Reference generators part 3

September 4, 2019

Some designers have to struggle with CMOS technologies that - at least officially - don't even have bulk diodes that are released to be used for building a bandgap. There are two ways out:

- Build a bandgap using bipolar diodes that are officially not released as a component.
- Build a bandgap using MOS transistors operating in weak inversion.

1 Weak Inversion Bandgap

In principal every kind of bandgap built with bipolar transistors can just as well be built using CMOS transistors operating in weak inversion. In bipolar bandgaps the voltage with the positive temperature coefficient is constructed with a delta Vbe.

$$\Delta V_{be} = \frac{k*T}{e} * ln(m) \tag{1}$$

with k being the Boltzmann constant, T being the temperature in K and e being the electron charge. m is the ratio of the current densities flowing in the bipolar transistors. Using MOS transistors operating in weak inversion the equation looks very similar.

$$\Delta V_{gs} = \frac{k*T}{e} * n * ln(m) \tag{2}$$

The only difference is the factor n. n is the inverse coupling factor of the gate voltage to the channel.

$$n = \frac{C_g + C_{bulk}}{C_g}$$

Next step the gate to channel capacity C_g and the channel to bulk capacity C_{bulk} must be calculated. (The calculation can be found in the chapter describing the MOS transistor. Here only the result is shown.)

$$n = 1 + \frac{\varepsilon_{si} * t_{ox}}{\varepsilon_{sio2} * \sqrt{\frac{2*\varepsilon_{si}*(\Phi - V_b)}{q*N_b}}} \tag{3}$$

Looks a bit ugly. Bottom line we have a multiplication factor n in the range of 1.1 to 1.6 for most technologies. What is not so nice is that this factor is not fully constant! It changes a little bit with the doping of the bulk and the work function Φ (at the surface of the silicon) and the bulk voltage (back gate). As a consequence bandgap circuits using MOS transistors in weak inversion by concept are less precise and more technology dependent than their bipolar counterparts. Literature reports weak inversion bandgaps with a 1σ spread in the range of 1% to 2%. The thinner the gate oxide the better it gets. (Well, a bipolar transistor can be regarded as a MOS transistor with a gate oxide thickness of $t_{ox} = 0$. So n approaches 1 and we exactly get the equation of ΔV_{gs} .)

The temperature coefficient of V_{gs} becomes:

$$\frac{dV_{gs}}{dT} = \frac{k * n}{e} * ln(m) \tag{4}$$

At 300K this is

$$\frac{dV_{gs}}{dT} = n * ln(m) * 86.1733 \mu V/K$$

To compensate n the resistor ratio just has to be adjusted a little bit (compared to the bipolar bandgap).

Since we have shown that a weak inversion bandgap by concept is less precise than a bipolar bandgap, why do people still use it? The answer is simple: MOS transistors in modern technologies are smaller than bipolar ones. So a weak inversion bandgap is just cheaper than a bipolar one. That is the only attractive feature of it.

1.1 Open loop weak inversion bandgap still using a bipolar diode

The most simple example of a weak inversion bandgap is shown below.

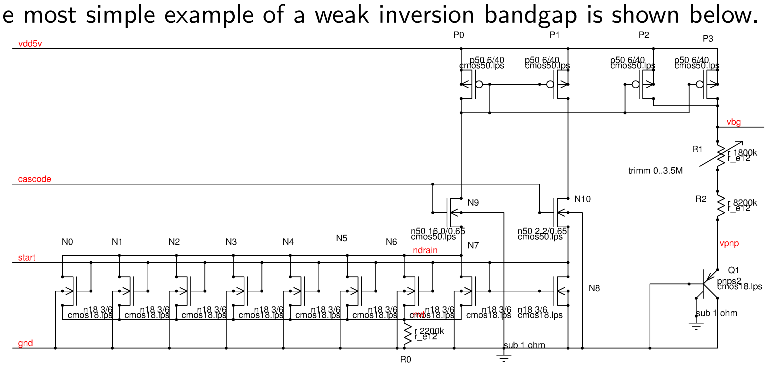

Fig.1.1: Example of a simple weak inversion bandgap

N0 to N8 operate in weak inversion. The voltage drop across R0 is in the range of 60mV. The current flowing is 30nA. The current flowing through R1 and R2 is 60nA.

