The Challenges Of Device Scaling

Mat. Res. Soc. Symp. Vol. 610 © 2000 Materials Research Society

JUNCTIONS FOR DEEP SUB-100 NM MOS: HOW FAR WILL ION IMPLANTATION TAKE US?

*H.-J. Gossmann, C. S. Rafferty, and P. Keys**
Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974

*University of Florida, Gainesville, FL 32611

ABSTRACT

We analyze the requirements that the International Technology Roadmap for Semiconductors (ITRS) implicitly imposes on the two-dimensional source/drain (SD) dopant profile and translate the results into implant parameters (energy, dose, peak concentration). We do this by determining the voltage drop that the SD current develops across the three main (exclusive of the channel) resistive components in the current path: the spreading resistance in the extension region; the metal-semiconductor contact; and the resistance in the link-up region where the SD-region meets the channel. The largest resistance occurs in the link-up region, followed by the resistance of the contact; the extension contribution is the smallest. The extension resistance requirement can be satisfied by ion-implantation for all generations of the ITRS. The link-up region requires very abrupt lateral profiles, not demonstrated so far by ion-implantation. It is found that such resistance cannot be reduced without impacting the intrinsic device behavior. The contact eventually necessitates dopant concentrations in excess of solid solubility and for NMOS in excess of the fundamental limit of dopant activation.

INTRODUCTION

Producing the junctions for transistors with sub-100 nm gate-lengths is a significant challenge. For many reasons, such as uniformity, reproducibility, and cleanliness, ion implantation is presently the method of choice for the formation of the source/drain (SD) junctions and the channel. On the other hand, ion implantation creates defects, giving rise to a host of undesirable effects, such as dopant clustering and transient enhanced diffusion. A clear-cut path to a solution that satisfies all constraints is by no means obvious.

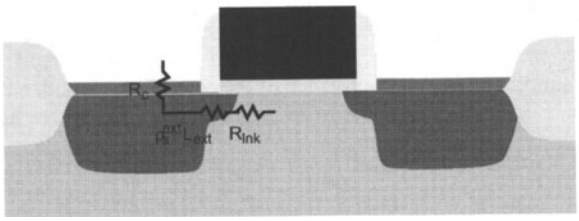

Figure 1. Cross-section of MOSFET (schematic); the three series-resistance components, contact resistance, R_c, extension resistance, $\rho_s^{ext} L_{ext}$, and link-up- or tip- resistance, R_{lnk}, are shown.

In this paper we analyze the requirements that the 1999 edition of the International Technology Roadmap for Semiconductors[1] (ITRS) implicitly imposes on the two-dimensional SD

dopant profile and translate the results into implant parameters (energy, dose, peak concentration). There are three main (exclusive of the channel) resistive components in the current path: (1) the spreading resistance in the extension region, R_{ext}; the metal-semiconductor contact, R_c; and the resistance in the link-up- or tip- region where the SD-region meets the channel, R_{lnk}. For each component we determine the dopant profile required to keep the voltage drop across that component on both sides of the transistor at 5% of the supply voltage, V_{DD}. As specified in the Roadmap, the SD current is taken as 0.75 mA/μm for NMOS and 0.35 mA/μm for PMOS, independent of the technology node. Other required parameters, such as sidewall length, junction depth or supply voltage, are taken from the Roadmap[1]. Since the Roadmap often only specifies ranges, all possible combinations of the upper and lower bounds are considered.

THE EXTENSION (SHALLOW JUNCTION)

Figure 2 shows the depth of the shallow junction and the length of the extension, as given in the Roadmap[1].

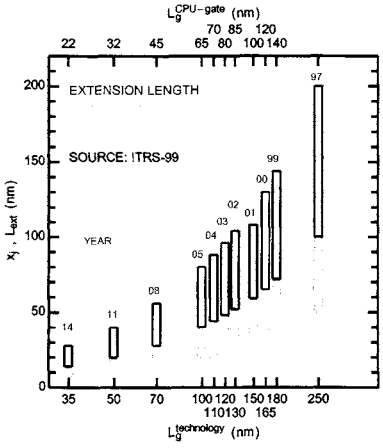

Figure 2. Junction depth and extension length, as a function of technology node (bottom abscissa) and CPU gate-length (top abscissa). Also shown is the year of introduction.

