Ultra-Low-Power Always-On Blocks in GF22nm Technology

Hasti Fattahi

August 2024

A thesis submitted to the Delft University of Technology
in partial fulfillment of the requirements for the degree of
Master of Science in Electrical Engineering

Supervisors:
Prof. Dr. K.A.A. Makinwa
Dr. D. Ruffieux
Dr. C.A.G. Salazar

TU Delft

csem

Abstract

Internet-of-Things (IoT) applications require nanowatt (nW) power references that are robust to process, voltage, and temperature (PVT) variations. This thesis presents the design of ultra-low-power (ULP) sub-10nW always-on blocks in GlobalFoundries 22nm (GF22nm) technology, including a Proportional to Absolute Temperature (PTAT) current reference, a bandgap reference, and a Low Dropout Regulator (LDO). These references are optimized to operate over the full automotive temperature range while consuming only 1nA of current per branch.

Given the high cost of GF22nm technology, achieving area efficiency is a critical aspect of this research. To address this, the design incorporates area-efficient components such as switched capacitors and duty-cycled resistors. The PTAT block achieves a line sensitivity of 2%/V and 5% spread (σ/μ) at 27°C consuming 4nW by utilizing MOSFETs in weak inversion and operates with an 800mV supply voltage while occupying a silicon area of 0.001mm^2. The bandgap reference is supplied from a battery with an end-of-life (EOL) voltage of 900mV. It achieves a maximum temperature coefficient (TC) of 140.6ppm/°C and a line sensitivity of 0.56%/V at 27°C with a supply range from 900mV to 1.98V. Without any trimming, the reference voltage spread due to process and mismatch variations is reduced to 2.9% (σ/μ) by using BJTs. The bandgap reference occupies a silicon area of 0.021mm^2 using duty-cycled resistors and has a nominal power consumption of 7.6nW. This voltage is used as the reference voltage for an LDO with unity-gain feedback to prevent multiplication of the reference voltage noise. The LDO maintains an output voltage line sensitivity of less than 1%/V with battery voltage variations from 900mV to 1.98V and load currents ranging from 100nA to 1µA.

Acknowledgements

Pursuing my Master's degree at TUDelft has been a pivotal experience that has profoundly shaped my journey as an engineer. The challenging academic environment and the innovative culture at the university have equipped me with essential skills and knowledge that I will carry forward in my career. I am especially grateful to TUDelft for the scholarship that made this experience possible, and I extend my sincere thanks to the faculty and staff for their support and encouragement throughout my time here.

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Kofi Makinwa. His unwavering support, patience, and invaluable guidance have been instrumental throughout the course of my thesis. I consider myself incredibly fortunate to have had the opportunity to work under his supervision, and it has been a great privilege to benefit from his exceptional knowledge and expertise.

I also had the privilege of collaborating with an outstanding team of researchers and mentors at CSEM SA. I am particularly grateful to Dr. David Ruffieux and Dr. Camilo Salazar, my supervisors at CSEM SA, for their warm welcome into their group and for the insightful mentorship they provided. My sincere thanks also go to all the members of the group, with special appreciation to our group leader, Dr. Paulo Augusto, for fostering an inspiring and collaborative environment.

Lastly, I would like to express my heartfelt thanks to my family. To my parents, whose endless love, support, and sacrifices have been the foundation of everything I have achieved—thank you. To my mother, your unwavering belief in me and your constant willingness to listen to my concerns and support me through every challenge have given me the strength to keep going, even when things were tough. And to my brother, Hooman, your constant support and belief in me have been a huge source of encouragement. The appreciation I have for you goes beyond words. You have always been, and will always be, the role model in my life. I am so grateful for the inspiration you have given me, pushing me to always do my best. For all of this, I am truly thankful.

Contents

List of Figures

List of Tables

1 Introduction

The advance of the Internet of Things (IoT) has greatly increased the number of connected devices, many of which operate in energy-constrained environments such as wearable technology, remote sensors, and mobile devices. These devices often require ultra-low power (ULP) integrated circuits that can function reliably for extended periods without recharging. Among these circuits, always-on components—such as real-time clocks (RTC), temperature sensors, and power management systems—are essential for maintaining continuous operation and system stability, even during deep sleep states where most of the system is powered down.

As these always-on components must remain active to support basic functions like timing, sensing, and low-power wake-up, they demand robust and energy-efficient voltage and current references. This is particularly challenging in advanced semiconductor processes such as GlobalFoundries 22nm Fully-Depleted Silicon-on-Insulator (FDSOI) technology. The primary goal of this thesis is to explore and develop ultra-low power, sub-10nW always-on blocks in GF22nm technology, focusing on Proportional to Absolute Temperature (PTAT) current references, bandgap (BG) references, and low-dropout (LDO) regulators.

In designing these ULP blocks, several challenges arise. The GF22nm technology presents unique advantages, such as reduced leakage currents, reduced parasitic capacitors, and wide body bias range, which help achieve the desired power efficiency. However, it also imposes constraints on area efficiency and component reliability, particularly in the face of temperature extremes common in automotive applications.

This research addresses these challenges through design strategies, including the use of switched-capacitor (SC) networks and duty-cycled resistors to replace traditional high-value resistors, which are typically area-inefficient at nA current levels.

The thesis provides a comprehensive exploration of design choices tailored to the constraints and opportunities of the GF22nm FDSOI technology, offering insights that can guide future developments in this and similar process nodes. Second, it demonstrates the feasibility of achieving sub-10 nW level power consumption in critical always-on blocks, such as PTAT and BG references and LDOs. The PTAT current reference and the bandgap reference designed in this work consume less than 4 nW and 12 nW respectively.

In conclusion, this research explores various solutions with a focus on BJTs for VREF generation and MOSFETs in subthreshold region for IREF generation. The results are expected to accelerate the design of ULP circuits in advanced process nodes, providing a robust foundation for the next generation of energy-efficient electronics.

2 Literature review

2.1 Ultra-low-power (nW) references

2.1.1 BJT based references

Conventionally, the design of bandgap references were limited to bandgap voltage (BG) references generated by adding the base-emitter voltage (V_{BE}) of a BJT to a resistor voltage with a PTAT current. This method produces the BG voltage of silicon (1.2V) with a zero temperature coefficient. Hence, the supply voltage should be higher than this voltage. In [15], an additional resistor was introduced, as shown in Figure 2.1, allowing for the generation of lower reference voltages. However, this approach has the disadvantage of adding another branch with a BJT and a third resistor, which is not area-efficient.

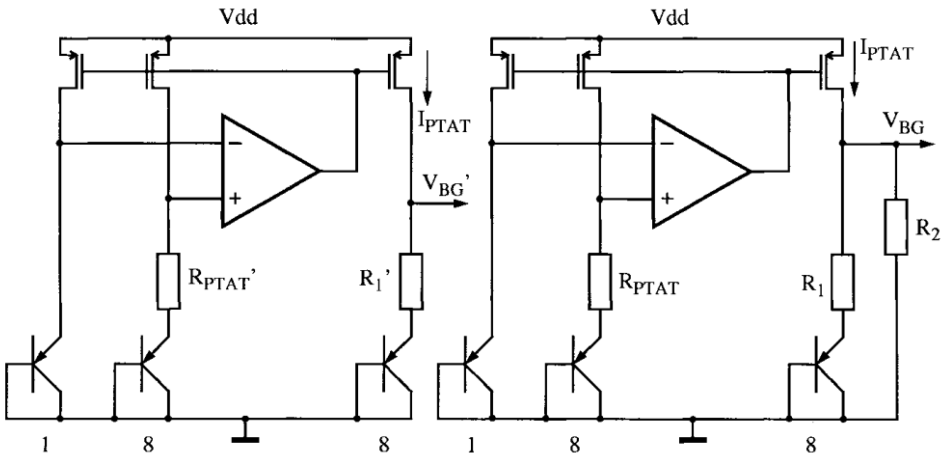


Figure 2.1: Conventional and low supply voltage bandgap circuits [15]

In [23], a 32 nW bandgap reference is introduced using charge pumps, BJTs, and most importantly, a switched capacitor network (SCN) to generate constants for scaling V_{EB} and ΔV_{BE}. As shown in Figure 2.2, V_{in} is 0.5V and the charge pump is required to generate a voltage above V_{EB}. Hence, with a 2x charge pump, the minimum required V_{in} is $V_{EB}/2$. Capacitor C_{L1} can be charged to $2 \times V_{IN}$ but Q_1 clamps it to V_{EB}. The difference between V_{EB1} and V_{EB2} is then stored on C_Δ. The SCN generates constants for scaling V_{EB} and ΔV_{BE} using a two-phase non-overlapping clock and the reference voltage V_{REF} is described as:

$$V_{REF} = 3\Delta V_{BE} + V_{BE1}\left(\frac{C_{a1}}{C_{a1} + C_{a2}}\right)$$

The capacitor values are selected to generate a temperature-independent constant V_{REF} of 500 mV at this work. The total power consumption is 32 nW and the area is 0.0264 mm^2.

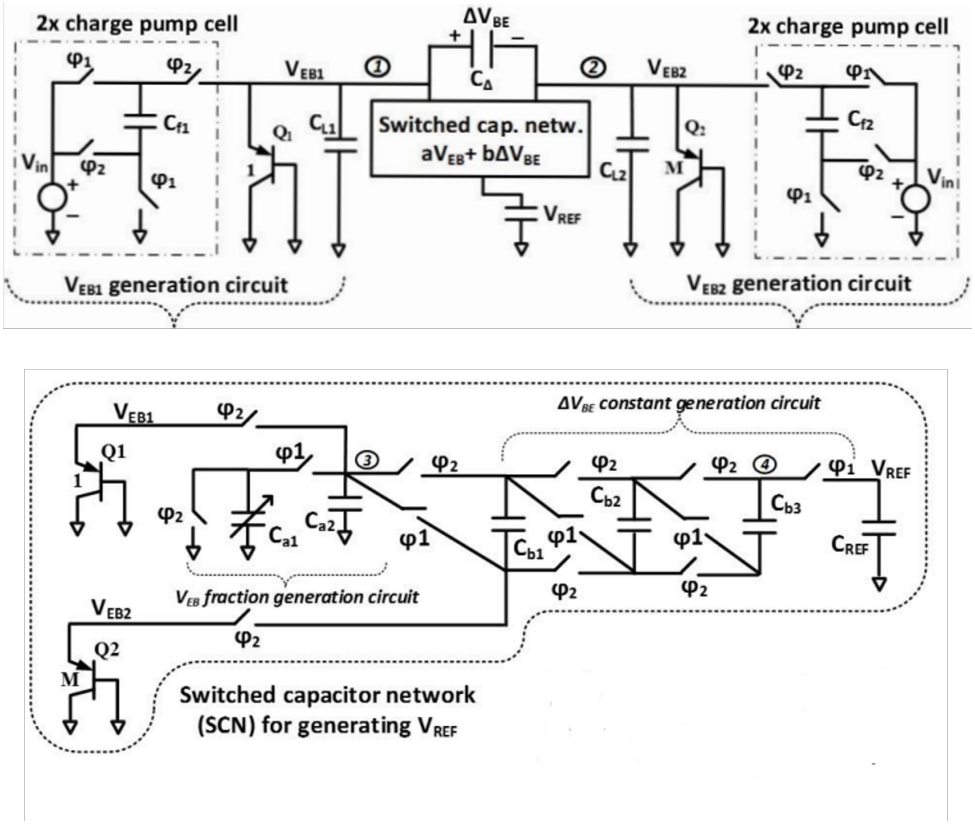


Figure 2.2: Bandgap reference circuit using charge pumps and a switched capacitor network [23].

This innovative approach reduced the supply voltage by half compared to other works utilizing BJTs and MOS current mirrors, by employing charge pumps. However, the nonlinearity of the BJT biasing and the leakage current in the proposed SCN limited the operating temperature range. This issue was addressed in [5] by implementing SCNs with low-leakage considerations. However, increasing the operating temperature range and lowering line sensitivity in [5] led to an area of 0.0522 mm^2.

In [3], a novel method is introduced that generates the reference current by combining proportional-to-absolute-temperature (PTAT) and complementary-to-absolute-temperature (CTAT) currents instead of voltages as shown in Figure 2.3. This current can then be distributed within the system to generate tunable sub-1V reference voltages. A primary advantage of this approach is its suitability for low supply voltages. Additionally, since the generation of voltage references depends on the ratio of resistors, process variations are effectively canceled out. However, the main disadvantages include the presence of more than two steady states due to parallel resistors R1 and R2 that shunt the diodes. Moreover, generating nA-range currents with this method requires a significantly large area due to large resistors.

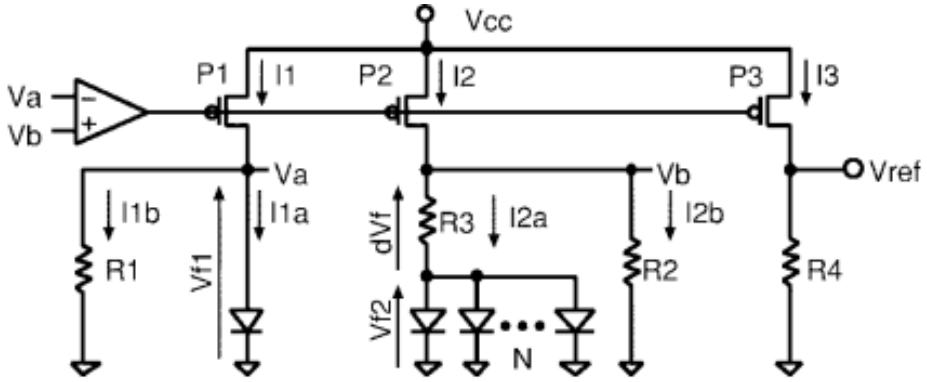


Figure 2.3: Schematic of the current and voltage reference with PTAT and CTAT currents [3]

In this work, the same topology is employed to generate CWT voltage and current reference. Instead of diodes, BJTs are used due to their improved matching and thermal stability. Additionally, the issues of unwanted steady states are mitigated and resistors are replaced with area-efficient components.

In [8], an innovative bandgap reference is proposed using capacitive bias applied to a pn-junction. Its work principle is based on the fact that when a pre-charged capacitor is discharged through a diode (shown in Figure 2.4) the resulting voltage over time can be obtained as:

$$V_D(t) = -nV_T \cdot \ln\left\{1 - \left[1 - \exp\left(\frac{-V_{dd}}{nV_T}\right)\right] \cdot \exp\left(\frac{-I_s}{C \cdot nV_T} \cdot t\right)\right\} \tag{2.1}$$

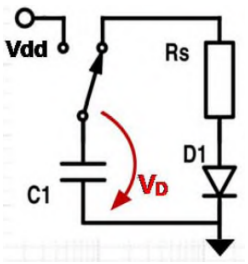


Figure 2.4: Capacitive bias of PN-junctions [8]

With a medium value of t, the above equation can be approximated as below, which varies with the natural log of time and does not depend on Vdd.

$$V_D(t) = -mV_T \cdot \ln\left(\frac{I_s}{C \cdot mV_T} \cdot t\right) \tag{2.2}$$

By precisely controlling the amount of time that the capacitor is discharged, a large dynamic range between the current densities of Equation 2.2 can be achieved, leading to a large current density ratio. This can be used to generate current ratios and a large PTAT voltage. The final Vref is achieved by switching capacitors in four phases (shown in Figure 2.5) with different sampling timings. The resulting Vref equation is derived as:

$$V_{\text{ref}} = \frac{C_2}{C_1 + C_2} \cdot \left[V_{d2} - \frac{C_1}{C_2} \cdot V_{d1}\right] \tag{2.3}$$

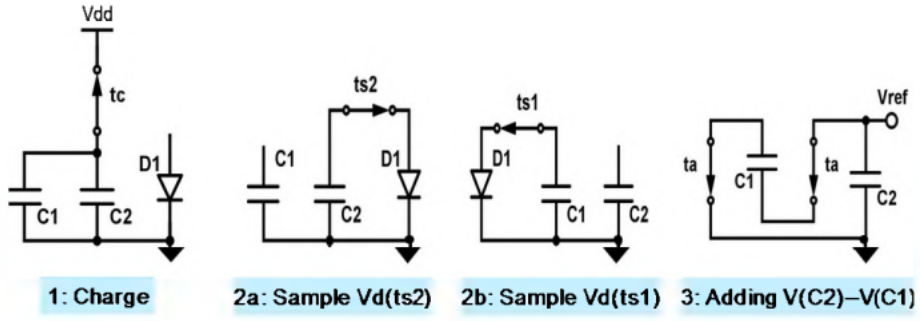


Figure 2.5: Capacitive switching phases to generate Vref [8]

This work operates with Vdd ranging from 0.85 V to 1 V and consumes 0.0022 mm^2 area.

Last but not least, a recent ultra-low-power voltage reference is presented in [7], fabricated using 22nm FDSOI technology. The proposed architecture, illustrated in Figure 2.6, bears similarities to the design in [16]. The operation principle relies on compensating the negative temperature coefficient of V_{CTAT} with a PTAT voltage generator. The PTAT voltage is generated using a resistor-less topology involving a differential pair with MOSFETs operating in the weak inversion region. To achieve voltage references lower than the silicon bandgap voltage (1.2 V), a voltage divider circuit is employed. Additionally, self-cascoded composites are used in the current mirrors to enhance the power supply rejection ratio (PSRR). Total power consumption is 45.6 nW with a 1.2V supply voltage and 0.0104 mm^2 active area.

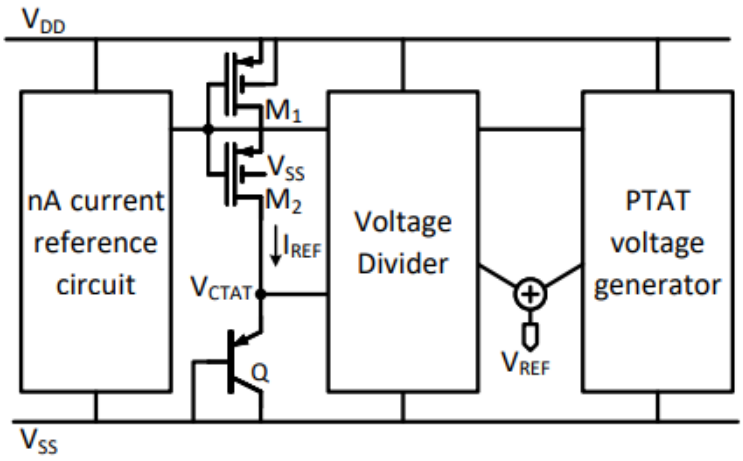


Figure 2.6: Proposed ultra-low-power voltage reference architecture in [7]

2.1.2 MOSFET based references

In [22], a native device with near-zero threshold voltage (V_{th}) and a thick oxide device with high V_{th} are used for transistors M_1 and M_2 respectively as shown in Figure 2.7. In practice, any combination of two devices with a significant difference in V_{th} can be used. The well-known subthreshold current is described by Equation 2.4, where μ represents mobility, C_{ox} is the oxide capacitance, and n is the subthreshold slope factor. By equating the currents through

M_1 and M_2, assuming both devices operate in weak inversion, the reference voltage V_{ref} can be derived, as shown in Equation 2.5. Since the MOSFET threshold voltage V_{th} is complementary to temperature, selecting appropriate transistor sizes can effectively cancel out the temperature dependence of the two terms. This design achieves pW-level power consumption with Vdd ranging from 0.5 V to 3 V over a narrow temperature range due to leakage currents. The area consumed in different processes is reported to be lower than 0.01 mm^2. To address process variations, a trimmable version is also proposed with 0.0093 mm^2 area.

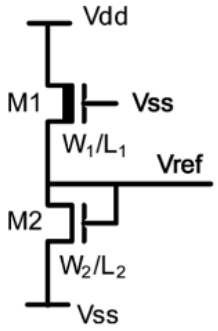


Figure 2.7: Two transistor (2T) voltage reference [22].

$$I_{sub} = \mu C_{ox} \frac{W}{L} (n-1) V_T^2 \exp\left(\frac{V_{gs} - V_{th}}{n V_T}\right) \left(1 - \exp\left(\frac{-V_{ds}}{V_T}\right)\right) \tag{2.4}$$

$$V_{ref} = \frac{n_1 n_2}{n_1 + n_2} (V_{th2} - V_{th1}) + \frac{n_1 n_2}{n_1 + n_2} V_T \ln\left(\frac{\mu_1 C_{ox1} W_1 L_2}{\mu_2 C_{ox2} W_2 L_1}\right) \tag{2.5}$$

On the other hand, when a current reference is required, the voltage reference must be converted to a current through a voltage-to-current converter and an op-amp-based feedback loop [13], as shown in Figure 2.8. Gate leakage transistors are used for generating picoampere (pA) current levels [25], and resistors are used for microampere (μA) current levels. However, for nanoampere (nA) current levels, as targeted in this project, using these components would consume a large chip area.

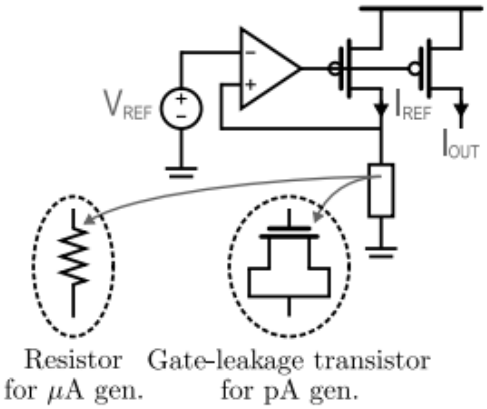


Figure 2.8: Current reference generation [13]

To achieve nA-range currents, self-cascoded MOSFETs are utilized in [13]. The design consumes an area of 0.0132 mm^2 and is fabricated in 22nm FDSOI technology. The authors propose a nanoampere-range constant-with-temperature (CWT) current reference that employs a self-cascoded MOSFET (SCM). The SCM is biased by a proportional-to-absolute-temperature (PTAT) voltage with a critical modification which is adding a CWT offset voltage to the PTAT bias obtaining $V_X = nU_T \log(K_{\text{PTAT}}) + \Delta V_T$ shown in Figure 2.9. This offset is derived from the threshold voltage difference between two transistors of the same type, with one being forward body-biased to reduce its threshold voltage. The current-voltage relationship of a SCM can be described as follows [6]:

$$I_D = I_S \cdot (i_f - i_r) \cdot S$$

where I_S represents the specific sheet current, $S = \frac{W}{L}$ is the transistor aspect ratio, and i_f and i_r are the forward and reverse inversion levels, respectively. By applying a constant-with-temperature (CWT) offset to the bias voltage of the SCM, i_f becomes complementary-to-absolute-temperature (CTAT). After tuning the offset, it effectively cancels out the proportional-to-absolute-temperature (PTAT) term in I_S. Lastly, to generate the offset based on the threshold voltage (V_{th}) difference between transistors M_6 and M_7, the body bias of M_6 is generated using a two-transistor (2T) voltage reference [22].

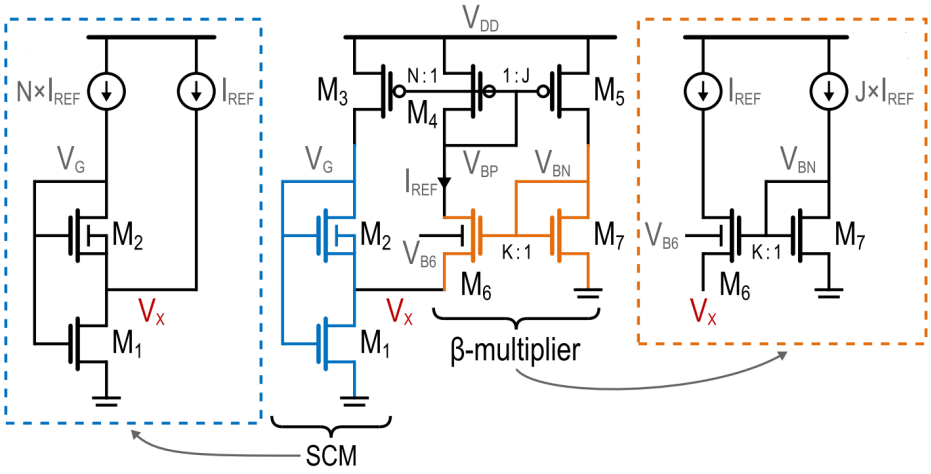


Figure 2.9: Self-cascoded MOSFETs used for nA-range reference current generation [13]

There are several limitations to the topology of Figure 2.9. First, the reference current I_{ref} is adversely affected by leakage currents, especially from the body parasitic p-well/n-well diode of M_2. Second, a temperature coefficient (TC) calibration circuit is necessary to ensure an acceptable TC across all process corners. Lastly, the value of the CWT offset depends only on technological parameters, offering no flexibility to tune V_{off} through transistor sizing. To address these issues, the topology shown in Figure 2.10 is proposed [12]. In this improved design, the body connection of M_2 is tied to ground to mitigate its leakage current, calibration is simplified by adjusting the width of either M_7 or M_9, and by using a four-transistor (4T) voltage reference, the offset voltage can be tuned by changing the ratio of S_7/S_6. The chip is fabricated in 0.11-μm bulk and 22-nm FDSOI technology and achieves a silicon area of 0.0106 mm^2 and 0.0026 mm^2 respectively.

