8

Analog Circuits in Weak Inversion

by Eric A. Vittoz

8.1 Introduction

As explained in Chapter 5, the characteristics of a transistor in weak inversion are very different from those in strong inversion. Weak inversion has some special properties that can be exploited, in particular for designing low-power and/or low-voltage analog circuits. But weak inversion also has some drawbacks, the most important being the poor current matching (see Section 5.8.1), a maximum noise content of the drain current (see Section 5.6.4), and of course the low speed. Therefore, and except for very low supply voltages (below 0.5 V), the best performance is obtained by combining transistors biased in weak and in strong inversion.

It must be remembered that the amount of inversion is controlled by the inversion coefficient IC defined in Section 5.3.4. In principle, any value of IC can be obtained for any value of saturation current by adjusting the specific current I_{spec} (5.32) through W/L. The only limits are the width W of the transistor (and the associated drain-substrate leakage) for I_{spec} very large, and its length L (and the associated leakage *underneath the channel*) for I_{spec} very small.

This chapter will explore where and how the various favorable features of weak inversion can be exploited in designing analog circuits.

8.2 Minimum Saturation Voltage

8.2.1 Current Mirrors

As illustrated by Figure 5.8, weak inversion provides the minimum possible saturation voltage. The most immediate application is that of low-voltage current mirrors. The output current becomes independent of the output voltage as soon as this voltage is larger than 4 to $6U_T$. It is thus possible, in principle, to build voltage amplifiers for a supply voltage as low as 8 to $12U_T$, provided the threshold voltage is sufficiently low.

The price to pay for this low voltage drop across the mirror is a low speed, a maximum spectral density for both channel noise and interface noise, and the worst matching of currents.

The only known possibility to drastically improve the matching and virtually eliminate the flicker noise, while keeping the low minimum output voltage, is to build the mirror with bipolar transistors. Those might be MOS transistors operated in the lateral bipolar mode, that are available in any CMOS process [159].

8.2.2 Cascode Mirrors

The residual output conductance of a mirror can be drastically reduced by cascoding it by means of a common-gate transistor, as illustrated in Figure 8.1.

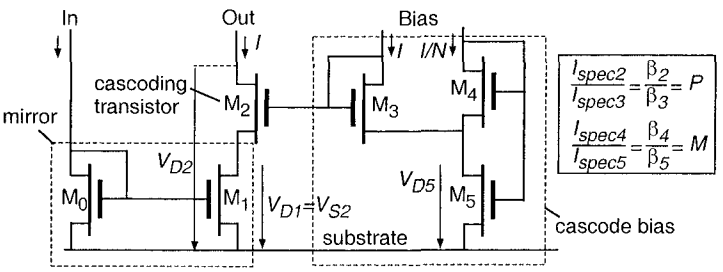

Fig. 8.1. Cascode mirror in weak or moderate inversion.

The common-gate transistor M_2 must be biased so as to ensure the saturation of the mirror transistor M_1. This is the purpose of the sub-circuit M_3-M_4-M_5 (cascode bias), itself biased by currents equal or proportional to the output current I of the mirror.

If transistor M_1 is in *weak inversion*, it is saturated for $V_{D1} = 4$ to $6U_T$ *proportional to the absolute temperature* T (PTAT). If all the other transistors are also in weak inversion, then equation (5.40) applied to each transistor gives [110, 106]

$$V_{D5} = U_T \ln [1 + (N+1)M], \tag{8.1}$$

$$V_{D1} = V_{D5} + U_T \ln P = U_T \ln [P(1 + (N+1)M)]. \tag{8.2}$$

The product MN should not be too large, to prevent the drain junction leakage of a wide transistor M_4 from dominating a very small bias current I/N.

Now, even for a large V_{D1}, the output conductance of M_1 cannot be lower than G_{ds}, the residual conductance due to channel shortening given by (5.50). We can therefore assume that saturation is obtained when the residual drain transconductance due to the remaining reverse current I_R is smaller than G_{ds}. Hence, according to (5.44):

$$G_{md1} = \frac{I_{R1}}{U_T} < G_{ds1} = \frac{I_{F1}}{V_{M1}} \quad \text{or} \quad \frac{I_{R1}}{I_{F1}} < \frac{U_T}{V_{M1}}. \tag{8.3}$$

By using the expression (5.39) of I_F and I_R, this gives

$$V_{DS1} > U_T \ln \frac{V_{M1}}{U_T}, \tag{8.4}$$

or by using (8.2) where $V_{S1} = 0$:

$$P(1 + M(N+1)) > V_{M1}/U_T. \tag{8.5}$$

For $V_{M1}/U_T < 200$, this condition can be fulfilled by choosing $N = 2$ and $P = M = 8$, which gives $V_{D1} = 5.3U_T$.

The circuit of Figure 8.1 can also be used for *moderate inversion*, provided a different approach is used for sizing the devices. This approach is based on the fact that the saturation of M_1 is ensured if $I_{R1} \ll I_{F1}$, *independently* of the inversion coefficient IC [160, 109]. For this purpose, we choose $M_5 \equiv M_1$ (two identical transistors) and $M_3 \equiv M_2$ ($P = 1$). Furthermore, $N \gg 1$ (which is possible in spite of the drain junction leakage, since the transistors are not in deep weak inversion).

Now M_2 and M_3 are both saturated with the same current density. Since they also have the same gate voltage they have the same source voltage, thus $V_{D1} = V_{D5}$. Transistors M_1 and M_5 are not fully saturated, but since $N \gg 1$, they have the same drain voltage and the same current density. Therefore:

$$I_{F1}/I_{R1} = I_{F5}/I_{R5}, \tag{8.6}$$

and the two transistors are in the same state of saturation.

The forward current I_{F4} of M_4 and the reverse current I_{R5} of M_5 are controlled by the same voltage V_{D5}. Since M_4 is M-times wider than M_5, then

$$MI_{R5} = I_{F4} = I/N \quad \text{whereas} \quad I_{F5} = I \tag{8.7}$$

since $I_{R5} \ll I_{F5}$. Thus

$$\frac{I_{F1}}{I_{R1}} = \frac{I_{F5}}{I_{R5}} = MN. \tag{8.8}$$

This ratio could be further increased by choosing $P > 1$, but the effect would then depend on the inversion coefficient [160].

If enough voltage is available, it is interesting to use this circuit with M_1 (and therefore M_0 and M_5) in strong inversion to improve current matching, but M_2 (and therefore M_3) in weak inversion to minimize V_{DS2} required for saturation.

