k (8–12) up to a true high *k* of 22–24. The gate electrode for some companies will remain poly ..." name="description">

SolidState TECHNOLOGY

August 2007 Exclusive Feature #2

VLSI SYMPOSIUM REPORT: Chipmakers, consortia reveal HK+MG integration

By John Borland, contributing editor, Solid State Technology

The VLSI Symposium meeting this year (June 12–14, Kyoto, Japan) revealed there will be not one, but many different solutions for the production implementation of hafnium-based oxides at the 45nm node and beyond, with Hf-based dielectric *k* values varying from a "medium"-*k* (8–12) up to a true high *k* of 22–24. The gate electrode for some companies will remain poly, while others will use a very thin metal/thick poly stacked layer (metal inserted polystack, "MIPS") and others will use a metal-only electrode. If metal is used, it can be single metal, or dual metal for nMOS and pMOS.

Various gate process flow integration approaches will also be used: gate-first; gate-first, but with disposable spacer (reverse source drain); gate-last (replacement gate); and a hybrid of gate-first for one type of MOS transistor, and gate-last for the other type of MOS transistor. The question is, how long can the industry support these multiple approaches? Some technologist say they must converge by the sub-32nm node, while others feel it may never converge.

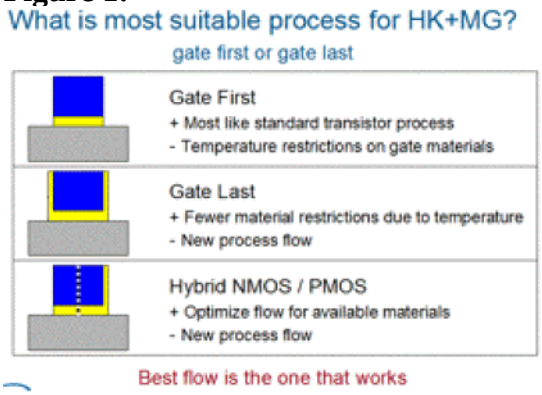
Handicapping the HK+MG field

Technologists from Intel, NEC, IBM, Toshiba, Samsung, and SEMATECH participating in an evening panel discussed the status and prognosis for high-*k*/metal gate (HK+MG) transistors, suggesting that there will be many different solutions for high *k* with *k*values varying from 8–24 in combination with gate electrodes including poly only, poly/metal stack, or pure metal?though interestingly, however, no one on the panel mentioned fully silicided (FUSI) as an option for metal gate.

Mark Bohr, Intel senior fellow in logic technology development, listed Intel's key HK+MG milestones, going back to late 2003, spanning early work on 45nm SRAMs to earlier this year touting working HK+MG processors. Main issues with gate-first are temperature restrictions and etching control of two different metals for nMOS and pMOS at the same time (the company uses ALD for control of thickness), he noted, adding that gate-last makes the process very flexible but requires new process flows. A third

option, he suggested, is a "hybrid" that allows gate-first for one type of MOS transistor, and gate-last for the other type of MOS transistor, providing more flexibility for process optimization **(Fig. 1)**. For all three models, the best flow is the one that works, he noted.

Figure 1:



Bohr said he believes the industry will see all three options over the next few years, until the industry perhaps converges on one approach. He also pointed out that HK+MG will crossover to become >50% of Intel's CPU shipments by 3Q08 **(Fig. 2)**, perhaps less than a year after entering into mass production. Bohr also noted that HK+MG adds 4% total wafer processing costs, the cost of metal + via layer, or 1/3 the cost of SOI wafers. The biggest benefit Intel sees with HK+MG is power reduction, with 1/10 the gate oxide leakage compared to 65nm with SiON and 1/5 the source drain leakage **(Fig. 3)**. The second-largest benefit is seen to be a performance increase (20% higher drive current), followed by scaling.

