Electron-spin-resonance transistors for quantum computing in silicon-germanium heterostructures

Rutger Vrijen,[1] Eli Yablonovitch,[1] Kang Wang,[1] Hong Wen Jiang,[2] Alex Balandin,[3] Vwani Roychowdhury,[1] Tal Mor,[1] and David DiVincenzo[4]

[1]*Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, California 90024*

[2]*Department of Physics, University of California, Los Angeles, Los Angeles, California 90024*

[3]*Department of Electrical Engineering, University of California, Riverside, Riverside, California 92521*

[4]*IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598*

(Received 17 September 1999; published 13 June 2000)

We apply the full power of modern electronic band-structure engineering and epitaxial heterostructures to design a transistor that can sense and control a single-donor electron spin. Spin-resonance transistors may form the technological basis for quantum information processing. One- and two-qubit operations are performed by applying a gate bias. The bias electric field pulls the electron wave function away from the dopant ion into layers of different alloy composition. Owing to the variation of the g factor (Si:$g=1.998$,Ge:$g=1.563$), this displacement changes the spin Zeeman energy, allowing single-qubit operations. By displacing the electron even further, the overlap with neighboring qubits is affected, which allows two-qubit operations. Certain silicon-germanium alloys allow a qubit spacing as large as 200 nm, which is well within the capabilities of current lithographic techniques. We discuss manufacturing limitations and issues regarding scaling up to a large size computer.

PACS number(s): 03.67.Lx, 85.30.Wx, 76.30.−v

I. INTRODUCTION

The development of efficient quantum algorithms for classically hard problems has generated interest in the construction of a quantum computer. A quantum computer uses superpositions of all possible input states. By exploiting this quantum parallelism, certain algorithms allow one to factor [1] large integers with astounding speed, and rapidly search through large databases [2], and efficiently simulate quantum systems [3]. In the nearer term such devices could facilitate secure communication and distributed computing.

In any physical system, bit errors will occur during the computation. In quantum computing this is particularly catastrophic, because the errors cause decoherence [4], and can destroy the delicate superposition that needs to be preserved throughout the computation. With the discovery of quantum error correction [5] and fault-tolerant computing, in which these errors are continuously corrected without destroying the quantum information, the construction of a real computer has become a distinct possibility.

Even with the use of fault-tolerant computing, a quantum computer engineer would still prefer a system that exhibits the smallest possible error rate on the qubits, the two-level systems that hold the quantum information. In fact, Preskill [6] (in a review of the subject) presented a requirement for fault tolerance: the ratio of the error rate to the computer clock rate has to be below a certain threshold.

Several systems have recently been proposed to obtain a physical implementation of a quantum computer. These systems include cold ion traps [7], nuclear magnetic resonance (NMR) systems [8,9], all-optical logic gates [10,11], Josephson junctions [12], and semiconductor nanostructures [13]. Successful experimental demonstrations of one- and two-qubit computers were reported for trapped-ion systems [14] and NMR systems [15].

Last year, Kane [16] proposed a very interesting and elegant design for a spin-resonance transistor (SRT). He proposed to use the nuclear spins of ^{31}P dopant atoms, embedded in a silicon host, as the qubits. At low temperatures the dopant atoms do not ionize, and the donor electron remains bound to the ^{31}P nucleus. The control over the qubits is established by placing a gate electrode, the so-called *A* gate, over each qubit. By biasing the *A* gate, one can control the overlap of the bound electron with the nucleus and thus the hyperfine interaction between nuclear spin and electron spin, which allows controlled one-qubit rotations. A second attractive gate, a *J* gate, decreases the potential barrier between neighboring qubits, and allows two nuclear spins to interact by electron-spin exchange, which provides the required controlled qubit-qubit interaction.

