eSAME 2015 Conference

LOW-POWER TECHNIQUES IN DIGITAL DESIGN, LANDSCAPE AND PROSPECT

Serge Backert, Cadence Design System, Inc.

Meylan, France

Abstract: There is a great variety of low-power techniques for digital integrated circuits, some are widely used and expending while others remain anecdotic. After listing most techniques with their costs and benefits, we discuss which could be the ones that will gain momentum in the coming years.

I. TRENDS IN THE SEMI-CONDUCTOR INDUSTRY

A. *Evolution of Power with Nodes*

Both leakage/static and dynamic power increase with lithographic resolution. Dynamic power is proportional to the total capacitance (gate and wire, see next section) and for this reason increases linearly with nodes. The leakage power which could be neglected in the past increases faster than dynamic power, became significant in 90nm and will be predominant under 20nm.

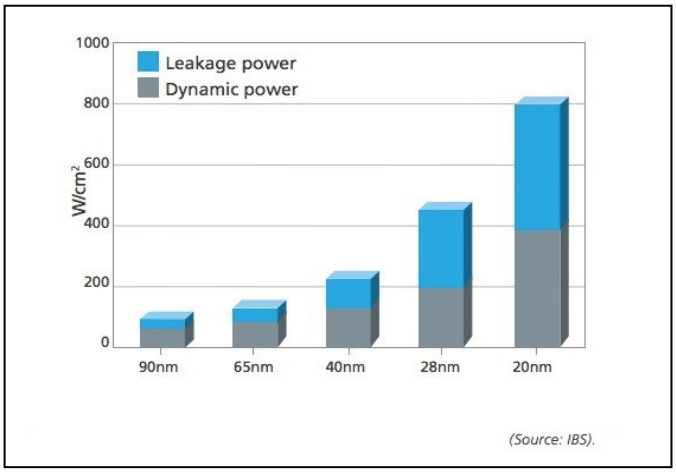

Fig. 1. Evolution of power with nodes

As a result, it can be expected that techniques wich aim at reducing leakage even on the expense of dynamic power will gain more attention in smaller nodes

B. *Evolution of Transistor Cost with Nodes*

As can be seen on following picture, the cost per device which was decreasing regularly for decades is now almost stable as a result of increasing manufacturing coast and decreasing yield.

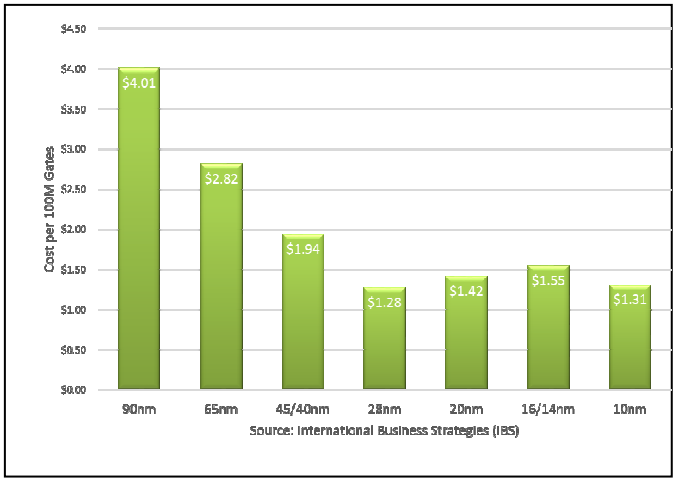

Fig. 2. Evolution of costs with nodes

As a result, it can be expected that a given node will have a longer life expectancy because the comparative advantage of the next nodes will be smaller. This will have a direct impact on the library developments as it means:

- More time to design custom cells, characterize the process and improve it
- That library development costs will be split over a higher number of chips

