Titanium Silicide/Silicon Nonohmic Contact Resistance for NFET's, PFET's, Diffused Resistors, and NPN's in a BiCMOS Technology

Terence B. Hook, Randy W. Mann, and Edward J. Nowak

Abstract—Self-aligned titanium silicide is often used to minimize the polysilicon and diffusion sheet resistances. Current is delivered to the channel of FET's, the body of diffused resistors, and into the active region of NPN's through the titanium silicide/silicon interface. This contact resistance can represent a significant fraction of the total device resistance for devices of small dimensions, and contributes to a loss in circuit performance. The impedance of this interface is a function of the doping level in the silicon immediately below the interface, and this doping level is a sensitive function of the heat applied to the structure after the formation of the silicide. The correspondence of FET series resistance, emitter resistance, the diffused resistor end effects and the non-ohmic nature of a contact after heat is applied is demonstrated. Use of a rapid thermal anneal to obtain the contact resistance is demonstrated for a 0.8-µm BiCMOS technology.

I. INTRODUCTION

The technologies discussed in this paper utilize 0.8-µm lithography for SRAM and logic applications. The BiCMOS device set consists of an NFET, a PFET, a polysilicon emitter NPN, and a p-type diffused resistor. The CMOS-only variant of this technology offers FET's identical to those in the BiCMOS version, but substitutes an arsenic-diffused resistor for the boron resistor employed in the latter. Both versions feature a damascene tungsten local interconnect, followed by four levels of wires, interconnected by damascene tungsten vias [4]. Chips fabricated in these technologies include a 300 K circuit CMOS logic family, a 200 K circuit BiCMOS logic family, and a 576 K 3.5-ns BiCMOS static RAM [5]-[7].

This paper addresses the TiSi₂/Si system and the effect of heat cycles (700°C-900°C) after the formation of the silicide on the contact resistance; the consequences for all the devices (PFET's, NFET's, resistors, NPN's) in the technology; and the requirements for successful silicidation of narrow polysilicon lines are described in conjunction with the requirements for the active devices. Data comparing a furnace anneal to a rapid-thermal-anneal (RTA) process are presented, as well as experiments directed to optimization of the RTA process.

The interaction of titanium silicide and common silicon dopants has been studied from several viewpoints. Using titanium silicide as a source of dopants has not been successful because metal-dopant compounds form, which limits the amount of dopant that can be introduced into the silicon from the silicide [8]-[10]. The formation of titanium silicide over heavily doped junctions has been found to be a function of the dopant type and concentration in the silicon [11], [12]. The thermal stability of the titanium silicide is a function of the dopants in the silicide [13], and rapid diffusion of dopants in titanium silicide has also been observed [14]. Interaction between the silicide and the doping in the silicon have been studied as well. When heat is applied to a silicided wafer, metal-dopant compounds (e.g., TiAs, TiB₂) may form. Diffusion of the dopants into the silicide may occur, depleting the silicon surface of the dopants. These effects cause an increase in the contact resistance between the silicide and the doped silicon [15]-[18].

The various active devices (FET's, NPN's, and resistors) all show a similar sensitivity to the contact resistance between the titanium silicide and the underlying silicon junction. This contact resistance is a function of the heat applied to the wafer after the titanium silicide formation, and can be highly non-ohmic. It was found that although the non-ohmicity of the n-type contacts could be substantially mitigated by an increased implant dose, a similar advantage could not be obtained for p-type contacts. Characteristics of devices processed with a rapid thermal anneal process that simultaneously combines the requirements of low-resistivity titanium silicide and minimizes the contact resistance are demonstrated. The net result with the improved process is a 20% improvement in emitter resistance; a 23% and 10% improvement in PFET and NFET series resistance, respectively; and a reduction by one-half of the contact resistance at the ends of the diffused resistors.