Figure 2:

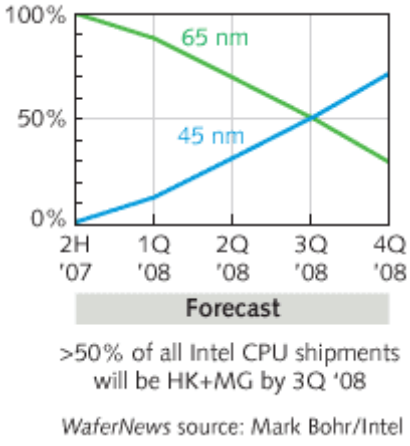
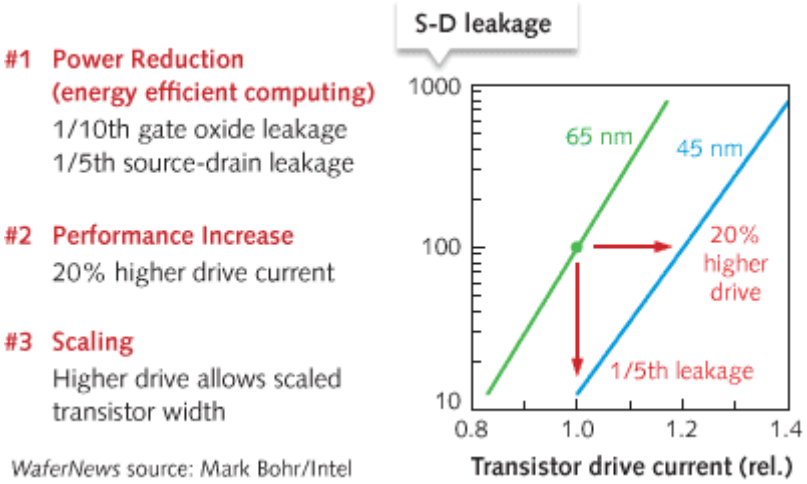


Figure 3:



K. Imai of NEC noted that his company introduced high *k* with poly at the 55nm node for LSTP devices, but that metal will not follow until 2009. They use gate-first, and believe work function tuning is the number one issue, with FUSI pattern dependency also a concern. With gate-last, though, the only issue is CMP, which is an easy, but more complex process, he indicated. Therefore, he said, NEC uses poly/HfSiON for nFET and poly/TiN/HfSiON for pFET transistors.

Samsung's U-I. Chung said that gate-first is the only option for memory for self aligned contacts, and he sees three possible gate-first approaches for DRAMs. Option No.1 is dual poly gate, and would use high-*k* gate dielectric with n+ and p+ poly electrodes (no metal). The second option is single metal gate, using HK+MG for pMOS and high-*k*/n+ poly for both nMOS and cell transistors. Option No.3 is a dual metal gate approach, with HK+MG#1 for pMOS and HK+MG #2 for nMOS, with the gate-first approach.

With respect to the use of MOCVD, Chung noted that Samsung uses MOCVD for >1.2nm EOT and ALD for 1.2nm EOT. The Toshiba panelist, Y. Tsunashima, added that his company utilizes MOCVD for >1nm EOT and ALD for 1nm EOT.

Mukesh Khare, senior manager of IBM Research, explained that gate-last is very flexible, but there are many integration challenges; while the gate-first material challenges, though complicated, are preferred because of the low cost. He prognosticated that 50% of the industry would be using high *k* by the 32nm node. For details, he pointed to IBM's paper presented later at the VLSI Symposium (11A-1, "High performance high-*k*/metal gates for 45nm CMOS and beyond with gate first processing"), in which researchers reported using a very thin single metal with a very thick poly on top for the HK+MIPS electrode, and said they dope the poly back for the desired n+ and p+ V_t targets **(Fig. 4)**. IBM's presentation was different from the published paper, however?most notably in that the spike temperature was >1080°C in the paper, but mentioned as 1000°C in the presentation. Also, in the paper IBM mentioned using spike + advanced annealing, but no mention of msec (advanced) annealing was made in the presentation.

In other papers discussed at the VLSI Symposium, Gen Tsutsui of NEC finally revealed the company's Hf-oxide (paper 10A-2, "A cost effective LOP/LSTP integrated CMOS platform utilizing multi-thickness SiON gate dielectrics with Hf for 45nm node"), reporting use of a stacked SiON/Hf dielectric layer with the same Hf deposition on four different SiON thicknesses, to realize EOTs of 1.35nm, 1.75nm, 3.2nm, and 7.0nm.

