Silicides and local interconnections for high-performance VLSI applications

As the minimum VLSI feature size continues to scale down to the 0.1–0.2-μm regime, the need for low-resistance local interconnections will become increasingly critical. Although reduction in the MOSFET channel length will remain the dominant factor in achieving higher circuit performance, existing local interconnection materials will impose greater than acceptable performance limitations. We review the state-of-the-art salicide and polycide processes, with emphasis on work at IBM, and discuss the limitations that pertain to future implementations in high-performance VLSI circuit applications. A brief review of various silicide-based and tungsten-based approaches for forming local interconnections is presented, along with a more detailed description of a tungsten-based "damascene" local interconnection approach.

Introduction
Over the last ten to fifteen years, silicides and local interconnections (LI) have become widely used in dense, high-performance VLSI circuits. Continuous chip size reduction through linear “shrinking” of ground rules and device scaling, both common industry practices, places an increased demand on the tolerable limits of interconnection resistance. Likewise, the shrinking feature size and circuit packing density heighten the sensitivity of performance to the resistivity of circuit conductors and the dielectric constants of associated insulators. This has led to an increase in the use of silicides with resistivity inherently superior to that of degenerately doped silicon in order to alleviate the resistance increase associated with reduced feature size [1–3].

While silicides have been introduced to preserve performance, local interconnections have been proposed primarily for density improvement. Local interconnection approaches involving one additional mask level such as HPSAC (high-performance silicided amorphous-silicon contact and interconnection technology) [4] and the use of a patterned TiN layer [5] were first proposed in 1985. Since then, several other modifications and promising approaches have been developed [6–11].

The state-of-the-art silicide and local interconnection processes used or considered for use in the VLSI industry are examined, with special consideration given to those that are considered to have the highest potential or are most dominant in the industry today. Some of the
Simulated gate delay per stage for an unloaded ring oscillator as a function of gate conductor sheet resistance. The simulation was performed for several channel lengths.

Silicide practices in VLSI applications

Silicides have become widely used in VLSI processing in order to avoid the increase in resistance otherwise associated with reduced polysilicon linewidths and film thicknesses. This is illustrated by examining the relationship between simulated gate delay per stage, as shown in Figure 1. The stage delay of a ring oscillator circuit is modeled as a function of gate conductor resistance for three different channel lengths. Although resistivity is an important factor, process integration issues also play a significant role in the selection of the most suitable interconnection materials. Thermal stability, oxidation properties, chemical reactivity, and diffusivity in silicon are extremely critical. We describe associated “polycide” and “salicide” processes and review the critical aspects.

As we discuss in more detail later, one of the more important factors for both salicide and polycide applications is the thermal stability of the silicide. The temperatures used in the circuit fabrication process generally require that the silicide exhibit good thermal stability. This inherently favors the refractory metal silicides with higher melting points and generally better thermal stability. It should be noted, however, that this thermal stability requirement may become less critical with the development of low-temperature (<600°C) planarization methods for achieving post-silicide isolation. The oxidation kinetics and chemical reactivity with various chemical species used in the fabrication of the integrated circuits can be either advantageous or problematic. A sufficiently low diffusivity in silicon is desired to prevent the metals from diffusing into the silicon p/n junction depletion regions and/or to the gate oxides.

Two types of silicide processes are currently used in the semiconductor industry: “polycide” and “salicide.” The polycide process is a method of patterning the silicide on the polysilicon gate electrode, and the salicide process is a method of self-aligning a silicide layer to all exposed silicon regions. A schematic of the salicide process sequence is shown in Figure 2. For the salicide process, the polycrystalline silicon gate is patterned and the sidewall spacers are formed prior to metal deposition. A metal layer is then deposited and reacted with the exposed silicon regions (gate and source/drain) to form the silicide. The silicide does not form over the isolation regions. A selective etch is used to “self-align” the silicide and remove the reaction products that are not desired. Often additional annealing is required to lower the resistivity of the silicide contacts.

In the polycide process, the silicide is deposited on the polysilicon gate at the stoichiometric composition of the desired silicide phase before it is patterned. When the gate is etched, the polycide “gate stack” is formed. Often an insulator material is included as the top layer of the stack.
for integration purposes. Annealing at a relatively high temperature is required to convert the amorphous silicide film to a lower-resistance polycrystalline phase. Once the sidewall spacers are formed, a salicide process is sometimes used to form silicide contacts to the source and drain regions. The sequence of steps used in a typical polycide process is shown in Figure 3. Table 1 summarizes some relevant materials parameters for silicides of most interest for VLSI applications.

