work, a prototype curvature-compensated monolithic CMOS bandgap reference, which achieves a temperature stability on the order of $10 ppm/^{\circ}C$ over the commercial temperature range without thin-film resistors and a precision laser trimming, is fabricated. The design features a curvature compensation, a simple trim up to 12-bit accuracy and a complete cancellation of the offset and the long-term offset drift of a CMOS op amp. A reference voltage is obtained by systematically adding the first-order and

the second-order temperature compensation voltages separately to the forward-biased diode voltage. Each temperature compensation voltage can be adjusted individually to minimize the reference output temperature drift and the reference output voltage can be set easily in the same manner as in the bipolar bandgap reference employing a room temperature trim procedure. The proposed reference can operate synchronously with other elements of the data acquisition systems of a 10^-12-bit accuracy.

APPENDIX I

MOS Threshold Voltages : Special Case of Step Profiles

In this appendix, the MOS threshold equations described in Section 2.2 are simplified for the step implantation profiles of Fig. 2.6.

1. Enhancement Mode Devices

By equating the area in Fig. 2.4(c) to $\Phi_E + V_{SB}$, the depletion depth in the enhancement device is given by

$$X_{dE} = \left[\frac{2\epsilon_s}{qN_a} (\Phi_E + V_{SB}) - \frac{N_{ai}}{N_a} X_i^2 \right]^{1/2} . \tag{1}$$

The threshold equation is from (2.1) and (2.2)

$$V_{TE} = V_{FB} + V_{SB} + \Phi_E - Q_{dE}/C_{OX} \tag{2}$$

where

$$\Phi_E = \frac{kT}{q} \ln \frac{N_{peak-E} N_a}{n_i^2} , \tag{3}$$

$$Q_{dE} = - qN_{ai} X_i - qN_a X_{dE} \tag{4}$$

and all the other terms have their usual meanings.

For the unimplanted device, (1) and (4) becomes

$$X_{dE} = \left[\frac{2\epsilon_s}{qN_a} (\Phi_E + V_{SB}) \right]^{1/2} \quad \text{and} \tag{5}$$

$$Q_{dE} = - qN_a X_{dE} . \tag{6}$$

2. Depletion Mode Devices

For the doubly-implanted device, equating the amount of the depleted charges on both sides of X_j as in Fig. 2.5(b) yields

$$(X_j - X_i)(N_{di} - N_{ai} - N_a) = (X_i - X_j)(N_{ai} + N_a) + (X_{dD} - X_i)N_a \quad . \tag{7}$$

Using (7) and equating the positive area of Fig. 2.5(c) to $\Phi_D + V_{SB}$, we obtain

$$X_1 = \frac{(N_{di}X_j - N_{ai}X_i)}{(N_{di} - N_{ai})}$$

$$- \left[\frac{2\epsilon_s N_a (\Phi_D + V_{SB})}{q(N_{di} - N_{ai} - N_a)(N_{di} - N_{ai})} - \frac{N_a N_{ai} N_{di} (X_i - X_j)^2}{(N_{di} - N_{ai} - N_a)(N_{di} - N_{ai})^2} \right]^{\frac{1}{2}} \quad . \tag{8}$$

Therefore, the threshold equation for the doubly-implanted device is from (2.1) and (2.2)

$$V_{TD} = V_{FB} + V_{SB} + \Phi_D - \Phi_X - \frac{Q_{dD}}{C_{OX}} \tag{9}$$

where

$$\Phi_D = \frac{kT}{q} \ln \frac{N_{peak-D} N_a}{n_i^2} \quad , \tag{10}$$

$$\Phi_X = \frac{q(N_{di} - N_{ai} - N_a)}{2\epsilon_s} X_1^2 \quad , \tag{11}$$

$$Q_{dD} = q(N_{di} - N_{ai} - N_a)X_1 \tag{12}$$

and all the other terms have their usual meanings.

For the depletion device, (8), (11) and (12) becomes

$$X_1 = X_j - \left[\frac{2\epsilon_s N_a (\Phi_D + V_{SB})}{q(N_{di} - N_a)N_{di}} \right]^{\frac{1}{2}} \quad , \tag{13}$$

$$\Phi_X = \frac{q(N_{di} - N_a)}{2\epsilon_s} X_1^2 \quad \text{and} \tag{14}$$

$$Q_{dD} = q(N_{di} - N_a)X_1 \quad . \tag{15}$$

These threshold equations for the depletion device are the same as the equation proposed by Sigmon [23] except the definition of the built-in potential. These equations also coincide with the equations widely used in the modeling of the depletion mode devices [63]-[65] and the buried channel CCDs (BCCDs) [66]-[68]. From (13), (14) and (15), and by setting $N_d = N_{di} - N_a$, we can easily obtain the forms used by Huang [63]-[64] ;

$$V_{TD} = V_{FB} + \Phi_{bi} + V_{SB} - \frac{qN_d X_j}{\bar{C}} \qquad (16)$$

$$+ \frac{1}{\bar{C}} \left[\frac{2q\epsilon_s N_a N_d (\Phi_{bi} + V_{SB})}{(N_d + N_a)} \right]^{1/2}$$

where

$$\frac{1}{\bar{C}} = \frac{1}{C_{OX}} + \frac{X_1}{2\epsilon_s} \qquad (17)$$

and by El-Mansey [39] ;

