# Constraints and Design Approaches in Analog ICs for Implantable Medical Devices

Fernando Silveira, Julián Oreggioni and Pablo Castro-Lisboa
Instituto de Ingeniería Eléctrica, Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay.
{silveira, juliano, pcastro}@fing.edu.uy

## ABSTRACT

Active implantable medical devices (AIMDs) are microsystems requiring ultra low energy operation, which is a characteristic increasingly shared by several other applications. On the other hand, AIMDs must comply with several specific constraints imposed by the medical implantable context.

This paper first summarizes, from the point of view of the analog IC designer, the state of the art of AIMDs and their specific constraints. Then, some general design techniques for analog ICs for AIMDs are highlighted and an analog front-end for neural devices is presented to illustrate current circuit and architecture approaches.

Keywords – implantable devices, ultra low power analog CMOS

## INTRODUCTION

Active implantable medical devices (AIMDs) are defined as those active devices (i.e. devices including a power source) intended to be introduced inside the body by a medical procedure and intended to remain there after the procedure. AIMDs, since the appearance in 1960 of the implantable cardiac pacemaker, required ultra low power consumption in order to allow for long enough periods between surgeries while having minimal battery size, which until present is a significant (around one half) part of the device total volume. Nowadays AIMDs are sophisticated embedded systems that take full advantage of the progress in microelectronics technology and still require ultra low power consumption techniques, particularly in the analog functions.

This tutorial paper presents the requirements and design techniques of ultra low power (ULP) analog CMOS integrated circuits for implantable medical devices.

The paper is organized as follows. First the main characteristics and needs of implantable medical devices, particularly from the point of view of the analog circuit designer, are described. The next section summarizes a design method for ULP analog CMOS. Finally, the technique is illustrated in an analog front-end for neural devices and conclusions are summarized.

## IMPLANTABLE MEDICAL DEVICE SYSTEMS

### *Therapies*

If we consider implantable medical devices that have actually reached the market and became successfully established therapies, after the cardiac pacemaker, a very slow appearance of other devices occurred. They main historical milestones are: cochlear implants (for treating audition disorders) emerged during the sixties, the implantable cardiac defibrillator in 1980 and finally implantable neurostimulators for pain (e.g spinal cord stimulation) and Parkinson's disease (deep brain stimulation) arose during the eighties and nineties. However, since the year 2000, multiple new therapies are in development and some of them in clinical testing in the cardiac and neural fields. They target conditions such as heart failure, hypertension, foot drop, obesity, sleep apnea and several others. Additionally the field of brain-computer interface definitely got established [1].

### *Functions*

All these therapies are built on top of platforms providing a few basic functions as follows.

*Stimulation*: Two main stimulation approaches exist: voltage mode and current mode. In voltage mode [2], which is the one applied in cardiac pacemakers, an approximately rectangular voltage pulse is applied to the tissue through a discharged series capacitor. This series capacitor then receives as much charge as the one delivered to the tissue. In a second phase this charge is given back to the tissue by discharging the series capacitor through the tissue, so that no net DC charge, that would damage the tissue, is delivered to it. In current mode, which is the most common method in neurostimulators, rectangular, biphasic, current pulses are delivered to the tissue. In both cases the required voltages range up to 7.5V to 16V, so a voltage multiplier is required to generate these voltages from the battery voltage (which at beginning of life is in the range from 2.8V to 4V depending on the battery type).

*Sensing*: The first cardiac pacemakers operated in open loop, only stimulating. Some current devices, like Deep Brain Stimulators for Parkinson, still do so or in other cases operate with a "patient in the loop", as in the case of neurostimulators for pain, where the patient can start or stop predefined therapies that are stored in the device. However, when it is possible to have a suitable sensing method, close loop operation is a better way to optimize the therapy. This is the way all current cardiac devices and some neural devices operate. The signals measured in order to "close the loop" include, on one hand, bioelectric signals generated in the body. On the other hand, signals that give information about the patient body state and that can be easily electrically measured from the implantable device are applied. The use of this kind of signals avoids the difficulty of dealing with implanted biosensors that measure non electrical magnitudes. In this second group are the signal of an accelerometer that provides information about the person movement or position and the impedance seen from the electrodes, which can be used to detect variations in organs (e.g. the use of the impedance of the thorax to detect the respiration cycle).