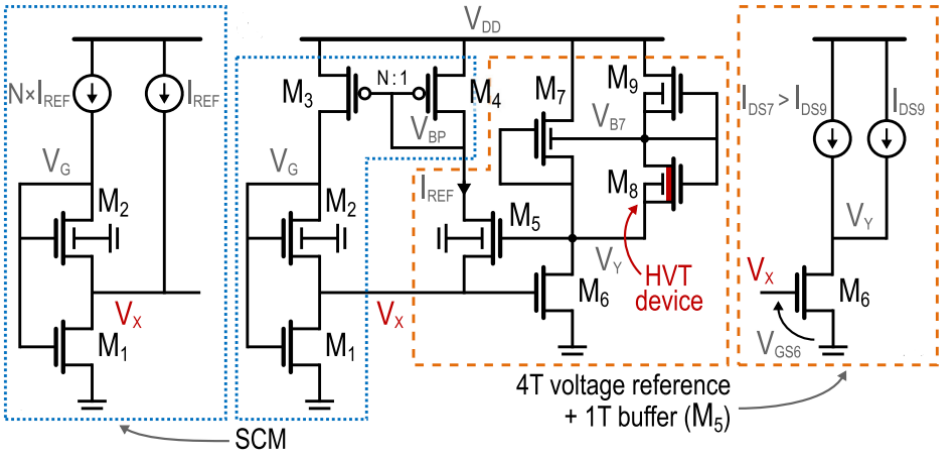


Figure 2.10: Schematic of the current reference with SCM and 4T voltage reference [12]

Both topologies for generating nA-range current references using SCMs [13] [12] are highly sensitive to leakage currents. Their performance has been reported to be reliable up to 85°C. However, beyond this range, particularly over the automotive temperature range, performance degrades significantly due to the increase in leakage currents.

2.1.3 Performance summary of the state of the art ULP references

	[23]	[8]	[13]	[12]	[5]	[22]	[7]
Technology	130nm	16nm FinFet	22nm FDSOI	22nm FDSOI	65nm	130nm	22nm FDSOI
Type	BJT+MOS	Diode+MOS	MOS	MOS	BJT+MOS	MOS	BJT+MOS
Vdd [V]	0.5 - 1.5	0.85 - 1	0.9 - 1.8	1 - 1.8	0.5	0.5 - 3	1.2 - 1.8
Vref [mV]	500	235	-	-	495	174.9	598
Iref [nA]	-	-	1.25	2.5	-	-	-
Power [nW]	32	40	7.8	15.5	38	0.022	45.6
Temp. Range [°C]	0 - 100	0 - 100	-40 - 85	-40 - 85	-40 - 120	-20 - 80	-40 - 120
TC [ppm/°C]	75	85	203	101	42	16.9 - 231	61
Line Sensitivity [%/V]	2	1.9	0.23	0.26	0.64	0.033	0.12
Area [mm^2]	0.0264	0.0022	0.0132	0.0025	0.0532	0.0013	0.0104
Spread (σ/μ) [%)]	0.67	0.82	6.66	4.7	1.03	0.85	0.37

Table 2.1: Performance summary of ULP reference designs

Among the prior art, the lowest area is achieved by the 2T voltage references in [22]. However, this design is limited by a narrow temperature range and a relatively high average temperature coefficient (TC). The current references presented in [13] and [12] are designed in 22nm FDSOI technology and manage to achieve low area and power consumption, though they exhibit a higher spread. In the category of BJT-based references, the design in [7] is notable for its implementation in 22nm FDSOI technology and operating over a wide temperature range. However, it requires a minimum Vdd of 1.2V. It is also shown that while BJT-based designs generally consume more area than MOSFET-only designs, they offer the advantage of operating over a broader temperature range. The performance analysis of this design with BJTs and area-efficient components instead of resistors will be shown in Table 6.1.

3 Proportional to absolute temperature (PTAT) current reference

The PTAT (Proportional To Absolute Temperature) current reference is a fundamental building block in analog and mixed-signal integrated circuits. PTAT current references are widely utilized in various applications, such as biasing amplifiers to achieve constant transconductance (g_m) and maintaining constant gain across temperature variations.

In this work, the goal is to design a ULP PTAT current reference that consumes a nominal current of 1 nA per branch. Achieving such a low current level per branch is challenging due to the leakage currents and area efficiency.

The following sections will briefly explain the design methodology, circuit implementation, and simulation results of the PTAT current reference.

3.1 Design methodology

3.1.1 Specifications

The design specifications are summarized in the Table 3.1:

Parameter	**Specification**
Supply Voltage	800 mV
Current per Branch	1 nA
Reference current	1 nA
Core Power Consumption	< 4 nW
Temperature Range	-40°C to 125°C
Area	$\leq 1200\,\mu m^2$

Table 3.1: PTAT Current Reference Design Specifications

The area is constrained to be less than $1200\,\mu m^2$, which is the area of the previous PTAT reference designed at CSEM in GF22nm technology generating a nominal current of $1\,\mu A$. Consequently, the goal of this project is to reduce the current consumption by a factor of 1000, while not exceeding the area of the previous design. This presents a challenge due to the significantly larger resistors required for this current scaling.

3.1.2 Design choices

The main decision in the PTAT reference involves selecting the appropriate MOSFETS. We have the option to choose between thin oxide and thick oxide transistors. Thin oxide transistors present a challenge due to gate and channel leakage currents, which can exceed 100 pA at high temperatures. An example of this is illustrated in Figure 3.1, which shows the exponential increase of channel leakage currents for temperatures above 85°C. Given our target output current of 1 nA, leakage currents exceeding 100 pA represent more than 10% of this, leading to temperature-related degradation in the core current.

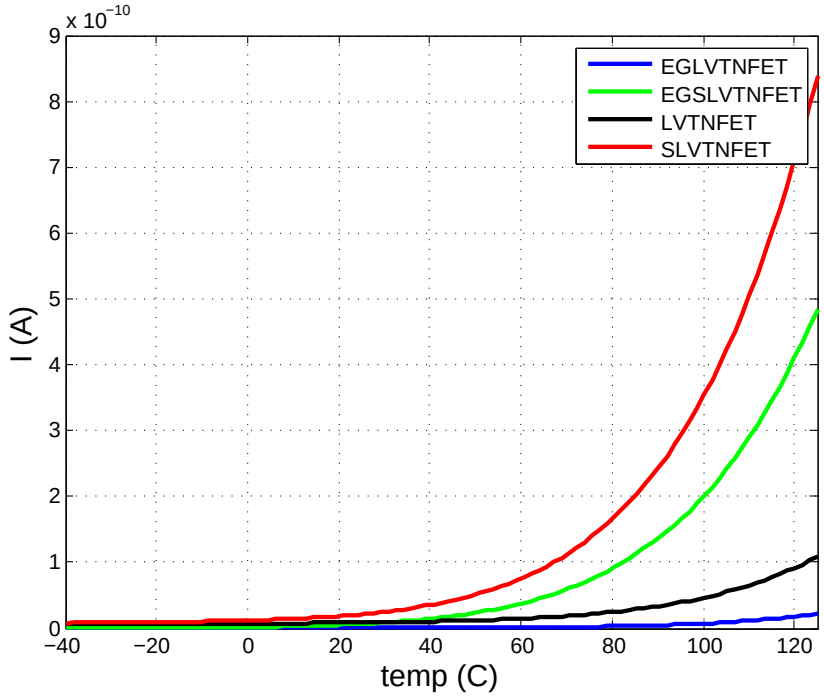


Figure 3.1: Channel leakage current as a function of temperature for unit transistors ($\frac{W}{L} = \frac{1u}{1u}$)

Therefore, thick oxide transistors must be used to mitigate leakage issues. Among the available options, super low Vt (EGSLVT) transistors were excluded due to their high leakage currents shown in Figure 3.1. Consequently, for all MOSFETs used in this design, thick oxide low Vt (EGLVT) transistors were selected to ensure robust performance.

3.1.3 Design steps

The topology of a simple beta-multiplier circuit is shown in Figure 3.2 [2]. MOSFETs in weak inversion (M1 and M2) are used instead of BJTs due to the low supply voltage of 800 mV to have more headroom for M3 and M4. If all the transistors are in weak inversion, the drain current given in Equation 2.4 is obtained as below in saturation [24]:

$$I_D = \frac{W}{L} I_S e^{\frac{V_{GS}-V_{th}}{nU_T}} \tag{3.1}$$

Voltage drop (VR) on Rptat is given by $V_{GS1} - V_{GS2}$, hence it can be written as [20]:

$$V_R = nU_T \ln \frac{\left(\frac{I_{D1}}{I_S\left(\frac{W}{L}\right)_1}\right)}{\left(\frac{I_{D2}}{I_S\left(\frac{W}{L}\right)_2}\right)} \approx U_T \ln(k) \tag{3.2}$$

where k is the ratio between the sizes of M_1 and M_2. This voltage is a PTAT voltage and generates a PTAT current equal to $\frac{V_R}{R_{\text{PTAT}}}$. After determining k, the value of R_{PTAT} can be obtained.

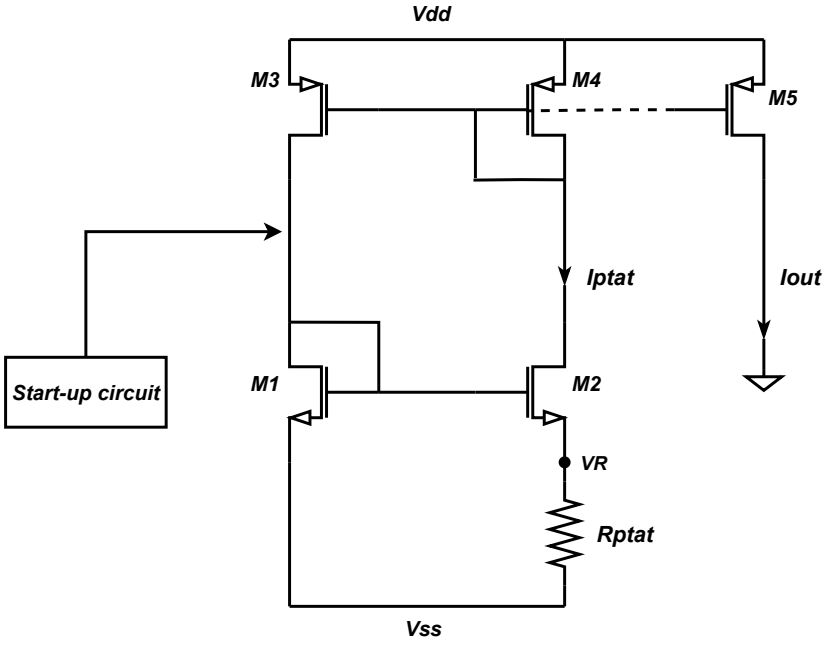


Figure 3.2: PTAT Current Reference Schematic

For the ratio between the sizes of M_1 and M_2, there is a logarithmic relation between k and I_{PTAT}. As k increases, its mismatch will cause lower variations in the PTAT current. However, increasing k also consumes more area. We can achieve higher k values by using different series and parallel topologies of unit transistors with $W/L = 1u/1u$ for M_1 and M_2 while maintaining the same area. Iterative simulations show the lowest spread of current was achieved with $k = 64$, using 8 series unit transistors ($W/L = 1u/1u$) for M_1 and 8 in parallel for M_2. Hence, we can calculate R_{PTAT} as follows:

$$I_R = \frac{U_T \ln(k)}{R_{\text{PTAT}}} = \frac{26\,\text{mV} \cdot \ln(64)}{R_{\text{PTAT}}} = 1\,\text{nA} \implies R_{\text{PTAT}} \approx 110\,\text{M}\Omega$$

This resistor is very large, hence it should be replaced by an area-efficient component such as a switched capacitor resistor, which will be explained later.

For the sizing of the PMOS mirror, minimizing mismatch is crucial. It is preferable to use large transistors in strong inversion, but with a 1 nA current, very long transistors are required. This will increase both the threshold voltage (V_{th}) and the required gate-source voltage (V_{GS}), limiting the available headroom. Therefore, the worst-case scenario should be considered, which occurs in SSSS-40, where V_{th} is at its maximum. Considering the DC operating points we have:

$$V_{dd} = V_R + Vdsat_{M2} + Vsg_{M4}$$

Keeping M2 in saturation leaves nearly 580 mV for Vsg, which translates to an IC factor equal to 0.24. The best matching is achieved with long transistors and a (W/L) of $0.25\,\mu\text{m}/10\,\mu\text{m}$ for the PMOS mirror. Due to the GF22nm technology limitation that restricts the area of the transistors to $2\,\mu\text{m}^2$ per finger, M3 and M4 are implemented with series composite of transistors.

Replacing Resistor with Switched Capacitor Resistor

The resistor required to generate a 1 nA current is 110 MΩ, which would cover a very large chip area. Therefore, it is necessary to replace this resistor with more area-efficient components, such as switched capacitor resistors. The resulting schematic is shown in Figure 3.3. The equivalent resistance of a switched capacitor is obtained as [26]:

$$R_{\text{sc}} = \frac{V_{\text{average}}}{I_{\text{average}}} \approx \frac{1}{Csf_{\text{clk}}} \tag{3.3}$$

Assuming a 32 kHz clock frequency available in the system from the crystal oscillator, the required capacitance is 280 fF.

In order to reduce the area, two branches with two-phase non-overlapping clocks are used. This setup effectively doubles the charge transfer and ripple voltage frequency which means that only half the capacitor value (140 fF) is needed to achieve the same equivalent resistance.

To dampen ripples, a large damping capacitor C_{d} is used in parallel with the switched capacitor. The ripple amplitude is roughly given by:

$$\text{Ripple Amplitude} = V_R * \frac{C_{\text{s}}}{C_{\text{d}}} \tag{3.4}$$

Since the effective capacitance C_{s} is halved due to the doubled clock frequency, the ripple amplitude is also reduced by half. Consequently, we only need half the damping capacitance C_{d} to achieve the same level of ripple reduction. Aiming for less than ±1% ripple in I_{out} nominal value (1nA) at TT27 results in a minimum C_{d} equal to 2.8 pF.

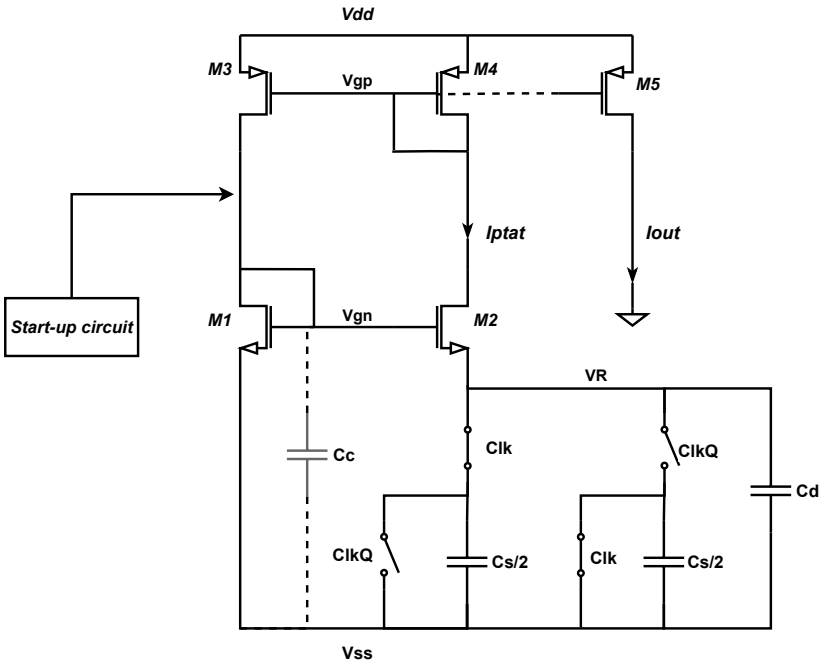


Figure 3.3: Schematic of PTAT current reference with switched capacitor resistor

The two-phase non-overlapping clocks required for the switches are generated using a cross-coupled RS flip-flop, as shown in Figure 3.4.

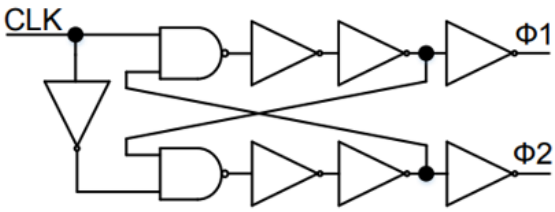


Figure 3.4: Schematic of the cross-coupled RS flip-flop

In the design of switched capacitor resistors, NMOS switches were selected due to the low input voltage ($V_R = 110$ mV). The (W/L) ratio of the NMOS transistors was chosen to be 0.5 μm/4 μm. This sizing ensures that the on-resistance (R_{on}) remains below 40 kΩ and the off-resistance (R_{off}) exceeds 80 GΩ across all process corners. This maintains a leakage current of less than 10 pA and limits the degradation in the PTAT current over the automotive temperature range.

Loop stability after adding the switched capacitor resistor

The loop containing M_1, M_2, M_3, and M_4 in Figure 3.2 forms a positive feedback loop. This loop remains stable as long as the loop gain is less than 1. The DC closed-loop gain is given by:

$$A_v = \frac{g_{m2}}{1 + g_{m2}R}\left(\frac{1}{g_{m4}} \parallel r_{o2}\right) g_{m3}\left(r_{o3} \parallel \frac{1}{g_{m1}}\right) \tag{3.5}$$

After adding the switched capacitor (C_s) and the damping capacitor (C_d), R in the equation above changes to $\frac{R}{1+RC_dS}$. This introduces a pole and a zero in the closed-loop frequency response. The pole is located at $\omega_{p1} = \frac{1+g_{m2}R}{RC_d}$, while the zero is at a lower frequency, located at $\omega_{z1} = \frac{1}{RC_d}$. The zero causes an increase in the loop gain and pushes it above 1, causing instability. In general, in this beta multiplier topology, the capacitor on the source of M_2 should be taken care of because it can lead to instability if it becomes too large [2].

To ensure stability, a capacitor (C_c) is added from V_{gn} to ground as shown in Figure 3.3. This changes r_{o3} in the Equation 3.5 to $r_{o3} \parallel \frac{1}{C_cs} \parallel \frac{1}{g_{m1}}$, adding another pole to counteract the zero introduced by C_d. The pole is located at $\frac{1+g_{m1}r_{o3}}{r_{o3}C_c}$. A C_c value of 2 pF is sufficient to ensure the loop gain remains below 0 dB and the circuit maintains stability across all process corners. .

3.2 Simulation results

3.2.1 DC Analysis

The output current over the temperature range is shown in Figure 3.5 for different process corners. The spread of the PTAT current over process corners is ±13%. This spread over process corners due to the switched capacitor resistor, PMOS mirror and NMOS transistors is about 10%, 3%, and 7% respectively.

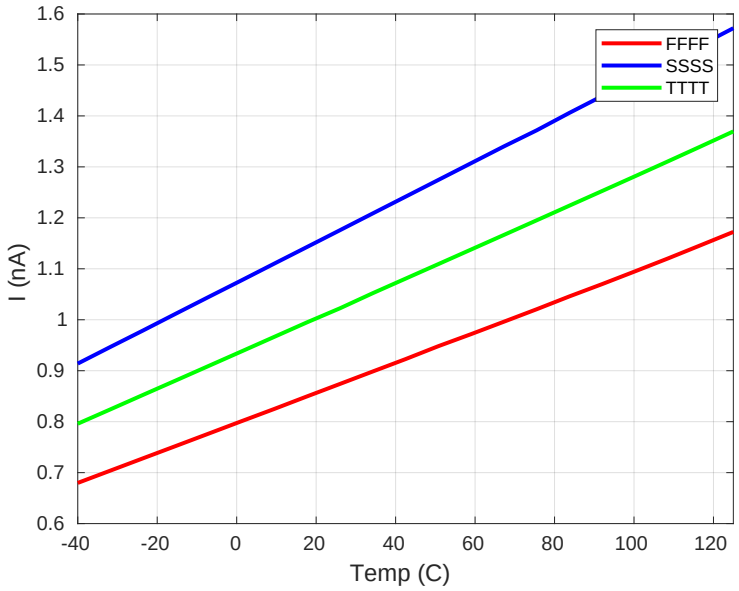


Figure 3.5: Output current over the temperature range for different process corners

The output current versus supply voltage at TTTT27 is shown in Figure 3.6. The line sensitivity is 2%/V calculated using the BOX method.

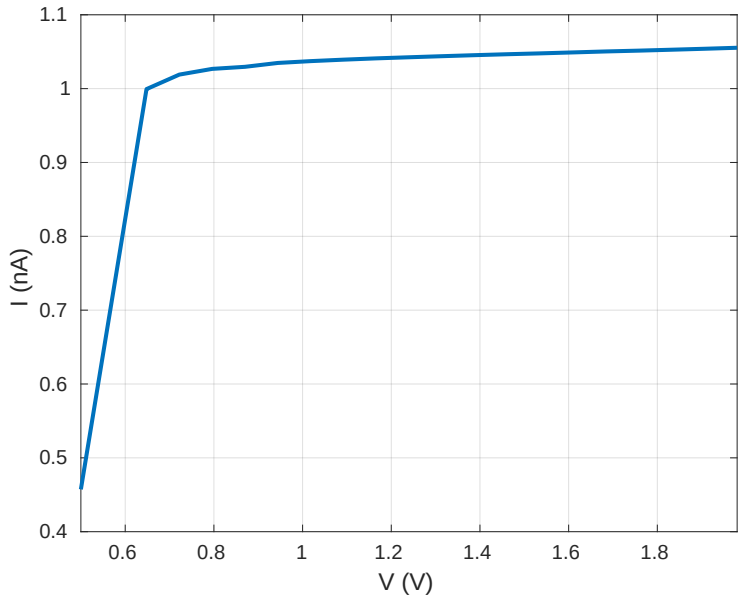


Figure 3.6: Average output current versus supply voltage

3.2.2 AC Analysis

The Power Supply Rejection Ratio (PSRR) is calculated as:

$$\mathrm{PSRR(dB)} = 20\log\left(\frac{\Delta I_{\mathrm{out}}}{\Delta V_{\mathrm{DD}}} \cdot \frac{V_{\mathrm{DD}}}{I_{\mathrm{out}}}\right)$$

The curves for different process corners are shown in Figure 3.7.

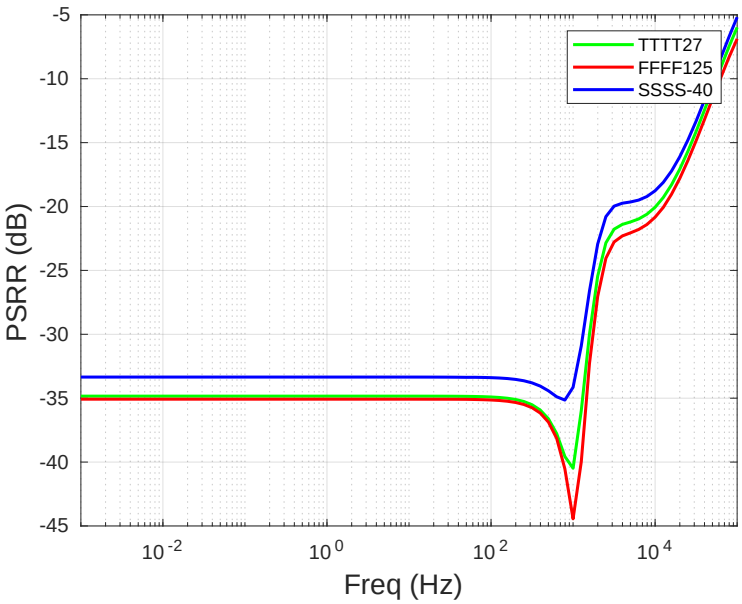


Figure 3.7: PSRR curves of output current

3.2.3 Stability Analysis

The frequency response of the PTAT closed loop is shown in Figure 3.8. The loop gain is always lower than 0 dB, ensuring that the positive feedback loop remains stable across process and temperature variations.

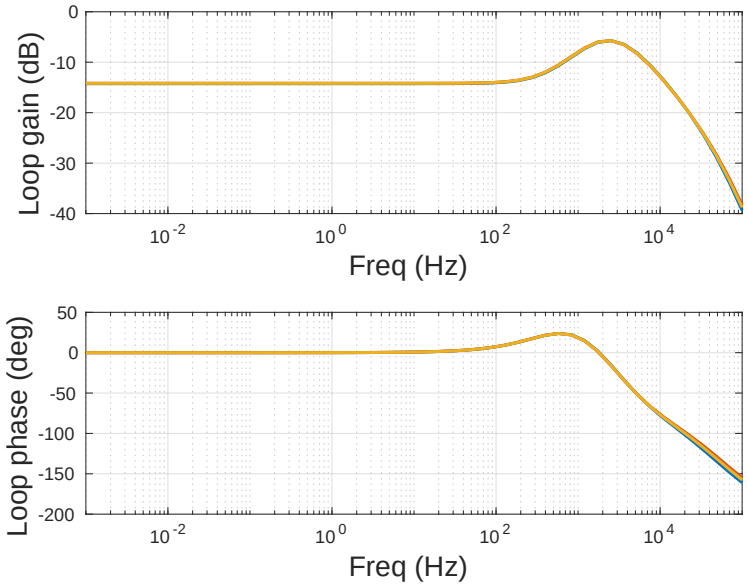


Figure 3.8: Frequency response of the PTAT closed loop.

