

1. Introduction

Tremendous progress in the Integrated Circuit (IC) technology brings to our mind digital circuits as microprocessors and digital-signal processors. Nowadays, it is possible to produce processors containing millions of transistors and performing billions of operations per second. CMOS technology has dominated over the market of digital IC for the last twenty years, and due to its good parameters it has also rapidly entered the field of analogue and mixed-mode integrated circuits. Relatively low cost of Very Large Scale of Integration (VLSI) production and development of software for the IC design make it possible to produce short series of Application Specific Integrated Circuits (ASIC) dedicated to very narrow applications of special use, as for example scientific experiments. There are many areas of scientific research (physics, neurobiology, medicine etc.) which need to process information from many sensors in parallel at the same time. Due to the high integration density of the modern CMOS technology that kind of parallel processing can be performed by multichannel integrated circuits. Because naturally occurring signals are analogue, to interface this real world an acquisition and initial processing analogue front-end electronic systems are necessary. The quality of these analogue parts on the chip is often a limiting factor to the overall system performance and the most difficult part of the design.

The work presented in this text is related to the parallel processing of small amplitude analogue signals, which can be done by using multichannel mixed-mode Application Specific Integrated Circuits. Because the amplitudes of the measured signals are very small, the aspect of low noise IC design becomes one of the most important. In addition to the small amplitudes of these signals, they are often accompanied by unwanted, out-of-band noise and disturbances. Therefore, it is necessary to amplify and filter natural signals in order to allow digitisation and further processing. The implementation of some digital processing directly on the chip increases the flexibility and reliability of the system, but on the other hand digital circuits can disturb the operation of analogue blocks and in that way degrade the analogue parameters of the entire system. Moreover, since we consider parallel processing of natural signals in a multichannel system, the designer should remember about the uniformity of analogue parameters for all channels. That is not an easy task because physical circuit components (i.e. resistors, capacitors, transistors) deviate from nominal values or design intentions due to the variety of statistical and deterministic effects being an inherent feature of the VLSI technology.

The choice from different VLSI technologies available on the market depends strongly on the ASIC application. Important parameters of the VLSI technology determining the choice can be its noise performance, integration density, radiation hardness, price etc.

CMOS technologies have rapidly dominated the market due to the low production cost and high density digital logic with the transistor dimensions that could be easily scaled down. On the other hand, the analysis of some general parameters (signal swing, noise, transconductance etc.) shows that the scaled down digital CMOS technologies become less compatible with analogue needs, and the best analogue performance for many circuits is expected for 0.6–0.8 μm gate length [1].

The second chapter analyses some aspects of designing multichannel mixed-mode ASICs in the already mentioned class of the CMOS technologies. A brief review of noise modelling in MOS transistors is presented, together with the effects of crosstalk between analogue and digital blocks placed on the same die. Since channel-to-channel matching of analogue parameters in multichannel IC is important part of the designing, problems related to random matching and systematic offsets are also discussed.

In the next two chapters we present examples of multichannel ASICs for which the discussed earlier problems have been solved at design stage. Performance of designed ASICs is verified by the measurements.

The third chapter describes the RX64 chip used for digital imaging in the Roentgen diffractometry. The chip processes signal from a silicon strip detector. Since the detector works with the low energy X-rays, amplitudes of the input signals are really low. For example a single 5 keV photon from the X-ray tube used to obtain digital image produces a charge of 1400 electrons in a silicon detector. The charge is collected by the readout electrodes in about 20 ns [2]. The integrated circuit consists of low noise analogue front-end channels for signal processing and digital blocks for data storage, control and testing.

The fourth chapter describes the NEURO64 chip used for the readout of signals from the neuronal cells, with typical amplitudes from 50 μV to 500 μV and the frequency spectrum in the range from 20 Hz to 2000 Hz [3]. The main parts of the chip are low noise pre-amplifiers, continuous time filters and an analogue multiplexer.

Both 64-channel mixed-mode chips are successfully applied in physical and biological experiments.

2. Some important aspects of low noise design of multichannel mixed-mode ASICs

Application Specific Integrated Circuits are designed taking into account very specific requirements of a given application. They are often used in modern scientific experiments where multichannel readout electronic systems open new perspectives for experimental technique. The main role of such an ASIC is to amplify and filter small analogue signals from many sensor elements in order to prepare analogue information for further processing. The information from the chip can be sent out to the external processing unit in an analogue or a digital form. In the first case the analogue signals from many channels are multiplexed to reduce the number of links for data output. In the second case analogue information is converted to the digital format directly on the chip and then is sent via a digital link to the computer.

Such multichannel chips are used in scientific experiments where the demand for the number of readout electronic channels starts from a hundred channels [4,5] and reaches several millions channels [6]. In most cases these systems are based on chips consisting of 64 or 128 readout channels on a single die. The number of readout channels on a single integrated circuit puts constraints on the maximum available area and power dissipation per single channel. Besides, to make the system effective the spread of analogue parameters from channel to channel must be minimised; so the good matching performance as the additional requirement should be taken into account starting from the early design stage.

The scientists aim to detect and collect extremely low input signals. Because the circuit noise represents a lower limit to the size of the input signal which can be processed with an acceptable quality, the low noise circuit performance becomes one of the main goals of the design. The noise can never be eliminated, since it is a fundamental property of the electrical circuit but it can be significantly reduced for example by filtering, proper circuit configurations, optimal transistor dimensions and correct choice of their operating points. In order to improve the functionality and the reliability of multichannel systems, the single chip should be equipped with digital blocks for control and testing the chip itself and even for some digital data processing directly on the chip. This results in extra crosstalk between sensitive analogue circuits and fast digital blocks, called usually as switching noise.

Taking into account above mentioned aspects, in this chapter we concentrate on three problems, important for such mixed-mode multichannel ASIC designs, viz.:

- SPICE modelling of noise in MOS transistors,
- crosstalk phenomena in mixed-mode integrated circuits,
- mismatch modelling in the MOS transistor and circuits.

2.1. Noise modelling in MOS transistors

Noise sets a lower limit to the accuracy of any measurements and to the amplitude of signal that can be processed electronically. By understanding the noise origin and models used by the circuit simulators one can significantly improve noise parameters of the circuit. Because we analyse the CMOS circuits, this chapter concentrates on a brief review of noise models for the MOS transistor, including simple models for first hand calculation, the HSPICE models and their limitation due to short channel effects.

There are several noise sources in MOS transistors:

- thermal noise in the drain-source channel,
- flicker noise in the drain-source channel,
- induced gate noise at high frequencies,
- shot noise due to the leakage current through the SiO₂ gate,
- shot noise associated with the leakage current of the drain (source) reverse biased diodes,
- thermal noise due to the source, drain and gate contact resistance.

In most cases only the first two items are important, so they are discussed in details in the next subsections. Induced gate noise becomes non-negligible at high frequencies when the MOS transistor must be considered as an RC distributed network, with the capacitive coupling to the gate representing a distributed capacitance and the channel itself representing a distributed resistance [7, 8]. The gate admittance has a capacitive and conductive components. The conductance has noise associated with it. To calculate the induced noise at higher frequency one should determine the degree of correlation between the gate and drain thermal induced noise current, since both noise currents have the same physical origin. The leakage current through the SiO₂ gate and drain (source) reverse biased diodes are in most cases negligible, and the shot noise associated with them can be also neglected. The source, drain and gate contact resistances contribute to the thermal noise and their power spectral densities are given by well known Nyquist equation.

