第 23 卷第 5 期　　半 导 体 学 报　　Vol. 23, No. 5
2002 年 5 月　　CHINESE JOURNAL OF SEMICONDUCTORS　　May, 2002

Direct Tunneling Currents Through Gate Dielectrics in Deep Submicron MOSFETs*

Hou Yongtian[1], Li Mingfu[1] and Jin Ying[2]

(1 *Silicon Nano Device Laboratory, Department of Electrical & Computer Engineering, National University of Singapore*)
(2 *Chartered Semiconductor Manufacturing Ltd., Singapore*)

Abstract: A direct tunneling model through gate dielectrics in CMOS devices in the frame of WKB approximation is reported. In the model, an improved one-band effective mass approximation is used for the hole quantization, where valence band mixing is taken into account. By comparing to the experiments, the model is demonstrated to be applicable to both electron and hole tunneling currents in CMOS devices. The effect of the dispersion in oxide energy gap on the tunneling current is also studied. This model can be further extended to study the direct tunneling current in future high-k materials.

Key words: MOSFET; direct tunneling current; quantum effect; gate dielectrics
EEACC: 2560R; 2560B
CLC number: TN386　　**Document code**: A　　**Article ID**: 0253-4177(2002)05-0449-06

1 Introduction

As MOSFETs are scaled into deep submicron regime, it is necessary to reduce the gate oxide thickness. When gate oxide is scaled below 3nm, direct tunneling current will dominate the gate leakage and the off-state power dissipation of the transistor[1]. Therefore, an accurate physical understanding of direct tunnelling through ultrathin gate oxide is important in new transistor technology development. At the same time, the channel doping is always increased in scaled CMOS devices[1]. This results in a large electric field at the Si/SiO_2 interface and leads to significant carrier quantization in the Si substrate.

Quantization[2~7] and direct tunnelling current of conduction electrons[7~14] have been extensively investigated. They rely on the traditional one-band effective mass approximation (EMA), where the effective mass values are extracted from the bulk Si[2~14]. Such a traditional one-band EMA was also applied to study hole quantization[3,15,16]. However this method is invalid in physics due to the valence band mixing effect in the presence of a strong surface electric field. A more rigorous method is by solving the multi-band Schrodinger-Poisson self-consistent equations. However, such a treatment is time-consuming and computationally prohibitive in routine device simulations[17,18]. In this paper, we present a simple and efficient model for quantization and direct tunnelling currents, which is appropriate for both n and p-type MOSFETs.

2 Experiments

In our experiments, the MOSFETs were fabricated by a standard dual-gate CMOS process. The

* Project supported by the National University of Singapore (Grant No. R263-000-077-112), and Singapore NSTB(Grant No. EMT/TP/00/001, 2)
Received 8 March 2002

oxide was grown by rapid thermal oxidation. The current-voltage (I-V) characteristics were measured using the HP4156A semiconductor parameter analyzer. For n-MOSFETs, the electron direct tunneling current was measured as the gate current at inversion (gate voltage $V_g > 0$). For p^+ polysilicon gate p-MOSFETs, the hole direct tunneling current was measured by a carrier separation method[19]. In the carrier separation measurement, the source and drain was tied together and grounded along with the substrate while a negative V_g was applied. The hole direct tunneling current is measured as source/drain current.

3 Description of the model

Figure 1 shows a schematic of the band diagram of a p-MOSFET at inversion. In MOS devices, when a gate voltage is applied, the carriers will be confined in a thin inversion layer and are quantized in energy. The tunneling current is the summation of the contributions from all the subbands.

