Report on the design of generic bandgap IP in CMOS technology

Dimitri GALAYKO

1 PTAT current generator

1.1 Self-biased PTAT source

Electronic circuits need stable reference voltage sources. The first cause of voltage instabilities is the variation of the supply voltage. The supply voltage is assumed to be unstable (because it is an external factor that the designer does not control), and its nominal variation is generally assumed to be ± 5%. Thus, the first objective of all reference sources is the independence of the currents or voltages generated with respect to V_{dd}. This is done using *self-biased* circuits exploiting the non-linearity of passive dipoles.

Consider two dipoles shown in fig. 1. The dipole (a) is a PN junction diode, named Q1. The dipole (b) is a PN junction diode, named Q2 with an area K times larger than that of Q1 and a resistor R_1 connected in series. As a reminder, diodes have current-voltage characteristics as follows :

$$I = I_S\left(1+\exp\left(\frac{V}{nV_T}\right)\right) \approx I_S \exp\left(\frac{V}{nV_T}\right), \quad (1)$$

for the diode Q1 when $V > 0.1V$

and

$$I = KI_S\left(1+\exp\left(\frac{V}{nV_T}\right)\right) \approx KI_S \exp\left(\frac{V}{nV_T}\right) \quad (2)$$

for the diode Q2 when $V > 0.1V$

Here I_S is the reverse current of the diode which is proportional to the area of the diode, V_T is the thermal voltage $V_T = kT/q$, n is a technological parameter which is worth 1 for the technology used.

For the following developments it is useful to invert these equations :

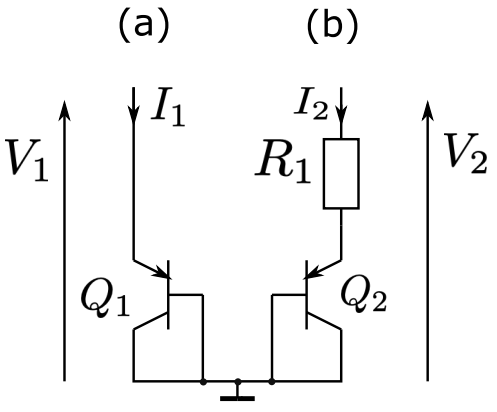

FIGURE 1 – Les circuits non-linéaires utilisés pour générer une polarisation indépendante de VDD et une source de courant PTAT.

$V \approx nV_T \ln\left(\frac{I}{I_S}\right)$ for the diode Q1

and

$V \approx nV_T \ln\left(\frac{I}{KI_S}\right)$ for the diode Q2

The current-voltage characteristics of the dipoles (a) et (b) read :

(a) :

$$V_1 = nV_T \ln\left(\frac{I_1}{I_S}\right) \quad (3)$$

which is the same characteristics as that of the diode Q1

(b) :

$$V_2 = nV_T \ln\left(\frac{I_2}{KI_S}\right) + R_1 I_2 \quad (4)$$

They are monotonic functions on the (I,V) plane which have two crossing points :

— (0,0) which is a trivial point and corresponds to two short-circuited dipoles [1]

— a point (I_0, V_0) that can be calculated by equalizing the last two equations :

$$nV_T \ln\left(\frac{I_0}{I_S}\right) = nV_T \ln\left(\frac{I_0}{KI_S}\right) + R_1 I_2 \quad (5)$$

We can deduce :

$$I_0 = \frac{nk_B T \ln\ K}{qR_1} \quad (6)$$

At ambient temperature $T = 300$, with $K = 8$ and $R = 50k\Omega$, we have $I_0 = 1.07\mu$A. We see that the current I_0 is inversely proportional to R_1. The voltage V_0 is always close to the "conventional" diode threshold value, 0.6-0.7V.

What is interesting here is the fact that the current I_0 does not depend on any parameter other than those participating in the formula (6), and therefore independent of the supply voltage. The current is proportional to the absolute temperature : this circuit is therefore a linear thermal sensor called a PTAT (Proportional To Absolute Temperature).