In all cases we are dealing with low frequency and amplitude signals. Frequency ranges from below 1Hz, in e.g. acceleration, impedance or some bioelectric signals, to at most 10kHz in the fastest neural signals. Amplitudes go from a few μVs (e.g. in some neural signals captured with cuff electrodes [3]) to some mVs. In addition, all these signals share the characteristic of their variability, among different patients and in the same patient along the time [4]. Due to this variability, it usually only makes sense to target a qualitative detection that enables the closed loop control of the system, opposed to a precise quantitative measurement. This low to medium resolution and low signal to noise ratio measurement has made the analog implementation the best choice when minimum power consumption is desired. Nevertheless, process scaling, with the associated reduction of consumption of digital signal processing blocks, is changing the balance in this trade-off.

*Communication (Telemetry)*: AIMDs include means for establishing bidirectional data communication with the outside of the body in order to receive configuration information and transmitting status data, statistics of operation and measurement results. Traditionally this has been done through an inductive (near field) link with up to 10 cm range. More recently a band around 403 MHz has been allocated for implantable device communication allowing ranges up to a couple of meters.

*Control*: The operation of the implantable device is commanded by a digital processor (microcontroller) running embedded firmware.

*Auxiliary functions*: In addition to the previous main functions the following auxiliary functions are included:

- Battery supervision: the voltage, impedance or consumed charge of the battery must be monitored in order to know when the battery is near depletion and the replacement of the implant needs to be scheduled (or the battery needs to be recharged in systems with rechargeable battery).

- The impedance seen from the device to the leads that connect to the body is measured as a way to diagnose the state of the connection of the device, leads and electrodes to the body.

- Most implantable devices include a magnet sensor (based on a reed switch or Hall sensor) used as a simple mechanism to alter the device operation when is not available an external device capable of communicating using the implant communication protocol.

- Some AIMDs nowadays apply rechargeable batteries; in this case, the required circuitry for battery recharge is needed.

### *Power Source (Battery) and Consumption*

AIMDs require batteries of proved reliability. Different lithium based chemistries have been applied. The capacity of the battery ranges from approximately 0.5Ah to 1.0Ah. One ampere hour is equivalent to 114 µA.year. This means that if a 10 year operating life is desired (e.g. in a cardiac pacemaker) a total average consumption of less than 11.4 µA is required. This consumption is typically distributed as follows: one third in the energy delivered to stimulate the heart, one third in the analog sensing and processing circuits and one third in the processor and other digital circuits. Therefore, the total consumption is much higher than the minimum energy which is the one required for stimulation, so still exists ample room for improvement in terms of power consumption.

Some therapies need to stimulate at higher frequency or with higher energy, thus intrinsically having higher power consumption. In these cases, rechargeable batteries are being used. These batteries are recharged through an inductive link with the exterior of the body. Some systems have powered the implant through RF energy, but the fact of not having a permanent source of power for the implant, limits this approach to particular, non critical, applications.

### *Safety and Reliability*

A medical device, besides being effective in providing the intended therapy, must be safe (i.e. to avoid harming the patient, operator or the environment) and reliable (i.e. with a low probability of failure). A key rule for assuring safety in the design of medical devices is that the device remains safe even under single fault conditions that could provoke critical failures. If we couple this approach with high reliability, where the probability of a single fault is low and therefore that of a double fault is extremely low, we assure that dangerous conditions are extremely unlikely. Tolerance to single faults calls for inclusion of redundant circuitry in key positions conditioning the device design. An example of this criterion, in the framework of analog IC design for AIMDs, is the design of input stages of biopotential amplifiers that connect to the body. In order to avoid damaging the tissues, the DC current that could flow to the body through the amplifier inputs, even under a single fault condition, must be limited, usually to 1 to 50 µA depending on where this current is applied. A method to assure this is to have an off-chip capacitor in series with the amplifier inputs, assuring that for DC current to flow towards the body a double fault must occur: one in the series capacitor and another one on the IC.

## ULP ANALOG CMOS DESIGN

Minimization of power consumption in the design of circuits for AIMDs involves taking advantage of the opportunities for optimization at all design levels from transistor level design through subcircuit architecture selection to system level design. We will analyze some techniques and ideas applied, starting with the transistor level analog design and then moving to the subcircuit architecture level.

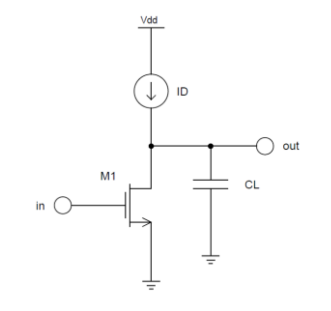
### *All inversion region design using the gm/ID design methodology*

Traditional CMOS IC design has only considered the above threshold or strong inversion region of operation of the MOS transistor, where the drain current in saturation varies quadratically with the gate voltage. In this approach the guiding variable for design has been the gate voltage overdrive equal to the difference between the gate voltage VG and the threshold voltage VT. Researchers, aiming at working with very low power, early on identified the potential of using the MOS transistor in the subthreshold or weak inversion region [5], where the drain current in saturation varies exponentially with the gate voltage. Nevertheless, an even better approach is to exploit all the possibilities that the MOS transistor give us by using indistinctly all the regions of inversion: weak and strong inversion but also the moderate inversion region (near or around the threshold region). Particularly because, as we will show next, the moderate inversion region in several cases provides the best compromise between transconductance generation and parasitic capacitances, leading to an optimum in power consumption.

