EXTENDING COPPER INTERCONNECT BEYOND THE 14NM NODE

IN THIS ISSUE
• Reducing Contact Resistivity With Implants into Silicide
• Selective Metal Capping With CVD Co
• Electrical Characterization of Through-Silicon Vias
Today the semiconductor industry finds itself about to embrace the first truly revolutionary change in its history—capitalizing on the third dimension to solve growing performance issues besetting 2D scaling. This paradigm shift is driving innovations in the transistor and the interconnect in response to new requirements that accompany scaling enabled by the new dimension. We survey a number of emerging memory technologies that could become high-speed, power-saving designs.

3D transistors require metal oxide semiconductor capacitors (MOSCAPs) that can characterize oxide quality and reliability, oxide-silicon interface, and metal work function in the vertical plane. We report on development of a sidewall MOSCAP for high-κ metal gate, spacer, and liner applications being adopted for 3D structures.

High contact resistance is a growing concern as contact widths shrink. Reducing specific contact resistivity in the NMOS source/drain has been especially challenging. Now, implants of selenium or phosphorus into nickel platinum silicide are achieving substantial reduction in NMOS contact resistance while preserving excellent junction characteristics and sheet resistance.

The shrink in pitch that is accompanying transistor scaling is also posing significant challenges in interconnect fabrication technologies. We discuss interface engineering in hard masks through precision materials modification to ensure pattern integrity, thereby enabling gap fill, reducing RC delay, and ensuring high reliability at the 14nm node.

We also present a method for combating electromigration (EM) failures at the copper-dielectric interface that can occur at very high electrical current densities of advanced interconnects. Selective CVD cobalt metal capping is a promising new technology that enables the best EM performance by eliminating fast diffusion at the interface. The ultra-thin cap minimizes the impact on capacitance, requires no pre- and post-cleaning, and does not affect resistivity by diffusing into the copper.

Through-silicon via (TSV) technology is maturing beyond process development and integration testing. We report electrical characterization that verifies capacitance, leakage, and breakdown performance consistent with healthy TSV operation. These results represent important progress towards successful implementation of TSV-based 3D integration.

We are pleased to share with you the diversity of innovations in these articles as we continue to expand the frontiers of knowledge, experimentation, and solutions for the semiconductor industry.

Cover: Increasingly advanced mobile devices rely on 11-15 layers of interconnect for the functionality, reliability, and endurance consumers demand. The 14nm node will be an inflection point in interconnect development beyond which disruptive solutions in materials and device designs will be needed to sustain scaling.
Reducing Contact Resistivity
With Implants into Silicide

REDDUCING CONTACT RESISTIVITY
Equation 1 defines the basis for modifying \( \rho_c \), where \( \Phi \) is the SBH, \( n \) is the dopant concentration at the silicide-silicon interface, \( q \) is the electronic charge, and \( C_1 \) and \( C_2 \) are constants.\(^{[3]}\)

\[
P_c = C_1 \exp\left(\frac{q \Phi}{kT} + C_2 \right)
\]

Several approaches for reducing NMOs \( \rho_c \) have been reported in the literature. One is DSS with phosphorus (P) or arsenic (As) implant to introduce a high dopant concentration and enhance the tunneling current at the silicide-silicon interface.\(^{[4]}\) A second approach is to reduce the effective SBH for electrons by implanting with sulfur (S), selenium (Se), tellurium, or aluminum and thermally driving the implanted species to the silicide-silicon interface.\(^{[5-7]}\) A third technique is to reduce the work function by using rare-earth metal silicides or by incorporating additional elements into the silicide.\(^{[8]}\) Similar concepts apply for PMOS \( \rho_c \) reduction.

To reduce \( \rho_c \) for silicide in both NMOS and PMOS with low cost of ownership, an important manufacturing requirement for silicides in logic applications.

EXTRACTING CONTACT RESISTIVITY
Experiments were evaluated with a test chip designed to investigate silicide- and junction-related electrical properties. The test structures included a van der Pauw transmission line model (TLM)\(^{[9]}\), structures, and diodes, all of which are testable after silicidation, to extract silicide phase and sheet resistance (\( R_s \)), silicide/SD external resistance (\( R_e \)), \( \rho_c \), and junction breakdown (\( V_b \)). In addition, the test chip also included a cross-bridge Kelvin resistor, testable after metal-1, to extract the resistance of single contacts for correlation with TLM structures.

PROCESS FLOW
Prior to silicidation, the p-type substrate was implanted with As (C >1E20/cm\(^3\)) and spike annealed for dopant activation. Ni (10% P) and TiN films (to prevent oxidation of Ni during silicidation) were deposited in a RF physical vapor deposition chamber specifically designed to reduce feature dependence of bottom coverage and to meet technology and productivity requirements for silicides in logic applications.

![Figure 1](image-url)
Figure 1. Extraction of \( \rho_c \) from TLM structures:
(a) Design of TLM structures; the number (n) of silicide segments determines \( L \).
(b) Cross-sectional TEM of a TLM structure.
(c) Fitting of \( R_c \) vs. silicide thickness to extract \( \rho_c \).

![Figure 2](image-url)
Figure 2. (a) Principle of reducing \( \rho_c \) by implanting into the silicide, and Figures 2b and 2c outline the separate process integration schemes employed in this work for incorporating either Se or P implants into NiPt silicide. Conventionally, the silicide process involves two anneal steps in a rapid thermal processor (RTP), wet strip of TiN and excess metal in sulfuric acid-hydrogen peroxide mixture (SPM), RTP(2). Se was implanted after the wet strip, but before RTP(2). However, P was implanted after RTP(2), with an added third anneal (RTP3) to thermally activate and drive P to the silicide-silicon interface to form DSS.
SILICIDE OPTIMIZATION

Before proceeding with implantation, the baseline silicide process was optimized in two phases. Phase 1 evaluated dilute hydrofluoric acid (DHF) wet clean vs. chemical plasma dry pre-clean prior to deposition of 10nm thick Ni or NiPt film and 10nm thick TiN film to form 23nm thick silicide film.1011 Concurrently, single-step and two-step RTP silicidation anneals were compared. For the single RTP process, 450°C (30 sec) was used. In the two-step process, RTP1 was at 300°C (1 min) and RTP2 was at 450°C (30 sec). The plasma pre-clean combined with the two-step RTP process resulted in improved $V_{bj}$ for both NiPt silicide and Ni silicide (Figure 3a); $R_s$ measurements (post-wet strip) on blanket wafers vs. temperature of RTP1 (N2, 30 sec) confirmed that NiPt silicide is stable over a wider temperature range than conventional Ni silicide (Figure 3b).

Figure 3. (a) NiPt silicide $V_{bj}$ vs. optimization scheme. (b) Silicide $R_s$ vs. temperature. NiPt silicide has wider temperature stability than NiSi.

Phase 2 of the optimization used the plasma pre-clean in combination with a 7.5nm NiPt layer to form thinner silicide (20nm) for compatibility with ultra-shallow junctions at advanced CMOS nodes. Furthermore, the two-step RTP process used a lower-temperature RTP1 (270°C, 30 sec), and replaced the RTP2 soak anneal with laser anneal (0.5 ms millisecond pulse). Within-wafer temperature splits were evaluated to reduce $R_s$ for the thinner silicide while also minimizing the thermal budget.

IMPLANT INTEGRATION

After the silicide module was optimized, the Se and P implants were integrated as shown in Figure 2. The implants were performed in a single-wafer high-current implanter, using a solid SeO source heated in a vaporizer for Se implants and a PH3 gas source for P implants. To avoid junction leakage, the implant energies were designed to place the peak and tail of the implant within the silicide. Se implant experiments used splits in energy (8-10keV) and dose (2E15-5E15/cm2) into 20nm NiPt silicide film. For P implants, energies of 4-6keV were evaluated with doses of 2E15-5E15/cm2. The implants were integrated into the silicide module using the plasma pre-clean and 270°C (30 sec) soak anneal for RTP1. For Se implant experiments, laser anneal was employed for RTP2 using within-wafer temperature splits (750-900°C) after the implant. However, for P implant experiments, RTP2 used a laser anneal at fixed temperature (800°C), followed by P implant, and an RTP3 laser anneal was added using within-wafer temperature splits (750-900°C).