1.2 Practical PTAT generator

1.2.1 An active short-circuit

In the previous example, the point (I_0, V_0) is found "by hand". We would like to create a circuit that would do this automatically. To do this, we need a two-port block (a quadrupole), which imposes on its two ports an equality between the voltages and the currents. Such a block can be seen

1. To obtain this operating point, it is not necessary to use the precise formula for the current of the diodes, because the approximation used in (1)and (2) is only valid for diode voltages greater than 0.1V

as a quadripole receiving a current at one of the terminals and receiving a voltage on the other (it is impossible to apply an arbitrary voltage and current to the same port). Such a function is represented in fig. 2. It can be called "an active short circuit".

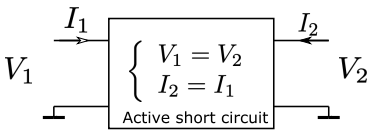

FIGURE 2 – Functional view of the active short circuit.

We note that this block is not a simple (passive) short-circuit implemented with a wire. Indeed, a short circuit is included in the functionality of this block, but this one is much wider. This is illustrated in fig. 3 : if a short circuit is used to achieve this block (a), the only possible (I, V) point for our problem is (0,0) . However, the point (I_0, V_0) is also possible if the more generic block fig. 3 is used, cf. fig. 3b. This block must then be active because it generates on the two dipoles the power equal to $2I_0V_0$, whereas a simple short-circuit (a wire) does not generate any power.

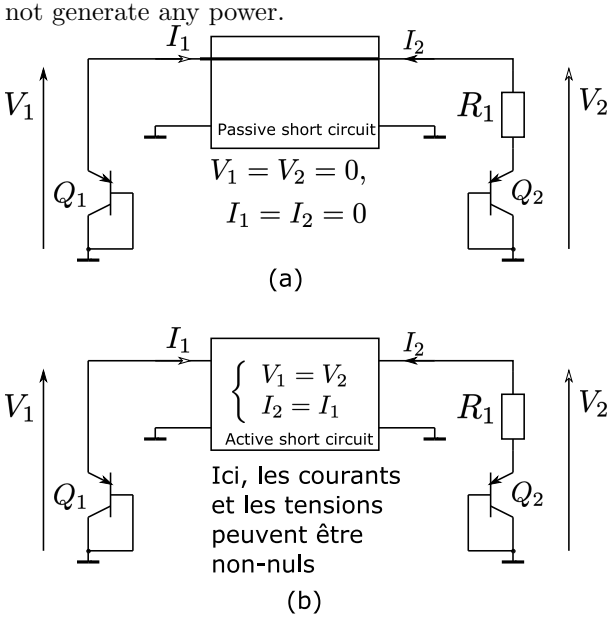

FIGURE 3 – Equalization of voltages and currents in the circuits of fig. 1 with a passive short-circuit (a) and with an active short-circuit (b).

How can we model this generalized block ? Circuit theory provides us with two abstract dipoles called "pathological" : a nullator and a norator.

A nullator is a dipole whose current and voltage are always zero simultaneously, cf. fig. 4. A norator is a dipole whose current and voltage can have any values independently of each other, cf. fig. 4.

It is obvious that neither of these two dipoles can be physically realized in an isolated way, even under ideal hypotheses. On the other hand, the association of a nullator and a norator effectively models a number of functions, such as the ideal operational amplifier. Thus, the quadrupole of fig. 2 is realized as given in fig. 5. We can see that $V_1 = V_2$ because the two nullators impose the equality of the voltages. And thanks to the presence of two norators, $I_1 = I_2$. This active short circuit imposes on the two dipoles one of the operating points corresponding to the equality of the current-voltage couple, $(0, 0)$ or (I_0, V_0). Which one ? It will depend on the initial conditions, as is the case for any bi- or multistable system.