Recalculating the condition (8.3) for $G_{md} < G_{ds}$ with the expression of transconductance (5.43) valid at all levels of current gives

$$\frac{I_{R1}}{I_{F1}} < \frac{U_T}{V_{M1}} \left(1 + \frac{U_T}{V_{M1}} IC_1\right), \tag{8.9}$$

which only departs from (8.3) for a very large inversion coefficient IC_1.

A simpler bias circuit can be used if both M_1 and M_2 are in strong inversion [161].

8.2.3 Low-Voltage Amplifiers

Thanks to the minimum saturation voltage, MOS transistors in weak inversion can provide some voltage gain A_v even at a very low supply voltage V_B. The maximum gain is achieved by the CMOS inverter as illustrated in Figure 8.2.

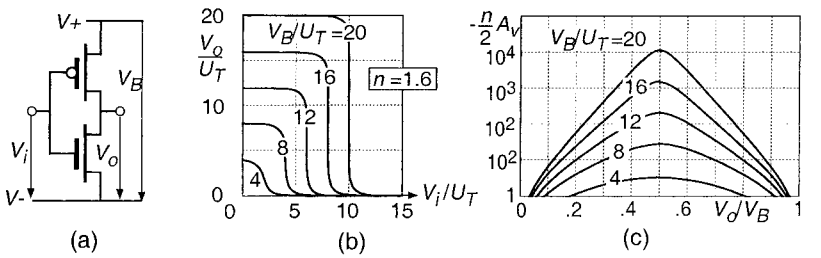

Fig. 8.2. CMOS inverter-amplifier: (a) circuit and definitions; (b) transfer characteristics $V_o(V_i)$; (c) gain as a function of the output voltage V_o.

Assuming that the two complementary transistors have the same values of n and $V_{T0} + nV_S$ in the equation (5.40) of the drain current, the unloaded transfer characteristics can be obtained by equating the current of the two transistors [53]. This gives

$$v_i = \frac{v_B}{2} + \frac{n}{2} \ln \frac{1 - \mathrm{e}^{v_o - v_B}}{1 - \mathrm{e}^{-v_o}}, \tag{8.10}$$

where v_B, v_i and v_o are the normalized supply, input and output voltages according to

$$\frac{V_B}{v_B} = \frac{V_i}{v_i} = \frac{V_o}{v_o} = U_T. \tag{8.11}$$

These characteristics are plotted in Figure 8.2(b) for several values of v_B.

The voltage gain A_v can then be obtained by differentiating (8.10), which yields

$$A_v = \frac{\mathrm{d}v_o}{\mathrm{d}v_i} = -\frac{2}{n} \cdot \frac{\mathrm{e}^{v_o - v_B} + \mathrm{e}^{-v_o} - \mathrm{e}^{-v_B} - 1}{2\mathrm{e}^{-v_B} - \mathrm{e}^{v_o - v_B} - \mathrm{e}^{-v_o}}. \tag{8.12}$$

It is plotted in Figure 8.2(c) for several values of v_B. As can be seen, a voltage gain larger than 100 can be obtained with a supply voltage of $12U_T \cong 300\mathrm{mV}$. For this symmetrical circuit, the gain reaches a maximum value for $V_i = V_o = V_B/2$ given by

$$|A_v|_{max} = (\mathrm{e}^{v_B/2} - 1)/n. \tag{8.13}$$

The residual output conductance in saturation (G_{ds} due to channel shortening or G_{mdsat} due to DIBL) has been omitted in this calculation and would put another limit on very high values of A_v.

By combining such a CMOS inverter with an adequate biasing circuit [40], it can be used as an analog amplifier.

In spite of the relatively low speed associated with weak inversion, transistors in RF front ends of integrated receivers can be biased close to weak inversion for low-voltage operation thanks to the availability of short channel lengths in submicron processes [162].

For digital applications, a supply voltage of $4U_T \cong 100\mathrm{mV}$ is sufficient to provide the necessary nonlinear transfer characteristics, as shown by Figure 8.2. Of course, the maximum current should be adjusted to obtain the required speed. This can be done, in principle, by choosing a very low value of V_{T0} and by controlling the current by the source voltage V_S [163, 53]. This should be done separately for P- and N-channel transistors in order to achieve the expected symmetrical characteristics.

8.3 Maximum Transconductance-to-Current Ratio

8.3.1 Differential Pair

As a consequence of the maximum value of G_m/I_D obtained in weak inversion, the difference of gate voltages required to compensate the mismatch of two transistors is minimum (see Figure 5.14). Moreover, the spectral density of input referred noise voltage is also minimum for a given value of drain current (see Section 5.6.4). These features can be exploited to minimize the offset and the noise of differential pairs, with the additional advantage of minimum saturation voltage.

It should be noticed that, as explained at the end of Section 5.8.1, the source voltage should be as small as possible in order eliminate the effect of Δn (mismatch of slope factors). This means that, for minimizing the input voltage offset, the two transistors should be put in a separate well connected to their sources, as illustrated in dotted line in Figure 8.3(a). In single-well processes, this is unfortunately only possible for one type of transistors (P-channel for N-well process).

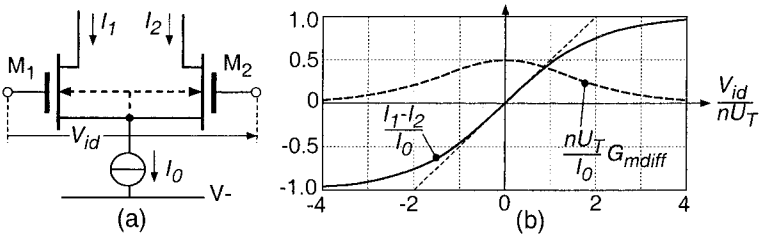

Fig. 8.3. Differential pair; (a) circuit; (b) transfer characteristics and transconductance.

The transfer characteristics of the saturated differential pair in weak inversion can be calculated by using equation (5.40) of the drain current, giving

$$I_1 = \frac{I_0}{1 + \exp\frac{-V_{id}}{nU_T}} \quad \text{and} \quad I_2 = \frac{I_0}{1 + \exp\frac{+V_{id}}{nU_T}}, \tag{8.14}$$

or for the difference of output currents

$$I_1 - I_2 = I_0 \tanh\frac{V_{id}}{2nU_T}, \tag{8.15}$$

which is represented in Figure 8.3(b). As can be seen, a major drawback of a differential pair in weak inversion is its limited input range of linearity. This is best illustrated by the differential transconductance

$$G_{mdiff} = \frac{\mathrm{d}(I_1 - I_2)}{\mathrm{d}V_{id}} = \frac{I_0}{2nU_T} \cdot \left(\cosh\frac{V_{id}}{2nU_T}\right)^{-2} \tag{8.16}$$

plotted in the same figure.