Figure 4 shows normalized $\rho$ for various implant splits vs. laser anneal temperature (RTP2 for Se implants and RTP3 for P implants), in comparison to $\rho$ of the unimplanted reference sample. Se implant (10keV, 5E15/cm2 dose) combined with 900°C laser anneal for RTP2 reduced NMOS $\rho$ by 45%. P implants reduced $\rho$ by 45% for 6keV, 5E15/cm2 dose with 900°C laser anneal for RTP3.

Neither implant adversely affected silicid $R_s$ (Figure 5a) for P implants. $V_{bj}$, (for 1nA/um2 applied current) over a wide temperature range, confirming a satisfactory process window (Figure 5b). Cross-sectional TEM images showed no agglomeration of NiPt silicide nor crystalline damage or piping in silicon with either implant, confirming the integrity of the silicide, interface, and the junction (Figure 5c).

Figure 4. Normalized $\rho$ vs. implant condition and laser anneal temperature (RTP2 for Se implants and RTP3 for P implants). $\rho$, was reduced by up to 45% with either implant into NiPt silicide.

Figure 5. (a) NiPt silicide $R_s$ vs. laser anneal temperature for Se and P implant splits compared to the unimplanted reference. (b) $V_{bj}$ vs. laser anneal temperature (RTP2 for Se, RTP3 for P). (c) Cross-sectional TEM image shows no damage to NiPt silicide or piping with implant.

CONCLUSION

NMOS contact resistance was reduced by up to 45% by implanting either Se or P into NiPt silicide film, followed by thermally activating and driving the dopant towards the silicide/silicon interface. The silicide module optimization included a plasma pre-clean, a thinner NiPt film (7.5nm), a 270°C soak anneal for RTP1, and laser anneals (750-900°C) for RTP2 and RTP3. Results demonstrated satisfactory process window, while maintaining excellent junction characteristics without degrading $V_{bj}$. 

Figure 5. (a) NiPt silicide $R_s$ vs. laser anneal temperature for Se and P implant splits compared to the unimplanted reference. (b) $V_{bj}$ vs. laser anneal temperature (RTP2 for Se, RTP3 for P). (c) Cross-sectional TEM image shows no damage to NiPt silicide or piping with implant.
REFERENCES


AUTHORS

Kalpatrum ‘Vivek’ Rao is a process applications manager in the Varian Semiconductor Equipment business unit of the Silicon Systems Group at Applied Materials. He holds his Ph.D. in materials science from the University of Arizona and an MBA from Babson College.

Chi-Nung Ni is a senior process engineer in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He earned his Ph.D. in materials science from the University of California, San Diego.

Fareen Adeni Khaja is a senior applications engineer in the Varian Semiconductor Equipment business unit of the Silicon Systems Group at Applied Materials. She received her M.S. in electrical engineering from North Carolina State University.

Adam Brand is the director of transistor technology in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He holds his M.S. in electrical engineering from MIT.

Naushad Variam is a director of strategic marketing in the Varian Semiconductor Equipment business unit of the Silicon Systems Group at Applied Materials. He earned his Ph.D. in chemical engineering from the University of Washington.

ARTICLE CONTACT

KV_Rao@amat.com

PROCESS SYSTEMS USED IN STUDY

Applied Endura® Sicon™ Pre-Clean Chamber
Applied Endura® Avern™ RF PVD
Applied Varius™ Trident High Current Implanter
Applied Radiance™ Plus RTP
Applied Vantage™ Astra™ DSA

With continued scaling, memory devices have been transitioning to 3D structures with gate dielectric on the sidewalls, and logic devices have already begun migrating from planar to fin field effect transistors (FinFET). A 3D equivalent of the traditional metal oxide semiconductor capacitor (MOSCAP) was therefore developed for characterizing oxide quality and reliability, oxide-silicon interface, and metal work function in the vertical plane. The sidewall MOSCAP performed successfully in high-k metal gate, spacer, and liner applications to be used in future 3D monolithic integration.

Gate dielectrics are the most heavily characterized as they determine transistor performance. In 3D transistors, critical dielectrics are frequently deposited or grown on sidewalls. Device processing technology is also becoming increasingly sophisticated, and the dielectrics on sidewalls often possess material properties that differ from those on a horizontal plane (e.g., composition, density, or thickness). Planar MOSCAPs are not able to accurately represent the properties of the dielectrics on the sidewall.

The 3D MOSCAP developed here for dielectric characterization can also be used in high-k metal gate, spacer, and liner applications for 3D monolithic integration of future-generation devices. It uses thick top and bottom dielectric layers, and a thin sidewall dielectric layer to separate the top and bottom electrodes.

Contact Resistivity

Reducing Contact Resistivity

Electrical Characterization of Sidewall Dielectrics

Extending Copper Interconnect

Selecting Metal Capping With CVD Co

Electrical Characterization of Through-Silicon Via

Developing 3D Architectures

With continued scaling, memory devices have been transitioning to 3D structures with gate dielectric on the sidewalls, and logic devices have already begun migrating from planar to fin field effect transistors (FinFET). A 3D equivalent of the traditional metal oxide semiconductor capacitor (MOSCAP) was therefore developed for characterizing oxide quality and reliability, oxide-silicon interface, and metal work function in the vertical plane. The sidewall MOSCAP performed successfully in high-k metal gate, spacer, and liner applications to be used in future 3D monolithic integration.

Gate dielectrics are the most heavily characterized as they determine transistor performance. In 3D transistors, critical dielectrics are frequently deposited or grown on sidewalls. Device processing technology is also becoming increasingly sophisticated, and the dielectrics on sidewalls often possess material properties that differ from those on a horizontal plane (e.g., composition, density, or thickness). Planar MOSCAPs are not able to accurately represent the properties of the dielectrics on the sidewall.

The 3D MOSCAP developed here for dielectric characterization can also be used in high-k metal gate, spacer, and liner applications for 3D monolithic integration of future-generation devices. It uses thick top and bottom dielectric layers, and a thin sidewall dielectric layer to separate the top and bottom electrodes. Measured electrical results are combinations of the sidewall capacitor and two parasitic capacitors at the top and bottom of the structure. The geometries are designed such that the sidewall capacitor dominates in current and capacitance measurements. Thermal oxide and a low temperature, cyclic plasma-enhanced chemical vapor deposition (PECVD) oxide were used to demonstrate sidewall dielectric characterization by electrical and materials analysis.

MOSCAP DESIGN AND FABRICATION

The 3D MOSCAP is conceptually similar to a planar MOSCAP, except that the dielectrics between the two electrodes are vertical. Three dielectrics are present: >25nm SiN on top of the silicon fin, ~8nm of sidewall dielectric, and >50nm bottom shallow trench isolation (STI) oxide. The sidewall dielectric is much thinner than the top SiN and bottom STI oxide; total sidewall area is much larger than the area of the top SiN and bottom STI oxide. Hence, the sidewall dielectric dominates the total measured current. Related measurements, such as leakage, breakdown voltage ($V_{BD}$), stress-induced leakage current (SILC), and time-dependent dielectric breakdown are also dictated by the sidewall capacitor.

Figure 1a illustrates the 3D MOSCAP process flow. After fin patterning, STI fill and CMP were carried out, followed by partially recessing STI oxide to expose the top portion of the fin. Sidewall dielectric ISG5 thermal oxide (in-situ steam generation for radical oxidation) was subsequently grown by rapid thermal processing on the exposed fin.
Modeled total capacitance contribution to total capacitance is reduced and sidewall area and thickness are appropriate.

EXPERIMENTS
To validate the 3D MOSCAP design, we tested high-temperature oxide (HTO) and conformal, low-temperature (down to 200°C) cyclic PECVD oxide, with ISSG oxide as the reference. HTO is not known for good conformity, but was used to calibrate and demonstrate the electrical testing and analysis methodology. Experiments showed that it had a lower breakdown field than ISSG.

Figure 5 shows almost identical C-V and current-voltage (I-V) plots for the cyclic PECVD film and ISSG oxide. $V_{bd}$ was also plotted against EOT and demonstrated comparable performance. Additionally, $V_{bd}$ vs. EOT on horizontal plane and on sidewall are comparable. Converting $V_{bd}$ to breakdown electric field also yielded comparable values for ISSG and cyclic PECVD oxide on sidewall surfaces.

