Proximity Effect Modeling

Colin McAndrew
Freescale Semiconductor
Tempe, AZ

Abstract

The book (*Operation and Modeling of the MOS Transistor, 3rd Edition*, by Yannis Tsividis and Colin McAndrew, Oxford University Press, 2011) discusses proximity effects (well proximity and stress effects) in Chap. 9. The models of these effects are highly empirical and are constantly changing, so no specific model forms were provided (see p. 581, Chap. 9); some details of presently used proximity effect models are provided here.

Well Proximity Effect Model

The well proximity effect (WPE) can significantly alter the characteristics of MOS transistors [1][2], and if not factored into design can cause circuits to fail. The cause of the effect and how it modifies the structure of a device are shown in Fig. 1 and Fig. 2, respectively. During implantation of the dopants that form the wells in which MOS transistors are formed, photoresist is patterned outside of these regions to prevent the ions penetrating into regions where they are meant to be excluded. However, ions that impinge on photoresist interact with the atoms that form the photoresist. These interactions are statistical, and "scatter" the ions at random angles. Multiple scattering events drain the ions of their kinetic energy, so they stop within the photoresist (which is subsequently stripped from the silicon wafer surface). For ions that impinge near the edge of the photoresist, they can scatter back out of the photoresist, and then become embedded in the silicon adjacent to the photoresist edge. Well doping is therefore not uniform across a wafer, but is enhanced in regions adjacent to the edges of a well, to a distance of about 1 μm [1][2]. This is shown diagrammatically in Fig. 2. Note that ion implantation is often done at a small angle to vertical to a wafer (as shown in Fig. 1), to prevent ions from avoiding scattering and penetrating too deeply by "channeling" along open paths in the silicon lattice. This may give the impression that the well doping enhancement is from ions that are "reflected" from the photoresist edge. This is incorrect; the scattering occurs within the photoresist and is observed even if the implantation angle is vertical.

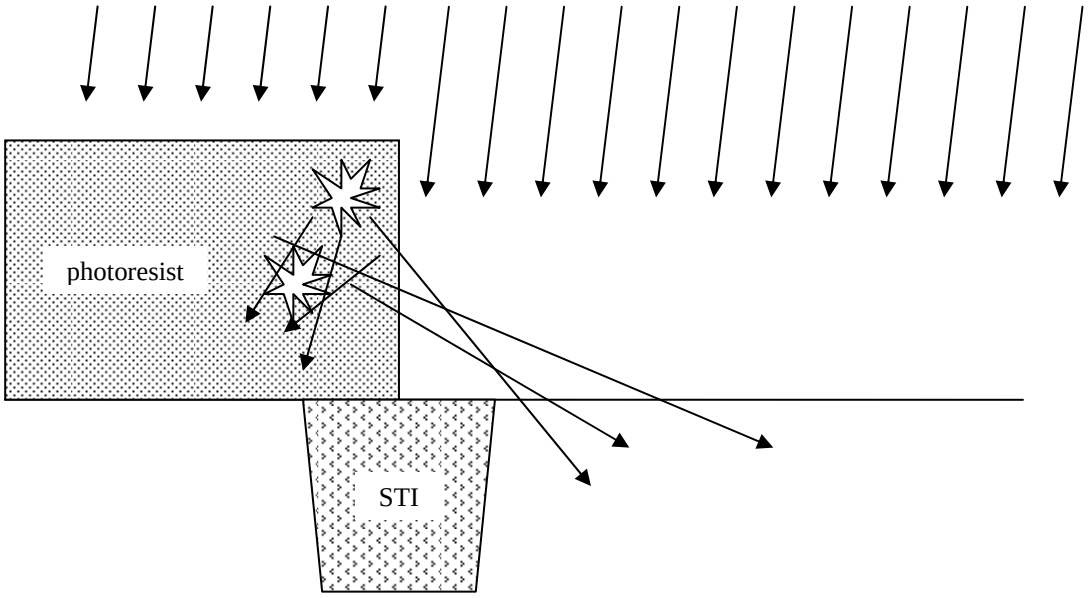

Fig. 1 Well proximity effect cause. Scattering in the photoresist augments well doping near the well edge.