**Silicides for salicide applications**

Although initially the range of possible choices for such applications may appear to be broad, once materials properties and integration issues such as thermal stability, self-alignment, and etch selectivity are taken into account, the field of potential candidates becomes significantly reduced. The two choices that have been given the most serious attention to date are CoSi₂ and TiSi₂. Since the resistivity is nearly the same for both, process integration considerations determine which is to be used. Titanium disilicide is the most commonly used silicide for salicide applications because of a combined set of characteristics including low resistivity, ability to be self-aligned, and relatively good thermal stability. Although titanium disilicide is heavily used for salicide applications, cobalt disilicide has recently been given more attention. This is because the low-resistivity phase of CoSi₂ is more easily obtainable than that of TiSi₂. For the purpose of this review we focus on TiSi₂ because of its widespread use. Cobalt disilicide is discussed where significant differences or appropriate contrasts to titanium disilicide are relevant.

- **Titanium disilicide polymorphism**

  When titanium and silicon are brought into contact and heated at temperatures above 500°C (in the presence of excess silicon) the higher-resistivity C₄₉-TiSi₂ phase forms before the low-resistivity C₅₄-TiSi₂ phase [12-15]. The C₄₉-TiSi₂ phase has an orthorhombic base-centered structure with 12 atoms per unit cell and a resistivity of 60–90 μΩ-cm [16]. The C₅₄-TiSi₂ phase has an orthorhombic face-centered structure having 24 atoms per unit cell and a significantly lower resistivity (12–20 μΩ-cm) than the C₄₉ phase [16]. Once the C₄₉-TiSi₂ phase has formed, further annealing is required to obtain the polymorphic phase transformation to the C₅₄-TiSi₂ phase. The activation energy required to convert a thin film of C₄₉-TiSi₂ to C₅₄-TiSi₂ is 4–8 eV [17-22], depending on processing conditions, substrate, and type of dopant. This high activation energy requires that high-temperature annealing be used to completely convert the C₄₉ phase into C₅₄-TiSi₂. This presents a significant practical problem in thin-film applications, in which too much thermal energy can cause morphological degradation.

- **Formation of C₄₉-TiSi₂ in salicide applications**

  During the formation of TiSi₂ in a salicide application, several different reactions involving the Ti layer occur simultaneously, as indicated in Figure 4 [23]. When silicon and titanium are reacted at the source, gate, and drain to form low-resistance TiSi₂ contacts, silicon can diffuse laterally over isolation regions and form thin filaments of TiSi₂. These thin filaments can electronically connect a gate electrode and source/drain regions to cause failure of a CMOS device. This defect mechanism is called “bridging.” The use of a nitrogen atmosphere during thermal annealing has been found essential in preventing this defect mechanism [24]. This is because nitrogen quickly diffuses into the grain boundaries and significantly reduces any long-range silicon diffusion in the titanium layer.

---

Table 1 Relevant silicide parameters.

<table>
<thead>
<tr>
<th>Type of silicide</th>
<th>Resistivity (μΩ-cm)</th>
<th>Thermal expansion (ppm/°C)</th>
<th>Si consumption (nm of Si)/(nm of silicide)</th>
<th>Melting point (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C₅₄-TiSi₂</td>
<td>12–25</td>
<td>12–13</td>
<td>0.904</td>
<td>1540</td>
</tr>
<tr>
<td>CoSi₂</td>
<td>18–25</td>
<td>~10</td>
<td>1.03</td>
<td>1326</td>
</tr>
<tr>
<td>MoSi₂</td>
<td>20–100</td>
<td>~8</td>
<td>0.99</td>
<td>1980</td>
</tr>
<tr>
<td>TaSi₂</td>
<td>10–50</td>
<td>8–11</td>
<td>0.92</td>
<td>~2200</td>
</tr>
<tr>
<td>Tetragonal WSi₂</td>
<td>50–120</td>
<td>6–8</td>
<td>0.98</td>
<td>2165</td>
</tr>
</tbody>
</table>

---

IBM J. RES. DEVELOP. VOL. 39 No. 4 JULY 1995
Simultaneous reactions that occur during the formation of TiSi, in a salicide application.

1. Ti + SiO₂ → TiO₂ + TiSi₂
2. Ti + Si₃N₄ → TiN₄ + TiSi₂
3. Ti + 2Si → C₄₉-TiSi₂
4. Ti + 1/2N₂ → TiN

The kinetics of the various reactions dictate that the temperature range for silicide formation be less than 700°C for furnace annealing [25-27]. The temperature can be somewhat higher with rapid thermal processing (RTP), but usually cannot exceed 775°C. This is because at higher temperatures TiSi₂, TiO₂, and TiN form over the isolation regions and cannot easily be etched in conventional salicide etching solutions. This can result in the formation of electrical shorts between adjacent circuit elements. In contrast to TiSi₂, CoSi₂ is considered to be inherently less susceptible to such “bridging,” since cobalt is the dominant diffuser in the formation reaction.