SILC measurements were also compared for cyclic PECVD and ISSG oxides. SILC measurement is an accelerated test of oxide quality widely used for tunnel oxide characterization in NAND Flash. It is the increase in leakage from inelastic trap-assisted tunneling that occurs at low gate voltage across an oxide layer after a high electric field stress. Traps can be present in as-grown dielectric; they can also be generated during electrical stress. An increase in traps causes higher leakage.

To obtain accurate EOT data for the sidewall dielectrics based on capacitance voltage (C-V) measurements of a 3D MOSCAP, three different test structures were fabricated as detailed in Table 1.
Figure 6 compares two different deposition conditions of cyclic PECVD with ISSG: Deposition temperature for condition 1 was higher than that for condition 2. SILC measurements were carried out by: 1) first I-V scan; 2) constant current density stress at 0.01A/cm² for 10 sec; 3) second I-V scan; 4) constant current density stress at 0.1A/cm² for 10 sec; 5) third I-V scan; 6) constant current density stress at 1A/cm² for 10 sec; and 7) fourth I-V scan. As expected, the cyclic PECVD oxide deposited at high-temperature condition 1 showed better SILC than condition 2 and is comparable to ISSG oxide.

CONCLUSION

Sidewall MOSCAP electrical structures were successfully demonstrated for process and materials development in high-k metal gate, spacer, and liner applications for 3D monolithic integration in future-generation devices.

ACKNOWLEDGEMENTS

The authors thank the Maydan Technology Center Group for fabricating the device structures and the DSM group for cyclic PECVD oxide process development. In particular, thanks go to M-P Cai, O. Chan, H. Chen, A. Dent, R. Hung, M. Jin, W. Lee, D. Mao, P. Nguyen, C-N Ni, A. Noori, M. Okazaki, R. Ramirez, P. Xu, and X. Xu. Thanks also to C. Olsen and X. Zhang for technical discussions.

AUTHORS

Bingxi Wood is a senior technology program manager in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. She holds her Ph.D. in physics from Rensselaer Polytechnic Institute.

Brendan McDougall is an integration engineer in the transistor technology group of the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He earned his Ph.D. in physics from Brandeis University.

Mei-Yee Shek is a technology manager in the Dielectric Systems and Modules business unit of the Silicon Systems Group at Applied Materials. She received her B.S. in chemical engineering from the University of California, Berkeley.

Chorng-Ping Chang is a director in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He holds his Ph.D. in nuclear engineering from the University of California, Berkeley.

ARTICLE CONTACT

Bingxi Wood, bingxi amat.com

PROCESS SYSTEMS USED IN STUDY

Applied Centura® Advantage™ Myos™ Etch
Applied Centura® PolyGen™ CVD
Applied Centura® SilGenPlus™ LPCVD
Applied Centura® URTima™ HDP CVD™
Applied Endura® Versa™ XLR W PVD
Applied Producer® API™ PECVD
Applied Reflexion® LK CMP
Applied Vantage® Radio™ RTP

The modern MOSFET era has seen numerous changes in transistor materials and fabrication, a trend expected to accelerate from the 2x nm node onward in response to expanding demands for mobility and connectivity. More so than at earlier nodes, lower power consumption, and higher speed and packing density are the key factors driving today’s transistor scaling and related design innovations. These, in turn, have spurred new interconnect processes and fabrication schemes. This first of two articles examines specific challenges in advanced interconnect scaling, the solutions for extending interconnect to 14nm, and the outlook beyond.

The primary driver for the continued success of the semiconductor industry is node over node reduction in transistor density, which is required to double transistor density every two years. As a direct result of technology innovations required to double transistor density, scaling over the last three nodes has been accompanied by significant changes in materials and architecture used to fabricate transistors. The 45nm node saw the adoption of high-k metal gate, while 22nm ushered in the era of 3D transistors.

Transistor scaling has a cascading effect on the interconnect. Increasing transistor density is accommodated by decreasing the interconnect pitch, increasing total interconnect length, and adding levels of interconnect. Greater interconnect lengths, especially for intermediate and global interconnects, result in higher voltage drop, and repeaters become necessary, occupying valuable chip area. Adding interconnect levels increases manufacturing complexity and slows down the yield ramp.

However, the biggest challenges result directly from pitch reduction. These include poor pattern integrity, the increase in resistive-capacitive (RC) delay, a narrowing of the process window for gap-fill, and degradation in reliability. Here we examine each of these challenges and solutions for next-generation interconnect fabrication.

PATTERN INTEGRITY

Industry has moved to hard-mask based process flows for lower damage and compatibility with double patterning at and below the 2x nm node. The most common hard mask (HM) used is TiN, primarily because of its high selectivity during low-κ etch. A key concern with TiN-based HMs, though, is the high compressive stress of the TiN combined with the low mechanical strength of ultra-low-κ (ULK) materials that causes buckling of the interconnect lines.

Lowering TiN HM stress is the logical solution. However, the biggest challenges result directly from pitch reduction. These include poor pattern integrity, the increase in resistive-capacitive (RC) delay, a narrowing of the process window for gap-fill, and degradation in reliability. Here we examine each of these challenges and solutions for next-generation interconnect fabrication.
conventional TiN deposition using bias power (DC) alone is accompanied by density reduction from ~4.8g/cc to 4.0g/cc, which in turn is strongly correlated to selectivity during dielectric etch. Figure 4c shows severe trench erosion in the lower-density TiN, leading to pattern fidelity issues. Thus, a lower-stress TiN film with minimal density degradation is needed to meet line buckling and selectivity requirements.

Ion bombardment is the major cause of compressive film stress. For DC-only TiN, high process pressure reduces film bombardment, but also creates more oblique deposition and reduces film density. Augmenting DC with very high frequency (VHF) source power increases ionization and reduces sheath potential. The result is collimated low-energy deposition with reduced bombardment. Figure 2 illustrates the high-density TiN (24.8g/cc) achievable with radio frequency (RF) based technology over a wide stress range. This high-density, low-stress technology avoids buckling and enables high selectivity as evidenced by good profile results with no line buckling for aspect ratios as high as 6:1 (Figure 3). The typical trench aspect ratio for etch is <4:1, these higher aspect ratios demonstrate the scalability of RF TiN HM films as the line-buckling process window is smaller as aspect ratio increases.

**RC DELAY**

Figure 4 shows RC delay as a function of technology node, increasing exponentially from the 14/10nm node as line resistance increases. Models based on bulk resistivity of copper predict that resistance will increase as the ratio of line length to area. But as copper widths drop below 40nm, a sharp rise in resistivity is expected due to mean-free-path-driven scattering at sidewalls, surfaces, and grain boundaries. At the 14nm node, the width of the metal lines will be as small as 22nm. Exacerbating the trend is the lag in developing dielectrics with suitably low dielectric constants, and hence, capacitance. The metal lines will be as small as 22nm. Exacerbating the trend is the lag in developing dielectrics with suitably low dielectric constants, and hence, capacitance.

**GAP FILL**

State-of-the-art gap-fill processes require liners (e.g., cobalt (Co) or ruthenium (Ru)) for the best gap-fill window. These new materials create their own integration issues (e.g., time dependent dielectric breakdown (TDDB) challenges with Ru). In the future, maximizing copper volume for best line resistance and to minimize additional interfaces for optimized via resistance will limit total barrier/liner thickness to 32nm. Thus, the maximum liner thickness allowed will be 28nm.

Two approaches can widen gap-fill process windows to reduce aspect ratios for electroplating. The first enhances seed deposition through thermal reflow, which capitalizes on capillary action to create bottom-up fill that reduces the plating aspect ratio and has been proven to fill 2x nm aspect ratio (Figure 5a). This process is also extendible to 16nm CDs (10nm node) with 1nm of optimized CVD Ru as the liner. The second approach is to develop a process that has twice the bottom coverage of state-of-the-art seed technology, followed by seed re-sputter to the sidewall, which will effect a substantial reduction in aspect ratios for plating. Combining the seed layer with advanced plating chemistry would enable liner-free gap fill down to 16nm CD trench and 2x nm CD dual damascene with today’s low-k dielectrics (κ=2.55).

**RESISTANCE REDUCTION**

Both of the above approaches help reduce line resistance through minimization or elimination of liners. All things being equal, up to 35% lower resistance can be achieved by thinning lines from 30nm to 10nm, and a further 20% can be achieved by eliminating liners in 14nm node interconnect.