As is discussed at length in the book, the net bulk doping concentration is one of the key physical parameters than controls the operation of MOS transistors. Therefore it should be apparent that the proximity of a transistor to a well edge, which changes its bulk doping, can affect its behavior significantly. Prior to about the 0.25 μm technology node the overall size of transistors meant they could not be placed close enough to a well edge for the enhanced doping level to affect their behavior. For technology nodes below that WPE has become an important issue.

There are two aspects to understanding and modeling WPE: the adjacency of a transistor to well edges, which affects the overall enhancement of the net doping in a device; and the effect of the net doping enhancement on transistor electrical behavior, via model parameters. Fig. 3 shows a MOS transistor in a simple rectangular well. The notation SC is for the spacing of a well edge to the edge of the gate, and there are four different spacings, $SC1$ through $SC4$ shown in Fig. 3. To integrate the effect of the additional well dose from each adjacent edge on every model parameter is a daunting task, and it blurs the line between layout extraction and modeling and simulation, which are generally separate steps in an IC design flow. Therefore the approach taken for WPE modeling has been to separate the steps of extracting effective indicators ("moments") of the additional well doping from the device layout, and of modeling the change in device behavior based on these moments. This enables implementation of WPE across different types of models; for example, for a surface potential-based model the WPE can be handled by making the V_{FB} parameter a function of the moments, and for a threshold voltage-based model V_{T0} can be made a function of the moments.

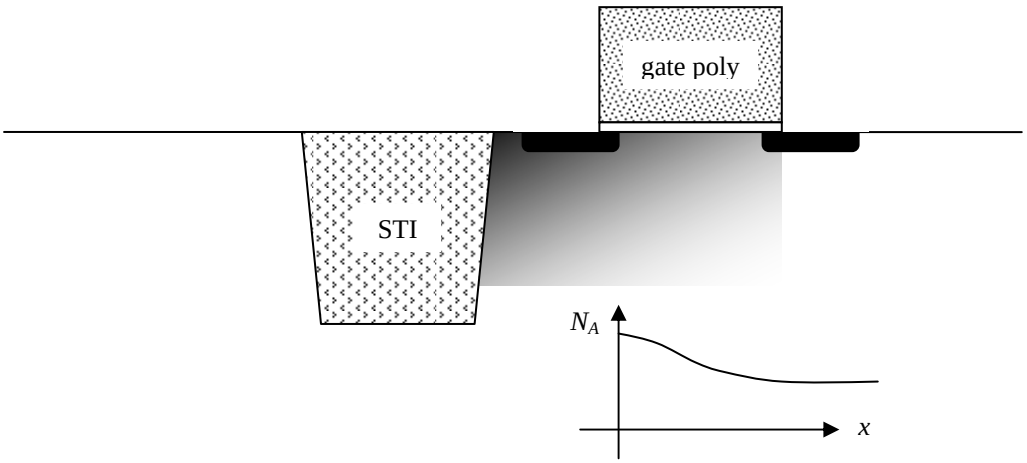

Fig. 2 Well doping varies with distance from the well edge.