$$V_{TD} = V_{FB} + \frac{N_d}{(N_d + N_a)} (\Phi_{bi} + V_{SB}) \qquad (18)$$

$$+ \frac{1}{C_1} \left[\frac{2q\epsilon_s N_a N_d (\Phi_{bi} + V_{SB})}{(N_d + N_a)} \right]^{1/2} - \frac{qN_d X_j}{C_2}$$

where

$$\frac{1}{C_1} = \frac{1}{C_{OX}} + \frac{X_j}{\epsilon_s} \quad \text{and} \qquad (19)$$

$$\frac{1}{C_2} = \frac{1}{C_{OX}} + \frac{X_j}{2\epsilon_s} \qquad (20)$$

Huang [64] considered the channel saturation effect at the drain side of the IGFET to get the value of $\bar{C}$. Also, in the BCCD application, Haken [20] wrote the same equation in the different form of

$$V_{TD} = V_{FB} + \Phi_{bi} + V_{SB} \qquad (21)$$

$$- \frac{qN_d X_j^2}{2\epsilon_s} \left\{ \left[1 + \frac{\epsilon_s}{C_{OX} X_j} - \frac{1}{X_j} \left(\frac{2\epsilon_s N_a (\Phi_{bi} + V_{SB})}{qN_d (N_d + N_a)} \right)^{1/2} \right]^2 - (\frac{\epsilon_s}{C_{OX} X_j})^2 \right\}$$

Even though they are expressed differently, (16), (17) and (21) are all identical.

APPENDIX II

Intrinsic Carrier Concentration and Mobility

1. Effective Mass

An electron in a periodic potential is accelerated relative to the lattice in an applied electric or magnetic field as if the mass of the electron were equal to an effective mass. The motion of an electron in an energy band is described as a wave packet in an applied electric field. Suppose that the wave packet is made up of the wave functions near a particular wave vector $\bar{k}$, the group velocity v_g of the wave packet is defined as

$$v_g = \frac{d\omega}{dk} = \frac{1}{\hbar} \frac{dE}{dk} \tag{1}$$

where $\hbar$ is the Planck constant $(h/2\pi)$, ω is the angular frequency, and E is the energy. Therefore, the time derivative of v_g yields

$$\frac{dv_g}{dt} = \frac{1}{\hbar} \frac{d^2E}{dkdt} = \frac{1}{\hbar} \frac{d^2E}{dk^2} \frac{dk}{dt} \ . \tag{2}$$

Since $\dfrac{dk}{dt} = \dfrac{F}{\hbar}$ holds, the force F is given by

$$F = \frac{\hbar^2}{\dfrac{d^2E}{dk^2}} \frac{dv_g}{dt} \tag{3}$$

which assumes the form of the Newton's second law. We define the effective mass m^* by

$$\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{d^2E}{dk^2} \ . \tag{4}$$

2. Density of States

The energy band structure , i.e., the energy-momentum relationship of a crystalline solid is obtained by the solution of the Schroedinger equation. The Bloch theorem states

that if a potential energy in a crystal is periodic with the period of the lattice, the wave functions satisfying the free-particle Schroedinger equation and the periodicity condition are of the form of a traveling wave ;

$$\psi(\bar{r}) = \exp(i\bar{k}\cdot\bar{r}) \tag{5}$$

where $\bar{r}$ is a space vector, the wave vector $\bar{k}$ satisfies

$$k_x = 0, \ \pm\frac{2\pi}{L}, \ \pm\frac{4\pi}{L}, \ \cdots \ ,$$

and similarly for k_y and k_z. L is the period of the x, y and z coordinates.

In the ground state of a system of N free electrons, the occupied states may be represented as the points inside a sphere in the $\bar{k}$ space. The energy at the surface of the sphere is the Fermi energy E_f. The wave vector at the Fermi energy surface has a magnitude k_f such that

$$E_f = \frac{\hbar^2 k_f^2}{2m^*} \ . \tag{6}$$

The total number of states in the volume of $4\pi k_f^3/3$ for the volume of $(2\pi/L)^3$ of the k-space is

$$N = 2\frac{\frac{4\pi k_f^3}{3}}{(\frac{2\pi}{L})^3} \tag{7}$$

where the factor of 2 represents the two allowed values of the spin quantum number. Using (7), the total number of states, N, is related to the energy $(\leqslant E)$ by

$$E = \frac{\hbar^2}{2m^*} (\frac{3\pi^2 N}{V})^{2/3} \tag{8}$$

where $V=L^3$. Therefore, the number of states per unit energy, $D(E)$, is from (8)

$$D(E) = \frac{1}{V}\frac{dN}{dE} = \frac{1}{2\pi^2}(\frac{2m^*}{\hbar^2}) E^{1/2} \ . \tag{9}$$

3. Intrinsic Carrier Concentration

The basic consideration of populating the allowed energy states with particles subject to the Pauli's exclusion principle leads to a distribution function for electrons that is called the Fermi-Dirac distribution of

$$f(E) = \frac{1}{1 + \exp(\frac{E - E_f}{kT})} \tag{10}$$

where E_f is a reference energy called the Fermi energy or level. The number of electrons excited to the conduction band at T can be calculated in terms of the Fermi energy E_f. The energy E is measured from the top of the valence band as shown in Fig. A.1. If $E - E_f >> kT$ is assumed for the conduction band of a semiconductor at the temperature of interest, the Fermi-Dirac distribution function reduces to the Maxwell-Boltzmann distribution of

$$f(E) = \exp(-\frac{E - E_f}{kT}) \ . \tag{11}$$

This is valid only when $f << 1$ and represents the probability that a conduction electron state is occupied. The Boltzmann function applies to the case that any number of electrons may exist in the allowed state.