Further reduction in line resistance requires technologies that enable large-grain copper or reduce sidewall scatttering. The reflow approach above can be optimized to improve grain size (Figure 5b). Even at a 10nm smaller CD, the reflow grain sizes are at least double those of the conventional approach. We believe that the higher mobility of copper with this new reflow approach results in larger grains. Normalizing to the same CD (22nm) would result in a 20-25% lower resistivity as calculated from electrical test measurements and transmission electron microscope images.
Reducing the copper electron scattering at the sidewall requires more disruptive and high-risk approaches. Conceptually, resistance models predict that the best interface for minimal electron scattering is a dielectric-copper interface.14 To realize this concept, a new materials system will be required, such as self-forming barriers, where a deposited material reacts in a self-limiting manner with the dielectric to form a barrier. To maintain a dielectric-copper interface, new fill techniques are required that are not dependent on having a conducting substrate.

CAPACITANCE REDUCTION

Ideally, the expected exponential increase in copper line resistance can be balanced by continuous reduction in effective $\kappa$ of the interconnect. Predictions of integrating an insulating material with a dielectric constant of 1.5 by 2015 have succumbed to tradeoffs between dielectric constant, mechanical integrity, and low-$\kappa$ damage. It is known that low-$\kappa$ films with higher carbon content show less integration-related damage. However, this robustness comes at the expense of mechanical strength (packaging integrity). Repairing the damage allows the integration of low carbon, high mechanical strength, high-porosity low-$\kappa$ films. A novel process has been developed for treating low-$\kappa$ films to restore chemical integrity and enhance mechanical strength. The process treats not only the surface of the film but also the bulk, enabling the lowest integrated $\kappa$, while enhancing key structural properties like modulus and hardness. Integrated on a x12.2:1 film, the treatment can reduce RC by as much as 6% (Figure 6a). Other advantages include TDDB improvement (Figure 6b) and strengthening of the dielectric to better withstand downstream processes.

RELIABILITY

TDDB and electromigration are key reliability concerns affecting the 14nm node.

Figure 6. (a) Treated low-$\kappa$ dielectric shows 4-6% lower RC. (b) Treated low-$\kappa$ dielectric shows improved TDDB. Treatment followed post-planarization CuO$_x$ reduction.

TDDB

Besides the interface and low-$\kappa$ damage-related issues that affect TDDB, double patterning has further narrowed the reliability process window. Double patterning is required to achieve sub-80nm pitch interconnects in the absence of extreme ultraviolet lithography. Techniques such as litho-etch-litho-etch (LLEE) require precise overlay (OL) for successful double patterning as the resist-to-resist space is defined by the OL at first and second exposures. Poor OL may lead to line-to-line and line-via shorting. While techniques such as self-aligned double patterning can be used for the metal to avoid intra-line issues, via patterning still requires LLEE. Line-via shorting and TDDB will become key limiters. Self-aligned via (SAV) schemes are therefore required. The success of a such a scheme depends on the integrity of the HM, and the ability of the dielectric-etch process to maintain HM corner integrity during dual damascene fabrication. While HM integrity can be improved by using dense, low-stress RF physical vapor deposition TiN, the etch process is equally critical. A highly selective etch process with minimal skew is desirable, and this has been achieved with 2–4 times higher conductance combined with symmetric RF, gas delivery, and pumping. Recent studies show that SAV can be achieved with TiN as thin as 150Å, which widens the process window for subsequent steps, such as copper gap fill, by enabling a lower aspect ratio.

While SAV resolves line-via OL issues at the same level, level-to-level alignment scaling remains a serious concern for TDDB. Based on the allowable OL tolerances, a simple calculation suggests that fields as high as 2MV/cm may be encountered at the 10nm node. This is typically where line-to-line leakage for the current set of dielectrics suddenly rises. Significant improvements in interfaces and bulk low-$\kappa$ materials must be achieved to pass stringent TDDB requirements for the 10nm node.

Electromigration

Electromigration (EM) lifetimes are expected to decrease node over node due to the increase in the maximum allowable temperature of the semiconductor junction and scaling-induced reduction in the critical void volume that can cause resistance increase and EM failures. A number of techniques offer EM improvement, ranging from a doped seed, to a self-aligned CuSi$_x$Ni film, to electroless selective metal caps. All have trade-offs, and, hence, scaling issues, including increased resistance (e.g., doped seed and CuSi$_x$Ni) or the inability to scale thickness due to pre- and post-clean that leads to unacceptable capacitance increases as metal height scales down. The selective metal cap approach (e.g., CoWP) shows the best EM performance by eliminating the fast-diffusion problem at the copper-dielectric barrier interface responsible for EM degradation. This approach can be implemented using a selective CVD metal cap layer that can scale as it does not require aggressive pre- and post-cleaning, does not increase resistivity by diffusing into the copper, and is thin ($\approx$5nm) to minimize capacitance impact. Studies show that EM can improve as much as 80X with this approach. This method also demonstrates scalability on a x$\approx$2.3 porous low-$\kappa$ dielectric with the repair treatment noted above, meeting TDDB lifetime requirements.

CONCLUSION

While many design and manufacturing challenges exist, there is a path to a manufacturable 14nm copper interconnect. For further extendibility, however, the industry must focus on disruptive solutions in materials and device designs.

ACKNOWLEDGEMENTS

The authors thank the process engineering, test and characterization, operations, and Defect and Thin Film Characterization Laboratory teams of the SSG-CTO group for enabling the above workflows and analysis. Thanks also to the Metal Deposition Products, Etch, and Semitool business units, and the blanket film deposition group in the Dielectric Systems and Modules business unit for process development.

REFERENCES


AUTHORS

Mehul Naik is a distinguished member of technical staff in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He holds his Ph.D. in chemical engineering from Rensselaer Polytechnic Institute.

Nikos Bekiaris is a senior member of technical staff in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He earned his Ph.D. in chemical engineering from the California Institute of Technology.

Zhenjiang “David” Cui is a senior member of technical staff in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He received his Ph.D. in chemical engineering from the University of Illinois, Chicago.

Alexandros Demos is a director in the Dielectric Systems and Modules business unit of the Silicon Systems Group at Applied Materials. He earned his Ph.D. in physical chemistry from Simon Fraser University, British Columbia, Canada.

Bencherki Mebarki is a program manager in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He received his Ph.D. in plasma physics from Paul Sabatier University, France.

Alan Ritchie is a senior member of technical staff in the Metal Deposition Products business unit of the Silicon Systems Group at Applied Materials. He is a candidate for an M.S. in system engineering from San Jose State University.

Jennifer Tseng is a senior program manager in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. She holds her M.S. in bio-meteorology from Utah State University.

ARTICLE CONTACT

Mehul_Naik@amat.com
Selective Metal Capping With CVD Co for Improved Interconnect Reliability

KEYWORDS
Cobalt Copper CVD Electromigration Interconnects Metal Capping Reliability Selective Metal Capping

Shrinking feature dimensions in copper (Cu) back-end metallization and low-κ dielectric integration present some unique challenges in maintaining the performance and reliability of interconnects under aggressive working conditions. At minimum pitch, these scaled interconnects withstand very high electrical current densities that can lead to electromigration (EM) failures at the Cu-dielectric interface. Selective CVD cobalt (Co) metal capping is a promising new technology for sub-3x nm nodes that prevents voiding is presumed to be exacerbated by normal stresses at the Cu-dielectric interface. In particular, EM-based interconnect failure has been shown to depend on interface bonding. Careful optimization of the surface preparation step prior to dielectric deposition has enabled good adhesion for the Cu-dielectric interface to the 2x nm node, thereby improving the desired interconnect reliability. However, as interconnect pitch scales further, adhesion marginalization at this interface can lead to voiding in Cu lines. Selective metal capping both reduces Cu mobility and promotes adhesion at the Cu-dielectric interface. Reduced Cu atom mobility enhances the EM performance of dense interconnects. Films of <20Å CVD Co significantly improve adhesion (Figure 2).

Using a selective metal cap poses some unique integration challenges. This layer should promote adhesion without degrading line resistance. It must also selectively bind to the Cu surface and not the low-κ surface as metal entrapment in low-κ will lead to time dependent breakdown (TDDB) failures. DRIVING SELECTIVITY WITH CVD Co METAL CAPPING

The use of an electroless CoWP (cobalt tungsten phosphide) metal cap has been reported to be helpful in improving degraded EM reliability. However, adoption of such a cap in general exacerbates the already problematic low-κ dielectric TDDB reliability due to metal entrapment in ultra-low-κ (ULK) dielectrics.[1,2] The selective CVD Co process is run on planarized Cu interconnect wafers. Prior to CVD Co deposition, the substrate is exposed to a dry cleaning step that removes CuO (copper oxide) and removes post-planarization interconnect wires that are isolated by dielectric layers through a series of deposition, fill, planarization, and passivation steps. Shrink pitch and higher packing densities subject interconnects below the 3x nm node to significant increases in current densities, making them more susceptible to EM failures.