The linear range can be extended by using the circuit of Figure 8.4(a) [164], which can be analyzed by means of the concept of pseudo-resistors introduced in Section 5.5, as shown by Figure 8.4(b). The saturated transistors M_1 and M_2 correspond to grounded resistors R_1 and R_2 in the resistor prototype. The linearization transistors M_{1a} and M_{2a} correspond to two resistors of K-times larger values. The difference of output currents can now be calculated by using the resistor prototype, giving

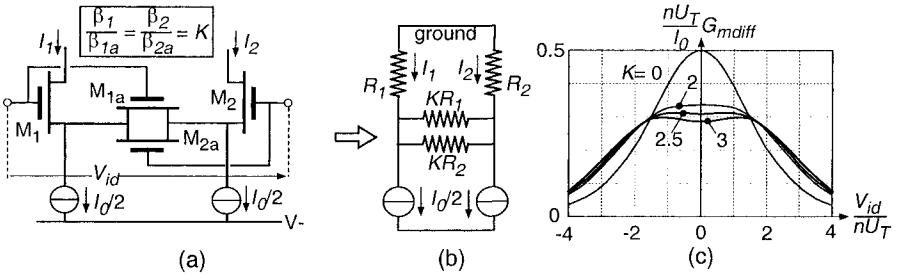

Fig. 8.4. Linearized differential pair; (a) circuit; (b) resistor prototype; (c) differential transconductance.

$$\frac{I_1 - I_2}{I_0} = \frac{R_2 - R_1}{R_2 + R_1 + K \cdot R_2 R_1/(R_2 + R_1)} = \frac{\mathrm{e}^{2x} - 1}{\mathrm{e}^{2x} + (K+2)\mathrm{e}^{x} + 1}, \quad (8.17)$$

where the last part has been obtained by replacing R_2/R_1 by e^x, with $x = V_{id}/(nU_T)$, according to the definition (5.54) of pseudo-resistors in weak inversion. The transconductance is obtained by differentiation of (8.17):

$$G_{mdiff} = \frac{\mathrm{d}(I_1 - I_2)}{\mathrm{d}V_{id}} = \frac{I_0}{nU_T} \cdot \frac{(2+K)\mathrm{e}^{2x} + 4\mathrm{e}^{x} + (2+K)}{\left(\mathrm{e}^{2x} + (2+K)\mathrm{e}^{x} + 1\right)^2}\mathrm{e}^{x}, \quad (8.18)$$

which reduces to (8.16) for $K = 0$. This result is plotted in Figure 8.4(c) for several values of K. As can be seen, the best linearity is obtained for $K \cong 2.5$, at the price of a 40% reduction of transconductance.

Another method for increasing the linear range is the "multi-tanh" technique, which was developed for bipolar transistors [165]. This approach uses the sum of the currents of differential pairs having symmetrical controlled offset voltages, as illustrated in Figure 8.5 for 2 pairs. The specific current

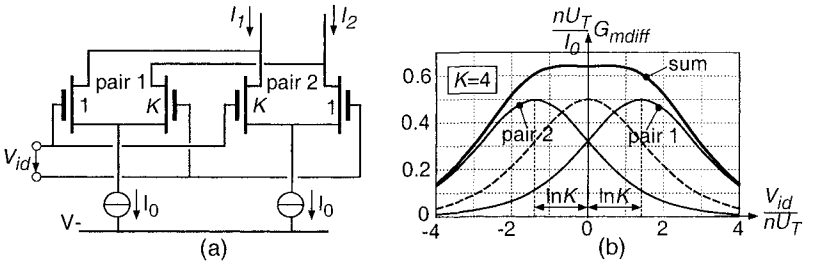

Fig. 8.5. Multi-tanh linearization: (a) circuit; (b) transconductance (single symmetrical circuit in dotted line for comparison).

of the two transistors of each pair are in a ratio K, which produces in weak inversion, according to (5.40), an input offset voltage

$$\Delta V_{id} = nU_T \ln K. \tag{8.19}$$

As shown in Figure 8.5(b), $K = 4$ (which can easily be implemented by 4 unit transistors) is an optimum value. The range of linearity can be further extended by choosing $K = 13$ and adding the output currents of a normal differential pair biased at $0.75I_0$ [165].

8.3.2 Single-Stage Operational Transconductance Amplifiers (OTA)

Another consequence of the maximum value of G_m/I_D obtained in weak inversion is a maximum value V_M/U_T of the intrinsic voltage gain, as was shown in Figure 5.10. If the channel is not too short, this voltage gain can be as high as 60 dB per stage and can be boosted to close to 120 dB by cascoding the current sources. Hence, Operational Transconductance Amplifiers (OTA's) can be implemented in a single cascoded stage, which eliminates the need for compensation.

In weak inversion, the unity gain bandwidth G_m/C_L is maximum for a given load capacitance C_L and a given current, therefore the settling time is minimum. But this is only true for small signals. Indeed, if a large voltage step is applied to the input, the differential pair saturates, hence the settling time is set by the slow slew-rate associated with the small bias current.

To circumvent this problem, the bias current I_0 of the differential pair must be increased whenever its input voltage V_{id} is large. One way to do this is to provide an increment of I_0 proportional to the difference of output currents [42] as illustrated in Figure 8.6(a), where

$$I_0 = I_b + A|I_1 - I_2| \quad \text{with} \quad I_0 = I_1 + I_2. \tag{8.20}$$

If the pair operates in weak inversion, the combination of (8.20) with (8.14)

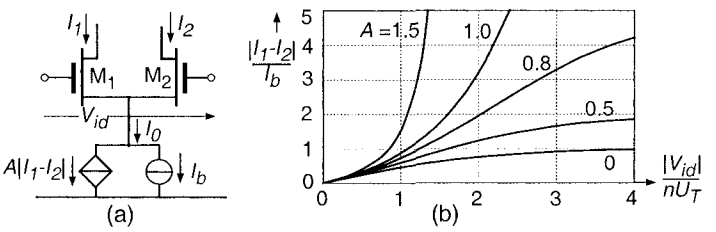

Fig. 8.6. Adaptive biasing of a differential pair: (a) circuit; (b) transfer characteristics for several values of feedback factor A.

gives

$$|I_1 - I_2| = I_b \cdot \frac{\exp\left|\frac{V_{id}}{nU_T}\right| - 1}{(A+1) - (A-1)\exp\left|\frac{V_{id}}{nU_T}\right|}, \tag{8.21}$$

which is plotted in Figure 8.6(b) for several values of A. With $A = 0$, it is the basic differential pair that saturates at $|I_1 - I_2| = I_b$. This saturation current increases for $0 < A < 1$. For $A > 1$, $|I_1 - I_2|$ tends to infinity (i.e. it is no longer limited by the bias current I_b) for a critical value of input voltage given by

$$|V_{idcrit}| = nU_T \cdot \ln \frac{A+1}{A-1}. \tag{8.22}$$

For the particular case $A = 1$:

$$\min(I_1, I_2) = I_b/2, \tag{8.23}$$

one of the two branch currents remains constant at $I_b/2$.

