The Best of Bob Pease

The Design of Band-Gap Reference Circuits: Trials and Tribulations

ABSTRACT

This tutorial will briefly discuss the designs of various band-gap references, with an emphasis on technical problems that caused serious troubles. Practical solutions are shown for problems, and a methodology is shown for solving problems.

TUTORIAL

The band-gap reference has been a popular analog circuit for many years. In 1971, Robert Widlar introduced the LM113, the first band-gap reference.[1] It used conventional junction-isolated bipolar-IC technology to make a stable low-voltage (1.220 V) reference. This type of reference became popular as a stable voltage reference for low-voltage circuits, such as in 5-volt data acquisition systems where zener diodes are not suitable. Band-gaps are also used in digital ICs such as ECL, to provide a local bias that is not adversely affected by ambient noises or transients.

The principle of the band-gap circuit is well known and will be mentioned here in the briefest terms. The circuit relies on two groups of transistors running at different emitter current densities. The rich transistor will typically run at 10 times the density of the lean ones, and a factor of 10 will cause a 60 millivolt delta between the base-emitter voltages of the two groups. This delta voltage is usually amplified by a factor of about 10 and added to a Vbe voltage. The total of these two voltages adds up to 1.25 volts, typically, and that is approximately the band-gap of silicon.

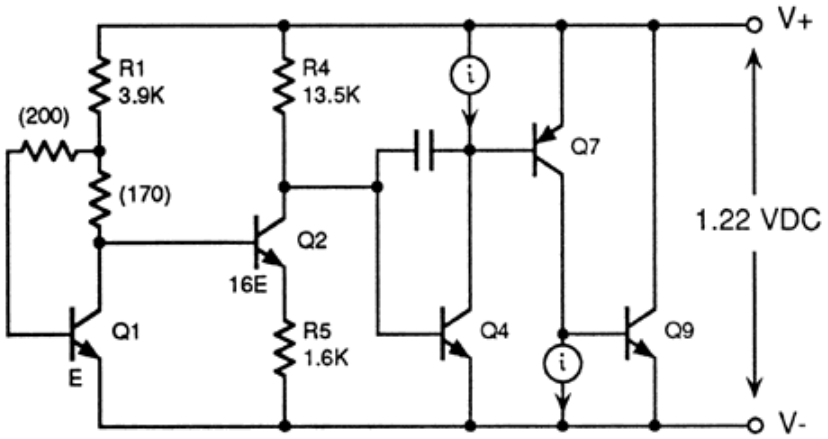


FIGURE 1
SCHEMATIC DIAGRAM, LM113 (SIMPLIFIED)

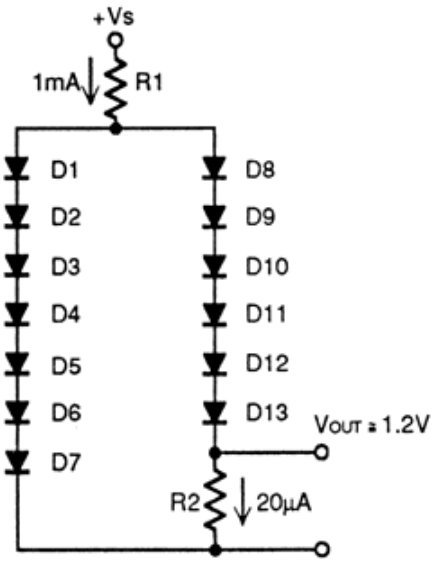
In figure 1, the LM113 schematic diagram shows a basic band-gap circuit. Q1 runs at a relatively high density, about 150 microamperes per square mil. Q2 is operated at a low density, about 10 microamperes per square mil, and so its Vbe is much less, about 70 millivolts. Now, Let's ASSUME that the circuit is at balance and the output is near 1.22 volts. Then the 70 millivolts across R5 is magnified by the ratio of R4 to R5, about 8:1, up to a level of 600 millivolts. This voltage is added to the Vbe of Q4 (about 620 millivolts at room temperature) to make a total of 1.22 volts, as required. Q4 then amplifies the error signal through Q7 and Q9, which provide enough gain to hold the V+ bus at 1.22 volts. The beauty of the band-gap reference is the summation of the Vbe term, which decreases at the rate of about -2 millivolts /°C, and the (delta-Vbe term) which grows at about + 2 millivolts /°C, to achieve an overall Temperature Coefficient (Tempco) that is substantially zero. All band-gaps employ this summation of a growing and a shrinking voltage, to make a