2.1.1. Channel thermal noise

Every resistive element generates thermal noise which is caused by random motion of charge carriers in it. A noisy resistor can be modelled as a noiseless resistance R with a series noise voltage generator of the power spectral density, given by

$$\overline{\frac{dv_{th}^2}{df}} = 4kTR \quad (2.1)$$

where:

- k – Boltzman constant,
- T – absolute temperature.

According to the Norton's theorem the voltage noise source, as above, can be represented as a parallel current source with the power spectral density equal to

$$\overline{\frac{di_{th}^2}{df}} = \frac{4kT}{R} \quad (2.2)$$

The MOS transistor has a resistive channel between the drain and the source, and the random thermal motion of carriers in the channel results in thermal noise. This noise depends on bias conditions, transistor dimensions and the properties of a given technology. For the first hand noise calculation the thermal channel noise in strong inversion in the saturation region can be expressed as

$$\overline{\frac{di_{th}^2}{df}} = \frac{8}{3}kTg_m \quad (2.3)$$

where g_m is gate transconductance of the MOS transistor. In the linear region for the case, when drain-source voltage V_{DS} is close to zero, the power spectral density is given as

$$\overline{\frac{di_{th}^2}{df}} = 4kTg_{ds} \quad (2.4)$$

where g_{ds} is source-drain small signal conductance of the MOS transistor. In the weak inversion region for $V_{DS} > 5kT/q$ the adequate expression for the thermal noise is

$$\overline{\frac{di_{th}^2}{df}} = 2qI_{DS} \quad (2.5)$$

where:

- q – elementary charge,
- I_{DS} – drain-source current.

There are several channel thermal noise models used by the circuit simulators. The HSPICE MOS transistor noise model has a parameter $NLEV$ which is used to select different equations for the noise calculation [9]. If the model parameter $NLEV$ is less than 3, then the power spectral density is given as

$$\overline{\frac{di_{th}^2}{df}} = \frac{8}{3}kTg_m \quad (2.6)$$

The above equation is used both in the saturation and the linear regions, but it can lead to wrong results in the linear region of operation. The model incorrectly predicts that the thermal noise of the MOS transistor falls to zero, when $V_{DS} = 0$ (because then $g_m = 0$). If

NLEV model parameter is set to 3, HSPICE uses a different equation which is valid in both the linear and the saturation regions

$$\overline{\frac{di_{th}^2}{df}} = \frac{8kT}{3} \mu C_{ox} \frac{W}{L} (V_{GS} - V_T) \frac{1+a+a^2}{1+a} GDSNIO \quad (2.7)$$

where:

- μ – mobility,
- C_{ox} – oxide capacitance per unit area,
- W, L – width and length of MOS transistor,
- V_{GS} – gate-source voltage,
- V_T – threshold voltage,
- $GDSNIO$ – SPICE channel thermal noise coefficient (default equal to 1),
- a – parameter defined as $a = 1 - V_{DS} / V_{DSsat}$,
- V_{DS} – drain-source voltage,
- V_{DSsat} – saturation drain-source voltage.

In the linear region with $V_{DS} = 0$, we have $a = 1$ that gives [10]

$$\overline{\frac{di_{th}^2}{df}} = 4kTg_{ds} \quad (2.8)$$

For the HSPICE *NLEV* equalling 3, the noise model works reasonably well for long channel devices, but is not adequate for short channel devices.

The other equations are used by the BSIM3v3 model from UC Berkeley which has been incorporated as HSPICE Level 49 MOS model [11]. There are also two channel thermal noise models, which can be selected by the model flag *NOIMOD*. For the default the *NOIMOD* value equals 1, the power spectral density of thermal noise given by formula

$$\overline{\frac{di_{th}^2}{df}} = \frac{8}{3} kT (g_m + g_{mb} + g_{ds}) \quad (2.9)$$

where g_{mb} – body transconductance of MOS transistor.

The above equation underestimates the noise power in the linear region. For example at $V_{DS} = 0$ according to the [7, 10], the power spectral density should be $4kTg_{ds}$, while formula (2.9) predicts only the two third of it. For the *NOIMOD* equalling 2, the power spectral density is given by

$$\overline{\frac{di_{th}^2}{df}} = \frac{4kT\mu_{eff}}{L_{eff}^2} |Q_{inv}| \quad (2.10)$$

where:

- μ_{eff} – effective mobility,
- L_{eff} – effective channel length,
- Q_{inv} – total inversion channel charge.

The inversion channel charge is computed from the charge-based capacitance model equations [10], taking into account the effective channel width and length, the operation point and the bulk charge effect.

The short channel devices can have significantly higher noise than the above formulae would predict. For that reason, the noise excess factor has been introduced and it is defined as a factor by which noise power spectral density is higher than that predicted by the long channel theory. Both the models and the measurements [12, 13] show that under certain bias condition, the noise for the short channel transistor can be even an order of magnitude higher than the value expected for the long channel theory. That is due to the following high-field effects:

- at high electric field the velocity saturates, resulting in the decrease of effective mobility;
- the electric field, which is especially high close to the drain, can produce hot carriers; the hot electron effect can be modelled by an equivalent “carrier temperature” which is higher than the lattice temperature and that results in higher noise [7, 8];
- hot electron may cause impact ionisation resulting in electron-hole generation; the generated electrons are collected by the drain, while the holes form a substrate current I_{DB} ; this current is a source of shot noise with the power spectral density of $2qI_{DB}$; at the higher value of I_{DB} this noise also exhibits an excess factor; the substrate current produces the voltage drop across the substrate resistance, and then through the substrate transconductance g_{mb} contributes to the drain current noise [10].

2.1.2. Flicker noise

It is well known that the noise of the MOS transistors in the low frequency range is high due to the flicker noise, also called the 1/f noise. There are two main theories which explain the physical origin of the flicker noise. The first one is the carrier number fluctuation theory [14, 15] that assumes, that the noise is caused by the random trapping and de-trapping of the mobile carriers in the channel by the oxide traps near the Si-SiO₂ interface. According to that theory the flicker noise can be modelled by the voltage noise generator in series with the transistor gate with the voltage noise power spectral density given by the formula [10]

$$\frac{\overline{dv_{1/f}^2}}{df} = \frac{K_f}{C_{ox}^2} \frac{1}{WL} \frac{1}{f^\alpha} \quad (2.11)$$

where:

- K_f – technology dependent constant (but independent of bias condition),
- α – exponent constant closed to unity (varies in the narrow range of 0.7 to 1.2),
- f – frequency.

In the other approach, called as the mobility fluctuation model [16], the flicker noise is attributed to the mobility fluctuation. The electrons (holes) are scattered by phonons of lattice vibrations and the density of phonons fluctuates with a 1/f spectrum. The result of that theory gives the equivalent voltage noise as

$$\frac{\overline{dv_{1/f}^2}}{df} = \frac{K_f(V_{GS})}{C_{ox}} \frac{1}{WL} \frac{1}{f} \quad (2.12)$$

where $K_f(V_{GS})$ is bias dependent factor. The inverse proportionality with C_{ox} is not universally accepted [10].

The above models are generally accepted for all regions of channel inversion. In most cases it is found that the PMOS transistors have significantly less noise than the NMOS ones (by one order of magnitude or more). That is attributed to the buried channel character of the PMOS transistor, because then the channel is farther away from the Si-SiO₂ interface and the charge carriers in the transistor channel are less affected by the interface traps. In modern technologies, where the buried channel effect is no longer present, the flicker noise for the PMOS transistors can be comparable with that of the NMOS ones [1].