3.2.4 Noise Analysis

The noise analysis of the PTAT output current is shown in Figure 3.10 for different process corners which are FFFF125, TTTT27, and SSSS-40. The maximum integrated output noise from 0.1 Hz to 100 kHz is simulated to be 7 pA_{RMS}. The main noise contributors to the integrated output noise in Figure 3.2 are shown in the pie chart of Figure 3.9.

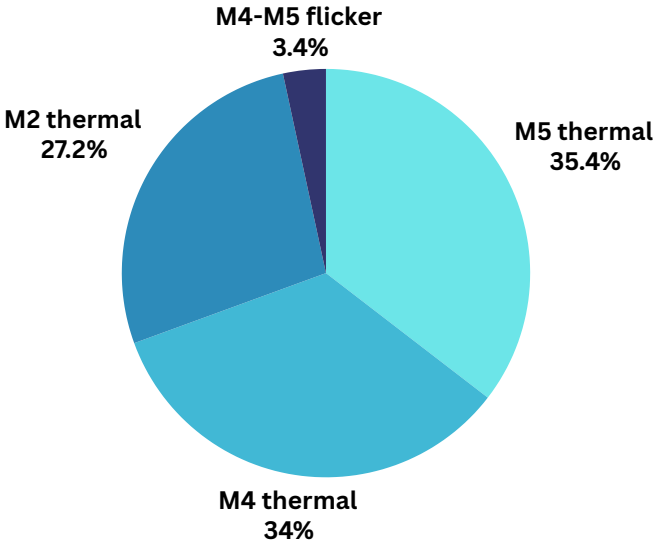


Figure 3.9: Pie chart of the main noise contributors to the PTAT current

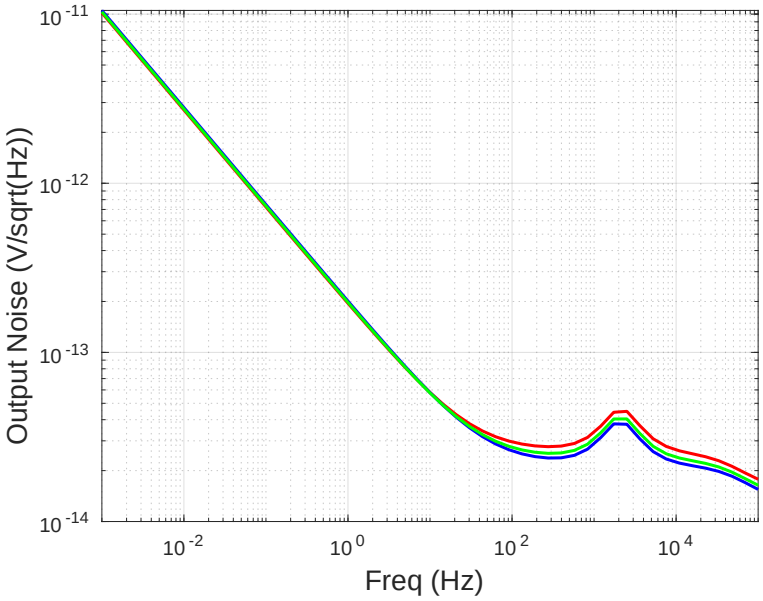


Figure 3.10: Output current noise density of PTAT reference

3.2.5 Monte Carlo Analysis

The Monte Carlo analysis was performed with 200 points at 27 degree, considering both mismatch and process variations. The histogram result for the average output current is shown in Figure 3.11. As illustrated, the σ/μ ratio is 5% for the output current. Output current spread over the temperature range is also shown in Figure 3.12. The spread (σ/μ) due to the PMOS mirror, switched capacitor resistor and NMOS transistors is about 3%, 4.3%, and 0.5% respectively.

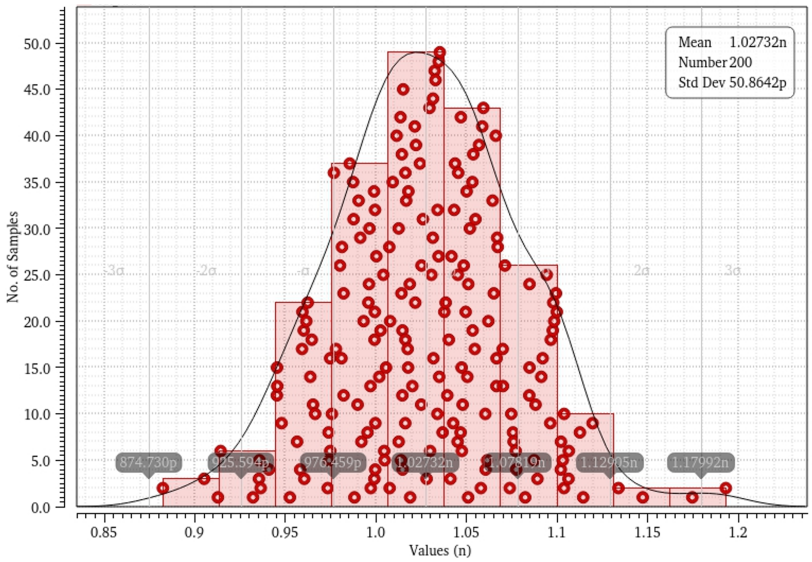


Figure 3.11: Histogram of the average output current

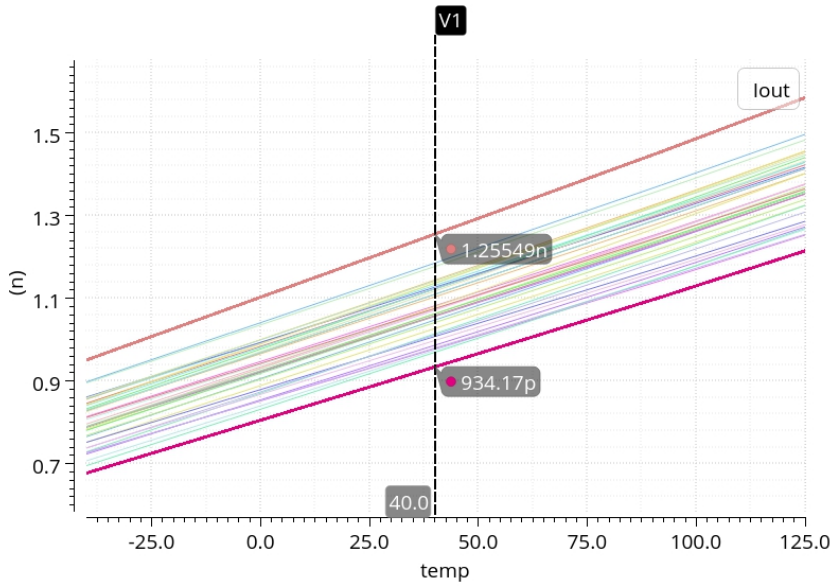


Figure 3.12: Monte Carlo simulation results for the output current over temperature

3.2.6 Table of results

The final results and specifications are summarized in Table 3.2. The design successfully meets all the specified requirements. The current consumption per branch has been reduced from 1 µA to 1 nA, achieving a substantial decrease in power consumption. This reduction was accomplished without compromising performance or increasing the overall area.

Table 3.2: PTAT Current Reference Summary of Performance Metrics

Metric	**Specification**	**Performance (TT27)**
Vdd	0.8 V	0.8 V
Power consumption	< 4 nW	3.4 nW
Area	$< 1200\ \mu m^2$	$1040\ \mu m^2$
Integrated noise [1m-100k]	-	$7\ pA_{rms}$
DC PSRR	-	-35 dB
Spread of output current (3σ/mean)	-	15%

4 Bandgap (BG) reference

4.1 Design methodology

A critical part of most analog circuits is the bandgap (BG) reference. This generates a stable reference voltage that should be independent of temperature variations and process spread. In our design, the primary objective is to achieve ultra-low power consumption while operating at a low supply voltage. Specifically, the target is to limit the current consumption to 1 nA per branch from an 900 mV supply. The design should also operate over the automotive temperature range.

After evaluating various topologies, we selected the Banba et al topology [3]. This is well suited for use in low-supply voltage applications and for generating reference voltages lower than the silicon bandgap voltage ($\sim 1.2V$). It works by combining proportional-to-absolute-temperature (PTAT) and complementary-to-absolute-temperature (CTAT) currents to realize a constant with temperature (CWT) current. This current is then converted into a reference voltage by a load resistor.

In the following sections, we will detail the design specifications and circuit-level design choices.

4.1.1 Specifications

The design specifications of the proposed BGR are summarized in the table below:

Parameter	**Specification**
Supply Voltage	900 mV
Reference Voltage	800 mV
Current per Branch	1 nA
Reference current	1 nA
Core Power Consumption	$<$ 6 nW
Temperature Range	-40°C to 125°C

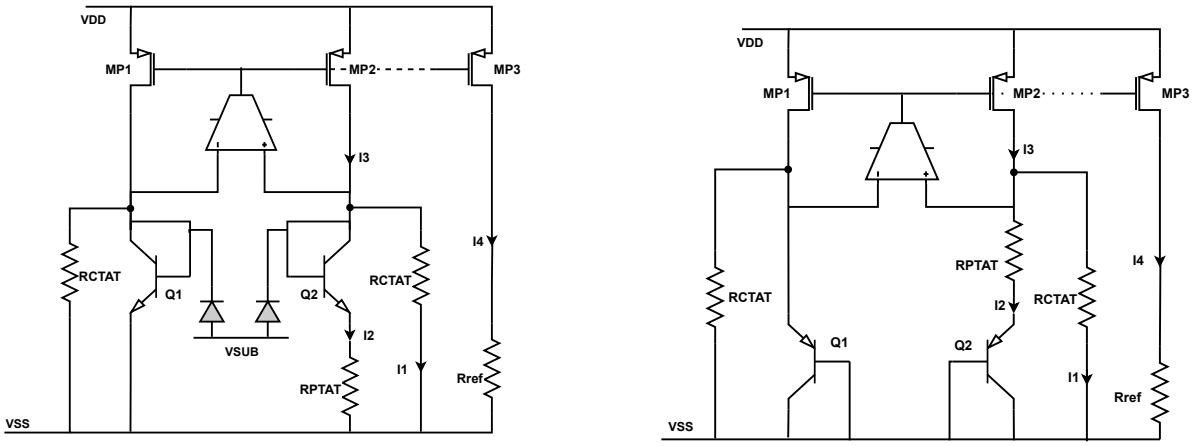
Table 4.1: Bandgap Current Reference Design Specifications

The supply voltage of 900mV is chosen based on the end-of-life voltage of alkaline, silver oxide, and zinc-air batteries [14]. The reference voltage is set to 800mV to match the nominal supply voltage required by the RF FETs in GF22nm technology and the core voltage of the system utilizing the bandgap reference. These specifications pose several challenges. Firstly, maintaining performance at a low supply voltage is difficult, particularly at low temperatures when V_{BE} increases to about 700 mV. Secondly, the wide automotive temperature range introduces considerable leakage currents at high temperatures, leading to errors in the nanoampere

(nA) range. These can compromise the accuracy and stability of the reference voltage. Lastly, generating nA currents requires the use of very large resistors. To avoid consuming significant silicon area, area-efficient alternatives such as duty-cycled resistors or switched capacitors will be used.

4.1.2 Design choices

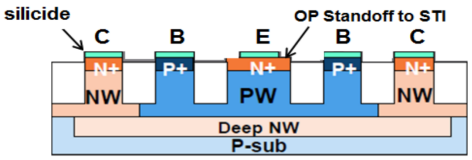
The first consideration is the choice of the BJTs. As shown in Figure 4.1, topologies based on either PNPs or NPNs can be used. NPNs have lower base-emitter voltage than PNPs with the same emitter area as shown in Figure 4.2d, which is important in low-voltage design. Substrate noise is a big concern in mixed-signal and RF integrated circuits, where digital switching and high-frequency signals can induce noise in the substrate that affects the performance of sensitive analog and RF components. In our design, the substrate noise rejection is very important as we are aiming for nA-level currents. The noise rejection from the substrate to the collector for both BJT types is illustrated in Figure 4.2c. In this testbench, a 1V ac source of noise is connected to the substrate, and collector voltage ac variations are plotted. NPNs have high substrate noise rejection. This is because their deep n-well acts as a shield, blocking noise from propagating through the substrate to the sensitive collector region. PNPs have very low noise rejection since their collectors are part of the substrate. However, they are less sensitive to packaging stress than NPNs. This structural difference is highlighted in the layouts shown in Figure 4.2.



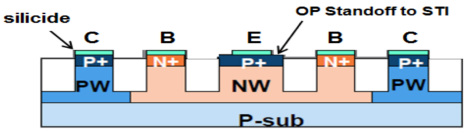
(a) BG schematic with NPNs

(b) BG schematic with PNPs

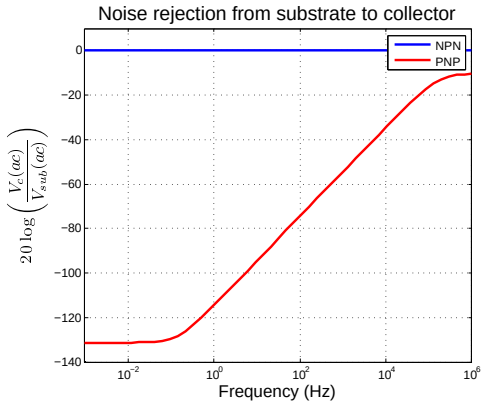
Figure 4.1: Comparison of BG schematic with NPNs (a) and PNPs (b)



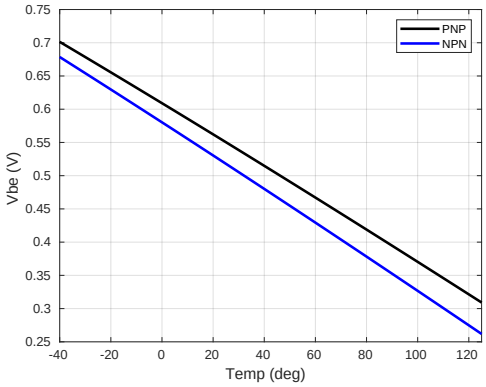
(a) Cross section of NPN



(b) Cross section of PNP



(c) Noise Rejection from Substrate to Collector



(d) Comparison between VBE of NPN and PNP with $3.2\mu m * 3.2\mu m$ emitter area

Figure 4.2: Comparison of NPN and PNP cross sections, noise rejection characteristics and base-emitter voltages

Additionally, the use of NPNs means that all the resistors can be connected to the ground, which should help enhance the linearity of any switches without the need to use transmission gates with bulk switching or clock bootstrapping which consumes more power.

However, as shown in Figure 4.2a the downside of NPNs is the existence of a parasitic diode from their collector to the substrate, which can leak more than 150pA of current, especially at temperatures above 85°C (4.1.7). Also, while NPNs have higher current gain (β), they also have greater variability in β across process corners. This variability contributes to a slightly higher process-induced spread in the final bandgap voltage, amounting to less than 1%. However, as will be demonstrated, the dominant source of spread in the designed bandgap voltage is the matching of the PMOS current mirrors, rather than the variation in β.

In this work, NPNs were chosen for the bandgap topology due to their lower base-emitter voltage and high substrate noise rejection. The contribution of β spread to overall error is minimal when compared to other error sources. To mitigate the effects of parasitic diode leakage, leakage compensation techniques will be employed.

The next decision involves selecting the appropriate MOSFETS used in the error amplifier and the BG core current mirrors. Based on the discussion in Section 3.1.2, thick oxide low Vt (EGLVT) transistors are selected to limit gate and channel leakage currents.

The last design choice before going into the detailed design steps is the selection of the error amplifier topology, which is essential for ensuring the stability and accuracy of the bandgap reference. Given the low supply voltage and wide input range, the two-stage Miller compensated and folded-cascode operational transconductance amplifiers (OTAs) were considered. Each of these topologies presents unique advantages and trade-offs in terms of performance, complexity, and power consumption. The final comparison table for the OTAs will be discussed later.

4.1.3 Design steps

First, we need to calculate the resistor values in the schematic of Figure 4.1a. The voltage across RPTAT is a proportional-to-absolute-temperature (PTAT) voltage which is given by Equation 4.1, as long as the collector currents of the BJTs are equal. Another approach is to scale Id1 up with respect to Id2, which is not desirable as it would also increase Vbe1 and result in lower available voltage headroom, as well as increasing power consumption.

$$V_{BE1} - V_{BE2} = U_T \ln\left(\frac{I_{C1}/I_{S1}}{I_{C2}/I_{S2}}\right) \approx U_T \ln\left(\frac{I_{S2}}{I_{S1}}\right) = U_T \ln(p) = \frac{kT}{q}\ln(p) \tag{4.1}$$

To define the value of p, we must consider mismatch. Although bipolar transistors exhibit excellent matching, the best results are achieved with a common centroid layout. For this configuration, p is defined as $p = n^2 - 1$. The minimum value of n for a common centroid layout is 3, which sets $p = 8$, resulting in 8 unit transistors for Q1. Increasing p further would only increase the area and parasitic leakage currents without significantly improving matching.

To define the emitter area for the BJT unit transistors, a testbench was set up with a 1nA ideal current injected into Q1 and Q2, with the defined p ratio equal to 8. The reference voltage, $V_{\text{ref}} = V_{\text{BE1}} + k\Delta V_{\text{BE}}$, was plotted for different emitter areas. The parameter k was tuned to achieve the lowest temperature coefficient for each area. Figures 4.3 and 4.4 show the plots of V_{ref} versus temperature for different corners and V_{BE} versus temperature, respectively.

Given that our output reference voltage is 800mV and the supply voltage is 900mV, we need to maintain at least a 100mV headroom for transistors M_{P1} and M_{P2}. Therefore, it is desirable to increase the emitter area. However, as shown in Figure 4.3, increasing the emitter area also increases the spread of V_{ref} over different corners. Additionally, since the saturation current I_{s} is proportional to the emitter area [18], it becomes comparable to the 1nA bias current, leading to nonlinearity in V_{BE}, especially at high temperatures.

Balancing the trade-off between headroom and spread over corners, the emitter area of the unit transistor was chosen to be $3.2\mu m \times 3.2\mu m$. This configuration provides a worst-case headroom of 120mV for M_{P1} and M_{P2}, while maintaining a reasonable spread of V_{ref} equal to 20 mV.

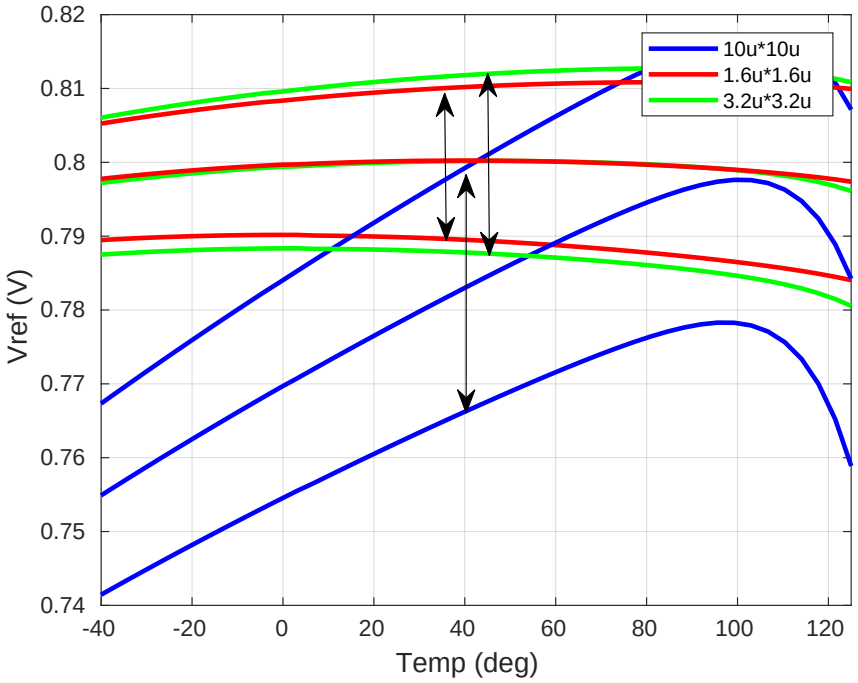


Figure 4.3: V_{ref} versus temperature for different emitter areas across FFFF, TTTT and SSSS corners

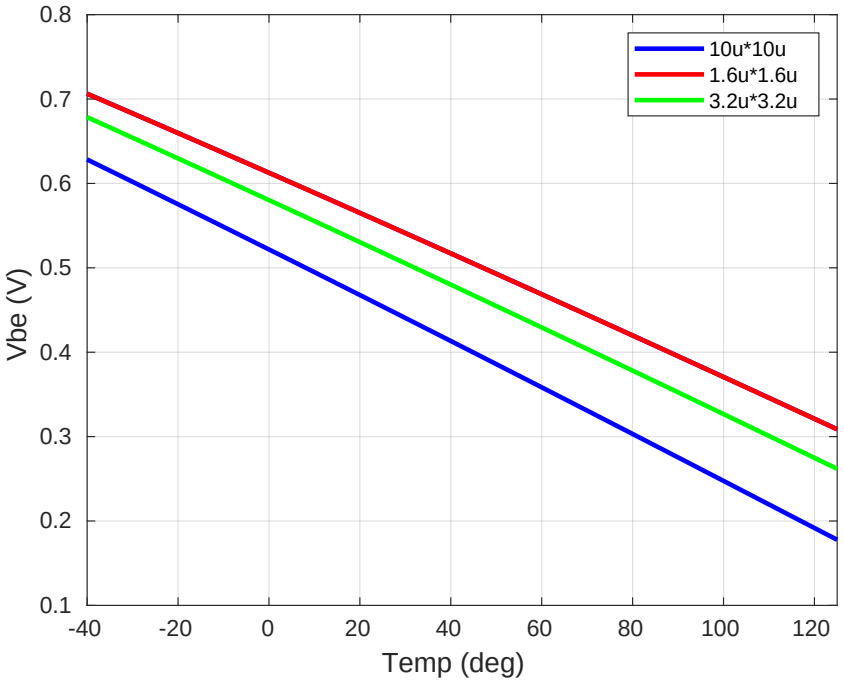


Figure 4.4: V_{BE} versus temperature for different emitter areas

The base-emitter voltage (V_{BE}) itself is a complementary-to-absolute-temperature (CTAT) and generates a CTAT current through R_{CTAT} (I_1) . By adding I_1 and I_2 and tuning their slopes such

that their temperature coefficients cancel out, we can generate a temperature-independent current, I_3. Therefore, the derivative of the following equation with respect to temperature should be zero:

$$\frac{di_{\mathrm{BG}}}{dT} = \frac{1}{R_{\mathrm{PTAT}}}\frac{d(\Delta V_{BE})}{dT} + \frac{1}{R_{\mathrm{CTAT}}}\frac{d(V_{BE})}{dT} = 0 \tag{4.2}$$

Given an emitter current of 1 nA, simulations show that $\frac{d(\Delta V_{BE})}{dT} \approx -2.5\,\mathrm{mV}/^{\circ}\mathrm{C}$. Solving for the ratio of the resistors, we find:

$$\frac{R_{\mathrm{CTAT}}}{R_{\mathrm{PTAT}}} = \frac{30}{\ln(p)} \tag{4.3}$$

Given $p = 8$, we can calculate the resistor values accordingly:

$$\Delta V_{BE} = \frac{kT}{q}\ln(8) = 54\,\mathrm{mV}$$

$$R_{\mathrm{CTAT}} = 1.3\,\mathrm{G\Omega}, \quad R_{\mathrm{PTAT}} = 90\,\mathrm{M\Omega}$$

With the resistor values determined, the next step is to size the PMOS mirror transistors before moving to the OTA design. To minimize mismatch, larger transistors are preferable. Additionally, a higher inversion coefficient, $I_C > 10$ (Equation 4.4) is desirable as it minimizes the effect of Vth spread. However, in this work, the target current is 1 nA, which inherently places the transistors in the weak inversion region unless very long channel transistors are used. On one hand, nearly 100mV headroom is available for the PMOS mirror. On the other hand, the gate voltage (V_{gp}) cannot drop below the output dynamic range of the OTA. These limitations restrict the gate-source voltage (V_{gs}) and the inversion coefficient. In the GF22nm technology node, there is an additional constraint where the gate area per finger should not exceed 2 $\mu\mathrm{m}^2$. In Figure 4.5, the drain current (I_D) versus the drain-source voltage (V_{ds}) for various long-channel transistors is shown, all having a (V_{dsat}) of 100 mV. Considering all the constraints, we selected the longest channel length of 8 µm to minimize the effects of channel length modulation. The width was chosen to be 0.25 µm, providing an inversion coefficient (I_C) of 0.2, as indicated by the green curve.

$$IC = \frac{I_D * L}{I_S * W} \tag{4.4}$$

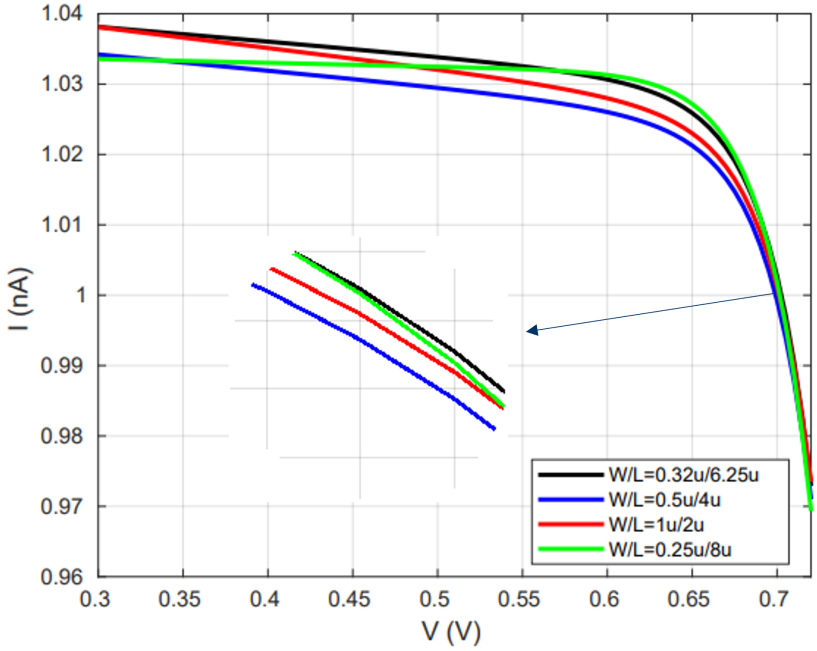


Figure 4.5: I_d of PMOS mirror vs V_S with different transistor sizings

So far, we have determined the values of the resistors, the ratio between the area of the BJTs, and the optimized I_C factor and sizing of the PMOS mirror. Before moving to the discussion about the design and considerations of the operational transconductance amplifier (OTA), we should specify error sources and error budgets for the main error sources.

