1 Process errors

1.1 General definitions

Fabrication of an integrated circuit is subjected to errors that make the final product different from the designed device. This problem, which is clearly typical of all industrial processes, needs to be well characterized in order to estimate the actual deviation that can be expected to occur from the ideal case. Let us start from very common definitions. We will focus on a component (e.g. a resistor) integrated on a silicon chip. Of that component, will consider a particular quantity (e.g. its resistance) that we will generically indicate with "A". The value of A assigned to the given component in the design phase is indicated as "nominal" value (A_N). Due to process errors, components integrated in the fabricated chips will show a value of A that differs from the nominal value. In addition, different realizations of the same component will show different value of A. The best way to represent the variability of the fabricated values (also indicated as "process spread") is using a histogram.

To build a histogram, we need to consider a large number of different specimens of the same component. Let us indicate the number of different samples with "n". Among this set, the quantity A assumes a minimum and maximum value. We divide the interval between the minimum and maximum into a series of uniformly sized sub-interval, called "bins", of width ΔA. For each bin, we count the number of samples whose quantity A falls into it. A graphical representation of a histogram is shown in Fig.1.1, where the quantity represented in the y-axis is the fractional number ($\Delta n/n$) of samples included in each bin.

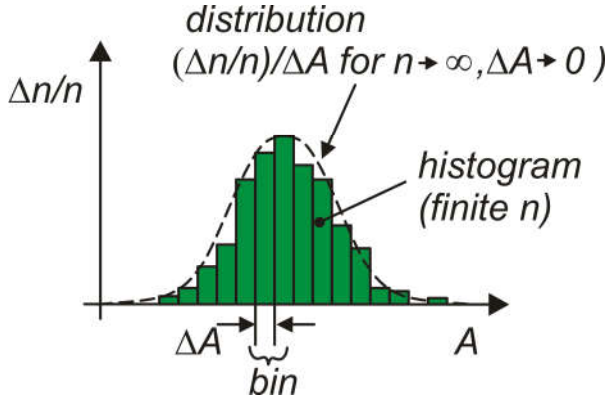

Fig.1.1. Example of histogram.

If we imagine to progressively increase the number of samples and, at the same time, increase the number of bins (reducing the width of each bin), the histogram tends to the ideal distribution that characterize the errors for the given fabrication process. To be more precise, the distribution is obtained by dividing the height of each bar in Fig. 1.1 ($\Delta n/n$) by the width of the bins (ΔA). Since A is a continuous variable, the distribution coincides with the probability density function.

The elements of the distribution that are of interest for the production process are illustrated in Fig. 1.2. These elements are summarized below:

A_N: this is the nominal value, defined in the design phase.

A_i: The value of quantity A for a generic i-th component.

$\langle A \rangle$: the mean of the distribution.

e_S: The systematic error = $\langle A \rangle - A_N$

e_R: Random error for the i-th component = $A_i - \langle A \rangle$.

The mean of the process can be estimated by taking the mean of A over a large set of components. The actual values of A taken on different components tends to group around the mean value. Differences from the mean value constitute the random error. The difference of the mean with respect to the nominal value is the systematic error. In a correct design, the systematic error should be negligible with respect to random errors. The presence of a non-negligible systematic error can be due to design errors, inaccurate or faulty fabrication process or from inaccuracy of the models used to represent the component behavior. For example, an excess systematic error may derive from neglecting the contact resistance of integrated resistors. In this case, the resistance of the fabricated resistors will be on average larger than the value set by design (nominal value).

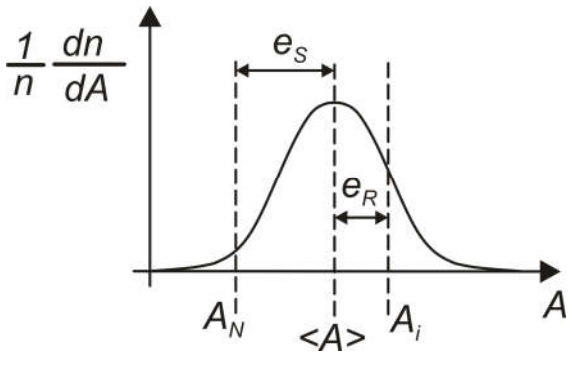

Fig. 1.2. Elements of the distribution.

The magnitude of random errors is well represented by the standard deviation (or standard error), which is the square root of the mean square value of the deviation from the mean. It is defined by:

$$\sigma_A = \sqrt{\left\langle \left(A - \langle A \rangle \right)^2 \right\rangle} \tag{1.1}$$

If we have a finite set of data (finite sample N dada), the best estimate (unbiased estimate) of the standard deviation of the whole fabrication process is given by:

$$\sigma_A = \sqrt{\frac{\sum_{i=1}^{N}\left(A_i - \mu_{A,N}\right)^2}{N-1}} \tag{1.2}$$

where $\mu_{A,N}$ is the mean calculated over the finite sample of N data. The square of the standard deviation is the variance.

The knowledge of the standard deviation is particularly important when the type of distribution is given, since it allows determining the fraction of data that fall with a given interval around the mean. Note that in most cases of interest for a fabrication process, the distribution is Gaussian. This occurs because fabrication process involve a large number of phenomena that contribute to the total random error. Generally, these phenomena are independent, so that the final distribution tends to a Gaussian even if the single distributions are not Gaussian (central limit theorem). A Gaussian distribution is perfectly determined when its man and standard deviation are given. The fraction of data that falls within an interval centered around the mean is given in the table 1.1:

Max deviation from the mean	± σ	± 2σ	± 3σ	± 4σ
Fraction of data within the interval	68.3 %	95.4 %	99.7 %	99.994 %
Fraction of data outside the interval	31.7 %	4.6 %	0.3 %	0.006 %

Table 1.1: Fraction of data that fall inside or outside an interval around the mean for a Gaussian distribution as a function of the maximum deviation from the mean.

1.2 Fabrication errors in a microelectronic process: global and local errors.

Figure 1.3 depicts the different scales of an integrated circuit (IC) fabrication process. At the smallest level there is the chip. At this stage, if we place several identical copies of the same component (nominally identical components) the differences among them are very small. For example, if we design a chip with different copies (instances) of a 1000 Ω resistor, we have good chances to get components that differ from each other by less than a few Ohms. At the second level of the fabrication process, there is the wafer, which collects hundreds or even thousands of dies (chips). The uniformity of process geometrical or physical parameters over a large wafer is much worse than over a single chip. Therefore, if we consider the set of components fabricated on the chip of the whole wafer, differences between these components begin to get significantly larger. Differences gets larger and larger as we consider the successive scale levels, that is the batch of wafers fabricated in a single run and, finally different runs. Differences between components fabricated in different runs can be very large, reaching even ± 20 %. If we consider again a resistor that is designed to have a resistance of 1000 Ω, we can likely get resistors of 800 Ω and 1200 Ω in different runs.

It is useful to introduce two new quantities:

$\langle A \rangle_{chip}$: The mean performed on all components integrated on a given chip. This value will change from one chip to another. Even if we cannot place an infinite number of copies of the same component on the same chip, we can imagine being able to reproduce the fabrication of that chip perfectly just in terms of mean values of all parameters (doping levels, oxide thickness, etc.). By this expedient, it is possible to justify the introduction of a mean, which is a property of a hypothetical process that led to the fabrication of that particular chip, and then refer to an infinite number of components.

$\langle A \rangle_{process}$ The mean performed over the totality of components fabricated by that process. Clearly, $\langle A \rangle_{process}$ is also the mean of $\langle A \rangle_{chip}$ calculated over all chips produced by that process.