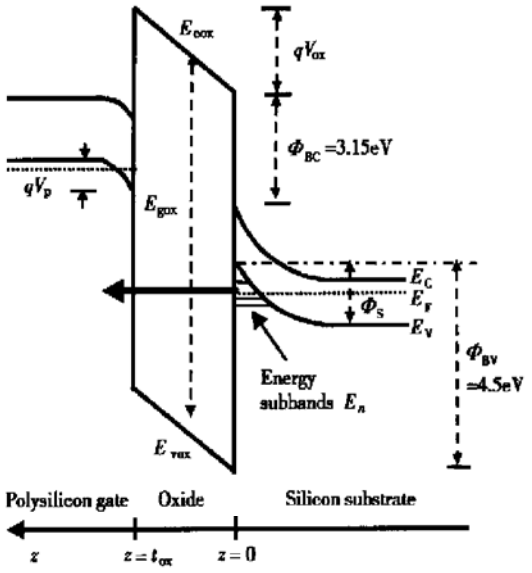

Fig. 1 A schematic of the band diagram of a p^+ polysilicon/SiO_2/n-Si MOS structure showing the hole quantization effect in the substrate and hole direct tunneling from the substrate inversion layer to the polysilicon gate

$$J = \sum_n N_n/\tau_n(E_n) \tag{1}$$

where N_n is the carrier density of the nth subband. τ_n is the lifetime of the carriers in the nth subband, which is determined by the subband energy E_n. Therefore the first step is to model the carrier quantization in the substrate.

3.1 Carrier quantization

For electron quantization, the traditional one-band EMA is applicable. From the widely used triangular well approximation, the subband energy and density are[2]:

$$E_n = \left[\frac{\hbar^2}{2m_{zn}^*}\right]^{1/3}\left[\frac{3}{2}\pi qF\left(n-\frac{1}{4}\right)\right]^{2/3} \tag{2}$$

$$N_n = \left(\frac{kT}{\pi\hbar^2}\right)m_{dn}^*\ln\left[1+\exp\left[\frac{E_F-E_n}{kT}\right]\right] \tag{3}$$

where m_z and m_d are the effective masses for quantization and density of states (DOS), F is the electric field. In order to achieve more accurate results compared with the rigorous self-consistent method, F is always to be replaced by an effective field F_{eff} in the inversion layer[5].

$$F_{eff} = \frac{q(N_{depl}+\eta N_{inv})}{\epsilon_0\epsilon_{Si}} \tag{4}$$

where N_{depl} and N_{inv} are the depletion and inversion charge density in the substrate. η is be found be 0.75 for electron inversion[5] and 0.5 for hole quantization[18].

Due to the valence band mixing in the presence of an electric field, this traditional one-band EMA is not appropriate for valence holes. The rigorous way to treat hole quantization is the multiband effective mass theory. We have developed a simple model for hole quantizaton in triangular well approximation based on a 6×6 EMA hole Hamiltonian including the heavy hole (hh), light hole (lh), spin-orbit (so) split-off hole and their spin degenerate bands[20]. It was found that, due to the valence band mixing, the traditional one-band EMA is not accurate to describe the electrostatics of the inversion layer, such as the subband energies and occupation factors of the subbands as shown in Fig. 2.

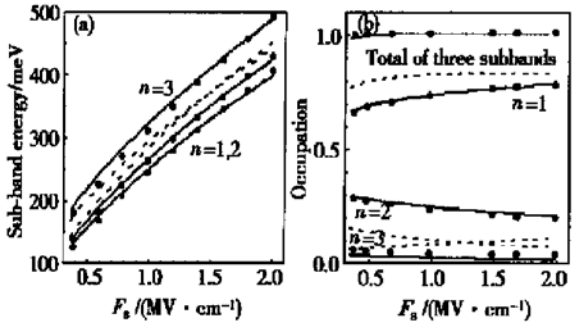

Fig. 2　Calculated (a) subband energies and (b) occupation factors of the three lowest subbands for a p-MOS device at inversion　Our results in six-band EMA are shown as solid circles. The dashed and solid curves are from the traditional and our improved one-band-EMA, respectively. The substrate doping is $5\times10^{17}\mathrm{cm}^{-3}$.

