Limits of Scaling MOSFETs

Grant McFarland and Michael Flynn

Technical Report CSL-TR-95-662

January 1995

This work was supported by the NSF under contract MIP93-13701 and by fellowship support from the IBM/CIS Fellow Mentor Advisor Program.

Limits of Scaling MOSFETs

by

Grant McFarland and Michael Flynn

Technical Report CSL-TR-95-662

January 1995

Computer Systems Laboratory

Departments of Electrical Engineering and Computer Science

Stanford University

Stanford, California 94305-4055

Abstract

The fundamental electrical limits of MOSFETs are discussed and modeled to predict the scaling limits of digital bulk CMOS circuits. Limits discussed include subthreshold currents, time dependent dielectric breakdown (TDDB), hot electron effects, and drain induced barrier lowering (DIBL). This paper predicts the scaling of bulk CMOS MOSFETs to reach its limits at drawn dimensions of approximately $0.1\mu m$. These electrical limits are used to find scaling factors for SPICE Level 3 model parameters, and a scalable Level 3 device model is presented. Current trends in scaling interconnects are also discussed.

Key Words and Phrases: MOSFET, device scaling, interconnect scaling, time dependent dielectric breakdown, hot electron effects, drain induced barrier lowering, spice models

Contents

List of Figures

List of Tables

1 Constant Field Scaling

Since MOSFET integrated circuits were first invented device performance has been steadily improved by scaling to smaller physical dimensions. This scaling of devices combined with scaling of the interconnects has also allowed higher levels of integration and has dramatically improved computer performance. The most famous systematic scheme for scaling MOSFET devices was written by Robert Denard in 1974 where he proposed constant field scaling [1]. In order to maintain the same qualitative behavior as we scale to smaller device sizes, Denard suggested the following scaling scheme.

Constant Field Scaling		
Description	Parameter	Scaling
Device Dimensions	L, W	$1/S$
Oxide Thickness	TOX	$1/S$
Channel Doping	NSUB	S
Power Supply	VDD	$1/S$
Junction Depth	XJ	$1/S$

Scaling the power supply down as channel doping is increased will cause the widths of the depletion regions to scale with the device dimensions. Since the oxide thickness and depletion widths are decreased by the same factor as the supply voltage, all the electric fields within the device will remain constant. We can see how closely industry has followed Denard's prediction by comparing constant field scaling on the base technology provided in Denard's paper to a modern fabrication process.

Technology Scaling Comparison			
Parameter	L=5μm[1] 1974	L=0.8μm Scaled	L=0.8μm[1]1993
TOX	1000 $\mathring{A}$	160 $\mathring{A}$	175 $\mathring{A}$
NSUB	$5 \times 10^{15} cm^{-3}$	$3 \times 10^{16} cm^{-3}$	$4 \times 10^{16} cm^{-3}$
VDD	$12V$	$2V$	$5V$

We see that constant field scaling predicts values for T_{OX} and N_{SUB} which are very close to those used today, but it also predicts a power supply voltage more than a factor of 2 below what is currently used. There are two principal reasons why industry has been reluctant to scale voltages.

[1] MOSIS HP CMOS26B Process

The first is the difficulties caused in board level design for chips designed to operate at different voltages. Secondly scaling the power supply hurts performance. This has caused chip foundries to avoid scaling power supplies as long as possible, and therefore while constant field scaling has been a good predictor of oxide thicknesses and doping levels, it has not been a good predictor of power supply voltages or device performance. Other criteria besides constant fields are needed to predict future scaling of devices.

2 Performance Scaling

Performance scaling seeks to scale each fabrication parameter to provide the highest performance device possible. The fact that current power supplies are much higher than predicted by Denard shows that the industry as a whole is not interested in maintaining constant fields. More realistic motivations for scaling device parameters are performance and reliability. Therefore, to predict the scaling of future technologies we should scale each parameter to provide the highest performance device possible while satisfying certain basic electrical and reliability requirements. These fundamental limits include the following:

Subthreshold Leakage Currents

To keep power consumption down we must limit leakage currents by maintaining reasonably high thresholds. For high performance operation the power supply voltage must be significantly above the threshold. These requirements of low leakage and high performance limit how small a power supply voltage may be used.

Time Dependent Dielectric Breakdown

High electric fields in the gate oxide can cause gradual deterioration of the oxide layer until eventually dielectric breakdown is reached. Long term reliability of the gate oxide limits how thin an oxide layer may be used.

Hot Electron Effects

High lateral fields in the channel can accelerate some electrons to high enough energies to pass into the gate oxide and change the device threshold over time. Long term threshold stability limits how short a channel length may be used.

Short Channel Effects

As we move the source and drain of a MOSFET closer physically together, it becomes more and more difficult to electrically isolate them. In deep sub-micron MOSFETs the depletion regions of the source and drain can significantly deplete the channel region making the threshold a function of the device length and drain voltage. These effects limit how low a doping concentration may be used.

In the following sections we will examine the nature of these limitations and how they effect the scaling of MOSFETs.

2.1 Subthreshold Leakage Currents

Simple MOSFET models assume that the device current is 0 for $V_{gs} < V_T$, however in reality the drain current decreases exponentially to 0 as the gate voltage drops below the threshold. The slope of this subthreshold current can be shown to be:

$$\frac{d(V_{gs})}{d(\log I_D)} = (\frac{kT}{q}\ln\ 10)(1 + \frac{C_D}{C_{OX}}) \tag{1}$$

Where C_D is the depletion capacitance in the channel. This means that at room temperature in the best case where $C_D \ll C_{OX}$ the subthreshold slope will be:

$$\frac{d(V_{gs})}{d(\log I_D)} = 60\ mV/decade \tag{2}$$

The on/off current ratio of a device is defined as the ratio of the current at $V_{GS} = V_T$ and the current at $V_{GS} = 0$. The larger this ratio the less significant leakage currents will be for a particular technology. For high performance chips we assume a current ratio of at least 10^5 is desirable [2], which assuming optimum subthreshold slope gives a minimum $V_T > 0.3V$.

Because currents drop off exponentially in the subthreshold regime, circuits designed to switch using primarily subthreshold currents will face severe performance penalties (see figure 1).