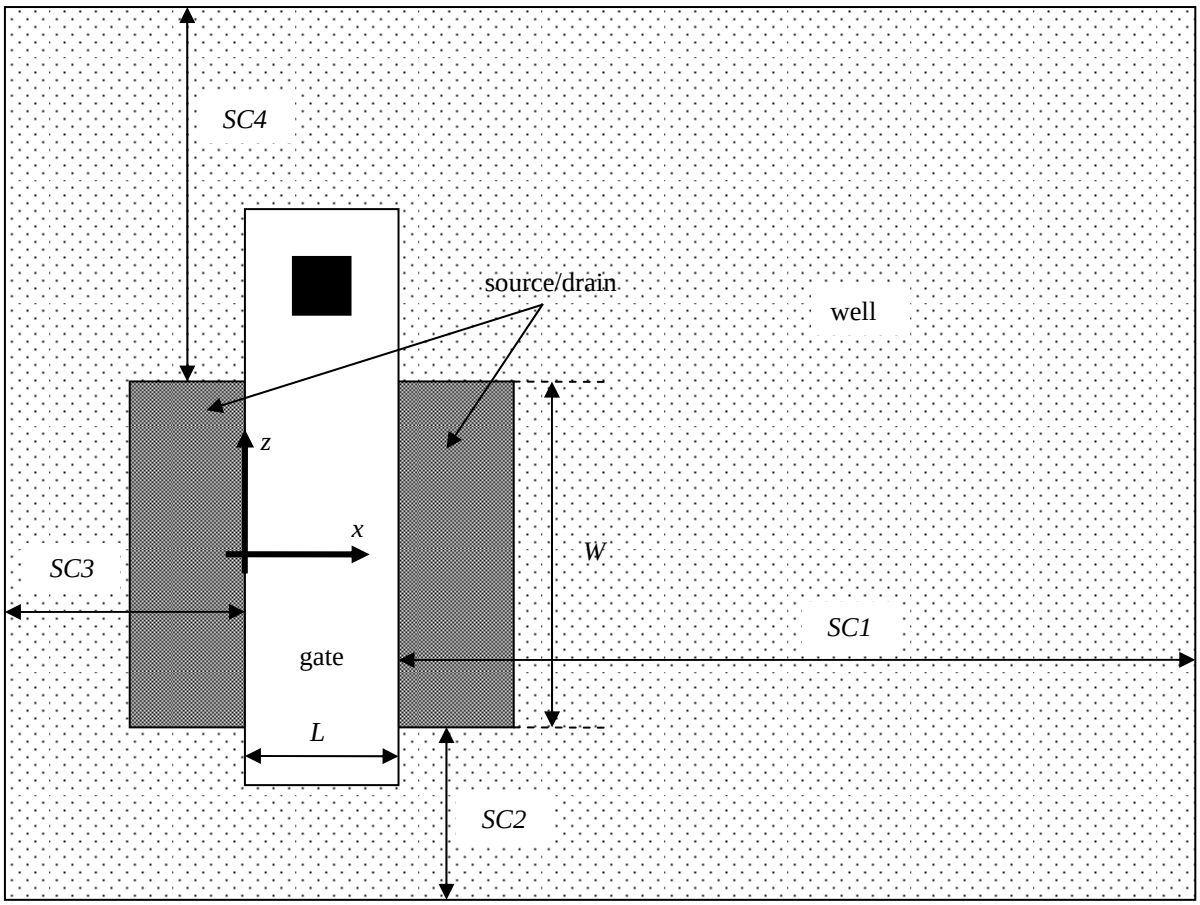

Fig. 3 Layout of a MOS transistor with respect to well edges.

Empirically, if spacing to a well edge is s (measured normal to the edge), and a reference well edge to gate edge length to normalize this distance is SC_{ref}, the following dose basis functions for the moment calculations have been found to provide the best fit to experimental data [3]

$$
\begin{aligned}
f_A(s) &= \left(\frac{SC_{ref}}{s}\right)^2 \\
f_B(s) &= \frac{s}{SC_{ref}} e^{-10s/SC_{ref}} \\
f_C(s) &= \frac{s}{SC_{ref}} e^{-20s/SC_{ref}}
\end{aligned}
\tag{1}
$$

where the first is the observed asymptotic behavior for large s and the other terms provide corrections for smaller s. To implement corrections for the WPE, MOS transistor models are to expected to accept as "instance" parameters three quantities, SCA, SCB, and SCC, that are the averages of the above basis functions. These are computed by integrating the functions over the area of the gate, which gives the total contribution from each dose function, and then normalizing each integral by dividing by the gate area,

$$
\bar{f} = \frac{1}{WL} \int_{-W/2}^{W/2} \int_{0}^{L} f(s(x,z)) dx dz \,. \tag{2}
$$

The normalization is necessary because although the additional dose from the left well edge in Fig. 3 will have certain maximum value, as the gate length of the transistor increases the portion of the transistor affected by this dose decreases. For the purpose of modeling the overall change in device behavior should also decrease, hence the need to divide by the gate area.