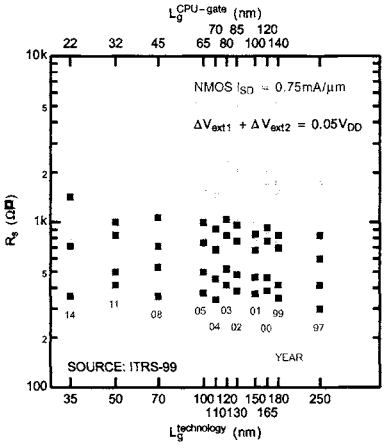

Figure 3. Required extension sheet resistance.

For an abrupt dopant profile the total resistance of the extensions on both sides of the transistor is

$$R = \rho^{ext} \frac{L_{ext}}{W x_j^{ext}} = \rho_s^{ext} \frac{L_{ext}}{W} \quad , \qquad (1)$$

where ρ^{ext} and ρ_s^{ext} are the resistivity and the sheet-resistivity of the extension, respectively, L_{ext} is the extension length, W the width of the device, and x_j^{ext} the extension depth. The required sheet-resistivity that keeps the voltage drop across the extension on both sides of the transistor at 5% of the supply voltage is then given by

$$\rho_s^{ext} = \frac{0.05}{2} V_{DD} \frac{W}{I_{ON}} \frac{1}{L_{ext}} , \qquad (2)$$

where I_{ON} is the on-current. It is plotted in Fig. 3. Using the implant simulator IMSIL[2] the corresponding implant energy of B and P, as well as the dose was determined for PMOS and NMOS, respectively. The dose is shown in Fig. 4.

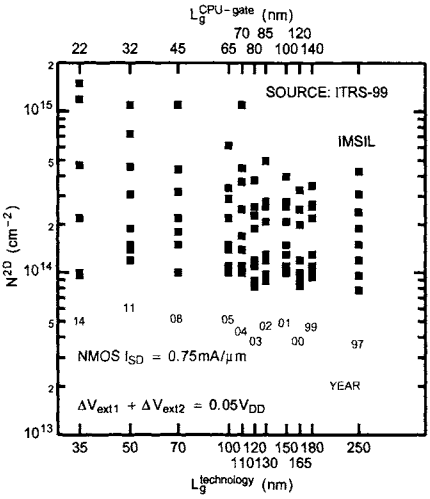

Figure 4. Required extension implant dose.

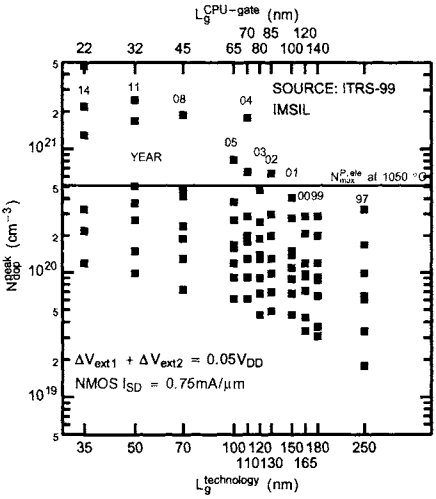

Figure 5. Required peak concentration in the extension.

A given dose and energy corresponds to a certain peak volume concentration, shown in Fig. 5, together with the maximum electrical activity[3] [4] in thermal equilibrium of B and P at 1050 °C. The large majority of calculated points lies below the appropriate maximum electrically active dopant concentration. Nevertheless, some points exceed it, most notably for NMOS; they correspond to the most "demanding" application for a given technology node, i.e. the shallowest junction and smallest supply voltage, but also the longest extension. The required implant energies, assuming elemental B implant, is shown in Fig. 6. Note that those represent really upper bounds, since the calculation did not include any diffusion. It is quite obvious that ultra-low-energy (ULE) implants are a necessity.