Based on the numerical results of the six-band model, an improved one-band EMA assuming a parabolic dispersion is also proposed[21]. Attention is first paid to the empirical energy quantization effective mass m_z, as shown in Fig. 3(a). They are determined inversely from the energies of the subband minima obtained by six-band EMA. For the lowest energy level $n=1$, the empirical energy quantization effective mass is $0.29m_0$ and is independent of electric field because this energy state is a pure heavy hole state at $k=0$. For $n=2$ and $n=3$ subbands, the empirical effective masses display an electric field dependent behavior due to the field dependence of band mixing effect. From the inversion charge density occupying the respective subbands, the obtained empirical DOS effective masses are shown in Fig. 3(b). Although some of these empirical effective mass values in Fig. 3 are electric field dependent, this dependence can be neglected in the first order approximation. The reason is that at room temperature most of inversion holes are occupied on the $n=1$ subband (over 70%) as shown in Fig. 2(b). From Fig. 3, both the energy quantization and DOS masses of $n=1$ subband have a weak electric field dependence. This leads us to propose a set of constant empirical effective mass values for an improved one-band EMA. They are found to be 0.29/1.16, 0.23/0.70, 0.23/0.60 m_0 for the three subbands. As shown in Fig. 2, such an improved one-band EMA can achieve consistent results on the subband energy levels and carrier occupations in comparison with the numerical results of six-band EMA.

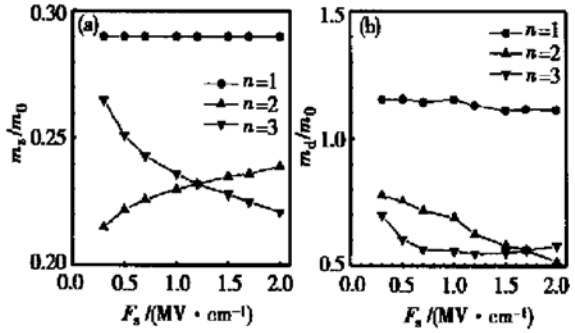

Fig. 3　Electric field F_s dependence of (a) empirical energy quantization effective mass m_z, and (b) empirical DOS effective mass m_d, determined from the numerical results of six-band EMA calculations (solid symbols)

3.2　Tunneling probability

In our direct tunnelling model, a modified WKB approximation, considering the reflections at the barrier discontinuities, is used for the transmission probability[9, 11, 12]. Combined with the improved one-band EMA for quantization, the current-voltage (I-V) characteristics can be obtained efficiently.

For carriers confined in the quasi-bound states in the inversion layer, the lifetime of an nth subband state is approximately given by:

$$\frac{1}{\tau_n(E)}=\frac{T(E)}{\int_0^{z_n}\sqrt{2m_{zm}^*/[E_n-E_{V(C)}(z)]}\,\mathrm{d}z} \tag{5}$$

where E_n is the subband energy for the nth quasi-bound state, $E_{V(C)}$ is the top of the Si valence band (bottom of the conduction band), and z_n the classical turning point for the nth bound state. $T(E)$ is the transmission probability of a particle. In the calculation of direct tunneling current, we adopt a modified WKB approximation with a correction factor accounting for reflections at the boundary of the oxide layer, such that

$$T(E) = T_R(E)T_{WKB}(E) \quad (6)$$

where T_{WKB} is the usual WKB approximation of the transmission probability and T_R the correction factor accounting for the reflections from boundaries of the oxide[11,13]. T_{WKB} is given by

$$T_{WKB}(E) = \exp\left(-2\int_0^{t_{ox}}\kappa(E,z)\,dz\right) \quad (7)$$

where κ is the imaginary wave number within the oxide gap energy and t_{ox} the oxide thickness, i. e. the tunneling distance.

Combining Eqs. (1) ~ (7) and the results from quantization calculation of the inversion layer, the tunneling current can be readily obtained.