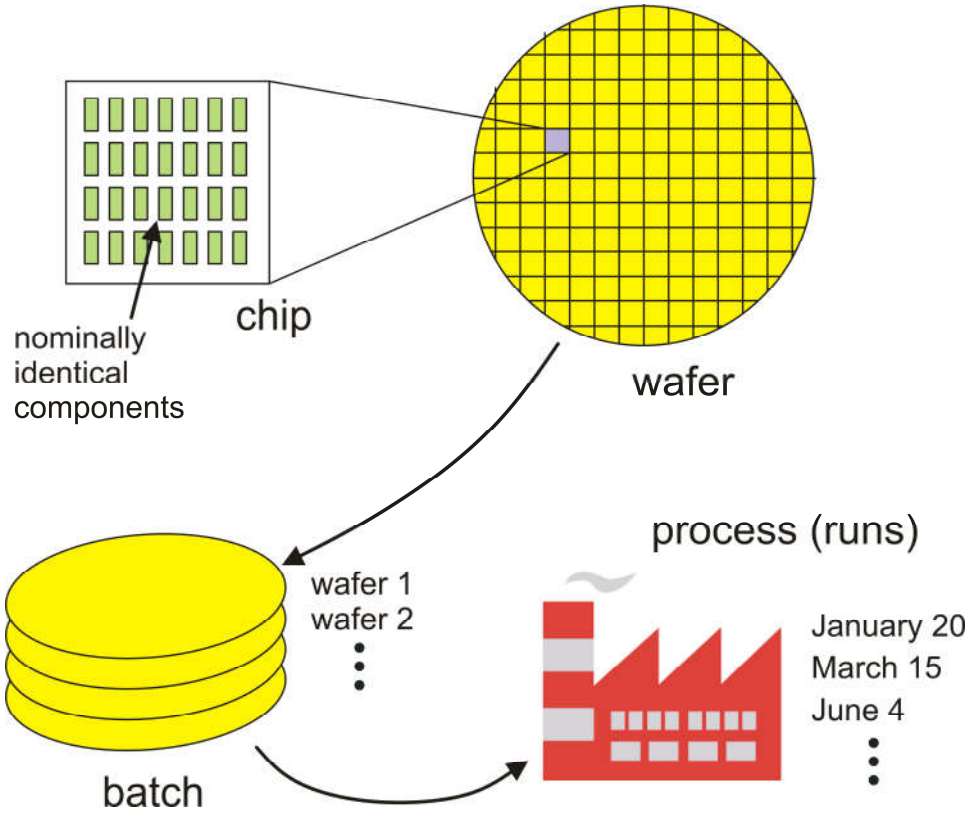

Fig. 1.3. Different scales of the fabrication process.

We can now divide the random errors into two different contributions:

-) **Local errors**, given by the difference between the value of the quantity of interest (A) assumed by a component with respect to the mean of the chip where it is located. Considering the discussion at the beginning of this paragraph, there is generally a good uniformity of parameters across a single chip, and then all components in that chip will exhibit values of A very close to $\langle A \rangle_{chip}$. In other word, local errors are generally very small. Symbolically, the local error for component i-th is given by:

$$e_{local} = A_i - \langle A \rangle_{chip} \tag{1.3}$$

where $\langle A \rangle_{chip}$ refers to the chip where component i-th is placed .

-) Global errors: given by the difference of the mean of a given chip with respect to the mean of the process. This error can be very large, since process parameters can vary much depending on; (i) the position of the chip in the wafer, (ii) the position of the wafer in the batch and, most importantly, (iii) the run the batch belongs to. (see Fig. 1.3). Symbolically, the global error for a given chip is given by:

$$e_{global} = \langle A \rangle_{chip} - \langle A \rangle_{process} \tag{1.4}$$

Figure 1.4 shows a graphical representation of the various error components. The random error is decomposed into a local and global error. The mean of single chips is distributed according to the global distribution shown at the bottom. The local distributions of two distinct chips ($chip_1$ and $chip_2$) are shown at the top of the figure. Decomposition of the random error is shown for a component belonging to $chip_1$.

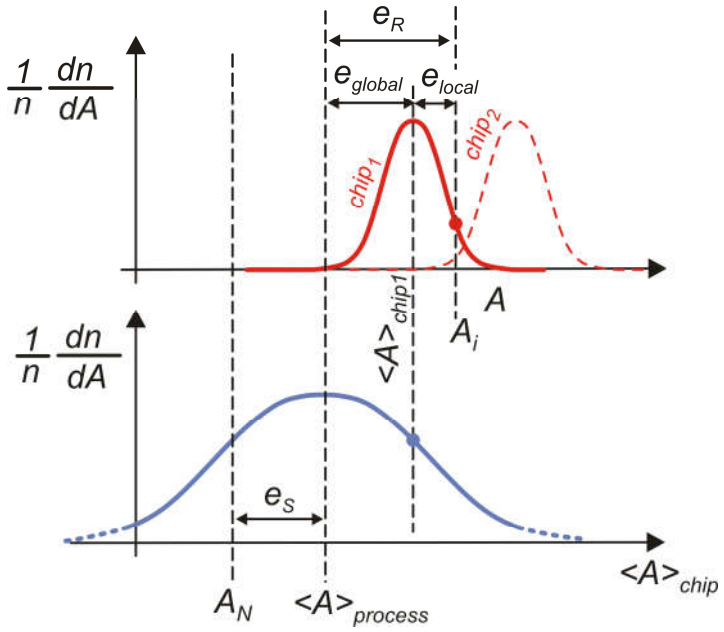

Fig. 1.4. Local (top) and global (bottom) errors. The width of local error distribution is comparatively much smaller than swon in the figure, where it has been artificially enlarged for visibility purpose.

Global and local errors are represented by distinct distributions, characterized by two distinct standard deviations, σ_{global} and σ_{local}, respectively. Different chips are characterized by different local means ($<A>_{chip}$), but all chips have the same standard deviation. This mean that distribution from different chips are simply shifted along the A axis, as shown in Fig. 1.4, but maintain the same shape and width. For the considerations made about the magnitude of global and local errors, we have:

$$\sigma_{global} >> \sigma_{local} \tag{1.5}$$

1.3 Matching errors.

A matching error is defined as the difference assumed by quantity *A* between two nominally identical components. In microelectronics, matching errors are considered only between components that are placed on the same chip. Therefore, matching errors are the consequence of local errors. If consider two component, identical by design, and indicate with A_1 and A_2 the value assumed by A on component 1 and component 2, respectively, then we can define the two quantities:

$$\begin{cases} \Delta A = A_1 - A_2 \\ \overline{A} = \dfrac{A_1 + A_2}{2} \end{cases} \tag{1.6}$$

Where ΔA is the matching error, while $\overline{A}$ is the midpoint value. Equations (1.6) can be solved to express A_1 and A_2 as a function of the matching error and midpoint value:

$$\begin{cases} A_1 = \overline{A} + \dfrac{\Delta A}{2} \\ A_2 = \overline{A} - \dfrac{\Delta A}{2} \end{cases} \tag{1.7}$$

There are two main causes of matching errors:

- Local granularity
- Gradients

1.4 Local granularity: The Pelgrom Model

Matching errors between identical components that are placed very close to each other into the same die are due to local non-uniformity ("granularity") of the material properties. To understand this, let us consider doping: dopant atoms are randomly distributed over the substrate and the number of dopant atoms that are present inside a given component will obviously vary, depending on the component location.

This phenomenon is clearly illustrated in Fig. 1.5, where the rectangle shows a portion of the chip area and the red crosses are dopant atoms. The yellow and green rectangles represent the area occupied by two nominally identical devices. Three possible placement for the two components are proposed. N is the total number of dopant atoms that fall inside the two components in a given location, while ΔN is the difference between the number atoms inside the yellow component and the number inside the green one. Note that the fluctuation occurring from one location to another is very large, reaching 38 %.

Repeating the experiment with larger components the relative fluctuation of the number of atoms is significantly reduced. This is due to the averaging effect that large areas operate on the local irregularity. Figure 1.6 represents a case in which the component width and height have been doubled, showing the considerable reduction of $\Delta N/N$. The same effect applies to other quantities, such as the gate oxide thickness, which exhibits local variations due to the unavoidable surface roughness.

As the examples in Figs. 1.5 and 1.6 clearly show, on large area devices, these short-length variations tend to have a smaller relative impact, since the device will include areas with both minimum and maximum levels of the physical quantities of interest, producing a sort of compensation.