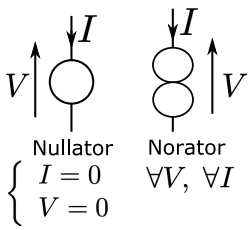

FIGURE 4 – The so-called "pathalogical" dipoles, nullator and norator.

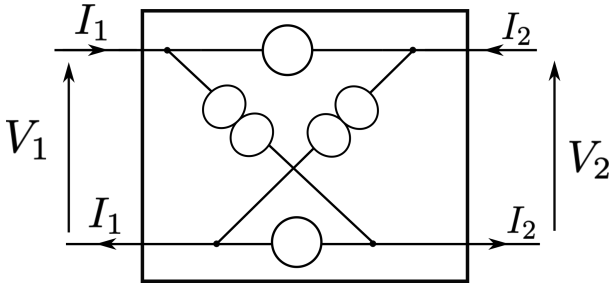

FIGURE 5 – Active short circuit modeled with pathological dipoles.

In the following paragraph we present a practical implementation of this device.

1.2.2 Quad CMOS based PTAT

The diagram of a possible realization of PTAT is given in fig. 6.

The supply voltage is 3.3V.

It uses a realization of "active short-circuit" composed of 4 MOS transistors. A unity-gain pMOS current mirror ensures equal currents in both branches. The (identical) currents of the two arms of the mirror pass through two identical nMOS transistors and generate identical voltages V_{gs}. Since the gates of the nMOS transistors are connected, the sources of the transistors (nodes 1 and 2) are the same.

2 Small-signal model

A CMOS active short circuit can be formally characterised as given in fig. 7. The input quantities are

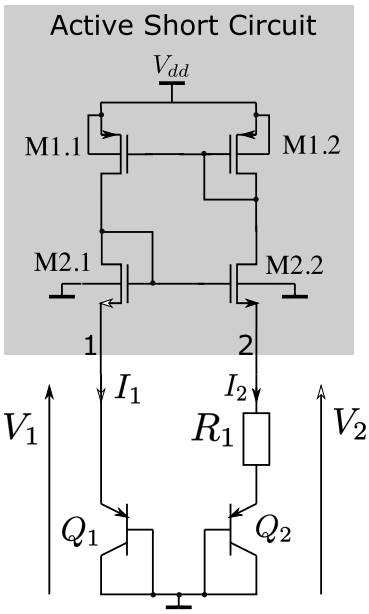

FIGURE 6 – A CMOS PTAT current generator

V_1 and I_2.

If the transistors are modeled as a transconductance (g_m) and output conductance (g_o), the active short-circuit is equivalent to a two-port linear model which can be expressed under the form of hybrid reverse parameters equation :

$$\begin{pmatrix} I_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \begin{pmatrix} V_1 \\ I_2 \end{pmatrix} \tag{7}$$

where

$$\begin{aligned} g_{11} &= -\frac{g_{op}g_{mn}}{g_{mn}+g_{op}} \\ g_{12} &= \frac{g_{mn}}{g_{mn}+g_{op}} \\ g_{21} &= \frac{g_{g^2_{mn}}}{(g_{mn}+g_{on})(g_{mn}+g_{op})} \\ g_{22} &= -\frac{g_{mn}g_{op}+g_{mp}g_{op}+g_{on}g_{op}}{g_{mp}(g_{mn}+g_{on})(g_{mn}+g_{op})} \end{aligned} \tag{8}$$

In the ideal case, when $g_o >> g_m$ and all transistors are identical, we have :

$$\begin{aligned} g_{11} &= -g_o \\ g_{12} &= 1 \\ g_{21} &= 1 \\ g_{22} &= -\frac{g_o}{g_m^2} \end{aligned} \tag{9}$$

Note that g_{12} is the input conductance and g_{21} is the output resistance.