4 Results and discussion

4.1 Electron tunneling currents

Electron direct tunneling was first studied. Figure 4 shows the results of electron direct tunneling current. The solid lines are the calculated results using an empirical Franz-type dispersion in the energy gap of SiO_2[10~12]. In the calculation, the conduction band offset between Si and SiO_2 is fixed at 3.15eV and $m_{ox} = 0.61m_0$[11,12]. The calculated results using a simple parabolic dispersion[8] ($m_{ox} = 0.50m_0$) are also displayed in Fig. 4 as dashed lines. The fitting results using the simple parabolic dispersion are only slightly degraded for thick oxides. It indicates that the parabolic dispersion is a good approximation for electron direct tunneling. The oxide thickness determined from the fitting of electron tunneling for our four samples are 1.81, 2.06, 2.34 and 2.66nm. They are close to the values determined from the C-V method (1.85, 2.07, 2.44 and 2.74nm, respectively). The maximum deviation is about 0.1nm and this is within the reported limits of different experimental methods, such as C-V, HRTEM and ellipsometer.

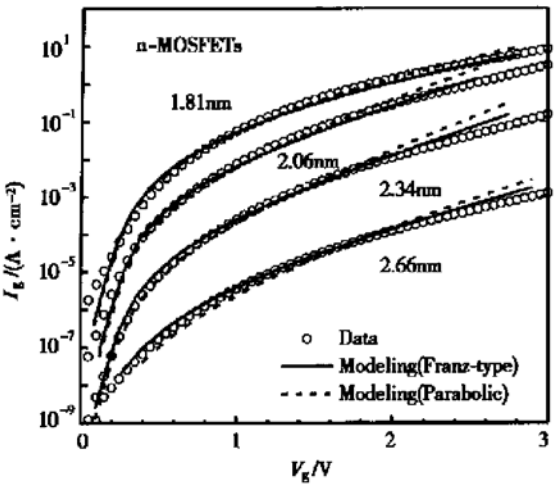

Fig. 4 Electron direct tunneling currents in n-MOSFETs The open circles are the measurements. The solid and dashed lines are the calculations by assuming the electron dispersion in SiO_2 band gap to be Franz-type ($m_{ox} = 0.61m_0$) and parabolic ($m_{ox} = 0.50m_0$), respectively.

4.2 Hole tunneling current

First, we shall discuss the hole direct tunneling using a parabolic hole dispersion in energy gap of SiO_2. To our knowledge, all previous works are based on such an approximation[22]. The results obtained by our physical model calculations are displayed as dashed lines in Fig. 5. In the calculation, the valence band offset between Si and SiO_2 is fixed at 4.5eV[19]. Assuming a parabolic dispersion, m_{ox}

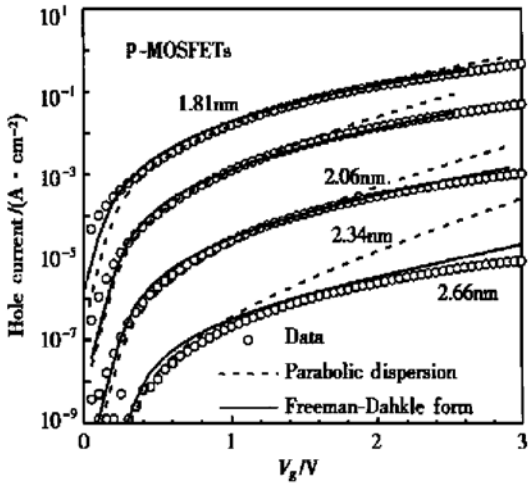

Fig. 5 Hole direct tunneling currents in p-MOSFETs The open circles are the measured values. The dashed lines are the calculations by assuming the hole dispersion in SiO_2 band gap to be parabolic ($m_{ox} = 0.41m_0$). The solid lines denote the results by assuming a Freeman-Dahlke form dispersion in SiO_2 band gap with $m_{cox} = 0.50m_0$ and $m_{vox} = 0.80m_0$.

is found to be $0.41m_0$ in order to get the best fitting. This value is close to the previous results reported for valence band electron or hole tunneling ($0.35 \sim 0.50m_0$)[19,22]. From Fig. 5, the parabolic dispersion can not fit the experimental data when the oxide thickness is larger than about 2nm. The deviation is more significant at higher gate voltage.