II. LOW POWER TECHNIQUES

The total power dissipated in a CMOS:

$$P_{Total} = P_{static} + P_{internal} + P_{switching} \quad (1)$$

$$P_{static} = V_{dd} \cdot I_{sub-threshold} = \mu C_{ox} V_{th}^2 \frac{W}{L} V_{dd} \cdot e^{\frac{V_{GS}-V_T}{nV_{th}}} \quad (2)$$

$$P_{internal} = t_{s.c.} \cdot V_{dd} \cdot I_{peak} \cdot P_{trans} \cdot f \quad (3)$$

$$P_{switching} = C_{load} \cdot V_{dd}^2 \cdot P_{trans} \cdot f \quad (4)$$

Where V_{dd} is the supply voltage, V_{GS} the ground source voltage, V_T the threshold voltage, V_{th} the thermal voltage, W and L the dimensions of the device, μ the mobility, n a factor between 1 and 2.5 that depends on the design process, C_{ox} the gate capacitance, $t_{s.c.}$ the short-circuit duration, I_{peak} the maximum current, P_{trans} the probability of transition and finally f the frequency.

One can try to classify low-power techniques according to the parameters in the above equations they target.

A. *Architectural optimizations*

When power is critical, decisions can be made on architectural level in both hardware and software, for instance splitting the memory in several clusters. These techniques are largely technology independent, one can therefore expect their frequency of use to be unaffected by the trends shown in part I.

B. *Less logic for less power*

Everything remaining equal, the less logic in a chip the less power it will dissipate. A compact design will also require less buffering and therefore dissipate less leakage and dynamic power. **Area optimization** can be in this sense considered as a power reduction technique.

Avoiding over-design due to safety-margin in multi-mode multi-corner (MMMC) timing analysis is beneficial for power, **statistical timing analysis** could be a solution even though its use is rather confidential today. Reasons for this are essentially a lack of confidence in this approach that would need to be confirmed on a larger number of productized chips for a given node. Note that different On Chip Variation (OCV) approaches have been proposed to mitigate the statistical approach CPU overhead.

Replacing flip-flops by **pulsed-latches** also decreases the total amount of logic. It requires however dedicated cells, pulse generators and/or clusters with a pulse generator and several D-latches. This technique has remained confidential until now because of additional difficulties to meet hold timing constraints and higher risks of congestion around sequential elements.

C. *Lowering the supply voltage*

The same way area optimization is an indirect way to decrease power, **timing optimization** also is. After reaching a higher frequency at a given voltage you may lower this voltage and consequently power while remaining within budget.

The clock tree typically dissipates half of the dynamic power, decreasing its voltage level while keeping the Boolean logic voltage unchanged reduces dramatically power. This requires however tailored flip-flops to handle this **low-swing clock**.

Multi-supply Voltage (**MSV**) is a static voltage level assignment to chunks of logic called “power domains”, dynamic voltage and frequency scaling (**DVFS**) or adaptive voltage and frequency scaling (**AVFS**) with feedback control are and extension of this techniques as it power or ground shutoff (**PSO**). These techniques require dedicated cells like level shifters, isolation cells, switches, always-on cells etc. and need attention throughout the verification and implementation flow involving all design stake-holders. Despite this overhead which can amount in the first trials to a 100% increase in development time, these techniques are now applied in nearly all low-power applications.

Body biasing like MSV also changes the operating range of transistors (back or forward bias).

The main issue with lowering the power supply is that speed (you need to lower also the threshold voltage) quickly conflicts with static power as shown by equation (2)

D. *Lowering frequency and/or transition probability*

Various optimizations of the **Boolean** logic like operand-isolation (prevent computation when its result is not used) or pin-swapping (assign input pins with smaller capacitance to the higher activity path) and now routinely used. They achieve dynamic power reduction with and limited cost in other criteria (timing, area and static power).

The same is true for **clock-gating**.

Asynchronous design has also been around for some time, it is very effective in term of power but suffers from a higher complexity in execution and has remained confidential.