4.1.4 Error sources

The first error type is the curvature of VBE. The discussion about compensating the TC of VBE with a PTAT voltage assumes VBE has only a first-order temperature coefficient. However, VBE is expressed as [11]:

$$V_{BE} = V_{g0} - (V_{g0} - V_{BE,T_r})\frac{T}{T_r} - (\eta - 1)V_T \ln \frac{T}{T_r} \tag{4.5}$$

The bandgap curvature or the temperature coefficient of the bandgap is mainly caused by VBE nonlinearity. With our chosen BJT, this error has a nominal value of $\pm 10mV$, translating into $\pm 1.25\%$ error contribution.

The current gain (β) of a BJT is limited and its spread can affect the slope of V_{BE}, which is crucial for the accuracy of our bandgap current reference. Although V_{BE} is primarily determined by the collector current (I_C), the PTAT current flows through the resistor from the emitter. The current gain (β) has a nominal value of 4 and varies by 50% due to process variations. The variations in V_{BE} with respect to the variations in β can be expressed as [11]:

$$V_{BE} = U_T \ln \left(\frac{I_E}{I_S} \cdot \frac{\beta + \Delta\beta}{1 + \beta + \Delta\beta} \right) \tag{4.6}$$

$$V_{\mathrm{ER}} = U_T \cdot \frac{\Delta\beta}{\beta} \cdot \left[\left(\frac{1}{1+\beta} \right) \Big|_{t=125^\circ\mathrm{C}} - \left(\frac{1}{1+\beta} \right) \Big|_{t=-40^\circ\mathrm{C}} \right] \tag{4.7}$$

The error in the temperature coefficient of V_{BE} can be calculated to be approximately $\pm 0.7mV$. This error is divided by R_{CTAT}, resulting in a final contribution to the reference current of about 0.5 pA and 0.4 mV to the reference voltage. This translates to $\pm 0.05\%$ error contribution to the TC. On the other hand, the error caused in the output voltage nominal value due to the spread of β over process corners is related to $\Delta\beta/\beta$ and the nominal value of the current gain and is equal to $\pm 1.25\%$.

Looking at Equation 4.8, it seems like all resistors appear in ratio form and the resistor spread will be canceled out. However, due to limited current gain, substituting $I_E = \frac{\Delta V_{BE}}{R_{PTAT}}$ in Equation 4.6 result in Equation 4.9 which is influenced by R_{PTAT} spread.

$$V_{ref} = \frac{R_{ref}}{R_{\mathrm{PTAT}}} \Delta V_{BE} + \frac{R_{ref}}{R_{\mathrm{CTAT}}} (V_{BE}) \tag{4.8}$$

$$V_{BE} = V_t \ln \left(\frac{I_E}{I_S} \cdot \frac{\beta}{\beta+1} \right) = V_t \ln \left(\frac{\Delta V_{BE}}{R_{PTAT}} \cdot \frac{1}{I_S} \cdot \frac{\beta}{\beta+1} \right) \tag{4.9}$$

The spread of narrow high r poly resistors is $\pm 30\%$ over process corners. These resistors were chosen due to their high sheet resistance and area efficiency. A testbench was created using the bandgap topology with ideal current mirrors and an ideal operational amplifier. When the corner spread is applied only to R_{PTAT}, it directly affects the reference current, both the temperature coefficient and the nominal value, as illustrated in Figure 4.6, and consequently impacts the reference voltage. However, when the spread is applied to R_{PTAT}, R_{CTAT}, and R_{ref}, the final error contribution to V_{ref} is reduced to $\pm 0.5\%$, as shown in Figure 4.7.

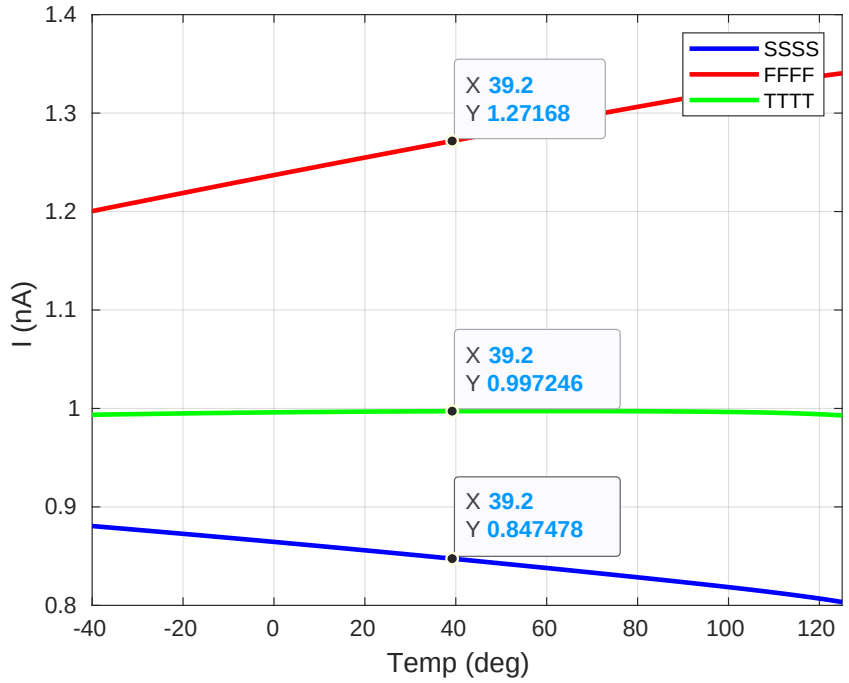


Figure 4.6: Effect of R_{PTAT} corner spread on reference current

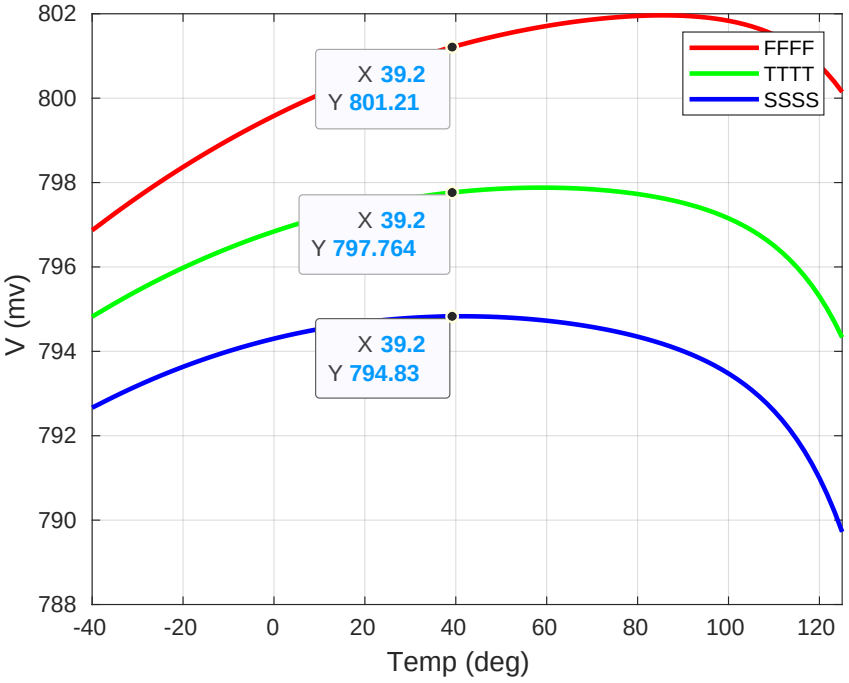


Figure 4.7: Error contribution on V_{ref} with corner spreads in R_{PTAT}, R_{CTAT}, and R_{ref}.

Error caused by opamp offset is expressed as Equation 4.10 and its contribution depends on the input-referred offset voltage of the OTA. Annotating 2 mV to the OTA input-referred offset results in about 15mV (2%) error contribution to V_{ref}:

$$V_{ref} = \frac{R_{ref}}{R_{\text{CTAT}}}(V_{BE} + \frac{R_{CTAT}}{R_{\text{PTAT}}}\Delta V_{BE} - V_{OS}(\frac{R_{PTAT} + R_{CTAT}}{R_{PTAT}})) \tag{4.10}$$

The error caused by opamp finite open loop DC gain translates into an input referred error voltage and error in drain currents. Annotating 1% error contribution to this error and not considering the spread of current caused by current mirrors result in having 40 dB power supply rejection [19], and the target open loop DC gain should be higher than 58 dB.

$$\Delta V_{ref} = R_{ref}\Delta I_d = R_{ref}\frac{\Delta V_{out}}{A * R_{PTAT}} \tag{4.11}$$

Where ΔV_{out} is the voltage error in the output of the amplifier.

Finally, as previously discussed, the PMOS current mirrors are the dominant contributors to the output voltage spread. The low headroom available in the design necessitated setting the inversion coefficient (IC) of the PMOS mirror to 0.2. This low IC value leads to significant mismatch between the MOSFETs, resulting in a substantial spread in the output current. Simulations indicate that the PMOS mirror contributes a one-sigma spread of 2.3% to the overall output voltage spread.

4.1.5 Operational transconductance amplifier (OTA)

First, the requirements of the OTA should be defined based on our core design. The most important requirements that can be crucial for choosing between topologies are the open-loop gain and input and output dynamic ranges.

Regarding the input dynamic range, V_{BE} of the BJTs varies from approximately 300 mV to 700 mV over the temperature range from -40°C to 125°C. This wide range necessitates an OTA with sufficient input dynamic range to accommodate these variations. Similarly, the required output dynamic range is from 200 mV to 500 mV, which corresponds to $V_{DD} - |V_{GSP}|$ over process and temperature variations. Ensuring that the OTA can handle these ranges is critical for maintaining the performance and stability of the overall design.

The two commonly used OTAs mentioned in Section 4.1.2 are shown in Figure 4.8. We can compare them by using the three main criteria established earlier: input dynamic range, output dynamic range, and DC gain.

The folded-cascode topology (pseudo-folded-cascode in Figure 4.8a) has the highest input and output dynamic ranges. The large input dynamic range is particularly critical in our design due to the significant variation of V_{BE} over temperature. It also provides high DC gain.

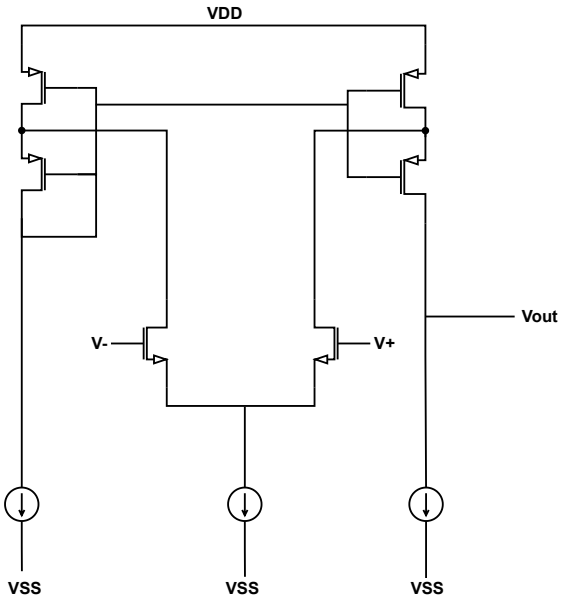
The two-stage Miller compensated topology (Figure 4.8b) provides a high output dynamic range but a low input range. It has excellent DC gain, making it suitable for applications requiring high gain and stability.

The table below provides a comparison of the two different OTAs:

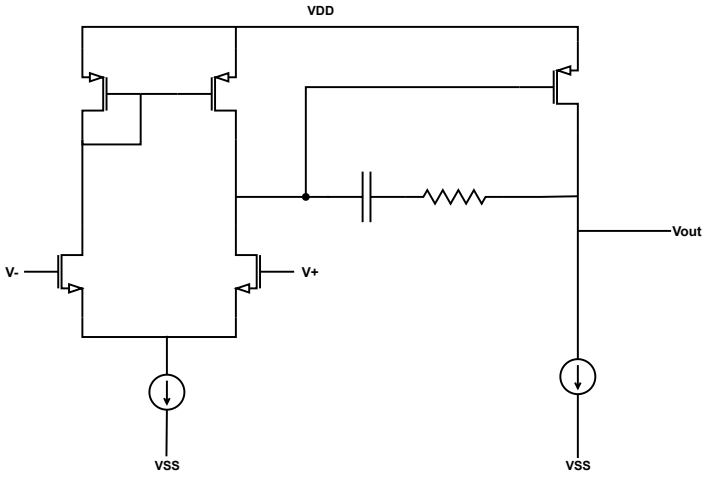
OTA Topology	**DC Gain**	**Vin Dynamic Range**	**Vout Dynamic Range**
Two-Stage Miller Compensated	Very High	Min: $V_{gsn} + V_{dsatN}$ Max: $V_{DD} - V_{sgP} + V_{gsN} - V_{dsatN}$	Min: V_{dsatN} Max: $V_{DD} - V_{dsatP}$
Pseudo-folded Cascode	High	Min: $V_{gsn} + V_{dsatN}$ Max: $V_{DD} - V_{dsatP} + V_{gsN} - V_{dsatN}$	Min: V_{dsatN} Max: $V_{DD} - V_{dsatP}$

Table 4.2: Comparison of OTA topologies based on DC gain, input dynamic range, and output dynamic range.

Based on the table above, the pseudo-folded-cascode OTA has the best overall performance, particularly in terms of input and output dynamic ranges. The conclusions drawn from these comparisons will be further elaborated upon and validated with simulations.



(a) Pseudo-folded-cascode OTA



(b) Two-Stage Miller Compensated OTA

Figure 4.8: Schematics of the two commonly used OTAs in bandgap reference design: (a) Pseudo-Folded-Cascode OTA, (b) Two-Stage Miller Compensated OTA

In most OTA designs, the main goal is to maximize the transconductance (g_m) of the input pair to achieve higher gain. Therefore, we aim for a low inversion coefficient (IC) with minimum length and maximum width for the input transistors. By using a large input pair, we can significantly reduce the input voltage offset. Additionally, for high gain, the output conductance (g_{ds}) of the current mirrors should be maximized. This is achieved by choosing the maximum length and minimum width for the transistors, leading to better matching and performance of the current mirrors.

To ensure stable performance, we employ different compensation techniques proper for each OTA topology while also optimizing for area efficiency.

Pseudo folded-cascode OTA

The pseudo-folded cascode schematic is shown in Figure 4.9. This topology has the advantage of a high input dynamic range since the voltage headroom of the input pair is limited by the V_{ds} rather than the V_{th} of the PMOS mirror. The design also employs self-cascoded transistors. A notable benefit of using PMOS self-cascoded composites is the ability to achieve effective compensation without the need for an additional series resistor with the compensation capacitor (to shift the right half plane (RHP) zero). Instead, we utilize the middle low-impedance node of the self-cascoded composites and size it appropriately to obtain the desired impedance. This approach, known as indirect compensation, is highly area-efficient and was first introduced by Vishal et al [21]

Regarding the PMOS self-cascoded transistors illustrated in Figure 4.9 (Mp1, Mp2 and Mp3, Mp4), a small signal analysis shows that their output impedance is given by:

$$R_{\text{out}} = g_{m2} r_{o2} r_{o1} + r_{o1} + r_{o2} \tag{4.12}$$

To increase the output impedance, the ratio of $(W/L)_4$ to $(W/L)_3$ should be increased. In our design, this ratio is set to 32, as depicted in the schematic, to achieve high output impedance and high gain.

For the NMOS mirror, we also used self-cascoded composites and tuned the sizes to exploit the low impedance node between M_{n1} and M_{n2} for compensation. This configuration enhances the overall performance of the OTA by providing an optimal balance between gain and stability.

For stability considerations, if we treat the OTA and bandgap core PMOS transistors as a two-stage amplifier (Figure 4.9, we would typically place a Miller capacitor between the input and output of the second stage (C1). However, as the gain of the second stage is not sufficiently high to benefit significantly from Miller compensation, it would be better to place this capacitor from the gate of the core PMOS to V_{DD} (C3). This will also improve PSRR by coupling high-frequency variations of V_{DD} to V_{GP}, keeping V_{GS} of the PMOS mirror constant. To optimize the area, we split this capacitor into two: one connected from the gate of the core PMOS to V_{DD} (C3) and the other used for indirect compensation (C2). The compensation capacitors are illustrated in Figure 4.9. The indirect compensation capacitor is connected from the output of the second stage (PMOS core), which is the input of the OTA, to the low-impedance node. This capacitor splitting approach reduced the passive area by a factor of four. We used two 0.5 pF capacitors to achieve a phase margin of 60 degrees, as shown in Figure 4.10. Moreover, The DC voltage gain achieved is 75dB.

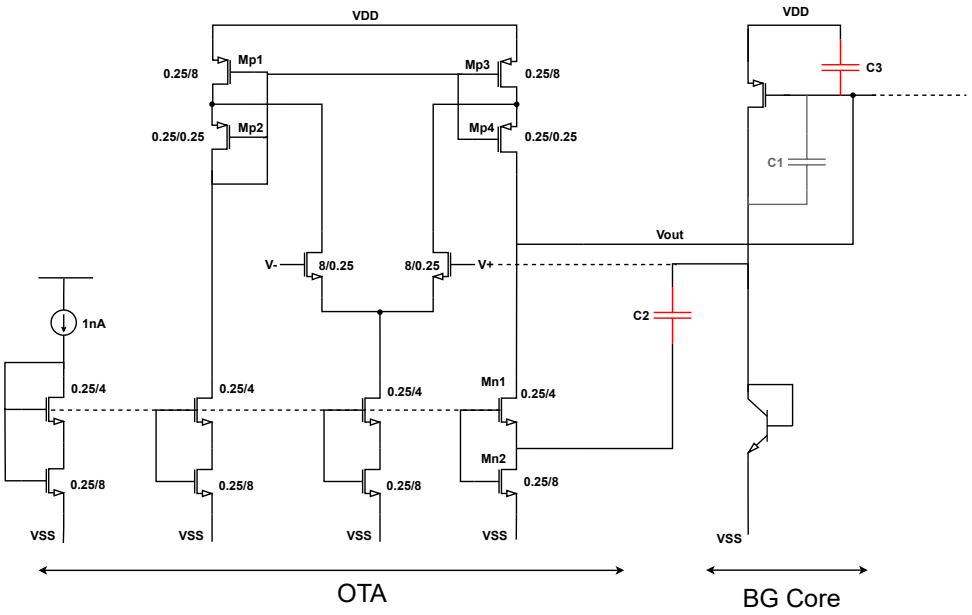


Figure 4.9: Schematic of the pseudo-folded cascode OTA with transistor sizes and compensation capacitors.

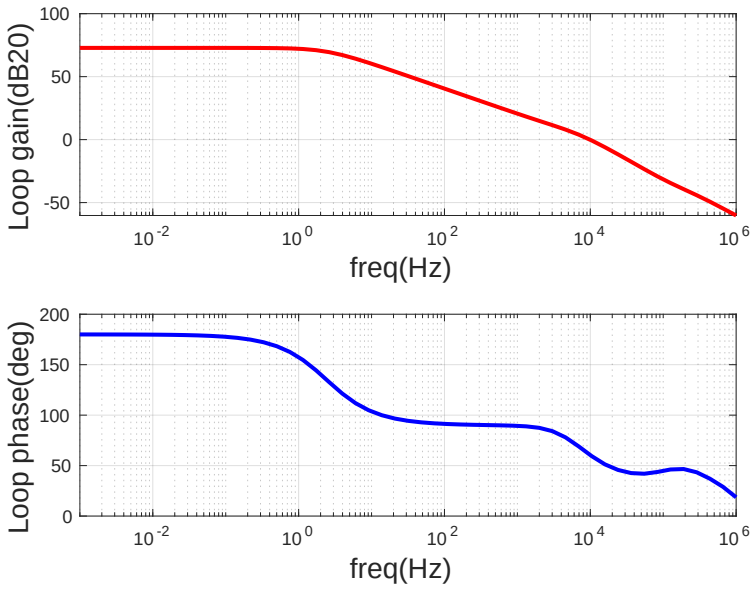


Figure 4.10: Stability analysis of pseudo folded cascode OTA in closed loop bandgap core

In summary, the pseudo-folded-cascode OTA design uses self-cascoded composites and indirect compensation to achieve high input dynamic range and stable performance. The capacitor splitting approach not only enhances stability but also optimizes area efficiency. The design meets the requirements for DC gain and phase margin and most importantly, input and output dynamic ranges. These performance metrics are the main reasons why we have chosen this topology for our bandgap design.

4.1.6 Bandgap startup

Regarding the bandgap startup, the topology of Figure 4.11 with Rs and RCTAT shunting the BJTs results in three steady states, two of which are undesirable: the zero state and the state where the current from the PMOS mirror only flows through Rs and RCTAT, with no current flowing to the BJTs and RPTAT. In the latter scenario, we would have a base-emitter voltage lower than expected. By removing Rs, we force all the current of its branch to flow through Q_1, hence eliminating the second unwanted state. This modification reduces the complexity of the startup. Since Ic1 is higher than Ic2, the PTAT voltage increases, increasing the bandgap current. Consequently, we must increase the resistor values to maintain a 1 nA current and the resistor ratios should be fine-tuned to get bandgap curvature. The new resistor values are 1.9 GΩ and 80 MΩ for R_{CTAT} and R_{PTAT}, respectively.

$$V_{BE1} - V_{BE2} = U_T \ln(k \cdot \frac{I_{C1}}{I_{C2}}) \tag{4.13}$$

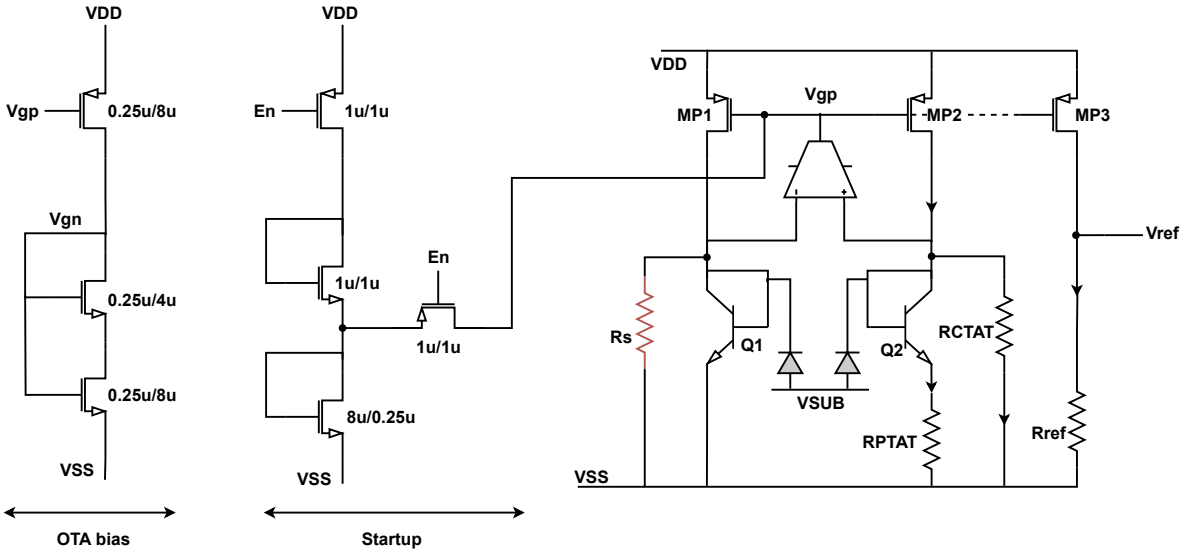


Figure 4.11: Schematic of the bandgap reference circuit with side resistor (Rs) and the proposed startup circuit

The bandgap voltage after removal of Rs and fine-tuning R_{CTAT} and R_{PTAT} is illustrated in Figure 4.12 with ideal resistors. The temperature coefficient (TC) of the bandgap voltage can be calculated using the box method as follows:

$$\mathrm{TC} = \frac{V_{\mathrm{REF,max}} - V_{\mathrm{REF,min}}}{V_{\mathrm{REF,avg}} \cdot (T_{\mathrm{max}} - T_{\mathrm{min}})} \times 10^6 \,\mathrm{ppm/°C} \tag{4.14}$$

The temperature coefficient of the bandgap voltage, simulated with ideal resistors, is 77 ppm/°C. Using real resistors, this TC will be lowered due to partial cancellation of the TC of the reference current and the resistor.

