Quantum Confinement and Electronic Properties of Silicon Nanowires

Xinyuan Zhao,[1] C. M. Wei,[1,2] L. Yang,[1] and M. Y. Chou[1]

[1]*School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332-0430, USA*
[2]*Institute of Physics, Academia Sinica, Nankang, Taipei, Taiwan 11529, Republic of China*

(Received 31 December 2003; published 11 June 2004)

We investigate the structural, electronic, and optical properties of hydrogen-passivated silicon nanowires along [110] and [111] directions with diameter d up to 4.2 nm from first principles. The size and orientation dependence of the band gap is investigated and the local-density gap is corrected with the *GW* approximation. Quantum confinement becomes significant for $d < 2.2$ nm, where the dielectric function exhibits strong anisotropy and new low-energy absorption peaks start to appear in the imaginary part of the dielectric function for polarization along the wire axis.

DOI: 10.1103/PhysRevLett.92.236805 PACS numbers: 73.21.Hb, 73.22.Dj, 78.67.Lt

Nanowires are one-dimensional nanostructures with electrical carriers confined in the other two (perpendicular) directions. They exhibit interesting physical properties that are noticeably different from those of quantum dots and the bulk. Of particular importance in technology are semiconducting nanowires that have potential applications in many fields such as optoelectronics, photovoltaic cells, and especially device miniaturization [1]. Recently, it has been possible to fabricate, for example, single-crystal silicon nanowires (SiNWs) with diameters as small as $\sim$1 nm and lengths of a few tens of micrometers [2–6]. Photoluminescence data revealed a substantial blueshift with decreasing size of nanowires [5,7,8]. Recent scanning-tunneling spectroscopy data [5,6] also showed a significant increase in the electronic energy gap for very thin semiconductor nanowires, explicitly demonstrating the quantum size effect.

Unlike other nanostructures such as quantum dots and nanotubes that have attracted considerable attention in recent years, relatively fewer theoretical investigations of quantum nanowires have been conducted to date. A conductance calculation for finite wires connected to aluminum electrodes has been reported [9], but most of the previous theoretical studies focused on interpreting the photoluminescence data of porous Si [7], and therefore concentrated on SiNWs oriented along the [100] direction with a hypothetical rectangular cross section. The diameters of the wires considered in previous first-principles studies [10–13] were $<$1.5 nm, too small to study the transition trend from the bulk. Studies using empirical or semiempirical methods have also been carried out [14–17]. In contrast, the SiNWs grown by recent experiments were mostly along the [110] direction [1,6], while wires along [112] were also observed [6]. Thus, the dependence of quantum confinement on the wire orientation can be uniquely investigated for these systems.

In this work, we study hydrogen-passivated SiNWs oriented along both [110] and [111] directions using the density-functional theory (DFT) in the local-density approximation (LDA) and the many-body perturbation method based on the *GW* approximation, focusing particularly on the electronic energy gaps, optical properties, and their dependence on wire diameter and orientation. SiNWs synthesized in experiments usually have a roughly cylindrical shape with chemical passivation on the surface. We construct our cylindrical wire models from the bulk and passivate all Si dangling bonds on the surface by H atoms in such a way that no complex of SiH_3 is present. Shown in Fig. 1 are the ball-and-stick models for the thinnest wire along the [110] and [111] orientations that we have studied.

Our calculations show that all [110] wires exhibit a "direct" fundamental band gap at $\bar{\Gamma}$ due to band folding, while the [111] wires exhibit a transition from an indirect gap in large wires to a direct one in small wires. The gap increases subquadratically as d decreases. More strikingly, when the diameter d becomes less than 2.2 nm the dielectric function exhibits strong anisotropy, and