**C₄₉ to C₅₄ phase transformation**

As stated previously, the resistivity of the C₄₉-TiSi₂ phase and the C₅₄ phase differ by about a factor of 4. For this reason, the phase transformation is very important. One of the most useful techniques for examining phase transitions in the Ti/Si thin-film systems is through *in situ* resistance probing. *In situ* resistance measurements of a Ti film deposited on polysilicon during the formation sequence in a rapid thermal annealing (RTA) tool are depicted in **Figure 5**. As the temperature is increased, there is an initial increase in resistance associated with silicon diffusion into the metal, followed by a sharp reduction in resistance as C₄₉-TiSi₂ is formed. Continued annealing leads to a transformation of the C₄₉-TiSi₂ to low-resistance C₅₄-TiSi₂ (around 800°C). At temperatures above 900°C the TiSi₂ becomes discontinuous due to agglomeration, which is accompanied by a rise in sheet resistance. Thus, there is an apparent process window of about 100°C for the formation of stable C₅₄-TiSi₂ before thermal stability problems become evident.

In practice, however, doping, reduced linewidth, and film thicknesses combine to dramatically reduce the process window. The measured activation energy for the C₄₉ to C₅₄ phase transformation on lightly doped (100) single-crystal silicon using a salicide process is about 5.7 eV [22]. Since the transformation is nucleation-limited, the transformation on narrow lines is more difficult. The distribution of silicide sheet resistance on heavily doped n⁺ polysilicon lines is plotted in **Figure 6** for two different linewidths (0.6 and 4.0 μm). The TiSi₂ was formed by reaction of Ti with polysilicon by annealing in a furnace. The tail of the distribution for the narrow lines compared to the wide lines is a result of the presence of fewer nucleation sites per unit length for the narrow lines (assuming constant nucleation density) and incomplete transformation; the latter can result in a mixture of C₅₄ and C₄₉ phases, contributing to the tail of the distribution. The use of higher formation temperatures increases the number of C₅₄ nuclei [28], but, as described previously, this desirable effect must be balanced against the increase in over-the-spacer shorts and reactions of the Ti with the dielectrics in the isolation regions.

The C₅₄ nucleation density is increased by higher-temperature annealing. This phenomenon can be explained by using classical nucleation theory [29], where the nucleation rate R may be expressed as

\[
R = N \exp \left( - \frac{\Delta G^*}{kT} \right) n_v p \exp \left( - \frac{E_a}{kT} \right),
\]

where \( N \) is the total number of nucleation sites per unit volume, \( n_v \) is the number of available atoms at the interface, \( \nu \) the jump frequency, and \( p \) the probability that the atomic jump will be in the proper direction. For polymorphic transformations, \( E_a \) is the activation energy...
for the atomic jump across the interface, and $\Delta G^*$ is the Gibbs free energy at the critical radius. From this equation, it can be seen that the nucleation rate increases exponentially as the temperature is increased. For a thin film this translates into a smaller C54 grain size as the formation temperature is increased (Figure 7).

These factors coupled with improved control over oxygen contamination have resulted in an increased use of rapid thermal processing (RTP) for TiSi$_2$ processing [30]. RTP permits processing at higher temperatures for short durations, mitigating the undesirable titanium/insulator reactions mentioned above. Sheet resistance histograms of RTP-formed TiSi$_2$ on 0.175-μm-wide and 0.15-μm-wide heavily doped n’ polysilicon lines are plotted in Figure 8. By using RTP, extension of the salicide process to dimensions unattainable by the furnace annealing process is possible. However, it is clear that the distribution is degraded for linewidths approaching 0.15 μm. The effect of the degraded TiSi$_2$ distribution on device performance is plotted in Figure 9. At channel lengths approaching 0.15 μm, the performance is degraded. Future TiSi$_2$ processes (for 0.1-μm CMOS devices) may require additional steps to enhance the nucleation density of the C54 phase or may need to bypass the C49 phase entirely. However, even extremely fast ramp rates of 2000°C/s obtained in laboratory settings have not made it possible to bypass the C49 phase [31].