II. PROCESS SEQUENCE

The final cross section of all the devices is shown in Fig. 1. Fig.1(a) is a cross section of all of the devices to be found in the BiCMOS process, and Fig. 1(b) is a schematic representation of the devices in the CMOS-only version. The FET's are identical between the two processes. Note that the resistors are formed differently in the two technologies. The contacts on the end of the boron-diffused resistors are formed by the same implant that provides the PFET source/drain, and contacts to the arsenic resistor are identical to that of the NFET. Junction depths for these 0.45-µm devices is...
approximately 0.30 μm. After the junctions are implanted and driven by a heat cycle, approximately 50 nm of titanium is uniformly deposited across the wafer and reacted in a nitrogen atmosphere to form titanium nitride over oxide regions, and titanium disilicide over the exposed silicon and polysilicon regions. After stripping the titanium nitride, high-resistivity (60-80 μΩ·cm) C49 phase titanium silicide remains over the exposed silicon regions. To obtain the maximum benefit in circuit performance, it is necessary to convert this silicide to the low-resistivity (12-18 μΩ·cm) C54 phase.

In a standard sequence a polysilicon etch stop layer is deposited. After phosphorus-silicate glass (PSG) is deposited and planarized, a two-step etch is performed to define the regions of the local interconnect. The polysilicon then “disappears” through a high-pressure oxidation step performed at 750°C. The etched holes are filled with tungsten and the local interconnection is completed by chemical-mechanical polishing [19]. The heat applied during this sequence converts the titanium silicide to the C54 phase.

For experimental cells, the same local interconnect was accomplished by using plasma-enhanced chemical-vapor-deposition (PECVD) nitride and an optimized etch process that achieved the necessary selectivity without the polysilicon etch stop. As virtually no heat is applied in this process, an anneal is performed prior to PECVD nitride deposition to accomplish the conversion of the silicide from the C49 phase to the C54 phase. Various manifestations of this anneal are described: a 750°C furnace anneal and rapid thermal processes varying from 800°C to 950°C. The physical structure of the local interconnect is shown schematically in Fig. 1, where the layer labeled as “etch stop” may be interpreted as either the oxidized polysilicon or the PECVD nitride.

Fig. 1. Final device cross sections. (a) BiCMOS technology. (b) CMOS-only technology.

III. DEVICE CHARACTERISTICS

The device series resistance is extracted by measuring the currents in individual devices of different lengths at various gate potentials; the series resistance is the total device resistance remaining when the channel length is extrapolated to zero [20]. In addition to the titanium silicide/silicon interfacial resistance, the lightly doped drain contributes to the series resistance. In the NFET, a shallow As “shunt” implant mitigates this effect without compromising the hot electron sensitivity. The PFET has no such implant, and the series resistance is a fairly strong function of the spacer width. Based on these considerations, there is a minimum series resistance due to the device design, and only that portion associated with the TiSi2/Si interface is addressed in this paper.

A similar method of extraction is applied to obtain the resistor “end resistance.” A resistor is formed in a manner analogous to the FET, except that a substantial boron (or arsenic) implant provides conductivity in the absence of gate drive. There is oxide and polysilicon over the resistor to prevent silicidation of the resistor region. The polysilicon dimension determines the resistor length and thick oxide determines the resistor width. The resistance of several devices of various lengths is measured, and the resistance remaining for (extrapolated) zero poly dimension is termed the “end resistance.” As in the FET’s above, there will be a nonzero contribution to the end resistance from the region beneath the spacer, but we concern ourselves here only with that portion attributable to the TiSi2/Si interface.

Emitter resistance is determined by holding the collector current at zero and forcing two different currents into the base contact [21]. The emitter resistance is defined as the change of
the collector–emitter voltage with respect to the emitter–base current. By measuring the potential at the collector the base resistance is eliminated, and only the resistance from the emitter–base junction to the tungsten contact is included in the measurement.

Histograms of the key device parameters are given in Figs. 2–5, for two different processes. The standard (STD) process is as described above, with the etch-stop polysilicon. The optimum (OPT) process utilizes a rapid thermal anneal of 850°C to accomplish the phase conversion and the nitride etch-stop process to form the local interconnect. The simultaneous improvement in all of the parameters is evident. The mean value of the NFET series resistance is reduced from 620 to 460 \( \Omega \cdot \mu m \), and the PFET series resistance is reduced from 1950 to 1750 \( \Omega \cdot \mu m \). The emitter and end resistances are not only nearly halved, but the variation has also been greatly reduced.

In the following sections these device characteristics are shown as a function of a variety of processing variables: junction implant dose, time and temperature of the heat applied after silicide formation, and titanium silicide thickness. The effect of these processes on the resistivity of narrow polysilicon lines is also investigated.

