Progress toward a 30 nm silicon metal–oxide–semiconductor gate technology

D. M. Tennenbaum
Bell Laboratories, Lucent Technologies, Holmdel, New Jersey 07733

Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974

W. Timp
University of Illinois, Champaign–Urbana, Illinois

(Received 16 June 1999; accepted 17 September 1999)

We report on progress toward scaling both N-metal–oxide–semiconductor (MOS) and P-metal–oxide–semiconductor MOS transistors to a gate length of 30 nm. We describe lithography and pattern transfer results that are suitable to meet this goal. Scanning capacitance microscopy is used to determine the effective channel lengths and source drain junction depths on cross-sectioned devices to optimize the fabrication process. We present interim electrical results obtained for high performance, down to $L_g = 57$ nm, N-MOS and P-MOS transistors made using this process. We have also used a device simulation program to predict subthreshold current for N-MOS transistors with gate lengths from 40 to 26 nm. The simulation provides insights into the effects of critical dimension control and edge roughness on leakage current and has implications for extending large scale integration of MOS technology beyond 50 nm. © 1999 American Vacuum Society.

[S0734-211X(99)15506-9]

I. INTRODUCTION

Reduction of the gate dielectric thickness, gate length, and area of transistors is being pursued to improve performance and circuit density. The key element in the continued progress in the performance of silicon metal–oxide–semiconductor (MOS) transistors is concentrated in the region between the source and drain and involves the formation and definition of the gate stack. Extremely high performance sub-100 nm transistors require both ultrathin gate dielectrics (<2 nm) and ultrashallow junctions (<50 nm), both to achieve a high saturation current with a low power supply voltage and to control short channel effects. While the most stringent requirements for high performance, sub-100 nm, complementary MOS (CMOS) technology are dictated by the P-MOS transistor due to boron diffusion through the gate dielectric and transient enhanced diffusion in the contacts, both N-MOS and P-MOS require similar extremes in gate level lithography and etch selectivity. Many process parameters must be varied to explore nonideal properties and the practical limits of each technology. To speed the convergence of this process, we employ a transistor test structure, requiring only one level of lithography, which is performed by direct write e-beam lithography. After completion, these devices are directly probed electrically and, along with other structures designed for cross-sectional preparation, analyzed by scanning capacitance microscopy (SCM) and transmission electron microscopy (TEM) to provide data on physical and electrical channel dimensions.

We report here on progress toward scaling a gate technology to 30 nm gate lengths. We discuss recent lithographic and pattern transfer related results from this scaled silicon transistor initiative and include both n- and p- channel transistors. We have targeted 30 nm physical gate devices since even optimal source–drain implantation conditions give rise to electrical channels that are smaller than the physical gate. Therefore, as our SCM data and modeling below will show, 30 nm represents the size biasing required to produce practical transistors. We report first on materials and processes used in fabrication, then on electrical characterization of recent transistors, and, finally, on simulation results based on these measurements which help predict the utility of large scale integration of such transistors.

II. FABRICATION

The epitaxial growth and subsequent cleaning steps prior to growth of the gate oxide are required to maintain the atomic force microscopy (AFM) measured root mean square (rms) roughness of 0.06 nm or less to preserve the quality of the Si/SiO$_2$ interface. The gate stack has been scaled for this work and consists of 80 nm of tetraethoxysilane (TEOS) hard mask over 80 nm of WSi$_2$ on 50 nm of polycrystalline silicon on gate oxides ranging in thickness from 0.55 to 2.0 nm thick on epitaxial (100) silicon substrates. Prior to gate oxide growth, the substrates used for the P-MOS transistors were implanted with $3 \times 10^{11}$ cm$^{-2}$ and $1.5 \times 10^{12}$ cm$^{-2}$ As at 7 and 50 keV and $8 \times 10^{12}$ and $1.4 \times 10^{13}$ cm$^{-2}$ P at 80 and 200 keV while the N-MOS substrates received 1.4 $\times 10^{13}$ and 2$\times 10^{13}$ cm$^{-2}$ B at 7 and 30 keV. The gate oxide was grown at 1000°C by rapid thermal oxidation (RTO) in pure oxygen. The 40–50 nm layer of amorphous silicon is
deposited by chemical vapor deposition (CVD) at 510 °C using silane and is in situ doped with phosphorus using PH₃. The as-deposited film is amorphous but is annealed to form polycrystalline material. The 90 nm WSIₓ (x ~ 2.7) is dc magnetron sputter deposited from a single target. The final layer of the gate stack is the 100 nm hard mask, which is a conformal, low temperature, low pressure CVD (LPCVD) oxide deposited using a TEOS decomposition process. Following the poly Si deposition, the gate stack was implanted with 4×10¹⁵ cm⁻² boron at 8 keV for P-MOS and 40 keV As at 4×10¹⁵ for the N-MOS.

