Chapter 1: Effective Oxide Thickness, Channel Length and Channel Width

1.1 Gate Dielectric Model

As the gate oxide thickness is vigorously scaled down, the finite charge-layer thickness can not be ignored [1]. BSIM4 models this effect in both IV and CV. For this purpose, BSM4 accepts two of the following three as the model inputs: the electrical gate oxide thickness *TOXE*[1], the physical gate oxide thickness *TOXP*, and their difference *DTOX = TOXE - TOXP*. Based on these parameters, the effect of effective gate oxide capacitance C_{oxeff} on IV and CV is modeled [2].

High-*k* gate dielectric can be modeled as SiO_2 (relative permittivity: 3.9) with an equivalent SiO_2 thickness. For example, 3nm gate dielectric with a dielectric constant of 7.8 would have an equivalent oxide thickness of 1.5nm.

BSIM4 also allows the user to specify a gate dielectric constant (*EPSROX*) different from 3.9 (SiO_2) as an alternative approach to modeling high-*k* dielectrics.

Figure 1-1 illustrates the algorithm and options for specifying the gate dielectric thickness and calculation of the gate dielectric capacitance for BSIM4 model evaluation.

1. Capital and italic alphanumericals in this manual are model parameters.

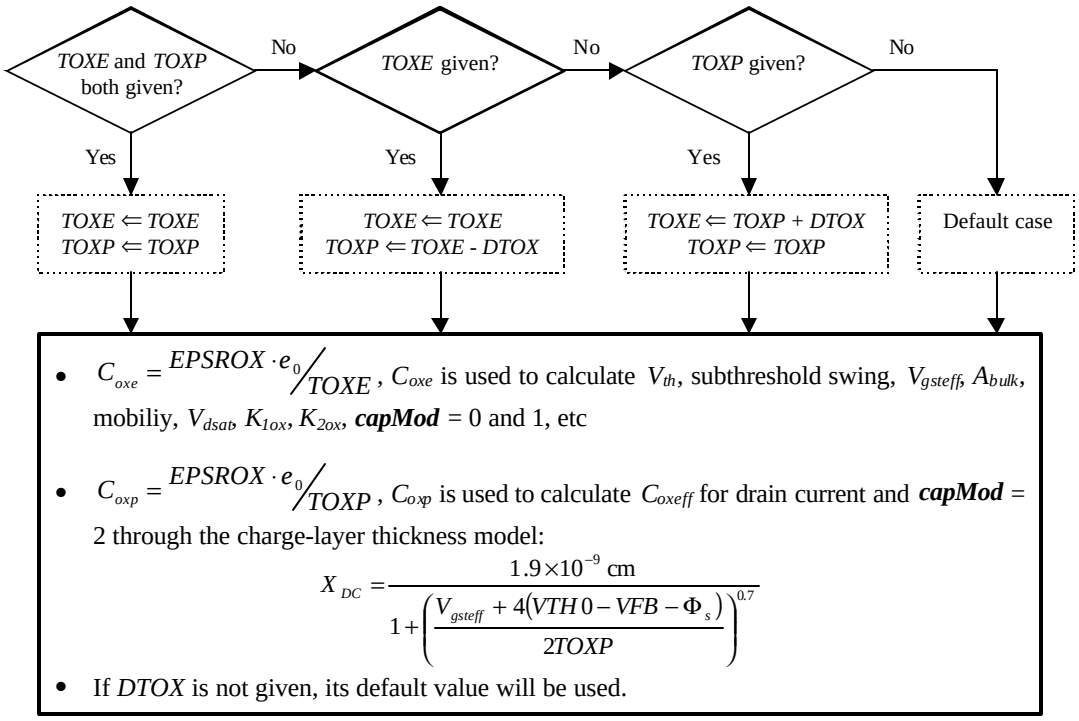

Figure 1-1. Algorithm for BSIM4 gate dielectric model.

1.2 Poly-Silicon Gate Depletion

When a gate voltage is applied to the poly-silicon gate, e.g. NMOS with n^+ poly-silicon gate, a thin depletion layer will be formed at the interface between the poly-silicon and the gate oxide. Although this depletion layer is very thin due to the high doping concentration of the poly-silicon gate, its effect cannot be ignored since the gate oxide thickness is small.

Figure 1-2 shows an NMOSFET with a depletion region in the n^+ poly-silicon gate. The doping concentration in the n^+ poly-silicon gate is *NGATE* and the

doping concentration in the substrate is *NSUB*. The depletion width in the poly gate is X_p. The depletion width in the substrate is X_d. The positive charge near the interface of the poly-silicon gate and the gate oxide is distributed over a finite depletion region with thickness X_p. In the presence of the depletion region, the voltage drop across the gate oxide and the substrate will be reduced, because part of the gate voltage will be dropped across the depletion region in the gate. That means the effective gate voltage will be reduced.