IV. ELECTRICAL BEHAVIOR OF THE DEGRADED CONTACT

A detailed look at the device characteristics indicates that a junction subjected to heat after the formation of the titanium silicide demonstrates diodic behavior. Fig. 6(a) shows the resistance of a short (0.5-\( \mu m \)) boron-diffused resistor as a function of applied bias, both before and after the standard local interconnect process (STD). A dramatic increase in resistance at low applied bias is apparent for the device subjected to heat. The resistance drops monotonically as the bias is increased to a moderate applied voltage. At high biases, local heating and velocity saturation cause the resistance to increase. For this short device, the channel itself contributes only about one-third of the total resistance; the remainder is associated with the end resistance. Arsenic-doped resistors demonstrate a similar nonohmic behavior (Fig. 6(b)).

Fig. 7 shows the resistance as a function of applied bias of a short boron resistor taken at room temperature and at 100°C. The resistance at a very low bias decreases as the temperature increases. At a moderate bias, the resistance increases as the temperature is increased. The behavior at moderate biases is attributed to the decrease of carrier mobility at increasing temperature [22]. The decrease of resistance at low applied biases is attributed to the diodic nature of the contact. The boron resistor has a "channel" resistance of 40 \( \Omega \), and a total end resistance of no less than 80 \( \Omega \). In the resistor, therefore,
Current in the emitter flows through several distinct regions: the polysilicon/monosilicon interface, the polysilicon emitter, the polysilicon/silicide interface, and finally, through the contact stud to the metal. The emitter resistance is a function of the impedance of each of these regions. It is well known that emitter resistance is increased by the presence of oxygen at the polysilicon/monosilicon interface [23], and it can surely be perturbed by errors in processing at the contact etch. However, the processing at the polysilicon emitter deposition and the contact etch were the same for both populations shown in Fig. 5, so these variables are not relevant in this case. In terms of the silicide/polysilicon interface, the doping in the polysilicon is similar to that of the NFET source/drain, and similar behavior may be anticipated.

V. IMPLANT DOSE EFFECTS

As contact resistance is an exponential function of the doping at the TiSi2/Si interface [24], it may be supposed that the dopant loss by diffusion into the silicide might be offset by an increase in the implanted dose. This was indeed found to be the case for As junctions. Fig. 8 shows resistance–current curves for short resistors with three different implant doses in the contact region. The increase of the implant dose from \(2 \times 10^{15}\) to \(5 \times 10^{15}\) has suppressed the diodic behavior for a given heat cycle, and a further dose increase to \(5.75 \times 10^{15}\) has virtually eliminated the negative differential resistance behavior. Other considerations (such as silicide quality and encroachment of the source/drain into the LDD region of the NFET) limit the dose that may be used in a practical application.

In contrast to the above case, no such advantage was found for the boron-doped contacts. Fig. 9 shows the boron-resistor end effect plotted against the PFET series resistance for a given chip. The variation in each parameter was accomplished by varying the implant dose. A reduction in the resistances is indeed found as a function of the dose, but the mechanism is not a mitigation of the diodic nature of the contact. A comparison of the change in resistor end effect with PFET series resistance reveals that the reduction occurs by an increase in the doping level under the spacer and not a reduction in the interfacial contact resistance. In the PFET, the sheet resistance of the region under the spacer is approximately 4500 \(\Omega\)/square, while that of the boron resistor is 450 \(\Omega\)/square. As the doping in this region is increased by outdiffusion from the source/drain implant, it would be expected that the series resistance would decrease at a rate 10 times that of the end resistance, while the contact resistance in the two devices should decrease at exactly the same rate. The data in Fig. 9 support the former hypothesis.

The different reaction of each of the types of contacts is likely due to the ability of the silicide to absorb the additional dopant and for the metal-dopant compounds TiB2 and TiAs. Hiu et al. [15] show a greater dependence of the specific contact resistivity on the implant dose for boron as opposed to arsenic. For the same heat cycle and implant dose, the resultant surface concentration of boron under TiSi2 is much less than that of arsenic [16], [18]. For the same processing, boron is
more readily drawn out from the silicon and bound up in a metal-dopant compound than is As.