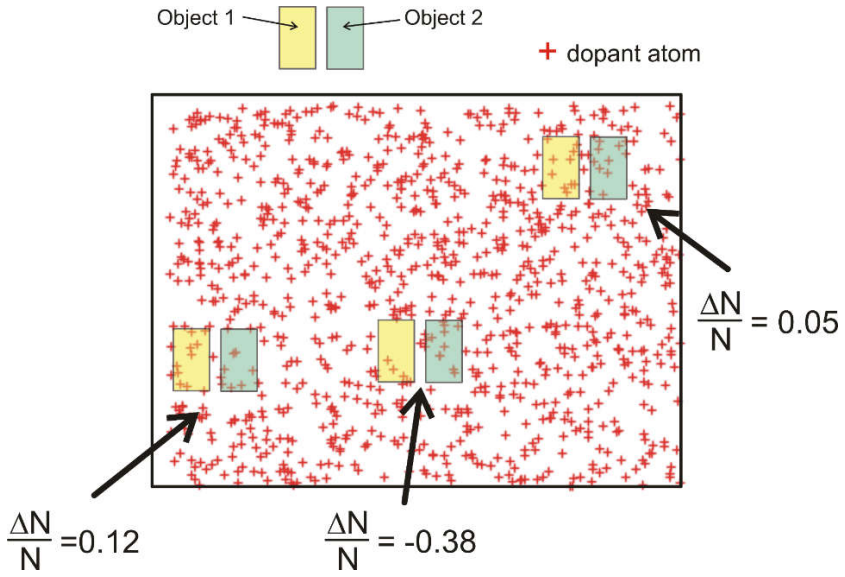

Fig.1.5. The red crosses represent dopant atoms, while the green and yellow rectangle represent the area occupied by two identical components. Different placements result in different atom distributions within the component areas.

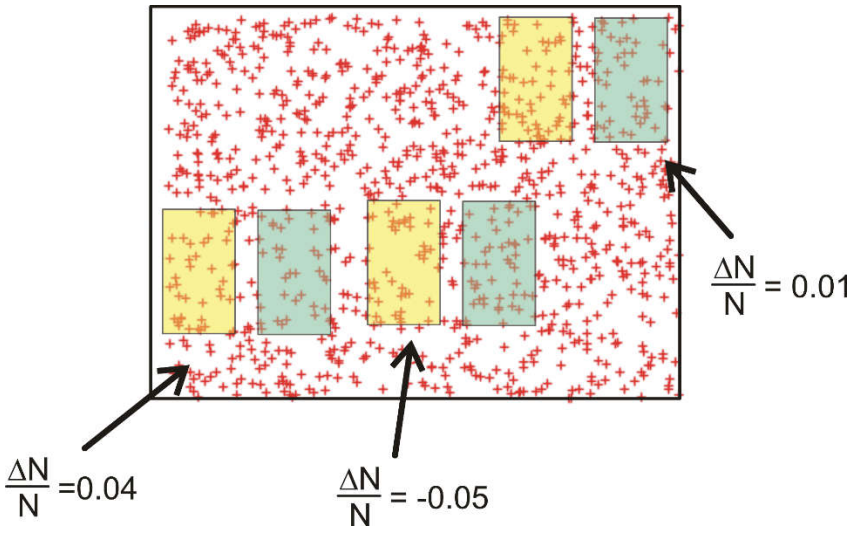

Fig.1.6: Same atom distribution as in Fig. 1.5, but with components of larger area. Note the relative fluctuation is much smaller than in the case of small components.

For this mechanism, matching errors will be smaller in large area devices. This intuitive idea is well represented in a quantitative way by the Pelgrom model [1] that express the standard deviation of the MOSFETS parameters as a function of the device gate area (WL) in the following way:

$$\begin{cases} \sigma_{\Delta V_t} = \dfrac{C_{Vt}}{\sqrt{WL}} \\ \sigma_{\frac{\Delta\beta}{\beta}} = \dfrac{C_{\beta}}{\sqrt{WL}} \end{cases} \tag{1.8}$$

where C_{Vt} and C_β are constant parameters that are typical of the fabrication process. These matching parameters can be found in the process DRM (Design Rule Manual). This model is generally valid also for other kind of devices, such as resistors, capacitors and bipolar transistors. For example, the standard deviation of the relative matching error of integrated resistors can be expressed by:

$$\sigma_{\frac{\Delta R}{R}} = \frac{C_R}{\sqrt{WL}} \tag{1.9}$$

where C_R is constant that depends on the process and on the type of resistor (polysilicon, high-resistivity polysilicon, diffusion etc.). Constants C_β and C_R are expressed in mm, while C_{Vt} is typically in mV·µm, so that W and L should be expressed in µm in expressions (1.8) and (1.9).

1.5 Gradients

Gradients indicate that important quantities that affect properties of devices are not uniformly distributed on a macroscopic scale. This means that these quantities gradually varies across the chip area.

Quantities of interest can be, for example:

-) Dopant density

-) Oxide thickness

-) Geometrical process biases (e.g. etching undercut)

-) Temperature (e.g. due to power devices present on the chip)

-) Mechanical stress (mainly due to the packaging process)

The effect of the gradient of a given quantity "A" (can be one of the list above) on the matching of two components is shown in Fig. 1.7

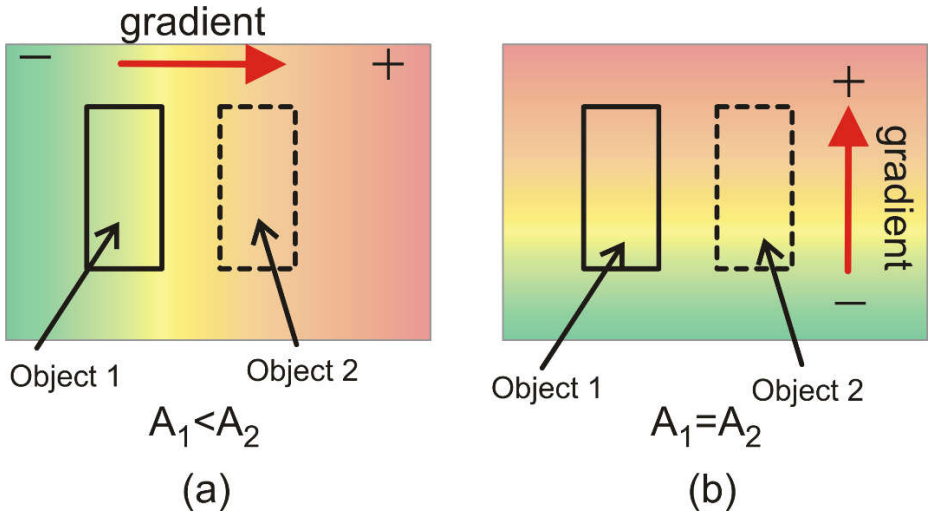

Fig. 1.7. Effect of gradients on device matching. Different colors represent different values of a given quantity "A". In (a) the average of quantity A is larger for component 2, while in (b) the two components receive the same average of A.

The effect can be different depending on the gradient orientation with respect to the line that join the two component locations. If the gradient and the joining line are parallel, as in the case of Fig. 1.7(a), the mismatch will be maximum; if the gradient and the joining line are orthogonal, the gradient does not cause mismatch, as shown in Fig. Fig. 1.7(b). Unfortunately, generally it is not possible to predict the gradient direction, since it will vary according to the position of the chip in the wafer, the wafer in the batch and so on. Only in the case of mechanical stress and temperature distribution, it is possible to have an idea of the gradients from the way the chip is mounted on the package and from the position on the chip of power devices, that can be important heat sources. However, even in these cases the prediction is fairly inaccurate, so that gradients are likely to produce mismatch.

