ECONOMIC IMPACT OF THE TECHNOLOGY CHOICES AT 28nm/20nm

Dr. Handel Jones
Founder and CEO, IBS, Inc.
Los Gatos, California

White Paper
June 2012
Overview

The semiconductor industry continues to be cost-driven, and the migration to 28nm has been slowed for many companies because die and product costs are higher than those for 40nm technology. The low yields at 28nm are due to high leakage, and leakage impacts parametric yields.

The yields for 28nm products will be increased over the next 12 months, and there will continue to be design-for-manufacturability challenges with planar 28nm technologies.

There are, consequently, major benefits to consider options to 28nm bulk CMOS that can give lower leakage, lower power consumption, and higher yields, which translates into lower cost.

For a technology to be utilized in high-volume production, costs must be lower than previous generations of technology.

It is appropriate to analyze the cost factors for the different versions of 28nm as a baseline for the decisions on 20nm.

20nm bulk planar CMOS technologies will have slow ramp-up in wafer volumes, and it is appropriate for the vendors and users of the wafers to evaluate other options, which can include enhancements to 28nm.

Making the wrong technology decisions at 20nm can cost wafer manufacturers and fabless companies billions of dollars.

The decision processes for the high-volume applications should be cost-driven, and it is important to use cost data that provides an accurate perspective on the real situation that is occurring at 28nm and what is expected to occur at 20nm.

Cost Analysis

A perspective on the wafer and die costs for 100mm² and 200mm² is shown in the following table.

<table>
<thead>
<tr>
<th></th>
<th>Bulk HKMG (HP)</th>
<th>Bulk HKMG (LP)</th>
<th>FD SOI (HP)</th>
<th>FD SOI (LP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100mm² Wafer cost ($)</td>
<td>2,867.12</td>
<td>2,786.33</td>
<td>3,066.93</td>
<td>3,088.21</td>
</tr>
<tr>
<td>200mm² Wafer cost ($)</td>
<td>2,867.12</td>
<td>2,786.33</td>
<td>3,066.93</td>
<td>3,088.21</td>
</tr>
<tr>
<td>Gross die/wafer</td>
<td>650.7</td>
<td>318.5</td>
<td>650.7</td>
<td>318.5</td>
</tr>
<tr>
<td>Yield (%)</td>
<td>56.8</td>
<td>42.9</td>
<td>60.7</td>
<td>46.5</td>
</tr>
<tr>
<td>Net die/wafer</td>
<td>369.6</td>
<td>136.6</td>
<td>395.0</td>
<td>148.1</td>
</tr>
<tr>
<td>Die cost ($)</td>
<td>7.76</td>
<td>20.98</td>
<td>7.05</td>
<td>18.81</td>
</tr>
</tbody>
</table>

The table above provides a comparison of the costs for different technology choices at 28nm, including bulk HKMG (High Performance) and FD SOI (FinFET) options.
The key issues in die costs include the following:

- HP is high performance, and LP is low power, which in turn gives low leakage. Both technologies are HKMG.
- The wafer costs are those of leading foundry vendors in Q1/2013 for 8LM. Selling prices of wafers will be higher and will include the gross profit margins of the foundry vendors.
- High-volume production with wafer fab at 95%+ utilization.
- The bulk CMOS cost model is based on $3V_t$ levels in the core of the chip. There is also the need to support the SRAMs and interfaces.

The FD SOI cost is based on $1V_t$ level for the core and use of body biasing. Body biasing can give two additional $V_t$ levels in the core, which is equivalent to bulk CMOS design options. There is the option of having $4V_t$ levels, and the cost of this option is less than 7% of wafer costs, which still gives lower cost at the die level compared to other options.

There is also the option of not using body biasing, which can still give superior performance and lower cost than bulk CMOS.

- The gross die per wafer is based on the usable area of the 300mm wafers.
- The net die per wafer is a combination of defect density and parametric yields.

All die yields for 28nm are based on $V_{DD}$ of 0.8V.

The net die per wafer is those that are expected in Q1/2013 (after many existing yield problems have been addressed).