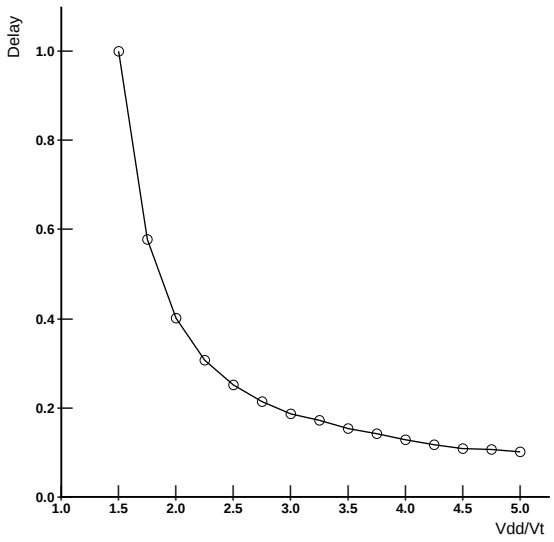

Figure 1: Fanout of 4 Inverter Delay

For high performance operation we must use a power supply voltage sufficiently greater than V_T so that devices will spend most of their switching time out of the subthreshold regime. Analyzing

performance as a function of V_T/V_{DD} shows that a good requirement for high performance operation is [3]:

$$V_{DD} > 4V_T > 1.2V \tag{3}$$

Therefore, subthreshold currents and the need for high performance limit the extent to which we can scale V_{DD} and V_T.

2.2 Time Dependent Dielectric Breakdown (TDDB)

MOSFET current drive can be increased and short channel effects reduced by using thinner gate oxides. However, if the power supply voltage is not scaled with the oxide thickness, fields in the oxide will increase. Experiments have shown that over time these high fields can damage the oxide layer until eventually breakdown occurs [4]. The time to breakdown is commonly written as:

$$t_{BD} = \tau_0 \exp^{G/E_{OX}} \tag{4}$$

where G is the breakdown acceleration factor, τ_0 is a time constant, and E_{OX} is the field in the oxide. At 25°C typical values for G and τ_0 are 350 MV/cm and 1×10^{-11} sec [5]. At 125°C t_{BD} is seen to be reduced by a factor of 20 compared to lifetimes at 25°C [6]. By solving for the electric field we are able to find that in order to have a 10 year lifetime at 125°C the field in the gate oxide should be below approximately 7 MV/cm. Assuming a worst case voltage of 10% over the nominal supply and adding 50% to the minimum thickness in order to compensate for process variation and defects we calculate the following condition:

$$T_{OX} > 1.5(1.1V_{DD})/(7\,MV/cm) \tag{5}$$

$$E_{OX} = V_{DD}/T_{OX} < 4.24\,MV/cm \tag{6}$$

Keeping this limit in mind we can look at how the gate oxide fields of SRAMs and processors have changed in recent years. Figure 2 shows that fields in the gate oxide have been steadily increasing until today dielectric breakdown is a serious concern, and it is common to find parts with fields very close to the limit we have given. Figure 3 compares this limitation with the oxide thicknesses projected in the SIA Roadmap [7] and shows that this restriction should be valid for future MOSFET generations as well.

2.3 Hot Electron Effects

Another effect of not scaling power supply voltages with device size has been increasing lateral electric fields in the channel. At sufficiently high fields electrons may gain enough energy to over come the oxide barrier and enter the gate oxide. This build up of negative charge will causes NMOS thresholds to increase over time. Eventually reduced current drive will cause the circuit to fail.

PMOS devices can also experience hot electron effects. In high fields holes can cause impact ionization in the channel producing an electron-hole pair. This free electron can then be swept

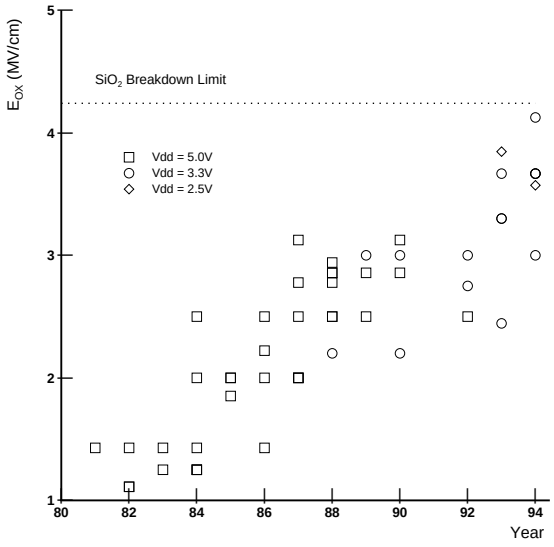

Figure 2: Gate Oxide Fields vs Year

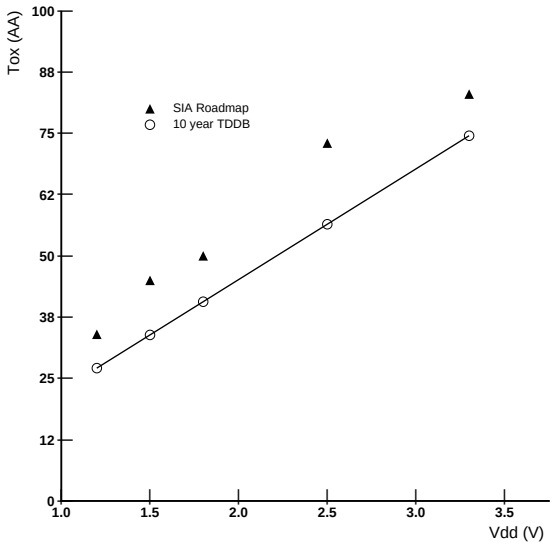

Figure 3: Minimum Gate Oxide Thickness

into the gate which will cause the PMOS threshold to become less negative (increasing the current drive). However, because of the lower mobility of holes and their reduced ability to cause impact ionization, hot electron effects are much less significant in PMOS devices [8].