For hole tunneling, an appropriate energy dispersion relationship in SiO_2 band gap is important. The existing models of hole DT all assume a parabolic dispersion of holes in the SiO_2 band gap[19,22]. However, it is questionable because the Si valence band aligns near the middle of the SiO_2 gap. To include the non-parabolic effect, a more appropriate Freeman-Dahlke dispersion[23] was used in our simulation.

$$\frac{1}{(\hbar k)^2} = \frac{1}{2m_{cox}(E_{cox} - E)} + \frac{1}{2m_{vox}(E - E_{vox})} \tag{8}$$

where E is the energy, E_{cox} and E_{vox} are oxide conduction and valence band edges respectively. m_{cox} and m_{vox} are effective masses of oxide conduction and valence bands. Equation (8) treats the SiO_2 conduction and valence bands on equal footing. Different from Franz-type dispersion where the same effective mass values are used, the Freeman-Dahlke form introduces different effective masses for conduction and valence bands.

The hole direct tunneling current calculated using the Freeman and Dahlke form of Eq. (8) is shown in Fig. 5 as solid lines. In the calculation, there are two effective mass values. From the electron tunneling and SiO_2 band structure calculation, we have $m_{cox} = 0.5m_0$[8]. m_{vox} is an adjustable parameter for best fitting. A t_{ox} independent value of about $m_{vox} = 0.8m_0$ can give the best results. From Fig. 5, it is apparent that a much better fitting of hole tunneling current to the experimental data can be achieved by using the Freeman-Dahlke form for the hole dispersion in SiO_2. It is also concluded that the simple parabolic dispersion can only be applied to thin oxide (< 2nm) or for low gate voltage.

4.3 Tunneling current in future high-k materials

Because of the excess tunneling current, new high-k materials will be needed to replace SiO_2 as the gate dielectric in future CMOS devices. Our direct tunneling model can also be extended to study the future high-k materials. Figure 6 is a typical study on the HfO_2 gate dielectric. The material parameters of HfO_2 are taken from[24]. At equivalent oxide thickness (EOT) of 1.2nm, the tunneling current through HfO_2 is about 4 orders lower than that of SiO_2, which is consistent with the recent experiments[25].

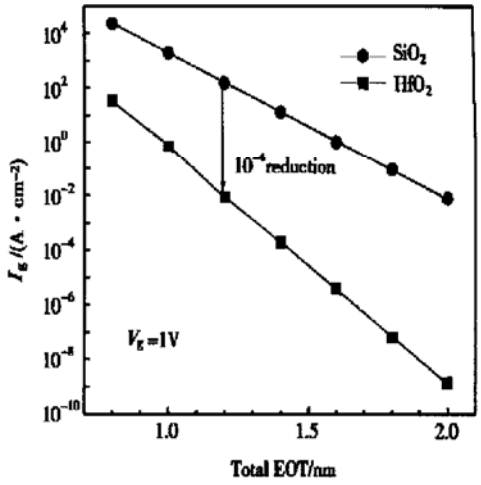

Fig. 6 Electron direct tunnelling currents at gate voltage of 1V through SiO_2 and HfO_2 as a function of equivalent oxide thickness

5 Conclusion

In conclusion, we present an efficient model for direct tunnelling current. Due to the valence band mixing, the hole quantization of the p-MOS devices is obtained by an improved one-band EMA in our model, which is more accurate than the traditional one-band EMA. The calculated results are in good agreements with the experiments. It is also found that the usual parabolic dispersion is applicable only for oxides thinner than about 2nm or in low gate voltage region. A Freeman-Dahlke dispersion is introduced in hole tunnelling current simu-

lation. It gives a better physical description reflecting the difference between the dielectric conduction band and valence band, while the simulation results have better agreement with the experiments.