stable low-tempco voltage. Further, it has been shown[2] that when a circuit has been trimmed to the correct voltage, the correct tempco will follow, despite process variations in parameters such as Vbe, beta, sheet resistivity, etc. Consequently, band-gap circuits are often trimmed to their ideal voltage so as to provide also a low tempco.

There are many other circuits that have been used for band-gap references, and each one has its own set of advantages and disadvantages. Figure 2 shows a simple brute-force scheme: the stack of Dl-D7 run at a rich current, and D8-Dl3 at a lean current, and the resultant output of 1.2 v has a low tempco. Unfortunately, this circuit is not very suitable for operation on low supply voltages, but its concept is clear!

The Brokaw cell (Figure 3), attributed to Paul Brokaw,[3] is often used for output voltages larger than 1.2 volts, as its Vout can be scaled by the ratio of two resistors, R3 and R4. A similar circuit is used in the LM117, an adjustable medium-power voltage regulator, which can be trimmed by two external resistors to any voltage in the range 1.25 to 57 volts. (Figure 4)

FIGURE 2
SCHEMATIC DIAGRAM
of SIMPLE BANDGAP



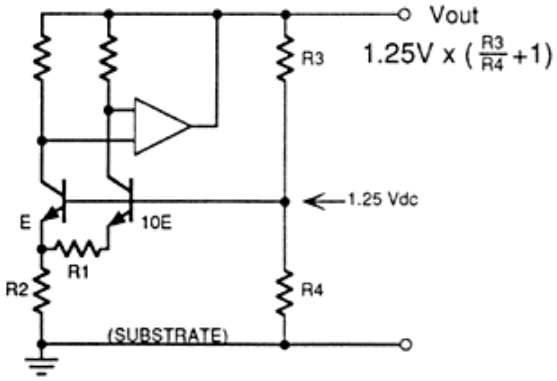


FIGURE 3
SCHEMATIC DIAGRAM, AD580 (SIMPLIFIED)

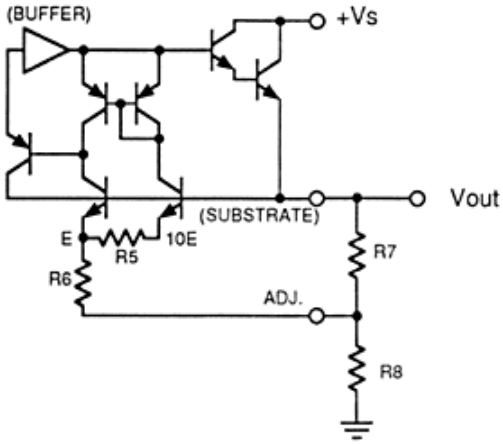


FIGURE 4
SCHEMATIC DIAGRAM, LM117 (SIMPLIFIED)

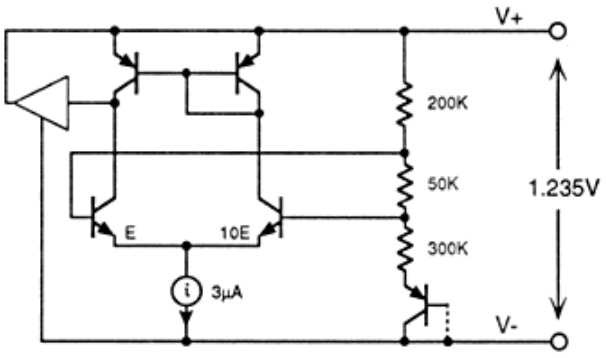


FIGURE 5
SCHEMATIC DIAGRAM, LM185 (SIMPLIFIED)