The HSPICE MOS transistor model for the flicker noise simulation uses a current noise source placed between the drain and the source. That allows one to calculate the power spectral density of the total noise as a sum of white and flicker noise components, given by

$$\frac{\overline{di_n^2}}{df} = \frac{\overline{di_{th}^2}}{df} + \frac{\overline{di_{1/f}^2}}{df} \quad (2.13)$$

To obtain the representation of the equivalent input noise voltage source in series with the transistor gate, one should use the following relation

$$\frac{\overline{dv_n^2}}{df} = \frac{1}{g_m^2} \frac{\overline{di_{total}^2}}{df} \quad (2.14)$$

For the flicker noise two parameters are used by the HSPICE models: flicker noise coefficient KF and flicker noise exponent AF (default equals 1). The model equations are selected by the $NLEV$ parameter. For the $NLEV = 0$ the power spectral density of the flicker current noise is equal to

$$\frac{\overline{di_{1/f}^2}}{df} = \frac{KF}{C_{ox}} \frac{I_{DS}^{AF}}{L_{eff}^2} \frac{1}{f} \quad (2.15)$$

For the $NLEV = 1$ L_{eff}^2 in the above equation is replaced by $W_{eff}L_{eff}$, where W_{eff} is the effective channel width.

For the *NLEV* equalling 2 or 3, the equation is the following

$$\frac{\overline{di_{Vf}^2}}{df} = \frac{KF}{C_{ox}} \frac{g_m^2}{W_{eff} L_{eff}} \frac{1}{f^{AF}} \quad (2.16)$$

The BSIM3v3 model uses a rather complicated formula for the flicker noise calculation with three parameters, taking into account the effective channel width and length and the effective mobility at given bias conditions [11]. The BSIM3v3 model uses different equations for strong and weak inversion regions.

The flicker noise is also sensitive to the small dimension effects. Hot carriers can degrade the quality of the Si-SiO₂ interface and introduce additional interface traps close to the drain region. If the transistor operates in nonsaturation region, these traps affect the channel charge and the current, and that results in significant increase of the flicker noise [17, 18].

Another effect that influences the noise performance of small area transistors is the capture and emission of carriers at single individual traps in the Si-SiO₂. Because for a small area transistor there are only a few traps which can exchange the charge with the channel, their individual effects become non-negligible and the discrete modulations of the source-drain conductance become visible in the form of random telegraph signals (RST) [19]. Variation of the drain current due to the RST effect can rise up to 0.1% or more of the DC current value [10]. For a large gate area with a large number of charge carriers and traps involved, neighbouring traps may interact and the channel current non-uniformity is smeared out [20].

2.2. Crosstalk in mixed-mode integrated circuits

With increasing integration density of VLSI circuits there appears a trend to integrate more and more functionality on a single chip. Such solutions require Application Specific Integrated Circuit designs, which contain both high performance analogue circuits and advanced high speed digital blocks on the same die. The implementation of a complex system on a single chip brings several advantages such as reduction in size, power dissipation, package, cost and increase in the speed of operation, flexibility and reliability of the system.

A major disadvantage in such a mixed-signal system is the increased interaction (crosstalk) between different parts of the circuit on the same die. A fast switching transient produced in the digital circuit can corrupt analogue sensitive components and impair the performance of the entire system. The problem is greatly aggravated in modern technologies, because of higher rates of circuits operation, greater density of elements per chip and smaller feature sizes [21].

There are three main rules to minimise the crosstalk on mixed-mode chips:

- reduce the amount of generated switching noise,
- increase the immunity of the analogue part,
- introduce proper isolation between the analogue and the digital part.

To follow these rules one should take into account the proper system solutions at an early stage of design (differential analogue section, low power digital design), the layout solutions (floor planning, power distribution, guard rings) and the aspects of packaging. For better understanding of these problems, let us analyse at first the mechanisms by which the switching noise is introduced and transferred in the mixed-mode integrated circuit.

2.2.1. Generation, transmission and reception of switching noise

The working digital cells are the sources of switching noise. The turning on or off the digital gates is connected with charging or discharging capacitors during a short period of time, and that generates current spikes in the circuit. A shorter switching time, a higher value of the total capacitance to overload and a larger voltage swing result in the higher switching noise. The noise can be transferred to the analogue blocks in two major ways: through the common power supply lines or through the common substrate shared by the analogue and the digital circuits.

The crosstalk through the common supply lines is often called a supply bounce [22]. When many gates change states, a large cumulative current spike flows through parasitic inductance and resistance of bias lines creating power supply voltage spikes. The main source of the series inductance in the integrated circuits is the parasitic inductance of the bond wires and package leads. The current spikes drawn from the supply lines of inductance L_{ind} generate the voltage drop $V=L_{ind} di/dt$, and as a consequence these supply lines can be very noisy. Because the analogue blocks have always a limited Power Supply Rejection Ratio (PSRR) the disturbances on the supply lines can corrupt their achievable accuracy. The main method to minimise the supply bounce is to reduce the value of the power supply connection inductance.

Even if power supply noise is significantly reduced, there is a second way of the crosstalk due to the non-ideal isolation provided by the common substrate. Every switching activity of the digital blocks injects a current into the substrate and causes the fluctuation of the substrate voltage. This is known as a substrate noise [23], although it is not a real noise. Because of the non-zero dielectric constant and the conductivity of the substrate material the parasitic currents can reach different parts of the chip.

There are three main mechanisms for injecting substrate noise. The first one is the capacitive coupling from switching nodes of active and passive devices. For example, in the case of the NMOS transistor made in p-type substrate, noise is coupled to the substrate via source/drain-to-substrate capacitance, while the vertical NPN transistor interacts with the substrate through the collector-to-bulk junction capacitance [24].

The second source of substrate noise is the coupling from the clock lines and power supply lines through the line-to-substrate capacitance. Moreover in most cases the digital ground is connected to the substrate in every standard digital cell, and then the supply bounce on digital ground is directly coupled to the substrate [25].

The third mechanism responsible for the noise injection is the impact ionisation in MOS transistors. For the short channel devices working in saturation, the electrical field strength in the drain-end of the channel can be sufficiently high to cause impact ionisation and to generate electron-hole pairs. The holes form a drain-to-substrate current which is always positive for both 1–0 and 0–1 transitions in the drain node. The effect of the impact ionisation is often taken into account by the SPICE transistors models provided by the

foundry. Whether or not the impact ionisation is an important source of the substrate noise depends on the technology, especially on the combination of the supply voltage and the channel length. For example, for the epi-type 0.5 μm 3.3 V technology presented in [25], the substrate current caused by impact ionisation is negligible.

Transfer of the switching noise through the substrate depends strongly on its electrical properties. In most cases modern CMOS technologies use one of two kinds of substrate: a low resistivity substrate or a high resistivity substrate, shown schematically in Figure 2.1. In a high resistivity substrate the bulk region of 200–400 μm thick is lightly doped silicon of resistivity about $\rho = 20 \Omega\cdot\text{cm}$ with a thin (1 μm) implant (or epi) layer on top. In a low resistivity substrate the bulk is heavily doped and has very low resistivity of the order of 1 $\text{m}\Omega\cdot\text{cm}$. Above it there is an epi-type layer of the thickness $T_{\text{epi}} = 5\text{--}10 \mu\text{m}$ with the resistivity in the range $\rho = 10\text{--}15 \Omega\cdot\text{cm}$. For the p-type bulk there is also a thin surface implant on the top (known as p-tub or channel stop) to avoid parasitic inversion of the silicon by potential of the lowest metal layers. Technology based on the low resistivity substrate is the one mostly used for the digital CMOS design due to its immunity to latch-up.