A special application of this adaptive biasing technique is the class AB voltage follower illustrated in Figure 8.7(a). Without the additional bias cur-

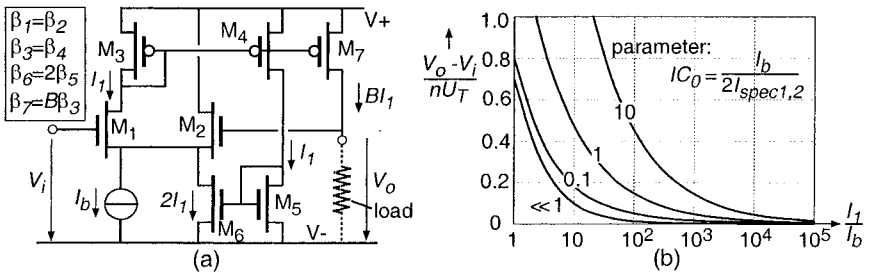

Fig. 8.7. Voltage follower for a resistive load: (a) circuit; (b) input-output offset.

rent $2I_1$ delivered by M_5-M_6, the circuit would be an elementary OTA with unity voltage feedback. It would only operate as a follower if the current delivered to the load were much lower than the maximum output current BI_b.

With the feedback through M_5-M_6, the total bias current of the differential pair M_1-M_2 is

$$I_0 = I_b + 2I_1 \text{ with } I_0 = I_1 + I_2, \text{ hence } I_2 = I_b + I_1. \tag{8.24}$$

The difference between the input voltage and the voltage applied to the load can then be expressed by using (5.35):

$$\frac{V_i - V_o}{nU_T} = \frac{V_{P1} - V_{P2}}{U_T} = \sqrt{1 + 8IC_0 \cdot I_1/I_b} - \sqrt{1 + 8IC_0(1 + I_1/I_b)}$$

$$+ \ln \frac{\sqrt{1 + 8IC_0 \cdot I_1/I_b} - 1}{\sqrt{1 + 8IC_0(1 + I_1/I_b)} - 1} < 0, \tag{8.25}$$

where $IC_0 = I_b/2I_{specI,2}$ is the inversion coefficient of M_1 and M_2 without feedback. These characteristics are represented in Figure 8.7(b) for several

values of IC_0. The output voltage is always *larger* than the input voltage, but the difference becomes very small for $I_1 \gg I_b$. For a given value of I_1/I_b, the difference is minimum if $IC_0 \ll 1$ (OTA without feedback in weak inversion).

A high current efficiency can be obtained by choosing $B \gg 1$ (ratio of mirror M_3-M_7).

As long as the differential pair remains in weak inversion, (8.25) reduces to

$$\frac{V_i - V_o}{nU_T} = -\ln\left(1 + \frac{I_b}{I_1}\right). \tag{8.26}$$

8.4 Exponential Characteristics

8.4.1 Voltage and Current Reference

The exponential dependency of the drain current on V_S/U_T makes it possible to extract a voltage proportional to U_T as shown in Figure 8.8. The basic

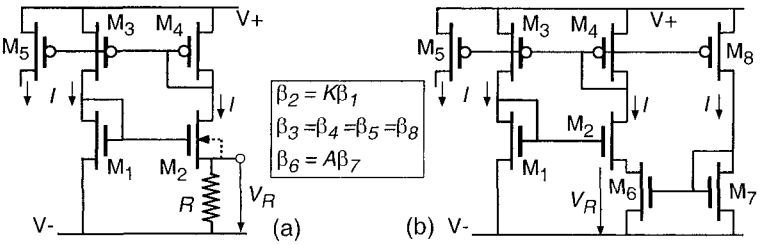

Fig. 8.8. Voltage and current reference: (a) basic circuit; (b) resistor-less current reference.

circuit [31] shown in part (a) of the figure contains a 1-to-K N-channel current mirror M_1-M_2, with the source of M_2 degenerated by a resistor R. A 1-to-1 P-channel current mirror M_3-M_4 (or any equivalent circuit) imposes the same current in the two branches. Therefore, a source voltage V_R builds-up across resistor R to compensate the ratio K of the N-channel mirror. If this mirror is in weak inversion, then

$$V_R = RI = U_T \ln K. \tag{8.27}$$

This voltage should be sufficiently larger than the threshold mismatch of M_1-M_2. In practice, it cannot be made much larger than $4U_T$, which corresponds to $K = 55$. It can be used as a PTAT voltage reference, or as a compensation voltage in a band gap voltage references [32].

A reference current I can be extracted by the additional mirror transistor M_5. Thanks to the small value of V_R a small current can be obtained with a reasonably low value of R.

If the transistor M_2 is in a separate well connected to its source, as shown by the dotted line, then $V_{S2} = V_{S1} = 0$. The factor K is then compensated by a difference of gate voltages, and U_T is multiplied by n in (8.27).

Figure 8.8(b) shows a variant of the basic circuit in which the resistor is replaced by transistor M_6 [166]. This transistor is the output transistor of a current mirror M_7-M_6 of ratio 1 to A, that operates in *strong inversion*, with the reference current itself as its input. If $A \gg 1$, then $I_{F6} = AI_{F7} = AI \gg I$. Hence, far from being saturated, M_6 is biased close to $V_D = V_S = 0$ where it behaves like a resistor of value $R = 1/G_{ms6}$ given by (5.45):

$$1/R = G_{ms6} = \sqrt{2n\beta_6 AI}. \tag{8.28}$$

By introducing this value in (8.27) we obtain, after arranging the result

$$I = 2n\beta_6 U_T^2 \cdot A(\ln K)^2 = I_{spec6} \cdot A(\ln K)^2. \tag{8.29}$$

This current is obtained without using any resistor, and is proportional to the specific current of transistor M_6. Therefore, as noticed at the end of Section 5.7, it becomes almost independent of the temperature if the mobility is proportional to $T^{-\alpha}$ with $\alpha \cong 2$.