The overall yields for bulk HKMG CMOS in Q2/2012 are 65% to 75% of the levels shown, which impacts die cost by the same percentage.

- The FD SOI die costs are lower than those of HKMG bulk CMOS at 28nm, even with the $500 SOI base wafer price versus the $129 bulk CMOS wafer price. There is the expectation that the SOI wafer prices will be reduced in the future.

Because of the characteristics of FD SOI, porting IP is almost identical to bulk CMOS, and the same design flows can be used.

- The 28nm analysis indicates that there is the need for the semiconductor industry to evaluate multiple opportunities for their technology options at 28nm and at 20nm.

There is the further factor that by using 28nm FD SOI, there can be the postponing of when 20nm will need to be ramped into high volume. The cost savings for the fabless IC vendors and the foundry vendors can be very large.

Intel has already diverged from the bulk CMOS road map by selecting FinFET technology at 22nm due to concerns with the problems with scaling bulk CMOS technologies. Intel, however, is unique in the level of resources that can be applied to advanced feature process technologies and ability to focus its design activities on a small number of high-volume products, but with very close linking between design and process.
If Intel considered that 22nm bulk CMOS was not viable, it is unlikely that any other company will have good success at 20nm based on a shrink of 28nm planar technologies. However to compete effectively, introducing a FinFET 2 generations after Intel is not the optimum response, finding a path to a fully depleted device faster and cheaper to pull in their lead is the more effective strategy.

20nm Assessment

Options that are being considered for high-volume production at 20nm include the following:

- HKMG planar bulk CMOS. Variants in metal pitch are being developed (eg, use of 80nm M1 versus 64nm).
- FD SOI. Can have process flow which is similar to bulk CMOS and is a natural migration from 28nm.
- FinFET. 18nm variants are also being developed. The analysis covers FinFET on bulk.

Wafer and die costs at different stages of maturity for the technology options are shown in the following table and figure.

**TABLE 2**

<table>
<thead>
<tr>
<th></th>
<th>Bulk HKMG CMOS</th>
<th>FinFETs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100mm²</td>
<td>200mm²</td>
</tr>
<tr>
<td>Wafer cost ($)</td>
<td>4,000.06</td>
<td>3,683.30</td>
</tr>
<tr>
<td>Gross die/wafer</td>
<td>650.7</td>
<td>650.7</td>
</tr>
<tr>
<td>Yield (%)</td>
<td>36.9</td>
<td>44.7</td>
</tr>
<tr>
<td>Net die</td>
<td>240.1</td>
<td>290.9</td>
</tr>
<tr>
<td><strong>Die cost ($)</strong></td>
<td><strong>16.66</strong></td>
<td><strong>12.66</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>FD SOI (HP)</th>
<th>FD SOI (LP)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100mm²</td>
<td>200mm²</td>
</tr>
<tr>
<td>Wafer cost ($)</td>
<td>3,918.69</td>
<td>3,608.37</td>
</tr>
<tr>
<td>Gross die/wafer</td>
<td>650.7</td>
<td>650.7</td>
</tr>
<tr>
<td>Yield (%)</td>
<td>60.5</td>
<td>67.4</td>
</tr>
<tr>
<td>Net die</td>
<td>393.7</td>
<td>438.6</td>
</tr>
<tr>
<td><strong>Die cost ($)</strong></td>
<td><strong>9.95</strong></td>
<td><strong>8.23</strong></td>
</tr>
</tbody>
</table>
The analysis of the wafer and die costs shows the following:

- **Wafer costs** are based on 8LM, with capacity utilization at 95% within a high-volume wafer facility.
- **Bulk HKMG CMOS wafer costs** are based on 3V levels in the core plus support of SRAM and 1.5V.
- 20nm **FinFET wafer costs** are based on a shrink of existing technologies within a low-cost, high-volume foundry environment and are based on FinFET on bulk. The cost takes into account the relatively long time for metrology checking in the process and also the manufacturing complexity related to the FinFET structures.
- The lowest-cost wafers at 20nm are for FD SOI options, which are based on SOI wafer prices at $500. There is the expectation that SOI wafer prices will decline as volumes increase. The reason for the lower wafer cost compared to the other technologies is lower mask count levels.
- **Die yields** shown are based on V_{DD} of 0.9V for all technologies. It is understood that if V_{DD} is 1.0V, die yields will be higher, but with the penalty of high power consumption (operating power is a function of CV^2).