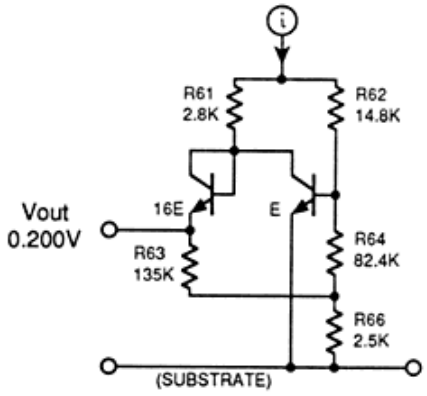


FIGURE 6
SCHEMATIC DIAGRAM, LM10 (SIMPLIFIED)

A different approach is shown in Figure 5, representing the LM136. The version shown is suitable for operation as a 1.2-volt shunt regulator. The circuit of Figure 6 is for the reference section of the LM10, which provides a reference voltage of 0.200 volts and operates on a supply voltage as low as 1.0 volts.

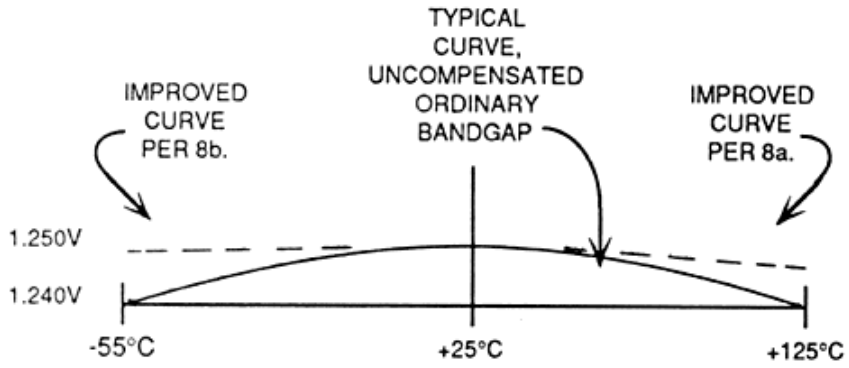


FIGURE 7
VREF VS. TEMPERATURE
BANDGAP REFERENCE

One of the major drawbacks of all these simple band-gaps is the curvature of their tempco. Ordinarily, all band-gaps have a negative tempco when hot, and a positive one when cold. See figure 7. Around room temperature, tempcos as good as 20 or 30 ppm/°C are typically attainable, but over a wide range, -1% shift is normal at hot and cold temperatures. Several techniques have been devised to neutralize this curvature. Figure 8a shows a little add-on circuit which can do a surprisingly effective improvement at warm temperatures. Sometimes 2 or more transistors are connected with slightly different biasses, for improved curvature correction. A circuit which can improve the tempco curvature at cold temperatures is shown in figure 8b.