The energy of an electron in the conduction band is

$$E = E_g + \frac{\hbar^2 k^2}{2m^*_{de}} \tag{12}$$

where m^*_{de} is the density-of-states effective mass of an electron. The same relationship applies to a hole with the density-of-states effective mass m^*_{dh}. The density of states at E is from (9)

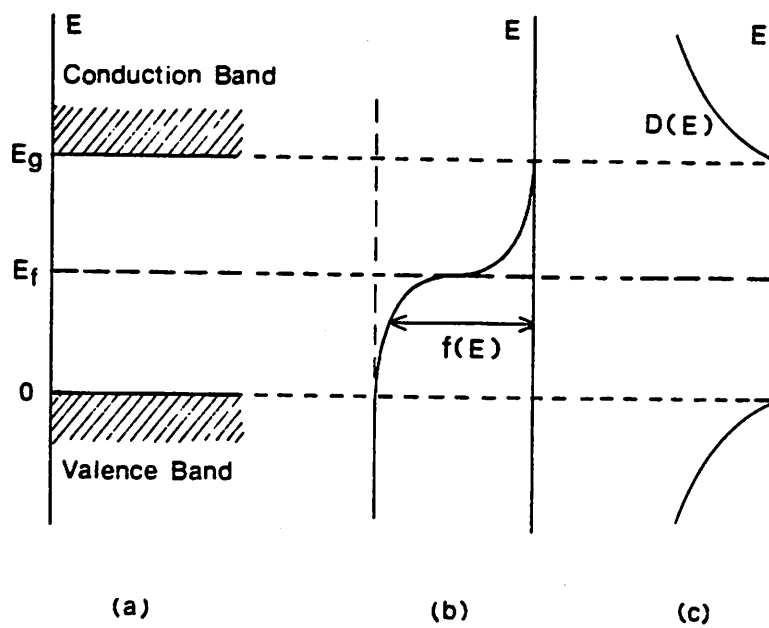$$D(E) = \frac{1}{2\pi^2}(\frac{2m^*_{de}}{\hbar^2})^{3/2} (E - E_g)^{1/2} \quad \text{for} \quad \text{electrons}$$

Fig. A.1. (a) Energy Band, (b) Fermi distribution function
and (c) density of state function.

$$= \frac{1}{2\pi^2}(\frac{2m^*_{dh}}{\hbar^2})^{3/2} (-E)^{1/2} \quad \text{for} \quad \text{holes} \quad . \tag{13}$$

Therefore, the electron concentration in the conduction band is given by

$$n = \int_{E_g}^{\infty} D(E)f(E)dE \tag{14}$$

$$= 2(\frac{m^*_{de}kT}{2\pi\hbar^2})^{3/2} \exp(\frac{E_f-E_g}{kT}) = N_c\exp(\frac{E_f-E_g}{kT}) \quad .$$

where N_c ($\approx 3.22\times10^{19}cm^{-3}$) is the effective density of states at the conduction band edge. Similarly, the concentration of holes in the valence band is

$$p = \int_{-\infty}^{0} D(E)[1-f(E)]dE \tag{15}$$

$$= 2(\frac{m^*_{dh}kT}{2\pi\hbar^2})^{3/2} \exp(-\frac{E_f}{kT}) = N_v\exp(-\frac{E_f}{kT}) \quad .$$

where N_v ($\approx 1.83\times10^{19}cm^{-3}$) is the effective density of states at the valence band edge. We multiply the expressions for n and p to obtain the equilibrium relation of

$$n_i^2 = np = 4(\frac{kT}{2\pi\hbar^2})^3(m^*_{de}m^*_{dh})^{3/2} \exp(-\frac{E_g}{kT}) \tag{16}$$

$$\approx 1.5\times10^{33}T^3\exp(-\frac{1.21}{kT})$$

where the approximation is from Morin [70].

The above equation holds in the presense of impurities as far as dopant densities are moderate enough for the Maxwell-Boltzmann distribution function to be valid.

4. Mobility

At a low electric field, the drift velocity $\overline{v}_d$ is proportional to the electric field strength $\overline{E}$ and the proportionality constant is defined as the mobility $(cm^2/V-sec)$ such that

$$\bar{v}_d = \mu \bar{E} \ . \tag{17}$$

For the nonpolar semiconductors such as Si and Ge, there are two scattering mechanisms which significantly affect the mobility. They are the scatterings due to the lattice vibration and to the ionized impurities. The mobility due to the lattice scattering μ_l is given by [42]

$$\mu_l = \frac{(8\pi)^{1/2} q \ \hbar^4 c_{11}}{3 E_{1g}^2 m^{*5/2} (kT)^{3/2}} \ . \tag{18}$$

where c_{11} $<100>$ is the longitudinal elastic constant which is approximately $1.67 \times 10^{12} dyne - cm^{-2}$, E_{1g} is the displacement of energy gap edge per unit lattice dilation, and m^* is the conductivity effective mass. The mobility due to the ionized impurities μ_i is given by [71]

$$\mu_i = \frac{124(2\pi)^{1/2} \epsilon_s^2 (kT)^{3/2}}{N_i q^3 m^{*1/2} \ln(1 + \frac{12\pi \epsilon_s kT}{q^2 N_i^{1/3}})^2} \ . \tag{19}$$

where N_i is the ionized impurity density and ϵ_s is the permittivity of Si.