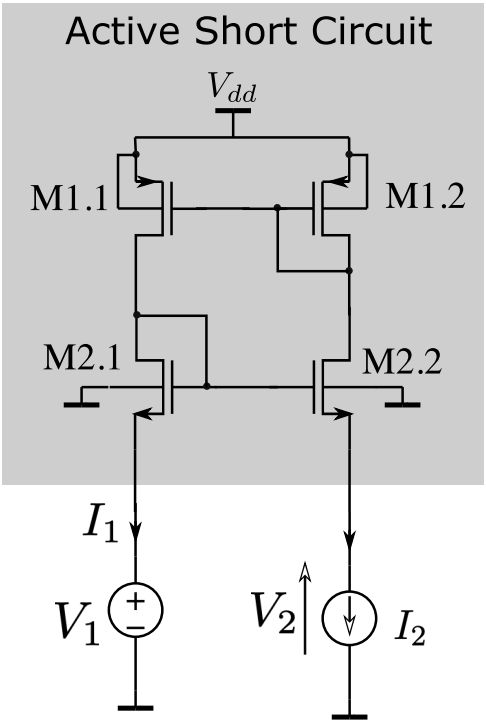

FIGURE 7 – Test of the circuit making an active short-circuit.

3 Bandgap generator

The PTAT circuit generates a current independent of the voltage of the power supply, but very sensitive to temperature. The 8 circuit, called "bandgap generator", uses the PTAT circuit to generate a stable output voltage V_{ref} close to 1.2 V, which happens to be the forbidden energy band width of the silicon (*bandgap*), hence the name.

3.1 Realization of the bandgap circuit

The voltage V_{ref} is obtained as follows. We copy the current I_2 which is a PTAT, using a current mirror M1.2-M1.3. This current is passed through a dipole composed of a resistor $R_2 = \gamma \cdot R_1$ and a diode Q3 identical to diode Q2. We have :

$$V_{ref} = I_2R_2 + V_{Q3} = I_2\gamma R_1 + V_{Q3} = \tag{10}$$

$$\frac{nk_B\gamma \ln\ K}{q}T + V_{Q3} \tag{11}$$

We now study the sensitivity of V_{ref} with respect to temperature :

$$\frac{\partial V_{ref}}{\partial T} = \frac{nk_B\gamma \ln\ K}{q} + \frac{\partial V_{Q3}}{\partial T} \tag{12}$$

We know that the diode voltage has a negative thermal coefficient $\frac{\partial V_{Q3}}{\partial T}$, while the first term of the

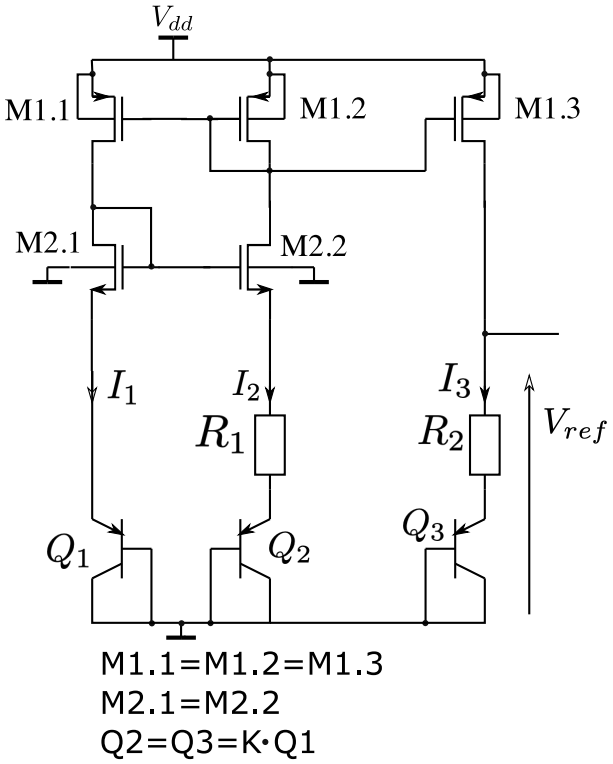

FIGURE 8 – Stable "bandgap" voltage generator.

equation (11) is proportional to the temperature with a positive factor. We can then choose γ (the ratio between resistances R_1 and R_2) so as to minimize the thermal variation of V_{ref} for a given temperature range. From experience, the good value for γ in the AMS 350nm technology is between 12 and 13.

