Reducing IC power consumption: Low-power design techniques

Sunil Deep maheshwariNaveen Srivastava,Rohit Ranjan, - September 24, 2015

Designers always look for ways to reduce unwanted components of power consumption, either by architecting the design in a fashion which includes low power techniques, or by adopting a process which can reduce the consumption. However, some of these solutions come at the expense of performance, reliability, chip area, or several of these. Eventually, one has to reach a compromise between power, performance, and cost. The article below aims to discuss some of those techniques. These techniques are divided into Architectural Techniques and Process Based Techniques.

1. Architectural Power Reduction Techniques: At the RTL level, one can take several steps to reduce the overall power consumption of the device. Typically, RTL based techniques minimize the dynamic power consumption of the device, however, using techniques like power gating, one can also reduce leakage power of a part of chip. Various popularly employed techniques are:

1.1. Clock Gating: This technique is a very popular Dynamic Power reduction technique. Dynamic power is the sum of transient power consumption (P\text{\scriptsize transient}) and capacitive load power (P\text{\scriptsize cap}) consumption. P\text{\scriptsize transient} represents the amount of power consumed when the device changes logic states, i.e. "0" bit to "1" bit or vice versa. Capacitive load power consumption as its name suggests, represents the power used to charge the load capacitance. Total dynamic power is as follows:

\[ P_{\text{\scriptsize dynamic}} = P_{\text{\scriptsize cap}} + P_{\text{\scriptsize transient}} = (C_L + C) V_{\text{dd}}^2 f N^3 \]

where \( C_L \) is the load capacitance, \( C \) is the internal capacitance of the chip, \( f \) is the frequency of operation, and \( N \) is the number of bits that are switching. As dynamic power consumption is directly linked to toggling of the MOS cells, gating the clock when not required helps reduce the dynamic current. This techniques help preserve the state of the design while only limiting the transient currents. Designers frequently use AND/NOR gates to gate a clock, however, latch based clock gating is the most favored technique as it also saves designs from hazards which can otherwise introduce additional power consumption, inherent in dynamic power consumption.

1.2. Variable Frequency/Frequency Islands: In a big chip, not all the blocks need to be clocked at highest possible frequency in order to achieve the desired level of performance. There can be few blocks which inherently work slow (e.g., slow communication blocks like I\text{\scriptsize 2}C, UART, etc.) and, therefore, can be clocked at slower clock than blocks like core/processor which require high frequency clock for maximum throughput. Therefore, by providing different frequency clocks to
different blocks, one can reduce localized dynamic consumption.

Figure 1: Frequency Islands

1.3. **Power Gating**: There can be applications where certain blocks of the chip might not be required to function in some of the low power modes like sleep, deep-sleep, standby mode, etc. and only a part of the device is required to function. In such cases, it makes sense to power off non-functional blocks so that device does not have to power unused blocks. This not only helps reduce the dynamic consumption but leakage power is also saved for such a power gated block. However, while dealing with such a technique, design has to make sure that signals coming in from power-gated blocks do not affect the functioning blocks while operating in low power. For this purpose, isolation blocks are placed in the path so that functionality corruption does not take place, as can be seen in figure 2. Please note that the isolation signals are not required for signals going out of always-ON domain to other power domains as they are never supposed to go non-deterministic.
1. Process Based Power Reduction Techniques: There are a lot of components of power consumption and not all can be targeted using Architectural Techniques alone. Power consumption due to effects like Drain Induced Barrier Lowering, Gate Induced Drain Leakage, sub-threshold leakage, etc. can be controlled most effectively using process based techniques. Below are some of the most commonly employed Process Based Techniques:

2.1 Multi Threshold Voltage CMOS Cells: A lot of MOS characteristics are governed by the threshold voltage of the cell. Sub-threshold current is the current between source and drain when the gate voltage is below threshold voltage. Mathematical expression for approximate value of this current is:

\[ I_{SUB} = I_o \cdot e^{\frac{V_{gs}-V_{th}}{\eta V_T}} \left[ 1 - e^{\frac{V_{DS}}{V_T}} \right] \]

As one can see, this current reduces when threshold voltage \( V_T \) is increased. Therefore, higher \( V_T \) cells can be placed to decrease this component. However, as we have seen in propagation delay above, increasing \( V_T \) has a negative impact on frequency of operation. Therefore, designers have to adopt a strategy to mix lower \( V_T \) and higher \( V_T \) cells in a way to reduce the leakage current while maintaining the desired frequency of operation. To implement this strategy, high \( V_T \) cells are used as sleep transistors which gate the supply to further low \( V_T \) based design when the block is supposed to be in standby mode. When device is in active mode, these sleep transistors are turned ON, and low-\( V_T \) blocks downstream of this sleep-transistor can get the power and work as usual. This helps in reducing the current in standby mode. Alternatively, various data paths are categorized in terms of timing critical versus non-timing-critical paths. Timing critical paths can be implemented with cells
of lower $V_T$ (known as LVT (lower $V_T$) cells) so that the same operation can be achieved in less amount of time as compared to a path implemented with high $V_T$ cells (also abbreviated as HVT cells). This mixed usage approach balances the leakage current even when the chip is in run mode.