Because some of the electrons which enter the oxide will actually pass all the way through to the gate node, gate current is a good measure of how severe hot electron effects are in a given device. We can write the ratio of gate current to drain current as [9]:

$$I_G/I_D \approx C_2 \mathrm{Exp}(\frac{\phi_B}{\lambda E_{MAX}}) \tag{7}$$

Where C_2 is a constant (4×10^{-3}), ϕ_B is the Si-SiO_2 barrier height (2.5V), λ is the hot-electron mean-free-path (78 Å), and E_{MAX} is the maximum lateral electric field in the channel. The only unknown in this equation is the lateral electric field. This has been determined empirically to be [10]:

$$E_{MAX} = \frac{V_D - V_{DSAT}}{0.2\, T_{OX}^{1/3}\, Xj^{1/2}} \tag{8}$$

Where V_{DSAT} is the potential at the pinch-off point in the channel and be modeled as [10]:

$$V_{DSAT} = \frac{(V_G - V_T) L_{EFF} E_{SAT}}{V_G - V_T + L_{EFF} E_{SAT}} \tag{9}$$

Where E_{SAT} is the carrier velocity saturation field of approximately $5 \times 10^4 V/cm$ for electrons. Note that thinning the gate oxide makes hot electron effects worse by moving the pinch-off point closer to the drain, and therefore increasing the lateral electric field.

A common means for reducing hot electron effects is the use of a lightly doped drain (LDD). By first performing a light implant and then using oxide spacers left on either side of the gate to mask a second heavy implant, the doping profile of the source and drain can be made to be much more gradual which will reduce E_{MAX}. How much E_{MAX} is reduced depends on the length of the lightly doped portion of the drain, however because of its light doping this region has a high resistance and as its length is increased performance will suffer. Typically E_{MAX} can be reduced to between 60% and 70% of its value in a comparable standard device without severe performance penalties [11].

By using these equations we can calculate the gate current to drain current ratios of recent fabrication technologies to see how hot electron effects have changed over time (see figure 4). Hu gives $I_G/I_D < 10^{-15}$ as a common criterion for designing a device to have less than a 10% shift in transconductance over 10 years [9], and our data shows that as of 1994 few parts have been made which violate this rule. Assuming a minimum gate oxide thickness based on the limits given in the previous section and that the junction depth at the channel edge is scaled with channel length[2], we can plot the minimum L_{EFF} to avoid serious hot electron effects. Figure 5 shows that this requirement matches well with SIA predictions for future technologies.

[2]Starting at a depth of $150\,nm$ for a $3.3V$ technology

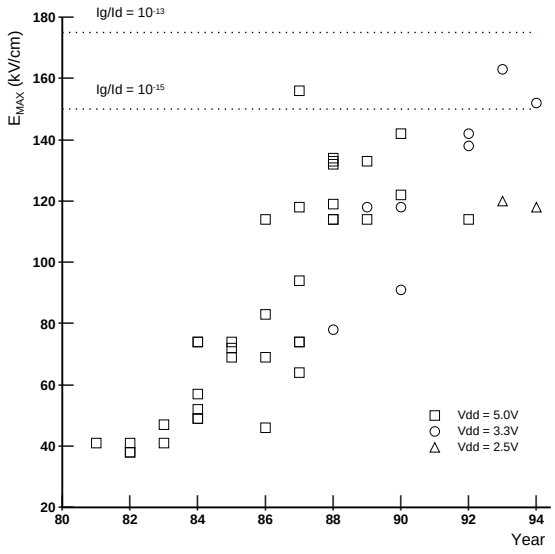

Figure 4: Channel Fields vs Year

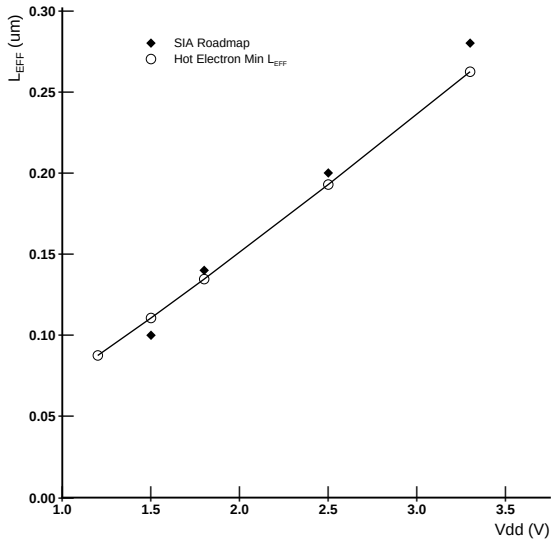

Figure 5: Minimum L_{EFF}

2.4 Short Channel Effects

As MOSFET channel length is reduced, the device threshold becomes dependent on L and V_{DS}. These deviations from the ideal threshold model are known as the short channel effect and drain induced barrier lowering (DIBL) respectively [12]. For sub-micron devices decreasing L and increasing V_{DS} will drive the threshold down as more of the channel region becomes depleted by the source and drain regions instead of the gate. At very high V_{DS} the depletion regions of the source and drain can touch causing large amounts of current to flow uncontrolled by the gate. This phenomenon is known as punch-through. All these effects can be reduced by increasing the channel doping to reduce the size of the source and drain depletion regions and by decreasing the gate oxide thickness to give the gate more control over the channel region. Since all these problems are related and solved in the same manner, this paper considers only the limitation of DIBL. The SPICE Level 3 model for DIBL is based on the work of Masuda and is as follows [2]:

$$\Delta V_T = \frac{-K \cdot ETA}{C_{OX} \cdot L_{EFF}^3} V_{DS} \tag{10}$$

Where $K = 8.14 \times 10^{-22}$ and ETA is a curve fitting parameter. However, Masuda's model was based on devices with channel lengths all over $1\mu m$. To better predict DIBL in sub-micron devices we need to use a more recent model. A model presented in 1993 by Liu and intended for use with devices between $0.1\mu m$ and $1\mu m$ is as follows [13]:

$$\Delta V_T = (e^{-L_{EFF}/l} + e^{-L_{EFF}/2l}(1 + \tfrac{V_{DS}}{PB+PHI})^{-1/2})V_{DS} \tag{11}$$

$$l = 0.1(X_J T_{OX} X_{DEP}^2)^{1/3} \tag{12}$$

$$X_{DEP} = \sqrt{(2\epsilon_{Si} PHI)/(q\, NSUB)} \tag{13}$$

Where X_{DEP} is the vertical depletion width in the channel. In order to be able to use SPICE to simulate scaled devices, we must model how curve fitting parameters such as ETA will vary as we scale other device parameters. By comparing the SPICE DIBL model with Liu's model we find that to make the two models give similar results we should scale ETA as follows:

$$ETA \propto \frac{S_{OX}^{0.5} S_{VDD}^{0.25}}{S_{SUB} S_L^{1.5}} \tag{14}$$

Where S_{OX}, S_{VDD}, S_{SUB}, and S_L are the scaling factors for gate oxide thickness, power supply voltage, channel doping, and channel length respectively. Figure 6 shows the fit between Liu's model and the SPICE model when scaling length alone and setting ETA accordingly. By using Liu's model with the minimum values of T_{OX} and L_{EFF} which meet the requirements for oxide breakdown and hot electron effects already discussed, we can calculate the minimum channel doping for a given variance in threshold. A reasonable requirement for the maximum threshold variance due to DIBL is [2]:

$$\Delta V_T < 0.25 V_T \tag{15}$$

Assuming this requirement the minimum channel doping for a given supply voltage is shown in figure 7.

