Review of Sub-1 V CMOS Bandgap References

Chris Biancone, *Member, IEEE*

***Abstract*—The continued push for lower supply voltages to VLSI circuits for reduced power consumption has led to a necessary adaptation for precision reference circuits. A bandgap reference circuit proposed by Banba et al. in 1998 is reviewed from its subsequent publication in 1999. This design adopts the current-mode method of generating a reference voltage, and by reconfiguring elements of a conventional design at the time, is adapted to operation under 1 V supply. This design has served as a foundation for countless low voltage bandgap designs since its publication, but it is found that the original hardware verification of this circuit is severely limited by the flash memory process used for its construction. Some avenues for reviewing and improving its performance are explored, backed by recent developments in this field.**

***Index Terms*—Bandgap, CMOS, voltage reference.**

I. Introduction

PRECISION voltage references are an essential component of solid-state circuits. Reference circuits are often used for appropriately and accurately biasing various circuit components, functioning as comparative standards for signal conditioning, and serving as a solid baseline for power regulation, voltage-controlled oscillation, and more. The bandgap reference, with its name deriving from its performance dependence on the material properties of its constituent devices, has served as the basis for many precision reference designs due to its extremely low output temperature coefficient, low internal resistance, and great long-term stability. The importance of a stable reference has been recognized since well before the inception of solid-state electronics design. Bell Labs' construction in 1950 of a diode with avalanche breakdown first predicted by Clarence Zener in 1934 afforded the miniaturization of the functionality provided by the Weston standard voltage cell. [1], [2]. This led to quick commercialization as solid state circuits became more widely available in the 1960s [3]; however, the lifetimes were typically not as long, and the breakdopwn voltage of single Zener diodes typically has non-negligible temperature dependence [4]. Significant advancement arose from the combination of highly-doped pn junction devices with inverse temperature dependence on current density, yielding voltage sources dependent almost exclusively on the semiconductor material [5].

Modern CMOS processes have been facing an overall push toward miniaturization, where lower threshold voltages enable reduce supply voltages and power consumption [6]. This has posed challenges for creating bandgap circuits as supply voltages have encroached upon and fallen below the bandgap of silicon at 1.12 V [7] while maintaining CMOS process compatibility. Various methods are exploited to achieve this, including subthreshold operation, digital calibration, additional curvature compensation, and others. The solution proposed by Banba et al., reviewed here in detail, employs the principle of a current-mode bandgap reference, where a steady current is used to generate a stable voltage across a device with internal resistance [8]. Material properties are leveraged to generate a current that is independent of supply voltage and temperature variation.

C. Biancone is with the Department of Electrical and Microelectronic Engineering, Rochester Institute of Technology, Rochester, NY, 14623 USA e-mail: cjb1402@rit.edu.

Manuscript received Mar. 05, 2024

A. Basic Bandgap Theory

For an ideal pn junction, it is known that the current follows an exponential curve related to the forward voltage, V_f, which is linearly dependent on the temperature:

$$V_T = \frac{kT}{q} \approx 8.617 \times 10^{-5}\,\text{V/K} \tag{1}$$

$$I = I_s\left[\exp\left(\frac{V_f}{V_T}\right) - 1\right] \tag{2}$$

The saturation current, I_s is shown by approximating the temperature dependence of intrinsic carrier concentration:

$$\begin{aligned} {n_i}^2 &= C_0 T^3 \exp\left(\frac{-E_G}{kT}\right) \\ I_s &\simeq C_{p^+n,n^+p} T^{\beta_{p,n}} \exp\left(\frac{-E_G}{kT}\right) \end{aligned} \tag{3}$$

where $\beta_{p,n}$ subsumes approximate temperature dependence of the diffusion coefficients and lengths within the junction, and C is an arbitrary constant of integration [9].

While at normal temperatures, standard junctions do not behave in the way described by Eq. (3) due to surface effects and depletion region generation/recombination [10], this model does accurately describe the V_{BE} of a BJT transistor as a function of I_C [11]. Expressing V_{BE} as a Taylor series expansion shows that as doping concentration increases, the temperature dependence decreases, similarly to increasing current density:

$$\begin{aligned} V_{BE} = \frac{kT_0}{q}\Bigg\{ &\ln\left(\frac{I_C}{I_s T_0}\right) + \left[\ln\left(\frac{I_C}{I_s T_0}\right) - \left(\beta + \frac{E_{G_0}}{kT_0}\right)\right]\left(\frac{T}{T_0} - 1\right) \\ &- \frac{\beta}{2}\left(\frac{T}{T_0} - 1\right)^2 + \cdots + \frac{\beta(-1)^{(n-1)}}{n(n-1)}\left(\frac{T}{T_0} - 1\right)^n + \cdots \Bigg\} \end{aligned} \tag{4}$$