4 Sizing procedure of a bandgap CMOS

The sizing of the bandgap may be done according to the following procedure.

Input parameters which should be provided by the designer :

- The bias current. This parameter defines the power consumption.
- The transisor length. The transistor length defines the precision of the bandgap which increases with L, at the price of the transistor size.
- The minimum supply voltage $V_{dd\,min}$ at which the bandgap needs to work.

These three parameters are strongly linked. The minimum supply voltage defines the nominal V_{gs} voltages of the transistors which, together with the bias current, define the transistor W/L and since L is fixed, W is directly deduced.

- Sizing of the bandgap dipole generator with the formula (6). The factor K is fixed by the designer, however, the usual value is 8 since it allows one to use a common centroid layout for the transistors Q_1 and Q_2 matching. The other parameters of the formula (6) comes from the technology (n) and the nominal absolute temperature T. This stage gives the value of R_1
- Sizing of the transistors of the active short circuit.
  - Calculation of the nominal V_{gs} voltage as $V_{dd\,min}/2$
  - For each transistor, the following parameters are known : $I_{ds} = I_0$, L, $V_{gs} = V_{dd\,min}/2$, $V_{ds} = V_{gs}$. These parameters define uniquely W with use of a CMOS calculator (see sec. 5).
- Calculation of the factor γ between R_1 and R_2. This factor is calculated so to minimize the V_{ref} variation over the working temperature range $[T_{min}, T_{max}]$. The advised solution is to use a numeric optimisation method which performs

$$\gamma_{opt} = \arg \min_{\gamma} (\max_{T} V_{ref} - \min_{T} V_{ref}) \quad (13)$$

5 Appendix : the CMOS calculator

The technique used for calculation of the technology component parameters (transistors, resistors, diodes...) is based on the simulation with the models proposed in the design kit.

Three python functions are based on the simulator :

- A function providing the transistors current I_{ds} when all voltages and dimensions of the transistor are given
- A function providing the voltage of a diode biased by a given current.
- A function simulating the bandgap circuit with given parameters and providing the V_{ref} voltage.

Other functions (not used in the bandgap generator script) include the calculation of the small-signal parameters of the transistors

These functions consist in running a simulator with a pre-written netlist including a single device (the MOS transistor or the diode) or a circuit (the full bandgap circuit) and the test sources. The parameters of the devices are copied from the function arguments to a parameter file which is used by the simulator.

The function calculating I_{ds} for a MOS transistor can be inverted in a python script. Typically, when the current is known and the transistor width W needs to be found, the equation $I_{ds}(W) = I_{ds\,0}$ needs to be solved numerically. However, other needs may exist, for instance, a selection of a L for a desired intrinsic gain g_m/g_{ds}.

We let the implementation of the function inversion for the designer. We believe that it is difficult to propose a generic function with a guaranty of robust operation in any technology node and which is adapted to all needs. In the bandgap generator, the implementation of the function $W(I_{ds})$ is done by finding the fixed point of the discrete map

$$W_{n+1} = W_n \frac{I_{ds\ 0}}{I_{ds\ n}(W_n)} \tag{14}$$

with $W_0 = W_{min}$, the minimum value of the technology. This method is based on the fact that I_{ds} is roughly proportional to W.

Since in many design kits there is a limit on the single transistor width (finger width), it is necessary to break the transistor in several devices for large W. This supposes a check in the iterations of (14) and an introduction of the parameters W_f, M (finger width, number of devices) so that the actual width is $W = M \cdot W_f$.

This paper was made possible with a donation from NGI Zero PET. NGI Zero PET is part of the European Commission's Next Generation Internet programme, established under the aegis of Directorate-General Communications Networks, Content and Technology (DG CNECT)

For more information, see: **https://nlnet.nl/PET** or **https://ngi.eu**

Horizon 2020 research and innovation grant agreement No 825310.