E. *Lowering capacitance*

Besides area optimization other techniques aim at reducing the total capacitance. **Clock-mesh** substitutes traditional clock tree by a clock mesh (chessboard-like for instance but many flavors exist) in order to primarily improve the skew and the On-Chip Variation (OCV) tolerance. If cautiously tailored it can also lead to a lower power. It is used occasionally, mainly in processor design.

Clustering several single-bit flip-flops into one larger **multi-bit** cell also decreases the logic and the total capacitance in the clock-tree. The price to pay is initially library development and during implementation a higher risk of congestion.

F. *At device level*

In order to conciliate low leakage and acceptable speed cells with lower threshold voltage (LVT, i.e. faster) are reserved to timing critical signals while those with higher threshold voltage (HVT, i.e. less leaky) are used everywhere it is possible. This **multi-VT** technique is now wide-spread and most founders offer now several flavors of cells ranging beyond LVT and HVT.

The device layout can be also adjusted like in Fully Depleted Silicon On Insulator (**FDSOI**) in which a buried oxide (BOX) layer isolates the channel from the substrate allowing a much lower leakage and an easier body-biasing. This technology initially suffered from the need to redesign libraries and from higher manufacturing costs but this later handicap may be fading away.

III. TRADE-OFF BETWEEN COSTS AND BENEFIT

Let us now try to measure the costs and benefits of these techniques. Costs in the subsequent tables are the sum of area, timing library development and complexity (turn-around-time, number of stakeholders, low predictability etc.), each is given a number ranging from 0 (no cost) to 3 (high cost). Symmetrically the benefit in static and dynamic power is weighted between 0 (no benefit) and 3 (high benefit). In the figures 3, 4 and 5 the cost corresponds to the X axis and the power improvement to the Y axis. A blue separation line has been drawn arbitrarily to separate a zone of interest in the upper left from a marginal zone in the bottom right.

A. *In the past*

With earlier nodes the static power was less critical, therefore in the table below only the benefit in dynamic power is considered and used in figure 3.

TABLE I. COSTS AND BENFITS

Techniques	Costs					Benefit		
	library	complexity	area	timing	Total	static	dynamic	Total
Area optimization	0	0	0	0	0	0	1	1
Timing optimization	0	1	0	0	1	0	1	1
Statistical T.A.	2	1	0	0	3	0	1	1
Boolean Optim.	0	1	0	0	1	0	1	1
Clock mesh	0	2	1	0	3	0	1	1
Low swing clock	3	2	1	0	6	0	2	2
Clock gating	0	1	0	0	1	0	2	2
Pulsed latch	1	2	1	0	4	0	2	2
Multi bit cells	2	1	0	0	3	1	1	2
Asynchronous	0	4	0	0	4	0	2	2
PSO	1	2	1	1	5	3	2	5
MSV	1	2	1	1	5	3	2	5
DVFS/AVFS	2	3	1	1	7	3	2	5
Multi-Vt	1	1	1	1	4	3	0	3
Body bias	2	1	0	0	3	3	0	3
FDSOI	3	1	0	0	4	3	0	3

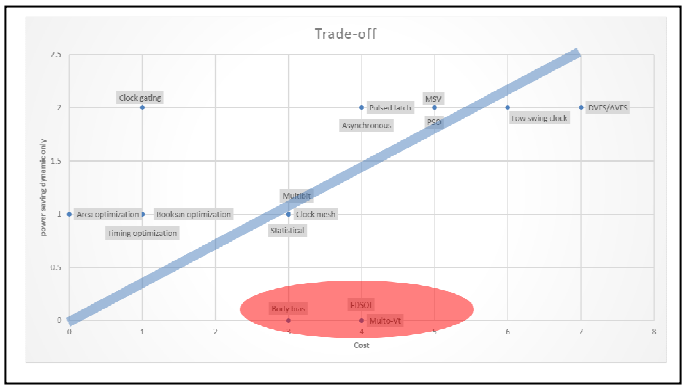

Fig. 3. Trade-off in the past

Body-biasing, Multi-Vt and FDSOI which target static power were marginal, while clock gating was already well accepted. Area and timing optimization were used anyway but not specifically for power reduction.