The combined mobility due to the above two scattering mechanisms is

$$\mu = \frac{1}{\frac{1}{\mu_l} + \frac{1}{\mu_i}} \ . \tag{20}$$

(19) shows that the mobility decreases as the impurity concentration increases. Also for a given impurity concentration, the electron mobility in (18) and (19) is larger than the hole mobility because the effective mass of electrons is lighter than that of holes. In a lightly-doped material, the mobility resulting from the ionized impurities is much higher than that from the lattice vibration. The theoretical result (18) indicates that the mobility should decrease in proportion to $T^{-3/2}$ when the lattice scattering is dominant. Experimentally, mobilities varies from $T^{-1.5}$ to T^{-3}, with $T^{-5/2}$ being common.

APPENDIX III

Input-Output Spectral Density Relation in a Linear System

The autocorrelation of a process $X(t)$ is defined as

$$R_{xx}(\tau) = E\ X(t+\tau)X^*(t) \tag{1}$$

The spectral density $S(\omega)$ is the Fourier transform of its autocorrelation $R(\tau)$;

$$S(\omega) = \int_{-\infty}^{\infty} R(\tau)e^{-j\omega\tau}d\tau \quad . \tag{2}$$

That is, the Fourier inversion formula gives the autocorrelation function of

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)e^{j\omega\tau}d\omega \quad . \tag{3}$$

Then the average power (variance) of the process $X(t)$ is

$$E|X(t)|^2 = R(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)d\omega \quad . \tag{4}$$

Consider a linear system with an impulse response of $h(t)$ as shown in Fig. A.2(a), the output of the system due to the input $x(t)$ is given by

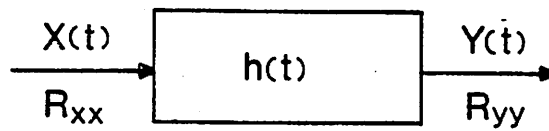$$y(t) = \int_{-\infty}^{\infty} x(t-\tau)h(\tau)d\tau \quad . \tag{5}$$

From (5), we have

$$x(t_1)y(t_2) = \int_{-\infty}^{\infty} x(t_1)x(t_2-\tau)h(\tau)d\tau \quad . \tag{6}$$
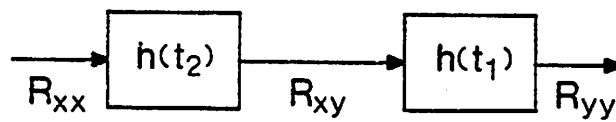
Taking the expected value of (6) yields

$$R_{xy}(t_1,\ t_2) = \int_{-\infty}^{\infty} R_{xx}(t_1,\ t_2-\tau)h(\tau)d\tau \tag{7}$$

Fig. A.2. Linear system with an impulse response h(t) (a) and
its equivalent system (b).

$$= R_{xx}(t_1, t_2) * h(t_2)$$

where $*$ stands for a convolution integral. From (5), we also have

$$y(t_1)y(t_2) = \int_{-\infty}^{\infty} x(t_1-\tau)y(t_2)h(\tau)d\tau \quad .$$

$$(8)$$

Also taking the expected value of (8) yields

$$R_{yy}(t_1, t_2) = \int_{-\infty}^{\infty} R_{xy}(t_1-\tau, t_2)h(\tau)d\tau$$

$$(9)$$

$$= R_{xy}(t_1, t_2) * h(t_1) \quad .$$

By combining (7) and (9), the autocorrelation of the output is given by a function of the autocorrelation of the input ;

$$R_{yy}(t_1, t_2) = R_{xx}(t_1, t_2) * h(t_2) * h(t_1)$$

$$(10)$$

$$= R_{xx}(t_1, t_2) * h(t_1) * h(t_2) \quad .$$

The interpretation of (10) is illustrated in Fig. A.2(b).

APPENDIX IV

Test Fixture

The following external circuits are necessary to test the prototype reference because the reference is clock-driven and the output V_{ref} is a pulse sequence :

1. Timing Circuit ; to supply three non-overlapping clocks.

2. Nulling Circuit ; to measure the pulse amplitude.

All circuits are made of CMOS IC's to avoid the interfacing problem between the analog circuit and the digital logic.