In practice, (8.28) is an acceptable approximation for $A > 5$.

It should be mentioned that the loop M_6-M_2-M_4-M_8-M_7 implements a positive feedback. However, the gain of this loop can be shown to be 1/2 at equilibrium.

8.4.2 Amplitude Regulator

The exponential characteristics of transistors in weak inversion are exploited in the amplitude regulator depicted in Figure 8.9(a) [31]. The sinusoidal signal of amplitude V produced by the oscillator enters the regulator through capacitor C_1, and transistor M_5 delivers the output current used to bias the oscillator. When no oscillation is present ($V = 0$), the circuit is reduced to the current

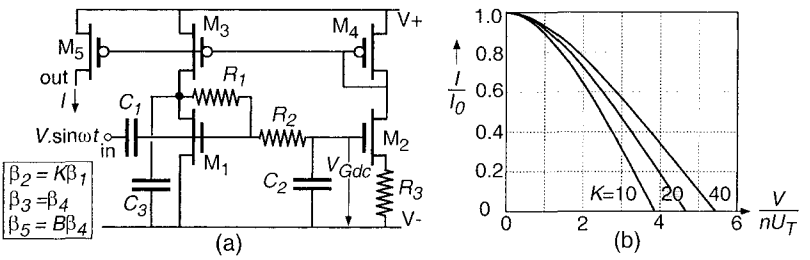

Fig. 8.9. Amplitude regulator for oscillators: (a) circuit; (b) transfer characteristics.

reference of Figure 8.8(a). According to (8.27), it delivers an output current

$$I = I_0 = BU_T \ln K/R_3, \tag{8.30}$$

which serves as the start-up current of the oscillator. As the oscillation voltage grows, it is superimposed on the DC component of gate voltages V_{Gdc} for M_1, *but not* for M_2, since it is blocked by the low-pass filter R_2-C_2. Because of the exponential function $I_D(V_G)$ of M_1, its average drain current should increase, which is not compatible with the 1-to-1 ratio imposed by the mirror M_4-M_3. Instead, V_{Gdc} decreases, resulting in a decrease of the output current I.

Assuming that all transistors remain saturated, that M_1 remains in weak inversion even during the peaks of its drain current, and that the residual oscillation amplitude at the gate of M_2 is much smaller than U_T (so that it has no nonlinear effect), the transfer characteristics are given by [31]

$$I = I_0 \cdot \left(1 - \frac{\ln \mathrm{I_{B0}}(\frac{V}{nU_T})}{\ln K}\right), \tag{8.31}$$

where $\mathrm{I_{B0}}$ is the 0-order modified Bessel function. They are plotted in Figure 8.9(b) for several values of K.

The amplitude of oscillation will stabilize when the regulator delivers exactly the bias current I required to produce the amplitude V.

If the peak drain current of M_1 leaves weak inversion, the transfer characteristics will be modified and may eventually loose their monotonicity, which must absolutely be avoided to maintain stable oscillation. A semi-empirical condition to ensure monotonicity is

$$\beta_1 > \frac{2\beta_3/\beta_4}{nU_T R_3}. \tag{8.32}$$

The role of capacitor C_3 is to keep the drain voltage of M_1 sufficiently constant to avoid de-saturation during the positive peaks of current.

At low frequencies, high (non-critical) values may be needed for resistors R_1 and R_2. Very high values have been obtained by using lateral diodes in the polysilicon layer [167, 37]. Lower values can be obtained by means of transistors adequately biased [46].

This amplitude regulator has been applied extensively to quartz oscillators in watches [37, 46, 168, 169, 170], but it can be used in different type of sinusoidal oscillators as well [38].

8.4.3 Translinear Circuits

Discovered for bipolar transistors, the translinear principle [171] is an outstanding application of the exponential characteristics of MOS transistors in weak inversion. Consider the loops of saturated transistors illustrated in Figure 8.10. They include an even number of transistors, half of which have their gate to source "junction" in the clockwise (cw) direction, the other half in the counter-clockwise (ccw) direction. Hence, for the whole loop:

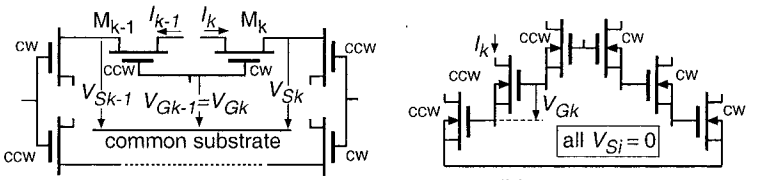

Fig. 8.10. Translinear loops: (a) alternated orientation of transistors in a common substrate; (b) non-alternated orientation: separate local substrates are necessary.

$$\sum_{cw}(V_{Gi} - V_{Si}) = \sum_{ccw}(V_{Gi} - V_{Si}). \tag{8.33}$$

If all transistors are in weak inversion, with negligible reverse current (saturated), then according to (5.41):

$$I_i = I_{D0i} \exp \frac{V_{Gi}/n_i - V_{si}}{U_T} \quad \text{or} \quad \frac{V_{Gi}}{n_i} - V_{Si} = U_T \ln \frac{I_i}{I_{D0i}}. \tag{8.34}$$

Now if cw and ccw transistors are *alternated* [172] as in Figure 8.10(a), then each gate voltage V_{Gi} is common to a pair cw-ccw of transistors. It appears therefore in both sides of equation (8.33), which can thus be rewritten as

$$\sum_{cw}(\frac{V_{Gi}}{n_i} - V_{Si}) = \sum_{ccw}(\frac{V_{Gi}}{n_i} - V_{Si}). \tag{8.35}$$

We can now introduce (8.34), divide by U_T (that is common to all transistors) and exponentiate both sides of the equation, which yields

$$\prod_{cw} \frac{I_i}{I_{D0i}} = \prod_{ccw} \frac{I_i}{I_{D0i}} \quad \text{or} \quad \frac{\prod_{cw} I_i}{\prod_{ccw} I_i} = \frac{\prod_{cw} I_{D0i}}{\prod_{ccw} I_{D0i}} = \lambda. \tag{8.36}$$

This result is independent of the temperature. If I_{D0} is the same for all transistors, then $\lambda = 1$. A circuit may include several loops sharing some transistors, each loop characterized by its value of λ. The mismatch of I_{D0i} is dominated by that of V_{T0i} and results in an error in the value of λ, with a standard deviation

$$\frac{\sigma(\Delta\lambda)}{\lambda} = \frac{1}{nU_T} \left[\sum \frac{1}{2}\sigma^2(\Delta V_{T0i})\right]^{1/2}, \tag{8.37}$$

where the factor 1/2 comes from the fact that $\sigma(\Delta V_{T0})$ is defined for a pair of transistors.