- **Dopant effects**
  The effects of dopants must be considered in practical applications in which it is not atypical to have as many as three dopant species present in the underlying silicon [32]. Each of the dopants (depending on the concentration) can affect silicide formation and phase transformation. The n-type dopants generally impede the silicide reaction more than do p-type dopants. In addition, titanium forms

---

**Figure 6**
Sheet resistance histogram for TiSi$_2$ formed on As-doped polysilicon. Formation annealing was at 650°C for 30 min., and transformation annealing was at 850°C. Results are shown for polysilicon lines having widths of 40 and 0.6 μm.

**Figure 7**
C54–TiSi$_2$ grain size as a function of formation temperature.

**Figure 8**
Sheet resistance histogram for TiSi$_2$ formed on As-doped polysilicon by RTP. Formation annealing was at 700°C for 30 s, and transformation annealing was at 900°C for 1 s. The widths of the polysilicon lines were 0.175 μm and 0.15 μm.
compounds with commonly used dopants and can reduce the concentration of active dopant at the silicide/silicon interfaces, resulting in high contact resistances. Figure 10 shows the increase in the series resistance of a p-type MOSFET device after TiSi₂ formation and subsequent annealing. The increase in resistance is due to boron diffusion into the silicide, lowering the concentration of boron at the TiSi₂/silicon interface. This affects the specific contact resistance between the silicide and the silicon (in this case, boron-doped). For the doping levels used in VLSI applications, this resistance is exponentially related to the doping [33] as

$$R_c \sim \exp \left(\frac{C}{\sqrt[3]{N_D}}\right),$$

where $C$ is a constant and $N_D$ is the substrate doping concentration. This is of particular concern as device junction depths are reduced or as higher temperatures are employed after silicide formation for phase transformation or glass planarization reflow.

**Agglomeration of thin silicide films**

The thermal stability of a titanium silicide film has been given much attention because of the potential to cause not only performance degradation but also yield loss [28, 34–40]. The morphological degradation often referred to as agglomeration occurs as the TiSi₂/Si system attempts to lower the overall surface energy, as shown schematically in Figure 11. Given sufficient thermal energy, silicon diffuses through the silicide film and epitaxially precipitates out at grain boundaries. The surface energy, interfacial energy, grain size, and film thickness are the dominant factors in determining the thermal stability of the film.

With extended annealing at sufficient temperatures, the silicide grains tend to spheroidize, and the film becomes discontinuous. This problem is of particular concern for narrow lines, for which the linewidth can be less than or approximately the same dimension as the diameter of the silicide grains.

Van den hove [34] first proposed using the grain boundary grooving model, as originally developed by Mullins [41, 42], to describe the thermal instability of TiSi₂ films. The grooving model [Equation (3)] provides a means of relating the groove depth $d$ to both time and temperature:

$$d = \tan \beta \left( D_0 \gamma' \Omega^3 C_0 n^3 \frac{t}{kT}\right)^{1/3},$$

where $\beta$ = groove angle (Figure 11), $D_0$ = diffusivity of Si in TiSi₂, $\gamma'$ = silicide/silicon interface energy, $\Omega$ = atomic volume of silicon, $C_0$ = equilibrium concentration of silicon in TiSi₂, $n$ = 3 for bulk diffusion or 4 for interface diffusion, $t$ = time, $T$ = temperature, and $k$ = Boltzmann’s constant.

In the case of silicide films, the silicide becomes discontinuous when the groove depth becomes equivalent to the film thickness. Because the grain-boundary grooving mechanism involves silicon diffusion through the silicide, RTP has proved useful in driving “diffusionless” polymorphic phase transformations and minimizing the diffusion of silicon [43].
Thickness scaling for salicided CMOS processes
Integration of the TiSi₂ into a CMOS process requires balancing the desire for a thicker TiSi₂ layer (for improved thermal stability and lower sheet resistance) against negative effects such as junction leakage. Figure 12 illustrates the impact of silicide thickness on the leakage of shallow (~150 nm) n⁺ and p⁺ junctions. The TiSi₂ thickness is varied by controlling the deposited Ti thickness or the annealing cycles. The final sheet resistance is a measure of the resulting thickness, which is typically thinner on n⁺ than on p⁺ silicon for a given process. The resulting perimeter-dominated leakage current is a strong function of TiSi₂ thickness and thus limits the thickness of silicide that can be formed on the p⁺ junction, e.g., to roughly 55 nm for the 150-nm junction depth.