Sheet resistivity is inversely related to junction depth, and an ideal curve for a perfect junction can be calculated based on the equilibrium solubility of the dopant and a box-shaped junction profile. Experimentally obtained resistivities are surprisingly close to the ideal targets, given that real profiles suffer from implantation straggle and transient enhanced diffusion (Fig 7). The data is from ULE implants of elemental B, annealed mostly with fast ramps up to 1050 - 1100 °C without dwell time ("spike anneal").[5] [6] [7] [8] In addition to the experimental and ideal data, there are sheet-resistance / junction depth points calculated from the 5% criterion. As long as there are data points (squares) below a calculated point (circle) it implies that an experimental demonstration of that particular extension exists. This is true for almost all Roadmap transistors. Thus it is concluded that implanted junctions will suffice to form the

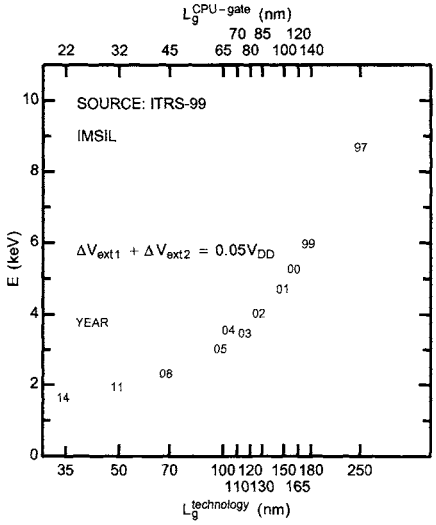

Figure 6. Required extension implant energy (PMOS, elemental B).

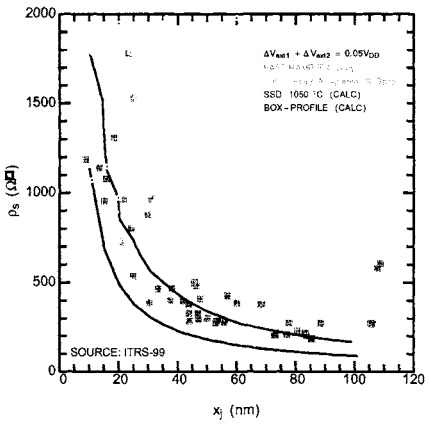

Figure 7. Comparison of actual data (squares) with the Roadmap requirements (circles); also shown are two idealized cases, solid source diffusion at 1050 °C and a box profile with concentration equal to the solid solubility at 1050 °C, both for B.

extension for the foreseeable future.

CONTACT RESISTANCE

Here we are only concerned with the metal-to-semiconductor contact; the metal bulk and metal-to-metal resistances of a plug, interfaces between barrier- and cladding-layers, and lines are considered small relative to the semiconductor and semiconductor-metal resistance. The geometry of the contact is shown in Fig. 8; the length of the metalization is denoted by L_C, that of the window by L_{W1}. For a non-silicided SD $L_C = L_{W1}$; in the case of silicided SD L_C may be significantly larger than L_{W1}. The contact resistance, R_c, depends[9] on the contact resistivity, ρ_c, and the sheet resistance of the semiconductor, ρ_s as

$$R_c = \frac{\sqrt{\rho_c \rho_s}}{W} \coth\left[L_C \sqrt{\frac{\rho_s}{\rho_c}}\right]. \quad (3)$$

Only if the transfer-length $\sqrt{\rho_c/\rho_s} \gg L_c$ is the contact resistance inversely proportional to the contact area

$$R_c = \frac{\rho_c}{L_c W}, \quad (4)$$

where W is the width of the contact. Note that this is not necessarily the case for all Roadmap transistors, and that assumption is not made in the subsequent calculations. The contact

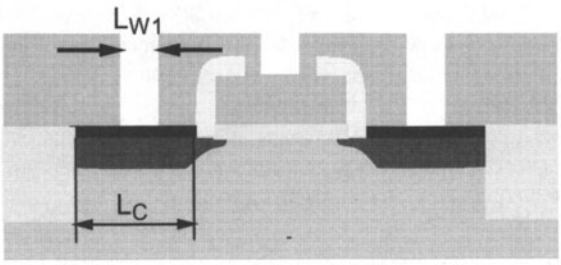

Figure 8. Cross-section of MOSFET (schematic) illustrating the contact geometry.