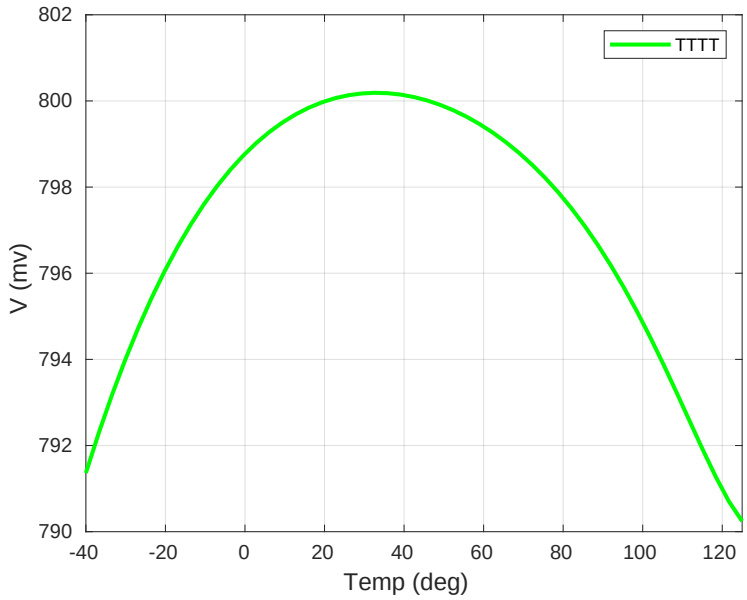


Figure 4.12: Bandgap reference voltage after removing Rs

Several methods can be employed for the design of bandgap reference startup circuitry. All these methods rely on pulling down the gate of the PMOS core to prevent it from staying equal to V_{DD} after V_{DD} ramps up from the ground. Here, the startup circuit shown in Figure 4.11 is used. When the enable signal is on, we connect an initial voltage lower than $V_{DD}/2$ to V_{gp}. This initial voltage generates a current into the bandgap branches, which is then mirrored to be used as the current source for the OTA. After this, we turn the enable signal off, disconnecting this voltage from V_{DD}. The generation of this enable signal will be explained in the system-level startup in Section 5.2.6. This operation allows the OTA to operate in the negative feedback loop with the initial bias we provided and reach the steady state voltages.

4.1.7 Parasitic diode leakage

As mentioned earlier, the NPNs have a parasitic diode from the collector to the substrate, which leaks at high temperatures. This can significantly affect the performance and accuracy of the bandgap reference circuit. To compensate for this, one effective approach is to mirror the leakage current and feed it back to the collector. As illustrated in Figure 4.13, a dummy diode (DNW layer on top of the PSUB in the layout) with the same area as the core NPN unit transistor generates the leakage current. This current is then mirrored and injected back into the core branch, effectively canceling out the leakage. Since Q2 consists of 8 parallel BJTs with 8 times of Q1 parasitic leakage current, the compensation current injected back to its branch should be 8 times, resulting in a mirror ratio of 8. However, due to the mismatch and g_{ds} effect of the current mirror and mismatch of the parasitic diode and the dummy diode, over- or under- compensation will appear as shown in the Monte Carlo result of Figure 4.14. Series transistors with $\frac{W}{L} = \frac{0.25\mu}{8\mu}$ were used for the current mirrors to improve their matching to less than 5% (σ/μ). Moreover, the current mirror ratio is trimmed at 125 degree and has ratios ranging from 7.75 to 8.5 with 2-bit trimming as shown in the mirror ratios of Figure 4.13.

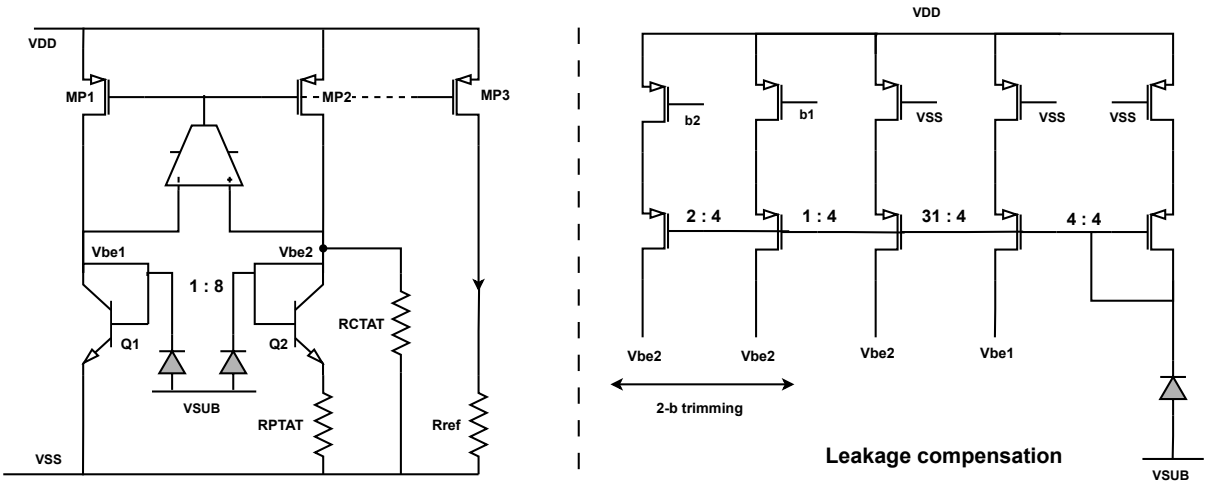


Figure 4.13: Parasitic diode leakage compensation circuit, dummy diode is generated in the layout with a DNW layer on top of the P-substrate

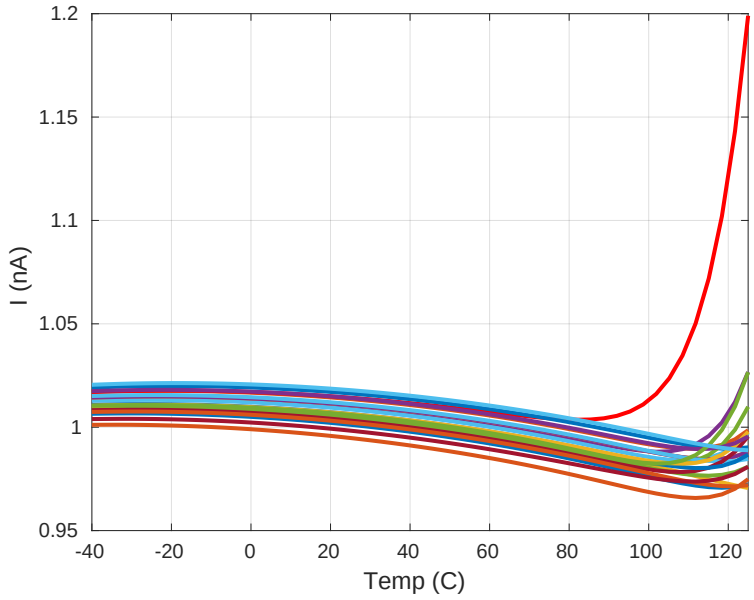


Figure 4.14: Bandgap current with and without (red) compensation.

In conclusion, compensating for parasitic diode leakage is necessary to keep the performance of the bandgap reference circuit with NPNs. By carefully designing the current mirror and trimming the multiplication factor, we can effectively mitigate the effects of leakage currents, especially at high temperatures.

4.1.8 Replacing resistors with area efficient components

As discussed earlier, the resistors required to generate a 1 nA bandgap current are $R_{\mathrm{CTAT}} = 1.9\ \mathrm{G\Omega}$ and $R_{\mathrm{PTAT}} = 80\ \mathrm{M\Omega}$. The area consumed by these resistors is significant and much

larger than the active area of the circuit. Therefore, we investigated area-efficient components such as the switched capacitor and the duty-cycled resistors shown in Figure 4.15.

Calculating equivalent resistance and ripple amplitude of switched capacitor [26] and duty-cycled resistor[4]

Before going into simulations, let us calculate the equivalent resistance and voltage ripple caused by these switched elements. As shown in Figure 4.15, a switched capacitor circuit operates in two phases: charging the capacitor from V_{in} and discharging to ground. Parasitic capacitors of the switches are negligible due to the FDSOI technology. The charge transferred from the capacitor during one switching cycle is transferred at a rate of f_{clk}.

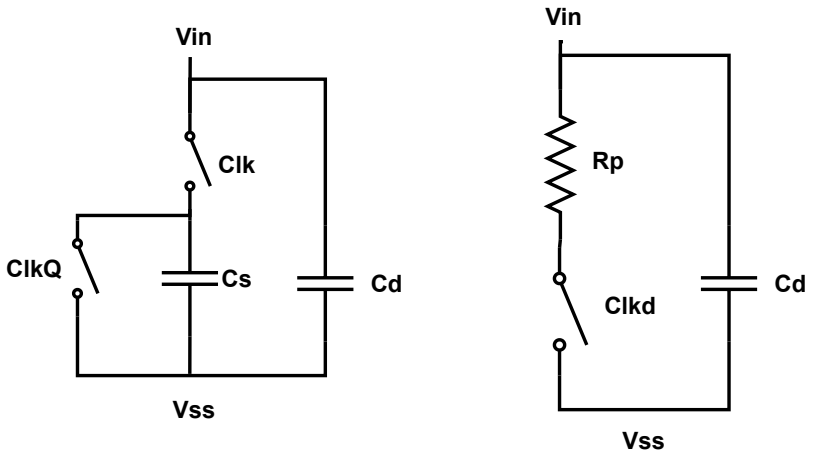


Figure 4.15: Switched capacitor resistor (left) and duty-cycled resistor (right)

The equivalent resistance of the switched capacitor is obtained as [26]:

$$R_{\text{sc}} = \frac{V_{\text{average}}}{I_{\text{average}}} \approx \frac{1}{C f_{\text{clk}}} \tag{4.15}$$

As shown in Section 3.1.3, the ripple amplitude of Vin is roughly given by:

$$\text{Ripple Amplitude} = V_{in} * \frac{C_{\text{s}}}{C_{\text{d}}} \tag{4.16}$$

For the duty-cycled resistor, the circuit operates in two phases. Similar to the switched capacitor approach, the parasitic capacitors of both the MOS switch and the resistors are negligible due to the use of FDSOI technology.

The equivalent resistor is obtained as [4]:

$$R_{\text{SR}} = \frac{V_{\text{in}}}{I_{\text{average}}} \approx \frac{R_{\text{p}}}{\delta}$$

Where δ is the clock duty cycle. Similar to the switched capacitor, the ripple amplitude can roughly be estimated as:

$$\text{Ripple Amplitude} = V_{in} * \frac{T_{\text{on}}}{C_{\text{d}} * Rp} \tag{4.17}$$

However, the exact voltage ripple can be obtained as:

$$\text{Ripple Amplitude} = \frac{1}{C_d} \int_0^{T_{\text{on}}} I_r(t)\, dt \tag{4.18}$$

From Equations 4.16 and 4.17, we can derive the general p-p ripple amplitude estimation for both the switched capacitor and the duty-cycled resistor as:

$$\text{Ripple Amplitude} = V_{in} * \frac{1}{C_{\text{d}} * f * Req} \tag{4.19}$$

Therefore, increasing the smoothing capacitor and clock frequency will result in a reduction of the ripple amplitude. Additionally, since the equivalent resistance (R_{eq}) is identical for both the switched capacitor and the duty-cycled resistor, they will have the same ripple amplitude for a given capacitance (C_d) and clock frequency.

Comparison and Selection Criteria

To determine whether to use a switched capacitor or a duty-cycled resistor in our design, we need to compare key parameters, including the robustness of the final bandgap voltage against PVT variations and area efficiency.

Regarding area efficiency, if we neglect the external required blocks (duty cycle generator for the duty-cycled resistor and oscillator for the switched capacitor), using narrow high-resistance polysilicon resistors and metal-oxide-metal (MOM) capacitors results in nearly the same area consumption for both components. Assuming a clock frequency of 1 kHz, 520 fF capacitance for Cs and 19 M Ω resistance for Rp with 1% duty-cycle is required to generate 1.9 G Ω resistor.

To dampen ripples, we use large damping capacitors (Cd) in parallel with both the switched capacitor and the duty-cycled resistor. As explained in Section 3.1.3, the required Cd can be halved with switched capacitors to get the same ripple reduction, leading to being more area efficient.

On the other hand, in the switched capacitor resistor, one switch is connected between a non-ground node and Vin. Hence, switch non-linearity with respect to the variations in Vin will cause the equivalent resistor to vary over PVT unless transmission gates with doubled clock signal swing [23] are used. To avoid design complexity and have less power consumption, duty-cycled resistors are used in this design.

Given the large required resistances, we need a very small duty cycle. For example, considering $R_{\text{CTAT}} = 1.9\ \text{G}\Omega$, a 1% duty cycle requires a poly resistor value of $\frac{1}{100} \times 1.9\ \text{G}\Omega = 19\ \text{M}\Omega$. Although the equivalent resistance is not related to the clock frequency, it must be chosen carefully. While higher clock frequencies make filtering the ripples easier, they consume more power and reduce the ON phase duration, potentially going below the RC time constant. Moreover, generating a very low duty cycle will consume more power as will be explained in Section 4.1.8. Balancing this trade-off, we selected a clock period of 1 ms and a pulse width of 10 µs.

Ring oscillator for duty-cycled resistor

We need to generate a 1% duty cycle with ultra-low power consumption. To verify this, a test-bench for the ring oscillator was built to achieve the duty-cycled clock from two consecutive clock edges as shown in Figure 4.16. In a ring oscillator, $T_{clk} = 2N \times T_{\text{delay}}$, where T_{delay} is the delay caused by each inverter. To achieve a 1% duty cycle, $T_{\text{delay}} = \frac{T_{clk}}{100}$, leading to $N = 50$. Using a NAND gate, we can derive a clock with a 1% duty cycle from this ring oscillator.

After determining the configuration, the next step is minimizing power consumption. The power consumption of the ring oscillator includes three parts: switching current, short-circuit current, and leakage current. The switching current is given by $I = NC_L \times F_{\text{clk}} \times V_{\text{dd}}$. To reduce this, we should minimize the supply voltage V_{dd} and the load parasitic capacitance of the inverters, which mainly consists of gate parasitic capacitances.

First, we reduced the supply voltage to 500 mV, the lowest possible without requiring a level shifter. We would like to use minimum-size MOSFETs for the inverters to minimize parasitic capacitance. However, lower lengths increase the current drawn from the supply and the clock frequency. Hence, we used maximum length and minimum width devices. This approach also minimized leakage currents.

To further reduce leakage currents, we explored different bulk and deep n-well (DNW) connections for PMOS and NMOS. In order to have an area-efficient layout, we are limited to two options of connecting both the DNW of NMOS and the bulk of PMOS to either V_{dd} or V_{ss}. Initially, connecting the bulk of PMOS to V_{dd} resulted in more than 50 pA leakage per inverter. Switching both the PMOS bulk and NMOS DNW to V_{ss} significantly reduced total power consumption by minimizing these leakage currents.

The final result signals are shown in Figure 4.17. The total current consumption is 3 nA in the nominal corner and 8 nA in the worst-case corner (FFFF125).

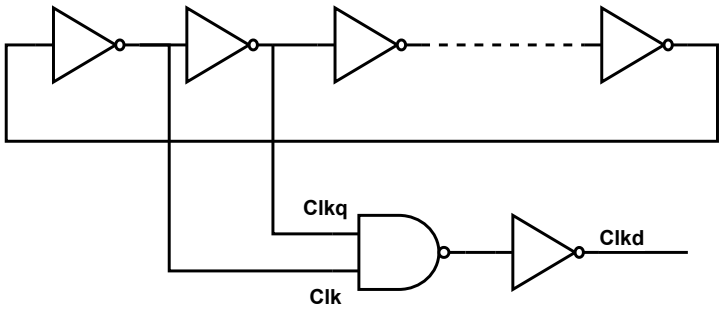


Figure 4.16: Ring oscillator with 51 inverters to generate 1% duty cycle

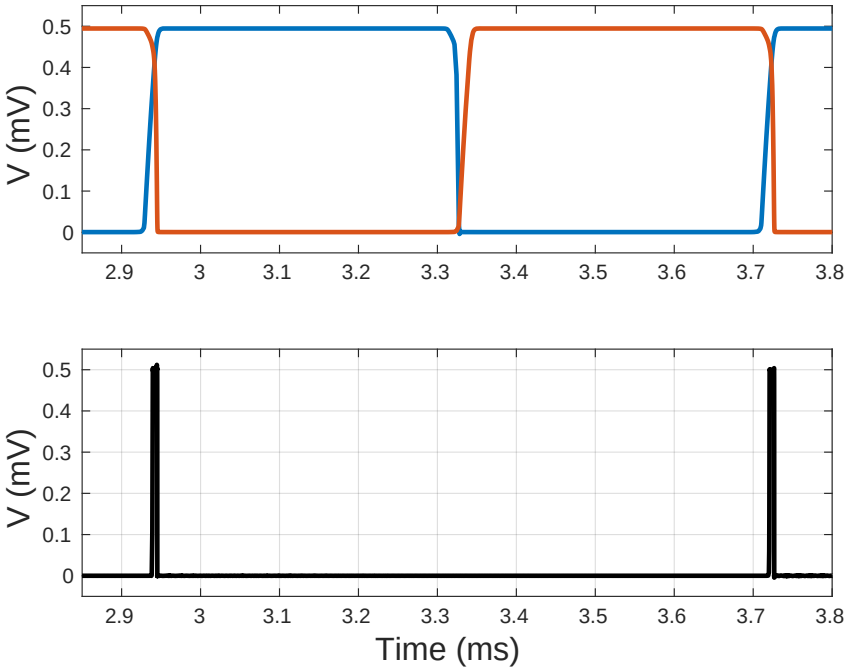


Figure 4.17: Ring oscillator two consecutive clock signals and the final clock with 1% duty cycle

In conclusion, using duty-cycled resistors with reasonable power consumption for external blocks is feasible.

4.1.9 Sizing of duty-cycled resistors in the bandgap design

As discussed earlier, to dampen the ripples in duty-cycled resistors, we use damping capacitors in parallel with them. These parallel capacitors charge during the off phase and discharge through the poly resistor during the on phase. Hence, the higher the current in one branch, the larger the capacitor required to maintain the same ripple level. The PTAT current is nearly twice the CTAT current at room temperature, necessitating a parallel capacitor nearly twice as large for the PTAT resistor. The ripple voltage in Vptat and Vbe will be attenuated by the closed loop gain of the OTA at 1 kHz (Figure 4.18) and their effect is negligible on Vref. On the other hand, the reference voltage will be connected to the LDO and the regulated voltage will be used to supply the analog blocks in the system. The target ripple at the output of the LDO is specified as $\pm 5mV$. Based on the system-level design and attenuation of the reference voltage ripple in the output of the LDO, the target ripple for Vref is $\pm 20mV$. This is achieved with a 20pF smoothing capacitor.

We can make the ratio between the sizes of the switches the same as the ratio between the poly resistors to maintain the same ratio between the R_{on} of their switches. However, as long as the R_{on} of the switches is in a few kΩ range, leakage currents are more important to be considered individually for each resistor. Single NMOS switches are used for the RCTAT, RPTAT, and Rref.

Before finalizing the values for the parallel capacitors and optimizing the switch sizes, the stability of the bandgap loop is reviewed considering the added capacitors. A stability (STB) analysis was done on the bandgap schematic with ideal resistors and 20 pF parallel capacitors with them. The loop became unstable after adding a small capacitor in parallel with R_{CTAT}. Consequently, the values of the compensation capacitors are adjusted to 2 pF for C_3 and 10 pF for C_2 (Figure 4.9). The new stability results are shown in Figure 4.18, and the phase margin is maintained more than 60 degrees.

With these new compensation capacitor values, the bandgap loop has sufficient margin to remain stable after adding parallel damping capacitors.

A corner analysis over the temperature range was done to fine-tune the switch sizes and damping capacitor values based on the target ripple and leakage currents to ensure minimal variations with temperature and process corners. With a clock signal of 1 kHz with a 1% duty cycle, the poly resistors were set to 800kΩ, 19 MΩ, and 8MΩ for RPTAT, RCTAT, and Rref respectively. The optimized switch sizes and parallel capacitor values are shown in the bandgap schematic of Figure 4.19.

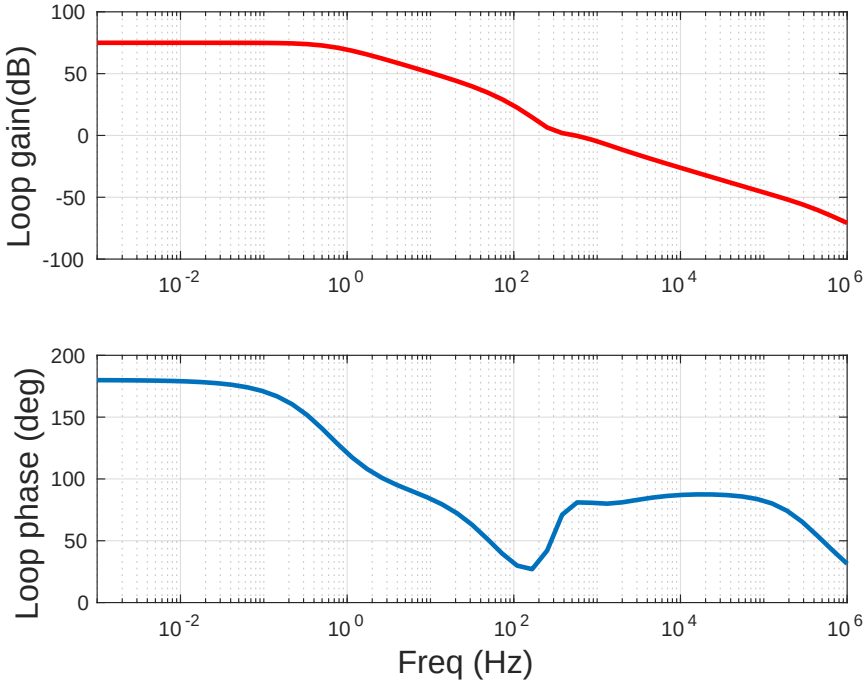


Figure 4.18: Stability results of the bandgap loop after adding parallel capacitors.

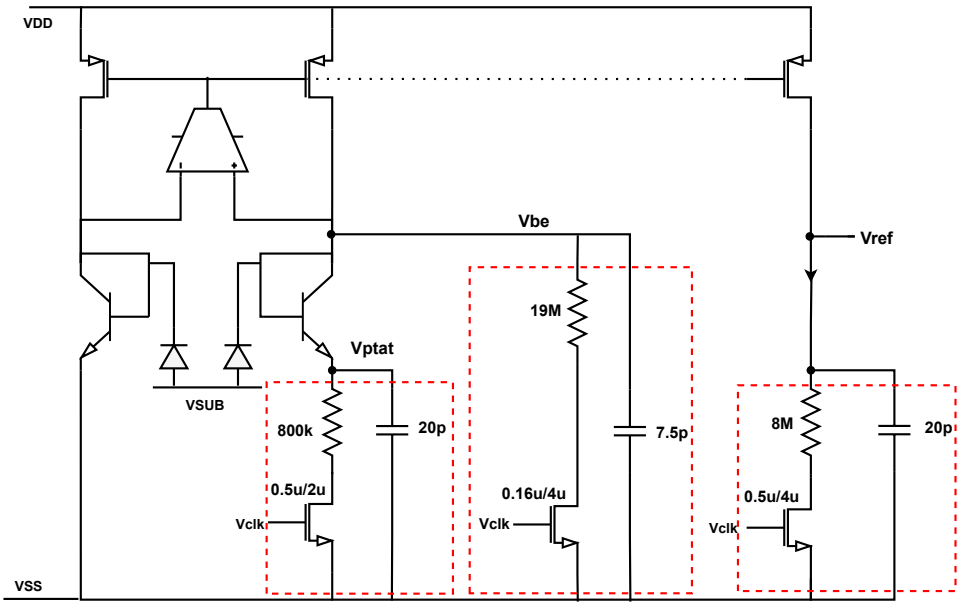


Figure 4.19: Bandgap schematic with optimized switch sizes and parallel capacitor values.