Fig. A.3(a) shows the timing circuit which generates three non-overlapping clocks with the duty cycles of 0.2, 0.2 and 0.3, respectively.   Fig. A.3(b) shows the nulling circuit. For a fast slew, a fast settle and a high input impedance, the JFET-input op amp LF356N is chosen.   Measuring procedures are as follows:

1. Adjust the offset of Op1 with S_1 off.

2. Adjust the offset of Op2 with S_2 off.

3. Measure V_{cal} by nulling V_{out}.

A diode is used to protect the op amp Op1 from being overdriven.
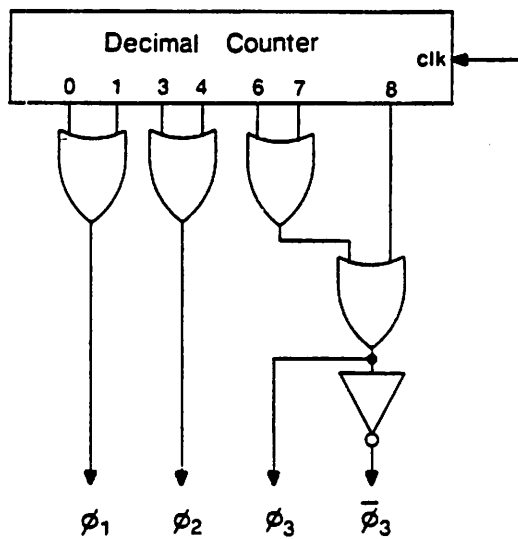
(a)
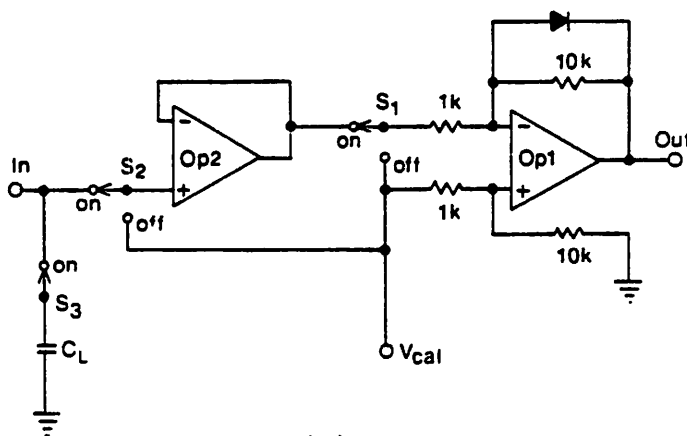
(b)

Fig. A.3. Test circuits : (a) timing circuit and (b)
nulling circuit (comparison circuit).

APPENDIX V

Self-Aligned Si-Gate CMOS Process [72]

1. Wafer Cleaning

2. Initial Oxidation (240nm)

    - 1000°C
      Push/N_2/5/3Min.
      Dry O_2/5/10Min.
      Wet O_2/4/25Min.
      Anneal/N_2/5/10Min.
      Pull/N_2/5/3Min.

3. Well Definition

    - Photolithography : Mask #1
    - Oxide etch : BHF/2Min. (HF/NH_4F=1/5)
    - Well implant : Phos./100KeV/8°/1.5×10^{12}
    - Well Drive : 1100°C
          Push/N_2/4/3Min.
          Dry O_2/15/280Min.
          1150°C
          Drive/N_2:O_2/3.5:15/720Min.
          Anneal/N_2/4/20Min.
          Pull/N_2/4/3Min.

4. Gate Oxidation (70nm)

    - Oxide etch : 7Min. (HF/DI=1/5)
    - Oxidation : 1000°C
          Push/N_2/4/3Min.
          Dry O_2/6.5/80Min.
          Anneal/N_2/4/10Min.
          Pull/N_2/4/3Min.
    - Nitride deposition : NH_3/SiH_4(600mT/100mT)/60Min.

5. NMOS Active Definition

    - Photolithography : Mask #2
    - Nitride etch : Preheat/N_2/1T/60W/70°C
          Descum/O_2/0.76T/10W/5Min.
          Etch/SF_6-O_2/0.1T/15W
          Purge/N_2/1T/3Min.
    - Field implant : Boron/100KeV/10^{13}
    - Backside implant : BF_2/200KeV/2×10^{15}
    - Drive : 1000°C

Push/N_2/4/3Min.
Dry O_2/6.5/20Min.
Anneal/N_2/4/5Min.
Pull/N_2/4/3Min.

6. PMOS Active Definition

- Photolithography : Mask #3
- Nitride etch

7. Local Oxidation (800nm)

- 920°C
  Push/N_2/3/3Min.
  Wet O_2/1/420Min.
  Anneal/N_2/4/20Min.
  Pull/N_2/4/3Min.

7. Capacitor Definition

- Photolithography : Mask #4
- Nitride etch
- Oxide etch : BHF
- Plasma bake : N_2/1T/60W/30Min.
- Capacitor implant : Phos./50KeV/8×10^{14}
- Photoresist removal : Ash/O_2/1T/100W/30Min.

8. Capacitor Oxidation (70nm)

- Oxidation : 1000°C
          Push/N_2/4/3Min.
          Dry O_2/6.5/80Min.
          Anneal/N_2/4/10Min.
          Pull/N_2/4/3Min.
- Nitride removal : Phos. Acid/155°C
- Threshold implant : Boron/50KeV/5×10^{11}

9. Poly Deposition and Doping

- Deposition : SiH_4/600mT/26Min.
- Doping : 950°C
          Push/N_2/4/3Min.
          Dry O_2:N_2/2.5:5/5Min.
          Add $POCL_3$/6/30Min.
          Anneal/N_2/4/5Min.
          Pull/N_2/4/3Min.

