Reference Generators part1

August 20, 2019

Now we are coming to the most disappointing and frustrating chapter of analog circuit design: The references!

1 Use of references

Almost every analog function we want to implement needs some kind of a reference. This can be a current or a voltage. Using a reference voltage is the most common.

We can build ADCs and DACs of almost unlimited precision. Here are some examples found in many electronic systems.

bits	class	quant. error	architecture	application	full scale	LSB
6-8	high speed	0.5% to 2%	flash	RF	3V	15..60mV
8	low performance	0.5%	succ. approximation	low cost	5V	20mV
10	medium performance	0.1%	succ. approximation	standard	5V	5mV
14..16	high performance	0.002%	delta sigma	audio, wide dynamic range	3.3V	$50..200\mu V$
24	very high perf.	$<$0.0001%	delta sigma, dual slope	measurement systems	3.3V	$0.2\mu V$

The table looks as if anything is possible provided you are willing to spend the money.

1.1 The sad truth

Resolution has little to do with absolute accuracy! It only means a change of the input signal leads to a change of the digital output if the LSB is exceeded.

The quantization error only applies to perfectly layouted ratiometric applications!

1.1.1 What your measurement system does

The ADC or DAC does nothing more than comparing a signal with a reference. Any change of the reference will become visible in the conversion result as a gain error. Even if you have a 24 bit ADC but your reference voltage is only 0.1% stable the

error of the full scale signal will be exactly these 0.1%. (Well, the accuracy close to the 0V crossing still is the LSB ($\pm 0.1\%$ reference error).

Example 1: The ADC has (ideal perfect) 16 bits and measures 0V:
The real signal is between $-100\mu V$ and $100\mu V$.

Example 2: The ADC has (ideal perfect) 16 bits and you measure full scale 3.3V. The reference has 0.1% accuracy:

The real signal has the quantization error of $-100\mu V$ to $100\mu V$ plus the reference error of -3.3mV to +3.3mV. The full scale error becomes $\pm 3.4mV$. The full scale error is determined by the error of the reference voltage!

2 Where does the reference come from

There are many ways of creating a reference voltage. Each of these methods has it's pros and cons and needs careful investigation.

2.1 Taking the reference from a voltage divider

The most simple way to create a reference voltage is to simply divide a supply voltage with a resistor divider. Assuming the resistors match as well or better than the network used in the ADC or DAC we can probably neglect the resistor errors. The most dominant gain error of the system will be caused by the production spread of the power supply.

Low cost voltage regulators often are specified to have about $\pm 4\%$. (production spread and temperature drift.). If you spend some cent more finding regulators with $\pm 1\%$ is possible.

If the system relies on a battery better expect supply voltage changes of easily $\pm 20\%$. As long as this just leads to a slowly changing gain error of an audio signal you may in fact get along with that provided there are no fluctuations in the the audible frequency range (These would be converted into AM of the audio signal!).

2.2 Using diodes

Using a diode or a MOS diode to create a reference voltage works for a very limited temperature range. A bipolar diode typically has a temperature coefficient of -2mV/K and a production spread of some 10mV (at one temperature, one current). The forward voltage is about 600mV. The resulting relative drift becomes

$$drift_{rel} = \frac{-2mV/K}{600mV} = -0.33\%/K$$

For an extremely narrow temperature range this may be acceptable. As soon as you want to cover for instance classical automotive temperatures (-40°C to 125°C ambient temperature) using a simple diode isn't an option. (Except if you want to

have exactly this temperature coefficient like in the low power oscillator example of my last post)

2.3 Using zener diodes

Producing a reference voltage using zener diodes is a classical solution of board design with discrete components of the 1960s and 1970s. Standard discrete zener diodes usually have specifications of $\pm 3\%$ for about 5V break down voltage [4]. Typical temperature coefficients are in the range of $\pm 0.02\%/K$. As soon as the required zener voltage deviates from about 5V the performance suffers. The following plot shows temperature coefficients (in %/K) of a standard precision zener diode versus the zener voltage.

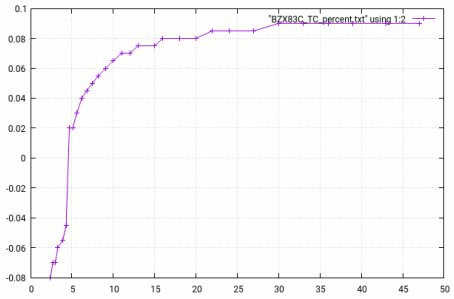

Fig.1: Temperature coefficient in %/K versus zener voltage

This relationship is very similar for almost all zener diodes because it relates to basic physical properties of silicon.