There are four edges, right, bottom, left and top, in Fig. 3 that contribute extra doping to the well under the gate of the transistor. The moments of the dose basis functions are then

$$
SCA = \frac{1}{LW}\left(W \int_{SC1}^{SC1+L} f_A(s)\, ds + L \int_{SC2}^{SC2+W} f_A(s)\, ds + W \int_{SC3}^{SC3+L} f_A(s)\, ds + L \int_{SC4}^{SC4+W} f_A(s)\, ds \right) \tag{3}
$$

$$
SCB = \frac{1}{LW}\left(W \int_{SC1}^{SC1+L} f_B(s)\, ds + L \int_{SC2}^{SC2+W} f_B(s)\, ds + W \int_{SC3}^{SC3+L} f_B(s)\, ds + L \int_{SC4}^{SC4+W} f_B(s)\, ds \right) \tag{4}
$$

$$
SCC = \frac{1}{LW}\left(W \int_{SC1}^{SC1+L} f_C(s)\, ds + L \int_{SC2}^{SC2+W} f_C(s)\, ds + W \int_{SC3}^{SC3+L} f_C(s)\, ds + L \int_{SC4}^{SC4+W} f_C(s)\, ds \right) \tag{5}
$$

where the spacing s is measured perpendicular to the well edge. Note that the quantities SCA, SCB, and SCC depend only on the layout geometries of the transistor and the well edges, and are not directly tied to parameters such as threshold voltage or body effect coefficient. They are extracted from the layout and then passed to a MOS transistor model, which is then expected to make appropriate modifications to parameters.

The expressions (3) through (5) are written in integral form to emphasize that they are averages over the gate area; when actually extracted from layout closed form solutions are used. Not all transistors have as simple layouts as shown in Fig. 3, and in practice corner contributions also need to be included. Details of how to compute SCA, SCB, and SCC for complex layouts, including multi-finger transistors, are provided in [4].

The second step in handling the WPE is for a model to use the SCA, SCB, and SCC instance parameters to make appropriate adjustments. Although it would appear that the simplest way of doing this is do modify the parameter associated with N_A this is not done in practice. Experimentally, the threshold voltage, body effect coefficient, and mobility are

observed to vary with well proximity. Rather than tie the first two together through N_A they are kept separate and have separate WPE variation models. This gives greater flexibility in fitting experimental data. The forms used in BSIM4, which have been adopted in other models, are [5]

(6) $$V_{T0} = V_{T0,noWPE} + K_{VT0WE}\left(SCA + W_{EB}SCB + W_{EC}SCC\right)$$

(7) $$\gamma = \gamma_{noWPE} + K_{2WE}\left(SCA + W_{EB}SCB + W_{EC}SCC\right)$$

(8) $$\mu = \mu_{noWPE}\left(1 + K_{U0WE}\left(SCA + W_{EB}SCB + W_{EC}SCC\right)\right)$$

where K_{VT0WE}, K_{2WE}, K_{U0WE}, W_{EB}, and W_{EC} are fitting parameters, and the added subscript "*noWPE*" is used to indicate the value of a parameter in the absence of the WPE. For charge sheet models, the threshold voltage shift (6) is implemented as a shift in V_{FB} which, from (4.7.19), is equivalent.

Fig. 4 shows the threshold voltage shift from the well proximity effect as a function of the well-edge to gate-edge spacing for a 90 nm technology [2]. The amount of the threshold shift reaches several 10's of mV, which can be significant if the supply voltage is of order 1.2 V, and is still a few mV even when the well edge is 1 μm from the edge of the gate.