$$V_{BE} > 4V_T, \qquad T < 2T_0$$

where E_{G_0} is the bandgap at 0 K. Compensation of the temperature relationship shown in Eq. (4) is performed by selecting an effective Θ such that:

$$V_{ref} = E_{G_0} + \Phi\frac{kT_0}{q}\left[1 - \frac{1}{2}\left(\frac{T}{T_0} - 1\right)^2\right] \tag{5}$$

$$\Theta = \Phi + \frac{E_{G_0}}{kT_0} \tag{6}$$

Since E_{G_0} is much greater than ΦV_{T_0}, V_{ref} is primarily a function of the semiconductor bandgap [5]. In a standard bandgap core, the temperature coefficient of V_{BE} in Eq. (4) of around $-2\,\text{mV/°C}$ becomes cancelled out by the ΔV_{BE} resulting from running the transistors at different current densities, around $2\,\text{mV/°C}$ [12], [13].

$$\Delta V_{BE} = V_T \ln\left(\frac{I_{C_2}}{I_{C_1}}\right), \qquad I_C = I_{C_0}\left(\frac{T}{T_0}\right)$$

$$\Delta V_{BE} = V_T \ln\left(\frac{T_0}{T}\right) \tag{7}$$

This principle has been extended to some widely-adopted bandgap designs, such as those proposed by Widlar, Kuijk, Brokaw, and many others, using a ratio of current densities with opposite temperature coefficients, with some additional series regulation through error amplification [14]–[16].

A downside to using a BJT-based bandgap design, though it provides superior temperature characteristics to a basic diode-based design, is that many standard CMOS processes do not provide means for the creation of true bipolar devices, limiting their application space. Generic process design kits (PDKs) can provide bipolar devices in the form of a vertical substrate BJT. Often, these devices have limited intrinsic current gain, and may be profitably employed in quasi-diode form for the creation of CMOS-compatible bandgap references. The additional temperature coefficients yielded by lighter doping require careful consideration for the design of the reference circuit.

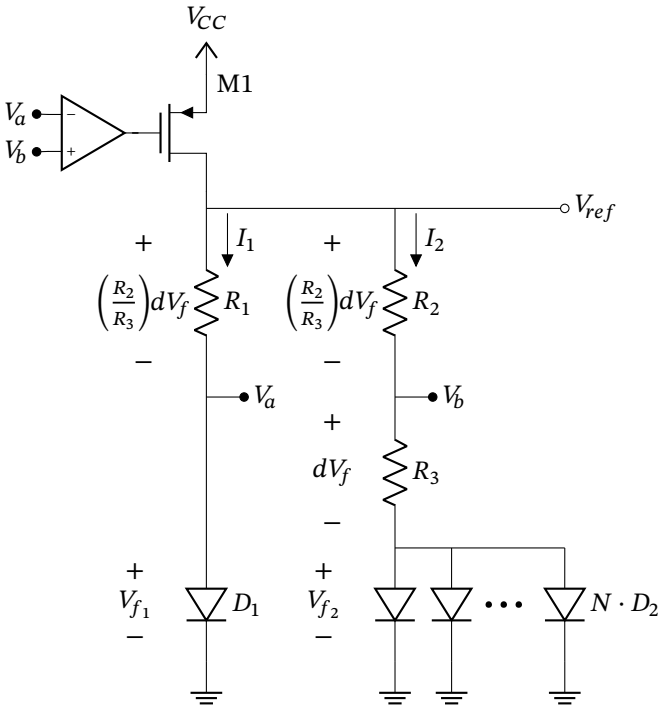

Fig. 1. Conventional bandgap reference circuit schematic used for comparison. Redrawn from [8].

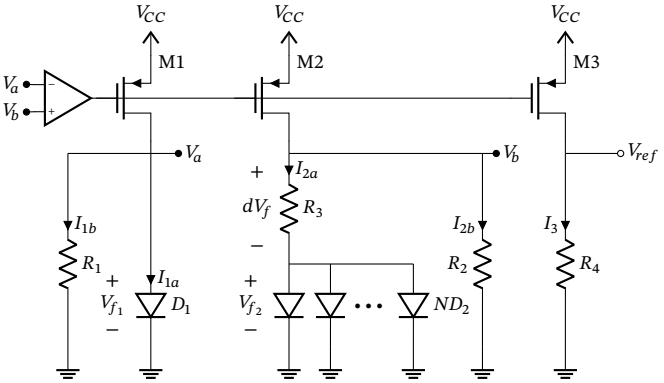

Fig. 2. Proposed bandgap reference circuit schematic. Redrawn from [8].