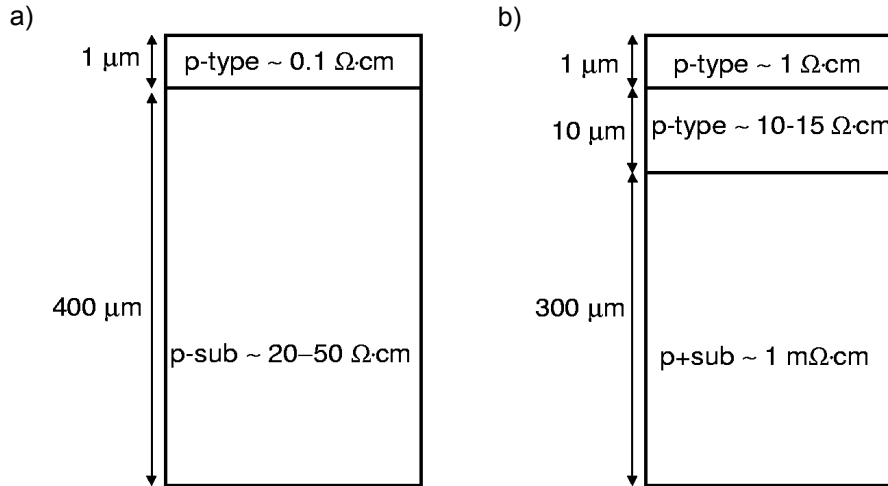


Fig. 2.1. Substrate cross-section: a) high resistivity substrate; b) low resistivity substrate

Current density J flowing through the substrate can be expressed by the formula

$$J = (\sigma + j\omega\epsilon_s)E \quad (2.17)$$

where:

- σ – conductivity,
- j – imaginary unit,
- ω – angular frequency,
- ϵ_s – dielectric permittivity of silicon,
- E – electric field.

For both, high and low resistivity substrates with a typical material resistivity from $1 \text{ m}\Omega\text{cm}$ to $20 \text{ }\Omega\text{cm}$, the capacitance of the substrate can be neglected for the frequency range below GHz [24]. Therefore, it is adequate to treat the substrate as a distributed network of simple resistors.

After fabrication a chip is mounted on board or in package. The mounting glue can be conductive or non-conductive, and then we are talking about conductive or non-conductive backside contact. The kind of backside contact has influence on substrate crosstalk.

One of the methods to simulate the substrate crosstalk is to use SPICE models of electronic elements and a proper model of the bulk. Such simulations have to be done for the specific layout of the circuit taking into account the positions of each electronic block and the distribution of supply and control lines. In the case of the high resistivity bulk the substrate is modelled by a simple resistive mesh. The effective resistance between two contacts on such substrate (which can represent, for example, the position of analogue and digital blocks) increases with the distance for non-conductive backside. For the conductive backside, some of the current flows through the backside contact and limits the isolation for the distance comparative with the wafer thickness [26].

The situation is different for the low resistivity bulk. The epitaxial layer is modelled as a set of vertical and horizontal resistors, while the heavily doped bulk region is considered as a single electrical node. For two contacts placed at the distance smaller than $4T_{epi}$, a significant portion of the total substrate current flows in the epitaxial layer. For the distance larger than $4T_{epi}$, the total current flows vertically in the epitaxial layer and then through the heavily doped bulk over the entire chip. For that reason the isolation between two contact ceases to be the function of distance for the separation greater than $4T_{epi}$ [23]. The situation is slightly modified by the skin effects for the frequencies above GHz [27].

The analogue blocks pick up the switching noise directly from the power supply lines or from the substrate by capacitive sensing or body effect. Power supply rejection ratio coefficients from positive and negative bias lines decide about the influence of the supply bounce on the analogue performance. The substrate noise is picked up by the components capacitively coupled to the substrate. This concerns transistors via the junction capacitances, passive components (polysilicon capacitance, resistors etc.), interconnect lines and input pads. For the MOS transistors the principal coupling mechanism occurs through the body effect. The threshold voltage depends strongly on the substrate potential and that makes the drain current dependent on the substrate noise.

The switching noise coupling degrades the performance of the analogue blocks. If the switching noise is of the order of magnitude or higher than thermal, shot or flicker noise of electronic devices, the analogue circuit losses accuracy, dynamic range, gain or bandwidth [24]. In addition to the above effects, one should take into account the possibility of circuit oscillation. The total on-chip power-to-ground capacitance and the inductance of the bond wires form a resonant LC circuit, which can oscillate at higher frequency (for example the total capacitance of 10 pF and the bond wires inductance of 3 nH give the resonant frequency about of 900 MHz). A remedy for the problem is to reduce the Q-factor of such a circuit by using series resistance or RLC parallel branch, as it has been proposed in [28]. There is also another source of possible oscillation when the high gain amplifiers are implemented. The substrate or the power supply lines on the chip can create a positive feedback path for the signal that leads to the loss of circuit stability. To cut the positive feedback path, proper isolation techniques discussed in detail in section 2.2.4 should be applied.

2.2.2. Reducing the noise generation

To minimise the crosstalk one should start from reducing the noise generation in the digital blocks. The switching noise is generated at the slopes of switching signals, so the peak-to-peak amplitude of generated noise strongly depends on the rise and fall times, and on the amplitude of switching signals [29]. The number of generated noise pulses increases with the switching frequency, and with the number of switching gates. The best solution but not practical would be to stop switching the voltage and the current in a digital block. The designer can control the time when switching occurs, and in that way can separate the critical analogue function from the moment of the noise generation in digital blocks. For example, for the proper timing in analogue to digital converter one should introduce a phase shift between a digital clock and an analogue sampling, to avoid comparison or sampling when large digital drivers switch [27]. All switching functions, logic blocks or drivers not in use should be turned off. The gated clock systems are preferable while the using of synchronous logic (with a large number of gates switching simultaneously) should be minimised.

There are several digital design styles which generate more or less switching noise. The static CMOS, for which maybe the main advantage is the low power consumption, is one of the most noisy circuit topologies: large rail-to-rail switching voltages are generated and large current spikes are drawn from the power supply. Similar situation is for the case of self-resetting logic which adds variable high frequency reset noise [30]. There are other design styles, that generate less noise. The best are those based on balanced current steering with the constant current consumption as ECL (Emitter-coupled Logic), CML (Current Mode Logic) or SCL (Source-coupled Logic). Preferable solutions are also those with smaller than rail-to-rail voltage swing like LVDS (Low Voltage Differential Standard). Because the low noise logics in most cases have non-negligible power consumption, they are not useful for making large digital integrated circuits and their application is limited to small blocks.

Special attention has to be paid to the nodes with large parasitic capacitance to the substrate (large fanout) like buses, I/O drivers, clock distribution networks. The rise and fall times for drivers of these nodes should be as large as the design constrains allow [22]. The voltage swing on these nodes should be minimised, while the distribution of the digital signals and clocks in complementary form reduces the net amount of coupling noise.

2.2.3. Increasing the immunity of analogue part

Analogue blocks should be designed to be as insensitive as possible to the injection of noise through the supply lines and the substrate. High PSRR (Power Supply Rejection Ratio) is necessary to minimise the coupling of supply bounce into the signal path. In the case when it is impossible to obtain sufficient PSRR, the weak path for the supply coupling must be identified to be able to apply isolation rules discussed in the section 2.2.4. Because the substrate noise can be treated as a common mode signal, fully differential circuits are preferred to single ended stages [29]. Such differential structures should have a high CMRR (Common Mode Rejection Ratio), so special attention has to be paid to the symmetry in the layout and matching of critical components. Mismatches between circuit elements convert a fraction of common mode noise into differential signals and significantly reduce the isolation. Sometimes other parameters of the design force the single ended signal processing. In that case the transmission of the signals to another subcircuit or to the outside world,

should be made concurrently with their reference through the dedicated path, rather than relay on the common ground [28]. All sensitive high-impedance nodes have to be kept inside the chip, because the chip parasitics are much smaller than those of package or printed circuit board. The use of the minimum required bandwidth for the signal processing in a switching noise environment is always an important aspect of design optimisation.