Salicide processing for sub-quarter-micron devices
Future devices and circuits with scaled vertical as well as horizontal dimensions will put additional pressures on narrow-line salicide formation because of the requirement for thinner silicide layers. Figure 13 illustrates the reduction of the process window for forming low-resistance TiSi₂ lines as the VLSI industry continues to scale linewidth and silicide film thickness. Thus, novel methods will be required in order to extend C54-TiSi₂ salicide processing beyond the 0.25-µm CMOS generation [44]. Examples are the use of Sb-doped Ti to lower the C49 to C54 phase transformation temperature [45] and nitrogen implantation to improve the agglomeration resistance of C54 silicide [46].

Silicides for polycide applications
Polycide processes have been employed in many instances in which the resistance of the gate conductor is critical.
and does not consume an appreciable amount of underlying silicon.

For the polycide process (Figure 3), the silicide is deposited at the stoichiometric composition during the fabrication of the gate stack, rather than being formed from a reaction between pure metal and silicon as in the salicide process. In the polycide process, an amorphous silicide film followed by an insulating film is deposited on top of heavily doped (usually n-type) polycrystalline silicon. The gate stack is etched and then heated to convert the amorphous silicide gate conductor into a crystalline silicide with a lower resistivity [48]. After the deposition of insulating sidewall spacers, the source and drain regions may be salicided as previously described. A TEM cross section of an etched polycide gate stack is shown in Figure 14. In the micrograph, the TiSi$_2$ is roughly 100 nm in thickness (colored blue). The silicide layer is between the thick oxide layer (purple) and the polysilicon layer (green).

The silicide that is used in a polycide application must have a low resistance and good thermal stability; choices include TiSi$_2$, WSi$_2$, TaSi$_2$, and MoSi$_2$ [46–51]. C$_5$4–TiSi$_2$ has a lower resistivity than the other refractory silicides, but it is also more susceptible to agglomeration during high-temperature heating cycles. The choice of which silicide to use for a polycide application is usually determined by design parameters and process complexity [50–52].

The resistance changes that occur in forming C$_5$4–TiSi$_2$ by heating an as-deposited amorphous TiSi$_2$ film in a polycide process are demonstrated in Figure 15. In this figure, the in situ resistance versus temperature at a heating rate of 50°C/s is plotted for the crystallization of a 100-nm-thick amorphous TiSi$_2$ film. For this film, a sharp decrease in the resistance occurs at 400°C, corresponding to the crystallization of the amorphous film into C$_4$9–TiSi$_2$. With further heating, the resistance of the polycide decreases slightly until just below 800°C, where there is a sharp decrease in the resistance of the film. This sharp decrease is due to the transformation of the high-resistance C$_4$9–TiSi$_2$ phase into the desired low-resistance C$_5$4–TiSi$_2$ phase (15–20 μΩ-cm). With further heating of the polycide to 900°C, the resistance of the film increases rapidly because of agglomeration. This high-temperature agglomeration is one of the major factors limiting the use of TiSi$_2$ in polycide applications [22].

The increase in the C$_4$9 to C$_5$4 transformation temperature due to geometrical patterning that is observed in salicide processing also occurs in submicron TiSi$_2$ polycide structures. As shown in Figure 15, for unpatterned polycide films, it is possible to study the C$_4$9 to C$_5$4 transformation in situ during rapid thermal annealing using resistance measurements. However, for submicron polycide structures, in situ resistance measurements cannot be used to study this transformation during annealing, because of difficulties in making and maintaining electrical contact to submicron polycide films at high temperatures. Another approach to studying the...
C49 to C54 transformation in situ during annealing in submicron polycide structures is to examine the X-ray diffraction spectrum of a chip having submicron polycide lines during heating. This is accomplished by using a synchrotron X-ray source to generate a large X-ray flux and a position-sensitive detector to measure the diffraction patterns with millisecond time resolution. Figure 16 shows an example of this type of analysis for a test chip having 0.4-μm-wide, 100-nm-thick TiSi₂ polycide gate lines that are from one to several hundred microns long. In this figure, the diffracted intensity (z-axis) is plotted versus the 2θ diffraction angle (x-axis) and temperature (y-axis). It is evident that the only titanium silicide phases present in the submicron polycide structures are C49-TiSi₂ at the lower temperatures and C54-TiSi₂ at the higher temperatures. The temperature for the onset of the C49 to C54 transformation is best determined by examining a color projection of the X-ray diffraction intensities of Figure 16 into the 2θ/temperature plane, as shown in Figure 17. It is apparent from this figure that the C54 (040) peak begins to form from the precursor C49 (131) and (060) peaks at about 910°C. This C49 to C54 transformation temperature is about 100°C higher than that observed for the corresponding unpatterned polycide film (Figure 15).