In order to exemplify the design method and key results we will consider the simplest analog CMOS circuit, shown in Fig. 1, a common source stage loaded by an ideal current source ID and a load capacitance CL, that we call “intrinsic gain stage” [6, 7]. The magnitude of the low frequency gain A0 and the transition frequency (gain bandwidth product, fT) of this circuit are given in Eqs. (1) and (2) in inset of Fig. 1, where gm and gd are the small signal transconductance and output conductance and VA is the “Early Voltage” that can be used as a first order approximation for ID/gd. CL includes the external load capacitance as well as the effect of the parasitic capacitances of the transistor.

Therefore, the “speed” or gain bandwidth product is related to the transconductance and the trade-off between speed and consumption is related to the transconductance to current ratio gm/ID.

Since gm is the derivative of the drain current with respect to the gate voltage, gm/ID is equal to the derivative of log(ID) with respect



$$A_0 = \frac{g_m}{g_d} = \frac{g_m / I_D}{g_d / I_D} = \frac{g_m}{I_D} . V_A \quad (1)$$

$$f_T = \frac{g_m}{2.\Pi.C_L} \quad (2)$$

FIGURE 1. Intrinsic gain stage applied as prototype of CMOS Operational Transconductance Amplifier (OTA) for showing the design method and its results

to the gate voltage, i.e. the slope of the ID vs. VG curve when ID is plotted in log scale. The maximum gm/ID (i.e. the best speed-power trade-off) occurs in the weak inversion region. However, as we will show next, to bias the transistor in this region is in many cases costly in terms of size and parasitic capacitances. In order to visualize this, let us consider the relationship between gm/ID, ID current and transistor aspect ratio W/L. When short channel effects are not significant, the transistor geometric features (W and L) appear in the expression of ID as a (W/L) multiplicative factor. In this case, it can be easily shown [6, 7] that gm/ID is determined by the "normalized current" or "current density" ID/(W/L). When short channel effects are significant, two or three of these curves, representing corresponding channel length L ranges need to be considered, as shown in Fig. 2 for a 90 nm process.

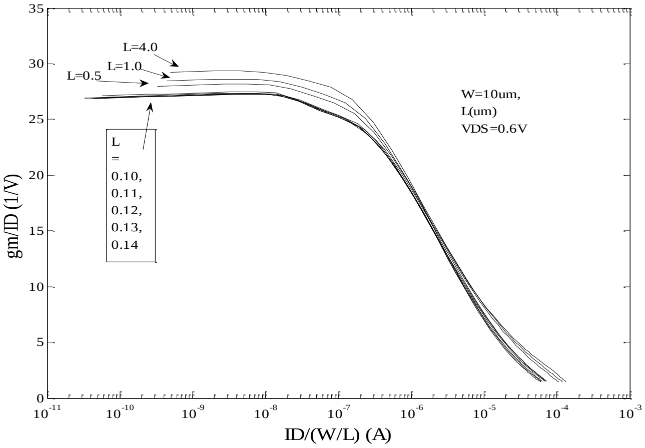


FIGURE 2. gm/ID vs. ID/(W/L) for nMOS transistors of a 90 nm process (Derived with data taken from [7])

### *Optimum of Power Consumption*

The existence of a power optimum for a given gain bandwidth product can be easily understood looking at the gm/ID vs. ID/(W/L) curve of Fig. 2. If for a gain bandwidth product requirement (which is given by Eq. (2)) we would like to reduce the consumption in a given design, we would need to move towards weak inversion in order to get higher gm/ID. However, since the scale in the ID/(W/L) axis of Fig. 2 is logarithmic, for a given increase in gm/ID, ID/(W/L) needs to decrease in a much larger extent. This implies to largely increase the (W/L) value, because ID cannot decrease so much while achieving the required gm. This effect of increase in (W/L) is more important as the operating point gets closer to the weak inversion region (quasi-constant gm/ID region). The large increase in (W/L) results in an increase of parasitic capacitances (Cpar) which increases the total CL capacitance. The higher CL capacitance requires a larger gm in order to maintain a constant fT and hence more current ID. Therefore a "sweet spot" occurs somewhere due to these opposing trends of increasing gm/ID and at the same time increasing parasitic capacitances. This optimum usually lies in the moderate inversion region.