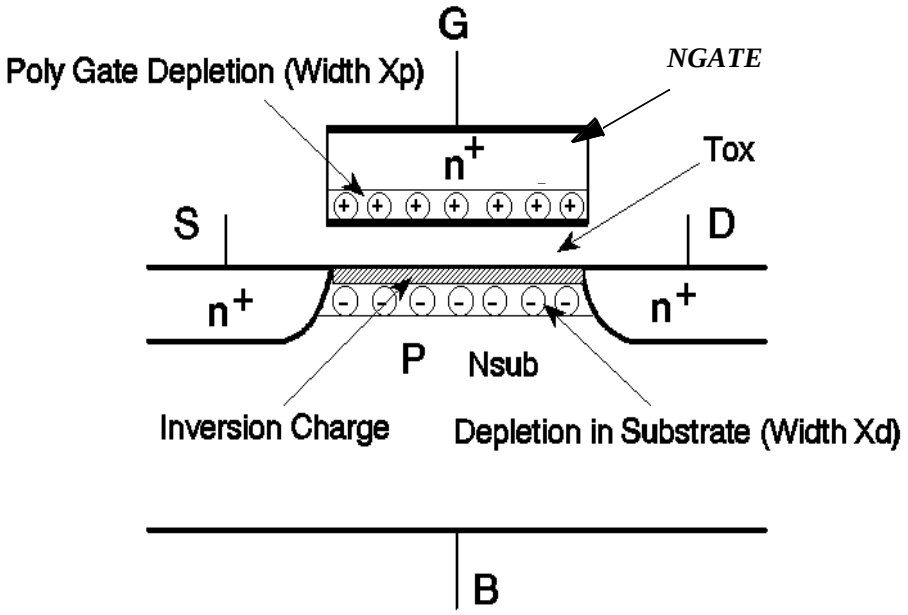

Figure 1-2. Charge distribution in a MOSFET with the poly gate depletion effect. The device is in the strong inversion region.

The effective gate voltage can be calculated in the following manner. Assume the doping concentration in the poly gate is uniform. The voltage drop in the poly gate V_{poly} can be calculated as

(1.2.1)

$$V_{poly} = 0.5 X_{poly} E_{poly} = \frac{qNGATE \cdot X_{poly}^2}{2e_{si}}$$

where E_{poly} is the maximum electrical field in the poly gate. The boundary condition at the interface of poly gate and the gate oxide is

(1.2.2)

$$EPSROX \cdot E_{ox} = e_{si} E_{poly} = \sqrt{2qe_{si} NGATE \cdot V_{poly}}$$

where E_{ox} is the electric field in the gate oxide. The gate voltage satisfies

(1.2.3)

$$V_{gs} - V_{FB} - \Phi_s = V_{poly} + V_{ox}$$

where V_{ox} is the voltage drop across the gate oxide and satisfies $V_{ox} = E_{ox} TOXE$.

From (1.2.1) and (1.2.2), we can obtain

(1.2.4)

$$a\left(V_{gs} - V_{FB} - \Phi_s - V_{poly}\right)^2 - V_{poly} = 0$$

where

(1.2.5)

$$a = \frac{EPSROX^2}{2qe_{si} NGATE \cdot TOXE^2}$$

By solving (1.2.4), we get the effective gate voltage V_{gse} which is equal to

(1.2.6)

$$V_{gse} = VFB + \Phi_s + \frac{qe_{si}NGATE \cdot TOXE^2}{EPSROX^2}\left(\sqrt{1 + \frac{2EPSROX^2\left(V_{gs} - VFB - \Phi_s\right)}{qe_{si}NGATE \cdot TOXE^2}} - 1\right)$$

1.3 Effective Channel Length and Width

The effective channel length and width used in the drain current model are given below where XL and XW are parameters to account the channel length/width offset due to mask/etch effect

(1.3.1)

$$L_{eff} = L_{drawn} + XL - 2dL$$

(1.3.2a)

$$W_{eff} = \frac{W_{drawn}}{NF} + XW - 2dW$$

(1.3.2b)

$$W_{eff}' = \frac{W_{drawn}}{NF} + XW - 2dW'$$

The difference between (1.3.2a) and (1.3.2b) is that the former includes bias dependencies. *NF* is the number of device fingers. *dW* and *dL* are modeled by

(1.3.3)

$$dW = dW' + DWG \cdot V_{gsteff} + DWB\left(\sqrt{\Phi_s - V_{bseff}} - \sqrt{\Phi_s}\right)$$

$$dW' = WINT + \frac{WL}{L^{WLN}} + \frac{WW}{W^{WWN}} + \frac{WWL}{L^{WLN}W^{WWN}}$$