FD SOI can support V_{DD} down to 0.7V, and while there is some reduction in performance, operating power is reduced, giving a very compelling performance-power advantage against other technologies.
Power density will become an increasing challenge for all technologies at 20nm, but specifically for FinFETs.

- Die yields for bulk HKMG CMOS are low due to the challenges in meeting performance levels that are established by FinFET metrics. A major source of yield loss for HKMG bulk CMOS is that of random dopant fluctuations from transistor implants. These implants are not required for FD SOI.
- Die yields for FD SOI (includes body biasing) are significantly higher than those for other technology options, resulting in much lower die cost.

There is also the power/performance advantage of FD SOI, compared to the other technologies, which is highly beneficial for the high-performance processors as well as for mobile platforms.

The power/performance characteristics of FD SOI with body biasing, and also without body biasing, are significantly superior to bulk CMOS at 20nm, a perspective of which is shown in the following figure.

**FIGURE 2**
Power/Performance Comparison

Comparing SOC power and performance for planar FD SOI vs. standard (planar) bulk silicon design at the 20nm node.

(a) (b) (c) Highlights the ability of FD SOI to go into high performance or low leakage modes just by adjusting the body bias voltage. This ability is available at any voltage operating point.

(1) 20nm bulk CMOS.
(2) 20nm FD SOI.
(3) Target frequency in 20nm bulk CMOS.
(4) Same frequency as 20nm bulk CMOS achieved in 20nm FD SOI highlighting total power reduction of approximately 30% to achieve same frequency as 20nm bulk CMOS.
FD SOI presents an excellent speed/power performance range. In fact, the FD SOI advantage increases significantly at low voltages, ie, greater than 5X boost in performance. Today, FD SOI is the only technology that can operate safely in the 0.6V to 0.7V range.

The time to reach defect density-related yields with allowance impact of parametric yields is 12 to 18 months for FD SOI versus 24 to 36 months for bulk HKMG CMOS and for FinFETs.

The faster ramp-up of wafer volumes combined with more predictable yield ramp-up provides additional cost benefits to using FD SOI over other options at 20nm.

Summary

- Bulk HKMG CMOS will have low parametric yields at 20nm, and the rate of ramp-up to 90% of $D_0$ yields will be 24 to 36 months.

- 20nm FinFET structures will be high-cost to manufacture, and parametric yields will be low.
  The time to ramp up to $D_0$ yields will be 24 to 36 months.
  Also, the operating voltage of FinFETs is likely to be 0.9V or potentially 1.0V.

- 20nm FD SOI gives the lowest wafer cost even allowing for the $500 price of the SOI wafer.
  Parametric yields for FD SOI will be 63% higher than bulk CMOS at 20nm for die size 100mm².
  Power/performance characteristics of FD SOI will be 30% to 40% superior to bulk HKMG CMOS at 20nm. Analog porting of FD SOI will be easier than with the other options because of the superior sub-threshold characteristics.

  While the real competition is likely to be between FinFETs and FD SOI at 20nm, FinFETs are a new technology (high-volume production perspective), with significant cost penalties even in Q1/2016.

Multiple factors need to be considered with the migration to 20nm, and the highly visible experience to date in attaining high yielding, volume production on 40nm and 28nm from the industry’s largest players provides visibility into what is likely to happen at 20nm bulk.

Cost per chip or per function is a key factor in deciding which technology to adopt at 20nm. The IC industry is highly cost-driven, and it is critical to have good baseline data for costs.

The semiconductor industry will, however, face additional challenges at 14nm and for sub-10nm technologies, but there is the near-term need to make the correct decision at 20nm. The costs associated with making the wrong decisions can be billions of dollars.