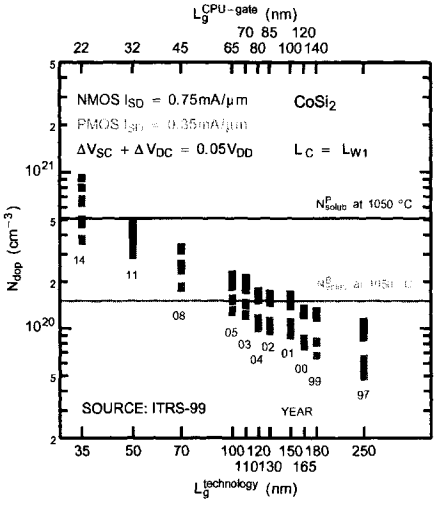

Figure 9. Required doping concentration in the contact.

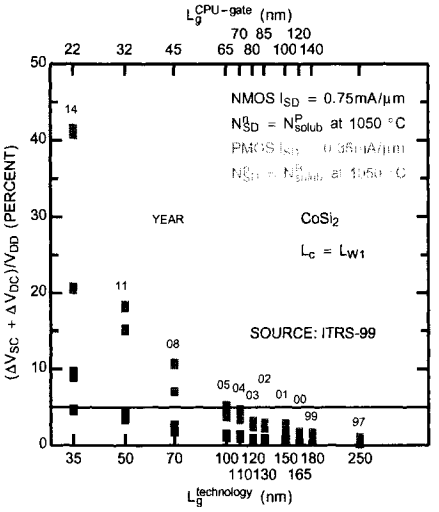

Figure 10. Voltage drop across source and drain contacts if the doping concentration is given by the maximum electrically active concentration at 1050 °C.

resistivity can be calculated analytically following Yu.[10] Depending on the semiconductor doping concentration, carrier transport occurs via thermionic emission over the barrier, field emission at the Fermi level, or a combination of the two. The appropriate equations and Eqn. 3 have been inverted numerically to yield a doping concentration that results in 5% V_{DD} voltage drop across both contacts. The result is shown in Fig. 9, for $L_C = L_{W1}$ and $CoSi_2$ as metalization, together with the maximum electrical activity[3][4] of B and P at 1050 °C. While NMOS

fulfills the 5% requirement at every node except 35 nm, PMOS violates it already at the 100nm node. Figure 10 shows the voltage drop expected across both contacts if the doping concentration is given by the maximum electrically active concentration at 1050 °C. In the most extreme case, 40% of V_{DD} will drop across the SD contacts.

To avoid significant contact resistance, epitaxial growth or annealing techniques (e.g. laser annealing), which increase activation at the silicide interface above equilibrium will be needed soon. Integration will not be easy, as meta-stable activation must be maintained through subsequent thermal processing.

LINK-UP (TIP) RESISTANCE

Under the gate, there is a region where the extension doping falls from its peak in the mid 10^{20} cm^{-3}, set by solubility limits, to a point where the carriers supplied by the inversion layer exceed those supplied by the doping, at which point the intrinsic channel can be considered to begin. (This point is well outside the metallurgical junction.[11]) Although this region, called the link-up region, is short, its relatively low doping can lead to a high resistance.

