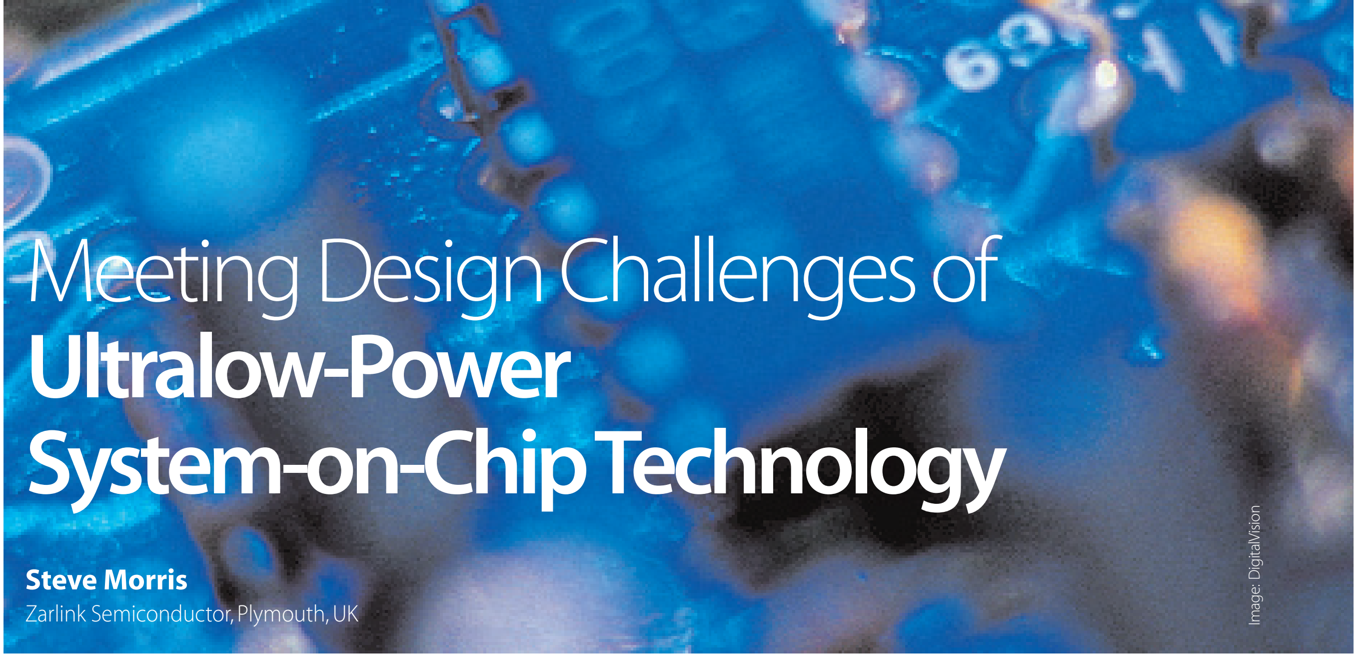


Image: DigitalVision

Meeting Design Challenges of Ultralow-Power System-on-Chip Technology

Steve Morris
Zarlink Semiconductor, Plymouth, UK

New-generation battery-powered products are required to provide increasingly greater performance. This article examines technology solutions and design techniques that can be employed to achieve ultralow-power medical devices.

Expanding demand for ultralow power

The recent upsurge in personal area network and wearable portable devices has placed great demand on microelectronic solutions that require ultralow-power (ULP) performance from a small battery. The terms "power aware" and "energy efficient" are increasingly influencing consumers' buying decisions and designers of integrated circuits (ICs) must meet the demand for ever-increasing performance in battery-powered systems such as personal digital assistants, mobile phones and headsets.

Despite these mass-market drivers, there is still a huge gap between power requirements for consumer devices, which have batteries that are rechargeable on a daily basis, and medical devices such as pacemakers, hearing aids, gastrointestinal camera capsules and telemetry systems, which are much more demanding applications. For example, a dual-chamber heart pacemaker consumes less than 10 µW, operates for more than seven years on a single battery and contains sophisticated digital and analogue circuitry on a single system-on-chip (SoC). A state-of-the-art, completely in-canal hearing aid draws less than 900 µA from a 1.25-V zinc-air battery and contains complex digital, digital signal processor (DSP), analogue and sometimes radiofrequency (RF) functions (Figure 1).

Designers of SoC, mixed-signal ICs for implantable and wearable medical equipment must use techniques and methodologies to improve power efficiency and thereby provide the longevity required by the end user. This entails going beyond what is normally achieved with low-power IC design. This article discusses some techniques and approaches to achieve ULP systems on chip and some areas of research in this field. It is known that special technologies can be employed to achieve ULP performance, but they can be expensive and have negative effects as well as positive advantages. For digital design, there is a conflict with analogue circuitry, while special solutions must be adopted for memory and RF.