There are some integrated "zener diodes" that consist of a zener diode plus a standard diode (in forward operation) to tweak the temperature coefficients. The negative temperature coefficient of the diode operating in forward direction is intended to compensate the positive temperature coefficient of a zener diode of about 7..8V.

2.3.1 High precision zener diodes

There are devices available that are called 'ultra high precision'. These are zener diodes that are exactly tweaked for the lowest possible temperature coefficient. The nominal voltage however still have a wide spread of $\pm 2\%$. What justifies the name 'ultra high precision' is the lower temperature coefficient often specified in the range of $\pm 0.001\%/K$. Since we want both, a high accuracy and a low temperature coefficient we have to trim.

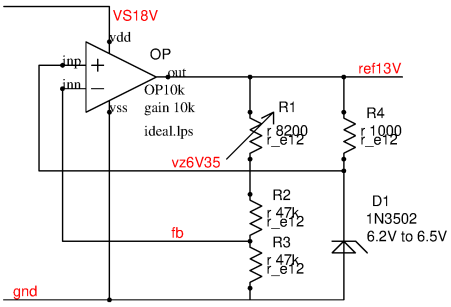

Fig.1: Precision reference with trim capability

In the circuit shown above a low drift zener diode is used. The production spread of the zener diode can be trimmed by adjusting R1. The reference voltage created is about twice the zener voltage. In this example this leads to a reference voltage of about 13V. To match the requirements of typical integrated ADCs this reference has to be divided down to 5V or 3.3V or whatever the ADC chips require.

2.3.2 Integrated zener diodes

Integrated zener diodes [1, pages 9, 27, 119] usually are not the components the process is optimized for. In most integrated circuits processes the zener diodes more or less perform like standard zener diodes. Even worse in many processes the zener diodes are components of minor interest. Typical examples are the base-emitter surface zener diodes that come for free in most bipolar processes. The junction is optimized for the performance of the NPN transistors but not for the characteristic of the zener diodes! Such BE-diodes used as zener diodes offer typical break down voltages in the range of 7V $\pm 0.5V$ and temperature gradients in the range of 0.04%/K. In addition surface zener diodes are susceptible for contamination and show long term drift of the break down voltage and the small signal resistance.

Making things worse a zener diode is more or less a combination of two different physical effects [3, pages 151ff]:

- Tunneling through a very narrow depletion zone usually has a negative temperature coefficient
- avalanche effects have a positive temperature coefficient
- avalanche produces noise that in extreme cases can reach up to $1V_{pp}$for a 7V surface zener diode

The zener diode achieves the best temperature stability if the temperature coefficients of the tunneling and the avalanching exactly cancel - which is rarely the case!

In some processes the junction acting as a zener diode can be moved away from the surface. This is called a buried zener diode. Buried zener diodes typically operate close to 5V and have about one magnitude less avalanching noise than 7V zener diodes. Long term drift (1000h operating time) can be expected between $+$100mV

and +500mV depending on the current the diode is operated with. Due to lattice defects caused by the avalanche current the small signal resistance can change by about one magnitude. (I have seen cases with an increase of the small signal resistance changing from 50Ω end of production to 800Ω after 1000h operation at 80°C.)

On chip zener diodes can't be recommended as a voltage reference unless the process is especially tweaked for the zener diode. (Tweaking the process for the zener often has negative side effects on the performance of other components on the chip. For this reason this usually isn't done in standard semiconductor processes.)

3 Bandgap reference

Each semiconductor has a certain bandgap voltage. This voltage is determined by the energy the electrons need to move from the valence band to the conduction band. At 0K the bandgap voltage exactly corresponds the forward voltage. However operating a bipolar diode at 0K will not work because a certain temperature is needed to mobilize the electrons. So the bandgap voltage at 0K must be regarded as something more theoretical.

With increasing temperature the energy available for the electrons comes from thermal movement and from the voltage applied over the junction. This leads to a forward voltage that decreases with temperature. (This is why silicon has a bandgap voltage of about 1.11V but the forward voltage of a diode at room temperature is about 0.7V. The difference is caused by the thermal energy of the electrons lowering the energy that has to be delivered by the electrical field.)