The reaction that occurs in heating an amorphous WSi₂ polycide film is demonstrated in Figure 18. In this figure, the in situ resistance versus temperature at a heating rate of 50°C/s is plotted for the crystallization of a 200-nm-thick amorphous WSi₂ polycide film. With heating there is a gradual decrease in the resistance of the film until about 450°C. At this temperature, there is a sharp increase in the resistance of the film owing to its crystallization into the low-temperature hexagonal form of WSi₂. The resistivity of the hexagonal WSi₂ phase is approximately 600 μΩ-cm. With further heating, the resistance of the WSi₂ polycide film remains constant until approximately 700°C. From 700 to 900°C, there is a gradual decrease in the resistance of the film owing to the polymorphic transformation of the hexagonal WSi₂ phase into the tetragonal WSi₂ phase. Further heating to 1050°C is required to completely transform the hexagonal WSi₂ into tetragonal WSi₂. Even with this type of elevated annealing condition, agglomeration of the WSi₂ polycide film is not observed. Future polycide gate conductor materials will most likely continue to be refractory silicide-based. Titanium disilicide offers the advantage of a low resistivity for improved gate delay for CMOS gates, but offers relatively poor thermal stability and increased process complexity compared to the other refractory silicides proposed for polycide conductors.
most silicide films [55], leading to dopant depletion at the silicide/silicon interface and, in certain cases, counterdoping at the interface. These effects can cause unpredictable threshold voltage shifts in MOSFET devices.

**Local interconnections**

The introduction of the use of silicides to produce local interconnection (LI) structures has provided a means of achieving circuit density improvement. Several schemes have been proposed over the last ten years for forming an additional level of interconnection for localized wiring purposes [4-11]. Because the local interconnection schemes require only one additional masking level and generally provide a 20–30% improvement in SRAM cell size, the productivity improvement is clear for SRAM and logic chips (e.g., microprocessor chips) requiring large amounts of on-chip cache. Although the line resistance is higher than that of conventional "global" wiring levels, the pitch is often significantly better, thus offering a density advantage in circuits where the increased resistance is acceptable because of the close proximity.

LI structures must be easily integrated with two or more underlying CMOS structures, which are usually silicided. There are four local interconnection approaches that have been evaluated for the manufacturing of CMOS circuits: silicide-based local interconnection, TiN-based local interconnection, selective CVD tungsten-based local interconnection, and tungsten-based "damascene" stud local interconnections. The process flows for each are detailed in Figures 19–22. The silicide-based (Figure 19) and TiN-based local interconnection (Figure 20) approaches are similar in that the silicide contacts to the source, gate, and drain regions and the local interconnections between those regions are formed concurrently. However, the silicide contacts to the source, gate, and drain regions and the local interconnections between those regions are formed simultaneously during thermal processing [4]. The selective CVD tungsten-based local interconnection approach detailed in Figure 21 is similar to the silicide-based and TiN-based local interconnection approaches in that the source, gate, and drain contacts and the local interconnections are all formed concurrently. However, the silicide source, gate, and drain contacts for the silicide-based and TiN-based local interconnection approaches are replaced by selective tungsten contacts to those regions. The tungsten-based damascene approach [10], detailed in Figure 22, is the only local interconnection scheme in which the formation of the local interconnections is decoupled from the formation of the silicide source, gate, and drain contacts. The silicide source, gate, and drain contacts are formed, and a layer of $Si_N$ and a thick overlying layer of PSG (phosphosilicate glass) are deposited. The PSG is then planarized and patterned to open holes to the source, gate, and drain regions, and a blanket tungsten layer is deposited by CVD. The excess W on top of the PSG is then removed by

A significant amount of work or perhaps invention will be required to integrate polycides into dual-work-function gate applications [22, 53, 54]. This is because at the temperatures required to form low-resistance polycide contacts (1000°C for $WSi_2$), dopants diffuse quickly through
chemical-mechanical polishing (CMP), isolating the W studs and creating the local interconnections. This decoupling of the salicide process and the formation of the local interconnections provides flexibility in process optimization and improved yields. For these reasons, the damascene stud interconnection is favored within IBM over the silicide, TiN, or selective tungsten interconnection schemes.

**Tungsten-based damascene interconnections**

The versatility of the tungsten-based damascene approach is demonstrated in an SRAM application depicted in Figure 23. The figure shows a low-angle SEM view of a portion of an SRAM array containing six-device cells, where the insulator material has been etched away to reveal the interconnections. The passivating oxide films have been removed to the nitride layer which protects the active devices. The word lines are salicided polysilicon gates. The damascene stud local interconnections are the structures which provide the cross-coupling for the cells and also act as the lower part of the contact studs to the first metal level. The tungsten studs are the upper part of the contact paths to the first global interconnection level. Note that these contact studs connect to the LI contact studs as well as to the LI cross-coupling bars.