Power challenges

It is well known by designers of digital complementary metal oxide semiconductor (CMOS) ICs that total power dissipation consists of dynamic and static components. Dynamic power is given by the expression

$$P_D = af \times C \times V^2 \qquad [1]$$

where af is switching activity as a function of clock frequency, C is the capacitance being driven and V is the supply voltage. Thus, dynamic power has a quadratic dependence on supply voltage. However, lowering V significantly reduces the switching speed of the gate. One solution is to use multiple voltage domains on the chip. Performance-critical functions would be located in a high-voltage domain, and noncritical functions would use a low-voltage domain.

Static power is determined by leakage current through each transistor and for most CMOS processes consists of diode reverse-bias junction leakage and subthreshold leakage (weak-inversion current between source and drain). This latter process can be dramatic because it increases exponentially with the reduction in the transistor's threshold voltage (V_T) and is made worse by short-channel effects such as threshold voltage roll-off and drain-induced barrier-lowering.

Consideration of the above factors (dynamic and static power) leads to the first paradox. Because propagation delay decreases with reduced V_T, V_T can be scaled to satisfy switching-speed requirements. However, this scaling would lead to an exponential increase in leakage current. Therefore, the scaling used in CMOS technology to allow more gates to be placed on a single chip (Moore's Law) causes static power to overtake dynamic power for feature sizes below approximately 150 nm (Figure 2).

Technology solutions

To deal with this, technology and circuit design innovations have been put in place. Technology solutions include double-gate, fully depleted silicon-on-insulator and multiple threshold voltage CMOS (MTCMOS). The latter is gaining popularity together with variable-threshold-voltage CMOS (V_TCMOS) because of the ease of realisation compared with the silicon-on-insulator (SOI) solution, which is known to suffer from self-heating effects.[1,2]

With MTCMOS, high-threshold transistors suppress subthreshold leakage current and low-threshold transistors provide high-switching performance. Because of processing complications, V_Ts are adjusted by selective ion implantation rather than using different gate oxide thicknesses, which means that only two V_Ts are normally available per process. It is desirable that a wider range of V_Ts is available on a single chip, thus work is underway to provide a viable solution by providing different gate-oxide thicknesses.

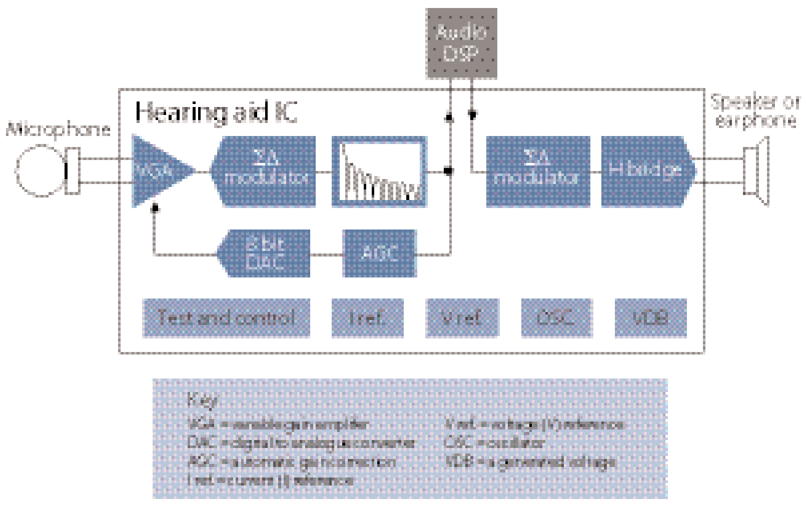
With V_TCMOS, a substrate-bias circuit is used to control the body bias and hence the threshold voltage. In standby (or sleep) mode a large reverse body bias is applied to increase V_T; in active mode a zero bias is applied. V_TCMOS and MTCMOS can be used in a cell-by-cell approach. Low-V_T transistors are used for critical paths and high-throughput modes; whereas leakage in idling circuits is cut off with high-V_T transistors (Figure 3).

Power saving by design

For ULP IC design it is essential to pay meticulous attention to detail at all levels and stages of the design process. The techniques described above are used in conjunction with a top-down approach for designs that are power aware. Everything must be optimised, from the specification to the architecture and down to details of implementation. Trade-offs must be made: performance versus power; area versus power; risk versus power; and design time (and effort) versus power. If a specification can be relaxed (yet still be compliant with system requirements), or a block of circuitry can be powered down when not needed, the return on power saving can be enormous. For digital and DSP design, the biggest gain in power reduction is to operate the circuitry at the lowest supply voltage and lowest possible speed while meeting system-throughput requirements. Resource sharing should be considered at the architectural level; concurrency and parallelism should be used as well as pipelining the data path.