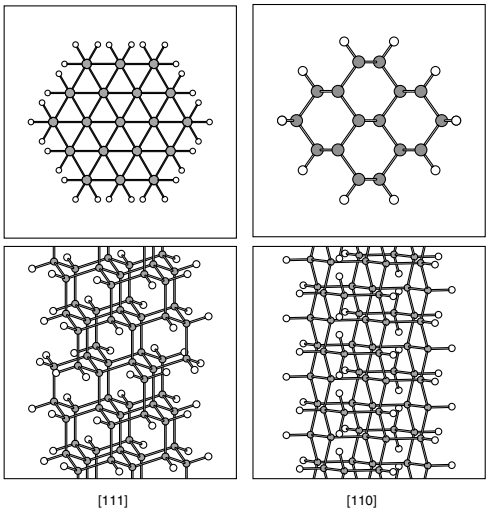

FIG. 1. Ball-and-stick models of silicon nanowires (SiNWs) along the [110] (right) and [111] (left) directions viewed from the top (upper panels) and the side (lower panels). Filled circles stand for Si atoms, and open circles for H atoms.

new absorption peaks show up in $\epsilon_2^{\parallel}(\omega)$, the imaginary part of the dielectric function along the wire axis.

Our calculations are carried out within DFT using norm-conserving pseudopotentials [18] with a plane-wave basis set. Periodic-boundary conditions are employed in the xy plane with supercells large enough to eliminate the interaction between neighboring wires. It is found that a 5-Å separation between two closest H atoms on neighboring wires is sufficient to make such interaction negligible (<1 meV per H atom on average). The energy cutoff for the plane waves is in the range of 15–20 Ry. The Monkhorst-Pack **k**-point meshes of $1 \times 1 \times 4$ to $1 \times 1 \times 8$ are found to provide sufficient accuracy in the calculation of total energies and forces. Summarized in Table I are the diameter of the wires and the number of atoms in the supercells used in our calculation.

We first relax the atomic configurations of our SiNWs, and the electronic structure is studied for the fully relaxed structures. Bulk Si is known to have an indirect band gap of 1.17 eV, with the conduction-band minima located at about 85% along Γ to X. Therefore, there are six equivalent conduction-band minima on $\pm x$, $\pm y$, and $\pm z$ axes, with a transverse mass (0.1905) much less than the longitudinal mass (0.9163). When [110] wires are formed, two of these minima on $\pm z$ will be projected onto $\bar{\Gamma}$ based on the effective-mass approximation, exhibiting both the large mass and the small mass in the confinement plane. The band edge associated with the large mass will be least upshifted due to confinement, giving rise to a direct gap. In contrast, the projection along the [111] direction is expected to produce an indirect gap in large [111] wires. Indeed, we do find an indirect gap for wires with diameter larger than 2 nm. However, the difference between the indirect and direct gaps is very small (less than 0.05 eV). Therefore, only the direct LDA energy gaps (E_g^{LDA}) are plotted in Fig. 2, which shows the size dependence of band gap for the [110] and [111] wires. Since the effective mass in the confinement plane for the [111] wires is smaller than the relevant counterpart (the longitudinal mass) for the [110] wires, the energy upshift is expected to be larger for the [111] wires. Hence, in addition to a size dependence, the energy gap also depends on orientation.

TABLE I. Diameter (d) and the number of Si and H atoms for the wires in our supercell calculation.

[110]	d (nm)	Si	H	[111]	d (nm)	Si	H
					0.92	38	30
	1.20	16	12		1.31	62	42
	1.60	42	20		1.69	110	54
	2.20	76	28		2.07	170	66
	2.60	110	36		2.46	218	78
	3.30	172	44		2.84	302	90
	4.02	250	52		3.23	398	102