10. N Poly Definition

- Deglaze
- Photolithography : Mask #5
- Poly etch
- NMOS S/D implant : As/$180 KeV$/3×10^{15}

11. P Poly Definition

- Photolithography : Mask #6
- Poly etch
- Plasma bake
- PMOS S/D implant : Boron/$60 KeV$/2×10^{15}

12. Passivation

- CVD Deposition : Undoped/$250 nm$
  Doped/$750 nm$ (6% PSG oxide)
- Reflow : $1050° C$
  Push/N_2/4/3Min.
  Flow/O_2:$POCL_3$/2.5:6/10Min.
  Anneal/N_2/4/10Min.
  Pull/N_2/4/3Min.
- Densification : $900° C$
  Push/N_2/4/3Min.
  Densify/Wet O_2/4/30Min.
  Anneal/N_2/4/5Min.
  Pull/N_2/4/3Min.

13. Contact Definition

- Photolithography : Mask #7
- Oxide etch : Repeat etch (BHF) and bake (10min./$130° C$) cycle

14. Metallization ($1 \mu m$)

- Palladium evaporation
- Sintering : $300° C$/15/5Min.
- Palladium removal : Gold etchant/60sec
  NH_4I/60sec
  RCA1/$75° C$/5Min.
- Al evaporation

15. Metal Definition

- Photolithography : Mask #8
- Metal etch : Al Tpye A/$40° C$
- Backside Al evaporation ($1 \mu m$)
- Sintering : $350° C$/15/20Min.

REFERENCES

[1]   D. A. Hodges, P. R. Gray and R. W. Brodersen," Potential of MOS Technologies for Analog Integrated Circuits," *IEEE J. Solid-State Circuits* , vol. SC-13, pp. 258-294, June 1978.

[2]   R. J. Widlar," New Developments in IC Voltage Regulators," *IEEE J. Solid-State Circuits* , vol. SC-6, pp. 2-7, Feb. 1971.

[3]   K. E. Kujik," A Precision Reference Voltage Source," *IEEE J. Solid-State Circuits* , vol. SC-8, pp. 222-226, June 1973.

[4]   A. P. Brokaw," A Simple Three-Terminal Bandgap Reference," *IEEE J. Solid-State Circuits* , vol. SC-9, pp. 388-393, Dec. 1974.

[5]   C. R. Palmer and R. C. Dobkin," A Curvature Corrected Micropower Voltage Reference," *Int. Solid-State Circuits Conf.* , pp. 58-59, Feb. 1981.

[6]   G. C. M. Meijer and P. C. Schmale and K. van Zalinge," A New Curvature-Corrected Bandgap Reference," *IEEE J. Solid-State Circuits* , vol SC-17, pp. 1139-1143, Dec. 1982.

[7]   M. E. Hoff," MOS Reference Voltage Circuit," *U. S. Patent* , No. 4100437, July 1978.

[8]   R. A. Blauschild, T. A. Tucci, R. S. Muller and R. G. Meyer," A New NMOS Temperature-Stable Voltage Reference," *IEEE J. Solid-State Circuits* , vol. SC-13, pp. 767-774, Dec. 1978.

[9]   F. P. Heiman and H. S. Miller," Temperature Dependence of n-Type MOS Transistors," *IEEE Trans. Electron Devices* , vol. ED-12, pp. 142-148, Mar. 1965.

[10]   L. Vadez and A. S. Grove," Temperature Dependence of MOS Transistor Characteristics below Saturation," *IEEE Trans. Electron Devices* , vol. ED-13, pp. 863-866, Dec. 1966.

[11]   R. Wang, J. Dunkley, T. A. DeMassa and L. F. Jelsma," Threshold Voltage Variations with Temperature in MOS Transistors," *IEEE Trans. Electron Devices* , vol. ED-18, pp. 386-388, June 1971.

[12]   F. H. Gaensslen and R. C. Jaeger," Temperature Dependent Threshold Behavior of Depletion Mode MOSFETs, Characterization and Simulation," *Solid-State Electronics* , vol. 22, pp. 423-430, 1979.

[13]   R. C. Jaeger and F. H. Gaensslen," Simple Analytic Models for the Temperature Dependent Threshold Behavior of Depletion-Mode Devices," *IEEE Trans. Electron Devices* , vol. ED-26, pp. 501-507, April 1979.

[14]   Y. P. Tsividis and R. W. Ulmer," A CMOS Voltage Reference," *IEEE J. Solid-State Circuits* , vol. SC-13, pp. 774-778, Dec. 1978.

[15]   E. A. Vittoz and O. Neyroud," A Low-Voltage CMOS Bandgap Reference," *IEEE J. Solid-State Circuits* , vol. SC-14, pp. 573-577, June 1979.

[16]   Y. P. Tsividis and R. W. Ulmer," A CMOS Voltage Reference," *IEEE J. Solid-State Circuits* , vol. SC-13, pp. 774-778, Dec. 1978.

[17]   E. A. Vittoz and O. Neyroud," A Low-Voltage CMOS Band-Gap Reference," *IEEE J. Solid-State Circuits* , vol. SC-14, pp. 573-577, June 1979.

[18]   M. C. Tobey, D. J. Gialiani and P. B. Askin," Flat-Band Voltage Reference," *U. S. Patent* , No. 3975648, Aug. 1976.

[19]   H. J. Oguey and B. Gerber," MOS Voltage Reference Based on Polysilicon Gate Work Function Difference," *IEEE J. Solid-State Circuits* , vol. SC-15, pp. 264-269, June 1980.

[20]   M. R. MacPherson," Threshold Shift Calculations for Ion-Implanted MOS Devices," *Solid-State Electronics* , vol. 15, pp. 1319-1328, 1972.