References

[1] Taur Y, Buchanan D A, Chen W, et al. Proc IEEE, 1997, 85: 486

[2] Stern F. Phys Rev, 1972, B5: 4891

[3] Moglestue C. J Appl Phys, 1986, 59: 3175

[4] Van Dort M J, Woerlee P H, Walker A J. Solid-State Electron, 1994, 37: 411

[5] Ma Y, Liu L, Yu Z, et al. IEEE Trans Electron Devices, 2000, 47: 1303

[6] Ma Y, Li Z, Liu L, et al. Solid-State Electron, 2000, 44: 401

[7] Liu X, Kang J, Guan X, et al. Solid-State Electron, 2000, 44: 1435

[8] Weinberg Z A. J Appl Phys. 1982. 53: 5052

[9] Rana F, Tiwari S, Buchanan D A. Appl Phys Lett, 1996, 69: 1104

[10] Lo S H, Buchanan D A, Tauer Y, et al. IEEE Electron Lett, 1997, 18: 209

[11] Register L F, Rosenbaum E, Yang K. Appl Phys Lett, 1999, 74: 457

[12] Yang N, Henson W K, Hauser J R, et al. IEEE Trans Electron Device, 1999, ED-46: 1464

[13] Cai J, Sah C T. J Appl Phys, 2001, 89: 2272

[14] Wang J, Ma Y, Tian L, et al. Appl Phys Lett, 2001, 79: 1831

[15] Hu C Y, Banerjee S, Sadra K, et al. IEEE Electron Device Lett, 1996, 17: 276

[16] Ma Y, Liu L, Yu Z, et al. Solid-State Electron, 2000, 44: 1335

[17] Jallepalli S, Bude J, Shin W K, et al. IEEE Trans Electron Devices, 1997, ED-44: 297

[18] Rodriguez S, Lopez-Villaneuva J A, Melchor I, et al. J Appl Phys, 1999, 86: 438

[19] Shi Y, Ma T P, Prasad S, et al. IEEE Trans Electron Devices, 1998, ED-45: 2355

[20] Hou Y T, Li M F. IEEE Trans Electron Devices, 2001, ED-48: 2893

[21] Hou Y T, Li M F. IEEE Trans Electron Devices, 2001, ED-48: 1188

[22] Yang K N, Huang H T, Chang M C, et al. IEEE Trans Electron Devices, 2000, ED-47: 2161

[23] Freeman L B, Dahlke W E. Solid-State Electron, 1970, 13: 1483

[24] Zhu W, Ma T P, Tamagawa T, et al. IEDM Technical Digest, 2001

[25] Gusev E P, Buchanan D A, Cartier E, et al. IEDM Technical Digest, 2001

深亚微米 MOS 器件中栅介质层的直接隧穿电流*

侯永田[1] 李名复[1] 金 鹰[2]

(1 新加坡国立大学电子与计算机工程系 硅纳米器件实验室, 新加坡 119260)

(2 特许半导体制造有限公司, 新加坡 738406)

摘要: 在 WKB 近似的理论框架下, 提出了一个 MOS 器件中栅介质层直接隧穿电流的模型. 在这个模型中, 空穴量子化采用了一种改进的单带有效质量近似方法, 这种方法考虑了价带的混合效应. 通过与试验结果的对比, 证明了这个模型可以适用于 CMOS 器件中电子和空穴的隧穿电流. 还研究了介质层能隙中的色散对隧穿电流的影响. 这个模型还可以进一步延伸到对未来高介电常数栅介质层中隧穿电流的研究.

关键词: MOSFET; 直接隧穿电流; 量子效应; 栅介质层

EEACC: 2560R; 2560B

中图分类号: TN386 **文献标识码**: A **文章编号**: 0253-4177(2002)05-0449-06

* 新加坡国立大学(No. R263-000-077-112)和新加坡 NSTB(No. EMT/TP/00/001, 2)

2002-03-08 收到