An effective solution is offered by the so-called common centroid configurations. The two devices that should match are split into different identical parts that are then placed in such a way that parts from object 1 are interleaved with parts from object 2. The requirement is that the centroids of the two devices coincides. This method is illustrated in Fig. 1.8: Component A_1 is divided into the two identical parts $A_{1,1}$ and $A_{1,2}$, while A_2 is divided into $A_{2,1}$ and $A_{2,2}$. The centroid of the two devices is indicated with C. Note that by splitting each device into two parts, the centroid is allowed to lie outside each convex shape that form the device (rectangles in the example). In this way, we can make the centroid to coincide even if the two devices does not overlap in any point.

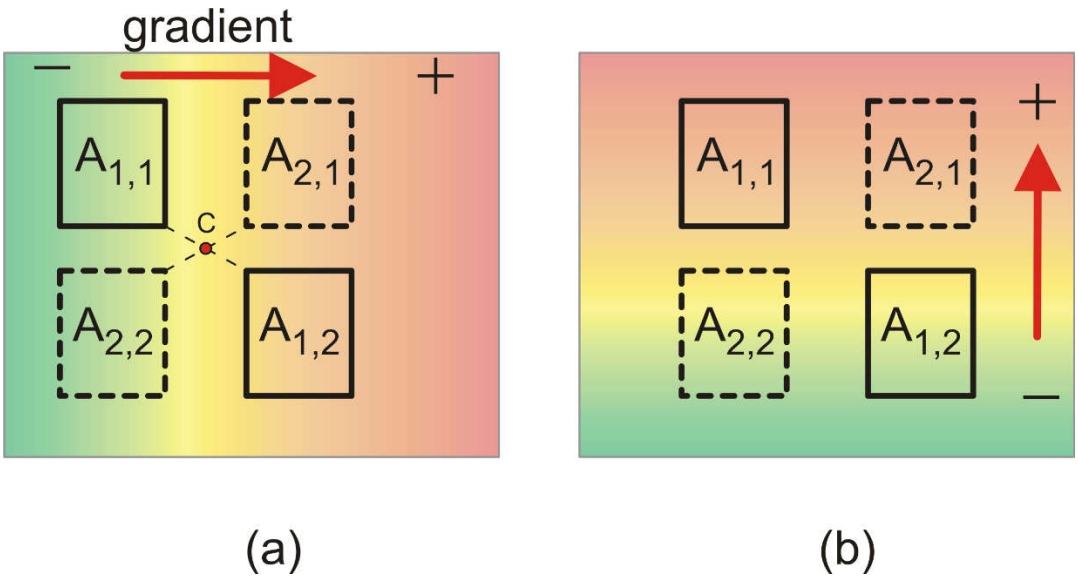

Fig. 1.8. Common centroid configuration in the case of two different gradient orientations.

From Figure 1.8 (a) and 1.8 (b) it is clear that now the two devices are affected by the quantity of interest (to which the gradient refers) in exactly the same way, independently of the gradient direction. In both cases depicted in Fig. 1.8 (a) and 1.8 (b), each device has a part that receives a larger value of the quantity while the other part receive a smaller value. On average, both devices receive the same value.

Note that Fig. 1.8 (a) and (b) represent two particular cases. Fig.1.9. represent the case of an oblique gradient. Object A_1, formed by parts $A_{1,1}$ and $A_{1,2}$ gets an intermediate value of the quantity. Object A_2 receives an higher value (part $A_{2,1}$) and a lower value ($A_{2,2}$). Again, we have a compensation and on average the two components A_1 and A_2 receive the same effective value of the quantity of interest. Differently from the cases depicted in Fig. 1.8 (a) and (b), the symmetry is not perfect for oblique

gradients and compensation is perfect only if the gradient is constant across the whole area occupied by the two components.

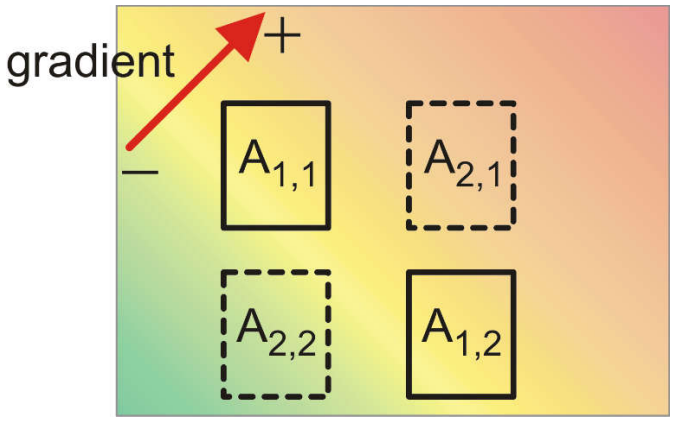

Fig. 1.9. Common centroid configuration in the case of oblique gradient.

In order to apply the common centroid approach, it is necessary to split each component into two parts that, properly connected, must still behave like the original component. The way components can be split and re-connected depends on the type of device. Figure 1.10 show the two options that can be adopted when the two components to match are resistors (R_1 and R_2, of nominal value R).

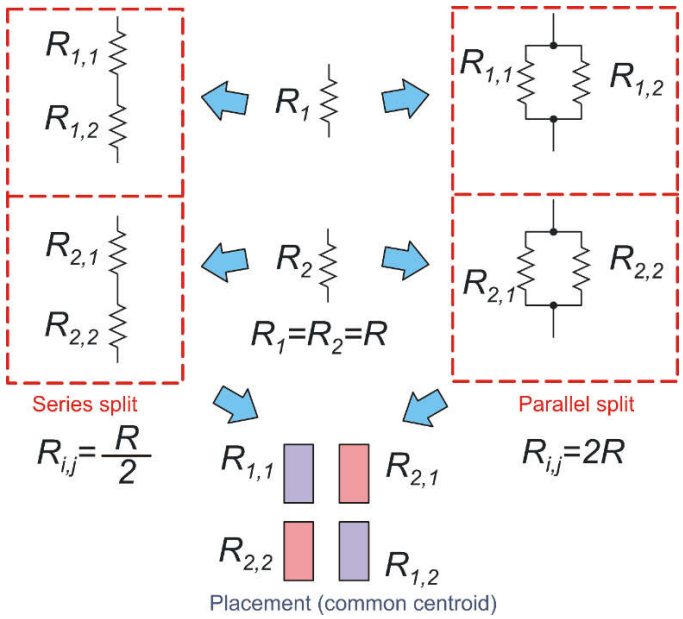

Fig. 1.10. Common centroid configuration of resistors implementd with series split and parallel split.

On the left, each one of the two resistors is split into two parts that are then reconnected in series. In order to maintain the original value, the individual parts should have half of the original value, i.e. their resistance should be R/2. On the right, the parts are connected in parallel. Then, to maintain the resistance of the original components, each part should have a resistance 2R. The placement is the same for the series and parallel split. The series split is advantageous in the case that R is large, resulting in particularly long resistors. The parallel split has to be preferred only in the case of small resistance (short resistors).

In the case of active devices, such as MOSFETs or BJTs, the only possible way to split the original components is the parallel split. Figure 1.11 illustrate the case of common centroid applied to MOSFETs. If the β (=$\mu_n C_{ox} W/L$) of the The original devices, M1 and M2, have the same nominal parameter β (=$\mu_n C_{ox} W/L$). Then the parts in which they are split are characterized by β/2. In practice, the parts have half the width (W) of the original MOSFETs. In a parallel connection of MOSFETs, the effective beta is the sum of the betas of the devices that form the parallel. The actual arrangement of the single parts is shown in Fig. 1.11, on the right. Series split configurations are not applicable to common centroid arrangement of MOSFETs. The reason is that in a series of MOSFETs, the two parts do not contribute in the same way the property of the composite device. The same applies to BJTs, for which the only possible split is the parallel one.