The current-mode multiplier/divider [173] shown in Figure 8.11(a) is a simple example of a single translinear loop with identical transistors ($\lambda = 1$) of alternated orientations.

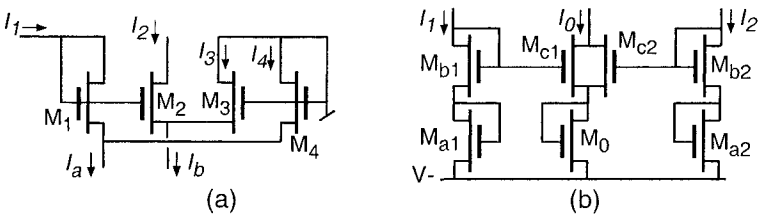

Fig. 8.11. Example of translinear circuits: (a) multiplier/divider; (b) vector length calculation.

The input currents are I_1, I_a and I_b whereas I_2 is the output. Using (8.36), we can write

$$I_1 I_3 = I_2 I_4 \quad \text{hence} \quad I_1(I_b - I_2) = I_2(I_a - I_1), \tag{8.38}$$

which gives after simplification

$$I_2 = I_1 I_b / I_a. \tag{8.39}$$

This result is valid as long as $I_1 < I_a$.

If cw and ccw transistors are *not* alternated, as in the loop example of Figure 8.10(b), then (8.33) cannot be rewritten as (8.35). The only way to make (8.34) compatible with (8.33) is to impose $V_{si} = 0$ by putting each transistor in a separate well connected to its source. Although each n_i is slightly dependent on the particular gate voltage, it can be approximated by a constant n multiplying U_T in (8.34).

In addition to requiring separate wells, non-alternated loops need a higher supply voltage, since they include stacks of at least two gate-to-source voltages.

An interesting example of a loop where transistors cannot be alternated is the circuit of Figure 8.11(b) that calculates the length of a vector in a N-dimensional space [174].

For this example with $N = 2$, the circuit contains two loops that share the transistor M_0. The corresponding current equations are

$$I_1^2 = I_{c1} I_0 \text{ and } I_2^2 = I_{c2} I_0, \text{ thus } I_0 = I_{c1} + I_{c2} = (I_1^2 + I_2^2)/I_0. \tag{8.40}$$

Hence, finally:

$$I_0 = \sqrt{I_1^2 + I_2^2}. \tag{8.41}$$

N loops are needed for N dimensions, each loop made of transistors M_{ai}, M_{bi} and M_{ci}, and all loops sharing M_0.

Although the basic translinear principle assumes that all the loop transistors are saturated, it is possible to include non-saturated transistors [175]. Consider M_k and M_{k-1} in the basic loop of Figure 8.10(a). They have the same gate voltage. Therefore, if they are connected in parallel, they represent the forward and reverse currents of a non-saturated transistor. This may help reducing the supply voltage needed by translinear circuits [176].

8.4.4 Log-Domain Filters [177, 178, 179, 180]

Consider the elementary circuit of Figure 8.12(a), where the transistor is assumed to be in weak inversion and saturated.

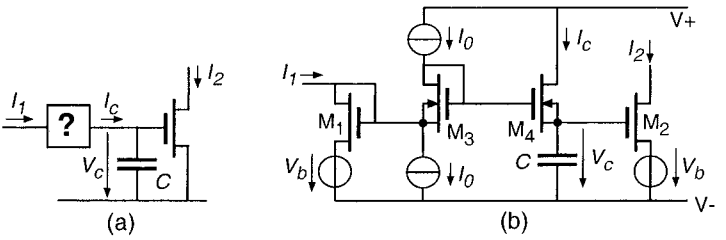

Fig. 8.12. Log-domain filters: (a) principle ; (b) direct implementation of the integrator (partial).

How should the input current I_1 be transformed into the capacitor current I_c in order to obtain a linear current-mode integrator, with

$$I_2 = \frac{1}{\tau}\int I_1 \mathrm{d}t, \quad \text{or} \quad I_1 = \tau\frac{\mathrm{d}I_2}{\mathrm{d}t}, \tag{8.42}$$

in spite of the exponential relationship $I_2(V_c)$? According to (5.41):

$$I_2 = I_{D0}\exp\frac{V_c}{nU_T}, \quad \text{hence} \quad V_c = nU_T \ln\frac{I_2}{I_{D0}}. \tag{8.43}$$

The current in the capacitor can therefore be expressed as

$$I_c = C\frac{\mathrm{d}V_c}{\mathrm{d}t} = \frac{CnU_T}{I_2}\cdot\frac{\mathrm{d}I_2}{\mathrm{d}t} = \frac{CnU_T}{I_2}\cdot\frac{I_1}{\tau}, \tag{8.44}$$

where the last term is obtained by introducing (8.42). Thus, the circuit will behave as a linear integrator according to (8.42) if

$$I_c I_2 = I_0 I_1 \quad \text{with constant} \quad I_0 = CnU_T/\tau. \tag{8.45}$$

This relation could be implemented by the translinear loop illustrated in Figure 8.12(b) [181]. However, in this direct implementation the transistors are not alternated, and two of them must be put in a separate well connected to their source. Notice that it is anyway only a partial implementation, since no current is available to discharge the capacitor.

A variant [182] with alternated transistors is depicted in Figure 8.13. The current I_c flowing through M_3 is mirrored into C by M_5-M_6. The bias voltage V_b provides the necessary headroom for the controlled current source M_7.

Here again, the circuit shown in full line can only charge the capacitor C. One way to circumvent the problem is to add the current source I_d sketched

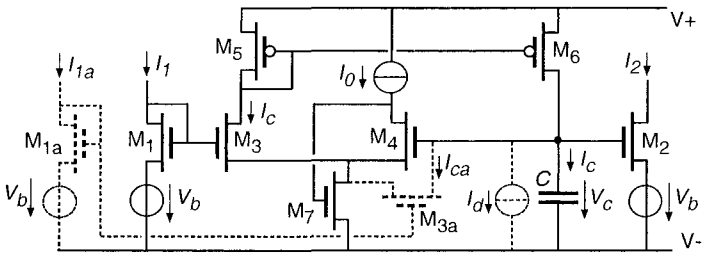

Fig. 8.13. Log-domain integrator implemented by an alternated translinear loop.

in dotted line. The current through M_3 is then $I_c + I_d$ and the loop equation becomes

$$I_1 I_0 = (I_c + I_d) I_2 \quad \text{or} \quad I_1 = \frac{I_c I_2}{I_0} + \frac{I_d}{I_0} I_2, \tag{8.46}$$

which corresponds to the s-domain transfer function

$$\frac{I_2(s)}{I_1(s)} = \frac{1}{I_d/I_0 + s\tau}. \tag{8.47}$$

It is thus a lossy integrator, that can be used to realize some specific filters [183]. The amount of damping could in principle be reduced by re-injecting an adequate proportion of I_2 at the input, but such a compensation is limited by the inaccuracy of current mirrors.