In the case of the p-type substrate, PMOS transistors are in n-well, which connected to a clean power supply can be used to shield the transistors from the substrate. Therefore the PMOS transistors are preferable for the signal handling [31]. The NMOS ones can be used as components of the DC current sources, and it is better to refer them to the substrate, than to the clean supply.

2.2.4. Isolation techniques

The proper isolation techniques should be used to minimise the switching noise transfer from digital to analogue blocks. The first step is to limit the supply bounce by reducing the parasitic inductance and resistance of connections between the chip and the external world [25]. These parasitics depend on the type of package, type of bonding and the number of pads used for the power supply lines. For the bonding wire connections the typical inductance is 2.6 nH per single wire (25 μm diameter and 2.5 mm long) while the flip-chip connection has the inductance below 0.2 nH [32]. To decrease the inductance of the bonding wire connection, one can use multiple bonding pads and the bonding wires for each power supply bus.

It can occur that even after the above steps the power supply lines on the chip still generate a large amount of noise. In that case, if there is enough place on the chip, it is reasonable to separate power supply buses, pads and bonding wires for noisy digital and sensitive analogue blocks. Some designs may require that sets of power and ground connections should be multiplied in order to further divide circuit blocks, for example fast output drivers [22].

The distribution of power supply on the chip is closely related to careful floor planning. The analogue blocks need to be categorised by the sensitivity to the switching noise, while the digital blocks need to be classified by the amount of generated noise. The most sensitive analogue circuit (i.e. high gain preamplifier) should be as far as possible from the noisy digital circuit (i.e. output buffers), and then the least sensitive analogue blocks should be placed next to the least offensive digital circuit [33]. The power supply lines must be sufficiently wide, especially in large chips, to avoid unwanted voltage drops across the chip while the current loop area should be kept small. The floor planning should ensure the proper signal routing and pads assignment. Digital signals should not be routed over an analogue portion of the chip or close to sensitive lines. Sensitive and noisy pads must be kept from each other as far as possible due to mutual inductance of the bonding wires or package pins.

Placement and connecting of substrate contacts are critical for proper isolation. Here techniques applied are different for high and low resistivity substrates, but in both cases every effort should be made to connect the substrate near the sensitive analogue region to a quiet supply [22]. The high resistivity substrate requires for the latch-up protection that the multiply substrate contacts should be tied to the ground periodically. The substrate needs to be split into quiet and noisy regions. Analogue and digital regions should have separate substrate contacts with separate bonding pads for the ground. The backside contact

to the low impedance ground is recommended, but to make it effective, it usually requires the wafer to be thinned [30].

For the low resistivity substrate the best solution is the placement of substrate contacts in the analogue part of the chip with a connection to a separate substrate pad, or if it is not possible, to the analogue supply [22]. In that case a low impedance connection for the substrate is of great importance, so that the isolation can be improved further on by thinning the wafer and adding the backside low impedance contact tied to the quiet ground by conductive epoxy. Some standard cell libraries connect automatically digital ground to the substrate, but it is not recommended because the low impedance path is formed between the quiet analogue and the noisy digital ground buses [30].

The using of guard rings can reduce transfer of switching noise. There are two kinds of these. Let us consider a p-type substrate. A guard ring may be a simply continuous ring of substrate contacts (p+ diffusion) that surrounds the circuit, providing a low impedance path to ground for the charge carriers produced in the substrate. Guard ring can be also formed as n-well ring, and then it stops the noise currents flowing near the surface. Methods of using both of them depend strongly on the kind of substrate.

In the high resistivity substrate, because the substrate current flow is lateral and concentrated near the surface, the guard rings are very effective, especially if biased via separate bonding pad. In the case of p-type substrate, the p+ ring diffusion can reduce the coupling of the switching noise by almost an order of magnitude [23]. The guard rings act as a very good current sink, and it is better to put them close to the protected objects. The best solution is to put two rings around analogue and digital blocks with separate package pins. The floating guard rings are less effective, while tied to the power supply buses they can even hurt. The n-well rings are also recommended, because they break the low resistivity surface implant layer and force the substrate current to flow in the substrate underneath the well where the resistivity is significantly higher [26].

In the low resistivity substrate most of the current flows vertically to the low resistivity bulk and then through the heavily doped bulk over the entire chip. The p+ guard rings reduce the substrate crosstalk only by about 20% when placed close to the analogue circuits and biased via the dedicated pins. The p+ guard rings can be used separately for analogue and digital blocks, but then the separate bonding pads for each of them are necessary. Connecting the p+ guard ring to the digital ground or large substrate contact results in the increase of the observed noise [28]. The n-well guard ring in the low resistivity bulk has nearly no effect [23].

Guard rings are not the only way of shielding. Diffusion, implant, polysilicon or metal layers tied by the low impedance connection to the non-contaminated with noise power supply can form vertical Faraday shielding. The n-wells can protect the device inside very effectively and for that reason the PMOS transistors are recommended for signal handling. Large area sensitive devices (even input pads) and noisy routing channels for clocks should be vertically shielded [34, 35].

Using of different power supply filters is also important. In most cases it is done by adding a large decoupling capacitance between the power supply and ground on the chip and off the chip on the board. The capacitance on the chip can be obtained by stacking supply rails, using MOS or polysilicon capacitors. Simple capacitance can be replaced by more efficient RLC filters [28]. There are other sophisticated techniques as “active supply bypass” [36] or “active guard ring” [37] which can be also taken into consideration.

2.3. Random matching and systematic offsets

The physical circuits components (i.e. resistors, capacitors, transistors) deviate from nominal values or design intentions due to the variety of statistical and deterministic effects. That results in the variation of electrical parameters for nominally identical devices and is a limiting factor in many precision analogue circuits [38], multiplexed analogue systems [39], digital to analogue converters [40], multichannel systems [41], reference sources [42], memory sense amplifiers [43], etc.

One should distinguish between systematic offset and matching. Systematic offset is the effect of asymmetry in circuit configuration, bias condition and layout. The obtaining of zero systematic offset is a matter of good engineering design. Mismatch is the process that causes time-independent random variations of physical parameters of identically designed devices. Mismatch as an inherent feature of the VLSI technology cannot be omitted, but it can be understood, and its influence on the performance of ICs can be minimised. The impact of mismatch of MOS transistors becomes more and more important as the dimensions of the devices are reduced and the available signal swing decreases.

There are numbers of theoretical approaches to mismatch modelling based on certain assumptions on the behaviour of defects that cause mismatch [42, 44, 45, 46, 47]. In the paper [44] the mismatch in capacitors and current sources is analysed in terms of the local and global variation. The random errors in MOS transistors are caused by random length and width transistor variations, surface state charge and ion implanted effects, fluctuation of the oxide thickness and variations of the effective channel mobility. Another paper [45] starts again from the possible physical causes of mismatch and describes the local variation of MOS transistors by means of threshold voltage and current factor standard deviations. A new approach proposed in [42] involves spatial Fourier transform technique to analyse mismatch effects. The variations in the threshold voltage, the current factor and the substrate factor of the MOS transistor are measured as a function of area, distance and orientation. There are also some new models proposed in [46, 47], but they are not so common as those proposed in [42, 45].