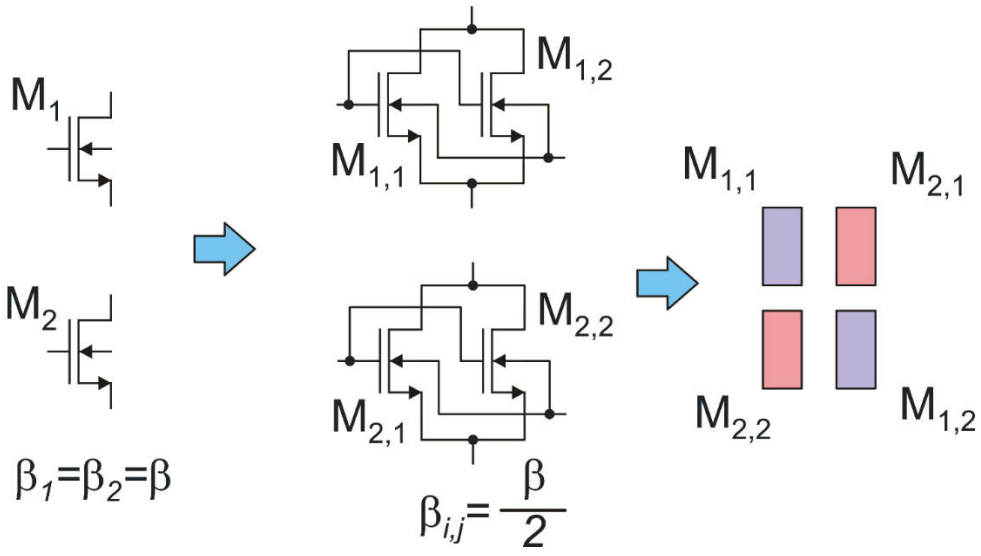

Fig. 1.11. Common centroid applied to MOSFETs.

Finally, common centroid configurations are widely used also for capacitors. A good matching between capacitors is a key element for the accuracy of switched capacitors circuits and in particular of charge-redistribution analog to digital converters. In principle, common centroid configurations can be applied to capacitors using both the series and the parallel split, just as for resistors. In practice, series connection of integrated capacitors has to be discouraged, because the dc voltage of intermediate nodes in a series of capacitors cannot be easily controlled. As a result, common centroid schemes are applied to capacitors using mainly the parallel split approach.

1.6 General rules for matching components

In addition to the rules introduced in paragraphs 1.4 (area of devices) and 1.5 (common centroid arrangement), there are also important rules that have to be mandatorily or optionally followed to reduce matching errors between component pairs. Figure 1.12 shows two mandatory rules. On the left, two objects with identical W/L ratios (e.g. two resistors or two MOSFETs) are shown. Setting the aspect ratios to be equal is not sufficient to obtain a good matching, even in the case that the expressions of the quantities of interest (e.g. resistance) include only the W/L ratios. The reason is that the properties of the materials that compose the devices tends to be different in the proximity of the boundaries of the device area (borders). For example, the resistivity of a conducting layer may be higher close to the borders due to reduction in dopant concentration or to increased scattering mechanisms. Since the extensions of the borders does not depend on the device dimensions, border-related effects will have a greater relative impact on the smaller device. For this reason, matched devices should be identical (same width and length). Note that the matching errors introduced by different device areas are systematic. Figure 1.12 shows two identical object (same lenghts and widths) that are placed along orthogonal directions. This may lead to poor matching since material properties can be anisotropic. The typica cause is mechanical stress due to the packaging procedure: packaging often occurs at a temperature that can exceed one hundred degrees. Successive cooling down produces mechanical stress through the different thermal expansion coefficients of the chip and package materials. Mechanical stress has generally a prevalent direction and this results in mentioned anysotropy. In addition, also the device dimensions are unevenly modified by the stress. A resistor subjected to mechanical stress that having a prevalent axis parallel to resistor length, will become longer and narrower than the original device. The opposite occurs if the stress is orthogonal to the resistor length. The stress will then cause different changes in the W/L ratios of devices width different orientations. As a result, matched devices should have the same orientation.

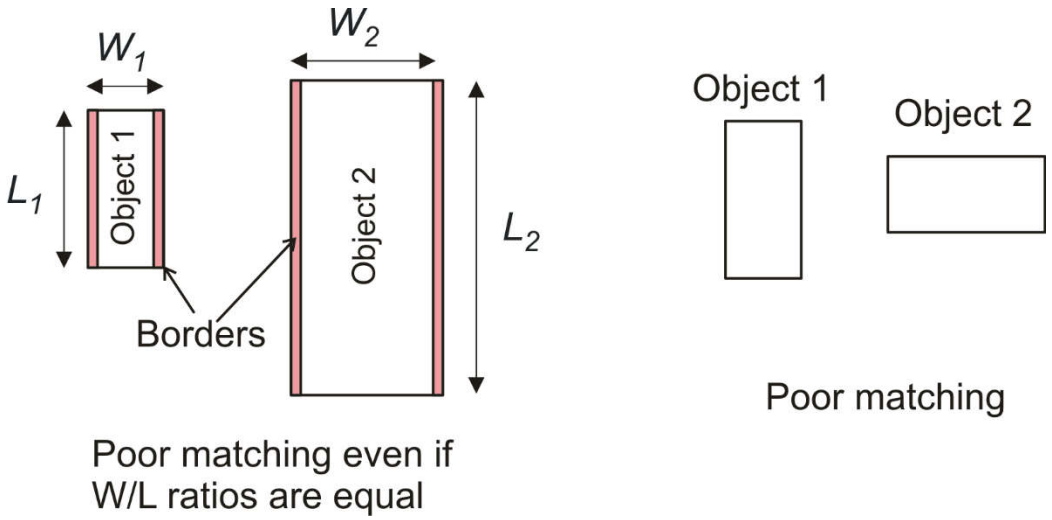

Fig. 1.12. Two common errors leading to poor matching: (left) fifferent device areas and (right) different device orientation.

.

Figure 1.13 illustrates two optional rules that have to be adopted when very low matching errors have to be achieved. The rule represented on the left regards the direction of current in the device. To obtain a good matching the direction of the current in the two devices should be the same. The reason is that unavoidable temperature gradients introduce an additional voltage drop whose sign depends on the

relative direction of the current with respect to the direction of the gradient. Another rule, indicated with "common surroundings" or "common environment" is illustrated in Fig. 1.13 (right).

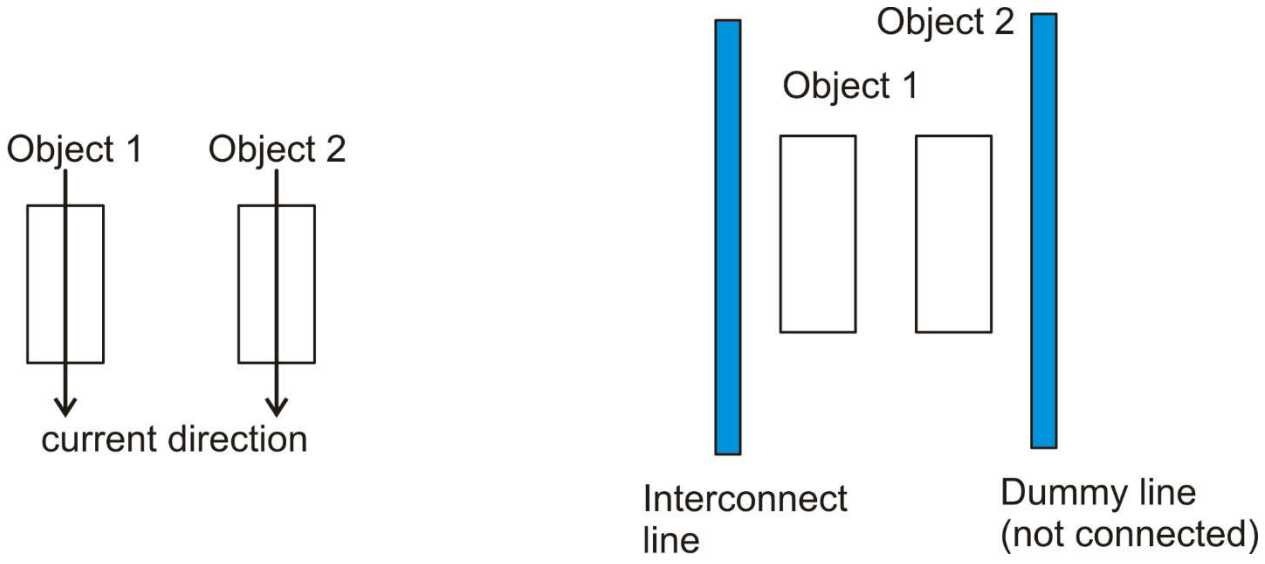

Fig. 1.13. Common direction (left) and common surroundings (right). .