Another solution is to dynamically change the $V_T$ of the cells as per the application requirement. This can be achieved by varying the well/body biasing voltage using a control circuit. This requires more complex MOS fabrication as it requires twin-well or triple-well fabrication technique. This is more commonly known as Variable Threshold CMOS (VTCMOS). However, one should note that lowering the $V_T$ also compromises with the reliability of the chip as even lower voltage swings can cause the logic to start functioning in an incorrect fashion. These voltage swings can arise due to various process or environmental variations. Therefore, one has to be very cautious while decreasing the $V_T$ of the cells so as not to compromise on the sanctity of the final application.

2.2 **Multi $V_{dd}$ Technique**: As we can see from the equation above, there is a quadratic relationship between device voltage $V_{dd}$ and dynamic power consumption. Therefore, one can reduce the dynamic voltage substantially by reducing the supply voltage. However, voltage reduction has its downside as well. Propagation delay of a cell is as below:

$$T_D = C \cdot \frac{V_{dd}}{k \cdot (V_{dd} - V_T)^2}$$

As one can see from equation above, reduction in the $V_{dd}$ increases the delay of the cell. As a result, the operating frequency of the cell reduces when one reduces the supply voltage. Therefore, one has to maintain a balance between voltage supply and associated performance.

A solution to this challenge can be to create voltage islands in the design where low performance slow peripherals can be powered using lower supply voltage and performance critical blocks can be powered using higher voltage. However, design has to make sure that appropriate voltage level
shifters are placed on those signals which talk across the voltage domains.

This technique also reduces Gate Induced Drain Leakage effect and associated power consumption in a device.

2.3 **Dynamic Voltage and Frequency Scaling**: Voltage Island technique, also known as Static Voltage Scaling presents few constraints while operating the device. This technique is not adaptive to the application needs and voltage supply to a block cannot be changed once designed. However, Dynamic Voltage Scaling technique liberates designer and customer of such limitations. This technique makes use of a regulator which can be programmed to deliver voltage levels as required. Therefore, various blocks can get configurable voltage and the customer/user can change the voltage settings as per the application settings. This can help save the power dynamically. Various solutions have also been used where the design freed the software to make changes to voltage scaling. The design itself senses the current-load requirement in the device and makes the voltage adjustments accordingly. This technique helps reduce power consumption in a more adaptive manner.

The same voltage scaling can also be clubbed with dynamic frequency scaling where the frequency of a block can be changed by the software as needed. Therefore, a block running on lower $V_{dd}$ can be clocked by a slower clock while maintaining the performance and functional requirements. This technique helps reduce dynamic as well as leakage power consumption in the device.

![Figure 4: Differential Voltage and Frequency Scaling](image)

2.4 **Fully Depleted Silicon on Insulator (FDSOI)**: This is another technique which helps reduce various components of leakage currents which are more of a menace at lower technology nodes. Leakage components like GIDL, Reverse bias current and Gate tunneling currents can be controlled very effectively using this technique. In this technique, the MOS sits over an ultra thin film of oxide which insulates the cell from rest of the body. On top of this oxide film, a very thin layer of silicon is deposited which acts as a channel. Due to its thinness, channel can be established in this layer without any additional doping of the same. For this reason, it is known as fully depleted SOI.
In another technique a small neutral region is deposited in the depletion region under gate. Here, channel thickness need not be as thin as required in FDSOI. This is known as partially depleted SOI (PDSOI). However, PDSOI tends to have higher $V_T$ (and therefore slower operation) and larger gate-effects as compared to FDSOI (hence, larger leakage currents). Therefore, due to better control over $V_T$ and drastically reduced leakage currents, FDSOI is a preferred choice for small process nodes (usually below 90nm).

The article discussed various solutions which can be adopted to achieve target current requirements. However, all of these have pros and cons. Therefore, designers often mix various solutions to reach the optimum level which meets not only current requirements but also takes care of performance and cost of the device to make the product sellable.

Also see:

- Aspects of IC power dissipation