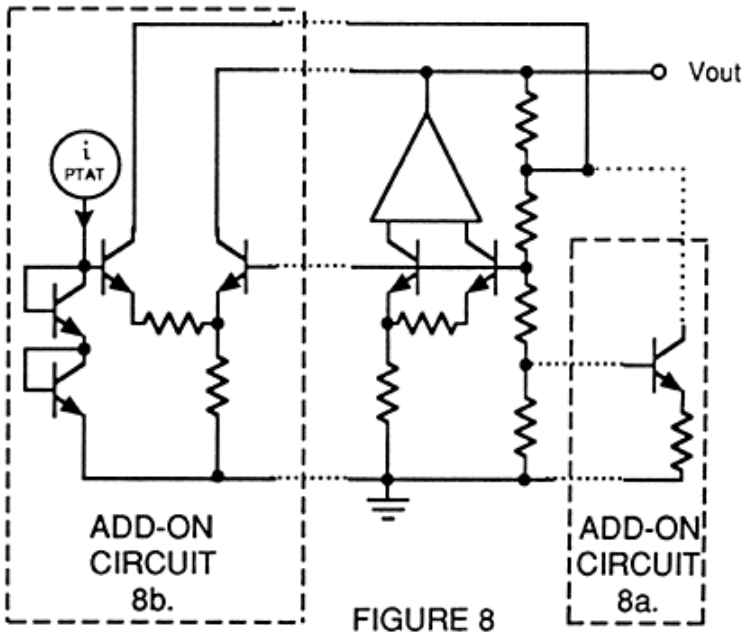


FIGURE 8
SIMPLE BANDGAP SHOWING
CURVATURE COMPENSATION

Other techniques rely on smoothly nonlinear techniques to cancel out the parabolic curvature. One example is shown in Figure 9. It is able to make an improvement of about 8:1. Other proprietary techniques and circuits are also used.

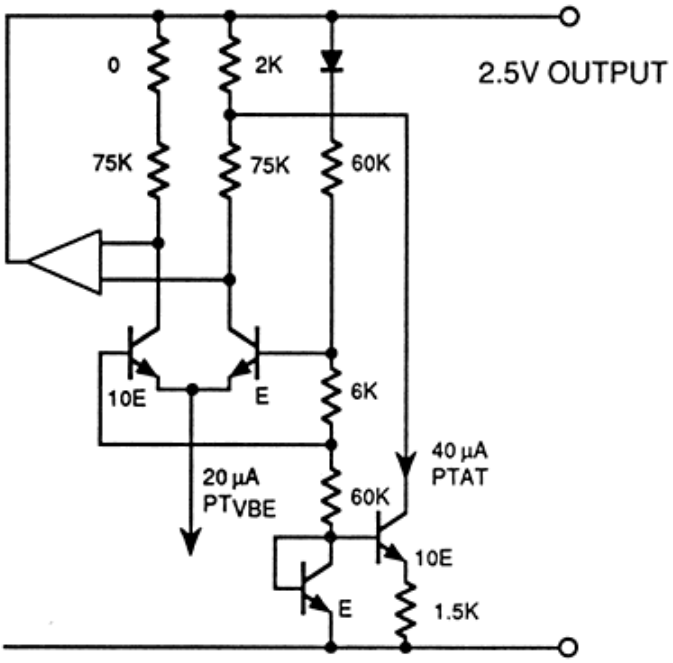


FIGURE 9
SMOOTH CURVATURE-CORRECTION CIRCUIT

Band-gap references are not normally thought of as low-noise devices, because of the high gains needed. Every 1.25 volt of output is made up of 0.6 volts of Vbe, plus a 60 mv signal magnified by a gain of 10. Consequently a 5-volt band-gap includes a 60 millivolt signal amplified by a gain of about 40. If you allow for a transistor running at 1 µA to have a theoretical amount of noise equal to 14 nv per square-root Hz, the output will have at least 800 nv per square-root Hz, and for a bandwidth of 10 kHz, about 80 µv rms, or 500 µv p-p. Obviously, a band-gap reference with 100 ppm of noise will not give good results with a 12-bit DAC. So although some band-gaps may claim advantages of low-power operation of only 11 µA total, for high-performance designs the circuits are more popular when biassed with 30 or 60 µA or more, for each group of emitters.

Many systems-on-a-chip depend on a bandgap to start up and bias all other circuits. Thus, the band-gap must be sure to start promptly under all conditions. It is not trivial to do this, as a band-gap that has not started may not have any bias currents to make sure it does start. Even the leakage of a junction capacitor may be sufficient to prevent the circuit from starting. A good solid sure-start circuit is advisable. If I can think of one, I'll present it. Figure 10, reserved.

FIGURE 10
SURE-START CIRCUIT for
BANDGAP REFERENCE

There are many other ways to design a band-gap reference that does not work well. There are two basic choices: with a computer, and without a computer.