If object 1 is close to a layout object (an interconnect line in the figure), then, to obtain an excellent matching, also object 2 should be close to a similar object. In other words, it is not sufficient that the objects are symmetrical, but also the environment where the object are placed must be symmetrical. If a metal is not passing close to object 2, we must place a metal (dummy line) that is not used for interconnection but only to make the environment symmetric. The dummy line can be left floating or, preferably, connected to gnd.

1.7 Rules for accurate ratios.

Frequently, important properties of electronic circuits are expressed as ratios of values of different components. A very simple example is shown in Fig. 1.14, depicting an inverting amplifier formed by an operational amplifier and two resistors.

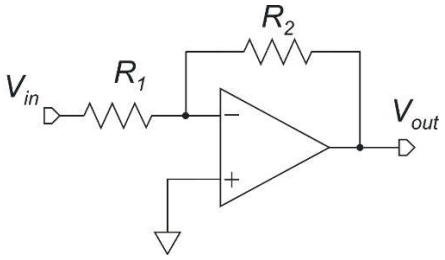

Fig. 1.14. Opamp-based inverting amplifier.

If the loop gain is large enough, the amplifier gain is simply given by:

$$A_V = -\frac{R_2}{R_1} \tag{1.10}$$

In many cases the gain magnitude (R_2/R_1) must be accurate, with differences from the ideal case of less than 1 %. If the gain magnitude is one, than R_1 should be equal to R_2 and obtaining a precise gain becomes simply a problem of good matching between the two resistors. If the gain to be obtained is different than one, than the problem is different. The more intuitive approach would be simply to introduce two resistance and set their value in order to obtain the required ratio. Unfortunately, many automated design kits assign the entered value of the resistance to the body of the resistor, leving out of the resistance computation the contact resistance,

Figure 1.15 (a) shows what happen if we simply try to obtain the required resistance ratio *r* by setting $r=L_2/L_1$. This sets the ratio of the resistor body approximately to the correct ratio. However, considering also the contribution of the contact resistances, as shown by the equivalent circuit of Fig. 1.15 (b), the actual ratio will be.

$$\frac{R_2}{R_1} = \frac{2R_C + rR}{2R_C + R} = r\frac{1 + 2R_C / rR}{1 + 2R_C / R} \tag{1.11}$$

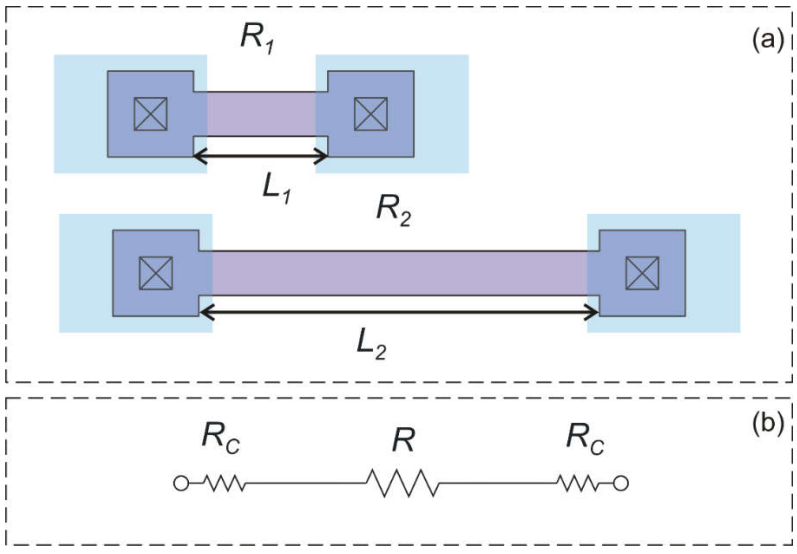

Fig. 1.15. Ratio obtained by simply scaling the resistor length (a). Equivalent circuit of a resistor with the contavt resistances (b).

The actual resistance ration is not r, unless r=1. For example, if r=3 and R_C/R=0.1, we would get: R_2/R_1=2.81, committing an error of nearly -6 %. In most cases, such an error is not acceptable. Clearly, it is possible to redesign the resistance values in order to take into account the contact resistance and obtain a more precise resistance ratio. Modern design kit of processes oriented to analog and mixed signal

design do so automatically. Unfortunately, contact resistance are not as accurate as resistor bodies, thus there would be still an important process-dependent inaccuracy. Furthermore, border effects that are not documented, makes also the resistor body close to both ends different from the central region. Again, border effects have an higher impact on the shorter resistor.

A much more accurate resistor ratio can be obtained by the so called modular approach. This is illustrated in Figure 1.16 (a) for a target ratio $r=R_2/R_1=3$. A single resistor module R_0 is used to form R_1, while R_2 is obtained by simply connecting three identical module R_0 in series. A possible layout for R_0 is shown in Fig. 1.16 (b), while the layout of the two resistors R_1 and R_2 is shown in Fig. 1.16 (c).

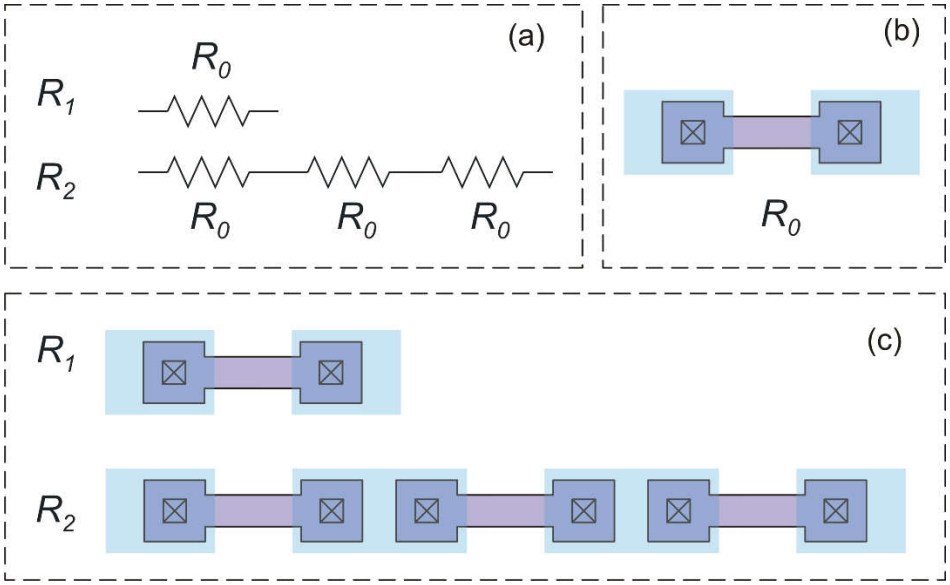

Fig. 1.16. Example of resistance ratio obtained with the modular approach.

By this arrangement, contact resistances and all other systematic non-idealities will affect in the same way all the instances of module R_0, so that the ratio R_2/R_1 will not change with respect to the ideal value ($r=3$ in the example). The method can be easily extended to non-integer ratios of the form M/N as shown in Fig. 1.17, where series of N and M modules (R_0) are used for R_1 and R_2, respectively.

In addition, it is possible to arrange the modules in parallel. This is advantageous when the resistances used to implement the ratio are particularly small. Combinations of series and parallel combinations of the same module R_0 can be used when a very large or a very small ratio has to be obtained. Using only pure series or parallel combinations would lead to the requirement of a large number of modules. For example, to implement a ratio $r=100$, the number of module involved is 101. The same occurs, obviously, if the ratio to be obtained is 1/100. Using a parallel of 10 R_0 modules for R_1 and a series of 10 R_0 modules for R_2 allows obtaining the required ratio of 100 with only 20 instances of module R_0. This approach is illustrated in Fig. 1.18.