A bandgap [1, page 289 ff] combines the temperature coefficients of the diode forward voltage and the thermal voltage.

There are hundreds of possible implementations of bandgaps. In the following only the Widlar bandgap and the Brokaw bandgap is shown to illustrate the concept. Other bandgaps follow the same concept. A bandgap reference adds the forward voltage of a bipolar diode and the temperature voltage created by the difference of two forward voltages operated at different current densities. The forward voltage of a bipolar diode (operated at a constant current) has a negative temperature coefficient of about -2mV/K. The difference of the forward voltages depends on the absolute temperature, the Boltzmann constant, the electron charge and the logarithm of the current density ratio.

Ideally at 0K the forward voltage of a bipolar diode exactly corresponds the bandgap energy of the semiconductor while the temperature voltage (the difference of two forward voltages) becomes zero (in other words: as long as the voltage is below the bandgap voltage all carriers are in the valence band. As soon as we exceed the bandgap voltage all carriers jump into the conductive band at 0K. The ideal diode is infinitely conductive at 0K).

Note: The theoretical bandgap voltage of silicon at 0K is 1.11V. Building bandgaps operating at room temperature however we usually end up with a bandgap voltage of about 1.16V to 1.25V. 1,23V usually is a good initial guess for most technologies.

$$V_t = \frac{k * T}{e} \tag{1}$$

$$k = 8.617332478 \times 10^{-5} eV/K$$

$$e = 1.60217656535x10^{-19} C$$

At 300K we get:

$$V_{t300k} = 25.852mV$$

Operating two diodes at different current densities we get:

$$\Delta V_{be} = V_t * ln(\frac{i_1}{i_2}) \tag{2}$$

The temperature coefficient becomes:

$$\frac{dV_t}{dT} = \frac{k}{e} = 86.1733\mu V/K$$

$$\frac{d\Delta V_{be}}{dT} = ln(\frac{i1}{i2}) * \frac{k}{e} \tag{3}$$

3.1 The Widlar bandgap

To compensate the -2mV/K of a diode we have to multiply this temperature coefficient to reach +2mV/K. The most simple circuit doing this job is the Widlar bandgap.

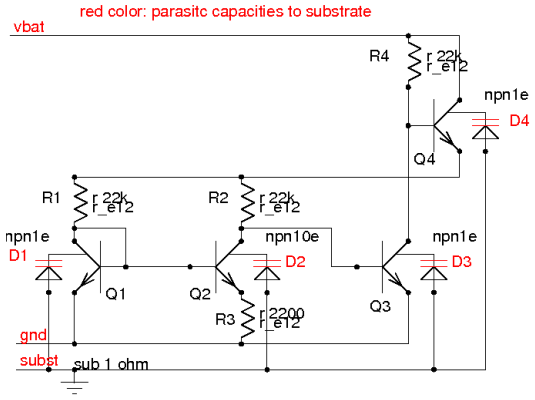

Fig.3.1.1: Widlar Bandgap

Ideally Q3 is operated at exactly the same current density as Q1. (Later we will have a look at errors caused by violating this condition.) So the voltage at the base of Q1 and the base of Q3 are equal and the currents trough E1 and R2 are equal. Since Q2 has more emitters than Q1 it operates at a different current density. (in out case the current density is 10 times less.) So at R3 we will find a voltage of:

$$V_{R3} = V_T * ln(n) \tag{4}$$

$$n = 10$$

$$V_{R3} = 25.852 * ln(10) = 59.526mV$$

at 300K. The voltage drop at R1 and R2 becomes:

$$V_{R1} = V_{R2} = 10 * V_{R3} = 595mV$$

The bandgap voltage becomes:

$$V_{bg} = V_{be} + V_{R2} \tag{5}$$

Ideally this should exactly be the bandgap voltage of the chosen material (1.23V in case of silicon). Note that the base emitter voltage of the above equation is part of the regulation loop. It is the base emitter voltage of Q3!