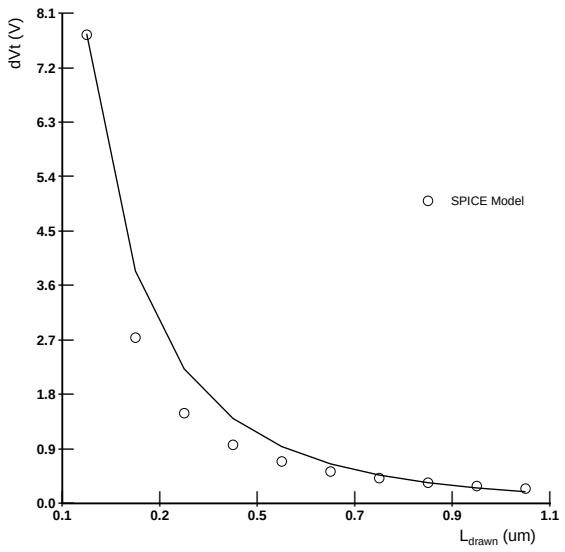

Figure 6: Drain Induced Barrier Lowering

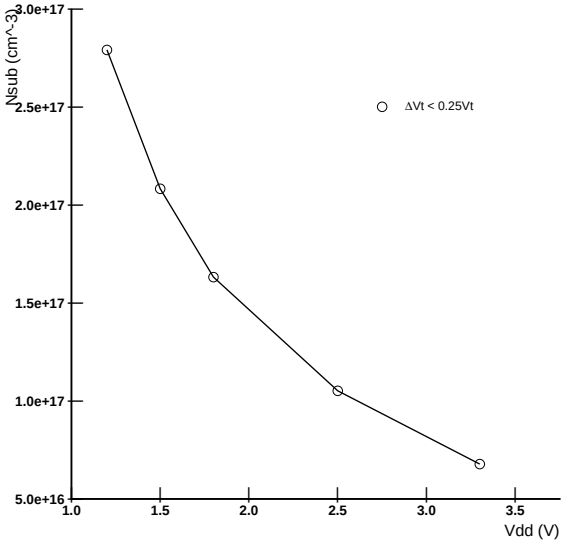

Figure 7: Minimum Channel Doping

2.5 Summary of Scaling Limits

By taking into account these various electrical requirements, we can construct the following list of scaling limits for any reliable high performance MOSFET.

Subthreshold Leakage Currents

To maintain at least 5 orders of magnitude between on and off currents and high performance operation we need:

$$V_T > 0.3V \qquad\qquad V_{DD} > 1.2V \tag{16}$$

Time Dependent Dielectric Breakdown

For a 10 year oxide lifetime at 125°C we specify:

$$E_{OX} < 4.24\,MV/cm \tag{17}$$

Hot Electron Effects

For less than a 10% shift in device transconductance over 10 years we specify:

$$I_G/I_D < 10^{-15} \tag{18}$$

Short Channel Effects

The minimum channel doping is determined by the DIBL requirement:

$$\Delta V_T < 0.25 V_T \tag{19}$$

By assuming that L_{DRAWN} is typically 25% larger than L_{EFF} and applying the limits above we can create the following table of likely future technology characteristics.

Performance Scaling Summary				
$VDD(V)$	$L_{DRAWN}(\mu m)$	$L_{EFF}(\mu m)$	$T_{OX}(\mathring{A})$	$N_{SUB}(cm^{-3})$
3.3	> 0.33	> 0.26	> 75	$< 5.0 \times 10^{16}$
2.5	0.24 - 0.33	0.19 - 0.26	55 - 75	5.0×10^{16} - 6.8×10^{16}
1.8	0.18 - 0.24	0.14 - 0.19	40 - 55	6.8×10^{16} - 9.1×10^{16}
1.5	0.14 - 0.18	0.11 - 0.14	35 - 40	9.1×10^{16} - 1.0×10^{17}
1.2	0.11 - 0.14	0.09 - 0.11	30 - 35	1.0×10^{17} - 1.1×10^{17}

Performance scaling predicts a minimum drawn channel length of $0.11\mu m$ for standard LDD devices, but this is not intended to represent the ultimate end of advancement in MOSFET technology. These constraints are merely trying to predict the limits of scaling current technologies without any radical changes in the construction or operation of the devices. New semiconductor materials, oxide materials, or device structures will likely carry us beyond the limits standard MOSFET technologies. However, the cost and uncertainty associated with incorporating large changes in device design or manufacturing make it unlikely these new fabrication techniques will be implemented until we have scaled very close to the limits of current technologies.

2.6 Performance Scaling of SPICE Level 3 Model

Using these scaling requirements for each of our fabrication parameters and writing all other model parameters as functions of these scaling factors, we can create a scalable SPICE model that will simulate devices meeting all of our electrical constraints for drawn lengths down to $0.11\mu m$. The scaling factors for the various parameters are as follows:

Scaling SPICE Level 3 Model			
Description	Parameter	Generalized Scaling	Performance Scaling
Device Dimensions	L, W	$1/S_L$	$1/S$
Oxide Thickness	TOX	$1/S_{OX}$	$1/S^{0.92}$
Channel Doping	NSUB	S_{SUB}	$S^{1.28}$
Power Supply	VDD	$1/S_{VDD}$	$1/S^{0.92}$
Junction Depth	XJ	$1/S_L$	$1/S$
Half Diff Width	HDIF	$1/S_L$	$1/S$
Junction Cap	CJ	$\sqrt{S_{SUB}}$	$S^{0.64}$
Sidewall Cap	CJSW	$\sqrt{S_{SUB}}/S_L$	$1/S^{0.36}$
Gate Sidewall Cap	CJGATE	$\sqrt{S_{SUB}}/S_L$	$1/S^{0.36}$
DIBL	ETA	$S_{OX}^{0.5}S_{VDD}^{0.25}/(S_L^{1.5}S_{SUB})$	$1/S^{2.09}$
Channel Modulation	KAPPA	1	1
Narrow Width Factor	DELTA	1	1
Diffusion Sheet Res	RSH	S_L	S

Note how close these scaling factors are to Denard's constant field scaling. This is because after years of not scaling power supply voltages with device dimensions, we have now reached the point

where gate oxide and channel fields are at their maximum tolerable levels. In order to continue to scale device dimensions and maintain reliability, we are now forced to do something very close to constant field scaling.