The wafers are subsequently patterned using only a single level of electron beam lithography to define the gate, source, and drain contacts simultaneously, where the gate electrode (325 μm × 325 μm) entirely surrounds both the source and drain contacts (both 25 μm × 25 μm) in a ring geometry. Isolation is provided by the long channel and high threshold associated with the parasitic portions of the transistor. The pattern is subsequently transferred using reactive ion etching (RIE) which stops on the gate oxide.

The TEOS hard mask is etched in CHF₃/SF₆ (65:3) and overetched in CHF₃/CF₄/Ar (30:4:60). The etch into the hard mask is performed in an Applied Materials AMI 5000 magnetically enhanced RIE system. The main etch end point detection is determined using an optical emission type signal. Following the hard mask etch, the resist is removed.

The remaining gate stack RIE is performed in a Lam Research TCP 9400SE transformer coupled etcher. The tungsten silicide is etched using Cl₂ in a 20% He/O₂ background. The gate structure is completed by transferring the pattern into polycrystalline silicon using a H₂O/He/O₂ mixture (100:208:2) in order to stop on the ultrathin gate oxide with very high selectivity. Optical emission end point detection is again used to determine a consistent stopping point. The conditions for the etch are similar to those previously reported, but the main poly-silicon etch step has been eliminated. The overetch step alone is sufficient due to the thinner polysilicon and the on-average thinner gate oxides used in this device design.

After RIE, the MOS field effect transistor (MOSFET), ultrashallow source-drain extensions are implanted using 4 keV. As for the N-MOS and 0.50 keV B for P-MOS with a dose of 4×10¹⁵ cm⁻². Ultrashallow junctions with high conductivity are necessary to diminish short channel effects without compromising device performance.

Following the deposition of a 150 nm thick undoped TEOS sidewall and a sidewall etch, a 3×10¹⁵ cm⁻² dose of B was implanted at 1 keV for P-MOS (for N-MOS, it was 3×10¹⁵ cm⁻² at 60 keV) to form the contacts. The implants were activated using rapid thermal annealing (RTA) at 1000 °C for 5 s. We formed either Ti or Co silicide in the contacts to reduce the sheet resistance to 4–8 Ω/sq.

We recently reported on promising lithographic results using negative tone resists from the Sumitomo, and here apply these materials to fabricate scaled transistor structures. We have found that NEB 22A from Sumitomo is especially useful since it exhibits high sensitivity (5–15 μC/cm² at 50 keV) and improved resolution over previous resists such as SAL603, a good process latitude, and improved etch resistance during RIE of the hard mask (3–1 μm/s, a 50% improvement over SAL603).

To optimize the thermal processing of the NEB 22 resist, the post exposure bake (PEB) time and temperature and post application bake (PAB) time were varied systematically and are tabulated in Table 1. From Table 1 an average sensitivity of the critical dimension (CD) control to process bake conditions can be determined. We find that a PEB temperature sensitivity of 4.8 nm/deg is an especially challenging process control issue. For linewidth effect of the PEB and PAB durations, we find 8 and 17 nm/min, respectively. It is somewhat surprising that the linewidth sensitivity to the PAB time is about twice that of the PEB time.

Table 1 also shows for various process protocols the effect on linewidth sensitivity to the e-beam dose. The best process at one linewidth may not be best for a smaller linewidth. For convenience, 75 and 120 nm lines are chosen to compare. Unfortunately, the more robust process may come at the expense of writing speed. In the example shown in Table 1, our nominal process (condition 5) allowed a 75 nm linewidth to be produced at a sensitivity of 12 μC/cm² but with a slope of 6.75 nm/10% dose variation, while condition 2 required a dose of 32 μC/cm² but exhibited an improved slope of 5.5 nm/10% dose variation, a modest but desirable improvement.

The etch sequence described above to transfer the gate pattern in the gate stack narrows the lithographic features by 10–20 nm. As a result, gate level lithography requires a minimum resist CD of 40–50 nm, which may be met using the NEB 22. As etch fidelity improves, however, 30 nm CD in gate lithography may be required. A candidate resist, which affords further resolution improvement, is Sumitomo NEB 31. Our highest resolution to date in NEB 31 is 28 nm for isolated lines. Figure 1 shows a 32 nm linewidth in a 0.1 μm thick NEB 31 resist. These are achieved by e-beam direct write in a JEOL JBX-600FS thermal field emission system operating at 50 kV using a Gaussian beam width in the range of 10–15 nm. These smallest features are written as single pass lines. Other attributes such as edge roughness also appear to improve with this higher resolution version of NEB resist.

The minimum linewidth for these resists is found to depend on thickness. As the resist thins from 500 to 120 nm,
the resolution improves from 75 to 36 nm. While thinner resist may be capable of even smaller linewidths, current etch selectivity during the hard mask etch dictates that a minimum resist thickness of 120 nm is necessary.