3.1.1 Practical considerations of the simple Widlar bandgap

Most important systematic errors Up to now we followed the assumption that the current through R4 is about equal to the current trough R1 and R2. If the supply voltage changes this assumption is no more true. With increasing supply voltage Q3 has to carry an increasing current. So the base emitter voltage of Q3 increases as well. The current flowing in R1 and R2 can be calculated (neglecting base currents) as:

$$I_{Q2} = \frac{V_T * ln(n)}{R_3} \tag{6}$$

While the current through Q3 is:

$$I_{Q3} = \frac{V_{bat} - V_{bg} - V_{beQ4}}{R_4} \tag{7}$$

Thus the difference of the base emitter voltages of Q3 and Q1 becomes:

$$V_{errR4} = V_T * ln(\frac{R_4 * V_T * ln(n)}{R_3 * (V_{bat} - V_{bg} - V_{beQ4})} \tag{8}$$

and the deviating bandgap voltage calculates as:

$$V_{bg} = V_{bgideal} + V_{errR4} \tag{9}$$

Even worse the error is supply voltage dependent. A simple octave script will calculate the error versus Vbat.

```
vbat=[1.4:0.1:12]
R1=22.000
R2=22.000
R3=2.200
R4=22.000
n=10
vt=25.8
vbe=0.65
vbg=1.23
vr3=vt*0.001*log(n)
ir3=vr3/R3
vr4=vbat-vbe-vbe
ir4=vr4/R4
verr=vt*log(ir4/ir3)
plot(vbat, verr)
```

The resulting plot shows the error voltage in mV versus the supply voltage vbat in V.

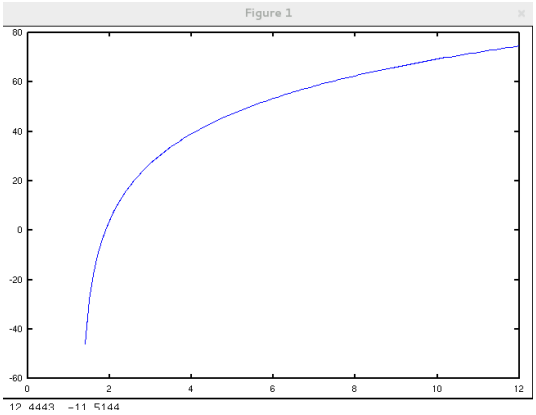

Fig.3.1.1.1: Error of the simple Widlar bandgap in mV with pure resistor bias.

Most important statistical errors The production spread of the bandgap depends on the following factors:

1. Matching of the transistors Q1 and Q2
2. Matching of transistors Q1 and Q3
3. Matching of resistors R3 versus R1 and R2

Mismatch of the base emitter voltage (at the same current density) of Q1 and Q2 will lead to a deviation of the current flowing in R2 and R3. So the error contribution of an offset to the bandgap voltage will be:

$$\Delta V_{bg1} = \Delta V_{be12} * \frac{R_2}{R_3} \tag{10}$$

An offset of Q3 will simply be added to the bandgap voltage:

$$\Delta V_{bg2} = \Delta V_{be13} \tag{11}$$

The offset of resistors typically is a relative error given in %. This error usually becomes big for small resistors (having a small silicon area). In most cases it is sufficient to observe the error of R3 while R1 and R2 occupy much bigger areas and their deviation can be neglected.

$$\Delta V_{bg3} = \frac{\Delta R_3}{R_3} * V_T * ln(n) * \frac{R_2}{R_3} \tag{12}$$

Since these are statistical errors that should be independent from each other we have to add the power:

$$\Delta V_{bg} = \sqrt{\Delta V_{bg1}^2 + \Delta V_{bg2}^2 + \Delta V_{bg3}^2} \tag{13}$$

EMC performance of the Widlar bandgap The Widlar bandgap is fairly robust against transients on the supply vbat. The only way to couple transients into the bandgap voltage is the miller capacity of Q4. Since the bandgap must have a frequency compensation anyway (for stability of the regulation loop) it is a good idea to use a parallel compensation (simply a capacitor from the base of Q4 to ground). On the other hand a miller compensation using a capacitor between the collector of Q3 and the base of Q3 is cheaper (but is less performant with regards to transients on Vbat).

The big drawback of the Widlar bandgap is its sensitivity to substrate noise. Since Q2 is the biggest active component it also has the biggest substrate capacity (collector of Q2 to substrate). Substrate noise picked up by the substrate capacity of Q2 will be peak rectified by the strongly non linear characteristic of Q3. So substrate noise will turn on Q3 and turn off the bandgap! The substrate noise sensitivity can be improved if the emitters of Q1 and Q3 and resistor R3 are tied to substrate rather than to metallic circuit ground. Modern CAD tools however will flag this design style as soft connect errors.