[21]   V. L. Rideout, F. H. Gaensslen and A. LeBlanc," Device Design Considerations for Ion-Implanted n-Channel MOSFETs," *IBM J. Res. Develop.* , pp. 50-58, Jan. 1975.

[22]   R. M. Swanson and J. D. Meindl," Ion-Implanted Complementary MOS Transistors in Low-Voltage Circuits," *IEEE J. Solid-State Circuits* , vol. SC-7, pp. 146-153, April 1972.

[23]   T. W. Sigmon and R. Swanson," MOS Threshold Shifting by Ion-Implantation," *Solid-State Electronics* , vol. 16, pp. 1217-1232, 1973.

[24]   J. R. Edwards and G. Marr," Depletion Mode IGFET Made by Deep Ion-Implantation," *IEEE Trans. Electron Devices* , vol. ED-20, pp. 283-298, Mar. 1973.

[25]   T. Tanaka," Shift of Gate Threshold Voltage of MOS Transistors due to the Introduction of Shallow Impurities," *Japan J. Applied Physics* , vol. 10, pp. 84-91, Jan. 1971.

[26]   R. Troutman," Ion-Implanted Threshold Tailoring for IGFETs," *IEEE Trans. Electron Devices* , vol. ED-24, pp. 182-192, Mar. 1977.

[27]   P. P. Wang and O. S. Spencer," Threshold Voltage Characteristics of Double-Boron-Implanted Enhancement-Mode MOSFETs," *IBM J. Res. Develop.* , pp. 530-538, Nov. 1975.

[28]   R. S. C. Cobbold," Temperature Effects in MOS Transistors," *Electronics Letters* , voi. 2, pp. 190-191, June 1966.

[29]   G. Giralt, B. Andre, J. Simonne and D. Esteve," Influence de la temperature sur les dispotifs semiconducteur du type," *Electronics Letters* , vol. 1, pp. 185-186, Sept. 1965.

[30]   H. C. De Graff and J. A. V. Nielen," Temperature Influence of Channel Conductance of MOS Transistors," *Electronics Letters* , vol. 3, pp. 195-196, May 1967.

[31]   W. E. Howard and F. F. Fang," Low Temperature Effects in Si FETs," *Solid-State Electronics* , vol. 8, pp. 82-83, 1965.

[32]   C. G. Rogers and A. K. Jonscher," Operation of Field Effect Transistors at Liquid Helium Temperature," *Electronics Letters* , vol. 3, pp. 210-211, May 1967.

[33]   H. C. Nathanson, C. Jund and J. Grosvalet," Temperature Dependence of Apparant Threshold Voltage of Silicon MOS Transistors at Cryogenic Temperature," *IEEE Trans. Electron Devices* , vol. ED-15, pp. 362-368, June 1968.

[34]   S. Horiuchi," Threshold Voltage Shift of n-Channel Si-Gate MOSFETs," *IEEE Trans. Electron Devices* , vol. ED-22, pp. 1083-1043, Nov. 1975.

[35]   M. Kamoshida," Electrical Characteristics of Boron Implanted n-Channel MOS Transistors," *Solid-State Electronics* , vol. 17, pp. 621-626, 1974.

[36]   W. Schemmert, L. Gabler and B. Hoefflinger," Conductance of Ion-Implanted Buried Channel MOS Transistors," *IEEE Trans. Electron Devices* , vol. ED-23, pp. 1313-1319, Dec. 1976.

[37]   T. Masuhara and J. Etoh," Low-Level Currents in Ion-Implanted MOSFET," *IEEE Trans. Electron Devices* , vol. ED-21, pp. 799-807, Dec. 1974.

[38]   S. E. Hansen, A. G. Gonzalez, D. A. Antoniodis and R. W. Dutton," SUPREM," *Stanford University.*

[39]   O. Leistiko, A. S. Grove and C. T. Sah,"Electron and Hole Mobilities in Inversion Layers on thermally Oxidized Silicon Surfaces," *IEEE Trans. Electron Devices* , vol. 12, pp. 248-254, May 1965.

[40]   W. H. White, D. R. Lampe, F. C. Blaha and I. A. Mack," Characterization of Surface Channel CCD Image Arrays at Low Light Levels," *IEEE J. Solid-State Circuits* , vol. SC-9, pp. 1-14, Feb. 1974.

[41]   J. W. Slotboom and H. C. DeGraaff," Measurements of bandgap narrowing in Si bipolar transistors," *Solid-State Electronics* , vol. 19, pp. 857-862, 1976.

[42]   J. Bardeen and W. Shockley," Deformation-Potentials of Mobilities in Non-Polar Crystals," *Phys. Rev.* , vol. 80, pp. 72-80, Oct. 1950.

[43]   K. L. Shaklee and R. E. Nahory *Phys.Rev. Lett.* , vol. 24, pp. 942, 1970.

[44]   H. Y. Fan,"Temperature Dependence of Energy Gap in Semiconductors," *Phys. Rev.* , vol. 82, pp. 900-905, June 1951.

[45]   G. G. Macfarlane, T. P. Mclean, J. E. Quarrington and V. Roberts," Fine Structure in the Absorption-Edge Spectrum of Si," *Phys. Rev.* , vol. 111, pp. 1245-1254, Sept. 1958.

[46]   Y. P. Varshni," Temperature Dependence of the Energy Gap in Semiconductors," *Physica* , vol. 34, pp. 149-154, 1967.