A typical back-end fabrication sequence involves planarizing the Cu interconnects followed by depositing a dielectric etch stop layer. The Cu-dielectric interface plays a critical role in EM performance (Figure 1). Mechanisms based on direct transport of atoms at the interface with a fast diffusion path for Cu atoms have been suggested as causes of EM failures.[3] Stress-induced voiding is presumed to be exacerbated by normal stresses at the Cu-dielectric interface. In particular, EM-based interconnect failure has been shown to depend on interface bonding.

Careful optimization of the surface preparation step prior to dielectric deposition has enabled good adhesion for the Cu-dielectric interface to the 2x nm node, thereby improving the desired interconnect reliability.[3] However, as interconnect pitch scales further, adhesion marginalization at this interface can lead to voiding in Cu lines. Selective metal capping both reduces Cu mobility and promotes adhesion at the Cu-dielectric interface. Reduced Cu atom mobility enhances the EM performance of dense interconnects. Films of <20Å CVD Co significantly improve adhesion (Figure 2).

Using a selective metal cap poses some unique integration challenges. This layer should promote adhesion without degrading line resistance. It must also selectively bind to the Cu surface and not the low-κ surface as metal entrapment in low-κ will lead to time dependent breakdown (TDDB) failures.

DRIVING SELECTIVITY WITH CVD Co METAL CAPPING

The use of an electroless CoWP (cobalt tungsten phosphide) metal cap has been reported to be helpful in improving degraded EM reliability. However, adoption of such a cap in general exacerbates the already problematic low-κ dielectric TDDB reliability due to metal entrapment in ultra-low-κ (ULK) dielectrics.[1,2] The selective CVD Co process is run on planarized Cu interconnect wafers. Prior to CVD Co deposition, the substrate is exposed to a dry cleaning step that removes CuO (copper oxide) and removes post-planarization interconnect wires that are isolated by dielectric layers through a series of deposition, fill, planarization, and passivation steps. Shrink pitch and higher packing densities subject interconnects below the 3x nm node to significant increases in current densities, making them more susceptible to EM failures.

A typical back-end fabrication sequence involves planarizing the Cu interconnects followed by depositing a dielectric etch stop layer. The Cu-dielectric interface plays a critical role in EM performance (Figure 1). Mechanisms based on direct transport of atoms at the interface with a fast diffusion path for Cu atoms have been suggested as causes of EM failures.[3] Stress-induced voiding is presumed to be exacerbated by normal stresses at the Cu-dielectric interface. In particular, EM-based interconnect failure has been shown to depend on interface bonding.

Careful optimization of the surface preparation step prior to dielectric deposition has enabled good adhesion for the Cu-dielectric interface to the 2x nm node, thereby improving the desired interconnect reliability.[3] However, as interconnect pitch scales further, adhesion marginalization at this interface can lead to voiding in Cu lines. Selective metal capping both reduces Cu mobility and promotes adhesion at the Cu-dielectric interface. Reduced Cu atom mobility enhances the EM performance of dense interconnects. Films of <20Å CVD Co significantly improve adhesion (Figure 2).

Using a selective metal cap poses some unique integration challenges. This layer should promote adhesion without degrading line resistance. It must also selectively bind to the Cu surface and not the low-κ surface as metal entrapment in low-κ will lead to time dependent breakdown (TDDB) failures.

DRIVING SELECTIVITY WITH CVD Co METAL CAPPING

The use of an electroless CoWP (cobalt tungsten phosphide) metal cap has been reported to be helpful in improving degraded EM reliability. However, adoption of such a cap in general exacerbates the already problematic low-κ dielectric TDDB reliability due to metal entrapment in ultra-low-κ (ULK) dielectrics.[1,2] The selective CVD Co process is run on planarized Cu interconnect wafers. Prior to CVD Co deposition, the substrate is exposed to a dry cleaning step that removes CuO (copper oxide) and removes post-planarization interconnect wires that are isolated by dielectric layers through a series of deposition, fill, planarization, and passivation steps. Shrink pitch and higher packing densities subject interconnects below the 3x nm node to significant increases in current densities, making them more susceptible to EM failures.

A typical back-end fabrication sequence involves planarizing the Cu interconnects followed by depositing a dielectric etch stop layer. The Cu-dielectric interface plays a critical role in EM performance (Figure 1). Mechanisms based on direct transport of atoms at the interface with a fast diffusion path for Cu atoms have been suggested as causes of EM failures.[3] Stress-induced voiding is presumed to be exacerbated by normal stresses at the Cu-dielectric interface. In particular, EM-based interconnect failure has been shown to depend on interface bonding.

Careful optimization of the surface preparation step prior to dielectric deposition has enabled good adhesion for the Cu-dielectric interface to the 2x nm node, thereby improving the desired interconnect reliability.[3] However, as interconnect pitch scales further, adhesion marginalization at this interface can lead to voiding in Cu lines. Selective metal capping both reduces Cu mobility and promotes adhesion at the Cu-dielectric interface. Reduced Cu atom mobility enhances the EM performance of dense interconnects. Films of <20Å CVD Co significantly improve adhesion (Figure 2).

Using a selective metal cap poses some unique integration challenges. This layer should promote adhesion without degrading line resistance. It must also selectively bind to the Cu surface and not the low-κ surface as metal entrapment in low-κ will lead to time dependent breakdown (TDDB) failures.
EM AND TDDB RESULTS
Relative to the baseline without CVD Co, EM performance improved by >10X to 100X when a <20Å CVD capping layer was used.37 A selective CVD Co capping layer was also evaluated using line resistance measurements. No measurable increase in line resistance was observed in test samples with and without CVD Co capping.37 Also, no degradation in TDDB was observed with selective metal capping integrated in the back-end metallization flow (Figure 5).

CONCLUSION
High current densities pose significant EM challenges for 2x nm node copper interconnects. Overcoming these requires selective CVD Co capping. A single-chamber, cyclical plasma-based process has been developed to selectively deposit a <20Å CVD Co capping layer to enhance EM performance without increasing line resistance or degrading TDDB.

ACKNOWLEDGMENTS
The authors appreciate contributions from the technology and engineering teams responsible for this work. The authors also recognize the contributions of the applied materials business unit and Maydan Technology Center.

REFERENCES

AUTHORS
Kavita Shah is a global product manager in the Metal Deposition Products business unit of the Silicon Systems Group. She holds a M. Eng. in chemical engineering from Cornell University.
David Sabens is a process engineer in the Metal Deposition Products business unit of the Silicon Systems Group. He has a Ph.D. in chemical engineering from Case Western Reserve University.
Joseph AuBuchon is a technology manager in the Metal Deposition Products business unit of the Silicon Systems Group. He received his Ph.D. in materials science and engineering from the University of California, San Diego.

ARTICLE CONTACT
Kavita.Shah@amat.com

PROCESS SYSTEM USED IN STUDY
Applied Endura® CVD

Electrical Characterization of Through-Silicon Vias for 3D Integration

In both the 3D and 2.5D architectures, the central element is the TSV itself, whose fabrication is subject to many unit-process and process-integration challenges.37 The primary steps in forming the via are deep silicon etch; chemical vapor deposition (CVD) of an insulting oxide liner; physical vapor deposition (PVD) of a metal barrier and seed layer; electrochemical deposition (ECD) of copper to fill the via hole; and chemical-mechanical planarization (CMP). Much of the difficulty arises from the high aspect ratio of the vias: they need to be deep to traverse the entire thickness of the silicon, and their diameters need to be small to avoid occupying valuable wafer space or affecting the performance of adjacent transistors. Advances in these unit processes and their integration have been described earlier.37

Several key TSV functional attributes can be evaluated electrically at the wafer level.37 In a 3D or 2.5D stack, it is important for the capacitance of the TSV to be low. Several of the TSV patterning processes may contribute to increased line resistance, leakage, and breakdown measurements give confidence that the necessary elements are in place for 3D integration to succeed. 3D integration with TSVs is an advance in microelectronic packaging that is poised to transform the semiconductor industry. Here we report on electrical characterization of TSVs that were fabricated in Applied Materials laboratories.