It is well known that the Kohn-Sham energy gap is not the quasiparticle gap, and that the LDA gap is always smaller than the observed value. This can be corrected by evaluating the self-energy operator in the GW approximation [19]. We have calculated the GW quasiparticle gaps (E_g^{GW}) for the two thinnest [110] wires as well as for bulk silicon. Both the LDA gaps and the GW corrections greatly increase as d decreases. We fit these band gap values with a function of $E_{g,\mathrm{bulk}} + \mathrm{const} \times (1/d)^{\alpha}$, where $E_{g,\mathrm{bulk}}$ is the bulk gap value from LDA or GW, and α is found to be approximately 1.7. The value $\alpha = 2$ is expected using an effective-mass particle-in-a-box approach [16]. For bulk silicon, $E_g^{\mathrm{LDA}} = 0.58$ eV, indicated by the solid line in Fig. 2, and the GW correction is about 0.5 eV, which uplifts the gap to 1.08 eV. The GW corrections are 1.62 eV (= 3.12–1.50)and 1.29 eV (=2.32–1.03) for the two [110] wires with $d = 1.20$ and 1.60 nm, respectively. These are more than twice or 3 times the correction found in the bulk. An interesting finding here is that the self-energy correction, which increases monotonically with decreasing diameter, exhibits a rather strong size dependence also. In the past, this important variation has been neglected by postulating a size-independent constant correction that is usually obtained from the bulk [13]. This will inevitably introduce significant errors in the calculated optical gaps.

Also shown in Fig. 2 are the measured band gaps [6] for SiNWs along [112] and [110]. Since most of the measured gaps were for wires along [112] in this experiment, the comparison with our theoretical prediction (the dotted line) should be viewed with caution. The single point for a [110] wire agrees very well with our theoretical value. Still, the experimental data shows generally good

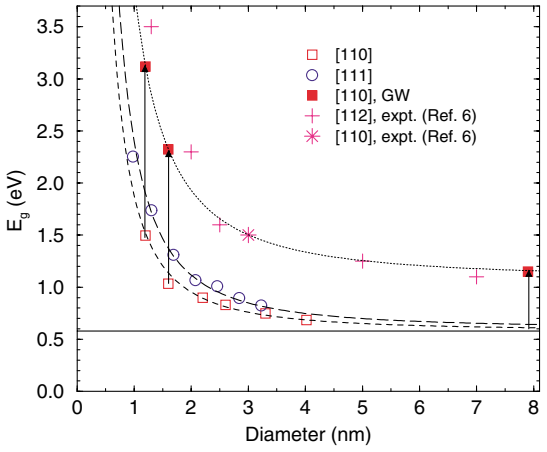

FIG. 2 (color online). LDA band gaps calculated for [110] (empty squares) and [111] (empty circles) wires, and the GW-corrected gaps (filled squares) for the two thinnest [110] wires and bulk Si, compared with the measured gaps for [112] wires (+) and a [110] wire (*). The dotted, dashed, and long-dashed lines are fitted to the data points (see text). The LDA band gap of bulk Si is indicated by the solid line, and the bulk GW gap is marked around $d = 8.0$ nm for convenience.

agreement with the *GW*-corrected curve, though for small wires the discrepancy becomes appreciable. This, in turn, confirms that the orientation dependence of band gaps become more significant in small wires, e.g., $d < 2.2$ nm in this case.

In Fig. 3, the characteristics of the electronic states in [110] wires is illustrated in a band-by-band fashion using the wire with $d = 1.2$ nm. (The structure of this wire is shown in Fig. 1.) The charge density is plotted in the xy plane after averaging along the axial direction. Shown are the results for the first six valence bands, the highest occupied band (band 38), and one band in between (band 18), as well as the lowest conduction band (band 39). The evolution can be semiquantitatively explained by the simple model of a particle confined in an infinite cylindrical potential well. If the radius of the cylindrical well is denoted by $r_0 = d/2$, the stationary solutions can be found analytically as $\psi_{nl}(r, \phi) \propto J_l(\beta_{nl} r/r_0)\cos(l\phi)$ or $J_l(\beta_{nl} r/r_0)\sin(l\phi)$, where $l = 0, 1, 2, \ldots$, and β_{nl} is the nth zero of the lth Bessel function $J_l(x)$ that determines the nodal structure in the radial direction. The eigenvalues are then $E_{nl} = \hbar^2\beta_{nl}^2/(2m^* r_0^2)$, where m^* is the effective mass. All energy levels except for $l = 0$ are therefore doubly degenerated. We find that the calculated charge density distribution for real wires indeed follows these features characterized by the quantum numbers $\{n, l\}$, as can be seen in Fig. 3. For instance, the values of $\{n, l\}$ for the first few bands in Fig. 3 are $\{1, 0\}$ for band 1, $\{1, 1\}$ for bands 2 and 3, $\{1, 2\}$ for bands 4 and 5, and $\{2, 0\}$ for band 6. For higher bands, such correspondence to the simple model becomes less precise.