VI. TIME AND TEMPERATURE EFFECTS

Utilizing the PECVD nitride etch-stop process to minimize the additional heat, experiments were run with various anneals to convert the silicide from C49 to C54 phase.

In Fig. 10 we see the effect of isochronal rapid thermal anneals. As the temperature is increased, the end resistance increases monotonically, and the spread in the data also increases. The absolute minimum temperature necessary to convert the silicide is that which should be applied to obtain the optimum process.

Fig. 11 illustrates the effect of increased time at temperature on the junction characteristics. At both 875°C and 900°C there is a profound difference on the mean value and spread in the resistance between the short and long anneals. The minimum time necessary to convert the silicide at a given temperature is,
that which should be applied to obtain the optimum process.

We have included data for a conventional furnace anneal at 750°C in both Figs. 10 and 11. This anneal, irrespective of whether it is acceptable from the perspective of converting the silicide to the C54 phase, allows substantial dopant diffusion and depletes the junction, even though it is performed at a much lower temperature than the RTA processes. Also, the spread in the data is very large relative to the rapid-thermal-anneal processes.

VII. EFFECT OF TiSi2 VARIATIONS

A further complication is introduced when it is recognized that both the minimum heat cycle necessary for conversion and the maximum temperature sustainable by the TiSi2/Si interface, are dependent on the titanium silicide formation process. Slight deviations in the silicide formation temperature and surface conditions can produce different thicknesses of resultant titanium silicide and alter the phase transformation kinetics [25], [26].

The results of Fig. 10 are again presented in Fig. 12, as well as another group of data taken from anneals with 800°C to 850°C for the same time. The Group A silicide sheet resistance was 1.85 Ω/square, and 2.25 Ω/square for Group B. There was no intentional deviation introduced into the process to produce the different silicides, but for the same time and temperature, the devices from Group A exhibit a greater degree of junction degradation than the Group B devices.

One interpretation is that the decrease in sheet resistance implies a thicker layer of titanium disilicide; there is deeper penetration of silicide into the diffused junction and the silicide/silicon interface forms at a point of reduced doping, increasing the impedance of the interface. An alternative interpretation might postulate that the diffusivity of the boron into the silicide is affected by the same process deviation that affected the sheet resistance, such as the ratio of C49 phase to C54 phase silicide. In any case, these data illustrate the causal link between the silicide characteristics and the interfacial impedance.

VIII. PROCESS OPTIMIZATION

The task of process optimization is to identify the combination of time and temperature which results in acceptable silicide characteristics and minimized impact on the device resistances. The criterion for successful conversion is stringent: the average resistivity of a long (200,000 μm) narrow (0.5 μm) serpentine line must be less than 4 Ω/square. Inadequate time at temperature can result in less-than-complete transformation; excessive time at temperature can lead to silicide agglomeration and a large effective resistance. Fig. 13 summarizes end-resistance data of the previous figures, accompanied by data from the long narrow polysilicon serpentine described above. To represent data for different silicide thicknesses, the data have been presented as a multiplicative factor. Unity was chosen to represent the lowest observed value, and only data from the short anneal time are shown. These data indicate a monotonic increase in end resistance with increasing temperature, and a monotonic decrease in polysilicon sheet resistance with increasing temperature. The optimum conversion process is a very brief time at 850°C. Both critical parameters suffer some degradation from their ideal values at this point, but neither is excessively compromised.

IX. CONCLUSIONS

It has been observed that a significant contribution to device resistance arises from the barrier at the TiSi2/Si interface. This effect has been observed on FET's of both types, NPN's, and both n- and p-type diffused resistors. This contact resistance is nonohmic in nature if substantial heat is applied to the devices after silicide formation. Heat may be required at this point for a variety of reasons: to achieve the low-resistivity C54 phase of TiSi2, for a local interconnect process as described above, or for reflowing oxide for insulator planarization. A heat cycle as low as 750°C produced substantial degradation.

Rapid thermal processing can achieve the requirements of the silicide conversion and minimize the adverse impact on the device characteristics, although the particulars of the required process will depend on the specifics of the junction design and
the silicidation process. For this technology, 850°C for a few seconds was found to be optimum to balance the requirements of device performance and silicide integrity.

REFERENCES