4.2 Simulation results

This section presents the simulation results for the designed bandgap reference circuit with duty-cycled resistors. The simulations include transient analysis, DC analysis, AC analysis, stability (STB) analysis, Monte Carlo analysis, and noise analysis.

4.2.1 Transient Analysis

The transient response shows the startup behavior and the time required to reach steady-state operation. In Figures 4.21 and 4.20, the transient response for important voltages is shown for different process corners: SSSS-40, TTTT27, and FFFF125. The first four letters represent different components including NMOS, PMOS, resistors/capacitors, and diodes/ bipolars, respectively. The following numbers show the temperature. In the simulation, V_{DD} is ramped up over 1 ms, and the enable signal is turned on for 100 ms afterward. The worst-case startup, which occurs for the SSSS-40 corner, takes nearly 40 ms to stabilize. Bandgap core branch current is also shown in Figure 4.22. The maximum current driven during startup is 80nA at FFFF125 and the minimum is 1nA at SSSS-40.

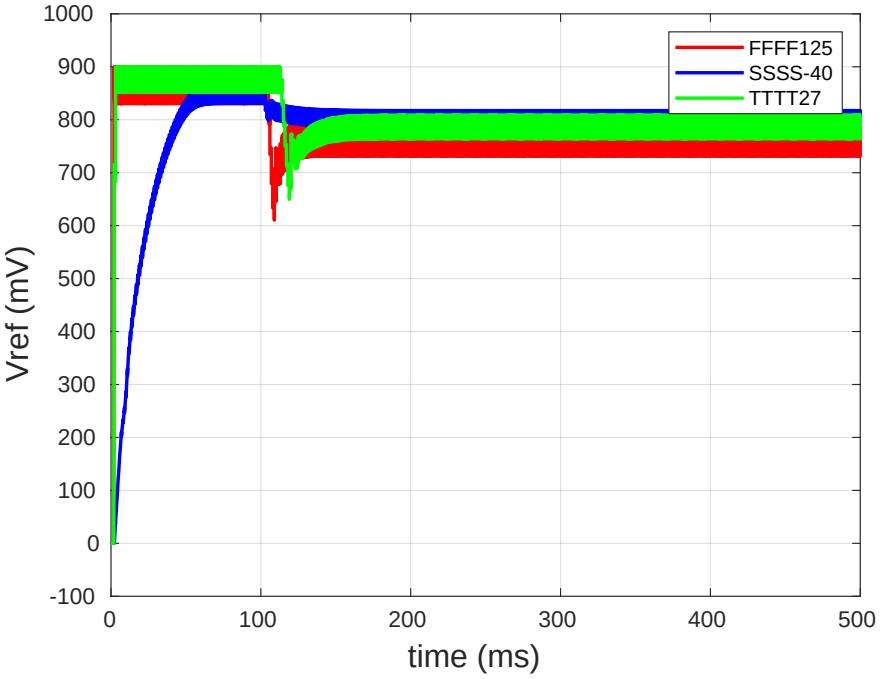


Figure 4.20: Transient response of V_{ref}

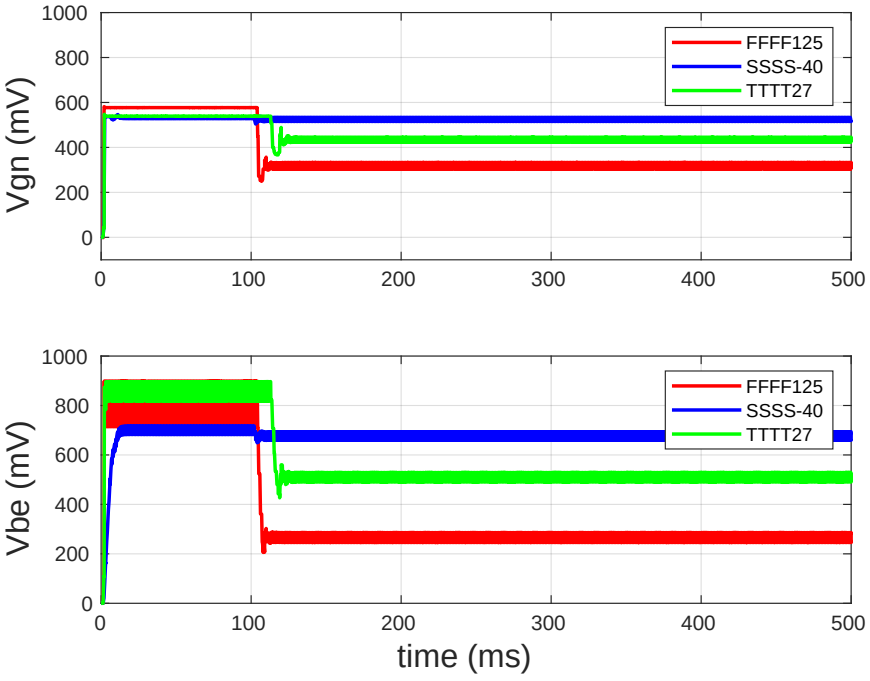


Figure 4.21: Transient response of V_{be} and V_{gn}

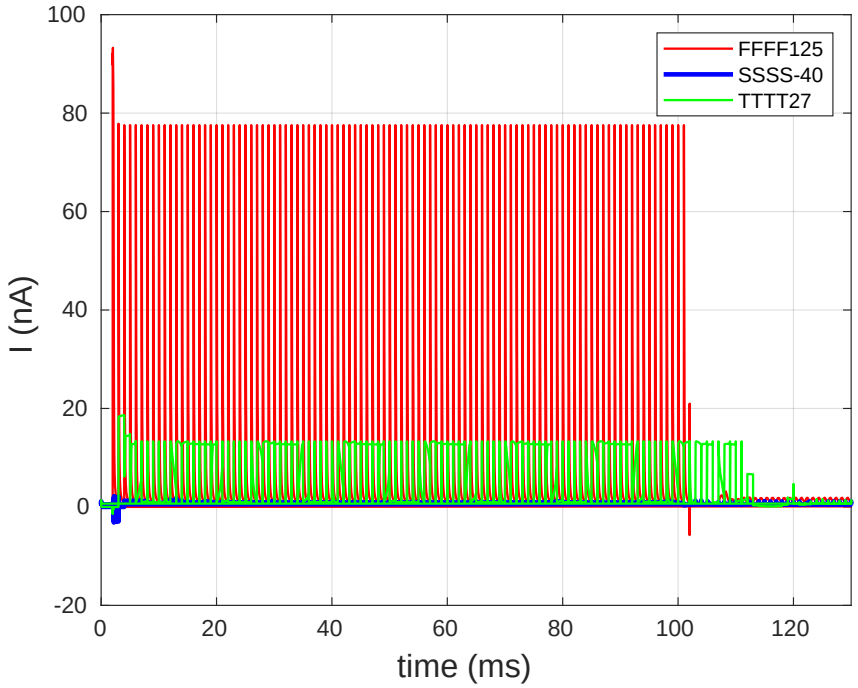


Figure 4.22: Transient response of bandgap current

4.2.2 DC Analysis

Figure 4.23 shows the bandgap current over the temperature range for different process corners. As discussed in the design methodology chapter, the output current varies across different process corners due to the resistor process variations, which are approximately $\pm 30\%$ for narrow high-resistance polysilicon resistors. However, since the same process variation affects the reference resistor, it cancels out in the bandgap voltage. Figure 4.24 shows the bandgap voltage reference over the temperature range for different process corners.

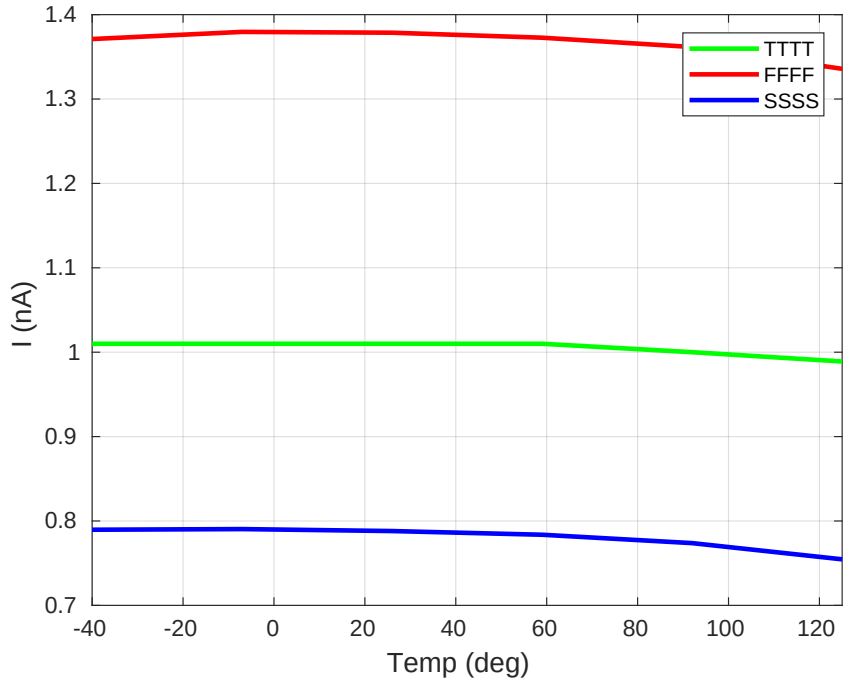


Figure 4.23: Output current over the temperature range for different process corners

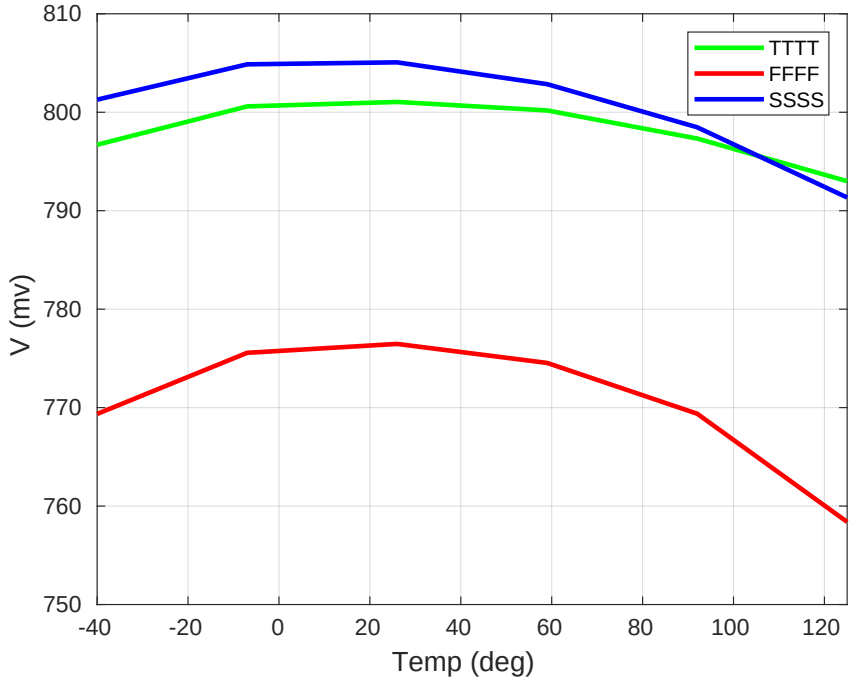


Figure 4.24: Bandgap voltage reference over the temperature range for different process corners

Table 4.3 shows the temperature coefficient (TC) of the bandgap voltage and current in different process corners obtained with Equation 4.14 using the box method. The maximum process variations of the bandgap voltage and current are also summarized in Table 4.4. As shown, the process variations of the bandgap current are significantly compensated in the output voltage, reducing from an average of 30% to 2%. The remaining variations are attributed to leakage currents and non-linearities caused by the switches in the duty-cycled resistors, V_{BE} of bipolar transistors, and offset of the OTA.

Process Corner	TC of I_{ref} (ppm/°C)	TC of V_{ref} (ppm/°C)
FFFF	189	140.6
TTTT	116	60
SSSS	260	103

Table 4.3: Temperature coefficient (TC) of bandgap voltage and current in different process corners.

Parameter	Max Process Variations
V_{ref}	+0.5%/-3%
I_{ref}	+36.8%/-23.5%

Table 4.4: Maximum process variations of bandgap voltage and current.

The line sensitivity (LS) of the voltage reference is defined using the box method, as follows:

$$\text{LS} = \frac{(V_{\text{REF,max}} - V_{\text{REF,min}})}{V_{\text{REF,avg}}(V_{\text{DD,max}} - V_{\text{DD,min}})} \times 100(\%/\text{V}) \tag{4.20}$$

The reference voltage variation over the supply voltage range is shown in Figure 4.25. Line sensitivity is measured to be 0.56%/V for a supply voltage range from 900 mV to 1.98 V.

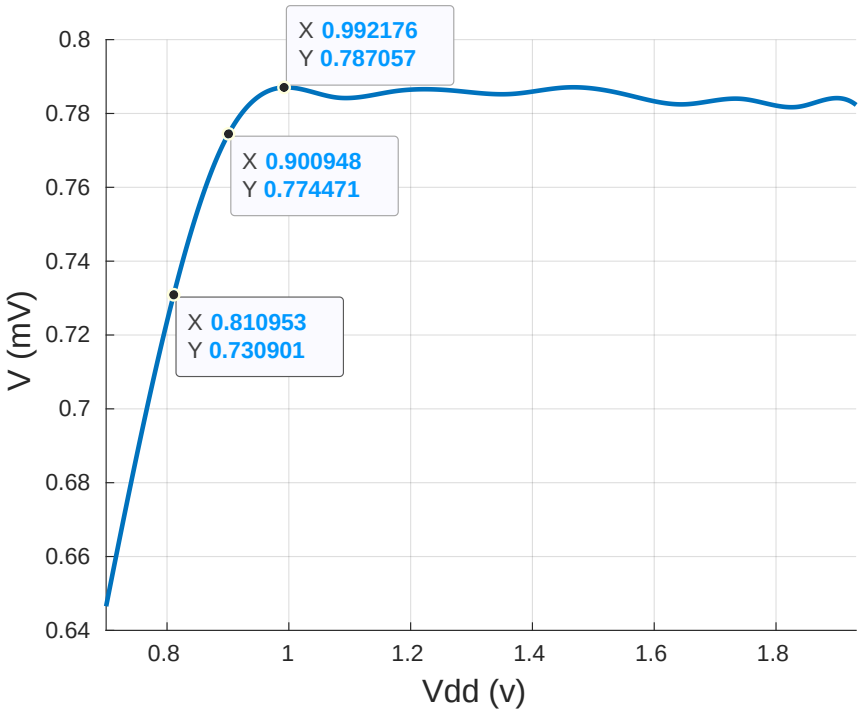


Figure 4.25: Reference voltage variation over the supply voltage

4.2.3 AC Analysis

With AC analysis, we evaluate the frequency response of the bandgap reference to measure the power supply rejection ratio (PSRR). The PSRR is calculated as:

$$\text{PSRR} = 20\log\left(\frac{\text{output voltage variation}}{\text{power supply variation}}\right) \tag{4.21}$$

The PSRR curves for FFFF125, TTTT27, and SSSS-40 are shown in Figure 5.6. Given our supply voltage of 900 mV and a reference voltage of 800 mV, there is a 100 mV headroom for the output PMOS mirror. The DC PSRR is -10 dB, which is primarily limited by the output conductance (g_{ds}) of the PMOS mirror branch. The PSRR curves for FFFF125, TTTT27, and SSSS-40 are shown in Figure 5.6.

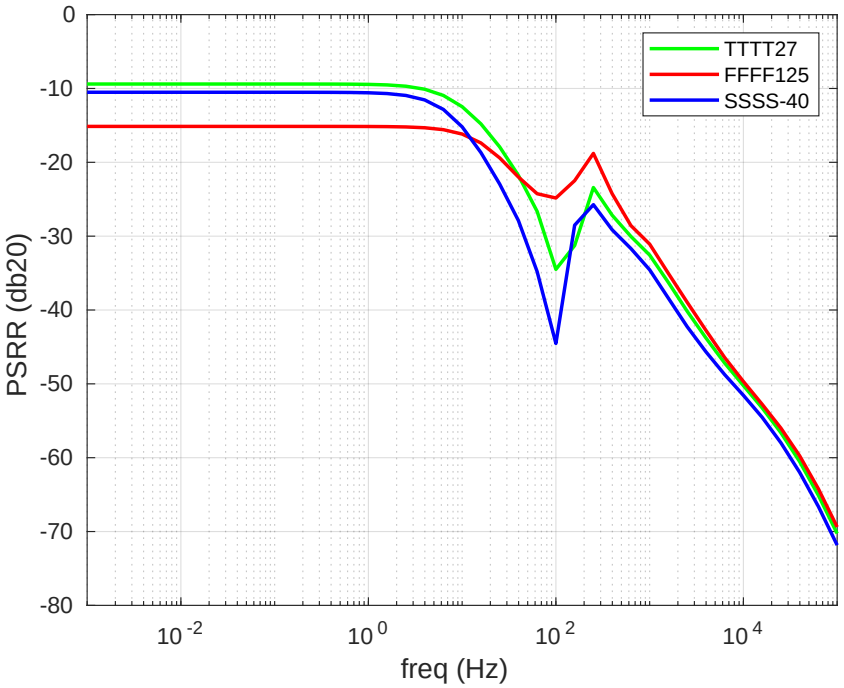


Figure 4.26: PSRR curves for different process corners

4.2.4 Stability Analysis

Figure 4.27 shows the stability analysis of the bandgap core loop comprising both negative feedback loop (MP2 and OTA positive input) and positive feedback loop (Mp1 and OTA negative input) in different process corners: TTTT27, FFFF125, and SSSS-40. The analysis indicates a phase margin of more than 60 degrees across all corners and temperatures, with the DC gain ranging from a minimum of 60 dB to a maximum of 75 dB. These results demonstrate the stable operation of the bandgap reference circuit under PVT variations.

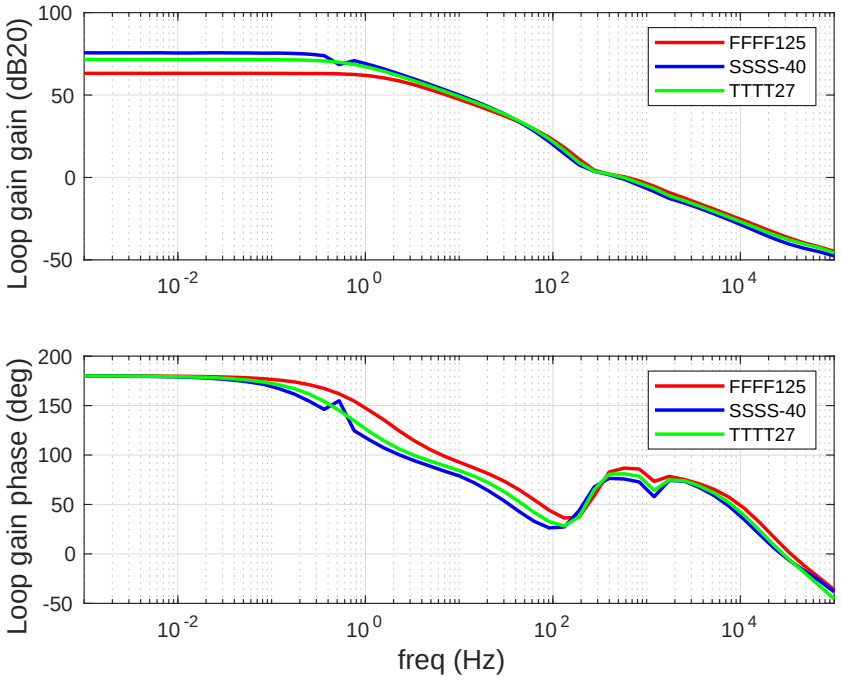


Figure 4.27: Bandgap core loop stability analysis in different process corners

4.2.5 Noise Analysis

The noise analysis of the bandgap output voltage is shown in Figure 4.29 for different process corners: FFFF125, TTTT27, and SSSS-40. The maximum integrated output noise from 0.1 Hz to 100 kHz is measured to be 940 μV_{RMS}. The main noise contributors are the PMOS mirror of the core bandgap and the PMOS self-cascoded composites of the OTA as shown in Figure 4.28.

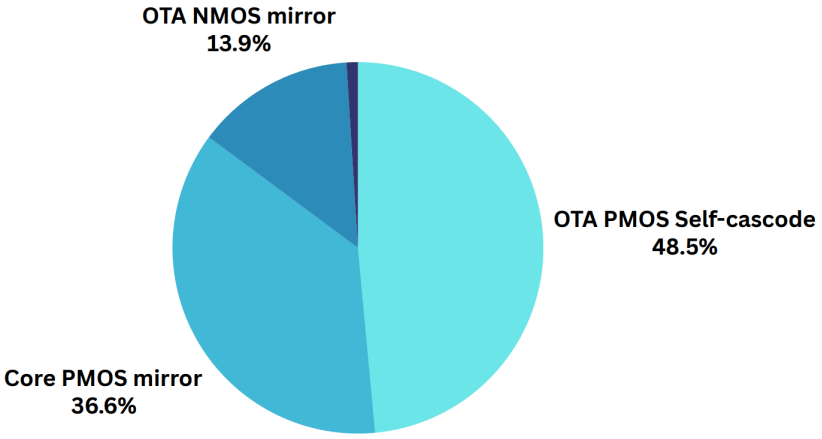


Figure 4.28: Pie chart of the main noise contributors to bandgap reference voltage

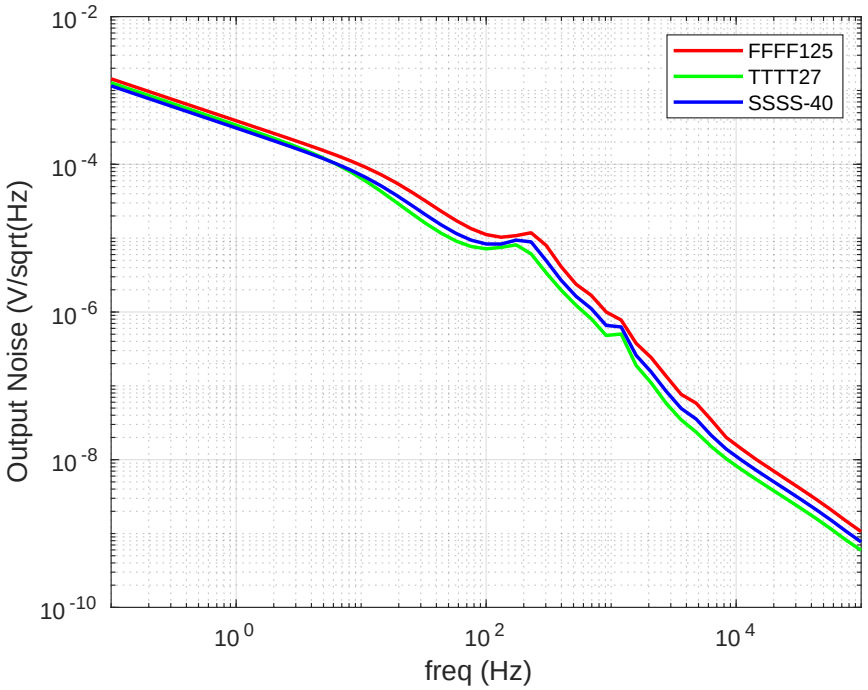


Figure 4.29: Output voltage noise density of bandgap reference

4.2.6 Monte Carlo Analysis

The Monte Carlo analysis was done with 200 points with both mismatch and process variations. The reference voltage $\pm 3\sigma$ spread over temperature and histogram results at 40 degrees for the CWT current and the PTAT, CTAT, and REF duty-cycled resistors are shown below. As shown in Figure 4.30 and Figure 4.31, the σ/μ ratio is 2.9% for the reference voltage and 11.6% for the output current. Also, the σ/μ ratio for all switched resistors is nearly 10% as shown in Figure 4.32.