If other components in the neighborhood of the bandgap inject electrons into the substrate the big area of Q2 will pick up these electrons. (It becomes the collector of a parasitic lateral NPN transistor. Minority carrier injection into the substrate close to Q2 will turn off Q3 and the bandgap voltage increases in an uncontrolled way!

Bottom line we must state that the Widlar bandgap can not be recommended for applications with high RF or electron injection into the substrate.

3.2 The Brokaw Bandgap

The Brokaw bandgap [2]solves the problem of matching the currents of Q3 of the Widlar bandgap with the currents inside the bandgap. This kind of bandgap became very attractive with the introduction of lateral PNP transistors into semiconductor design libraries end of the 1960s. (Brokaw published 1974). The most simple way of building a Brokaw bandgap is shown below:

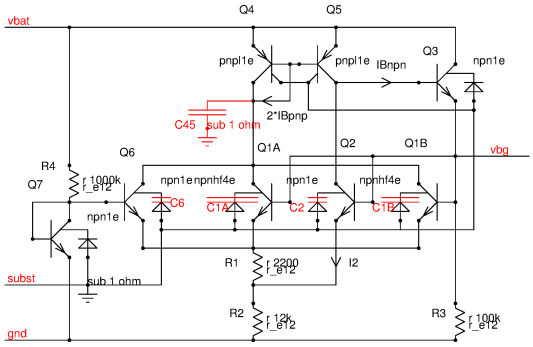

Fig.3.2.1: The most simple version of the Brokaw bandgap

The bandgap published by Brokaw was a bit more complex solving errors caused by the base currents of the transistors. This simplified version however serves well to explain the concept. If current mirror Q4, Q5 and output stage Q3 is designed using MOS transistors this will solve most of the systematic errors too.

Different from the Widlar bandgap the Brokaw bandgap needs a starter (R4, Q7, Q6). Once the bandgap is running the emitter of Q6 is lifted to about 600mV and starter Q6 turns off.

R2 now carries twice the bandgap current. This reduces the are needed for high precision resistors. R3 and R4 are simple pull up or pull down resistors that can be designed using minimum width devices. Q1 is split in two devices (Q1A and Q1B) placed left and right of Q2 for better matching. In this example the ratio of the current densities becomes 4+4=8. So the ideal bandgap voltage calculates:

$$V_{R3} = V_T * ln(n) \tag{14}$$

In the example shown we have:

$$n = 8$$

So the voltage at R1 becomes:

$$V_{R1} = V_T ln(n)$$

At 300K we get:

$$V_{R1300K} = 25.852mV * ln(n) = 53.758mV$$

Neglecting base currents assuming an ideal current mirror Q4, Q5 we will find twice the current of R1 flowing in R2. So the voltage across R2 is:

$$V_{R2} = 2 * \frac{R_2}{R_1} * V_T * ln(n) \tag{15}$$

In the example shown this calculates:

$$V_{R2} = 586.45mV$$

Assuming Vbe=650mV we get an ideal bandgap voltage of:

$$V_{BGideal} = V_{be} + V_{R2} = 650mV + 586.45mV = 1.2365V$$

To calculate the frequency compensation it is necessary to know the transconductance of the bandgap. To concentrate on the essential things in the following figure the current mirror is built using PMOS transistors. So we don't have any influence of base currents in the mirror. The starter has be omitted as well. To make things a bit more convenient the output current of the bandgap is defined to flow into the circuit. Since the current mirror doesn't loose any current the output current becomes

$$I_{out} = I_2 - I_1 \tag{16}$$

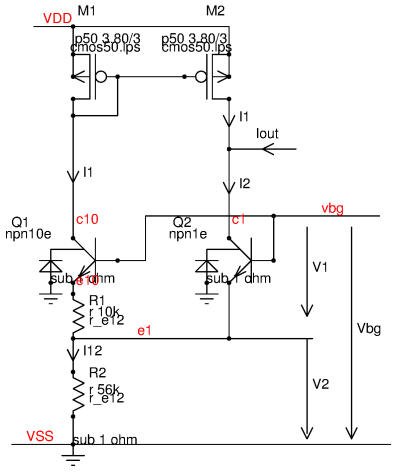

Fig.3.2.2: The core of the Brokow bandgap

Ta calculate the transconductances it is more convenient to start with V1. In the operating point the currents through the transistors are equal.