II. Banba et al. Design

A. Architecture

The design in this paper, originally presented at the 1998 Symposium on VLSI Circuits, is proposed as a means of operating under the constraints of both CMOS compatibility and reduced supply voltage [17]. It is based on a conventional bandgap circuit, shown in Fig. 1 which generates two currents, one proportional to dV_f, the difference in forward voltage between a diode and N other parallel diodes, and another proportional to V_T, the thermal voltage described by Eq. (1). Resistive voltage dividers are placed in parallel with these diodes, and an op-amp in single feedback configuration measures the voltages generated by each current. The output voltage from this reference circuit is defined by:

$$V_{ref} = V_{f_1} + \frac{R_2}{R_3}V_T \ln\left(\frac{R_2}{R_1}N\right) \tag{8}$$

which typically has a lower limit of 1.25 V [18].

The presented design remedies this by placing R_1 and R_2 in parallel with the diodes, and by adding additional current sources, shown in Fig. 2. The gates of the sources are all controlled by the op-amp monitoring the differential voltage between nodes a and b, so that the desired equilibrium state is where I_1, I_2, and I_3 are all equal. The output voltage in this case is:

$$V_{ref} = \frac{R_4}{R_2}V_{f_1} + \frac{R_4}{R_3}V_T \ln(N) \tag{9}$$

Assuming that the resistor and diode parameters are the same in both conventional and proposed designs, the output voltage is reduced by a factor of R_2/R_4. With $M1$, $M2$, and $M3$ in saturation, VCC can be theoretically lowered to V_f as long as V_{ref} is below V_f.

B. Results

To verify this design, it was first simulated against the conventional design using low-V_{th} transistors, having $V_{th_p} = -0.3\,\text{V}$ and $V_{th_n} = 0.4\,\text{V}$, performing amicably against a conventional bandgap circuit. It was then fabricated in a $0.4\,\mu\text{m}$ n-well flash memory process with single polysilicon, single silicide, and 2 metal layers. This process has $V_{th_p} = -1.0\,\text{V}$ and $V_{th_n} = 0.7\,\text{V}$, so the input transistors to the op-amp were native NMOS with $V_{th_{n,nat}} = -0.2\,\text{V}$. An additional control signal is added for startup of the bandgap on chip along with stabilizing capacitors. Although the proposed circuit's experimental results show an achieved $V_{ref} = 518\,\text{mV}$ with 3σ variation of $15.2\,\text{mV}$, the high threshold voltages for the chosen fabrication process limited the operation of the op-amp to $2.1\,\text{V}$ V_{CC}, even with the native input devices.

III. Discussion

A. Merits

The bandgap circuit topology proposed by Banba et al. has some undoubted benefits over the conventional design it is compared to in their paper. Much of the elegance of this design is in its simplicity; simply rearranging the resistor configuration and leveraging the properties of MOSFET current mirroring allows for operation under considerably less supply voltage than before. The fabricated design shows a $\pm 1\,\text{mV}$ variation when varying V_{CC} across the tested range above the turn-on voltage, and a $\pm 3\,\text{mV}$ variation from $27\,^\circ\text{C}$ to $\pm 125\,^\circ\text{C}$. This performance is good and is on par with what may be expected from a bandgap of this type of architecture. The proposed design should allow for more advancements in the realm of CMOS technologies with lower and lower voltage supplies, which there is an increasing push toward.

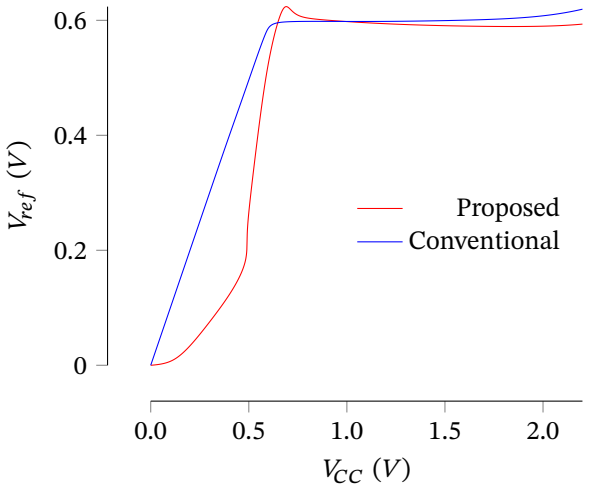

Fig. 3. Results from simulating the two circuits in [8] in the Cādence gpdk045_v5.0 with 1 V low-threshold devices – $V_{th_n} = 0.28\,\text{V}$, $V_{th_p} = -0.34\,\text{V}$. Operation below 1 V V_{CC} is easily demonstrated.