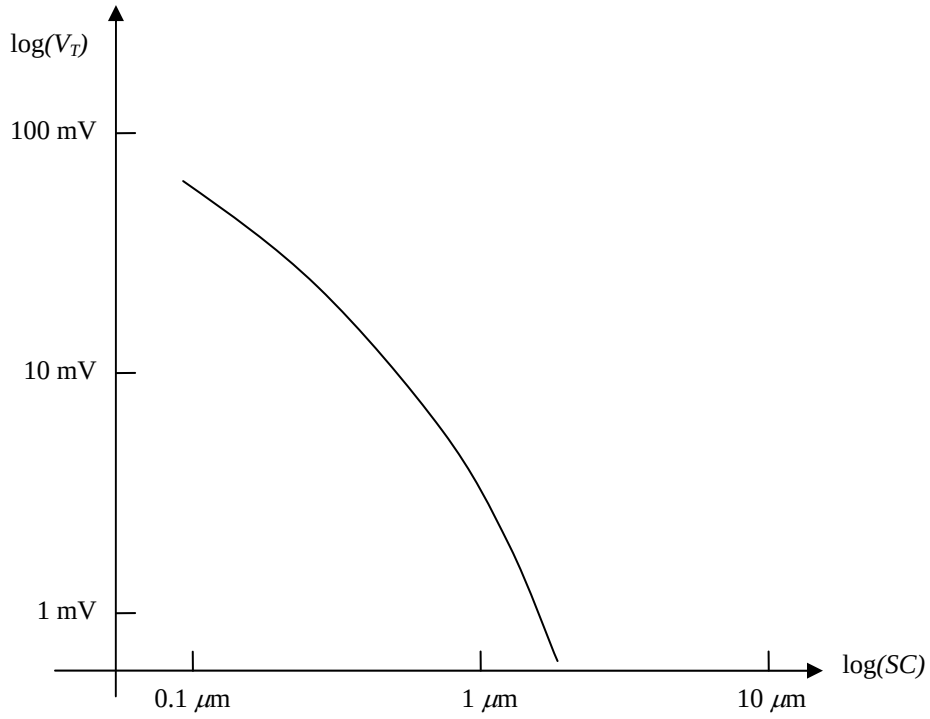

Fig. 4 Threshold voltage shift vs. well spacing.

Shallow Trench Isolation (STI) Stress Effect Model

Shallow trench isolation was introduced at around the 180 nm technology node, and because of the difference in lattice spacing between the silicon semiconductor regions and the SiO_2 isolation regions induces stress.. Although the first-order effect of stress is to alter the band-structure of a semiconductor, which changes the effective mass of the mobile electrons or holes and therefore alters their mobility, it has been experimentally observed that stress also alters the saturation velocity, and because stress affects the spacing of atoms in the lattice structure it changes the solid-state diffusion rates of dopants through the silicon lattice, which causes changes in the final doping profiles in the body after thermal annealing steps. The threshold voltage and body effect coefficient, which depend on effective body doping, can therefore also be indirectly altered by stress effects.