### *Design Space Exploration*

In order to explore the design space, considering all inversion regions, and in order to determine the optimum design point what is needed is a) a way of modeling the MOS transistor in continuous fashion in all inversion regions and b) a suitable guiding variable to "sweep" the whole design space.

Regarding the models, when short channel effects are not significant, analytical compact models such as EKV [8] and ACM [9] are good options. When short channel effects are significant, though these effects can be included in the EKV and ACM models, both the model and the associated parameter extraction become too cumbersome. The same occurs with the BSIM model. The final design will be checked and fine tuned based on simulation with the model (in many cases BSIM) and parameters provided by the foundry, but BSIM is not a practical model for simple design space exploration and optimization routines as we are looking for. In current short channel technologies a practical approach is to rely on what we may call a "semi-empirical" model [6, 7] where the key relationships are extracted from measurements or simulation with model and parameters provided by the foundry. These key relationships are the following: gm/ID vs. ID/(W/L); gd/ID vs. gm/ID and Cxx/(W.L) vs. gm/ID, being Cxx the intrinsic capacitances of the MOS transistor.

Regarding the guiding variable, gm/ID proves to be a good choice since: a) it varies in a small range where the significant variations can be covered with a grid of a small number of points, b) it gives indication of the inversion region where the transistor is operating and c) the key performance aspects are related to gm/ID. We have just exemplified the case of the gain bandwidth product, but most analog design aspects such as matching, noise, output swing and others [8, 9, 10] can be stated as a function of gm/ID.

Fig. 3 shows an example of the result of a design space exploration performed with this technique [11]. It shows how the optimum point location for the input differential pair transistors of a Miller amplifier changes with changing gain-bandwidth product and load capacitance value.

## NEURAL ANALOG FRONT-END EXAMPLE

Fig. 4 shows the core architecture [12] of the input amplifier for an analog front-end for neural signals. The main challenge of this kind of amplifiers is to reach a very low equivalent input noise (around 2 µVrms in a band from 250Hz to 8 kHz in this case) with minimum power consumption. This architecture improves previous works that also applied a Differential Difference Amplifier (DDA) structure by making it asymmetric and thus much reducing the noise contribution of the Gm2 block to the input. For Gm2 a symmetrical OTA structure was applied, as shown in [12]. The $1/K^2$ factor that contributes to this asymmetry is implemented through the copy factor of the current mirrors of the symmetrical OTA.

We will consider now the main trade-offs in the design of a standard input differential pair for Gm2. Next we will show how these trade-offs can be dealt with through the design space exploration approach that was previously presented.

When the inversion level (or equivalently gm/ID ratio) of the input transistors of Gm2 is changed, it impacts on the following

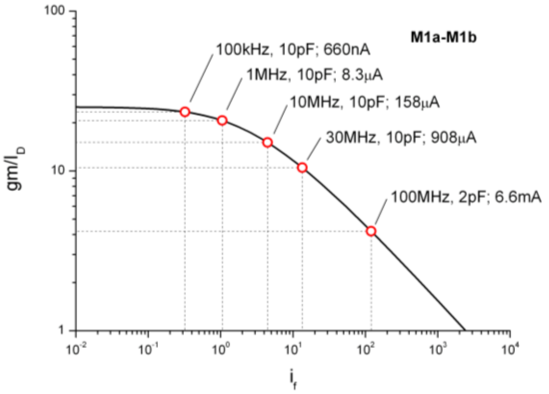


FIGURE 3. Design space exploration for the input differential pair transistors of a Miller amplifier in a 0.35 µm process. Each point shows the optimum at a given gain bandwidth product and load capacitance and the total amplifier consumption [11].

aspects. First the key aspect of how much noise is contributed by Gm2 to the output and, hence, to the equivalent input noise. This aspect is much relaxed by the 1/K$^2$ factor (equal to 9 in [12]), but must anyway be taken into account. Second, the input linear range of the differential pair, which must be enough to handle the maximum expected output amplitude (around 250 mVp in this case). Finally, the selection of gm/ID impacts the power consumption required to achieve the required transconductance value for the Gm2 block [12].

Fig. 5 illustrates how the presented method allows the designer to assess the trade-offs and select the best compromise. In [12] a gm/ID value of 6.6 1/V was selected.