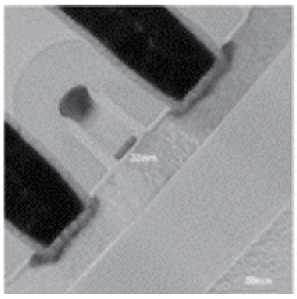


Figure 4. TEM image of the HK/MG nFET at 32nm L gate.

Meanwhile, Renesas reported on using n+ poly/metal/high *k* for pMOS while nMOS remained poly with an n+ poly/high-*k* (paper 12A-1, "Advanced poly-Si NMIS and poly-Si/TiN PMIS hybrid gate high-*k* CMIS using PVD/CVD stacked TiN and local strain technique"). Unique to Renesas' process is an in situ phosphorus-doped n+ poly for both the nMOS and pMOS poly gate electrode, rather than a separate p+ poly for the pMOS electrode. The researchers also mentioned using spike RTA + laser annealing.

Consortia pushing HK+MG agendas

For the 45nm node, SEMATECH first looked at FUSI, then replacement gates, then finally settled on a gate-first high-*k* (22–24) metal-inserted poly stack (MIPS) electrode, according to Byoung Hun Lee, program manager for advanced gate stack (and IBM assignee), in an interview with *SST*. Currently, 50% of the equivalent oxide thickness (EOT) is from the interfacial oxide layer, so this is scalable to 32nm node and it is compatible with high temperature processing, he said. The issues for pMOS can be solved with SiGe channel material for pMOS and $MoAlO_2$ for low-standby power (LSTP) pMOS, as presented in their papers 9A-1 and 9A-4.

Lee sees 45nm and 32nm as straightforward scaling with no roadblocks, but the 22nm node presents challenges. SEMATECH has a great deal of experience with etching to solve the metal etch problems with HK+MG, though, and he believes the challenges are known, and they simply need to optimize the process. Lee revealed, however, that he is personally concerned that the industry needs to come to a consensus on either gate-first or gate-last, because both approaches cannot be supported.

For frontend processing, SEMATCH is focused on process manufacturing at the 32nm node, stated Raj Jammy, director frontend process (also an IBM assignee). The group's view is that for scaled devices to be stable, tight control of gate length variation due to lithography and etching will be required, as will control of dopant placement and number of dopant atoms in the channel to prevent fluctuation, assuming no diffusion. The question is whether to control gate length scaling, or use more strain-Si, while also reducing contact resistance. To address these questions, SEMATECH is looking into Ge and SiGe channels for the 32/22nm nodes. For high *k*, the consortium's HfSiON material can go up to a *k* of 16, and its approach is gate-first because replacement gate scaling is limited. However, Jammy noted that, while addressing HK+MG for nMOS is easy, pMOS presents a problem because of the roll-off of V_{fb}. He added that for its gate-first approach, SEMATECH has developed a high-*k* Hf dielectric that can tolerate a 1075° C spike anneal.

SEMATECH has looked at 270 metal gate candidates for high-*k* metal gate work function tuning, and if the work function starts higher, it rolls off when EOT 2nm. For the 32nm node, the researchers have targeted both HP (high-performance) and LSTP devices; therefore, the issue is how to scale the gate stack. With a replacement gate, the challenge is controlling the sidewall to prevent the keyhole filling problem. Additionally, a high-*k* dielectric on the vertical sidewall will cause the spacer *k* value to increase.

Jammy noted that SEMATECH is also working on activities involving FinFET and source/drain doping.

The group has made 13nm × 60nm Fins, and wants to achieve a smooth 92° vertical Fin process that is reproducible.

Michael Polcari, SEMATECH president and CEO, told *SST* about other areas that are receiving attention, including a major effort in lithography (EUV and immersion lithography), plus looking at new materials for memory and strain-Si technology, and 3D wafer/chip bonding.

The area in which SEMATECH is seeing a great deal of interest, Polcari noted, is in its manufacturing program on fab productivity and benchmarking of fabs for optimum high-volume/low-mix, and high-volume/high-mix. To address this topic, they look at their members' fab equipment and do process equipment matching. Since they changed their model to be more flexible and allow member companies to be a program-based participant, Polcari said that the result has been that new companies have joined the consortia, including NEC, Renesas, Panasonic, TSMC, Samsung, and Micron.