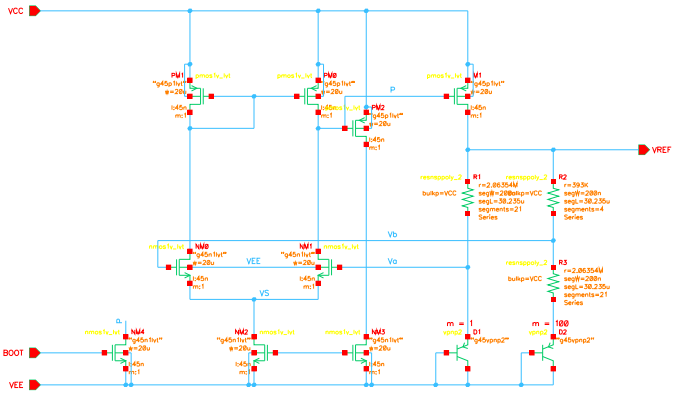

Fig. 4. Schematic of conventional bandgap in Cādence Virtuoso.

B. Downsides

While the circuit proposed by Banba et al. represents a significant step forward, the methods used to verify the performance in simulation and hardware leave a lot to be desired in solidifying their argument for this circuit's benefits. Firstly, the simulation results presented for low threshold devices seem to plot a truly simulation version of the proposed circuit against a theoretical model for the conventional circuit, evidenced by the perfect shape of the plot. This is fine for illustrating the general point that the new design has a much lower turn-on voltage, but the disparity between the curves is notable.

The choice of process used to verify the circuit operation in hardware leaves a lot to be desired. The gigantic disparity in the threshold devices of the simulated process and fabricated one prevented the circuit from functioning anywhere close to its stated performance in practice, with a lower V_{CC} limit of 2.1 V – worse than a conventional bandgap in a traditional process. The results from the hardware evaluation actually start at a $V_{CC} = 1\,\text{V}$, while the paper is titled for operation sub-1 V. Simulating both circuits in a process with low-threshold devices in a 45 nm process yields results that much better reflect the stated performance of the device, as shown in Fig. 3. However, similar performance was also seen with the conventional circuit at low voltages, with the proposed circuit holding steady at higher supply voltages. The schematics constructed for the conventional and proposed designs are shown in Fig. 4 and Fig. 5, respectively.

The authors attempted to circumvent the threshold voltage issue by using native NMOS devices on the input to the op-amp, but this was not reflected in the schematic shown in their paper, nor did it appear to assist in bringing down the V_{CC} requirement. The authors also did not report on the PSRR of this circuit in comparison with the conventional design, which is an important metric to consider when designing a reference voltage that should be free from dependence on power supply noise. In reporting the distributions from the fabricated conventional and proposed circuits, the conventional design data was taken across 34 points, whereas the proposed design was reported using only 23 points.

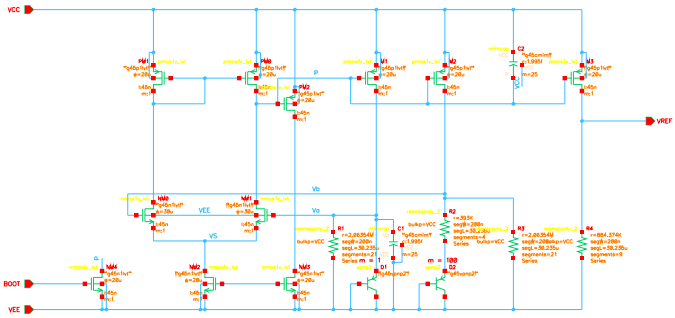

Fig. 5. Schematic of proposed bandgap in Cādence Virtuoso.