The lithography is subsequently transferred into the gate stack using the RIE sequence above. To date we have successfully transformed the e-beam written NEB 22 patterns into physical gate lengths of 35 nm. Figure 2 shows a TEM cross section through a 40 nm gate n-MOSFET. We observe that the structure is close to vertical and that the etch has sufficient selectivity to stop on a gate oxide here only 1.3 nm thick. Using many wafer splits, we have systematically explored gate stacks with gate dielectric layers of 0.55–2.0 nm. A TEM image of a gate region with a 0.7 nm gate oxide shown in the inset demonstrates the uniformity of the dielectric growth even in these extremely thin layers.

In addition to difficulties associated with achieving the 40–50 nm linewidth in 120–200 nm of resist, we have encountered additional failure mechanisms in these resists. First, we observe that high quality resist profiles often resulted in rough or even serpentine gate patterns after the gate etch. This is illustrated in the left half of Fig. 3 for a 30 nm gate viewed by scanning electron microscopy (SEM) after the hard mask etch step. This distortion was found with increasing frequency as the resist linewidth decreased below 70 nm. We also found that for linewidths of 60–70 nm for which resist layers of 200 nm were used, the effect was mitigated by adding an additional postdevelopment bake step, as shown in the right half of Fig. 3. The thinner resists, which were required to achieve the finer linewidths (e.g., 28–30 nm), however, conspire to produce an additional failure mechanism. We have observed a falling over of these lines both during development and during the postdevelopment bake cycle. This effect is suggestive of a depression of the glass transition temperature measured recently in other polymers78 when thinner layers are used. A remedy to this failure mechanism may lie in working with development at reduced temperatures. That work has been investigated and is reported elsewhere.79

As a final gate formation complication, the very smallest gate length structures (~35 nm) have also been observed to distort even when the patterned hard mask is of good quality. This is illustrated in a plan view TEM image, Fig. 4, comparing a $L_e = 35$ and 160 nm etched gate stack. The wiggles along the small gate appear to be correlated with individual grains of the $W_{Si}$. perhaps indicating a relaxation process when $L_e$ approaches the individual grain sizes. The grains in the larger gate appear to be held in place by the matrix of neighboring grains.

**III. SCM ANALYSIS**

In scanning capacitance microscopy, an atomic force microscope is used with an electrically conductive probe, which acts as a resonant capacitance sensor. The signal in SCM is a measure of $dC/dV$ and thus, for an applied ac voltage, the sensor maps the change in capacitance, which is interpreted as changes in the depletion (or accumulation) layer. In an edge cleaved transistor this signal is readily interpreted as changes in the carrier concentration, revealing both the magnitude and the sign of the carrier concentration. SCM, in conjunction with ultrasharp silicon tips, has been used to
A TRANSISTOR MEASUREMENTS

...
Fig. 7. dc electrical performance data for two sets of matched N-MOS (blue) and P-MOS (red) transistor pairs. For the upper set, \( I_{\text{on}} = 64 \, \text{nm}, V_{th} = V_{gs} = 0.32 \, \text{V}, \) and \( t_{ox} = 1.5 \, \text{nm} \); while for the lower set, \( I_{\text{on}} = 57 \, \text{nm}, V_{th} = V_{gs} = 0.31 \, \text{V}, \) and \( t_{ox} = 1 \). The top set shows very good \( I_{\text{on}} \) and threshold voltages but modest \( I_{\text{off}} \) characteristics. The bottom set shows very good \( I_{\text{on}} \) performance but much larger \( I_{\text{off}} \) current and a lower threshold current.

Semiconductor Industry Association (SIA) projections of 0.6 mV/µm drive current for N-MOS and 0.25 mV/µm drive current for P-MOS with a 0.3–0.6 V power supply voltage and a subthreshold leakage < 10 nA/µm.

V. SIMULATION AND DISCUSSION

The narrow process window to attain these results makes the effect of variations in gate linewidth a critical issue for lithographers working at this extreme size scale. We therefore have also used our simulation program to predict the effect of linewidth variations on subthreshold current for N-MOS transistors with gate lengths below 40 nm. The model assumes a gate oxide of 1.3 nm and the same process parameters incorporated in our scaled process described above. Since the process is identical for all cases, only the gate length varies. Figure 8 illustrates the dramatic increase that occurs in \( I_{\text{off}} \) at room temperature as the gate length is decreased down to 26 nm. We note that the general behavior of \( I_{\text{off}} \) is given as

\[
I_{\text{off}} \propto I_{\text{on}} e^{(V_{gs} - V_{th})/S}
\]

where \( V_{gs} \) is the gate voltage, \( V_{th} \) is the threshold voltage, and \( S \) is the subthreshold voltage slope (in mV/decade). We have