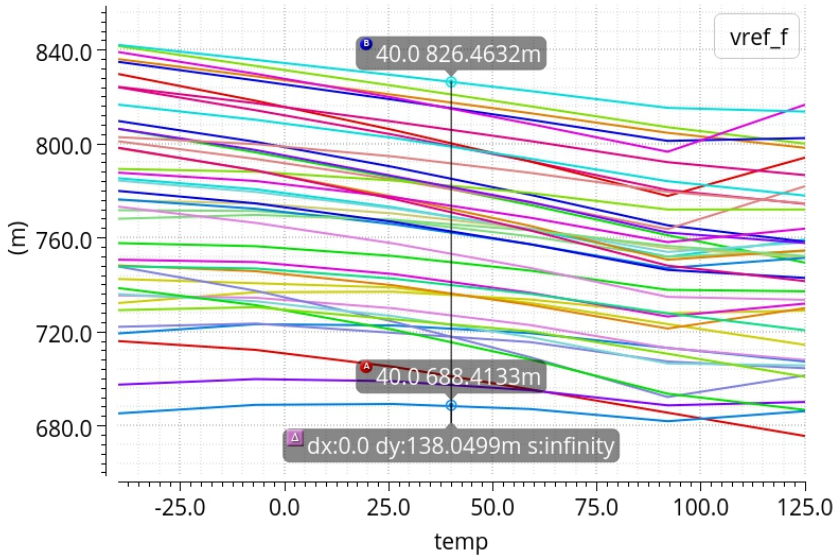


Figure 4.30: Monte Carlo simulation results for the reference voltage over temperature

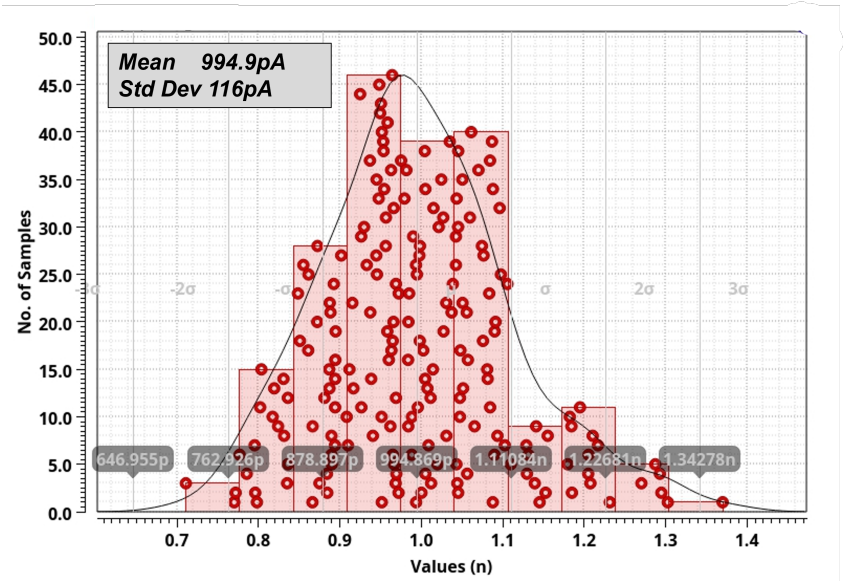
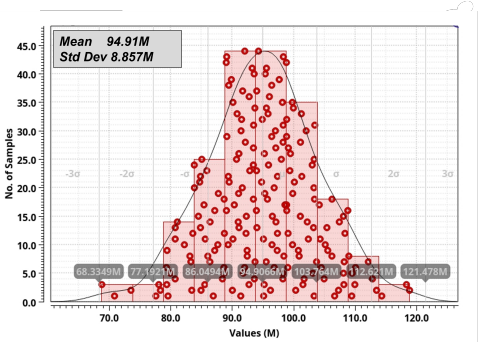
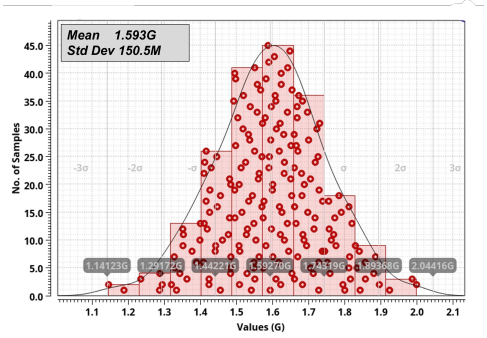


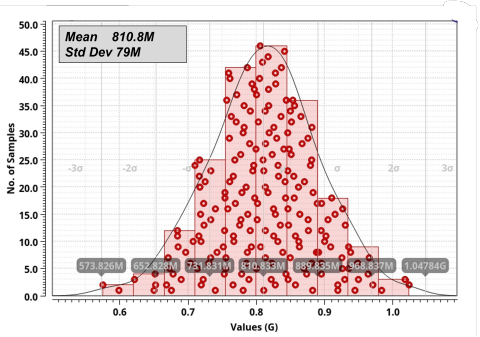
Figure 4.31: Monte Carlo simulation results for the output current.



(a) RPTAT Histogram



(b) RCTAT Histogram



(c) RREF Histogram

Figure 4.32: Monte Carlo simulation results for the duty-cycled resistors: (a) RPTAT, (b) RCTAT, and (c) RREF histograms.

4.2.7 Power Consumption Distribution

The power consumption distribution with the nominal supply voltage of 900mV is as follows:

	Min	Typ	Max
Bandgap core	2.2 nW	2.7 nW	3.5 nW
OTA and biasing	2.7 nW	3.4 nW	4.5 nW
Leakage compensation	3.2 pW	4.5 pW	227 pW
Ring oscillator	650 pW	1.5 nW	4 nW
Total	5.5nW	7.6 nW	12 nW

Table 4.5: Power consumption distribution of the bandgap reference

5 Low-dropout regulator (LDO)

5.1 Design methodology

Low Dropout Regulators (LDOs) are necessary components in many electronic systems, providing stable and efficient voltage regulation with minimal power dissipation.

LDOs function by maintaining a constant output voltage even when the input voltage fluctuates, ensuring that sensitive electronic components receive a steady power supply. Unlike traditional linear regulators, LDOs can operate with a very small difference between the input and output voltages, making them ideal for battery-powered devices and systems with high efficiency.

In our design, the first objective is to develop an LDO with ultra-low power consumption while operating at a low supply voltage. The design specifications, choices, and results will be detailed in the next section.

5.1.1 Specifications

The design specifications for the LDO are shown in the table below:

Parameter	Specification
Supply voltage range	0.9 V to 1.98 V
Load current range	100 nA to 1 μA
Load capacitor	100 pF
Output voltage	0.8 V
Accuracy	$\pm$2.5%
Temperature range	-40°C to 125°C

Table 5.1: LDO Design Specifications

The initial step in the design process is to select an appropriate topology that aligns with these specifications.

5.1.2 Design choices

System-Level view

The first design consideration involves the system-level integration of the bandgap reference and the LDO as shown in Figure 5.1. The decision revolves around whether to place the bandgap reference before or after the LDO, or whether it should be connected to the battery voltage or regulated voltage.

Placing the bandgap reference after the LDO results in a more complicated startup scenario. First, reference voltage should be generated from the battery voltage and then LDO can operate and BG reference can be supplied from the output of the LDO. On the other hand, the noise from the bandgap output voltage will also be amplified by the resistive ratio, leading to increased noise in the regulated output. Also, there will be less headroom available for PMOS mirrors in bandgap topology (Figure 4.1a) as it is supplied from 800mV regulated Vout.

Due to these considerations, the configuration where the bandgap reference is placed before the LDO and supplied from the battery voltage is the preferred option. This setup avoids the startup complications and noise amplification associated with placing the bandgap reference after the LDO and using resistive feedback and offers more headroom for bandgap voltage.

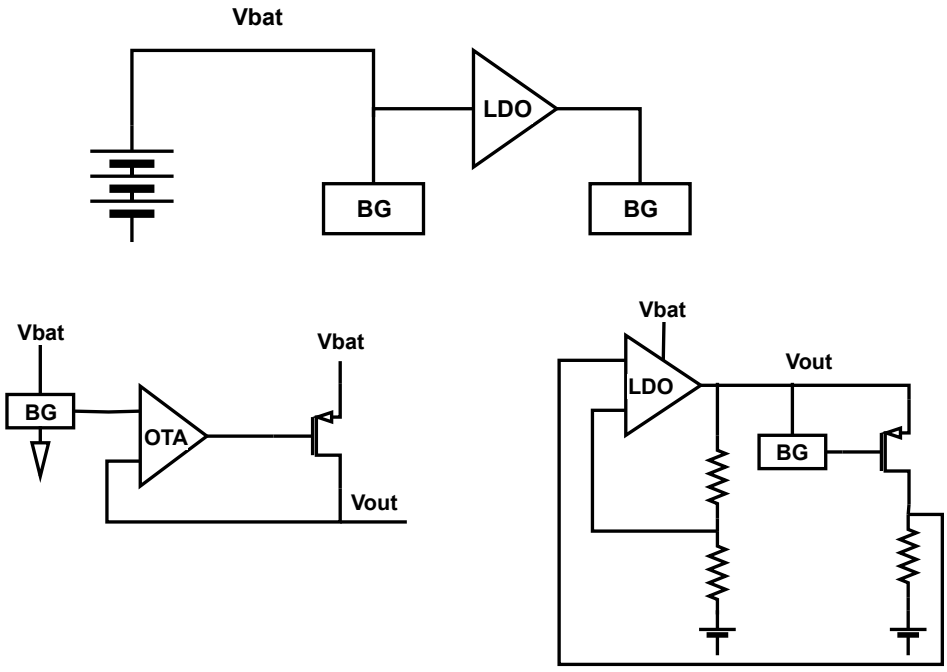


Figure 5.1: System-level options for integrating bandgap reference and LDO

Choice of Pass Transistor

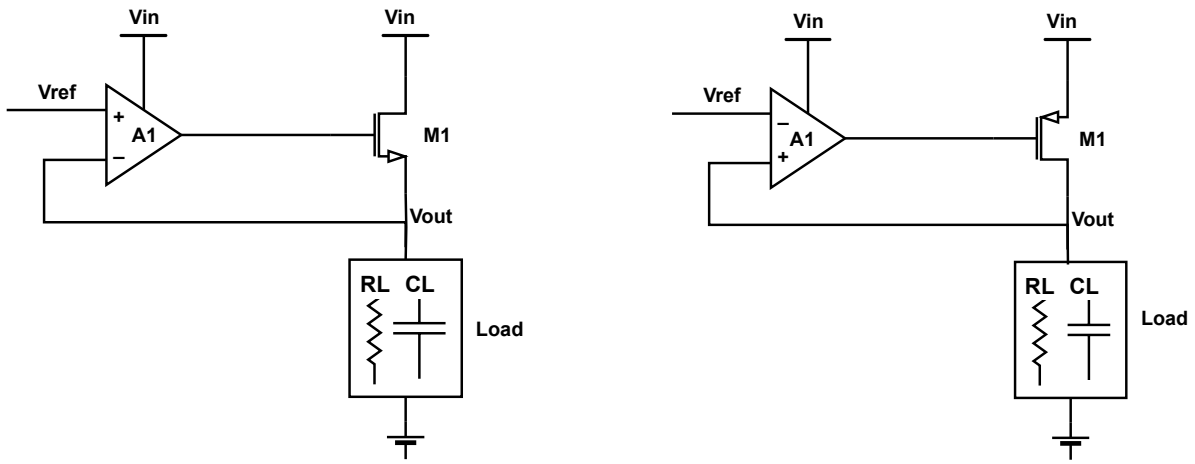
The next design choice is the selection of the pass transistor, M_1. The basic LDO topology is shown in Figure 5.2. The unregulated voltage, V_{in}, is applied to M_1 and controls the current flow through it by its V_{gs}, which is regulated by the operational amplifier (opamp) such that the two inputs, V_{out} and V_{ref}, remain close to each other.

M_1 in Figure 5.2 can function as a controlled current source or as a source follower, as illustrated in Figures 5.2a and 5.2b, respectively. In the case of the source follower topology (Figure 5.2a), neglecting channel length modulation, $\frac{\partial V_{out}}{\partial V_{in}} \rightarrow 0$ since the changes in the drain voltage do not affect the source voltage as long as the transistor remains in saturation. This provides very high PSRR. However, this topology results in a much higher dropout voltage ($V_{out} - V_{in}$). This is because the gate voltage can be at most equal to V_{in}, and $V_{out} = V_{in} - V_{gs}$, which includes the threshold voltage of M_1.

Given the design specifications, the lowest dropout voltage required is:

$$V_{\text{dropout}} = V_{\text{supply(min)}} - V_{\text{out}} = 0.9\,\text{V} - 0.8\,\text{V} = 100\,\text{mV}$$

Therefore, we should employ the current source topology for M_1.



(a) Source follower topology

(b) Current source topology

Figure 5.2: Pass transistor topologies: (a) Source follower, (b) Current source

OTA topology

Based on the OTA comparison done in the bandgap reference design section, the pseudo-folded-cascode topology is chosen. This topology offers several advantages, including area-efficient compensation techniques with available low-impedance nodes, sufficient DC gain, and a high output dynamic range. Having already designed this topology for the bandgap reference, we now need to optimize the compensation capacitors to ensure a stable closed-loop operation in the LDO design.

5.1.3 Design steps

To size M_1 depicted in Figure 5.2b, it should be considered that it must handle currents ranging from 100 nA to 1 µA while the supply voltage varies between 900 mV and 1.98 V. For a fixed input voltage of 900 mV, the W/L ratio should be large enough to achieve a reasonable V_{GS} to provide the maximum load current of 1 µA and small enough to keep the gate voltage of M_1 in the OTA output dynamic range. This translates to a W/L ratio greater than 9u/1u, with maximum and minimum values of V_{GS1} equal to 680 mV and 140 mV respectively, over load current range and PVT variations.

Other important parameters to consider when designing the LDO include power supply rejection ratio (PSRR) and load regulation. Referring to the LDO schematic in the Figure 5.2b, we begin by defining line regulation through the closed-loop gain, which can be expressed as:

$$A_{LG} = A_1 \cdot g_{m1} \cdot R_L \tag{5.1}$$

where R_L is the load resistance. If we consider the pass transistor as a common gate stage with an open-loop gain of $g_{m1} \cdot R_L$, placing it in a negative feedback loop results in:

$$\frac{V_{out}}{V_{in}} = \frac{g_{m1} \cdot R_L}{1 + A_{LG}} \approx \frac{1}{A_1} \tag{5.2}$$

Thus, line regulation can be improved by increasing the opamp gain, assuming infinite PSRR for the opamp. Considering the finite output impedance of the pass transistor (ro), the propagation from V_{in} to V_{out} can be viewed as a resistor divider:

$$\frac{V_{out}}{V_{in}} = \frac{R_{out}}{R_{out} + r_o} = \frac{1/g_{mA1} \parallel R_L}{r_o + 1/g_{mA1} \parallel R_L} \tag{5.3}$$

Load regulation is defined as $\frac{\partial V_{out}}{\partial I_{load}}$, which corresponds to the output impedance of the LDO, R_{out}. Considering the pass transistor as a diode-connected transistor with boosted transconductance, the output impedance is:

$$R_{out} = \frac{1}{g_{mA1}} \tag{5.4}$$

Both load and line regulation degrade as the opamp gain drops after the 3 dB bandwidth, leading to increased output impedance at high frequencies. This causes significant variations in the regulated voltage due to changes in the load current. To mitigate this, a smoothing capacitor (C_L) in Figure 5.2b is necessary at the output to provide a low-impedance path for transient currents. Based on our design specifications, the load capacitor is 100 pF.

The stability of the loop is evaluated under the worst-case scenario, which occurs at the lowest load current (100 nA). This is due to the low transconductance (g_m) and high output impedance of the LDO, which create a low-frequency pole. Given the presence of two high-impedance nodes, we employ Miller compensation (C1) for pole splitting and indirect compensation (C2 and C3) to introduce left-half-plane (LHP) zeros. Optimal stability was achieved using the capacitor values depicted in Figure 5.3. This configuration ensures having more than 60 degree phase margin and 60dB DC gain for all load currents, as will be shown in the results section.

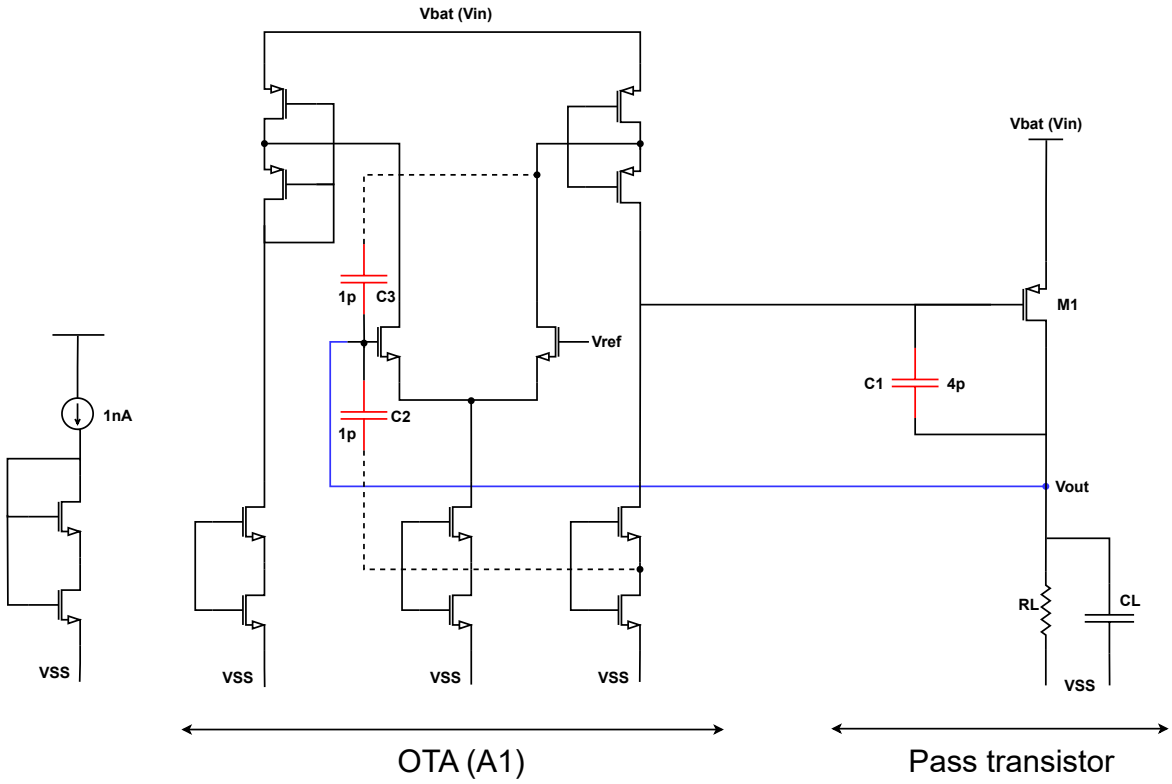


Figure 5.3: LDO schematic with capacitor values used for optimal stability.

5.2 Simulation and results

In this section, the simulation results of the LDO output voltage are presented when the reference voltage from the bandgap block is connected to the LDO. Additionally, system-level startup signals will be explained at the end of this section.

5.2.1 DC simulation

The difference between the output voltage of the LDO and the bandgap reference voltage is illustrated in Figure 5.4. The results are over the temperature range and in different process corners and indicate a maximum variation of 2 mV between the reference voltage (V_{ref}) and the LDO output voltage fitting within the accuracy specification ($\pm$2.5%). For this testbench, V_{DD} was set to 900 mV and the load current was maintained at 1 µA.

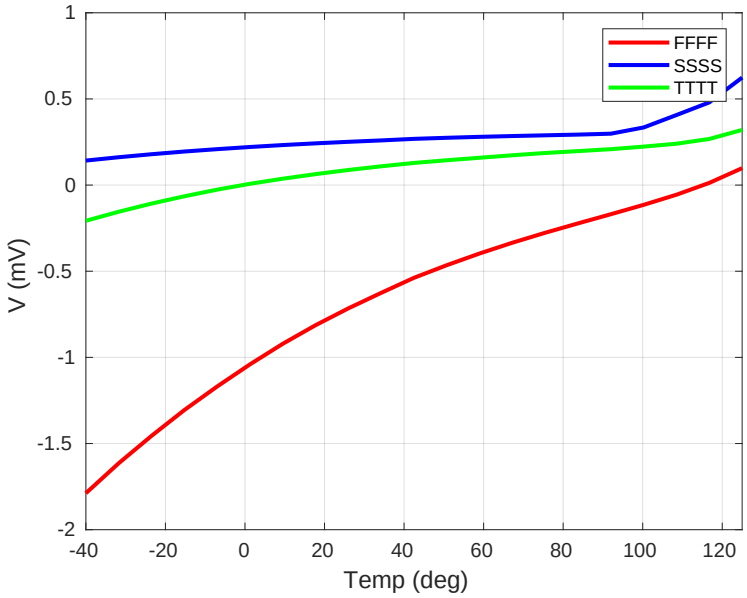


Figure 5.4: V_{ref} - V_{out} over temperature range and process corners

The output voltage of the LDO as a function of the input voltage (V_{dd}) is illustrated in Figure 5.5 for both maximum and minimum load currents. As depicted, the variation in output voltage is less than 1 mV across the load current range. Furthermore, the average line sensitivity, calculated using Equation 4.20, remains below 1% throughout the input voltage.

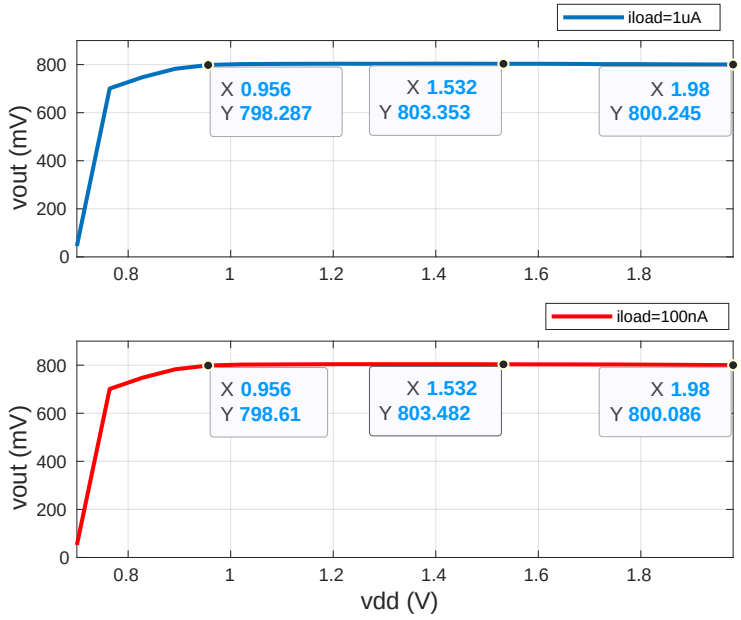


Figure 5.5: LDO output voltage versus input voltage (V_{dd}) for maximum and minimum load currents.

5.2.2 AC Analysis

The PSRR curves are presented in Figure 5.6. Both the LDO and the bandgap reference are powered from V_{in}, and the resulting PSRR for maximum and minimum load currents and different supply voltages is depicted. The minimum DC PSRR achieved is -12 dB, which happens at the end-of-life voltage of the battery (V_{in}). It is primarily constrained by the output conductance (g_{ds}) of the pass transistor and the fluctuations in the reference voltage.

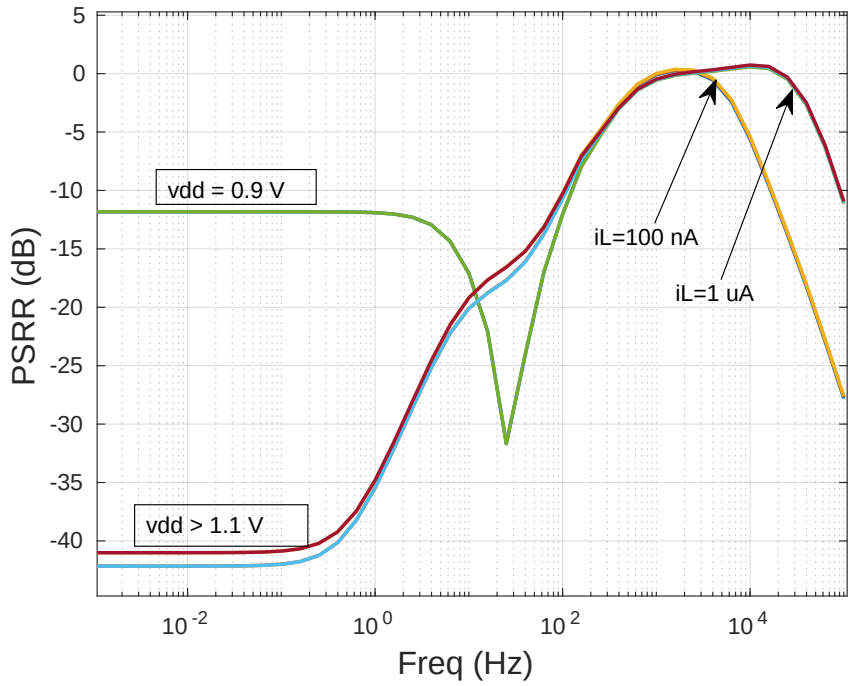


Figure 5.6: PSRR curves for different supply voltages and load currents

5.2.3 Stability Analysis

Figure 5.7 shows the stability analysis of the LDO core loop consisting of the OTA and pass transistor in different process corners, load currents, and supply voltages. The analysis indicates a phase margin of more than 60 degrees across all curves, with the DC gain ranging from a minimum of 63 dB to a maximum of 75 dB. These results demonstrate the stable operation of the LDO under supply voltage, load current, and process variations.