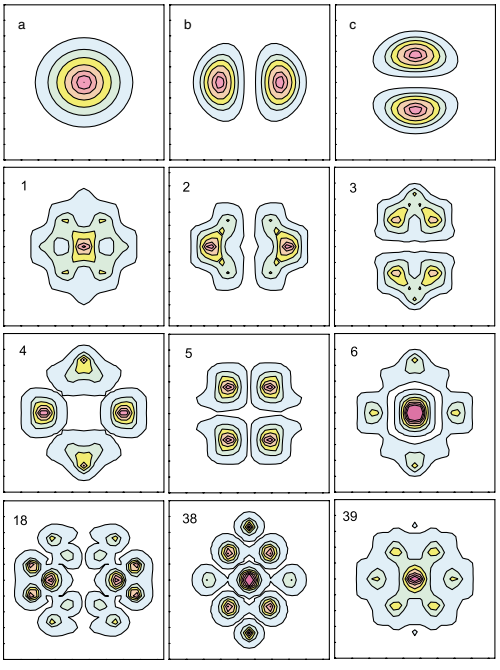

FIG. 3 (color online). Electronic charge distribution in the xy plane (integrated along the wire axis) for the [110] wire with $d = 1.20$ nm. The band indices are shown in each figure. The top panels marked with a, b, and c are the results of a simple model for bands 1, 2, and 3, respectively (see text). The high (low) density region is indicated by red/dark (blue/light).

The optical properties are studied by computing the complex dielectric function, $\epsilon(\omega) = \epsilon_1(\omega) + i\,\epsilon_2(\omega)$, in which the imaginary part $\epsilon_2(\omega)$ is closely related to the optical absorption at a given frequency ω. A knowledge of $\epsilon_2(\omega)$ over a wide frequency range allows one to obtain $\epsilon_1(\omega)$ using the Kramers-Kronig relation. $\epsilon_2(\omega)$ can be written as [20]

$$\epsilon_2(\omega) = \frac{4\pi^2 e^2 \hbar}{m^2\omega^2} \sum_{ij} \frac{2}{(2\pi)^3} \int_{\mathrm{BZ}} d\mathbf{k} |M_{ij}(\mathbf{k})|^2\, \delta[\omega - \omega_{ij}(\mathbf{k})],$$

where the integral is over all states in the Brillouin zone and the sum is over all combinations between the valence band i and the conduction band j, whose wave functions and energy levels are denoted as $\{\psi_i, E_i\}$ and $\{\psi_j, E_j\}$, respectively. The matrix element between bands i and j at $\mathbf{k}$ can then be written as $M_{ij}(\mathbf{k}) = \langle\psi_j(\mathbf{k})|\hat{\mathbf{e}}\cdot\mathbf{p}|\psi_i(\mathbf{k})\rangle$, and $\hbar\omega_{ij}(\mathbf{k}) = E_j - E_i$. Here $\hat{\mathbf{e}}$ and $\mathbf{p}$ denote the polarization vector and the momentum operator, respectively. Using our *GW*-corrected band gaps (Fig. 2), a scissor operator is applied to move all conduction bands up by the amount of the *GW* correction in the calculations of $\epsilon_2(\omega)$. The spin-orbit interaction and excitonic effect are not included in this calculation.