3 Delay of Scaled Devices

Using our scaled SPICE models we can simulate device performance over a wide range of drawn lengths.

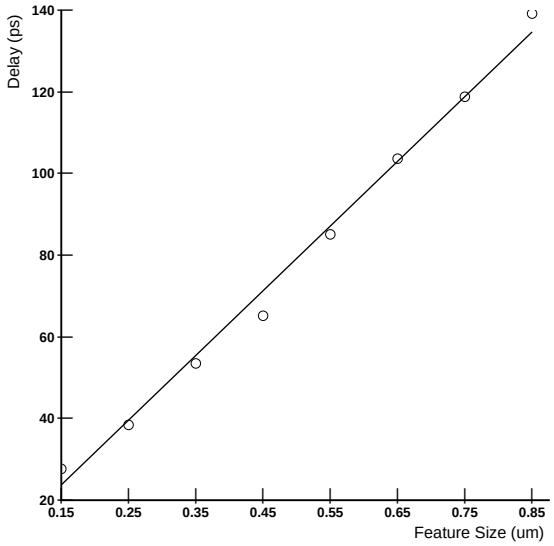

Figure 8: Fanout of 4 Inverter Delay

Figure 8 shows the delay of an inverter driving a fanout of 4 over a range of drawn lengths. The improvement in performance is basically linear as we would expect from heavily velocity saturated devices where delay is as follows:

$$t \propto \frac{Fanout \cdot L}{V_{MAX}} \tag{20}$$

4 Scaling of Interconnects

In addition to the scaling of the devices, it is also important to examine the scaling of the device interconnects and how their changing resistance and capacitance will effect wire delay. The resistance of the line per unit length can be written as

$$R_L = \rho/(W_{INT}T_{INT}) \tag{21}$$

Where ρ is the resistivity of the interconnect material, W_{INT} is the interconnect's width, and T_{INT} is its thickness.

To properly model the capacitance per unit length, we must take into account area and fringe capacitance to the substrate as well as coupling capacitance to adjacent wires. A good empirical model of capacitance is as follows [15]:

$$C_{SUB}/\epsilon_{ox} = 1.15\frac{W_{INT}}{T_{FOX}} + 2.8(\frac{T_{INT}}{T_{FOX}})^{0.222} \quad (22)$$

$$(23)$$

$$C_{COUP}/\epsilon_{ox} = (0.03\frac{W_{INT}}{T_{FOX}} + 0.83\frac{T_{INT}}{T_{FOX}} - 0.07(\frac{T_{INT}}{T_{FOX}})^{0.222})(\frac{W_{SP}}{T_{FOX}})^{-1.34} \quad (24)$$

$$(25)$$

$$C_{TOTAL} = C_{SUB} + 2C_{COUP} \quad (26)$$

Where T_{FOX} is the thickness of the field oxide, and W_{SP} is the width of the space between adjacent lines. Given these models for resistance and capacitance we can compare the three basic interconnect scaling schemes suggested by Bakoglu, ideal scaling, quasi-ideal scaling, and constant-R scaling [14].

Scaling Local Interconnects					
Description	Parameter	Ideal Scaling	Quasi-Ideal Scaling	Constant-R Scaling	SIA Scaling
Width	W_{INT}	$1/S$	$1/S$	$1/\sqrt{S}$	$1/S$
Spacing	W_{SP}	$1/S$	$1/S$	$1/\sqrt{S}$	$1/S$
Thickness	T_{INT}	$1/S$	$1/\sqrt{S}$	$1/\sqrt{S}$	$1/S^{0.41}$
Field Oxide	T_{FOX}	$1/S$	$1/\sqrt{S}$	$1/\sqrt{S}$	$1/S^{0.43}$
Aspect Ratio	Ar	1	$\sqrt{S}$	1	$S^{0.61}$
Res per Length	R_L	S^2	$S^{1.5}$	S	$S^{1.60}$
Cap per Length	C_L	1	$0.9 + 0.1S$	1	$S^{0.31}$
RC per Length	R_LC_L	S^2	$0.9S^{1.5} + 0.1S^{2.5}$	S	$S^{1.91}$

Figure 9 shows that constant-R scaling is clearly superior in minimizing delay, however this scheme means that the wire widths and spacing will not scale as quickly as the device sizes[3]. Very quickly circuit area will become wire limited and integration will suffer. Ideal scaling has the advantages of scaling wire pitch with device size and keeping a constant aspect ratio, but it has the worst delay

[3]This figure assumes that for a scaling factor of 1, $W_{INT} = W_{SP} = T_{INT} = T_{FOX} = 1\mu m$

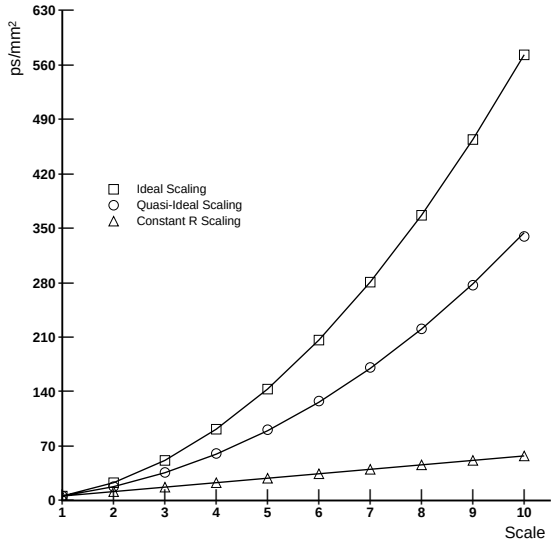

Figure 9: Wire Delay

of the three schemes. Therefore, in order to avoid becoming severely wire limited while keeping delay to a minimum, quasi-ideal scaling is a good middle ground. The major disadvantage is the increase in aspect ratio which can cause step coverage problems, but the inclusion of via-hole filling and dielectric planarization steps in modern processes has allowed increasing aspect ratios to be tolerated. The table above shows that the interconnect scaling predicted by the SIA roadmap matches very closely with quasi-ideal scaling.