The model of random matching is applied to the parameters of equally designed devices on the same chip as the result of unavoidable variation during IC fabrication. The matched devices should be of equal nature, have the same layout, identical surrounding and be used identically (bias voltages, currents and temperature). Two kinds of errors can be distinguished, local and global variations.

Short distance variations related to the short distance effects have the following features [42]:

- the total mismatch of the parameter P is composed of many single events of the mismatch generating process,
- the effects of a single event on a parameter are so small that the contributions of many events to the parameter can be summed,
- the events have a correlation distance much smaller than the device dimensions.

The amplitude of these short distance variations has a normal distribution with mean value around zero.

The second class of mismatch is related to the different gradients across the wafer (for example in oxide thickness), which originates from wafer fabrication process. That can be modelled as an additional stochastic process with a long correlation distance.

Consequently, with the above assumptions the variation of parameter ΔP can be expressed as [42]

$$\sigma^2(\Delta P) = \frac{A_p^2}{WL} + S_p^2 D^2 \quad (2.18)$$

where:

- A_p – area proportionality constant for variation of parameter P ,
- W, L – width and length of rectangular devices,
- S_p – constant of variation of parameter P with distance,
- D – spacing distance.

Such definition of random matching excludes batch to batch or wafer to wafer variations. The above equation illustrates two main rules of matching:

- the devices with the larger area $W \times L$ match better,
- matched devices should be placed at small spacing distance D .

Increase of the size of the devices is always a good option for better matching, but is often in opposition to the high speed and small chip area requirements. So, at first one should find matching sensitive devices in the circuit and then consider a compromise between matching and other design constraints.

2.3.1. The mismatch parameters of MOS transistors

There are three main parameters the variations of which cause the mismatch of MOS transistors at DC and low frequencies: threshold voltage V_T , body factor γ and current factor β . The threshold voltage of the NMOS transistor built on p-type substrate can be expressed as [48]

$$V_T = V_{T0} + \gamma \left(\sqrt{2|\phi_F| + V_{SB}} - \sqrt{2|\phi_F|} \right) \quad (2.19)$$

with V_{T0} and γ equalling:

$$V_{T0} = \phi_{MS} + 2|\phi_F| - \frac{Q_{ox}}{C_{ox}} + \frac{\sqrt{4q\epsilon_s|\phi_F|N_B}}{C_{ox}} \quad (2.20)$$

$$\gamma = \frac{\sqrt{2q\epsilon_s N_B}}{C_{ox}} \quad (2.21)$$

where:

- V_{T0} – threshold voltage for $V_{SB} = 0$,
- V_{SB} – source-bulk voltage,

- γ – body factor of MOS transistor,
- ϕ_F – Fermi potential in the bulk,
- ϕ_{MS} – gate-semiconductor work function difference,
- N_B – bulk doping density,
- Q_{ox} – fixed oxide charge density,
- ϵ_s – silicon permittivity,
- C_{ox} – gate oxide capacitance per unit area equalling ϵ_{ox}/t_{ox} ,
- ϵ_{ox} – silicon dioxide permittivity,
- t_{ox} – oxide thickness.

Let us analyse the contributions of possible errors to the threshold voltage mismatch:

- because ϕ_{MS} and ϕ_F have logarithmic dependence on doping in the substrate [49], they can be considered as constants and not contributing to any mismatch,
- in the well controlled process the oxide thickness t_{ox} is reproducible and its contribution to the variation of V_T can be ignored,
- the density of fixed oxide charges Q_{ox} in the modern VLSI MOS processing is much smaller than the depletion charge density equal to $\sqrt{4q\epsilon_s\phi_F N_B}$ (see the third and fourth term in equation (2.20)), and one can say that the variation of bulk doping level N_B is a dominant source of mismatch,
- the variation of body factor γ (caused mainly by the variation of the bulk doping level N_B and oxide thickness t_{ox}) is important only for the transistors bias with non-zero V_{SB} voltage.

In the case of the PMOS transistor, additional term equal to qD_I/C_{ox} should be added in equation (2.20), where D_I is the threshold adjust implant dose [45]. Because the main factors causing mismatch of the threshold voltage and bulk factor satisfy in the first order approximation the assumptions of the above model of random matching, according to equation (2.18) the standard deviation of V_{T0} and γ can be written as [42]

$$\sigma^2(V_{T0}) = \frac{A_{VT0}^2}{WL} + S_{VT0}^2 D^2 \quad (2.22)$$

$$\sigma^2(\gamma) = \frac{A_\gamma^2}{WL} + S_\gamma^2 D^2 \quad (2.23)$$

where:

- W, L – width and length of MOS transistor,
- A_{VT0} – area proportionality constant of variation of threshold voltage V_{T0} ,
- S_{VT0} – constant of variation of threshold voltage V_{T0} with distance,
- A_γ – area proportionality constant of variation of body factor γ ,
- S_γ – constant of variation of body factor γ with distance.

The current factor β is given by

$$\beta = \mu C_{ox} \frac{W}{L} \quad (2.24)$$

The mobility μ , oxide capacitance per unit area C_{ox} and the dimensions W and L of the MOS transistor are all independent. For example, the definition of width W and length L is determined by different steps in different conditions during IC processing, so they may be treated independently. With that assumption the variance of current factor β can be expressed as

$$\frac{\sigma^2(\beta)}{\beta^2} = \frac{\sigma^2(\mu)}{\mu^2} + \frac{\sigma^2(C_{ox})}{C_{ox}^2} + \frac{\sigma^2(W)}{W^2} + \frac{\sigma^2(L)}{L^2} \quad (2.25)$$

Physical effects responsible for the variation in the mobility and oxide capacitance can be treated according to equation (2.18) while the variation in transistor dimensions requires some additional comments. Variations of the width and length originate from the variations in photolithographic process. From one dimensional analysis of random error due to the edge roughness one can find that $\sigma^2(L) \sim 1/W$ and $\sigma^2(W) \sim 1/L$. Then equation (2.25) can be re-written as

$$\frac{\sigma^2(\beta)}{\beta^2} = \frac{A_\mu^2}{WL} + \frac{A_{Cox}^2}{WL} + \frac{A_W^2}{W^2L} + \frac{A_L^2}{WL^2} + S_\beta^2 D^2 \quad (2.26)$$

where:

- A_μ, A_{Cox}, A_W, A_L – technology dependent constants,
- S_β – constant of variation of current factor β with distance.

The paper [42] suggests that for W and L large enough, the above equation can be approximated as

$$\frac{\sigma^2(\beta)}{\beta^2} \approx \frac{A_\beta^2}{WL} + S_\beta^2 D^2 \quad (2.27)$$

where A_β is area proportionality constant of variation of current factor β .

Some other authors [45, 50] come to conclusion that the relative variation of current factor should be rather scaled with $1/(W^2+L^2)$ than with $1/(WL)$. Moreover the experimental results [51] show that even for equal area transistors, shorter channel lengths and wider channel widths result in poorer matching than longer channel lengths and narrower channel widths.

The variation of gate oxide capacitance is a common factor in V_T and β , so one could expect the correlation in the mismatches in V_T and β . However, both the theoretical and experimental values of the correlation coefficient are close to zero [45], so it can be assumed that the variations of V_T and β are almost independent.

2.3.2. Matching in various processes

Exemplary matching constants according to the equations (2.22), (2.23) and (2.27) for NMOS and PMOS transistors (with) are shown in Table 2.1 [42].