Figure 4 shows the calculated $\epsilon_2(\omega)$ for three thinnest wires along [110] and bulk Si. $\epsilon_2(\omega)$ are evaluated for the polarization along the axial (z) direction [$\epsilon_2^{\parallel}(\omega)$] and in the xy plane [$\epsilon_2^{\perp}(\omega)$]. Although the fundamental gap is direct in the wires, the matrix elements are still vanishingly small near the gap. Even for the smallest wire ($d = 1.2$ nm), the allowed dipole transition does not show up until 1 eV above the fundamental gap. It is ready to see that $\epsilon_2(\omega)$ demonstrates strong anisotropy in two polarization directions for $d < 2.2$ nm, where the overall spectrum of $\epsilon_2^{\perp}(\omega)$ is shifted to higher frequency compared to $\epsilon_2^{\parallel}(\omega)$, and the relative shift becomes larger as d decreases. At $d = 2.2$ nm, $\epsilon_2^{\parallel}(\omega)$ and $\epsilon_2^{\perp}(\omega)$ almost merge together and exhibit features in bulk Si except for an energy shift. Similar anisotropy is also found in a previous semiempirical tight-binding study [14] for a Si wire of 0.77 nm oriented in the [100] direction. Here we confirm this anisotropy from first-principles studies and illustrate its evolution as the wire diameter decreases. Our calculation predicts that a size of about 2 nm or smaller is needed to exhibit the anisotropic absorption in SiNWs.

Another interesting feature in Fig. 4 is that the absorption peak in $\epsilon_2^{\parallel}(\omega)$ associated with the main peak (black dots) in bulk Si seems to remain approximately 3.0 eV above E_g^{GW} in the nanowires, while the absorption edge (vertical dashed lines) moves toward E_g^{GW} (arrows) as d decreases. The latter is a consequence of mixing the bulk states in finite-sized wires, leading to an enhancement of

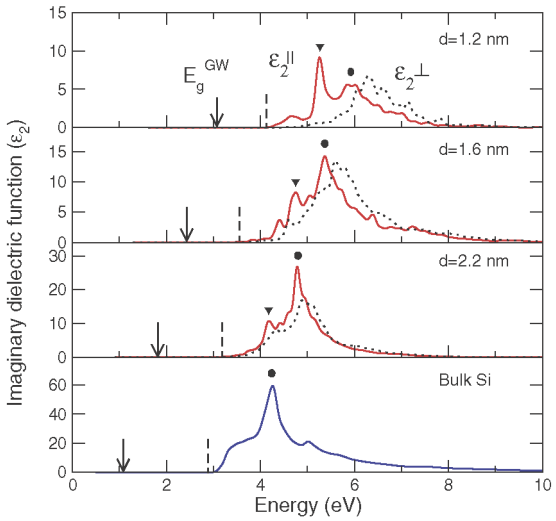

FIG. 4 (color online). Imaginary dielectric functions $\epsilon_2(\omega)$ polarized along the z direction [$\epsilon_2^{\parallel}(\omega)$, solid lines] and in the xy plane [$\epsilon_2^{\perp}(\omega)$, dotted lines], for silicon [110] wires with $d =$ 1.2, 1.6, and 2.2 nm, respectively. Shown in the bottom panel is $\epsilon_2(\omega)$ for bulk c-Si. The dielectric functions are calculated with the scissor operator to fix the band gap at the GW values (E_g^{GW}) and the energy zero is set at the top of the valence band. The arrows and vertical dashed lines mark E_g^{GW} and the optical absorption edges, respectively. The black dots indicate the original absorption peak in bulk Si, and the inverted triangles show the new peaks developed in nanowires.

the dipole transition. This also gives rise to other low-energy peaks in smaller and smaller wires, including a major peak (triangles) that is about 0.7 eV below the original peaks (black dots) and becomes the most prominent one for $d = 1.2$ nm. In a previous theoretical study conducted for Si [001] wires ($d < 1.56$ nm, Ref. [11]), certain peaks were found not to change with the system size and were assigned to bulklike excitations. After including the size-dependent quasiparticle gap correction, we did not find such features in our studies.