5 Conclusions

This paper has shown that as MOSFET channel lengths are scaled, the ranges of oxide thickness, channel doping, and power supply voltage which will produce a reliable device are all reduced. For MOSFETs with dimensions of greater than $1\mu m$ different manufacturers could produce working devices with very different oxide thicknesses and channel dopings. However, as device dimensions move below $0.5\mu m$ toward $0.1\mu m$ all manufacturers will be forced to converge on the very narrow range of parameters which will produce a working device.

In the past constant field scaling has been an excellent predictor of trends for gate oxide thickness and channel doping but not for supply voltages. However, recently fields in the gate oxide and the channel have become high enough to force scaling of the supply voltage. Therefore, future scaling trends should closely follow constant field scaling.

If the proper means of economical fabrication can be developed bulk CMOS MOSFETs can be scaled

down to drawn lengths of approximately $0.10\mu m$. The route cause of this limit is the inability to scale the slope of the subthreshold current.

There are at least three basic approaches for scaling devices below $0.10\mu m$. The subthreshold slope could be improved by cooling devices significantly below room temperature or by using Silicon-On-Insulator (SOI) devices. Alternatively, new gate oxide materials with higher permittivities, higher breakdown fields, and which are more resistant to hot electron effects could reduce the need for further supply voltage reductions. However, even though such devices are electrically feasible it may not be possible to economically mass produce them.

Appendix A – SIA Roadmap

1994 SIA Roadmap Summary					
1st DRAM Year	1995	1998	2001	2004	2007
Feature Size (μm)	0.35	0.25	0.18	0.13	0.10
$L_{EFF}(\mu m)$	0.28	0.20	0.14	0.10	< 0.10
TOX($\AA$)	83	73	50	45	34
Vdd (V)	3.3	2.5	1.8	1.5	1.2
XJ (nm)	70-150	50-120	30-80	20-60	15-45
$W_{INT}(\mu m)$	0.40	0.30	0.22	0.15	0.11
$W_{SP}(\mu m)$	0.60	0.45	0.33	0.25	0.16
$T_{INT}(\mu m)$	0.60	0.60	0.55	0.45	0.39
Wire Aspect Ratio	1.5	2	2.5	3	3.5
Wire Res ($\Omega/\mu m$)	0.15	0.19	0.29	0.82	1.34
Wire Cap ($fF/\mu m$)	0.17	0.19	0.21	0.24	0.27

Appendix B – Scalable HSPICE Device Models

```
* To use model simply set Length parameter to desired drawn device length

.param BaseLength = 0.8u
.param s = 'BaseLength/Length'

* The parameters which can be set the fabrication process are
* Length, Tox, Nsub, and Vdd

.param Slen = s
.param Stox = 'pwr(s, 0.92)'
.param Svdd = 'pwr(s, 0.92)'
.param Ssub = 'pwr(s, 1.28)'

.param vSupply = '7 / Svdd'

.param Seta = 'sqrt(Stox)*pwr(Svdd, 0.25)/(pwr(Slen, 1.5)*Ssub)'

.MODEL TN NMOS LEVEL=3
*
+ * Parameters which can be scaled directly
+ TOX  = '135e-10 / Stox'  * Gate oxide thickness
+ NSUB = '5e16 * Ssub'    * Channel doping
*
+ * Parameters which are functions of scaled parameters
+ XJ     = '0.16u / Slen'       * Junct depth/Short channel effects
+ HDIF   = '1u / Slen'          * Dis from gate to drain center
+ CJ     = '9e-5  * sqrt(Ssub)' * Bottom junction capacitance
+ CJSW   = '5e-10 * sqrt(Ssub)/Slen' * Sidewall junction capacitance
+ CJGATE = '3e-10 * sqrt(Ssub)/Slen' * Gate-edge sidewall capacitance
+ ETA    = '0.03 * Seta'             * DIBL
+ RSH    = '40 * Slen'               * Diff Sheet Resistance
*
+ * Parameters which are independent of scaling
+ TPG    = 1      * Gate poly same type as source and drain
+ ACM    = 3      * Source/drain Diffusion area model
+ KAPPA  = 0.2    * Channel Length Modulation
+ THETA  = 0.12   * Surface Mobility Reduction
+ DELTA  = 0.1    * Narrow Width Effects
+ DELVTO = 0.419  * Zero bias threshold offset
+ U0     = 580    * Carrier mobility
+ VMAX   = 2e5    * Velocity Saturation
```

```
.MODEL TP PMOS LEVEL=3
*
+ * Parameters which can be scaled directly
+ TOX  = '135e-10 / Stox'  * Gate oxide thickness
+ NSUB = '5e16 * Ssub'    * Channel doping
*
+ * Parameters which are functions of scaled parameters
+ XJ     = '0.16u / Slen'       * Junct depth/Short channel effects
+ HDIF   = '1u / Slen'          * Dis from gate to drain center
+ CJ     = '9e-5  * sqrt(Ssub)' * Bottom junction capacitance
+ CJSW   = '5e-10 * sqrt(Ssub)/Slen' * Sidewall junction capacitance
+ CJGATE = '3e-10 * sqrt(Ssub)/Slen' * Gate-edge sidewall capacitance
+ ETA    = '0.03 * Seta'             * DIBL
+ RSH    = '110 * Slen'              * Diff Sheet Resistance
*
+ * Parameters which are independent of scaling
+ TPG    = 1      * Gate poly same type as source and drain
+ ACM    = 3      * Source/drain Diffusion area model
+ KAPPA  = 0.2    * Channel Length Modulation
+ THETA  = 0.12   * Surface Mobility Reduction
+ DELTA  = 0.1    * Narrow Width Effects
+ DELVTO = -0.419  * Zero bias threshold offset
+ UO     = 175    * Carrier mobility
+ VMAX   = 2e5    * Velocity Saturation
```