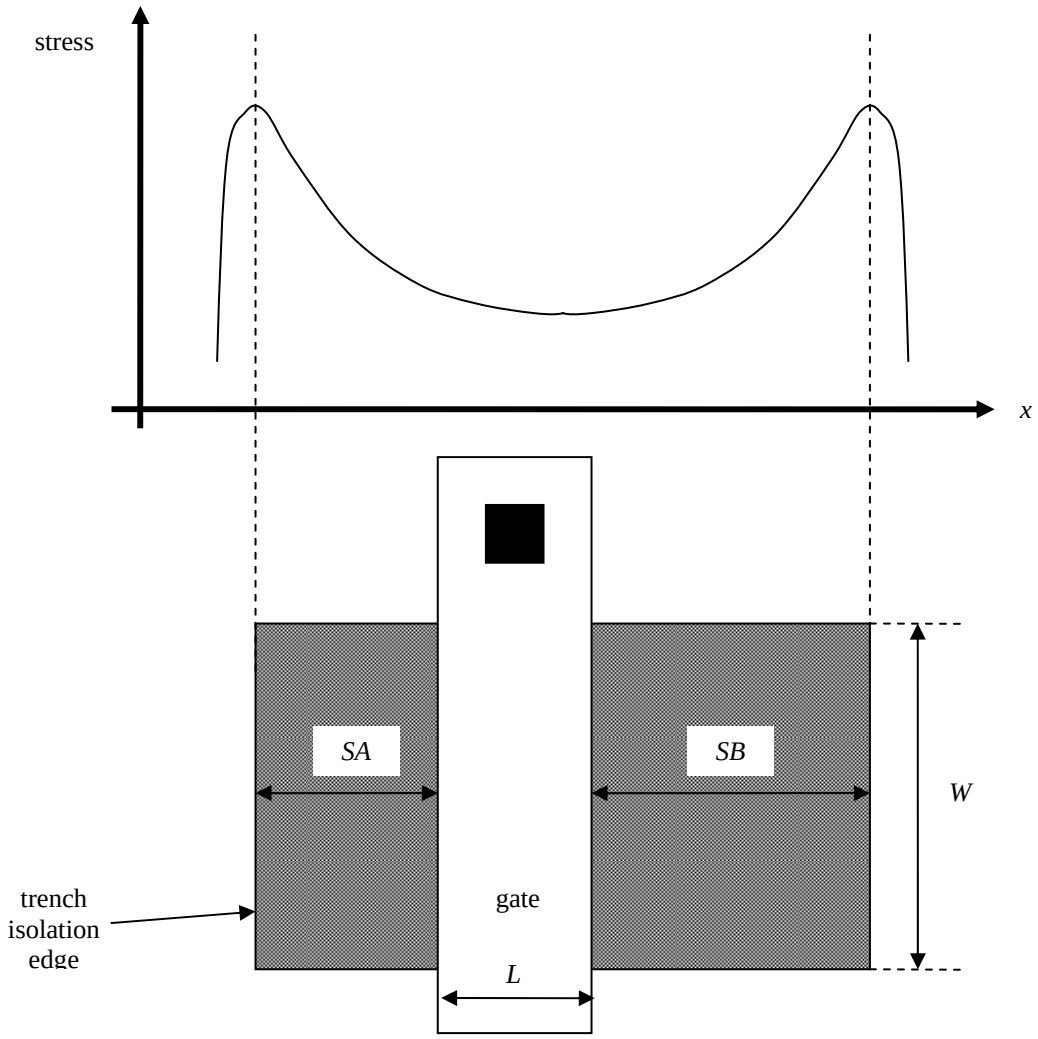

Fig. 5 Layout of a MOS transistor with respect to stress-inducing trench isolation edges.

Originally, shallow trench isolation (STI) was the first source of stress observed to affect transistor performance [6], and it is the source that varies with layout and so has been incorporated into models. Fig. 5 shows a top view of a MOS transistor, and each edge of the shallow trench isolation that defines the boundary of the region in which the transistor is formed is a source of stress, because of the difference in lattice spacings of silicon and the oxide that fills the trench. The stress is observed to decrease approximately inversely with distance from the trench edge [7]. Unlike the well proximity effect, which can be reduced to zero by moving the well edges sufficiently far from a device, STI defines the boundaries of a device in the width direction, and so cannot be completely removed from affecting a device. The effect of stress on a transistor is therefore modeled with respect to a reference layout. As just noted, there is no flexibility in adjusting the STI edges with respect to the width of a transistor, therefore only the STI edges that run parallel to the width dimension of a transistor can be changed in

layout, by varying the spacings from the gate edges to the STI boundary. The layout parameters SA and SB, with reference values SA_{ref} and SB_{ref}, are introduced to characterize these spacings.

Mobility is known to vary with stress, so taking the center of the gate as the position at which to calculate the stress, where the distances to the STI edges are $SA+L/2$ and $SB+L/2$, and assuming to first order that mobility varies linearly with stress, which decreases inversely as the distance from the STI edge, gives

$$\mu = \mu_{ref} \frac{1+K_\mu\left(\frac{1}{SA+L/2}+\frac{1}{SB+L/2}\right)}{1+K_\mu\left(\frac{1}{SA_{ref}+L/2}+\frac{1}{SB_{ref}+L/2}\right)} \tag{9}$$

where μ_{ref} is the mobility for the reference layout and K_μ is the linear proportionality coefficient. This form matches the experimentally observed variation of mobility with STI edge spacing [7], however it does not match the experimentally observed behavior that for a fixed STI region size the change in mobility increases as the channel length increases. From the stress vs. position curve in Fig. 5 it can be seen that, for a fixed $SA+L+SB$, as L increases the average stress level of the channel increases, and if $SA+L+SB$ and L are both fixed the average stress level in the device changes depending upon the position of the gate within the isolation region. As with WPE modeling, it would seem that introducing moments of the stress, integrated and averaged over the gate area, would be useful. However, chronologically STI effect modeling was introduced before WPE modeling, and so preceded the development of the concept of characterization via moments. Therefore an additional geometry dependence of the STI variation of mobility was included by making K_μ an empirical function of length and width [7]

$$K_\mu = \frac{K_{UO}}{1+\frac{L_{KUO}}{L^{LLODKUO}}+\frac{W_{KUO}}{W^{WLODKUO}}++\frac{P_{KUO}}{L^{LLODKUO}W^{WLODKUO}}} \tag{10}$$

where K_{UO}, L_{KUO}, $LLODKUO$, W_{KUO}, $WLODKUO$, and P_{KUO} are introduced as fitting parameters. The second and third terms in the denominator enable modeling of the dependence on length and width, respectively, and the fourth term enables flexibility in fitting short and narrow devices. Additional geometry offsets and empirical temperature dependence are added in practice, and a similar stress dependence model is applied to the saturated carrier velocity. Further details, and information on how to compute STI effects for multi-fingered devices and more complex layouts than shown in Fig. 5, are available in [5][7].

