A 32nm SoC Platform Technology with 2nd Generation High-k/Metal Gate Transistors Optimized for Ultra Low Power, High Performance, and High Density Product Applications


Logic Technology Development, Intel Corporation, Hillsboro, Oregon, USA

Abstract

A leading edge 32nm high-k/metal gate transistor technology has been optimized for SoC platform applications that span a wide range of power, performance, and feature space. This technology has been developed to be modular, offering mix-and-match transistors, interconnects, RF/analog passive elements, embedded memory, and noise mitigation options. The low gate leakage of the high-k gate dielectric enables the triple transistor architecture to support ultra low power, high performance, and high voltage tolerant I/O devices concurrently. Embedded passives include high density (0.148 um^2) and low voltage (0.171 um^2) SRAMs as well as secure OTP fuses. Analog/RF SoC features include high precision and high quality passives (resistors, capacitors and inductors) and deep-nwell noise isolation.

Introduction

As Moore’s Law continues to guide CMOS scaling down to the 32nm node, SoC (System-on-Chip) integration of numerous functional circuit blocks has become the mainstream IC manufacturing trend. However, new challenges to accommodate diverse requirements from integrating different system components constantly arise. Examples include high performance cores, low standby power always-on circuitry, high voltage I/O, high frequency RF, and precision analog circuits. This paper reports on a leading edge 32nm high-k/metal gate SoC technology with a mix-and-match triple transistor architecture that enables independent optimization of transistor characteristics to meet the needs of different SoC circuit blocks.

Transistor Architecture and Process Flow

Three transistor families are offered simultaneously – logic (HP or LP), ultra low power (ULP), and high voltage I/O (Fig. 1). The process flow and device characteristics are shown in Tables I and II. Historically, the triple gate transistor architecture is very expensive to implement; the extremely low gate leakage of the high-k dielectric has enabled a much simpler triple gate implementation by allowing the logic and ULP transistors to share the same high-k dielectric layers. I/O transistors utilize a composite gate dielectric stack with a pre-patterned thermal CDO layers are used in interconnect stacks except for top two layers [2].

Logic Transistors – High Performance/ Low Power

The logic transistor family features the most aggressive 112.5 nm gate pitch at the 28/32nm nodes with two options- high performance (HP) or low power (LP). Either can be paired with the ULP transistors. The HP option is based upon the CPU process with a 30 nm Lgate [2]. The ULP option uses a 4 nm longer gate length and employs low damage implants and junction grading to lower the subthreshold leakage. NMOS/P莫斯 drive currents are 1.09/0.87 mA/um at 1 V and 1 nA/um Ioff (Fig. 2) which is 15% faster than 45 nm MP process [3]. Figs. 3 and 4 show well controlled transistors Vt roll-off, subthreshold, and transfer characteristics. All devices exhibit dynamic ranges down to a low 0.75V for low active power modes.

Ultra Low Power Transistors

Ultra low power (ULP) transistors are required for low standby/always-on circuit applications. Longer channel lengths combined with high threshold voltages enable reduction of the subthreshold component (Fig. 5). However, in the 10 pA/um region, the subthreshold component is no longer an adequate indicator of the device leakage behavior. All leakage components, including gate, junction, and subthreshold must be suppressed. High-k dielectrics provide the advantage of extremely low gate leakage. Co-implant species optimization and junction grading techniques are used to minimize junction leakage (Fig. 6). A leakage metric Ileakage is used to guide the process optimization:

\[ I_{\text{leakage}} = \frac{1}{2} \left( I_{\text{on}} \right) + \frac{1}{2} \left( I_{\text{off}} \right) + \frac{1}{2} \left( I_{\text{off}} \right) + \frac{1}{2} \left( S.F. \times I_{\text{on}} \right) + I_{\text{junction}} \]

The improvement over 45nm high-k is from the composite dielectric thickness scaling to improve short channel effects (Fig. 8). The TDDB reliability for high-k/metal gate is as robust as the SiON process (Fig. 9).

Interconnects

A total of seven to eleven layers of metals are supported with tighter pitch upper metal layers to improve routing density whereas CPU interconnects are focused on RC performance (Fig. 10). Multiple 1x and 1.5x pitch layers are used for local routing while 3x and 4x pitch layers are reserved for semi/global/global routing. A thick 7 um top metal layer is offered for on-die power distribution (Fig. 11). Lower-k etch stop and CDO layers are used in interconnect stacks except for top two layers [2].

Mixed Signals and RF Features

Resistors supported include well resistors, trench contact linear resistors, and precision linear resistors capable of < 0.5% matching and low TCR. The low resistivity top metal enables high Q inductors with a quality factor of 20 (Fig. 12). Capacitors supported include MIS and MOM metal finger capacitors with Q > 100 in the RF frequency regime, Fig. 13. Other critical SoC features supported include varactors, vertical BJTs, bandgap diodes, and high density secure fuses/OTP memory.