Appendix C – Scalable HSPICE Wire Models

```
* To use model simply set Length parameter to desired drawn device length

.param BaseLength = 0.8u
.param s = 'BaseLength/Length'

* The parameters which can be set in the metalization process are
* field oxide thickness, and wire thickness, wire width, and
* wire spacing

* Bakoglu's quasi-ideal scaling
.param Swid = s
.param Sfox = 'sqrt(s)'
.param Sthk = 'sqrt(s)'
```

```
* Assume chip side and therefore average wire length increase as
* the sqrt of scaling factor
.param Schp = 'sqrt(s)'

.param rhoAl = 0.03      * Resistivity of Al wires (Ohm)(um)
.param E0 = 8.854e-18    * Permittivity of free space (F/um)
.param Eox = '3.9*E0'    * Permittivity of SiO2 (F/um)

.param Basechip = 12500  * Chip side in um for base tech
                         * Avg 486DX2, microSPARC, R4000
.param BaseWireWidth = '2*BaseLength/1u'  * Base min wire width (um)
.param Basefox = 0.7     * Base tech field oxide thickness (um)
.param Basethk = 0.7     * Base tech wire thickness (um)
.param Cfringe = 0.12f   * Fringe cap (F/um) for base tech

* Pi3 wire model, see Bakoglu p.200

.subckt wireRC in out
+  L='0.1*Basechip*Schp'
+  W='BaseWireWidth / Swid'

    .param Fox   = 'Basefox / Sfox'
    .param Thick = 'Basethk / Sthk'

    .param Res = 'rhoAL * L /(W * Thick)'
    .param Cap = '(W*Eox/Fox + Cfringe)*L'

    R1 in A   'Res/3'
    R2 A  B   'Res/3'
    R3 B  out 'Res/3'

    C1 in  gnd 'Cap/6'
    C2 A   gnd 'Cap/3'
    C3 B   gnd 'Cap/3'
    C4 out gnd 'Cap/6'

.ends wireRC

.subckt wireC in out
+  L='0.1*Basechip*Schp'
+  W='BaseWireWidth / Swid'

    .param Fox   = 'Basefox / Sfox'
```

```
    .param Cap = '(W*Eox/Fox + Cfringe)*L'

    R1 in out 0.001

    C1 out gnd 'Cap'

.ends wireC
```

Appendix D – Mathematica Models

TOXBREAK

```
<< rules.m

(* For a given power supply return the minimum safe gate oxide  *)
(* thickness to avoid oxide breakdown in time tbd.              *)

ToxBreak[vdd_, tbd_:(20*10*year)] := 1.5(1.1vdd/Ebreak[tbd]);

(* Return the maximum electric field for a given oxide thickness *)
(* and desired time to breakdown.                                 *)
(* Lifetime is measured at 25C, to calculate breakdown field for *)
(* 125C, multiply TBD by 20.                                      *)

Ebreak[tbd_:(10*year)] :=
  Block [ { t0 = 10^-11*sec,
            G = 350*10^6*(V/cm) },

    Return[ N[ G/Log[tbd/t0] ] ]
]

(* K. Schuegraf and C. Hu, "Hole Injection SiO2 Breakdown Model *)
(* for Very Low Voltage Lifetime Extrapolation", IEEE Tran. on  *)
(* Electron Devices, May 1994, p.761.                           *)

(* K. Schuegraf and C. Hu, "Effects of Temperature and Defects  *)
(* on Breakdown Lifetime of Thin SiO2 at Very Low Voltages",    *)
(* IEEE Tran. on Electron Devices, July 1994, p.1227.           *)
```

LEFF

```
<< rules.m
<< toxbreak.m

(* Return the minimum Leff which will not suffer severe hot *)
(* electron effects at a given supply voltage assuming the  *)
(* minimum gate oxide thickness for that voltage and that   *)
(* a lightly doped drain structure is being used.           *)

(* Hu, C., "Hot-Electron Induced MOSFET Degradation -- Model, *)
(* Monitor, and Improvement", IEEE Tran on Electron Devices,  *)
(* 1985, p. 375.                                              *)

(* FRF is from Mayaram, "A Model for the Electric Field in *)
(* Lightly Doped Drain Structures", ITED, 1987, p. 1509    *)

Leff[vdd_] :=
  Block [ { FRF = 0.7 (* Field Reduction Factor for LDD *) },

  Ldep = 0.2*(m^(1/6)) ToxBreak[vdd]^(1/3) Xj[vdd]^(1/2);

  vdsat = vdd - Ecrit[]*Ldep/FRF;

  Return[ Lsat[vdsat, vdd] ];
];

(* Return value for source/drain junction depth at channel *)
(* for a given value of vdd.  These values reflect the     *)
(* worst case (deepest junctions) from the SIA roadmap.    *)

Xj[vdd_] := 150*10^-9*m / (3.3V/vdd)^(1/0.8)

(* Return the lateral electric field in the channel needed to *)
(* produce the given ratio of gate current to drain current.  *)
(* A ratio of 10^-15 is sometimes given as the maximum        *)
(* acceptable ratio.                                          *)

(* Hu, C., "Hot-electron Effects in MOSFET's", IEDM, 1983, p. 176 *)

Ecrit[ iratio_:(10^-15)] :=
  Block [ { C2 = 4*10^-3,
            PhiB = 2.5*(V),
```

```
            lambda = 78*10^-10*(m) },

  Return[ N[ -PhiB/(lambda Log[iratio/C2]) ] ];
];

(* For a given vdsat and vdd return the leff which will produce *)
(* that pinch off voltage at that supply.  Assume threshold is  *)
(* at a minimum.  For heavily velocity saturated devices the    *)
(* effect of vt is small anyway.                                *)

Lsat[vdsat_, vdd_] :=
  Block [{ esat = 5*10^4*(V/cm),  (* electron vel sat field *)
           vt = 0.3*(V) },

  Return[ vdsat*(vt - vdd)/(esat*(vt - vdd + vdsat)) ];
];
```