The tungsten-based damascene approach, depicted in Figure 22, begins after the active devices are fabricated.

---

**Figure 20**

Process flow for forming TiN-based local interconnections. After forming the devices (a), a blanket layer of Ti is deposited (b). The salicide-formation annealing is carried out in nitrogen, forming a layer of TiN on the titanium surface. A resist mask is then applied (c), and the TiN layer is etched in the unmasked areas. The resist is removed, completing the local interconnections (d).

**Figure 21**

Process flow for forming local interconnections by selective CVD tungsten. A single polysilicon layer forms the gate conductor and infrastructure of the local interconnections (a). The masking step is carried out to permit removal of the sidewall spacers where desired (b). A selective CVD tungsten layer is grown on the exposed silicon surfaces, forming interconnections across the thin gate oxide regions and electrical contacts to the source, gate, and drain regions of the devices (c).

**Figure 22**

Process flow for forming tungsten-based "damascene" local interconnections. The conventional salicide process is followed by nitride and PSG depositions, and the PSG is planarized by CMP (a). A masking layer is used to pattern base contacts and local interconnections (b). Tungsten is deposited via CVD to fill the etched regions in the PSG layer, followed by CMP planarization, thus defining the contacts and local interconnections.
The LI mask is defined, and the PSG is etched with an etch that is selective to the Si,N,. The nitride is both parasitic capacitances and the etching process under-etching produces high resistance or open contacts. Film thicknesses, etch rate ratios, and etch control is limited to one width, e.g., 0.6 µm, but may be any length.

At this point they are passivated by PECVD silicon nitride followed by a PSG film. The nitride acts both as an ionic contamination barrier and as an etching stop. The 4–8 wt.% PSG is then polished by CMP to planarize the PSG, leaving the desired thickness over the highest topography. The thickness of the PSG is chosen in order to optimize both parasitic capacitances and the etching process window. The LI mask is defined, and the PSG is etched with an etch that is selective to the SiN₄. The nitride is then removed with an etch that is selective to both SiO₂ and TiSi₂. These two steps are critical: Over-etching or loss of selectivity can lead to junction leakage, while under-etching produces high resistance or open contacts. Good integration of the PSG, silicon nitride, and silicide film thicknesses, etch rate ratios, and etch control is necessary in order to consistently provide acceptable contacts and interconnections. Next, a thin Ti/TiN composite liner is deposited by collimated sputtering, followed by a thick CVD tungsten deposition. Since the tungsten must completely fill the local interconnection “troughs” and “holes,” the LI images on the mask are limited to one width, e.g., 0.6 µm, but may be any length. This allows complete stud filling while minimizing the amount of tungsten that must be deposited. CMP is then used to remove the tungsten and liner above the PSG, resulting in studs of tungsten embedded in a layer of surrounding PSG. Passivation is then achieved through use of a PECVD oxide, and contacts are etched to the underlying studs. A sputtered liner and CVD tungsten film are then deposited and polished to form the completed stacked contact. The LI studs are borderless to both polysilicon word lines and diffusion regions; that is, an LI stud need not be completely surrounded by diffusion regions, but can lap over onto the isolation region, which in this case is a shallow trench isolation (STI). The ability to cross the isolation-region boundaries is a very significant and essential LI feature which allows for smaller cell sizes due to relaxed overlap tolerances.

There are several other advantages of the tungsten damascene approach in addition to the decoupling of the silicide and location interconnection processes. First, since chemical–mechanical polish is used for planarization rather than glass reflow, all post-silicide processing is at low temperature (below 550°C), which minimizes problems due to thermal agglomeration. Second, a high degree of planarity is also achieved, making subsequent processing less complex, and the contacts to both gates and diffusion regions can extend across the silicide/field oxide (STI or LOCOS) boundaries without yield detraction. Finally, because of the resistivity of tungsten and the thicknesses used, this LI approach provides a relatively low sheet resistance, typically less than 1 Ω/□.