Architecture. For DSP implementations, it is worth using custom-developed architectures to reduce power and achieve high utilisation. The result is that nonconventional architectures are produced (probably nonHarvard), but the low-complexity implementation will use optimally low coefficient word lengths and, therefore, low data word lengths. This will reduce storage requirements and compound power saving.

Figure 1: A DSP-based hearing aid contains many complex functions.



Switching activity. For arithmetic operations, switching activity can be reduced by choosing a suitable number representation (for example, sign magnitude over 2s complement because the number of transitions on busses will be reduced) and by selecting an appropriate adder or multiplier (for example, Conditional Sum Adder rather than Wallis Tree). Multipliers should be judiciously shared, with efficiency obtained by using shift-and-add when multiplying by constants.

IIR filters. At the algorithmic level, consider Infinite Impulse Response (IIR) filters rather than Finite Impulse Response and radix-4 Fast Fourier Transforms rather than radix-2. Further reductions can be achieved by producing highly efficient DSP blocks and reusing these for many operations.

Digital techniques. A number of other digital considerations should be borne in mind. Clock gating should be used so that clocks are turned off when not needed and gate–cell mapping should be considered to reduce switching activity. Register Transfer Level (RTL) code should also be structured with power in mind and, at the time of writing, there are a number of tools becoming available to read the RTL as it is being developed, apply checks and indicate in schematic and RTL where potential power problems are occurring.

Transistor size. At the transistor level, consideration should be given to transistor sizing because this yields high gains in power savings. By determining the ratio of parasitic capacitance to gate input capacitance, minimum size cells can be used and drivers, which are sized only as appropriate, between blocks and for off-chip signals. Clock tree buffering can also use up power unless carefully controlled.

Placement. During physical implementation, care is needed with floor planning and placement so that high-activity clocks and busses do not suffer from, or cause, cross-talk, resistive losses or parasitic loading. High-activity areas should be kept together on the

➔

chip and interconnections kept as short as possible and close to pads. Care is also needed for these areas and potential interactions with sensitive analogue blocks and nodes. To correctly implement multiple voltage domains it is necessary to separate the different power meshes for each domain. Power-aware cell placement based on activity-weighted nets can be used to minimise dynamic power consumption. Results from voltage drop analyses may be used to determine ideal locations for inserting buffers and special care is needed to avoid floating nets.

Power saving in medical devices

Most of the above applies to digital circuitry. ULP medical devices will inevitably also contain analogue circuitry and require different power-dissipation considerations. For analogue circuits, down-scaling supply voltage or process feature size will not automatically reduce power consumption and, in fact, usually has the opposite effect. In analogue circuits, power is consumed to maintain the signal energy above the thermal noise floor in order to achieve the required signal-to-noise ratio (SNR) or dynamic range (DR). Because minimum power consumption is also proportional to the ratio between supply voltage and signal amplitude, power-efficient analogue circuits should be designed to maximise the voltage swing. Reducing the supply voltage while maintaining SNR and bandwidth for an analogue circuit therefore requires transconductance to be increased. This is normally done at the expense of power. The approach for analogue parts must therefore be different.

Firstly, architectural or circuit implementation areas can be examined for power optimisation. For a given supply current power goes up if the supply voltage is increased, but more noise is acceptable at a higher voltage when DR is considered and this allows a reduction of bias currents and total power may be reduced. For amplifiers, a higher supply voltage allows telescopic rather than folded structures, thus the current can be lower. For mixed-signal ICs, a dual supply can be provided on-chip by dividing down the chip's supply voltage and this is used on other parts of the chip for digital and using the higher supply for analogue blocks. In general, lowering the frequency gives immediate return on power savings; this applies to digital circuits and clocking sampled data and sigma-delta circuits.

At transistor level. Here, devices should be operated in the optimum region by selecting the inversion coefficient and trade-offs should be examined. For example, in a switched capacitor integrator, the input capacitance of the amplifier can be minimised by a low W.L product of the differential pair (W is the width of the MOSFET transistor and L is the length of the transistor); but enlarging W/L maximises the transconductance and hence the bandwidth. Thermal noise versus flicker noise can also be traded and keeping down flicker noise (large W.L) can allow higher thermal noise and therefore lower device current. Use of detailed matching data[3] is important in analogue design for low power, because the better the modelling of mismatched devices, the smaller the area that can be used to guarantee a satisfactory yield. Consequently, the circuit will consume less power for the specified accuracy and bandwidth. The different matching properties for transistors operated in weak or strong inversion should be taken into account too.