NSUBMIN

```
<< rules.m
<< toxbreak.m
<< leff.m

(* Return the minimum substrate doping which will give a    *)
(* threshold variation of less than 1/16 the power supply   *)
(* for a given vdd.  Assuming minimum values for gate oxide *)
(* thickness and effective channel length.                  *)

Nsubmin[vdd_] :=
  Block[ { startsub = 10^12*(cm^-3),
           incsub = startsub/20 },

  tox = ToxBreak[vdd];
  leff = Leff[vdd];

  For[ nsub = startsub,
       dVtDIBL[vdd, tox, leff, nsub]/V > vdd/(16V),
       nsub += incsub,
       incsub = nsub/20;
  ];

  Return[ N[nsub] ];
];
```

```
(* This DIBL model is based on Liu, Z., "Threshold Voltage Model  *)
(* for Deep-Submicrometer MOSFET's", IEEE Tran. on Elec. Devices, *)
(* Jan '93, p.86.                                                 *)

dVtDIBL[vds_, tox_, leff_, nsub_] :=
  Block[ { pb = 0.8V,
           l = Lchar[tox, Xj[vds], nsub] },

  dVds = Exp[-leff/l] + Exp[-leff/(2l)]/Sqrt[1 + vds/(pb + phi[nsub])];

  Return[ vds*dVds];
];

Lchar[Tox_, XJ_, nsub_] :=
  Block[ { tox = Tox/(10^-10*m),
           xj  = XJ/(10^-6*m),
           wch = Wch[nsub]/(10^-6*m) },

  Return[ 0.1(tox * xj * wch^2)^(1/3) * (10^-6*m) ];
];

Wch[nsub_, vsb_:(0)] := Sqrt[2Esi(phi[nsub] + vsb*V)/(q nsub)];

phi[nsub_] = 2*0.026V*Log[nsub/ni];
```

WIREDELAY

```
<< rules.m

(* T. Sakurai, "Simple Formulas for Two and Three Dimensional
   Capacitances", IEEE Tran. on Electron Devices, 1983, p. 183. *)

(* All lengths are in microns *)
(* w = wire width *)
(* t = wire thickness *)
(* h = oxide thickness *)
(* sp = wire spacing *)

GroundCap[w_:(1), t_:(1), h_:(1)] :=
  Eox*(1.15(w/h) + 2.8(t/h)^0.222)

CoupleCap[w_:(1), t_:(1), h_:(1), sp_:(1)] :=
```

```
  2Eox*(0.03(w/h) + 0.83(t/h) - 0.07(t/h)^0.222)(sp/h)^-1.34

WireCap[w_:(1), t_:(1), h_:(1), sp_:(1)] :=
  GroundCap[w, t, h] + CoupleCap[w, t, h, sp]

WireRes[w_:(1), t_:(1)] := rhoAl/(w*t*um^2)

WireDelay[w_:(1), t_:(1), h_:(1), sp_:(1)] :=
  WireRes[w, t] * WireCap[w, t, h, sp]
```

RULES

```
(* This file provides rules for changing units and simplifying *)
(* expressions.  As well as some useful physical constants.    *)

(* Factor square roots *)
Unprotect[Sqrt]
Sqrt[x_*y_] := Sqrt[x]*Sqrt[y]
Sqrt[x_/y_] := Sqrt[x]/Sqrt[y]
Sqrt[x_^n_] := x^(n/2)
Protect[Sqrt]

(* Factor Powers *)
Unprotect[Power]
Power[x_*y_, n_] := Power[x,n]*Power[y,n]
Power[x_/y_, n_] := Power[x,n]/Power[y,n]
Power[x_^m_, n_] := Power[x,m*n]
Protect[Power]

(* Change centimeters to meters *)
cm := 0.01m

(* Change microns to meters *)
um := 10^-6*m

(* Change angstroms to meters *)
AA := 10^-10*m

(* Change Farads into Coloumbs per Volt *)
F := C/V

(* Change Ohms *)
```

```
Ohm := V*s/C

(* Change electron volts *)
eV := V*(1.602*10^-19*C)

(* Change kilograms to electrical units *)
kg := V*C*sec^2/m^2

(* Change years *)
year := 365*24*60*60*sec

(* Simplify Absolute Voltages *)
Unprotect[Abs]
Abs[x_*V] := Abs[x]*V
Protect[Abs]

(* PHYSICAL CONSTANTS *)

E0 = 8.854*10^-14*(F/cm); (* Permittivity of free space *)
Esi = 11.7*E0;            (* Permittivity of Si *)
Eox =  3.9*E0;            (* Permittivity of SiO2 *)
q = 1.602*10^-19*(C);     (* Electron charge *)
ni = 1.45*10^10*(cm^-3);  (* Intrinsic carrier concentration *)
rhoAl = 3*10^-6*(Ohm*cm); (* Resistivity of Al wires *)
```

References

[1] R. Denard. "Design of Ion-Implanted MOSFETs with Very Small Dimensions", *IEEE Journal of Solid State Circuits*, 1974, p. 256.

[2] H. Masuda. "Characteristics and Limitation of Scaled-Down MOSFETs Due to Two-Dimensional Field Effect", *IEEE Transactions on Electron Devices*, 1979, p. 980.

[3] J. Pfiester. "Performance Limits of CMOS ULSI," *IEEE Transactions on Electron Devices*, 1985, pp. 333.

[4] K. Yamabe. "Time-Dependent Dielectric Breakdown of Thin Thermally Grown SiO_2 Films", *IEEE Transactions on Electron Devices*, 1985, p. 423.

[5] R. Moazzami. "Projecting the Minimum Acceptable Oxide Thickness For Time-Dependent Dielectric Breakdown", *International Electronic Devices Meeting*, 1988, p. 710.

[6] K. Schuegraf. "Effects of Temperature and Defects on Breakdown Lifetime of Thin SiO_2 at Very Low Voltages", *IEEE Transactions on Electron Devices*, 1994, p. 1227.

[7] "The National Technology Roadmap for Semiconductors", Semiconductor Industry Association, 1994.

[8] T. Hayashi. "Hot Carrier Injection in PMOSFETs", *OKI Technical Review*, Sept. 1991, p. 59.

[9] C. Hu. "Hot-Electron Effects in MOSFETs", *International Electronic Devices Meeting*, 1983, p. 176.

[10] C. Hu. "Hot-Electron Induced MOSFET Degradation – Model, Monitor, and Improvement", *IEEE Transactions on Electron Devices*, 1985, p. 375.

[11] Mayaram. "A Model for the Electric Field in Lightly Doped Drain Structures", *IEEE Transactions on Electron Devices*, 1987, p.1509.

[12] R. Troutman. "VLSI Limitations from Drain-Induced Barrier Lowering", *IEEE Transactions on Electron Devices*, 1979, p. 461.

[13] Z. Liu. "Threshold Voltage Model for Deep-Sub-micrometer MOSFETs", *IEEE Transactions on Electron Devices*, 1993, p. 86.

[14] H. Bakoglu. Circuits, Interconnections, and Packaging for VLSI, Addison-Wesley, 1990.

[15] T. Sakurai. "Simple Formulas for Two and Three Dimensional Capacitances", *IEEE Transactions on Electron Devices*, 1983, p. 183.