Table 2.1
Matching data for NMOS and PMOS transistor pairs
in 2.5 μm n-well process, with 50 nm gate oxide

Parameter	NMOS	PMOS	Unit
A_{VT0}	30	35	mV μm
A_β	2.3	3.2	% μm
A_γ	16×10^{-3}	12×10^{-3}	$\text{V}^{0.5} \mu\text{m}$
S_{VT0}	4	4	$\mu\text{V} / \mu\text{m}$
S_β	2	2	$10^{-6} / \mu\text{m}$
S_γ	4	4	$10^{-6} \text{V}^{0.5} / \mu\text{m}$

The above table shows the tendencies which are common for many n-well CMOS processes:

- the matching of V_T and β is better in the case of NMOS transistors compared to the PMOS ones (n-well technology),
- the mismatch of the NMOS transistors is more sensitive to the non-zero V_{SB} voltage (larger A_γ factor),
- the relative effect of the increased distance between the matching devices is significant only for large area devices with a considerable spacing.

By scaling the technology down we not only change the minimum dimensions of the transistor but also reduce the oxide thickness. Let us consider the V_T matching for NMOS transistors (for $V_{SB} = 0$). Mismatch of V_T is dominated by the variation in bulk doping level N_B (the fourth term in the equation (2.20)) and according to [1] can be expressed as

$$\sigma(V_{T0}) \approx \text{const} \frac{t_{ox}^4 \sqrt{N_B}}{\sqrt{WL}} \quad (2.28)$$

Improvement of matching due to the thinner oxide layer is slowed down by the increase of the doping level under the oxide. For the PMOS transistors the situation in V_{T0} matching is more complicated and strongly depends on the technology used. Table 2.2 presents some data for the modern CMOS technologies.

By scaling the technology down we don't observe any improvements in β variation, i.e. for NMOS transistors scaled down to the oxide thickness from 50 nm to 12 nm the A_β parameter is still around 2% μm [52]. Generally, in submicron process for minimum dimension transistors matching stays roughly the same and it improves for larger than minimum dimension transistors.

Table 2.2
Matching of V_T for NMOS and PMOS for different technologies

Technology / oxide thickness	A_{VT} for NMOS [mV μm]	A_{VT} for PMOS [mV μm]	Reference
0.8 μm /15 nm	10.7	18	[1]
0.6 μm /12 nm	11.0	8.5	[1]
0.25 μm /6 nm	6.0	20	[52]
0.25 μm /4.4 nm	3.6	–	[43]
0.18 μm /3.3 nm	3.4	–	[52]

2.3.3. Current matching in MOS transistors

In the analogue circuits the transistors operate mostly in the saturation region in strong inversion where the drain current is given by

$$I_{DS} = \frac{\beta}{2} (V_{GS} - V_T)^2 \quad (2.29)$$

As the correlation coefficient between mismatch in V_T and β is nearly equal to zero, the relative source-drain current I_{DS} mismatch may be written as

$$\frac{\sigma^2(I_{DS})}{I_{DS}^2} = 4 \frac{\sigma^2(V_T)}{(V_{GS} - V_T)^2} + \frac{\sigma^2(\beta)}{\beta^2} \quad (2.30)$$

At a low value of $(V_{GS} - V_T)$, the dominant factor in the mismatch comes from the variance of V_T , while for the higher $(V_{GS} - V_T)$ the variance of current factor β dominates. The better matching is obtained at the higher difference $(V_{GS} - V_T)$, which means the higher current. At the low value of V_{GS} the mismatch does not go to infinity, because transistors enter the weak inversion region where square law model (see equation (2.29)) is no longer valid and I_D depends exponentially on V_{GS} . The relative drain current mismatch $\sigma(I_D)/I_D$ in weak inversion is independent of V_{GS} and thus of the current, and typically stabilises on the level of 3%–3.5% [53].

2.3.4. Random matching in circuits

One of the main requirements for practical completion of a multichannel VLSI chip is the uniformity of analogue parameters for all channels. The spread of parameters as gain, offset, cut-off frequency, noise, etc., should be minimised to the acceptable by a given application level. The spread of these analogue parameters depends not only on the geometry of the layout, but also on the bias conditions. Solutions with the possibility of tuning some of these analogue parameters for each channel independently are also possible [54], but they cost an extra area of silicon and additional complications of chip architecture.

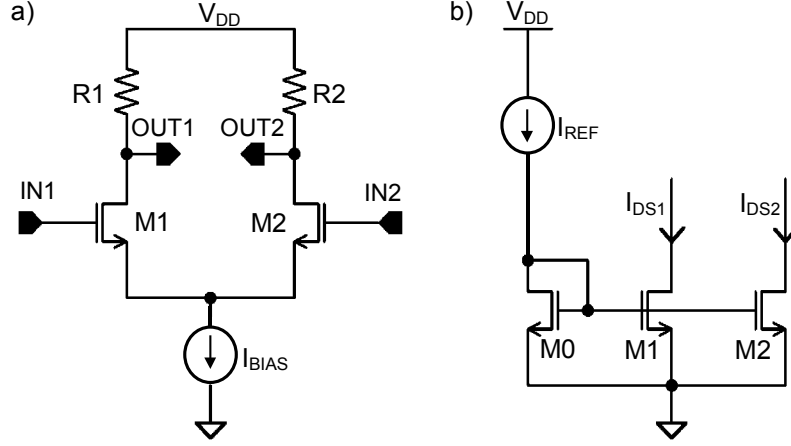


Fig. 2.2. Mismatch in circuits: a) differential pair; b) current sources

To illustrate the dependence of matching of analogue parameters on the bias conditions, let us consider two exemplary circuits shown in Figure 2.2, a differential pair and a current mirror which are often used in analogue circuits. For a differential pair both the input transistors and the load resistors suffer from mismatching such as ΔV_T , $\Delta\beta$, ΔR . The device mismatches are incorporated as $V_{T1}=V_T$, $V_{T2}=V_T\pm\Delta V_T$, $\beta_1=\beta$, $\beta_2=\beta\pm\Delta\beta$, $R_1=R$, $R_2=R\pm\Delta R$. When both transistors operate in strong inversion, the total random offsets V_{osr} referred to the input can be expressed as [50]

$$V_{osr} = \pm\Delta V_T + \frac{V_{GS} - V_T}{2} \left(\pm \frac{\Delta R}{R} \pm \frac{\Delta\beta}{\beta} \right) \quad (2.31)$$

Equation (2.31) shows that random offset depends both on device mismatches and bias conditions. The contributions of mismatches ΔR and $\Delta\beta$ increase with the transistor overdrive ($V_{GS}-V_T$), while the threshold voltage mismatch ΔV_T is directly referred to the input. To minimise the offset it is desirable to use low values of ($V_{GS}-V_T$) by lowering the bias current or increasing the W/L ratio. The random mismatch results not only in the random offset but it also reduces the common-mode rejection ratio of the differential amplifier [50].

Let us analyse the mismatch of the nominally identical current sources M1 and M2 shown in the Figure 2.2b. Let us assume that both transistors work in strong inversion with the same drain source voltage $V_{DS1}=V_{DS2}$. The device mismatch is incorporated again as $V_{T1}=V_T$, $V_{T2}=V_T\pm\Delta V_T$, $\beta_1=\beta$, $\beta_2=\beta\pm\Delta\beta$, that results in different currents $I_{DS1}=I_{DS}$ and $I_{DS2}=I_{DS}\pm\Delta I_{DS}$. The relative variation of the output currents can be written as [50]

$$\frac{\Delta I_{DS}}{I_{DS}} = \pm \frac{2}{V_{GS} - V_T} \Delta V_T \pm \frac{\Delta\beta}{\beta} \quad (2.32)$$

To minimise the random current variation the overdrive voltage ($V_{GS}-V_T$) should be maximised. That is in the opposite to the random offset in the differential amplifier (see equation (2.31)). In practice there is also a systematic current variation for the case when $V_{DS1} \neq V_{DS2}$, coming from the finite output resistance of the current source.