If you rely on breadboards, you may trim to get a good tempco but the resistors may not be matched in their tempco if you just use RN55D or RN55Cs. If you actually get film resistors with a matched tempco, the tempco of your circuit may still not agree with your final circuit, due to the gross tempco of the actual diffused (or film) resistors in your circuit. Example, diffused resistors have 1600 ppm per °C; implanted silicon resistors are even steeper. Some SiChrome resistors can have +300 ppm/°C, and they will give a completely different tempco than 50 ppm/°C. Also, diffused resistors sometimes have a different tempco depending on how many volts they are biassed below their tub's voltage, so the tempco of a voltage divider may not be as good as theoretical, unless each resistor sees about the same amount of tub bias; this can be done (in theory, if you have enough space) by giving each resistor its own tub.

The dynamic stability of a band-gap IC is often inferior to its breadboard due to stray capacitances in the breadboard. Consequently the use of SPICE and similar computer simulation schemes are becoming popular, as the capacitances can be well defined without false influences or strays. However, as with any other computer usage, a "sanity-check" is recommended to insure rational results in a known situation, before the computer can be trusted in unknown cases.

However, there are many ways to get false results in SPICE and other simulation schemes. Most transistor models do not accurately model the shape of the curve of Vbe vs. temperature. It is sometimes possible to tweak the characteristics of the model of a transistor until the tempco of the breadboard or analog model matches that of the computer model. However, after you have done this, it is not safe to assume that reasonable changes in the operating points will cause reasonable changes in the tempco. The actual changes may be different from the computed changes, even for a minor tweak. For example, a minor change of a resistor, which actually gives a change of +10 ppm per °C may be predicted by the computer to give a change of +5, with no plausible reason for the discrepancy. Often, a SPICE run will simply fail to converge at cold temperatures. If you take the observed operating points from a run at -34°C and insert them as the node-set voltages for a run at -35°C, don't be surprised if you still get no convergence. In a circuit, being close to the right answer helps, but in SPICE, convergence depends on the matrix yielding a valid answer, and that has almost nothing to do with the circuit or reality. For example, on a recent SPICE run, I had a resistor connected only to ground and to a capacitor which also went to ground. If I used an * to comment-out the resistor and capacitor, the circuit refused to converge, but if I put them back into the circuit, they somehow helped the matrix converge. So, in the future we may be able to insert useless meaningless resistors around the circuit, whose sole function is to help the matrix give good convergence.

Other practical examples will be given of how to make a band-gap work badly, and how to give it a chance of working well.

Advice to the Engineer

While it is not practical or suitable to show examples of How to Design a Good Band-gap Reference, it is possible to show examples of circuits that do not work well. By avoiding the thicket of bad design practices, a good design may be seen to be feasible. A list of examples of bad design will be given here, grouped in categories.

Layout Problems

- Bad cross-coupling. As in a good op-amp, the critical transistors should be laid out with cross-coupling -- example, 1/2 of Q1 on one side of Q2 and the other half on the other side of Q2, so the centroid of Q1 and of Q2 will be at the same place. Likewise, critical resistors should be arranged so that 1/2 of R1 is on one side of R2, and the other half on the other side -- again, common centroid. This is especially important in a power regulator where large temperature gradients can be expected. Also it is very advantageous to reject gradients in sheet rho, beta, etc. In a typical band-gap, this cross-coupling should be done for the delta-Vbe transistors and also for the PNP transistors that serve as their collector loads, unless you can prove it to be unnecessary.
- Layout of Delta-Vbe circuit vs. Vbe. In some circuits, the transistors that form the delta-Vbe are not the same ones that make the Vbe. These must be laid out to reject gradients from all expected directions.
- Thermal regulation errors. When a band-gap regulator or reference has a step change of dissipation, the thermal gradients across the die may cause significant errors. A good layout with adequate cross-coupling and attention to detail is normally required to keep these errors to acceptable levels. A good power regulator can do 0.005%/W; a good precision reference can do 0.002%/W. Thoughtful layout techniques as mentioned above are especially important for a library cell that will be used in an ASIC, because when it is used, you can never guess where the thermal gradients will come from; you have to lay the circuit out to reject gradients from all directions.
- Rejection of I x R drops in power busses. If some parts of a circuit are connected to a power bus at one point, and other parts at another point, significant errors can be caused when current flows through the bus. In some cases, you can specify that no such current can flow through the bus; in cases where the bus current must be expected to fluctuate, the connections to the power bus must be arranged to reject the I x R drops.
- Thermal stress in corners of the die. When a precision circuit is laid out far off the center-line of the die, near a corner, the thermal stresses can cause errors. For best results, avoid putting precision circuits in corners or far off-axis. This applies to ASIC cells, of course.