FABRICATION AND EXPERIMENTAL SET-UP
TSVs were formed in p-doped silicon wafers with a doping concentration of 1015/cm3. Before TSV patterning, 2kÅ oxide was deposited to simulate the pre-metal dielectric layer on a device wafer. Via 5um in diameter and 50um deep were created using the Bosch etch method in a unique inductively-coupled plasma etch chamber. The oxide liner was deposited using a proprietary ozone/TEOS-based CVD process with a highly conformal coverage. It helps to provide a smooth TSV sidewall. Liner oxide thickness as deposited was 250nm on the field and 220nm on the lower sidewall. The PVD barrier

KEYWORDS
3D Integration Breakdown Voltage Capacitance Leakage Current Through-Silicon Via TSV
has a field thickness of 24Å and an approximate thickness of 50Å at the thinnest point. The seed layer, PVD Cu, is 68Å thick in the field and no less than 50Å thick inside the via. The vias were filled by copper ECD in a chamber with special agitation and field-shaping features, using a programmed controlled-current-waveform. After plating, the wafer was annealed at 400°C for 30 minutes in forming gas. CMP was performed by a three-platen sequence with endpoint control. The resulting total thickness of oxide on the field was 320nm (a combination of the blanket HDP oxide and the TSV liner oxide that remained after CMP). Figure 1 shows focused ion beam (FIB) images of a sample via from the study.

A comb pattern was made by PVD and etch of aluminum traces 19µm wide above the copper TSVs, leading to 210×180µm probe pads. Figure 2 shows schematics of the aluminum traces and the TSVs connected by them, as well as an optical micrograph of a portion of the structure at probe test. Three different comb-pair structures were used, each pair consisting of one comb with TSVs descending from it and a corresponding “blank” comb without TSVs. Each of the three pairs had a different pitch or center-to-center spacing between the TSVs: 10µm, 25µm, and 50µm. During measurement, the terminal pad was contacted by one probe tip, and the grounded metal chuck holding the wafer provided the return path to the LCR (inductance-capacitance-resistance) meter. Capacitance measurements were made on the TSV-populated comb and its corresponding blank comb so that the net capacitance of the vias themselves could be estimated by the difference between the two values.

**ELECTRICAL CHARACTERIZATION**

TSV capacitance was measured by sweeping the bias voltage from -10V to 10V with a small-signal AC component at 1MHz. Figure 3b shows the wafer map with a total of 76 dies. Five dies representing the center, mid-radius, and edge locations were selected for measurement. Figure 3b shows the region of the mask design where the three comb-pairs were located. As noted above, TSVs were located in three regions where they were 10µm, 25µm, and 50µm apart. The comb structures were designed such that for each of these three pitches, the total number of comb fingers and the comb finger length were all different. Corresponding TSVs totaled 522, 256, and 240 for the 10µm, 25µm, and 50µm pitches, respectively.

The quasi-static capacitance-voltage (QSCV) technique is frequency-independent and has high measurement sensitivity. We used this method to extract accumulation capacitance ($C_{ox}$). Figure 4 shows the QSCV measurement results for TSV-populated combs and blank combs for 50µm-pitch TSVs in the accumulation region. By subtracting the two and dividing by the number of TSVs, we obtained a $C_{ox}$ of 157fF per TSV.

It has been suggested that operating within the minimum depletion capacitance regime can effectively reduce the total TSV capacitance, which is desirable for 3D integrated circuits. Considering that the TSV bottom region does not contribute much to the total capacitance, a TSV can be modeled as a cylindrical metal-oxide-silicon structure. Thus, the minimum depletion capacitance is given by solving:

$$C_{dep\,min} = \frac{2\pi \varepsilon_0 \varepsilon_r \varepsilon_{Si}}{2kT} \frac{q^2}{4\pi^2 \varepsilon_0^2} \frac{N_a}{n_i} \left( \frac{2kT}{q} \right)^{\frac{1}{2}}$$

where $\varepsilon_0$ is permittivity of vacuum, $\varepsilon_{Si}$ is TSV depth, and $R_m$ is outer radius of the oxide liner. The maximum depletion radius, $R_m$, is given by solving:

$$R_m = \frac{1}{2} \left[ \frac{1}{2} \sqrt{1 + \frac{2kT}{q} \frac{N_a}{n_i}} \right]$$

Using the above $C_{ox}$ and $C_{dep\,min}$ values, the minimum per TSV capacitance is calculated to be 67fF.

TSV capacitance uniformity at different pitches is illustrated in Figure 5. Nine dies across the wafer were checked for each of the three pitches. It is clear that in all cases the capacitance shows very little variation.
Table 1 summarizes the C-V data in the minimum capacitance region. For all three pitches, the mean capacitance per TSV is identical and is in good agreement with the calculated TSV capacitance from Equation 3.

Table 1. Total and per via capacitances for different TSV pitches.

<table>
<thead>
<tr>
<th>TSV Pitch (µm)</th>
<th>Total Capacitance of Comb (pF)</th>
<th>Capacitance per TSV (pF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>522</td>
<td>52.2</td>
</tr>
<tr>
<td>25</td>
<td>256</td>
<td>25.6</td>
</tr>
<tr>
<td>50</td>
<td>240</td>
<td>24.0</td>
</tr>
</tbody>
</table>

Figure 3a). Three scenarios are apparent: the 10µm-pitch sub-die undergoes a “soft” breakdown, depicted by a jump in current at approximately 42V followed by gradually increasing current as the bias ramps up. For the 25µm-pitch comb, leakage current increases at a low rate until “hard” breakdown occurs at 79V. The 50µm-pitch comb leakage curve is initially almost identical to the 25µm-pitch comb, but breakdown does not occur. For all three cases, leakage current at 20V bias falls in a narrow range (106-143pA) for the three TSV clusters. The corresponding average leakage current per TSV is less than 0.5pA, which matches published leakage-current data for TSVs of similar dimension. The leakage current does not depend appreciably on TSV pitch.

Figure 6 illustrates the statistics of Vbd for all three TSV pitches in nine dies. The Vbd is seen to be dependent on TSV pitch: Vbd is lower when TSVs are closer. In the case of the highest pitch, all but three of the combs survived up to 100V without breaking down. A linear fit for each case gives the Weibull slope, and the shape parameter, which is greater than 1 in each case, as would be expected. While there are not enough data points to draw a conclusion for 50µm-pitch combs (not many dies breaking down), curve fits for the 10µm-pitch and 25µm-pitch combs indicate similar failure rates.

CONCLUSION

TSV structures fabricated by a baseline integrated process sequence were electrically characterized. Capacitance, leakage, and breakdown results generally conform to expectations and to healthy TSV characteristics. Further work remains to be done, however. This includes: 1) extending electrical testing to ongoing improvements in TSV unit processes and integration, especially improvements leading to greater wafer-level uniformity in electrical characteristics; 2) further understanding the observed breakdown behaviors and dependence on via-to-via spacing; 3) extending characterization tests to include time dependent dielectric breakdown at elevated temperature and other reliability tests; and 4) characterizing electrical properties as TSV size scales down. Such learning—with continuous feedback to unit-process, equipment, and integration development efforts—will facilitate the successful implementation of TSV-based 3D integration.

ACKNOWLEDGMENTS


REFERENCES


AUTHORS

Minrui Yu is a process engineer in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He holds his Ph.D. in chemical engineering from the University of Wisconsin-Madison.

John Dukovic is a distinguished member of technical staff in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He earned his Ph.D. in chemical engineering from the University of California, Berkeley.

Bharat Brusahan is a process engineer in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He received his M.S. in electrical engineering from IIT Bombay, India.

Niranjan Kumar is a product marketing manager in the Silicon Systems Group at Applied Materials. He holds his bachelor’s degree in technology from IIT Kanpur, India, and a degree from Stanford University, both in electrical engineering.

Sachin Ramaswami is managing director, TSV and advanced packaging in the Chief Technologist Office of the Silicon Systems Group at Applied Materials. He earned his M.S. in chemical engineering from Syracuse University and MBA from San Jose State University.