Meanwhile, Serge Biesemans, director of CMOS device technology research at IMEC, told *SST* that to meet all the needs of the consortium's clients, his group is tasked with providing all three high-*k* gate integration options: 1) gate- and metal-first (MIPS), 2) gate-first and metal last (FUSI), and 3) gate- and metal-last (replacement gate). After four years and 25 papers, IMEC's FUSI approach is now complete. IMEC had eight papers at the VLSI Symposium out of a total of 86 papers, and all its HK+MG papers were on FUSI—and so work will now focus on developing the other two areas, he noted. Biesemans said the *k* value for IMEC's Hf-based oxide is 15 and it can be maintained even with laser annealing processing.

With the MIPS gate-first process, etching two different metals at the same time can be a problem resulting in etch variations. Therefore, gate/metal etch is an issue to maintain precise sidewall profile control and not variation which would increase device variability, Biesemans explained. He added that the gate-last approach avoids the gate metal etch issues. Another issue is with pFET flat band voltage roll-off when the EOT 2.0nm.

At the 45nm node, Biesemans said the majority of companies will still use SiON/poly gate stacks, and then switch to FUSI/high *k* at the 32nm node. Moving on to 22nm, his personal view is that because of parasitic capacitance, the industry may need to go into the third dimension by stacking the device, as with multi-core processors.

Other noteworthy papers at the VLSI Symposium

IBM paper 4A-2, "Strained Si channel MOSFETs with embedded silicon carbon formed by SPE," reported a new way for creating localized tensile strain for nMOS avoiding embedded SiC (eSiC) selective epitaxy by using 1–3% carbon implantation into the nMOS S/D (source/drain). The best results were obtained using 1.65% carbon, where a 35% mobility improvement was seen and a 6% drive current improvement was also seen at the expense of a 50% increase in sheet resistivity (Rs) and a 10× degradation in junction leakage **(Fig. 5)**.

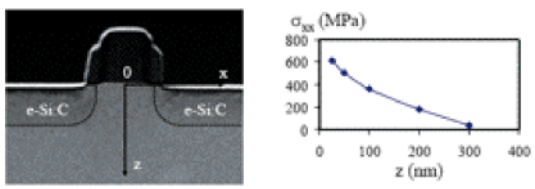


Figure 5. Cross-sectional STEM image (left) of an NFET with e-Si:C before silicidation, and (right) the uniaxial tensile stress (σ_{xx}) measured by CBED as a function of depth (z) at the center of the channel (x = 0). The stress at z = 25nm is 615 MPa.

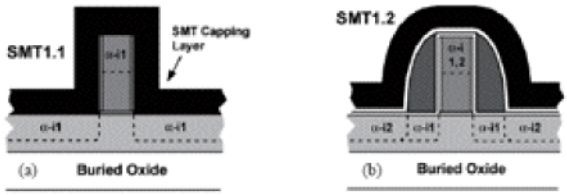


Figure 6. AMD reported using just a pre-amorphization implant with either the nSDE (n-type source/drain extension) or a deep n+S/D followed by a low-temperature solid phase epitaxy in the temperature range of 580–850°C.

In a similar paper about nMOS strain enhancement, "Multiple stress memorization in advanced SOI CMOS technologies" (12A-2), AMD reported using just a pre-amorphization implant (PAI) with either the nSDE (n-type source/drain extension) or a deep n+S/D followed by a low-temperature solid phase epitaxy (SPE) in the temperature range of 580-850°C **(Fig. 6)**. This scheme improved drive current by 10%—and adding poly stress effects with an RTA anneal, the additive effect was a 27% improvement in drive current.

In paper 9B-1, Samsung reported "Improved cell performance for sub-50nm DRAM with manufacturable bulk FinFET structure," saying the recess channel array transistor (RCAT) and sphere-shaped RCAT (SRACT) transistor designs will be extended until 2009 or 2010, at which time the company plans to introduce the FinFET for DRAMs.

John Borland is founder of J.O.B. Technologies and a member of *SST*'s Editorial Advisory Board.