Recently, the energetics of silicon nanowires and possible ground-state structures have been discussed [21]. In this work we have adopted a wire geometry with a bulklike structure in the interior, as observed in experiments [1,6]. The energetics of the nanowires can be investigated through the energy difference between the nanowires and the bulk with the same number of atoms N. Apart from a constant term which represents the energy of the edges, we find that the energy difference between the wire and the bulk is linear in $N^{1/2}(\sim d)$ (with a linear correlation factor of 0.99). This is consistent with a regular surface energy contribution to the total energy of the wires.

In conclusion, we have studied from first principles the structures of SiNWs oriented in the [110] and [111] directions and their electronic and optical properties as a function of diameter. These properties are strongly influenced by quantum confinement and the electronic states can be described by those associated with a cylindrical potential well. Direct fundamental band gaps are found at $\bar{\Gamma}$ for [110] and small [111] wires, which increase subquadratically as diameter shrinks. It is also found that [111] wires have overall a larger gap than [110] wires, as expected from the effective-mass difference. The underestimation of band gaps by LDA is rectified by the GW self-energy correction, which turns out to be surprisingly sensitive to wire size. In addition, anisotropy in the dielectric function is found in the wires with $d \leq 2.2$ nm, where extra peaks in the absorption coefficient also start to develop.

This work is supported by the National Science Foundation (Grants No. DMR-02-05328 and No. SBE-01-23532) and the National Science Council of Taiwan (Grant No. 92-2112-M001-052). The computation used resources of the National Energy Research Scientific Computing Center (NERSC), which is supported by the U.S. Department of Energy (Grant No. DE-AC03-76SF00098). We acknowledge the use the PARATEC package at NERSC.

[1] Y. Cui and C. M. Lieber, Science **291**, 851 (2001); X. Duan, Y. Huang, Y. Cui, J. Wang, and C. M. Lieber, Nature (London) **409**, 66 (2001).
[2] A. M. Morales and C. M. Lieber, Science **279**, 208 (1998).
[3] J. D. Holmes *et al.*, Science **287**, 1471 (2000).
[4] Y. Cui *et al.*, Appl. Phys. Lett. **78**, 2214 (2001).
[5] D. Katz *et al.*, Phys. Rev. Lett. **89**, 086801 (2002).
[6] D. D. D. Ma *et al.*, Science **299**, 1874 (2003).
[7] L. T. Canham, Appl. Phys. Lett. **57**, 1046 (1990).
[8] X. Duan *et al.*, Appl. Phys. Lett. **76**, 1116 (2000).
[9] U. Landman *et al.*, Phys. Rev. Lett. **85**, 1958 (2000).
[10] A. J. Read *et al.*, Phys. Rev. Lett. **69**, 1232 (1992).
[11] F. Buda *et al.*, Phys. Rev. Lett. **69**, 1272 (1992).
[12] M. S. Hybertsen and M. Needels, Phys. Rev. B **48**, 4608 (1993).
[13] B. Delley and E. F. Steigmeier, Appl. Phys. Lett. **67**, 2370 (1995).
[14] G. Sanders and Y. Chang, Phys. Rev. B **45**, 9202 (1992).
[15] C. Delerue *et al.*, Phys. Rev. B **48**, 11 024 (1993).
[16] C.-Y. Yeh *et al.*, Phys. Rev. B **50**, 14 405 (1994).
[17] J. Xia and K. Cheah, Phys. Rev. B **55**, 15 688 (1997).
[18] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[19] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986), and references therein.
[20] M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors* (Springer-Verlag, New York, 1988), 2nd ed.
[21] Y. Zhao and B. I. Yakobson, Phys. Rev. Lett. **91**, 35501 (2003).