In contrast to the STI effect on mobility, which for a given $SA+L+SB$ decreases as L decreases, the STI effect on threshold voltage is observed to increase as L decreases. This is attributed to the increased sensitivity of short devices to the halo doping profiles, which are changed by stress via its impact on dopant diffusion [7]. Slightly different empirical forms are therefore used,

$$V_{TO} = V_{TO\,ref} + K_V\left(\frac{1}{SA+L/2}+\frac{1}{SB+L/2}-\frac{1}{SA_{ref}+L/2}-\frac{1}{SB_{ref}+L/2}\right) \tag{11}$$

where

$$K_V = \frac{K_{VTO}}{1+\frac{L_{KVTO}}{L^{LLODKVTO}}+\frac{W_{KVTO}}{W^{WLODKVO}}++\frac{P_{KVTO}}{L^{LLODKVTO}W^{WLODKVTO}}} \tag{12}$$

and K_{VTO}, L_{KVTO}, $LLODKVTO$, W_{KVTO}, $WLODKVTO$, and P_{KVTO} are parameters. This is an additive, rather than multiplicative, correction, and similar to the mobility parameter STI model is included empirically to enable flexibility in fitting observed STI related variations in threshold voltage with geometry. For charge sheet models, the threshold voltage shift, the term involving K_V in (11) is implemented as a shift in V_{FB} which, from (4.7.19), is equivalent.

Similar expressions to the above have also been used to include changes with STI stress for the body effect coefficient, DIBL, and velocity saturation.

At present all modeling of stress effects is empirical and is not based on first-principles theoretical analysis. It is unlikely that this situation will change. The innovation in technologies in ways to enhance device performance by deliberately introducing stress continues to advance rapidly, and theoretical modeling of the effects of stress on device behavior lags the technological innovation rate both because of the increasing complexity of the geometric relationships between stress sources and devices and because of the immaturity of direct modeling of how stress sources alter physical parameters that control device performance. Empirical modeling based on proximity of a device to main sources of stress will continue to be the necessary for the foreseeable future.

References

[1] T. B. Hook, J. Brown, P. Cottrell, E. Adler, D. Hoyniak, J. Johnson, and R. Mann, "Lateral ion implant straggle and mask proximity effect," *IEEE Transaction on Electron Devices*, vol. 50, no. 9, pp. 1946-1951, September 2003.

[2] Y.-M. Sheu, K.-W. Su, S.-J. Yang, H.-T. Chen, C.-C. Wang, M.-J. Chen, and S. Liu, "Modeling well edge proximity effect on highly-scaled MOSFETs," *Proceedings of the Custom Integrated Circuits Conference*, pp. 831-834, September 2005.

[3] J. Watts, K.-W. Su, and M. Basel, "Netlisting and modeling well-proximity effects," *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 2179-2186, September 2006.

[4] M.Basel, J. Xi, J. Watts, P. Humphries, K. Su, S. Moinian, R. Thoma, and K. Onozawa, "Guidelines for extracting well proximity effect instance parameters," CMC web site [Online]: http://www.geia.org/index.asp?bid=1333

[5] BSIM web site [Online]. http://www-device.eecs.berkeley.edu/~bsim3/bsim4.html

[6] G. Scott, J. Lutze, M. Rubin, F. Nouri, and M. Manley, "NMOS drive current reduction caused by transistor layout and trench isolation induced stress," *Technical Digest of the International Electron Devices Meeting*, pp. 827-830, December 1999.

[7] K.-W. Su, Y.-M. Sheu, C.-K. Lin, S.-J. Yang, W.-J. Liang, X. Xi, C.-S. Chiang, J.-K. Her, Y.-T. Chia, C. H. Diaz, and C. Hu, "A scaleable model for STI mechanical stress effect on layout dependence of MOS electrical characteristics," *Proceedings of the Custom Integrated Circuits Conference*, pp. 245-248, September 2003.