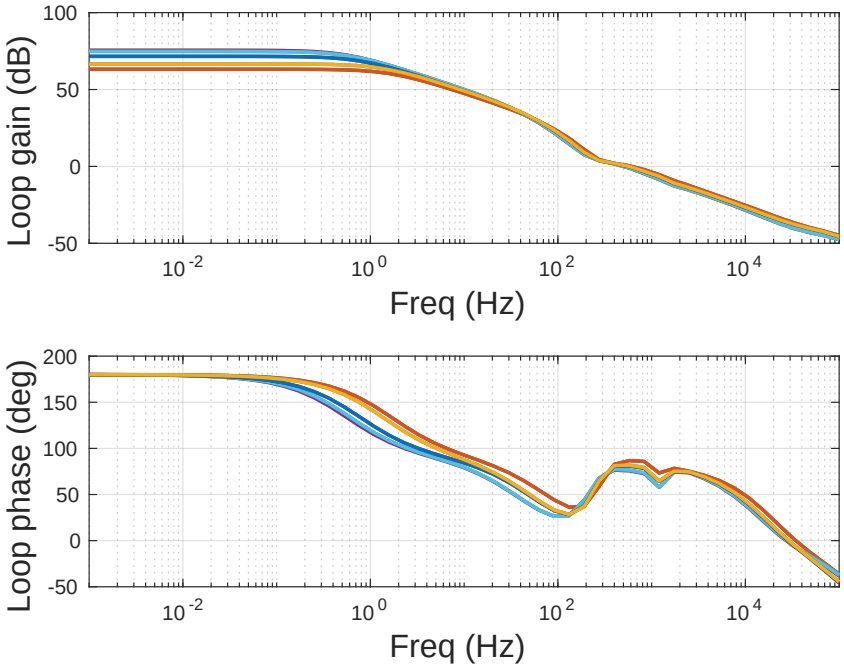


Figure 5.7: LDO loop stability analysis

5.2.4 Noise Analysis

The noise analysis of the LDO output voltage is shown in Figure 5.8 for different process corners, 100 nA to 1 uA load currents, and 0.9 V to 1.98 V supply voltages. The maximum integrated output noise from 0.1 Hz to 100 kHz is measured to be 1.3 mV_{RMS} at FFFF125.

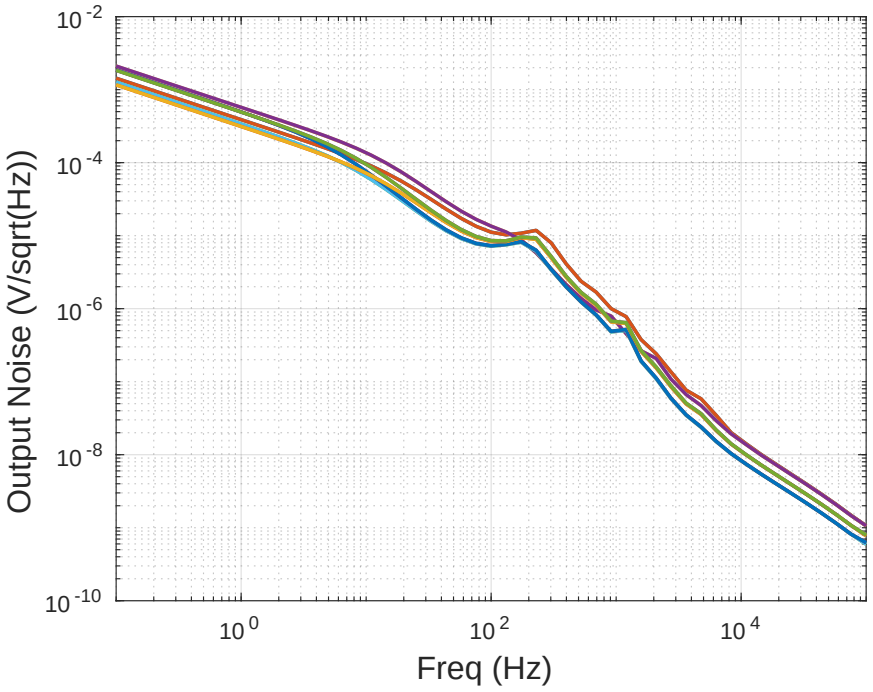


Figure 5.8: Output voltage noise density of LDO

5.2.5 Monte Carlo Analysis

The Monte Carlo analysis was done with 200 points with both mismatch and process variations specified for LDO components. The histogram result for output voltage is shown in Figure 5.9. The σ/μ ratio is 0.8% for the output voltage excluding the bandgap reference voltage variations.

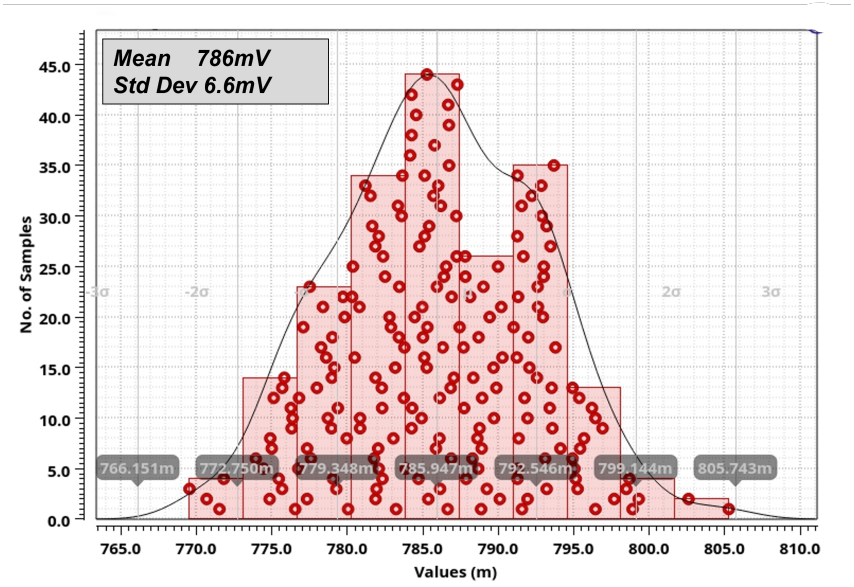


Figure 5.9: Monte Carlo simulation results for the LDO output voltage.

5.2.6 System level startup

Regarding the system-level startup, as V_{DD} ramps up, the Power-On Reset (POR) signal also ramps up, and the NPOR signal rises as soon as the POR turns off. NPOR is used as the enable signal to start the ring oscillator, which generates a 1kHz clock. This clock is then connected to a counter including a series of 9 D-type flip-flops (DFFs) to generate a clock frequency 512 times that of the ring oscillator (RCO). We allow 32 cycles (32 ms) for the RCO to stabilize and produce a precise duty cycle.

At this stage, a bandgap enable signal, which is a pulse signal with a width of 32 ms, is generated and connected to the bandgap startup circuitry. The RCO is also connected to the bandgap circuit and provides the required duty-cycled clock for the switches. An additional 200 ms is allocated for the bandgap to reach the steady state, which is more than twice the worst-case steady state time at SSSS-40, discussed in Section 4.2. Once stabilized, a bandgap-ready signal is generated, and the bandgap reference voltage and the bias current of the OTA are connected to the LDO.

The corresponding waveforms are shown in Figure 5.10. The bandgap reference voltage and output voltage of the LDO are shown in Figure 5.11 in extreme corners. The maximum startup time required for the entire system of RCO, bandgap reference, and LDO is 300 ms.

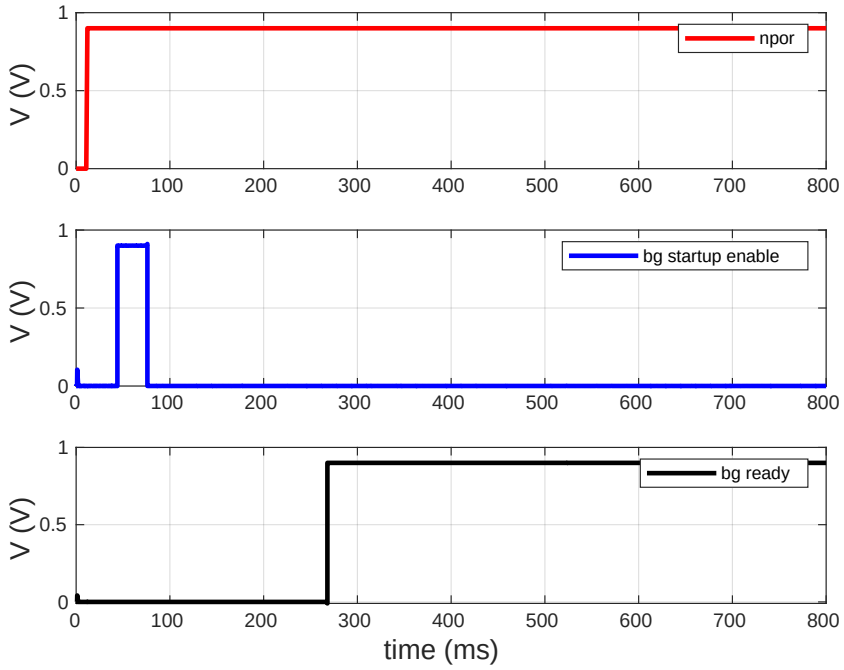


Figure 5.10: System-level signals

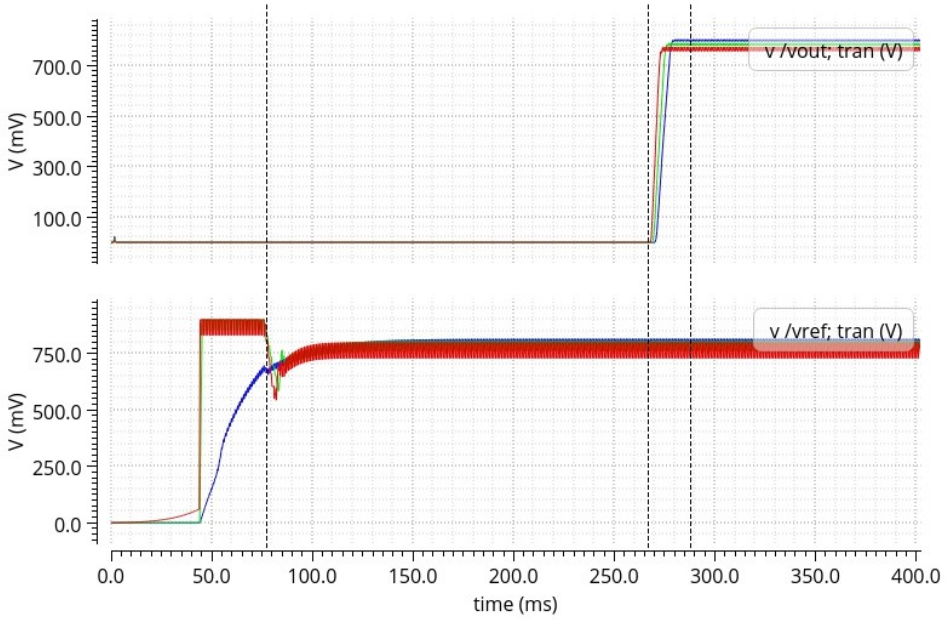


Figure 5.11: Transient Vref and Vout in FFFF125 (red), TTTT27 (green), SSSS-40 (blue)

6 Conclusion and future work

6.1 Conclusion

In this thesis, we presented the design and implementation of ultra-low-power (ULP) sub-10nW always-on blocks in GlobalFoundries 22nm (GF22nm) technology. This includes a Proportional to Absolute Temperature (PTAT) current reference, a bandgap reference, and a Low Dropout Regulator (LDO). These components were optimized to operate over the full automotive temperature range while the references maintained a current consumption of only 1nA per branch. We explored various solutions with a focus on BJTs for VREF generation and MOSFETs in subthreshold region for IREF generation and the results are expected to accelerate the design of ULP circuits in GF22nm technology.

The PTAT block achieved a line sensitivity of 2%/V at 27°C and σ/μ of 5% with a power consumption of 4nW, utilizing MOSFETs in weak inversion. This block operated with an 800mV supply voltage and occupied a silicon area of 0.001mm^2. The bandgap reference, supplied from a battery with an end-of-life (EOL) voltage of 900mV, achieved a maximum temperature coefficient (TC) of 140.6ppm/°C and an average line sensitivity of 0.56%/V at 27°C across a supply range of 900mV to 1.98V. The reference voltage spread due to process and mismatch variations was minimized to a σ/μ ratio of 2.9% by employing BJTs and with no resistor trimming. This bandgap reference consumed a total nominal power of 7.6nW and occupied a silicon area of 0.021mm^2 by using duty-cycled resistors. A performance comparison with prior ultra-low-power (ULP) references is presented in Table 6.1. This design demonstrates the lowest power consumption across the automotive temperature range. However, it exhibits a higher spread, which is a trade-off resulting from the 0.9 V supply voltage and 800 mV Vref. This bandgap voltage was used as the reference for an LDO with unity-gain feedback to prevent the multiplication of reference voltage noise. The LDO maintained an average output voltage line sensitivity of less than 1%/V and an accuracy of $\pm$2.5% with battery voltage variations from 900mV to 1.98V and load currents ranging from 100nA to 1µA. These always-on blocks can be used in IoT applications to maintain the performance of IoT sensor nodes.

	[23]	[8]	[13]	[12]	[5]	[22]	[7]	**This work**
Technology	130nm	16nm FinFet	22nm FDSOI	22nm FDSOI	65nm	130nm	22nm FDSOI	**22nm FDSOI**
Type	BJT+MOS	Diode+MOS	MOS	MOS	BJT+MOS	MOS	BJT+MOS	**BJT+MOS**
Vdd [V]	0.5 - 1.5	0.85 - 1	0.9 - 1.8	1 - 1.8	0.5	0.5 - 3	1.2 - 1.8	**0.9 - 1.98**
Vref [mV]	500	235	-	-	495	174.9	598	**800**
Iref [nA]	-	-	1.25	2.5	-	-	-	-
Power [nW]	32	40	7.8	15.5	38	0.022	45.6	**7.6**
Temp. Range [°C]	0 - 100	0 - 100	-40 - 85	-40 - 85	-40 - 120	-20 - 80	-40 - 120	**-40 -125**
TC [ppm/°C]	75	85	203	101	42	16.9 - 231	61	**60 - 140.6**
Line Sensitivity [%/V]	2	1.9	0.23	0.26	0.64	0.033	0.12	**0.56**
Area [mm^2]	0.0264	0.0022	0.0132	0.0025	0.0532	0.0013	0.0104	**0.021**
Spread (σ/μ) [%)]	0.67	0.82	6.66	4.7	1.03	0.85	0.37	**2.9**

Table 6.1: Comparison of ULP reference designs

6.2 Future Work

There are several areas for future research and development of the designed references.

6.2.1 Lowering the area of bandgap reference

The bandgap reference design consumes a total area of 0.021 mm^2. Notably, 67% of the total area is occupied by capacitors, which are primarily used in parallel with duty-cycled resistors to smooth the voltage ripples. At the same time, after adding these parallel capacitors, larger compensation capacitors were required to stabilize the loop leading to even more area consumption. Hence, to reduce the area we should reduce the value of the capacitors. As discussed in Subsection 3.1.3 during the PTAT reference design, by using two branches with two-phase non-overlapping clocks instead of one branch, the ripple amplitude is halved while its frequency is doubled. Hence, half the switched capacitor is required for the same equivalent resistor, resulting in half the damping capacitor for the same ripple amplitude. One can continue this trend by making more phases of the clock until the switched capacitor value is not too small to be affected by parasitic capacitors of the switches. This will lead to N times reduction in the capacitor area with N phase clocks. We also showed in Section 4.1.8 the feasibility of having as large as 50 different phases with a sub-10 nW ring oscillator.

The remaining challenge of this part is to keep the linearity of the switches connected to the VBE of BJTs and minimize their ON resistance variations over the temperature range as well as their leakage currents. Transmission gates are simulated to help the conductance of the switches by 2 times, and the bulk switching technique [9] helps by 3 times. Although it costs more power and area, clock boosting seems a reasonable solution. This will reduce ON resistance without affecting OFF resistance. In [23] and [10], area-efficient ways of doubling the clock swing are introduced. However, the clock doubler circuit will add at least 4 nA more current consumption [23] to the circuit which is the same power consumption as the bandgap core. Hence, going toward lowering the area will increase power consumption.

Another potential approach to slightly reduce the area and achieve a more stable reference voltage with lower ripple is to incorporate a switched capacitor notch filter [1] [17]. By sampling the ripple at the clock frequency (f_{clk}), this method can effectively attenuate ripples in the bandgap reference voltage.

6.2.2 Lowering the spread of the bandgap reference voltage

As shown in Section 4.2.6, the spread of the bandgap reference voltage is $\sigma/\mu = 2.9\%$. This spread is mainly due to variations in the PMOS current mirrors, which contribute 2.3% to the overall spread. The low inversion coefficient of 0.2 used in the PMOS mirrors, due to limited headroom, results in a significant spread in the output current.

The primary goal of this work was to explore the feasibility of extreme specifications and push the limits of power and area efficiency without trimming the resistors. However, to reduce the voltage spread, trimming could be an effective solution.

Voltage trimming could be used to fine-tune the reference voltage after fabrication, helping to reduce the spread caused by component variations. This could be done with a 2-bit trimming system that adjusts the reference resistance (R_{ref}) by small steps, such as 300 kΩ. In

addition, resistor trimming could improve the temperature coefficient of the bandgap reference. By adjusting R_{PTAT} with small 10 kΩ steps, it is possible to fine-tune the temperature coefficient.

In future work, implementing these trimming methods could significantly reduce the spread in the reference voltage, while maintaining the ultra-low power and covering a slightly larger die area than this design.

Another potential improvement to provide more headroom for the PMOS mirrors is to replace the NPNs with NMOS transistors operating in weak inversion, as discussed in Chapter 3. By doing so, the IC of the PMOS mirror can be increased, resulting in a reduced spread in the output current. However, it is necessary to minimize the contribution of the NMOS transistors to the current spread. This can be achieved by increasing their area and optimizing the ratio between their sizes.

Bibliography

[1] Xiaomeng An, Sining Pan, Hui Jiang, and Kofi A. A. Makinwa. A 0.01 mm2 10mhz rc frequency reference with a 1-point on-chip-trimmed inaccuracy of $\pm$**0.28%** from $-$**45°C** to **125°C** in 0.18m cmos. In *2023 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 60–62, 2023.

[2] R. Jacob Baker. *CMOS: Circuit Design, Layout, and Simulation*, chapter 23: Voltage References, pages 745–772. John Wiley & Sons, Ltd, 2010.

[3] Hironori Banba, Hitoshi Shiga, Akira Umezawa, Takeshi Miyaba, Toru Tanzawa, Shigeru Atsumi, and Koji Sakui. A cmos bandgap reference circuit with sub-1-v operation. *IEEE Journal of Solid-State Circuits*, 34(5):670–674, 1999.

[4] Francesco Centurelli, Alessandro Fava, Giuseppe Scotti, and Alessandro Trifiletti. A detailed model of the switched-resistor technique. *IEEE Open Journal of Circuits and Systems*, 2:497–507, 2021.

[5] U Chi-Wa, Wen-Liang Zeng, Man-Kay Law, Chi-Seng Lam, and Rui Paulo Martins. A 0.5-v supply, 36 nw bandgap reference with 42 ppm/° c average temperature coefficient within- 40° c to 120° c. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 67(11):3656–3669, 2020.

[6] Ana Isabela Araujo Cunha, Márcio Cherem Schneider, and Carlos Galup-Montoro. An mos transistor model for analog circuit design. *IEEE Journal of solid-state circuits*, 33(10):1510–1519, 1998.

[7] Adilet Dossanov, Christian Ziegler, and Vadim Issakov. Ultra-low-power high psrr sub-1 v voltage reference circuit in 22 nm fdsoi cmos. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2024.

[8] Matthias Eberlein, Georgios Panagopoulos, and Harald Pretl. A 40nw, sub-iv truly 'digital'reverse bandgap reference using bulk-diodes in 16nm finfet. In *2018 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, pages 99–102. IEEE, 2018.

[9] Acacio Galhardo, João Goes, and Nuno Paulino. Novel linearization technique for low-distortion high-swing cmos switches with improved reliability. In *2006 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 4–pp. IEEE, 2006.

[10] Mingyang Gu, Yunsong Tao, Xiyu He, Yi Zhong, Lu Jie, and Nan Sun. A 3.7 mw 11b 1gs/s time-interleaved sar adc with robust one-stage correlation-based background timing-skew calibration. In *ESSCIRC 2023-IEEE 49th European Solid State Circuits Conference (ESSCIRC)*, pages 145–148. IEEE, 2023.

[11] Ge Guang, Zhang Cheng, G Hoogzaad, et al. A single-trim cmos bandgap reference with a 3 inaccuracy of$\pm$015% from 40c to 125c. *JSSCC*, 2011.

[12] Martin Lefebvre and David Bol. A 2.5-na area-efficient temperature-independent 176-/82-ppm/ ° c cmos-only current reference in 0.11- μ m bulk and 22-nm fd-soi. *IEEE Journal of Solid-State Circuits*, pages 1–15, 2024.

[13] Martin Lefebvre, Denis Flandre, and David Bol. A 1.1-/0.9-na temperature-independent 213-/565-ppm/° c self-biased cmos-only current reference in 65-nm bulk and 22-nm fd-soi. *IEEE Journal of Solid-State Circuits*, 58(8):2239–2251, 2023.

[14] Fahad Masood. How to choose the best battery for your healthcare medical device design. *Analog Devices*, December 2023.

[15] Harry Neuteboom, Ben MJ Kup, and Mark Janssens. A dsp-based hearing instrument ic. *IEEE Journal of Solid-State Circuits*, 32(11):1790–1806, 1997.

[16] Yuji Osaki, Tetsuya Hirose, Nobutaka Kuroki, and Masahiro Numa. 1.2-v supply, 100-nw, 1.09-v bandgap and 0.7-v supply, 52.5-nw, 0.55-v subbandgap reference circuits for nanowatt cmos lsis. *IEEE Journal of Solid-State Circuits*, 48(6):1530–1538, 2013.

[17] Pyoungwon Park, David Ruffieux, and Kofi A. A. Makinwa. A thermistor-based temperature sensor for a real-time clock with $\pm$ 2 ppm frequency stability. *IEEE Journal of Solid-State Circuits*, 50(7):1571–1580, 2015.

[18] Micheal AP Pertijs and Johan Huijsing. *Precision temperature sensors in CMOS technology*. Springer Science & Business Media, 2006.

[19] B. Razavi. *Design of Analog CMOS Integrated Circuits*. McGraw-Hill Higher Education, 2016.

[20] S Rekha et al. A 0.5 v, 1 na switched capacitor ptat current reference circuit. In *2019 IEEE Conference on Modeling of Systems Circuits and Devices (MOS-AK India)*, pages 51–54. IEEE, 2019.

[21] Vishal Saxena and R Jacob Baker. Indirect compensation techniques for three-stage fully-differential op-amps. In *2010 53rd IEEE International Midwest Symposium on Circuits and Systems*, pages 588–591. IEEE, 2010.

[22] Mingoo Seok, Gyouho Kim, David Blaauw, and Dennis Sylvester. A portable 2-transistor picowatt temperature-compensated voltage reference operating at 0.5 v. *IEEE Journal of Solid-State Circuits*, 47(10):2534–2545, 2012.

[23] Aatmesh Shrivastava, Kyle Craig, Nathan E. Roberts, David D. Wentzloff, and Benton H. Calhoun. 5.4 a 32nw bandgap reference voltage operational from 0.5v supply for ultra-low power systems. In *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, pages 1–3, 2015.

[24] Eric Vittoz and Jean Fellrath. Cmos analog integrated circuits based on weak inversion operations. *IEEE journal of solid-state circuits*, 12(3):224–231, 1977.

[25] Hui Wang and Patrick P Mercier. A 14.5 pw, 31 ppm/° c resistor-less 5 pa current reference employing a self-regulated push-pull voltage reference generator. In *2016 IEEE international symposium on circuits and systems (ISCAS)*, pages 1290–1293. IEEE, 2016.

[26] Yuanmao Ye and Ka Wai Eric Cheng. Analysis and optimization of switched capacitor power conversion circuits with parasitic resistances and inductances. *IEEE Transactions on Power Electronics*, 32(3):2018–2028, 2017.