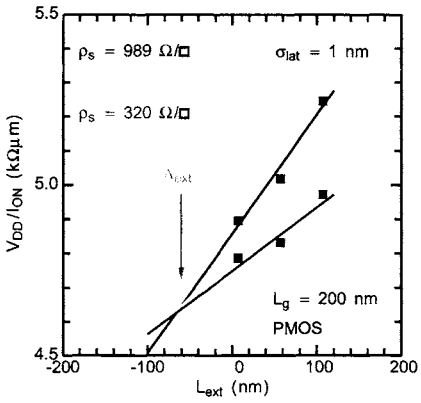

Figure 11. Transistor resistance, V_{DD}/I_{ON}, for V_{DD} = 1.5 V as a function of extension length, with the extension sheet resistance as parameter, for an extension lateral dopant profile variance of σ_{lat} = 1 nm.

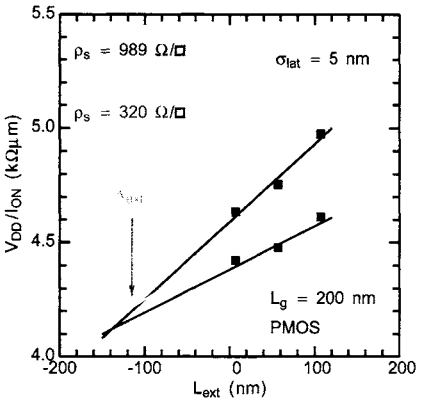

Figure 12. Transistor resistance, V_{DD}/I_{ON}, for V_{DD} = 1.5 V as a function of extension length, with the extension sheet resistance as parameter, for an extension lateral dopant profile variance of σ_{lat} = 5 nm.

This resistance is inherently two-dimensional in nature and is not easily estimated analytically as were the other two components. To make a first estimate of its significance, a device simulation is performed of an 200nm PMOS device with uniformly doped channel at concentration 7×10^{17} cm^{-3}, gate oxide thickness 2.3 nm, and junctions that are abrupt in the vertical direction (depth 15 nm, concentration 2×10^{20} cm^{-3}, sheet resistivity 539 Ω/□) and have a lateral Gaussian straggle of 1nm. The device has a gate overlap of 7 nm, i.e. the dopant concentration of the junction is constant in the lateral direction to a point 7 nm underneath the gate, where it begins to roll off. In a device simulator, it is possible to prolong the contacts down to a

"zero-extension-length", meaning that the contact equi-potential extends to the point where the doping first begins to drop. The total "transistor resistance", V_{DD}/I_{ON}, for $V_{DD} = 1.5$ V is seen to fall linearly as the extension length is reduced to zero (Fig. 11). If the same calculation is carried out using different junction concentrations (and corresponding sheet resistivities), the extrapolated "device resistance" lines meet not at zero, but at a negative intercept of about 60 nm (Fig. 11). The interpretation is that the link-up region is providing a resistance equivalent to about a 60 nm long extension. This "extra extension length" Λ_{ext} is a function of the lateral doping slope; increasing the lateral gaussian straggle to 5 nm increases Λ_{ext} to about 120 nm (Fig. 12). This is to be expected since the resistance arises from the gradual decrease of doping from extension to channel; the more abrupt the less the influence of the transition layer.

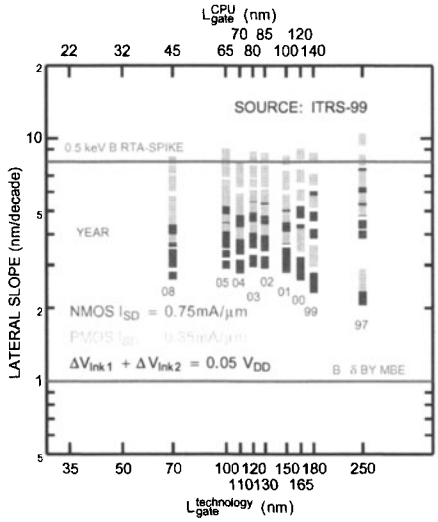

Figure 13. Required lateral dopant profile slope.