A better solution [182] is to implement a second loop formed of M_{1a}-M_{3a}-M_4-M_2 as also depicted in dotted line in Figure 8.13. The current through M_{3a} can now discharge C. The two input currents I_1 and I_{1a} are delivered by a signal conditioner that separates the positive and negative half-waves of the input signal.

8.5 Pseudo-Resistor

8.5.1 Analysis of Circuits

The concept of pseudo-resistor can facilitate the analysis of current-mode circuits by transforming them in a resistive prototype. In general, this possibility is limited to circuits made of transistors sharing the same gate voltage, like attenuators or special current mirrors [113, 109]. But, as was explained in Section 5.5, this limitation does not exist in weak inversion. If several transistors (in a common substrate) have the same gate voltage, the corresponding pseudo-resistances are proportional to their respective specific currents. For a difference ΔV_G of gate voltages, the ratio of pseudo-resistances is multiplied by $\exp(-\Delta V_G / nU_T)$ according to (5.54).

We have already seen an example of application in the linearized differential pair of Figure 8.4. Another example is illustrated by Figure 8.14. Consider

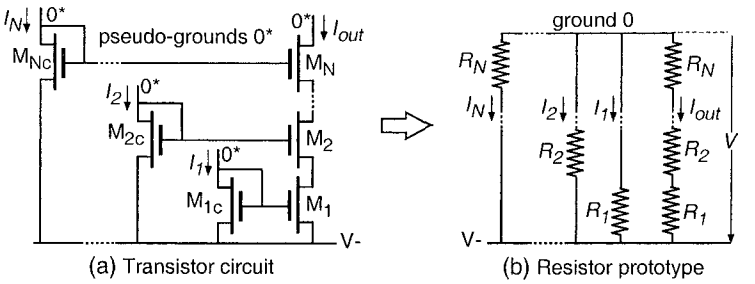

Fig. 8.14. Calculation of the harmonic mean of N currents.

the transistor circuit of Figure 8.14(a). It has N input currents I_1 to I_N and produces the output current I_{out}. All transistors are identical, some of them saturated, others not (M_1 to M_{N-1}). This circuit could be analyzed by introducing equation (5.41) for each transistor and by solving the resulting set of equations. Another possibility is to use the translinear principle, with N loops. The non-saturated transistors would have to be split in their forward and reverse components, as explained at the end of Section 8.4.3.

A much simpler approach uses the resistor prototype (or resistor equivalent) shown in in Figure 8.14(b). Since all transistors are identical, each pair M_i-M_{ic} corresponds to a pair of resistors of value R_i. As explained in Section 8.4.3, the saturated side of a transistor is a pseudo-ground (labeled 0*), that corresponds to a real ground (common node 0) is the resistor prototype (notice that, since we have N-channel transistors, all voltages are negative in the prototype). By inspection of this simple circuit, we obtain

$$I_{out} = \frac{V}{\sum_{i=1}^{N} R_i} = \frac{V}{\sum_{i=1}^{N} (V/I_i)} = \frac{1}{\sum_{i=1}^{N} (1/I_i)} = \frac{I_{hm}}{N}. \tag{8.48}$$

The output current is proportional to the *harmonic mean* I_{hm} of the N input currents I_i. If the sum of these input currents is forced to be constant, I_{out} is maximum when all I_i are equal (coincidence or "bump" circuit [184]).

8.5.2 Emulation of Variable Resistive Networks

Thanks to the concept of pseudo-resistors, any network of variable linear resistors can be implemented by transistors operated in weak inversion, provided only currents are considered.

Each pseudo-resistor implemented by a transistor M can be controlled by a current by associating a control transistor M_c of same size, as illustrated in Figure 8.15. The pseudo-conductance $G^* = 1/R^*$ is then proportional to the control current I_c [112]. If the bias voltage V_b is the same for all pseudo-resistors of a circuit, then $G_i^*/G_j^* = I_{ci}/I_{cj}$ (same proportionality constant).

Both transistors must remain in weak inversion. Thus, for M≡M_c:

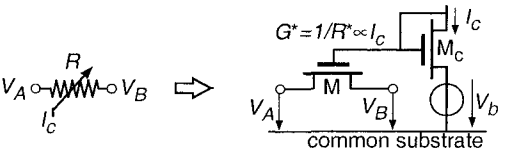

Fig. 8.15. Current control of a pseudo-resistor R^* corresponding to a real resistor R.

$$I_c \ll I_{spec} \quad \text{and} \quad V_A, V_B \geq V_b. \tag{8.49}$$

If needed, several pseudo-resistors may be controlled by the same control current through the same control transistor.

One of the most immediate applications of controlled resistors is the multistage variable linear attenuator depicted in Figure 8.16. The amount of at-

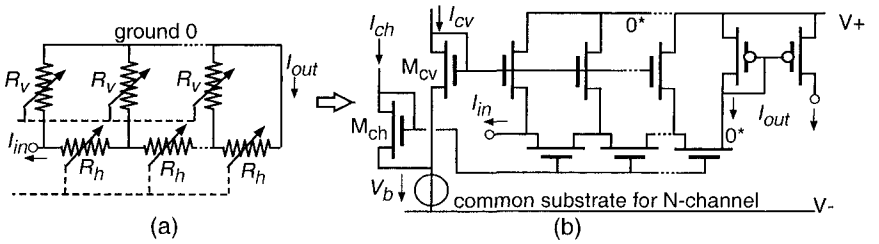

Fig. 8.16. Current-mode variable linear attenuator: (a) resistor circuit; (b) pseudo-resistor implementation.

tenuation depends on the number of stage and on the ratio $R_h/R_v = I_{cv}/I_{ch}$. Notice that the bias voltage V_b is needed to provide sufficient voltage headroom for the source of input current. The output current is extracted by means of a P-channel current mirror, which does not affect the current if the last horizontal transistor (hence all vertical transistors as well) are saturated (pseudo ground 0*). If needed, the current through the vertical branch of each cell can also be extracted by a mirror.