$$I_1 = I_2 = \frac{I_{12}}{2} = I \tag{17}$$

Since both transistors operate at the same current we can simply assume they have the same transconductance gm.

$$gm = \frac{I}{V_T} = \frac{I * e}{k * T} \tag{18}$$

For Q2 that has no resistor in the emitter we find the small signal transconductance

$$\frac{dI_2}{dV_1} = gm \tag{19}$$

At Q1 we have an emitter impedance in series with R1 leading to

$$\frac{dI_1}{dV_1} = \frac{1}{\frac{1}{gm} + R_1} \tag{20}$$

Thus the small signal transconductance of the bandgap becomes

$$\frac{dI_{out}}{dV_1} = \frac{dI_2}{dV_1} - \frac{dI_1}{dV_1} = gm - \frac{1}{\frac{1}{gm} + R_1} \tag{21}$$

This is an important result. The higher we chose R1 the higher the loop gain we can achieve. Of course we can not chose R1 arbitrarily. To be able to use a high R1 we have to use a high size ratio between the two transistors. On the other hand R1 following the logarithm of the areas of the emitter the increase of R1 is limited. Practical values for the size ratio of the transistors are about 8 to 15.

The second important observation is that the transconductance is a difference between a bipolar transconductance and a degraded transconductance (resistor in the emitter of the bigger transistor). The transconductance of the Brokaw bandgap is always lower than the transconductance of a bipolar differential amplifier. Therefore it is very important to have a good current mirror. In the circuit shown this current mirror is built using MOS transistors in stead of bipolar PNP transistors. This avoids the systematic error of the PNP transistor base currents. If the MOS current mirror is replaced by bipolar PNP transistors we add a significant systematic error!

Up to now we referred everything to V1 in stead of the bandgap voltage. So we have to find the ratio (small signal) between V1 and Vbg. The impedance at the emitter is

$$Z_e = \frac{1}{gm} \tag{22}$$

The (small signal) admittance found from node vbg to node e1 becomes

$$Y = \frac{1}{Z_{e1}} = gm + \frac{1}{\frac{1}{gm} + R1} \tag{23}$$

$$Z_{e1} = \frac{1}{gm + \frac{1}{\frac{1}{gm} + R_1}} \tag{24}$$

Now the divider consisting of the impedance at e1 and R2 can be calculated. Since this is a small signal consideration the equation is written in its differential form.

$$\frac{dV_1}{dV_{bg}} = \frac{Z_{e1}}{Z_{e1} + R_2} = \frac{1}{1 + R_2 * gm + \frac{R_2}{\frac{1}{gm} + R_1}} \tag{25}$$

Referring everything to Vbg the transconductance of the bandgap becomes

$$\frac{dI_{out}}{dV_{bg}} = \frac{dI_{out}}{dV_1} * \frac{dV_1}{dV_{bg}} = \frac{gm - \frac{1}{\frac{1}{gm} + R_1}}{1 + R_2 * gm + \frac{R_2}{\frac{1}{gm} + R_1}} \tag{26}$$

Using the expression for the delta Vbe (depending on the emitter ratio n) and the current we can determine R1.

$$R_1 = \frac{V_T}{I} * ln(n) \tag{27}$$

Furthermore replacing gm by the current and Vt simplifies the equation for the transconductance of the bandgap

$$\frac{dI_{out}}{dV_{bg}} = \frac{I}{V_T} * \frac{1 - \frac{1}{1+ln(n)}}{1 + \frac{I*R_2}{V_T} * \frac{2+ln(n)}{1+ln(n)}} \tag{28}$$

R2 is intentionally left in the equation because sometimes we might want to build a bandgap that has a temperature coefficient or someone wants a double bandgap. This can at low cost be built stacking further diodes and multiplying up R2. But this of course reduces the transconductance.