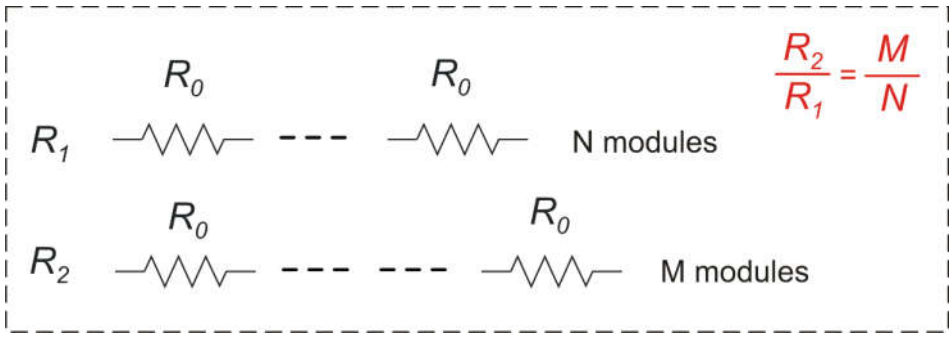

Fig. 1.17. Non-integer ratios R_2/R_1 obtained by the modular approach. .

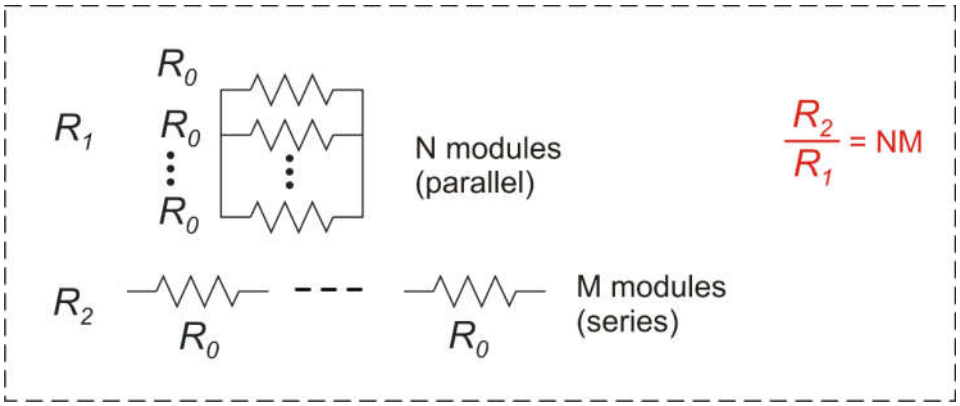

Fig. 1.18. Combination of series and parallel connections of modules to obtain very large ratios with a smaller number of components than pure series or parallel modular approaches.

Precise capacitance ratios are generally obtained with parallel connections, for the same reason mentioned for the common centroid configurations.

In the case that the requirement of precise ratios refers to the beta factor (β) of MOSFETs, high accuracy can obtained only with the parallel approach, although series or mixed connections have been proposed in the scientific literature [2]. An example of modular approach applied to a MOSFET-based current mirror is shown in Fig. 1.19. In both the input and output branch of the mirror, composite MOSFETs formed by parallel of a different number of nominally identical modules (M_0) are used. The composite MOSFET in the input branch, M1, is formed by *N* modules, while in the output branch M2 is formed by *M* modules. This corresponds to set the nominal ratio β_2/β_1 exactly equal to *M*/*N*. Doing the same using single MOSFETs for M1 and M2 and varying the aspect ratio (*W*/*L*) to obtain the required β_2/β_1 ratio results in a significant difference with respect to the target value due to the mentioned border effects (effective W and L are different from the drawn dimensions) and by unwanted effects that both *W* and *L* has on the threshold voltage. However, in all cases that a precise ratio is not required, it is preferable to act on the aspect ratio since it generally allows saving silicon area.

In the case of BJTs precise ratios of the saturation currents (I_S), are obtained only using parallel connections of a single module.

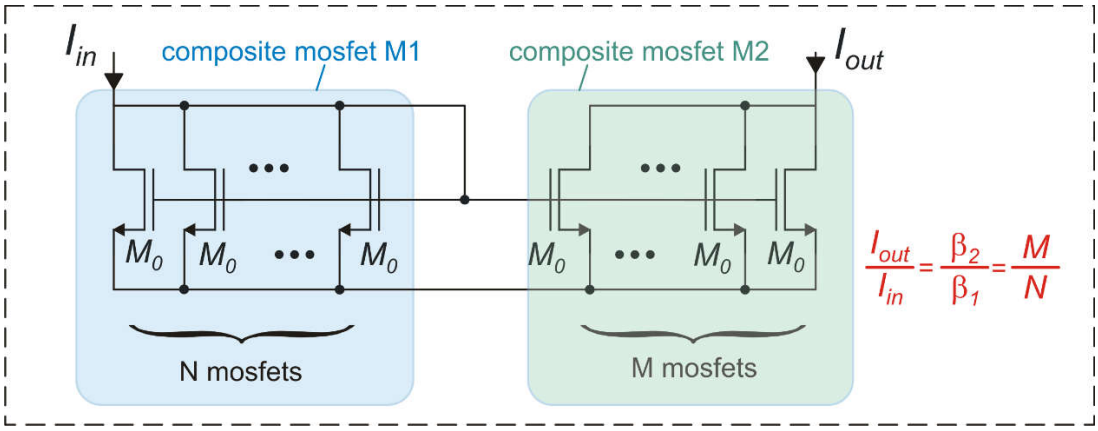

Fig. 1.19. Modular approach applied to a MOSFET-based current mirror.

1.8 Error propagation elements applied to matching errors

Generally, expressions that allow prediction of process errors (both global and local) are available for a few important parameters of the main devices of the process. An example is given in paragraph 1.4, where expressions for the standard deviations of matching errors are given for the beta factor and threshold voltage of MOSFETs. The problem that has frequently to be solved is finding how global properties of a circuit (e.g. an amplifier gain), are affected by the errors on the parameters of each component of the circuit. This is a particular case of a general problem called error propagation, which consists in finding the error on a quantity G that depends on variables A, B, C and so forth , resulting from the errors on the variables it depends on.

In this paragraph, a few remarkable cases that are easy to remember and that recur often in analog electronics. The focus will be on matching errors, but the results can be directly applied to a much wider spectrum of cases.

One-dimensional case

Let us start with a simple problem, illustrated in Fig. Quantity G depends on a single quantity A in a non-linear fashion. We are interested at finding the difference of the values assumed by G for two distinct values of A, indicated with A_1 and A_2. We introduce the following definitions:

$$\begin{aligned} &G_1 = G(A_1);\ G_2 = G(A_2) \\ &\Delta G = G_1 - G_2\ ;\ \Delta A = A_1 - A_2 \end{aligned} \tag{1.12}$$

In the case of matching errors, we have two object, named object 1 and object 2, that should be as equal as possible. Thus, A_1,G_1 refer to object 1 while A_2, G_2 to object 2. However, this model can represent also other useful situations. Examples are:

- we have a single component and A_1, G_1 refer to the nominal case, while A_2, G_2 to the real case that will be affected by both global and local errors.

- A_1, G_1 and A_2,G_2 are simply two different statuses in which a given real component can be found.

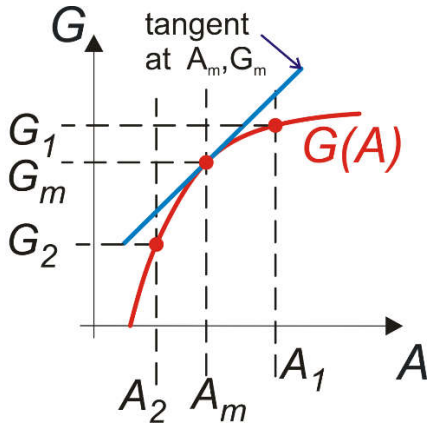

Fig. 1.20. Simple example of error propagation problem used to introduce the basic definitions.