While threshold parameters for the process used were given, it may have also been useful to indicate certain parameters for the vertical BJTs used as quasi-diodes for the process used, since typical doping profiles and junction areas will have a fundamental impact on the bandgap performance. Likewise, since the resistors used in this design are fabricated from the n-type diffusion layer, they will have a temperature coefficient that is interdependent with the bipolar and MOS devices of the process, which is typically a known value, and could have been stated in some way.

Something of note is that the original presentation of the V_{ref} equations in the paper leaves something to be desired for allowing the reader to easily draw connections between the designs compared in the paper. Care has been taken to rewrite them here in such a manner.

IV. Recommendations

Based on the stated critiques, a first course of action to truly verify this design's performance could be to construct it in a process that is designed with low voltage operation in mind, where threshold voltages closer to those used in their simulated version would allow truly sub-1 V V_{CC} operation. The same number of data points should be used in reporting distribution data from the new and conventional designs, or reasoning should be given for why there is a difference in number.

The reliance on a ratio of resistance in the proposed design does not lend well to the process variation of resistors created in standard CMOS processes. The values used in circuit are large and would not be easily created in a polysilicon layer with reduced temperature coefficient from the n-implant layer. Possible options may be to use a process that can implement low-tempco thin film resistors, use trimmed resistor networks, and even integrate these with trimmed BJT networks to further correct the temperature curve [16], [19], [20]. Another option is to forego the use of resistors entirely and utilize subthreshold MOS devices as resistive elements to leverage their temperature coefficient. This has an additional advantage of reduced coupled noise while retaining CMOS compatibility [21]. Switched capacitor designs, while offering better ratio matching in CMOS processes than resistive networks, must be carefully implemented to avoid noise injection into the reference circuit.

Recently, the idea of using native CMOS devices with low threshold has been adapted to a fully-CMOS based design, with switched-capacitor offset cancellation and additional noise suppression, achieving a lower V_{CC} limit of 0.5 V [22]. This design is robust with a temperature coefficient of 45 ppm/° C over 10-50 ° C shown after construction in a 180 nm process. In the future, precision trimming applied to this circuit may further increase its precision.

V. Conclusion

The bandgap reference design proposed by Banba et al. in 1998 has served as the basis for countless modifications of the circuit in an attempt to adapt it to lower supply voltages. Its impact on the industry is irrefutable with over 800 collective citations for the two original publications, which will likely continue to grow as the push to squeeze as much power savings as possible out of CMOS processes increases. The original publications leave something to be desired from the presented validation of the circuit design, by constructing the circuit in a process that was not only not designed for low voltage operation and did not reflect the simulated process parameters, but completely prevented the design from being validated under the conditions for which it was designed. The robustness of the Toshiba engineers' approach has been well-proven in literature since their publication, but it is conceivable that the impact of this design may have been further increased by presenting hardware data that supported its goals.

Acknowledgment

The author would like to thank Dr. Mark Pude for providing the paper to be analyzed in his course ***Mixed-Signal IC Design*** at Rochester Institute of Technology.

References

[1] E. Weston, "Voltaic cell," US Patent US494 827A, 1893. [Online]. Available: https://patents.google.com/patent/US494827

[2] C. Zener and H. H. Wills, "A theory of the electrical breakdown of solid dielectrics," ***Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character***, vol. 145, pp. 523–529, 7 1934. [Online]. Available: https://doi.org/10.1098/rspa.1934.0116

[3] M. Ltd., L. G. Cripps, L. B. Johnson, L. H. Light, E. Wolfendale, E. G. Cooper, M. J. Gray, K. Holford, L. Jannson, H. Kemhadjian, A. F. Newell, J. R. Nowicki, J. F. Pawling, A. J. Rees, W. L. Stephenson, P. Tharma, T. B. Watkins, and P. G. Williams, ***Reference Manual of Transistor Circuits***, 1st ed. Mullard Overseas, Ltd., 1960, vol. 1.

[4] R. P. Baker and J. Nagy, "An investigation of long-term stability of zener voltage references," ***IRE Transactions on Instrumentation***, vol. I-9, pp. 226–231, 9 1960. [Online]. Available: https://doi.org/10.1109/IRE-I.1960.5006922

[5] D. Hilbiber, "A new semiconductor voltage standard." IEEE, 1964, pp. 32–33. [Online]. Available: https://doi.org/10.1109/ISSCC.1964.1157541

[6] R. Gonzalez, B. Gordon, and M. Horowitz, "Supply and threshold voltage scaling for low power cmos," ***IEEE Journal of Solid-State Circuits***, vol. 32, pp. 1210–1216, 8 1997. [Online]. Available: https://doi.org/10.1109/4.604077