RF IC design. Here, optimising the architecture for a given application is a key requirement when considering ultimate low-power consumption. For example, using direct conversion with a high-modulation index for low data rate applications and, when the design freedom exists, choosing a modulation scheme that allows constant envelope techniques rather than high-level modulation, that is, a modulation scheme in which the bits/symbol are greater than 1. Using DSP within complete receiver or transmitter sections can also yield power-efficiency benefits, even when considering the data conversion requirement.

Circuits. Techniques such as switching power amplifiers rather than linear or semilinear variants can also be beneficial for low-power consumption. In RF circuits, achieving low-power operation is highly dependent on the use of passive components for matching or biasing, for example, the choice

Figure 2: Energy and gate delay as a function of supply voltage.

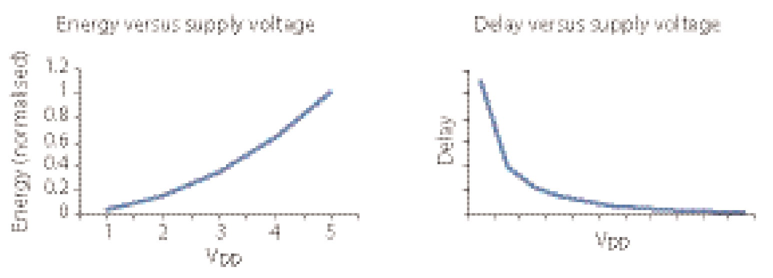
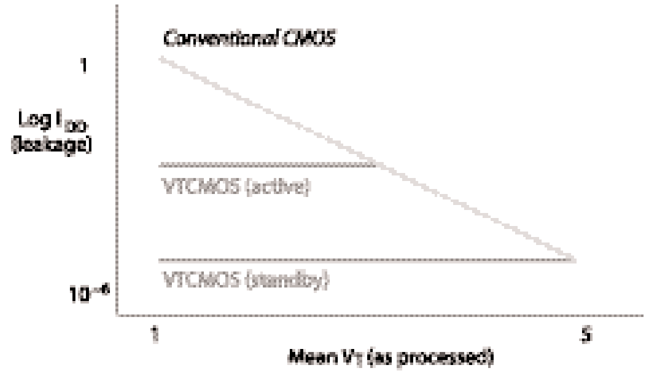


Figure 3: Leakage current for a device in conventional CMOS versus VTCMOS.



of integrated resistive or current source loads or an external inductive load.

Components. Power savings are possible by careful component selection and by moving some components off-chip. An obvious example is using just one external "bias" resistor to avoid over-dimensioning the analogue circuitry, to allow for the inherent ±30% variation that comes with the on-chip bias resistor. Sometimes it is worthwhile trimming circuit elements by automatic tuning techniques (in RF) or by trimming at wafer probe. This requires the use of a one-time programming feature such as antifuse or polyfuse or Electrically Erasable Programmable Read Only Memory on the process, which can be worthwhile for power-efficient circuits. If a special CMOS process is available, then low-V_T transistors, high-resistive polysilicon and double-poly or polysilicon over diffusion capacitors are useful to keep circuitry on-chip and avoid the necessity of wasting power by interconnection to external components.

Memory. ULP equipment usually requires nonvolatile and high-density memory. Low power and high performance are mandated and therefore lowering the operating voltage is not always viable because of the requirement for stability, speed and radiation tolerance. Low-voltage design techniques are used for the basic circuits such as sense amplifiers, charge pumps (for FLASH), and precharge circuits (for Dynamic Random Access Memory) but work continues in the quest for ULP memories with operating voltage below approximately 1.5 V.[4]

References

1. T. Kuroda et al., "A 0.9 V 150 MHz 10 mW 2-D Discrete Cosine Transform Core Processor with Variable Threshold Voltage Scheme," IEEE International Solid-State Conf. (1996).
2. L. Wei, Z. Chen and K. Roy, "Design and Optimisation of Double-Gate, Fully Depleted, SOI MOSFETs for Low-Power High-Performance Designs," IEEE SOI Conf. (1997).
3. M. Pelgrom et al., "Matching Properties of MOS Transistors," *IEEE J. of Solid State Circuits*, **24**, 5 (Oct 1989).
4. M. Pedram and J. Rabaey, "Low-Power Design Methodologies," Kluwer (1996). mdt

Dr Steve Morris
is Director of Research and Development,
Ultra Low-Power Communications Division, at
Zarlink Semiconductor, Tamerton Road,
Roborough, Plymouth PL6 7BQ, UK,
tel. +44 1752 693 769, fax +44 1752 693 743,
e-mail: steve.morris@zarlink.com
www.zarlink.com

This article was first published in *Medical Device Technology* magazine.

➔