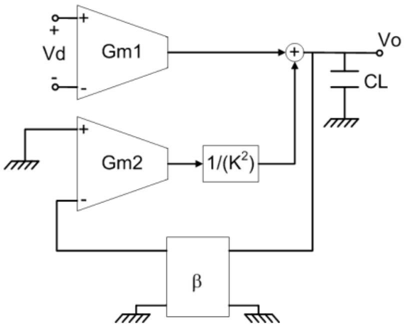


FIGURE 4. Asymmetric DDA architecture. Feedback factor β is taken as 1 and Gm1 includes a local feedback loop at the output for implementing high pass characteristic [12, 13].

## CONCLUSIONS

AIMDs often require minimization of energy consumption in order to increase operating life at a given battery size or to minimize size at a given operating life. This goal involves design decisions at the system, subcircuit and transistor level. At the transistor level, adequate design methods, that allow to efficiently explore the design space in all inversion regions, are key for rapidly obtaining a suitable initial design that could be fine tuned by simulation. The use of gm/ID as the guiding variable for this process has proved to be a good choice that, furthermore, allows to easily showing the existence of an optimum of power consumption.

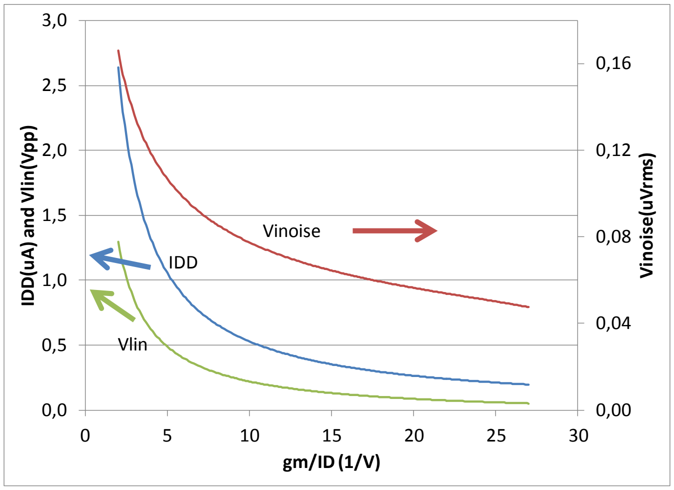


FIGURE 5. Design space exploration of input differential pair of Gm2, showing total supply current (IDD), peak to peak input linear range and contribution to the total amplifier input equivalent noise as a function of the gm/ID ratio of the differential pair transistors.

## REFERENCES

[1] R.R, Harrison, "The Design of Integrated Circuits to Observe Brain Activity," Proceedings of the IEEE, vol.96, no.7, pp.1203-1216, July 2008

[2] F. Silveira, D. Flandre, Low Power Analog CMOS for Cardiac Pacemakers, Springer, 2004.

[3] J.A. Hoffer et al, "Initial results with fully implanted Neurostep FES system for foot drop", 10th Annual Conference of the International FES Society, July 2005.

[4] J. Webster, Medical Instrumentation. Application and Design, Sec. 1.4 , Medical Measurements Constraints, John Wiley and Sons, New York, 1995.

[5] E. Vittoz, J. Fellrath, "CMOS analog integrated circuits based on weak inversion operation", IEEE Journal of Solid-State Circuits, Vol SC-12, pp. 224-231, June 1977.

[6] F. Silveira, D. Flandre, P. Jespers, "A gm/ID Based Methodology for the Design of CMOS Analog Circuits and its Application to the Synthesis of a Silicon-on-Insulator Micropower OTA", IEEE Journal of Solid State Circuits, Vol. 31, No. 9, Sept. 1996, pp. 1314 - 1319.

[7] P. G. A. Jespers, The gm/ID Methodology, A Sizing Tool for Low-voltage Analog CMOS Circuits, Springer, 2010.

[8] C. C. Enz and E. A. Vittoz. Charge-Based MOS Transistor Modeling - The EKV Model for Low-Power and RF IC Design. John Wiley, 2006.

[9] M.C. Schneider, C. Galup-Montoro, CMOS Analog Design Using All-Region MOSFET Modeling, Cambridge University Press, 2010.

[10] D. Binkley, Tradeoffs and Optimization in Analog CMOS Design, Wiley, 2008.

[11] P. Aguirre, F. Silveira, "CMOS op-amp power optimization in all regions of inversion using geometric programming", Proc. SBCCI 2008, pp. 152-157, ACM

[12] P. Castro, F. Silveira, "High CMRR power efficient neural recording amplifier architecture." in Circuits and Systems (ISCAS), 2011 IEEE International Symposium on, pp. 1700-1703.