Fig. 14 shows 1/f flicker noise performance for high-k/metal gate is equivalent to 65nm oxide/poly process. This is achieved by dielectric and silicon engineering to eliminate interfacial traps. Fig. 15 shows the optimal deep-nwell process to reduce substrate noise by 50dB at low frequencies by capacitive decoupling. Low substrate noise is crucial for RF/analog applications such as LNA, ADC/DAC, and PLL.

Low Voltage and Low Standby Power SRAM

Several 6T SRAM are supported, including high density (0.148um^2) and low voltage/high performance (0.171um^2) bit cells (Figs. 16). Fig. 17 shows the low standby bit cell leakage (10-20pA/cell) is achieved at 0.6V. Fig. 18 demonstrates SRAM operating frequencies of 3.8GHz, 2.9GHz and 2GHz for HP, LP and ULP processes, respectively at 1.1V. The ULP SRAM frequency is 2x faster than the 65nm LP SRAM [3]. Healthy low Vmin and high volume manufacturing yield on these SRAM cells have been routinely collected from 291Mbit test chips.

Summary

A leading edge 32nm high-k/metal gate technology optimized for ultra low power and high performance SoC products has been developed. The triple transistor architecture enables logic transistors with extremely high drive currents and record low leakages in a single chip.

References

Table I. Transistor options, design rules and characteristics summary

<table>
<thead>
<tr>
<th>Transistor Type</th>
<th>Logic (option of HP or LP)</th>
<th>ULP</th>
<th>High Volt I/O (option of 1.8 or 3.3 V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Options</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vdd (V)</td>
<td>0.75 / 1</td>
<td>0.75 / 1</td>
<td>0.75/1.2</td>
</tr>
<tr>
<td>Gate Pitch (nm)</td>
<td>112.5</td>
<td>112.5</td>
<td>126 min, 338, 675</td>
</tr>
<tr>
<td>Lgate (nm)</td>
<td>34</td>
<td>34</td>
<td>46 min, 140, 320</td>
</tr>
<tr>
<td>NMOS/PMOS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Idsat (mA/um)</td>
<td>1.56 / 1.9</td>
<td>1.99 / 0.88</td>
<td>0.72 / 0.55 / 0.68 / 0.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>@ 1 V, 100 nm/um, 1 nA/um, 100 nA/um</td>
</tr>
</tbody>
</table>

Table II. Interconnects (left) and process flow (right) summary

<table>
<thead>
<tr>
<th>Layer</th>
<th>Pitch (nm)</th>
<th>CPU Layer</th>
<th>SOC # of Lyrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>MT 1x</td>
<td>112.5</td>
<td>MT 1/2/3</td>
<td>2-5 layers</td>
</tr>
<tr>
<td>MT 1.5x</td>
<td>168.8</td>
<td>MT 4</td>
<td>2-3 layers</td>
</tr>
<tr>
<td>MT 2x</td>
<td>225</td>
<td>MT 5</td>
<td></td>
</tr>
<tr>
<td>MT 3x</td>
<td>337.5</td>
<td>MT 6</td>
<td></td>
</tr>
<tr>
<td>MT 4x</td>
<td>450</td>
<td>MT 7</td>
<td>MT Top-2</td>
</tr>
<tr>
<td>MT 5x</td>
<td>504</td>
<td>MT 8</td>
<td>MT Top-1</td>
</tr>
<tr>
<td>MT Top</td>
<td>19.4 um</td>
<td>MT 9</td>
<td>MT Top</td>
</tr>
</tbody>
</table>

Fig. 1 High-k/metal gate triple transistor architecture

Fig. 2 Idsat vs. Ioff of Logic (HP/LP) transistors at 1 V

Fig. 3 Vgs-Vs-L (right) and sub-threshold characteristics (left) for logic (HP/LP options) and ULP transistors

Fig. 4 Typical Id-Vds of logic and ULP transistors

Fig. 5 Idsat vs. Ioff of ULP transistors at 0.75 V and 1 V

Fig. 6 Ioff vs. Ijunction and total leakage (Itotal) optimization for the final process

Fig. 7 I/O transistors (1.8 V) Idsat vs. Ioff comparison of 65 nm oxide/poly vs. 45/32nm high-k/metal gate

Fig. 8 XTEM images of I/O transistor (left) and the composite gate stack (right) of the I/O transistor
Fig. 9 Logic/ULP and high voltage NMOS/PMOS TDDB

Fig. 10 Interconnect architecture comparison of CPU and SoC technology

Fig. 11 Thick top metal layer for power gating

Fig. 12 Thick metal inductor image and quality factor plot

Fig. 13 Metal finger capacitor and quality factor plot

Fig. 14 1/f flicker noise trending for 65/45/32 nm

Fig. 15 Improved noise isolation by deep N-well structure

Fig. 16 6T SRAM options: high density, low leakage and low voltage/high performance

Fig. 17 SRAM bit cell leakage vs. voltages

Fig. 18 SRAM operating frequency Schmoo plot