![Diagram showing Ioff vs. Vgs for N-MOS and P-MOS transistors with various gate voltages and drain voltages.](image)


Fig. 8. Simulation of subthreshold current variation with drain current for a N-MOS transistor with gate lengths from 40 nm to 26 nm for three different power supply voltages. The blue line is the estimated requirement from the SIA Technology Roadmap. None of these cases meets this target for leakage current, indicating that the power consumption may be too high for VLSI applications.
plotted a large number of devices with gate lengths varying
from 40 down to 26 nm in steps of 1 or 2 nm. A complete set
is plotted for three different supply voltages, 0.5, 1.0, and 1.5
V, the probable useful range for circuits. We see that at 1.5
V, where the \( I_{on} \) target is most readily achievable, even the
40 nm transistors are nearly two orders of magnitude too
"leaky" to meet very large scale integration (VLSI) stan-
dards. Nearly a factor of 10 improvement in \( I_{on} \) is obtained
by decreasing the supply voltage to 0.5 V. The accompa-
nying slope change associated with this decrease in supply volt-
age makes tolerances for variation in linewidth even tighter
(steeper slope). We note that none of these devices meets the
SIA roadmap requirements for subthreshold leakage current
(10 nA/\( \mu m \)).

The plots in Fig. 8 also illustrate the critical role that
linewidth variations may play in the practicality of transistor
designs and the implications for lithographic control are very
worrisome. The logarithmic vertical scale causes small vari-
ations in gate length to result in large departures from accept-
able leakage currents. In effect, the variation in gate length

\phantom{.}

\( \text{can be thought to "average" for } I_{on} \text{, but for } I_{off} \text{ even a low}\)

\( \text{density of shortened gates will dominate. This same principle should also apply to variations within a single gate and therefore make edge roughness during lithography and pattern transfer especially stringent, especially if low power supply voltages (e.g., 0.5 V) are adopted.} \)

Ultimately these factors affect the density of transistors
which can packed into a given chip. Without further en-
heartments, it may be that high performance below
40 nm will have to develop a niche application or be inserted sparingly into VLSI designs.

The N-MOS simulations also predict that 20 nm physical
gate length transistors are shorted. Therefore even when im-
provements in the dopant process are made to optimize the
source/drain extensions and junction depths, the shortest
channel high performance transistor will likely require gate
lengths of the order of 30 nm [with allowance for small (~5
nm) differences between N-MOS and P-MOS].

VI. SUMMARY AND CONCLUSIONS

We have demonstrated an improved MOS transistor pro-
cess capable of producing high performance devices down to
\( L_{\text{eff}} = 20 \text{ nm}. \) Direct SCM measurements of \( I_{on} \) for our n-
and p-channel MOSFETs determine that a best effort may
allow \( L_{\text{eff}} = 30 \text{ nm to produce working transistors. To this end we have explored the various gate technologies needed to fully exploit this capability. Using direct write e-beam lithography we have investigated use of the NEB series of

\( \text{resists from Sumitomo as high throughput, high resolution candidates, since they have produced suitable resist profiles down to 28 nm. The new scaled gate stack comprises several thinner layers. These include a thinner gate dielectric (down to 0.55 nm) to improve drive current capability and a thinner hard mask and polysilicon layers to accommodate both the thinner resist layers available and to allow use of a less ag-
gressive reactive ion etch process. The new RIE sequence permits formation of nearly vertical profiles yet stops on gate oxides ~1 nm thick. The etch has successfully demonstrated gate formation down to 35 nm. We have presented interim device results, which show high performance N-MOS, and
P-MOS transistors with gate lengths near 60 nm. Further
work will be needed to simultaneously meet the high drive current and low leakage current requirements dictated by the
SIA Technology Roadmap, however. Using measured results we have calibrated a device simulation program to model the
leakage current in N-MOS transistors with gate lengths in the
range 26-40 nm and with power supply voltages from 0.5 to
1.5 V. None of these satisfies the SIA required leakage curent
specification. A second observation from these simulations is
the very steep slope evident in Fig. 8. It indicates that
small variations in gate length will result in large departures
from acceptable leakage currents. This applies to both varia-
tions within a single gate as well as among the gates on a
given chip. This must be overcome if CMOS technology at
this size scale is to be used in VLSI type designs.} \)

3Muller, T. Sorsich, S. Moccio, F. Baumann, K. Evans-Lutterodt, and G. L.
6Sumitomo Chemical Co.
7Trademark of Shipley Corp., Newton, MA.
8D. S. Fryer, J. J dePablo, and P. F. Nealey, Proc. SPIE 3333, 1031
(1994).
12R. Kleinman, M. L. O'Malley, F. H. Baumann, J. P. Gano, W. G. Temp,