2.3.5. Layout rules for good matching

Symmetry is the main rule in the layout to improve matching. Symmetry means not only the same layout of matched devices, but also the same temperature, bias, parasitics, metal coverage and the surrounding environment.

There are several layout design rules that help maintain the symmetry.

The same bias and signal delay. Asymmetry in the circuit configuration or bias point is a source of systematic offset. One should realise that the parasitic components as resistance of paths or contacts and parasitic capacitance modify the electrical scheme of the circuit. Voltage drops on parasitic resistance of the power supply lines, in the source of a differential pair, on the voltage reference distribution lines, etc., can be devastating. The unwanted voltage drops can be minimised by proper metal width, extra contacts and using the star connection to maintain symmetry of current paths. In digitally controlled analogue circuits particular attention should be paid to the clock and signal skew due to parasitic resistance and capacitance [55]. Sometimes manual correction must be made in electrical scheme for the proper evaluation of these effects.

The same temperature. The current of the MOS transistor or the resistor value depend on temperature. For the MOS transistor the main temperature dependent parameters are the mobility μ and the threshold voltage V_T . Near room temperature the temperature coefficient for the threshold voltage is in the range 0.5–3 mV/°C [10] while the mobility depends on the temperature as $\mu \sim T^{-1.5}$ [49]. So, the matched devices must operate at the same temperature, which is not a problem if the power dissipated by each block on the chip is low. Otherwise one should identify the large power sinks on the chip and the distribution of isotherms on the chip. If it is possible, the sensitive circuits should be placed at the largest possible distance from power blocks and realigned on the same isotherms.

The same orientation. IC processes exhibit anisotropy which results from certain processing steps or lattice orientation and has great influence on matching. For that reason it is important how the devices are oriented on the layout relative to one another and what are the relative directions of currents flow. Comparative measurements of the standard deviation of current factor β matching for transistor placement rotated by 90 degrees show that the matching is several times worse than in the case of parallel transistor placement [42]. Because the threshold and substrate factor mismatch is identical for the rotated and parallel placements, so the local mobility variations can be a possible explanation for the rotation dependent mismatch. In the case of parallel placement the better matching is for the transistors with the same direction of current flow. This is due to the subtle effect called “gate shadowing” [56]. To avoid the channelling effect during the drain/source ion implantation the implant beam (or the wafer) is tilted by 7–9°. For that reason one obtains small asymmetry between the source and drain diffusion resulting in different capacitive coupling as well as different transconductance between the transistors with opposite current directions.

Common centroid layout. Because of the gradients in temperature, oxide thickness and other process variations, it is difficult to ensure symmetry for large transistors. To reduce these global errors the common centroid layout is the most effective solution, however, the routing of interconnections is not so simple. It is important to keep the symmetry also in the routing of interconnections on the layout and to avoid differences in the parasitic resistance or capacitance [33].

Unit cells. All devices to be matched should have the same area to the perimeter ratio. In the case of two devices of different dimensions it is a good practice to base the layout of both of them on the same unit cell, which can be a simple capacitor, resistor, transistor or even a more complicated structure as a current cell [57]. Let us consider a DAC using the binary weighted current source architecture. After setting the minimum dimension of the unit current source according to the required accuracy and matching parameters of the technology [58], each binary weighted current source is constructed by putting the appropriate amount of N unit current sources in parallel (for example a source of weight equalling 16, is implemented as 16 unit cells together). To minimise the effect of linear gradient, the cells which switch simultaneously in a binary array are arranged in common centroid geometry [59]. Of course using only translated copies is allowed with no rotation or mirroring. Excellent examples of the use of unit cells can be found in [38, 60].

The same surrounding. The adjacent structures have a systematic influence on the matched devices, so symmetry must be applied not only to the devices under consideration but also to their surrounding environment. Each line of metal, diffusion, polysilicon etc., has width variation during IC processing. There is also a line width variation dependent on the location of adjacent structure. This is the proximity effect due to the variable light interference in exposure and to the variations in chemical flow for photoresist, developers and etchants [61]. So the environment should be identical outside to a distance of at least 30–50 μm and the matched devices should be “infinitely far away” from the crystal edge. The remedy for that problem is to surround the matched devices with dummy structures. The dummy device should be of equal nature as the matched device, for example for matched transistors (cells) the dummy structures are also transistors (cells). How many rings of dummy cells are required depends on the requested accuracy and technology [60]. Of course this solution costs an extra area of silicon.

Metal coverage. Symmetry is also important in metal coverage. The asymmetrical metal lines over identical MOS transistors which cover one of the transistors partially or completely can be disastrous for matching performance. The papers [62, 63] show that the shape and the degree of metal 1 overlap influence the relative mismatch current dramatically. The rule is simple, the best is to avoid metal lines over matched components. Sometimes it is impossible, for example for high accuracy DAC [60], and then it is better if the matched cells are covered by higher metal layers than metal 1, and to obtain the best symmetry all of them should be covered with metal.

2.3.6. Matching on multichip modules

After fabrication and cutting a wafer, the integrated circuit is glued, bounded and packaged. Packaging introduces not only parasitic components (resistance, capacitance, inductance), but also is a source of mechanical stress and heat non-uniformity distribution

that adversely affects the matching. It is especially important for multichannel systems, where several identical chips are mounted together on the same custom design multilayer board. So the design of such a board and the process of mounting chips should be handled with a special care to limit all the above effects [64].

The mismatch between identical components from different dies, wafers or batches is always bigger than between components placed on the same chip. As a result the spread of channel parameters inside the chip is small, while the differences from chip to chip on the module are too large to be accepted. The solution is to design on each chip the correction DACs for tuning the gain, passband, discriminator level, etc., and implement on the chip the local address set during bonding a die on a module [54]. Then after the calibration measurements the correction DACs are loaded individually one by one. That protocol guarantees the cancellation of unwanted offsets from chip to chip, and ensures proper work of the whole module.

2.3.7. Mismatch simulation using the Monte Carlo analysis

Evaluation of the matching performance of a design can be done employing the Monte Carlo simulation [65]. In that way one can quickly estimate channel-to-channel matching performance, identify critical devices and then optimise their dimensions or bias conditions for them. One should remember that specified in usual way matching parameters, for example for transistors, are valid under certain assumptions (common centroid layout, identical surrounding, no metal crossover, etc.). Having these parameters specified for a given technology, one can calculate variations of V_{70} and β for each transistor dimension used in the design. The variations of the threshold voltage and the current factor can be incorporated into the BSIM3v3 transistor model [11] for each transistor independently as parameters of the Gaussian distribution of the corresponding HSPICE parameters: zero-bias threshold voltage V_{TH0} and surface mobility U_0 . In the Monte Carlo simulations each of these parameters is sampled independently from a given distribution. Since the equation (2.22) and (2.27) specify matching of a pair of devices, the standard deviation for single devices used in the Monte Carlo simulation should be by factor $1/\sqrt{2}$ smaller. The Monte Carlo simulation performed for a single channel gives an estimate of random channel-to-channel matching. However, one has to remember that they do not cover long distance systematic effects, which for multichannel architectures may be important.