[7] C. C. Hu, ***Modern Semiconductor Devices for Integrated Circuits***, 1st ed. Pearson, 3 2009. ISBN 978-0136085256

[8] H. Banba, H. Shiga, A. Umezawa, T. Miyaba, T. Tanzawa, S. Atsumi, and K. Sakui, "A cmos bandgap reference circuit with sub-1-v operation," ***IEEE Journal of Solid-State Circuits***, vol. 34, pp. 670–674, 5 1999. [Online]. Available: https://doi.org/10.1109/4.760378

[9] J. Brugler, “Silicon transistor biasing for linear collector current temperature dependence,” ***IEEE Journal of Solid-State Circuits***, vol. 2, pp. 57–58, 6 1967. [Online]. Available: https://doi.org/10.1109/JSSC.1967.1049790

[10] C. tang Sah, R. Noyce, and W. Shockley, “Carrier generation and recombination in p-n junctions and p-n junction characteristics,” ***Proceedings of the IRE***, vol. 45, pp. 1228–1243, 9 1957. [Online]. Available: https://doi.org/10.1109/JRPROC.1957.278528

[11] C.-T. Sah, “Effect of surface recombination and channel on p-n junction and transistor characteristics,” ***IRE Transactions on Electron Devices***, vol. 9, pp. 94–108, 1 1962. [Online]. Available: https://doi.org/10.1109/T-ED.1962.14895

[12] R. Widlar, “An exact expression for the thermal variation of the emitter base voltage of bi-polar transistors,” ***Proceedings of the IEEE***, vol. 55, pp. 96–97, 1 1967. [Online]. Available: https://doi.org/10.1109/PROC.1967.5396

[13] R. Pease, “The design of band-gap reference circuits: trials and tribulations.” IEEE, 1990, pp. 214–218. [Online]. Available: https://doi.org/10.1109/BIPOL.1990.171166

[14] R. Widlar, “New developments in ic voltage regulators,” ***IEEE Journal of Solid-State Circuits***, vol. 6, pp. 2–7, 2 1971. [Online]. Available: https://doi.org/10.1109/JSSC.1971.1050151

[15] K. Kuijk, “A precision reference voltage source,” ***IEEE Journal of Solid-State Circuits***, vol. 8, pp. 222–226, 6 1973. [Online]. Available: https://doi.org/10.1109/JSSC.1973.1050378

[16] A. Brokaw, “A simple three-terminal ic bandgap reference,” ***IEEE Journal of Solid-State Circuits***, vol. 9, pp. 388–393, 12 1974. [Online]. Available: https://doi.org/10.1109/JSSC.1974.1050532

[17] H. Banba, H. Shiga, A. Umezawa, T. Miyaba, T. Tanzawa, S. Atsumi, and K. Sakui, “A cmos band-gap reference circuit with sub 1 v operation.” IEEE, 1998. ISBN 0-7803-4766-8 pp. 228–229. [Online]. Available: https://doi.org/10.1109/VLSIC.1998.688094

[18] B. Razavi, “The bandgap reference [a circuit for all seasons],” ***IEEE Solid-State Circuits Magazine***, vol. 8, pp. 9–12, 6 2016. [Online]. Available: https://doi.org/10.1109/MSSC.2016.2577978

[19] P. Malcovati, F. Maloberti, C. Fiocchi, and M. Pruzzi, “Curvature-compensated bicmos bandgap with 1-v supply voltage,” ***IEEE Journal of Solid-State Circuits***, vol. 36, pp. 1076–1081, 7 2001. [Online]. Available: https://doi.org/10.1109/4.933463

[20] P. Mok and K. N. Leung, “Design considerations of recent advanced low-voltage low-temperature-coefficient cmos bandgap voltage reference.” IEEE, 2004. ISBN 0-7803-8495-4 pp. 635–642. [Online]. Available: https://doi.org/10.1109/CICC.2004.1358907

[21] A. Buck, C. McDonald, S. Lewis, and T. Viswanathan, “A cmos bandgap reference without resistors,” ***IEEE Journal of Solid-State Circuits***, vol. 37, pp. 81–83, 1 2002. [Online]. Available: https://doi.org/10.1109/4.974548

[22] A. Ria, A. Catania, P. Bruschi, and M. Piotto, “A low-power cmos bandgap voltage reference for supply voltages down to 0.5 v,” ***Electronics***, vol. 10, p. 1901, 8 2021. [Online]. Available: https://doi.org/10.3390/electronics10161901