B. *Today*

Keeping the same number but this time depicting the total power improvement (static + dynamic):

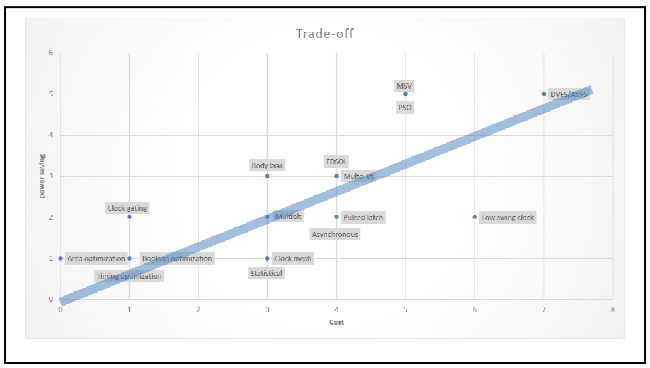

Fig. 4. Trade-off today

One can see that body-biasing, multi-Vt and FDSOI have moved toward the zone of interest. MSV, PSO and DVFS/AVFS are now well established despite their higher costs. This corresponds to the trend observed in the industry. Let us now attempt to make some predictions using the same approach.

C. *In the near future*

For the future we assume that we can neglect the library development cost which is of course a gross simplification. This yields the following and last figure

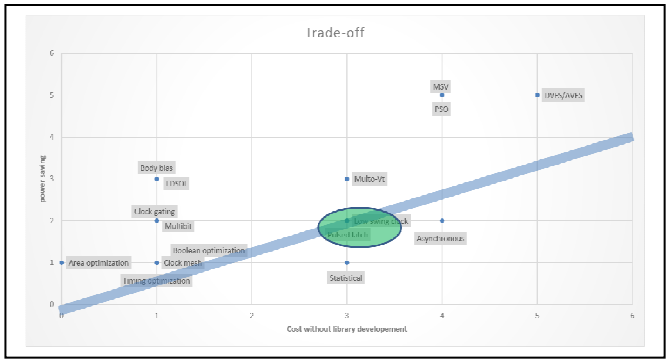

Fig. 5. Trade-off in the future

With these assumptions low-swing clock and pulsed latch move closer to the zone of interest. Asynchronous design and statistical timing analysis seem to be lagging in the marginal zone. There may be an effect on clock mesh but this technique has not benefited from its dwelling in the zone of interest in the past decades so it is unlikely that it will gain popularity for the only sake of power.

D. *Conclusion*

Increasing difficulties to manufacture thinner devices and the simultaneous increase of the power dissipation create a big engineering challenge. The demand for faster and more complex low-power (mobile) system-on-chips showing no sign of decline there is therefore a great opportunity for smarter solutions. Some techniques exist and will proliferate, other are still to be developed…

REFERENCES

Col. "A Practical Guide to Low-Power Design" regularly updated, http://www.powerforward.org

Handel Jones, "Whitepaper: Semiconductor Industry from 2015 to 2025" International Business Strategies (IBS), http://www.semi.org/en/node/57416

Scotten Jones, "Is SOI Really Less Expensive? ", June 24th 2014 https://www.semiwiki.com/forum/content/3599-soi-really-less-expensive.html

Michael Keating, David Flynn, Robert Aitken, Alan Gibsons and Kaijian Shi, "Low Power Methodology Manual for System on Chip Design", 2007 Springer Publications, NewYork

Henry Samueli, "As Moore's law slows, better design engineering will win out" May 10, 2015, http://venturebeat.com/2015/05/10/as-moores-law-slows-better-design-engineering-will-win-out/