[47]   C. D. Thurmond," The Standard Thermodynamic Functions for the Formation of Electrons and Holes in Ge, Si, GaAs and GaP," *J. Electrochem. Soc. Solid-State Sci. Tech.* , vol. 45, pp. 1846-1848, 1974.

[48]   W. Bludau, A Onton and W. Heinke," Temperature Dependence of the Bandgap in Si," *J. Appl. Phys.* , vol. 45, pp. 1846-1848, 1974.

[49]   Y. P. Tsividis," Accurate Analysis of Temperature Effects in $I_C - V_{BE}$ Characteristics with Application to Bandgap Reference Source," *IEEE J. Solid-State Circuits* , vol. SC-15, pp. 1076-1084, Dec. 1980.

[50]   P. R. Gray and R. G. Meyer," Analysis and Design of Analog Integrated Circuits," *New York, Wiley* , pp. 84, 256, 1977.

[51]   W. L. Kauffman and A. A. Bergh," The Temperature Dependence of Ideal Gain in Double Diffused Silicon Transistors," *IEEE Trans. Electron Devices* , vol. ED-15, pp. 732-735, Oct. 1968.

[52]   D. Buhanan," Investigation of Current-Gain Temperature Dependence in Silicon Transistors," *IEEE Tran. Electron Devices* , vol. ED-16, pp. 117-124, Jan. 1969.

[53]   H. J. J. De Man," The Influence of Heavy Doping on the Emitter Efficiency of a Bipolar Transistor," *IEEE Tran. Electron Devices* , vol. ED-18, pp. 833-835, Oct. 1971.

[54]   R. U. Martinelli," The Temperature Dependence of the DC Base and Collector Currents in Silicon Bipolar Transistors," *IEEE Trans. Electron Devices* , vol ED-23, pp. 1218-1224, Nov. 1976.

[55]   J. L. McCreary, "Matching Properties, and Voltage and Temperature Dependence of MOS Capacitor," *IEEE J. Solid-State Circuits* , vol. SC-16, pp. 608-616, Dec. 1981.

[56]   R. Gregorian, G. A. Wegner and W. E. Nicholson, Jr.," An Integrated Single-Chip PCM Voice Codec with Filters," *IEEE J. Solid-State Circuits* , vol. SC-16, pp. 322-333, Aug. 1981.

[57]   J. S. Brugler," Silicon Transistor Biasing for Linear Collector Current Temperature Dependence," *IEEE J. Solid-State Circuits* , vol. SC-2, pp. 57-58, June 1967.

[58] S. M. Sze," Physics of Semiconductor Devices," *Wiley, New York* , pp. 102-104, 1969.

[59] R. W. Brodersen and S. P. Emmons," Noise in Buried Channel Charge-Coupled Devices," *IEEE j. Solid-State Circuits* , vol. SC-11, pp. 147-155, Feb. 1976.

[60] R. J. Kansy," Response of a Correlated Double Sampling Circuit to $1/f$ Noise," *IEEE J. Solid-State Circuits* , vol. SC-15, pp. 373-375, June 1980.

[61] D. B. Estreich," The Physics and Modeling of Latch-Up and CMOS Integrated Circuits," *Stanford Electronics Laboratories* , Tech. Report No. G-201-9, Nov. 1980.

[62] R. D. Thornton, D. DeWitt, P. E. Gray and E. R. Chenette," Characteristics and Limitations of Transistors," *New York, Wiley* , pp. 27-31, 1966.

[63] J. S. T. Huang, "Characteristics of Depletion-Type IGFET," *IEEE Trans. Electron Devices* , vol. ED-20, pp. 513-514, May 1973.

[64] J. S. T. Huang and G. W. Taylor," Modeling of an Silicon-Gate Depletion -Mode IGFET," *IEEE Trans. Electron Devices* ,vol. ED-22, pp. 995-1001, Nov. 1975.

[65] Y. A. El-Mansey," Analysis and Characterization of the Depletion-Mode IGFET," *IEEE J. Solid-State Circuits* , vol. SC-15, pp. 331-340, June 1980.

[66] J. S. T. Huang," On the Design of Ion-Implanted Buried Channel Charge Coupled Devices (BCCDs)," *Solid-State Electronics* , vol. 20, pp. 665-669, 1977.

[67] R. A. Haken," A Comparison of Static Bulk-Channel CCD Characteristics using Uniform, Gaussian and Measured Impurity Distributions," *Solid-State Electronics* , vol. 20, pp. 789-797, 1977.

[68] R. A. Haken," Analysis of Deep Depletion MOSFET and the Use of the DC Characteristics for Determining Bulk Channel CCD Parameters," *Solid-State Electronics* , vol. 21, pp. 753-761, 1978.

[69] C. Kittel," Introduction to Solid-State Physics," *New York, Wiley* , pp. 157-233, 1976.

[70] F. J. Morin and J. P. Maita," Electrical Properties of Silicon Containing Arsenic and Boron," *Physical Review*

[71] E. Conwell and V. F. Weisskopf," Theory of Impurity Scattering in Semiconductors," *Phys. Rev.* , vol. 77, pp. 388-390, Feb. 1950. , vol. 96, pp. 28-35, Oct. 1954.

[72] R. Kaneshiro and T. Choi," Berkeley CMOS Process," *U. C. Berkeley* , 1982.