ARTICLE CONTACT

Minrui_Yu@amat.com

PROCESS SYSTEMS USED IN STUDY

Applied Centura® Silica™ Ethox
Applied Producer® Invia™ CVD
Applied Endura® CubS PVD
Applied Raider®-S LCD
Applied Reflexion® LK CMP
Developing 3D Architectures for Future Memory

As 2D semiconductor device scaling approaches its limit, the third dimension offers the means to extend scaling. Will 3D NAND be the first high-volume 3D technology or will other promising alternatives establish a foothold? 3D technology holds the promise of lowering costs by enabling denser device packing, the most fundamental requirement for memory. Yet uncertainties arise in achieving high speed, low power, and superior endurance—characteristics of universal memory—as devices transition from conventional 2D to 3D. 3D X-bar devices, when made compatible with established CMOS fabrication processes, could be viable in parts of the memory market at 1x nm and beyond.

Fulfilling Moore’s Law, the semiconductor industry has gone through 40+ years of rapid scaling and is now approaching the 10nm range, a dimension in which atoms are counted by the hundreds. Quantum phenomena, such as energy band discrete and confined charge transportation, become more dominant at these dimensions. Consequently, well-established CMOS technology—especially memory, for which array architecture is typical—will face an exponential increase in challenges. In the early 2000s, DRAM led the scaling trend; by the end of the decade, NAND Flash had surpassed DRAM by scaling beyond the 20nm node. However, given its fundamental limitations, the 2D NAND device is expected to reach its limit at the 10-12nm node in the next 3 to 5 years. Besides dimension scaling, higher speed and lower power are vital for memory devices in the new mobility and connectivity era.

In the research community, 3D architectures have emerged a decade ago when both the stacked 2D NAND and stacked 2D dieo + SiO fuse-one-time programmable (OTP) memory were proposed. However, 3D architecture has not been successfully commercialized for high-density memory, with the exception of stacked OTP memory. Complex process challenges, such as creating high-mobility silicon channels for vertical NAND, and etching and depositing on extremely high aspect ratio (HAR) features (>50:1), must be overcome for 3D memory architecture to be viable. Besides 2D stacking, emerging phase-change random-access memory (PCRAM), resistive RAM (ReRAM), and spin-transfer torque RAM (STT-RAM) technologies also hold promise for high speed and low power.

**EMERGING MEMORY TECHNOLOGIES**

**PCRAM**

First proposed by Stanford R. Ovshinsky in the 1960s, PCRAM uses the physics of material phase change when heat is applied to chalcogenide glass by an electric current (Figure 1). The structure sandwiches the phase-change material between two metals. The phase-change material transforms from a crystalline (low resistance, binary 1) phase to an amorphous (high resistance, binary 0) phase when a fast electrical pulse at higher voltage is applied to generate needed heat for melting and quenching it. The reverse change occurs when a low-voltage electrical pulse is applied for a relatively long time, allowing for lattice ordering. Resistance is read using another electrical pulse at lower voltage for the 0 and 1 states. All the electrical pulses are applied with the same polarity, which is called unipolar operation. PCRAM has demonstrated high speed, good endurance, and scalability. However, its high program current density, thermal disturbance for small geometry, thermal instability during integration, and material reliability during device operation are all hurdles to overcome if it is to be competitive with existing technologies. Proposed solutions include improving heater design for better heat transfer efficiency to reduce high current density, enhancing thermal isolation using materials with low thermal conductivity, and refining deposition and integration processes to achieve better cell reliability.

**ReRAM**

ReRAM originated from the negative-resistance property discovered in several material systems in the 1960s. During the last decade, researchers applied this property to high-density memory by integrating a metal-metal oxide-metal sandwich structure into the CMOS process. Recent experiments confirmed earlier suggestions that the ReRAM mechanism is a filamentary phenomenon (Figure 1), in which a conductive path is formed by a controlled dielectric electrical breakdown process called forming. Oxygen atoms move away from metal atoms, creating a path of conductive-phase material. As electric current passes through the conductive path, the heat generated causes the oxygen to move back into the conductive path, restoring the material to its non-conductive phase. This exchange process is further enhanced by applying the electric field in two different polarities to move the negatively-charged oxygen toward one interface so that only a small amount of oxygen is needed to form a high-resistance thin insulating layer within pico- or nanoseconds. Reapplying the electric field restores the conductive path. These repeated operations constitute a memory device.

ReRAM has demonstrated high speed, reasonable endurance, lower power than PCRAM, and scalability to 10nm. However, passage through the nm-scale filament of the high-density current (typically 10-20µÅ) induces high local temperature and poses the risk of further interaction between materials. Capacitive discharge from fast filament activation accelerates dielectric wear, degrading reliability and increasing variability related to filament initialization (forming) and size over time (both are highly random). These issues reduce cell-to-cell cycle-to-cycle stability. Remedies include optimizing a forming-less flipstack (forming voltage decreases linearly with decreasing oxide thickness) and using high work function metal to suppress tunneling current and increase the “On” resistance, thus lowering the reset voltage and programming current.

**STT-RAM**

Magnetic memory was proposed about 20 years ago after the giant-magneto-resistance phenomenon was discovered, revealing that the resistance of stacked magnetic materials changes when the magnetic dipole polarization is aligned parallel or anti-parallel among the different materials. During the past decade, several RAM technologies employing magnetic materials have been proposed. One uses a magnetic tunneling junction (MTJ) for higher resistance change between parallel and anti-parallel states (Figure 1). The MTJ consists of top and bottom ferromagnetic materials separated by a thin insulating layer. The resistance delta increases in the presence of the middle dielectric layer.
which can tunnel the polarized current. An STT-RAM device is formed when one of the ferromagnetic electrodes is pre-magnified with a particular polarization and the polarization of the other can be changed through electron spin induced by an electric current. STT-RAM demonstrates scalability beyond 20nm, excellent endurance (exceeding 10^12 cycles), lower power, and high speed. However, fabrication poses challenges ranging from material complexity of the structure to process integration. The semiconductor industry is not familiar with magnetic film processing and the device consists of many ultra-thin layers (on the order of nm) of materials with widely varying characteristics. Deposition and etch processes will have to achieve virtually atomic-scale control to ensure extreme uniformity and negligible surface roughness essential for a high tunnel magneto-resistance ratio. Ultra-clean processes will be needed to avoid re-deposition of by-products or residue that could lead to shorting.

EMERGING 3D ARCHITECTURE

2D Stacking

The first 3D approach stacked 2D devices by connecting the layers at the memory array level while using the same circuit (Figure 2A). Through-silicon via (TSV) technology connects only the circuit input/output by vertical metal vias (Figure 2B). The former method was successfully demonstrated with NAND, using epitaxial silicon to stack the NAND string. This approach increases bit density by per unit silicon area by a factor of n (the number of layers), but it depends on a reliable silicon on insulator process, which is expensive. Consequently, this approach has not been commercialized.

3D X-Bar Architecture

The second approach proposed was to stack the functional element (cell) with a local control device (selector) to prevent sneak path current and globally connect these layers of cells to the circuit on the underlying silicon. Using this approach, a pure silicon PN junction has been demonstrated for PCRAM, and PIN diodes using polysilicon have also been explored (Figure 3).[8,9] The latter enables 2D PCRAM + diode cell stacking with anti-fuse can be modified to include a metal-insulator-metal anti-fuse which is expensive. Consequently, this approach has not been commercialized.

Both PN and PIN diodes are compatible with unipolar switching materials, such as PCRAM materials. Silicon PN diode offers limited stackability as high quality crystalline silicon is needed for low reverse current and high breakdown voltage. PIN diodes formed using low-pressure chemical vapor deposition offer the requisite stacking, but heat penetration accompanying traditional high-temperature rapid thermal anneal dopant activation makes this method unsuitable for PCRAM. By comparison, laser annealing, which applies high temperature for a very short time, avoids heat transfer to the phase-change material and is a promising alternative to investigate further.

Attempts have also been made to create the PN junction diode from metal oxides, which can be stacking and processed at low temperature. However, current density was found to be much lower than 1MA/cm^2, thus insufficient for PCRAM (10-50MA/cm^2) and current ReRAM cells (1-10MA/cm^2). Although PN and PIN diodes have been long studied, no reports exist of scaling them below 20nm. As the sensitivity to surface recombination states increases, the scaling of the diode becomes difficult.