Start-up Circuit Problems

- Bad start-up. Normally applies only to series-mode circuits, not shunt-mode.
- Start-up circuit too weak. This often happens when the start-up current (high-value resistor or EPI-FET) puts out too little current. This problem is often exacerbated by high temperatures, leaky diodes, leaky capacitors, or excessive substrate currents if one of the terminals is pulled below the substrate. You can never absolutely be free of this, but you can avoid it if you have a good start-up test. Do not expect the computer to be of much help.
- Start-up circuit too strong. This can happen when the EPI-FET puts out too much current and overwhelms the start-up circuit. It's not easy to model this in SPICE, but you can think about this as a worst-case to be avoided.
- Dynamic start-up with no dc start-up. The circuit can start on the dv/dt of the input, but may not

start if dv/dt is small. Use a start-up test.
- Start-up too slow. You may avoid this by good worst-case design and good modelling in SPICE or breadboarding.

Oscillations

- Oscillation due to capacitive load. You avoid this by good circuit design. Sometimes the condition can be helped by pre-load currents, or by a series R-C damper network to ground. Use Pease's Principle to make sure that ringing is not lurking nearby which will turn into oscillation when you turn your back. Check for ringing at all relevant temperatures.
- Oscillations at some temperatures and not others. Check for this by watching the Vout for ringing as the part's temperature is SWEPT from one extreme to the other. Note, monitoring the dc output voltage is not necessarily sufficient to insure freedom from oscillation.
- Oscillations due to improper start-up circuit. Make sure the start-up circuit is well-designed and well-behaved in all worst cases.
- Oscillations because the breadboard had enough strays but the IC does not. This is a matter of good modelling. The breadboard can be expected to lie about this. SPICE may be helpful if applied thoughtfully.
- Obscure oscillations. The computer often lies badly about these. It may refuse to admit that they happen, and refuse to show them happening.

DC Output Voltage Errors (Room temp)

- Excessively broad distribution of Vref. When you expect a tolerance of ±3% and the observed distribution is excessive, the problem is usually either badly-matched resistors or transistors. The geometries must be identical as drawn and as masked. If your small resistor is short, your large resistor should be made of a group of short resistors. Sometimes it is advantageous to draw the Rs and Qs as cells, so that the mask-making process acts identically on each cell.
- Parts cannot be trimmed. Make sure that every voltage can be trimmed by at least one combination of trims; avoid any possible "holes" in your trim scheme.
- Interaction of trims. It is a good idea to make sure that in your plan, the size of each trim is (substantially) invariant of whether any of the previous trims have been done.
- Dependence of pre-trim Vref on beta. In general, a higher beta transistor has a lower Vbe. In some designs, a pinch resistor (whose resistance is a linear function of beta) is used to compensate for the shift of Vbe and improve the room-temp accuracy. In other cases, a pinch resistor is used to compensate for tempco, even as it degrades the room-temp accuracy.
- Band-gap "narrowing". With high-speed processes, the band-gap voltage (and the voltage for zero tempco) is decreased vs. ordinary transistors. A good breadboard can help determine the right place to operate. The computer cannot.