We now introduce a third value of A, indicated with A_m. whose position with respect to A1 and A2 is completely arbitrary. Particular cases are when Am coincide with either A_1 or A_2 or is placed just in the middle of them. Values A1 and A2 can be expressed through their deviations with respect to A_m:

$$\begin{cases} A_1 = A_m + \Delta A_1 \\ A_2 = A_m + \Delta A_2 \end{cases} \tag{1.13}$$

Using a first order approximation of the $G(A)$ function around point A=A_m, we can express G(A_1) and G(A_2) as:

$$\begin{cases} G_1 = G\left(A_m + \Delta A_1\right) \cong G\left(A_m\right) + \Delta A_1 \left.\dfrac{dG}{dA}\right|_{A=A_m} \\ G_2 = G\left(A_m + \Delta A_2\right) \cong G\left(A_m\right) + \Delta A_2 \left.\dfrac{dG}{dA}\right|_{A=A_m} \end{cases} \tag{1.14}$$

$$\Delta G = G_1 - G_2 \cong \Delta A \left.\frac{dG}{dA}\right|_{A=A_m} \tag{1.15}$$

where $\Delta A = A_1 - A_2 = \Delta A_1 - \Delta A_2$.

The result represented in (1.15) is the well-known first order approximation of the relationship between the increment in the dependent variable G and the corresponding increment in the independent variable A. What we want emphasize is that the approximation shown in (1.15) is independent on the point where the derivative is calculated (point A=Am). Changing the point affects the accuracy of the approximation, but not the form of the latter.

Multi-dimensional case

In the general case, the quantity G depends on several independent variables, indicated with A, B, C and so forth. For the sake of simplicity, we will consider the case of three independent variables. We are interested at two points in the space of the independent variables and we will indicate these points as $\boldsymbol{P_1}=(A_1,B_1,C_1)$ and $\boldsymbol{P_2}=(A_2,B_2,C_2)$. Repeating the considerations made for the one-dimensional case, we can find a linear approximation of the function $G(A,B,C)$ around an arbitrary point $\boldsymbol{P_m}=(A_m,B_m,C_m)$. In the same way as in Eq. (1.13) we can express points $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$. through their deviations with respect to $\boldsymbol{P_m}$:

$$\begin{cases} A_1 = A_m + \Delta A_1 \, , B_1 = B_m + \Delta B_1 \, , C_1 = C_m + \Delta C_1 \\ A_2 = A_m + \Delta A_2 \, , B_2 = B_m + \Delta B_2 \, , C_2 = C_m + \Delta C_2 \end{cases} \tag{1.16}$$

Then, the first order approximation of $G(A_1,B_1,C_1)$ and $G(A_2,B_2,C_2)$ can be written in the form:

$$\begin{cases} G_1 = G(A_m, B_m, C_m) + \Delta A_1 \left.\dfrac{\partial G}{\partial A}\right|_{P_m} + \Delta B_1 \left.\dfrac{\partial G}{\partial B}\right|_{P_m} + \Delta C_1 \left.\dfrac{\partial G}{\partial C}\right|_{P_m} \\ G_2 = G(A_m, B_m, C_m) + \Delta A_2 \left.\dfrac{\partial G}{\partial A}\right|_{P_m} + \Delta B_2 \left.\dfrac{\partial G}{\partial B}\right|_{P_m} + \Delta C_2 \left.\dfrac{\partial G}{\partial C}\right|_{P_m} \end{cases} \tag{1.17}$$

The difference $\Delta G=G_1-G_2$ can then be easily obtained from (1.17) as a function of the deviations $\Delta A=A_1-A_2$, $\Delta B=B_1,B_2$ and $\Delta C=C_1-C_2$.

$$\Delta G = G_1 - G_2 = \Delta A \left.\frac{\partial G}{\partial A}\right|_{P_m} + \Delta B \left.\frac{\partial G}{\partial B}\right|_{P_m} + \Delta C \left.\frac{\partial G}{\partial C}\right|_{P_m} \tag{1.18}$$

The advantage of using an arbitrary point for the calculation of the derivatives is that we can use the point that is more advantageous for the particular situation. Clearly, for the approximation to be accurate enough, point $\boldsymbol{P_m}$ should be close to both $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$. As already stated for the one-dimensional case, Pm may be one of the two points $\boldsymbol{P_1}$ and $\boldsymbol{P_2}$, or $(\boldsymbol{P_1}+\boldsymbol{P_2})/2$. Another possible choice that can be convenient in some occasions is using the nominal value for $\boldsymbol{P_m}$, since it is the only value that we know *a priori*.

Useful examples of relationships that occurs frequently

The first relationships that will be analyzed are linear relationships. The following properties, where k is a constant, can be easily demonstrated:

$$\begin{cases} G = A + B \Rightarrow \Delta G = A_1 + B_1 - \left(A_2 + B_2\right) = \Delta A + \Delta B \\ G = kA \;\Rightarrow \Delta G = kA_1 - kA_2 = k\Delta A \end{cases} \tag{1.19}$$

Then, linearity can be applied to deviations. After that, it is interesting to analyze the so-called posynomial function, defined by the following formula:

$$G(A,B,C) = A^{\alpha} B^{\beta} C^{\gamma} \tag{1.20}$$

where α, β and χ are constant coefficients. Applying (1.18), we find:

$$\Delta G = \alpha A_m^{\alpha-1} B_m^{\beta} C_m^{\gamma} \cdot \Delta A + \beta A_m^{\alpha} B_m^{\beta-1} C_m^{\gamma} \cdot \Delta B + \gamma A_m^{\alpha} B_m^{\beta} C_m^{\gamma-1} \cdot \Delta C \tag{1.21}$$

This is not an expression that can be easily remembered. The formula becomes much simpler and meaningful if we calculate the relative error, $\Delta G/G$, the particular form $\Delta G/G_m$, where G_m is given by:

$$G_m = G(A_m, B_m, C_m) = A_m^{\alpha} B_m^{\beta} C_m^{\gamma} \tag{1.22}$$

With simple calculations, we find:

$$\frac{\Delta G}{G_m} = \alpha \frac{\Delta A}{A_m} + \beta \frac{\Delta B}{B_m} + \gamma \frac{\Delta C}{C_m} \tag{1.23}$$

This expression can be summarized by saying that the relative error of a posynomial dependent variable is the sum of the relative errors of the independent variables, weighted by the respective exponents. Clearly, in the case that we need to calculate the absolute error ΔG, then we can simply obtain by multiplying expression (1.23) by Gm, given by expression (1.22).

Finally, we will consider the following remarkable case:

$$G(A, B, C) = \ln\left(A^{\alpha} B^{\beta} C^{\gamma}\right) \tag{1.24}$$

Defining a variable Z equal to the argument of the logarithm in (1.24), i.e. $Z=A^{\alpha}B^{\beta}C^{\gamma}$ we can write:

$$\Delta G = \Delta Z \left.\frac{dG}{dZ}\right|_{Z=Z_m} = \frac{\Delta Z}{Z_m} \qquad with\ Z_m = A_m^{\alpha} B_m^{\beta} C_m^{\gamma} \tag{1.25}$$

Considering that *Z* is a posynomial, its relative error $\Delta Z/Z_m$ us given by (1.23), then:

$$\Delta G = \alpha \frac{\Delta A}{A_m} + \beta \frac{\Delta B}{B_m} + \gamma \frac{\Delta C}{C_m} \tag{1.26}$$

We can summarizing expression (1.26) saying that, in the case that the posynomial expression is the argument of a natural logarithm, it is the absolute error ΔG to be the sum of the weighted relative errors of the independent variables.

1.9 References

[1] M. J. M. Pelgrom, A. C. J. Duinmaijer and A.P.G. Welbers, "Matching Properties of MOS Transistors", *IEEE J. Solid State Circuits*, vol. 24, No. 5, pp. 1433-1440, October 1989.

[2] C. Galup-Montoro, M. C. Schneider and I. J. Loss,. "Series-parallel association of FET's for high gain and high frequency applications" *IEEE Journal of Solid-State Circuits*, vol. 29, No 9, pp. 1094-1101, September 1994.