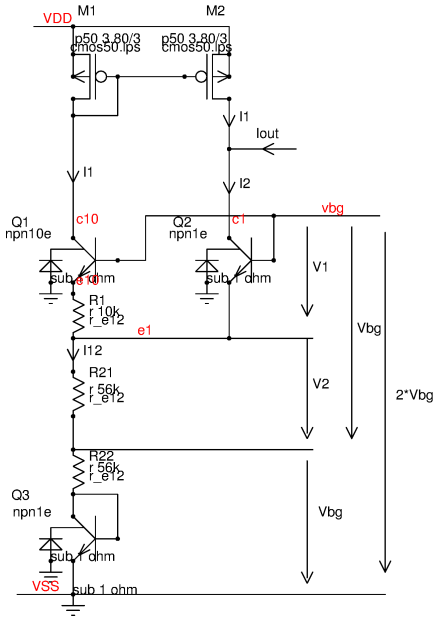

Fig.3.2.3: Stacking a diode more to create a double bandgap at low cost

Supply transient response of the Brokaw bandgap: The bigger NPN transistor with the multiple emitters has a higher capacity to ground than the smaller single emitter transistor. Therefore at a fast rising edge of the supply the Brokaw bandgap usually produces an overshot at the output. At a fast falling edge the bandgap turns off for some microseconds.

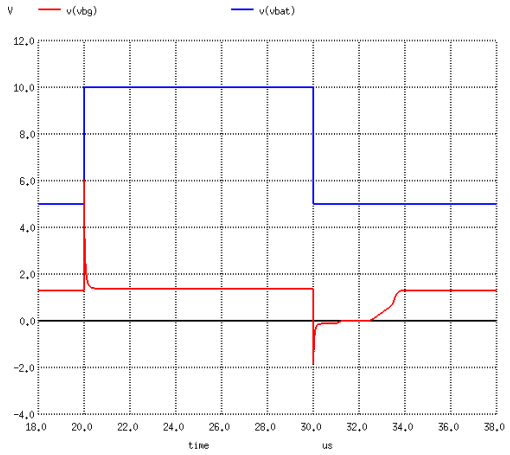

Fig.3.2.4: Response of the Brokaw bandgap to a transient on the supply rail

Replacing the PNP transistors by PMOS transistors helps to reduce the parasitic capacity to ground at the base of the current mirror. But the improvement is limited because we still have the capacity of the big NPN transistor that can't be avoided in the Brokaw bandgap.

EMC performance of the Brokaw bandgap: The Brokaw bandgap is most sensitive to RF injected into the emitters of the bandgap transistors. Q1 and Q2 are extremely fast rectifiers. Usually the base of both transistors is more or less low resistive tied to the ground node of the bandgap (often there is a frequency compensation tying the base to circuit ground).

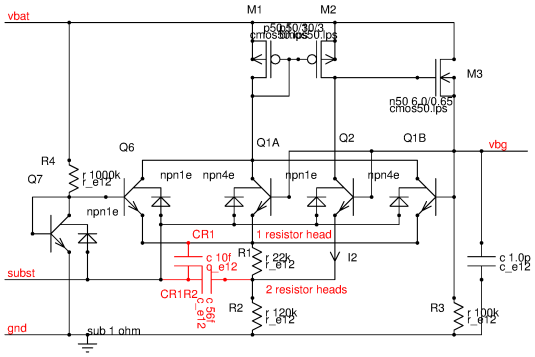

Fig.3.2.5: Parasitic capacities CR1 and CR1R2 can couple substrate noise into the emitters of the bandgap transistors if the resistors are placed over substrate.

Substrate noise coupled into the emitters will be amplified and peak rectified (negative peak) by Q1 and Q2. Usually Q2 wins because it sees the capacity of two resistors. This will pull down the bandgap even if the injected RF frequency is higher than the transit frequency of the transistors (common base circuit like the UHF tuners of the old days!).

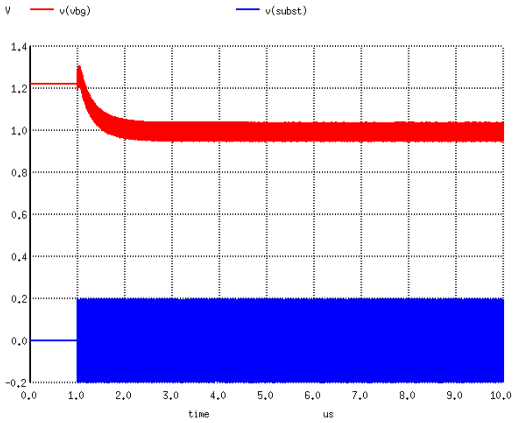

Fig.3.2.6: 200mV of substrate bounce at 200MHz pulls down the bandgap

The effect of RF coupling through the parasitic capacity at the bottom side of the resistors becomes stronger when the circuit is designed with higher impedances (bigger resistor area AND higher node impedance square the problem reducing current consumption of the bandgap!)