1.2 Weak inversion bandgap using MOS transistors only

Q1 can be replaced by an NMOS transistor. In weak inversion the threshold voltage like Vbe of a bipolar transistor has a negative temperature coefficient. Normally a copy of N0 to N8 will be used as a "diode". Unfortunately Replacing Q1 by a MOS diode adds further error sources to the design. The threshold of a MOS transistor can be engineered (intentionally as well as unintentionally) by the doping of the gate poly silicon and by the doping at the surface of the channel.

$$V_{th} = -\varphi_{bi} + 2 * \varphi_b + \sqrt{4 * \epsilon_{si} * q * N_a * \varphi_b}/c_{ox}$$

In this equation $\varphi_{bi}$is the built in voltage that depends on the work function of the gate material. In case of a poly silicon gate this value can be adjusted in a certain range by the gate doping. For n-doped poly silicon gates of a NMOS transistor $\varphi_{bi}$is in the range of -0.1V to -0.2V [1] . Using p-doped poly silicon for the NMOS transistor the built in voltage φ_{bi} can becomes positive leading to a threshold of 0V or even a negative threshold. (This would be a somewhat unusual process, but it can be done.) If the gate material differs from silicon (for instance if a real metal gate is used) the difference of the work functions of the gate material and silicon has to be used. This means transistors using a gate that is not poly silicon may have completely different thresholds! (The work function of silicon is 4.05eV. The built in voltage is the difference between the work function of the gate material and silicon. A nice table can be found at [2])

φ_b is the "distance" between the Fermi level of the doped bulk semiconductor and the intrinsic Fermi level.

$$\varphi_b = V_{th} * ln(\frac{N_a}{n_i})$$

N_a is the acceptor doping of the substrate. q is the elementary charge. The choice of the substrate doping can modify the threshold in a range of 1V to 2V. (c_{ox} is the specific capacity t_{ox}/ϵ_{ox} of the gate dielectric.)

Since there are so many possibilities to modify the gate voltage this is done to tune the thresholds for the best performance of the logic. For the weak inversion bandgap this means the bandgap voltage depends on the way the threshold of the transistor used is tuned. Here are some examples of my own experience:

process usage	tox	Vth	Vbg	remark
5V	25nm	1.7V	3.9V	simulated, not used in a product due to spread seen in corner simulation
5V	15nm	1.3V	3.3V	only used for cascode bias. Never used as a reference.
3.3V	7nm	0.9V	2.3V	very poor accuracy
1.2V	3nm	0.4V	1.4V	used transistor without halo implant

The table already shows that the weak inversion bandgap voltage deviates a lot from the bandgap voltage found using bipolar transistors. Replacing Q1 of figure

1.1 by a MOS diode means you are at the mercy of the process engineer. If the process is tuned for better digital performance (I will be tuned. You can bet on it!) your bandgap will change it's behavior and it's typical voltage.

1.3 Conclusion

If you can, use a bipolar bandgap. If you are forced to use a ΔV_{th} instead of a ΔV_{be} to create your PTAT voltage try to at least add the PTAT voltage to a bipolar diode forward voltage. Generating both, the PTAT voltage and the NTAT voltage from MOS diodes is the LAST DESPERATE ESCAPE and you should have a wide trimming range to accommodate process changes in the future that are not at all in the models you have today! (his leads to a product that needs readjustment of the trimming each time the process went through a yield adjustment)

The design process is in four steps:

1. Do a first guess of the operationg points by manual calculation
2. Simulate the threshold and it's temperature coefficient at the guessed operating point.
3. Correct the initial guess using the first simulation result
4. Run the final optimization by simulation (hoping the models are correct) and add a trimming network giving you at least twice the trimming range the corner simulation suggests.

References

[1] "Bauelemete der Halbleiter Elektronik (Halbleiter Elektronik 2)", R. Mueller, Springer 1987

[2] "Metal-dielectric band alignment and its implications for metal gate complementary metal-oxide-semiconductor technology", Yee-chia Yeo, Tsu-Jae King, Chenming Hu, Journal of applied physics volume 92 No.12 2002