(1.3.4)

$$dL = LINT + \frac{LL}{L^{LLN}} + \frac{LW}{W^{LWN}} + \frac{LWL}{L^{LLN}W^{LWN}}$$

WINT represents the traditional manner from which "delta *W*" is extracted (from the intercept of straight lines on a $1/R_{ds}$~W_{drawn} plot). The parameters *DWG* and *DWB* are used to account for the contribution of both gate and substrate bias effects. For *dL*, *LINT* represents the traditional manner from which "delta *L*" is extracted from the intercept of lines on a R_{ds}~L_{drawn} plot).

The remaining terms in *dW* and *dL* are provided for the convenience of the user. They are meant to allow the user to model each parameter as a function of W_{drawn}, L_{drawn} and their product term. By default, the above geometrical dependencies for *dW* and *dL* are turned off.

MOSFET capacitances can be divided into intrinsic and extrinsic components. The intrinsic capacitance is associated with the region between the metallurgical source and drain junction, which is defined by the effective length (L_{active}) and width (W_{active}) when the gate to source/drain regions are under flat-band condition. L_{active} and W_{active} are defined as

(1.3.5)

$$L_{active} = L_{drawn} + XL - 2dL$$

(1.3.6)

$$W_{active} = \frac{W_{drawn}}{NF} + XW - 2dW$$

$$dL = DLC + \frac{LLC}{L^{LLN}} + \frac{LWC}{W^{LWN}} + \frac{LWLC}{L^{LLN}W^{LWN}} \tag{1.3.7}$$

$$dW = DWC + \frac{WLC}{L^{WLN}} + \frac{WWC}{W^{WWN}} + \frac{WWLC}{L^{WLN}W^{WWN}} \tag{1.3.8}$$

The meanings of *DWC* and *DLC* are different from those of *WINT* and *LINT* in the I-V model. Unlike the case of I-V, we assume that these dimensions are bias-dependent. The parameter dL_{eff} is equal to the source/drain to gate overlap length plus the difference between drawn and actual POLY CD due to processing (gate patterning, etching and oxidation) on one side.

The effective channel length L_{eff} for the I-V model does not necessarily carry a physical meaning. It is just a parameter used in the I-V formulation. This L_{eff} is therefore very sensitive to the I-V equations and also to the conduction characteristics of the LDD region relative to the channel region. A device with a large L_{eff} and a small parasitic resistance can have a similar current drive as another with a smaller L_{eff} but larger R_{ds}.

The L_{active} parameter extracted from capacitance is a closer representation of the metallurgical junction length (physical length). Due to the graded source/drain junction profile, the source to drain length can have a very strong bias dependence. We therefore define L_{active} to be that measured at flat-band voltage between gate to source/drain. If *DWC*, *DLC* and the length/width dependence parameters (*LLC*, *LWC*, *LWLC*, *WLC*, *WWC* and *WWLC*) are not specified in technology files, BSIM4 assumes that the DC bias-independent L_{eff} and W_{eff} will be used for the

capacitnace models, and *DWC*, *DLC*, *LLC*, *LWC*, *LWLC*, *WLC*, *WWC* and *WWLC* will be set to the values of their DC counterparts.

BSIM4 uses the effective source/drain diffusion width W_{effcj} for modeling parasitics, such as source/drain resistance, gate electrode resistance, and gate-induced drain leakage (GIDL) current. W_{effcj} is defined as

(1.3.9)

$$W_{effcj} = \frac{W_{drawn}}{NF} - 2 \cdot \left(DWJ + \frac{WLC}{L^{WLN}} + \frac{WWC}{W^{WWN}} + \frac{WWLC}{L^{WLN} W^{WWN}} \right)$$

Note: Any compact models have their valided design region, so does BSIM4. The warning limits for effective channel length, channel width, and gate oxide thickness are substantially decreased to avoid a large number of warnings when BSIM4 is used beyond its design region. For meaningful results, it is recommended to keep these variables/parameters within the BSIM4 design region, while decreasing warning limits will allow users to extend the model beyond with fewer warnings. The table below table shows the design limit and warning limit. For users' reference, the fatal limitation in BSIM4 is also shown.

TABLE 1. Geometry limitation for BSIM4

	Designed Limitation(m)	Warning Limitation(m)	Fatal Limitation(m)
Leff	1e-8	1e-9	0
LeffCV	1e-8	1e-9	0
Weff	1e-7	1e-9	0
WeffCV	1e-7	1e-9	0
Toxe	5e-10	1e-10	0
Toxp	5e-10	1e-10	0
Toxm	5e-10	1e-10	0