**Concluding remarks**

Although the trends regarding MOSFET device length and minimum feature size are essentially predetermined through the year 2000 by the availability of photolithography tools and demand for increased performance, those regarding local interconnections are less evident. The trend toward dual-work-function gate materials will almost certainly continue, thereby increasing the use of the salicide process. The most common silicides used in that process are TiSi₂ and CoSi₂. We have discussed in some detail the factors that limit the extendability of TiSi₂. Although CoSi₂ is less thermally stable on polysilicon than TiSi₂, it is much easier to obtain the low-resistance CoSi₂ phase than the C54–TiSi₂ phase. Lower post-silicide processing temperatures should permit CoSi₂ to become a more valid option. For polycide applications, WSi₂, TaSi₂, and MoSi₂ will probably continue to be used because of their refractory nature, and TiSi₂ should begin to find acceptance in applications in which lower resistivity becomes critical and thermal stability less an issue because of the use of lower post-processing temperatures.

The proliferation of local interconnection approaches that occurred in the last ten years will most likely decline with the emergence of a few viable ones. Although certain SRAM cell approaches such as those involving the use of thin-film transistor p-type MOSFETs may mitigate their
benefits, local interconnections will probably continue to find acceptance for imbedded cache applications, in full-CMOS SRAM, and in logic circuits.

Acknowledgments
The authors wish to thank F. M. d’Heurle, J. M. E. Harper, G. Miles, and J. Nakos for their many helpful discussions and insights over the years that we have all worked together. The authors would also like to acknowledge G. Gifford for his contributions to the polycide section of the paper, and G. B. Stephenson, S. Brauer, C. Cahral, Jr., J. Jordan-Sweet, G. Moráics (Boston University), and K. Ludwig, Jr. (Boston University) for their contributions in the development of synchrotron X-ray analysis to measure silicide formation, in situ, during rapid thermal annealing. The synchrotron analysis was performed at beamline X20C at the Brookhaven National Laboratory.

References


35. C. A. Sukow and R. J. Nemanich, “Comparison of the


Received December 22, 1994; accepted for publication June 21, 1995
Randy W. Mann  IBM Microelectronics Division, Burlington Facility, Essex Junction, Vermont 05452 (rmann@btvlabm.vnet.ibm.com). Mr. Mann is an advisory engineer in the Process Development and Integration Department at the Essex Junction, Vermont, facility of the IBM Microelectronics Division. He received his B.S. from the University of North Carolina at Greensboro and his M.S. in materials science and metallurgical engineering from the University of Notre Dame in 1982. In 1982 he joined IBM in Essex Junction, where he worked on CMOS and bi-CMOS process development for high-performance SRAM and logic applications. In the last eight years he has worked on the development and process integration of half-micron, sub-half-micron, and quarter-micron device technologies. His current interests include silicides, silicide-device interactions, and thin-film transistors. Mr. Mann is a member of the Materials Research Society and the Electrochemical Society; he is an author or co-author of 22 publications.

Larry A. Clevenger  IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (larrycl@watson.ibm.com). Dr. Clevenger is a research staff member in the Thin Film Metallurgy and Interconnections Department at the IBM Thomas J. Watson Research Center. He received his B.S., Summa Cum Laude, from the University of California, Los Angeles, in materials engineering in 1984 and his Ph.D. degree in electronic materials from the Massachusetts Institute of Technology in 1989. He then joined the IBM Research Division, where he has worked on developing thin-film materials for submicron interconnections, diffusion barriers, and silicide contact applications. Dr. Clevenger’s recent interests have focused on the in situ measurement of submicron-patterned thin-film reactions during rapid thermal annealing, using synchrotron X-ray diffraction and resistance measurement techniques. He is a member of the Materials Research Society and the American Vacuum Society, holds four patents, and is an author or co-author of 61 publications.

Paul D. Agnello  IBM Research Division, Advanced Silicon Technology Center, Hopewell Junction, New York 12533. Dr. Agnello received his B.S., M.S., and Ph.D. degrees, all from the Electrical, Computer and Systems Engineering Department of Rensselaer Polytechnic Institute. In 1988, after completing his dissertation, he joined the IBM Thomas J. Watson Research Center, where he has worked on low-temperature Si and SiGe epitaxy for heterostructure devices and silicides for deep-submicron CMOS devices. Dr. Agnello is currently working at the Advanced Silicon Technology Center, where he is involved in the integration of salicide and polycide technology into deep-submicron CMOS devices and circuits for advanced logic and SRAM applications.

Francis R. White  IBM Microelectronics Division, Burlington Facility, Essex Junction, Vermont 05452 (fwhite at BTVMOS, fwhite@vnet.ibm.com). Mr. White received his B.S. and M.S. degrees in electrical engineering from the University of Maine at Orono in 1977 and 1978, respectively. He joined IBM in 1978 and has worked in the areas of dry etching of silicon compounds and process development. He had a sojourn in management and is a member of the American Vacuum Society, where he has been active in local chapter activities. Mr. White currently works in the area of process development for advanced CMOS technologies.