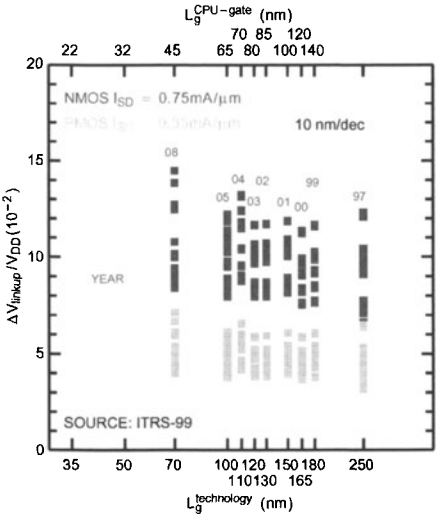

Figure 14. Voltage drop due to the link-up resistance for a lateral dopant profile slope of 10 nm/dec

An analytical calculation of the link-up resistance[9] provides an estimate of the doping steepness required to meet a 5%V_{DD} criterion. Again using Roadmap junction depths, oxide thicknesses and power supply voltages, this estimate is shown in Fig. 13, together with two experimental values from a 0.5 keV, 2×10^{14} cm^{-2} B implant,[12] spike-annealed to 1050 °C, and a B-δ marker layer, grown by MBE.[13] The steepness experimentally obtained from a manufacturable technology (ion implantation) invariably fails to meet the criterion, suggesting that devices are already significantly limited in their drive due to link-up resistance. Figure 14 indicates that 10% V_{DD} is presently lost to link-up resistance assuming a "typical-best" lateral steepness of 10nm/decade. These results suggest pursuing an experimental program to find techniques which introduce ultra-steep junctions, for instance by epitaxy. However, unlike contact resistance, the extension profile cannot be modified without impacting the intrinsic device.

Preliminary results show that any gain in sheet resistance is offset by increased leakage, due to carrier spilling into the channel. A longer channel is then required to offset the increased leakage, leading to reduced on-current, cancelling the improvement in drive obtained by reducing sheet resistance. Figure 15 shows the trade-off between I_{ON} and I_{OFF} in a super-halo 50nm device design known as the "well-tempered MOSFET".[14]

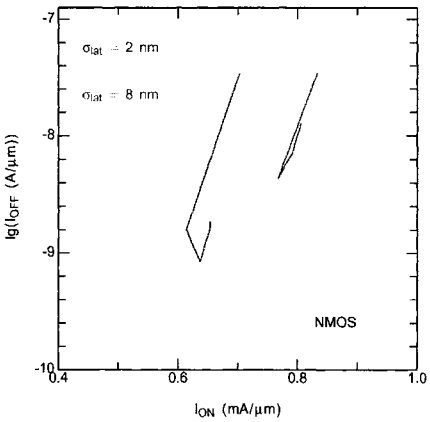

Figure 15. Simulated off-current as a function of on-current. Simulated transistors differ in lateral dopant profile variance.

The only modification with respect to ref. [14] is the oxide thickness (1.5 nm was used) and the operating voltage (V_{DD} = 0.9 V.) Each point on each curve corresponds to a device simulation of a different channel length. Each curve corresponds to a different lateral doping slope. Only the lateral slope, not the vertical slope is varied. For any target off-current, the lowest on-current arises for the device with steepest profile. The off-current loss due to increased short-channel effects is apparently more significant than the reduction in resistivity. This is similar to the factors governing choice of extension depth, where short-channel effects are the first consideration and sheet resistivity a secondary consideration. To be sure, a steeper profile allows target drive current to be reached in a shorter device, and has better packing density, shorter transit time, reduced capacitance, and flatter Vt-rolloff curve. The optimal lateral steepness is therefore a function of many technology variables and the discussion is outside the scope of this article. However it should be clear that sheet resistance is far from the only consideration and that steeper is not automatically better.

SUMMARY

We have investigated the series resistance of a MOSFET within the bounds of the ITRS Roadmap. For each of the three components of the series resistance, the extension, R_{ext}, the contact, R_c, and the link-up region, R_{lnk}, we calculated the dopant profile required to drop 5%V_{DD} total across both sides.

- The requirements for the extension implant can be fulfilled with current state-of-the-art technology.