Bipolar operation in which the directions of the set (from low to high resistivity) and reset (from high to low resistivity) voltages are opposite exhibits more robust ReRAM performance than unipolar operation. STT-RAM also needs bipolar operation. Therefore, ReRAM and STT-RAM are not compatible with PN or PIN diodes with one exception. A PIN diode can be used if one of the operations (either set or reset) requires very low current density (<0.1MA/cm^2) for a brief period at its reverse polarity near the breakdown. Recent study has shown that a N-P-N diode with its strong non-linearity can be effective for bipolar operation. Simulation of N-P-N has shown that the required 1MA/cm^2 current density can be achieved at 3V (programming) and 10A/cm^2 at 15V (reading and unselecting), a factor of 10^3 high.

3D Vertical Architecture

In recent years, several approaches to scaling that do not rely on increasingly costly high-resolution lithography have been proposed (Figure 4). Each basically rotates the NAND string from planar to vertical to increase bit density. As the number of cells in the string grows, so does bit density. The unit string device layout is typically 6T (2F*3F) for 3D architecture, in which the 2F direction is the vertical channel and the 3F is the lateral word line (WL) and isolation. Table 1 compares the effective bit areas of 2D NAND multi-level cell (MLC) and 3D NAND single-level cell (SLC) at 50 and 40nm nodes with 24-, 32-, 48-, and 64-cell vertical strings.
Vertical architecture was proposed for 3D ReRAM, whose cell allows higher bit density by sharing the WL to yield a 4F² vertical cell string; at the same time, two physical bits are made (Figure 5). The effective cell size of this design can be related to 2D NAND as follows:

$$4F^2/2 = 4F^2/n = n = (F/f)^2$$

Therefore, the bit area of a 16-cell string of 3D ReRAM at the 40nm node is the same as that of a 10nm 2D MLC NAND. ReRAM operation voltage is also much lower, enabling greater savings in circuit area.

However, the necessary device performance is not yet available for this 3D ReRAM design. The first requirement is that the device fabricated inside the 40nm hole must be both switching and self-rectifying. Selecting and unselecting occur through a biasing scheme that minimizes sneak path current. For example, the selecting bit or bits will be biased at high voltage (both positive and negative) for programming and the unselecting occur through a biasing scheme that must be both switching and self-rectifying. Selecting is that the device fabricated inside the 40nm hole is available for this 3D ReRAM design. The first requirement is that the device fabricated inside the 40nm hole is available for this 3D ReRAM design. The second requirement is low-current operation. The diameter of the vertical metal wiring is in the range of 20-30nm after the switching material has been deposited. Wiring resistance will be on the order of 100Kohm (considering the lateral wiring as well). When operating current is high at 10µA, an additional 7V will be needed for far access bits. When programming current is 1µA, the voltage difference will be only 0.7V, which most circuits can handle. The third requirement is a higher On/Off switching ratio. This not only helps the sensing circuit by providing a large window, but also increases the array block size through greater rectifying if a certain percentage of the bits are in the “Off” state. CHALLENGES IN NEW MEMORY ARCHITECTURE

**Bit Cost**

As NAND has scaled, the manufacturing cost of the bit has decreased. Manufacturing experience has also improved efficiency (high device yield and lower operation costs). New architectures will require new learning for high device yield and new equipment will be introduced as needed. This naturally leads to higher manufacturing cost. However, new architectures should ideally lower existing costs by 30%. Therefore, the first generation of the new architecture must be 30% cheaper to manufacture for the same die size or, if manufacturing cost remains constant, the die size must be 30% smaller.

2D stacking architecture yields higher density per unit of silicon. The number of process steps remains constant as all the layers are needed. Some savings might be gained in the periphery as the circuit is needed only once, but the fundamental requirement of the crystalline silicon layer limits cost reduction. If the silicon layer can be made by means such as laser melting anneal, additional arrays could be made as the circuit (which typically occupies 30-40% of the die) is not needed for the second layer and up. However, these gains are not large enough to drive commercialization.

3D X-bar architecture realizes 2D stacking without a crystalline silicon layer. Different memory layers are connected by metal interconnect. The circuit can be made underneath the memory layer to save die size, or to introduce a more sophisticated circuit (e.g., to introduce a controller) as a system solution. Overall, the cost of this architecture depends strongly on sub-20nm fine-resolution lithography. When a simple and cost-effective lithography approach is proven, such as imprint, the cost of the 3D X-bar architecture will be greatly reduced and commercialization will be viable.

3D vertical architecture has shifted stringent lithography requirements to the third dimension, which poses major HAR challenges. Although future lithography costs may decrease, HAR challenges will reduce process productivity, especially in etch. Additional processes, such as new material deposition with extreme step coverage and HAR structure post-etch cleaning, will initially increase manufacturing cost. Ultimately though, as process technology improves, the manufacturing cost of HAR structures is expected to diminish.

**Performance**

2D stacking does not change device performance but increases memory density. Currently, NAND device performance improvement is achieved largely by system-level solutions, such as programming algorithms and controllers. TSV has demonstrated strong performance improvement and has emerged as commercially viable. 3D X-bar architecture is not limited to slow NAND, but can be adapted to much faster operating PCM and ReRAM for dramatic performance improvement. In addition, the large silicon area underneath the memory array offers the opportunity to design good circuit architecture and a sophisticated controller to make performance even better. Future computer systems will be able to benefit from such a sub-system level device that integrates and optimizes power, speed, and density.

Vertical architecture and 3D are expected to overcome some of the challenges in HAR. The 3D NAND structure can overcome 2D NAND limitations, such as cell-to-cell interference, as the device shrinks. It also uses the charge trap approach rather than floating polysilicon gate approach to overcome the scarcity of electrons at smaller geometries. However, the structure itself has introduced the fundamental issue of channel mobility. Similar to 2D NAND, as channel size decreases, resistance grows and reading becomes more difficult, slowing performance. In 3D NAND, the polycrystalline silicon channel is more resistive and has less mobility. Thus, better circuitry, algorithms, and controllers will be needed to match 2D NAND performance.

For 3D ReRAM, the individual device has positive attributes, such as speed and voltage, but the high programming current must be reduced for low-power operation. Low current is also required to offset the increase in resistivity of the metal wiring. For best reliability, the current circuit scheme must be simplified to ensure predictable switching.

In computer systems, DRAM is used as buffer memory for the central processing unit and NAND is used for storage, such as solid state drives (SSD). The entire system operates at optimum speed when these components are properly balanced, stage by stage. Today, the biggest gap (often called the “memory wall”) is between DRAM and SSD—DRAM operates at tens of nanoseconds while NAND operates at hundreds of microseconds. This mismatch induces speed and power inefficiencies that could be greatly reduced by incorporating additional non-volatile memory operating at microsecond speeds. Employing 3D X-bar memory architecture has the potential to meet these requirements for 3D-40nm technology through non-volatile operation, sub-microsecond speed, and high density at 1-4GB.

---

### Table 1. Bit area comparison between MLC 2D NAND and SLC 3D NAND

<table>
<thead>
<tr>
<th>Node (nm)</th>
<th>40</th>
<th>400</th>
<th>200</th>
<th>150</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit Area (nm²)</td>
<td>24</td>
<td>24</td>
<td>32</td>
<td>48</td>
<td>64</td>
</tr>
<tr>
<td>Bit Area (nm²)</td>
<td>625</td>
<td>469</td>
<td>313</td>
<td>234</td>
<td>150</td>
</tr>
</tbody>
</table>

---

**Figure 5.** Compact design proposal for 3D ReRAM.
CONCLUSION
By using established devices, 2D stacking is effective in increasing bit density. But the fundamental requirement for thin crystalline silicon formation reduces its commercialization potential, minimizing the cost savings of this approach compared with others. 3D vertical architecture offers an effective path to reduce bit manufacturing cost without depending on extreme ultraviolet (EUV) lithography. Implementation will entail overcoming challenges that fundamental device requirements of such architecture pose for processes (novel material properties) and equipment (accurate control to the atomic level). 3D X-bar stacking offers an effective option for overcoming the memory wall at the 30-40nm node; it also enables bit scaling with finer lithography, such as EUV and imprint. Its device requirements will also pose new process and equipment challenges as features continue shrinking.

REFERENCES

AUTHOR
Er-Xuan Ping is a managing director in the Silicon Systems Group at Applied Materials. He holds his Ph.D. in electrical engineering from Iowa State University.

ARTICLE CONTACT
Er-Xuan.Ping@amat.com