It is good engineering practice to use resistors over wells to isolate the resistors from substrate noise.

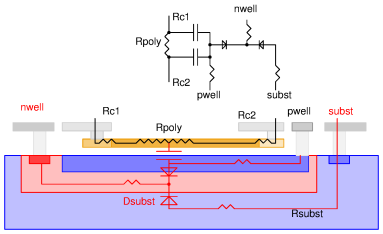

Fig.3.2.7: double well isolation of the poly silicon resistor

pwell has to be connected to the ground node of the bandgap (exactly where the bandgap voltage refers to). nwell usually is connected to the supply of the bandgap to prevent latch up. This way substrate noise is attenuated twice before it reaches the back side capacity of the resistor.

3.3 Common Properties of all Bandgap Circuits

Although there are hundreds of possible implementations of a bandgap there are certain common properties that apply to all of them.

1. The compensation of the temperature gradients of Vt and Vbe must always be fitted for a certain operating temperature
2. The resulting output voltage of the bandgap is a parabola like function.
3. Even trimming doesn't remove this parabola. It only shifts it up or down.
4. If the deviation caused by the parabola can't be accepted a second order compensation may further improve accuracy by about factor 3. This kind of compensation requires extremely precise models of the bipolar transistors.
5. Alternatively the parabola can be characterized and compensated changing the trim values with chip temperature. (This requires test insertions at various temperatures and makes the product significantly more expensive!)
6. If all these measures aren't sufficient the last resort is to operate the bandgap in a thermostat.
7. Since the bandgap energy can be modified by applying mechanical stress a bandgap voltage can change some mV to some 10mV exposing the chip to mechanical stress. (I have seen up to 20mV at 1.23V bandgap voltage)
8. For extreme requirements (μV- range) microphoning of the chip (pick up of mechanical excitation of the board) should also be considered as an error source.

Achievable precision of a bandgap is in the range of 2% untrimmed, 0.2% to 0.5% after trimming (for a temperature range of -50°C to 150°C). If you need a higher precision you will have to measure each individual bandgap and calibrate your board depending on this result. This calibration must be repeated if the solder joints were heated to the melting temperature because this may have changed mechanical tensions inside the package.

The following plot shows the result of 500 Monte Carlo simulation runs of a double bandgap (before trimming). This is already a high precision design! Trimming can remove the spread, but not the parabola. After trimming the remaining error over the whole temperature range caused by the parabola characteristic will still be in the range of $\pm 8mV$ for a temperature range of -40C to 150C.

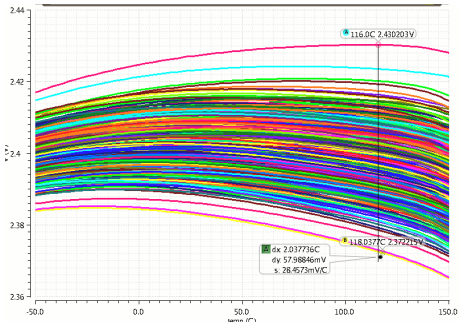

Fig.3.3.1: parabola characteristic of 500 double bandgaps before trimming

This plot nicely shows the spread as well as the change of the temperature behavior to be expected in high volume production. It does not include mechanical stress. On the chip place the bandgap in the center of the die if possible to reduce mechanical stress.

Extreme measures as described in items 5 to 8 usually can only be done for high precision measurement equipment that is produced in low volume and at a high sales margin. For high quality measurement systems consider non standard packages to reduce drift caused by the mechanical stress (low stress mold, Poly Imid coating, ceramic packages, metal can packages).

In other words: If you need a full range accuracy exceeding 8bit (better than 0.5% gain error of the DAC or ADC) the reference generation will start to have a significant impact on the price of the product because trimming at the chip manufacturer alone won't be sufficient anymore to reach this target.

References

[1] Paul R. Gray, Robert G. Meyer, "Analysis and Design of Analog Integrated Circuits", John Wiley & Sons, 1984

[2] A. Paul Brokaw, "A simple Three-Terminal IC Bandgap Reference", IEEE Journal of solid state circuits, VOL. sc9, no. 6, December 1974

[3] R. Mueller, "Grundlagen der Halbleiter Elektronik", Springer, 1987

[4] "Dioden", Siemens AG, 1980