The same circuit may be configured in a two-dimensional array to obtain a diffusion network. Some elementary spacial processing can then be carried-out on an image: low-pass spatial filtering is obtained by injecting the current of each pixel into the local node and by extracting it from the local vertical branch. Edge enhancement can then be obtained by subtracting each output from each input [50, 111]. A modification of the cell can provide local adaptation in order to eliminate gradients of illumination [185, 186].

An interesting novel application of pseudo-resistors in weak inversion is the on-line minimization of the energy spent by a multiprocessor system-on-chip to execute a set of related tasks [187].

According to Maxwell's heat theorem [188], at steady state, any network of linear resistors driven by a constant current minimizes its power dissipation. The basic idea is to emulate the total energy E_{tot} required by the system to execute the whole set of tasks within a fixed duration D_{tot} by the power P_{Rtot} dissipated by a resistor network driven by a current I_{tot}.

Consider the example illustrated in Figure 8.17. Part (a) of the figure shows the task graph of 5 tasks T_1 to T_5 executed by 2 processing elements PE1 and PE2. The duration of the whole process is D_{tot}.

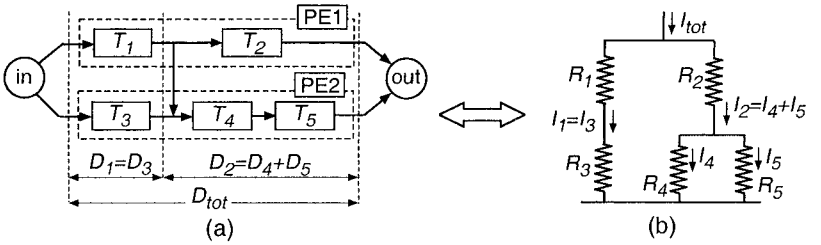

Fig. 8.17. Example of 5 tasks executed by 2 processing elements: (a) task graph; (b) corresponding network of resistors.

Now, each task T_i requires a given number N_i of cycles, and each cycle consumes an amount of energy that can be reduced if the supply voltage V_i is reduced. But the cycle time T_{ci} increases with V_i decreasing, and should therefore take the maximum value compatible with the available duration (no waiting time):

$$T_{ci} = D_i/N_i. \tag{8.50}$$

In this example, the task T_4 can only start after T_1 is ended, hence $D_1 = D_3$. The rest of the available duration D_{tot} should be entirely used for T_2 and for the two consecutive tasks T_4 and T_5. Hence $D_2 = D_4 + D_5$.

Figure 8.17(b) shows the corresponding network of resistors, in which each duration D_i of a task T_i corresponds to the current I_i in a resistor R_i. The total duration D_{tot} corresponds to the total current I_{tot} driving the circuit.

Now, the energy E_i per task and the power P_{Ri} per resistor are given respectively by

$$E_i = P_{Pi}D_i = \frac{P_{Pi}}{D_i}D_i^2 \Longleftrightarrow P_{Ri} = R_i I_i^2 \tag{8.51}$$

where P_{Pi} is the (average) power consumption of the processor executing the task T_i. Hence, since the current I_i is proportional to the duration D_i

$$\frac{P_{Pi}}{D_i} \Longleftrightarrow R_i \quad \text{thus} \quad G_i = \frac{1}{R_i} \propto \frac{I_i}{P_{Pi}} \tag{8.52}$$

Each resistor R_i is implemented as a pseudo-resistor R_i^*, so that its value can be adjusted proportionally to this ratio by means of a feedback loop that

includes a calculation of P_{Pi}. This loop is illustrated in Figure 8.18. A key el-

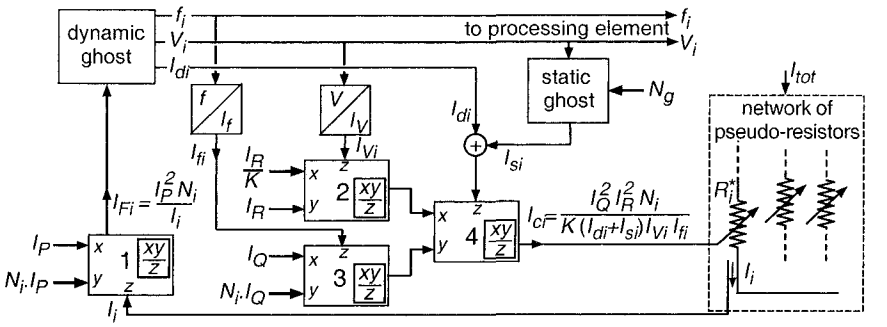

Fig. 8.18. Block diagram of a single control loop for energy minimization.

ement of the loop is a dynamic *ghost* circuit that mimics the maximum speed of the processing element built on the same chip and operated at the same voltage V_i. This ghost circuit is essentially a ring oscillator made of the same gates as the processing element. It is forced to oscillate at the required frequency $f_i = 1/T_{ci}$ by the input current $I_{Fi} \propto N_i/I_i$ in accordance with (8.50). It delivers the corresponding supply voltage V_i (which is the voltage needed to reach the frequency f_i) and the associated dynamic current consumption I_{di}. The voltage V_i and the frequency f_i are transmitted to the processing element. They are also converted to currents I_{Vi} and I_{fi} for further processing in the loop.

If necessary, a static ghost circuit is added to mimic the static current consumption of the processor, proportional to the total number of gates N_g (that may be different for different processors). This current is added to the dynamic current I_{di}.

Current-mode processing is carried out by multipliers/dividers (labeled xy/z). Each of them is implemented by the simple translinear loop of Figure 8.11(a) with some fixed bias currents I_P, I_Q and I_R.

The result is the current I_{ci} that controls the corresponding pseudo-resistor as illustrated in Figure 8.15. Hence

$$G_i = \frac{1}{R_i} \propto I_{ci} = \frac{I_Q^2 I_R^2 N_i}{K(I_{di} + I_{si}) I_{Vi} I_{fi}} \propto \frac{I_i}{P_{Pi}} \tag{8.53}$$

as in (8.52). Notice that I_i (that becomes the input of the dynamic ghost through multiplier 1) cannot be used directly to produce the control current I_{ci}. It must be replaced by $N_i/I_{fi} \propto I_i$ that is the output of the dynamic ghost (response to the input).

The factor K introduced by multiplier 2 is proportional to the equivalent switching capacitance $P_P/(fV)$, that may be different for different processors.