Tempco Errors

- Wrong "V-magic". Normally there is one value of Vref, which (if you trim to that voltage) gives the best tempco. But there may be circuit problems, masking problems, or process problems that can cause tempco to vary even when the Vref is well trimmed. The deviations are typically + or - 5 or 10 ppm/°C, but in a poor circuit can be considerably worse.
- Inability to trim to Vmagic. If there are "holes" in your trim scheme and you cannot get the Vref to trim to Vmagic with good precision, the tempco will be degraded by about 3 µv/°C per millivolt of error.
- Tempco curvature-correction circuit does not work well? MOST tempco curvature correction circuits do not work well the first time. Some are still bad after several tries. Computer models are of little help, or usually of much harm. Breadboards often are no better. Trial-and-error is usually needed.
- Leakages cause high-temp errors. You have to watch out for device leakages, tub leakages to substrate, and capacitor leakages. These are usually hard to model.
- Tempco does not agree with breadboard. If you used discrete low-tempco resistors in your breadboard instead of the monolithic (diffused or implanted) resistors, you can expect bad results.

Computer Modelling Problems

- Bad model of the Vbe versus temp. Most transistor models are poor. It is possible to tweak various parameters to get decent match of Vbe vs. temp, but the derivatives may not be realistic.
- Mis-typed data input. Be extremely meticulous about typing errors and copying errors. Strange results may occur otherwise.
- Failure to converge at cold temperatures. Yes, that is a problem...
- Failure to converge despite accurate "Node-set". Sometimes if Nodeset is TOO accurate, convergence is worse than if Nodeset is approximate.
- Failure to converge due to 35 volts across a diode (Vf) which does not conduct any current. Maybe you need a bigger diode...

- False current due to dv/dt across a hidden capacitance. When the collector voltage has a dv/dt, the "collector current" can show the current through the c-b junction, but the current through the c-substrate junction may not be included.
- False dv/dt due to .PLOT command truncating the .TRAN command. This has been observed to happen, when the end of the .PLOT is not at the same time as the end of .TRAN.

Miscellaneous Problems

- Low-voltage-lock-out problems. YOU must engineer carefully to get good results at low voltages, as the system may require good behaviour at very low supply voltages, and the reference may not want to give it. The breadboard works better than SPICE, here. Check at all temperatures.
- Thermal limit circuit works badly, poor errors, soft knee. Try to use a circuit that has been successful in the past. Hysteresis is often a good feature.
- High noise due to starvation. Be sure to check the breadboard. (Unless you are sure the computer gives good answers.)
- High noise due to high resistance. Check the breadboard carefully, and use realistic transistor samples. High-beta parts will have higher rbb' than low-beta ones.
- Bad matching due to buried layer. If the actual buried layer causes crystal growth to fall in the middle of a critical transistor, Vbe matching may suffer. Note, the crystal growth at the surface is shifted from where the buried layer appears to be.
- Saturation due to insufficient buried layer. Transistors do not run well at warm temperatures when asked to run near saturation, especially at high temp, when the buried layer has been omitted.
- Assembly shift. This is usually caused by stress sensitivity. A good layout can help minimize this.

Czardom

At National, we appointed a Czar to oversee the design of all band-gap circuits, and to monitor and to log all good and bad results. This has helped cut down on the number of repetitive foolish errors.

Bibliography

1. LM113 Data Sheet, National Semiconductor Linear Data Book, 1972, 1976.
2. R. J. Widlar, "An Exact Expression for the Thermal Variation of the Emitter-Base Voltage of Bipolar Transistors", Proc. IEEE, Jan. 1967.
3. AD580 Data Sheet, Analog Devices, Inc., 1975.

NOTICE: THE MENTION OR DESCRIPTION OF ANY CIRCUIT HEREIN DOES NOT CONFER ANY RIGHTS TO THE USE OF SUCH CIRCUITS, IN VIEW OF PATENT RIGHTS THAT MAY APPLY TO SUCH CIRCUITS.

Robert A. Pease
Mail Stop D2597A
National Semiconductor
P.O. Box 58090
Santa Clara, CA 95052-8090
E-mail: rap@webteam.nsc.com

Originally published in the IEEE Proceedings of the 1990 Bipolar Circuits and Technology Meeting, September 17-18 1990 in Minneapolis, Minnesota.