The rate of loss of phase coherence between qubits in a quantum system is typically characterized by the dephasing time T_2. The T_2 dephasing time of the nuclear spins in silicon is extremely long. The silicon host efficiently isolates the nuclear spins from disturbances [17]. A quantum computer based on semiconductors offers an attractive alternative to other physical implementations due to compactness, robustness, the potentially large number of qubits [18], and semiconductor compatibility with industrial scale processing. However, the required transistors are very small, since their size is related to the size of the Bohr radius of the dopant electron. Furthermore, after the calculation is completed Kane's SRT requires a sophisticated spin transfer between nuclei and electrons to measure the final quantum state.

We suggest using the full power of modern electronic band-structure engineering and epitaxial growth techniques, to introduce a more practical field-effect SRT transistor design that might lend itself to a near-term demonstration of qubits on a silicon wafer. We alter Kane's approach by the implementation of these spin-resonance transistors in engi-

neered germanium/silicon heterostructures that have a controlled band structure. Si-Ge strained heterostructures, developed by IBM and other companies, are in the mainstream of silicon technology, and are currently used for high-frequency wireless communication transistors and high-speed applications.

In Si-Ge heterostructure layers, we can control the effective mass of the donor electron to reduce the required lithographic precision, and to permit the SRT transistors to be as large as ≈ 2000 Å. The Bohr radius of a bound electron in Si-Ge can be much larger than in silicon due to the very small effective mass in strained Si-Ge alloys, and their higher dielectric constant. This places the lithographic burden well within the practical range of electron beam lithography, and almost within range of contemporary optical lithography.

Among the other simplifications, we will employ an electron spin rather than a nuclear spin as the qubit. Owing to the difference in the electronic g factor, $g=1.998$ for Si and $g=1.563$ for Ge, the electron-spin-resonance transition can be readily tuned by an electrostatic gate on a compositionally modulated Si-Ge epilayer structure. By working with electron spins rather than nuclear spins, we avoid the requirement of a sophisticated spin transfer between electrons and nuclei, for read-in–read-out of quantum data and for the operation of two-qubit gates. In addition, due to their higher Zeeman energy, electron spins will eventually permit a clock speed up to 1 GHz compared to a speed of ≈ 75 kHz projected for the nuclear spins. Likewise, isotopic purity is not critical for electron spins.

In order to read out the final result of a quantum calculation we will need to be able to detect single electron charges. Individual electrostatic charges are readily detected by conventional field-effect transistors (FET's) at low temperatures, which obviates the need for the sophisticated single-electron transistors. In this paper, we illustrate our design for an electron-spin-resonance (ESR) transistor.

II. ELECTRON-SPIN DEPHASING TIME IN SILICON AND GERMANIUM

Electron spins benefit from the same protective environment provided by the silicon host as nuclear spins. Indeed, the ESR line in doped silicon at low temperatures turns out to be exceptionally clean and narrow compared to other ESR lines.

Feher and co-workers [19–21] found that the Si:^{31}P ESR line is inhomogeneously broadened by hyperfine interactions with neighboring nuclear spins. But the nuclear spin flip T_1 relaxation times were measured [20] to be in the 1–10-h range. Thus the nuclei can be regarded as effectively static on the time scales needed for quantum computing. Likewise the direct electron spin-flip T_1 is also around [20] an hour.

On the question of the critical transverse T_2 ESR dephasing linewidth there was only a little information. Feher and Gere studied some heavily doped *n*-Si:P samples, and found that the ESR linewidth actually narrowed [22] at high doping, down to a 1-MHz linewidth at the 9-GHz ESR frequency, for the heavy doping level $n=3\times 10^{18}/\mathrm{cm}^3$. This unusual behavior was clearly the result of exchange narrowing of the hyperfine inhomogeneity. For quantum computing, the issue is the linewidth of a single-electron-spin transition, rather than a heavily doped inhomogeneous ensemble.

TABLE I. Conduction-band effective masses relative to m_0, and the corresponding Bohr radii and g factors in Si and Ge.

Material	ϵ	m_{xy}	m_z	$a_{B,xy}$	$a_{B,z}$	$g_\parallel$	$g_\perp$
Germanium	16	0.082	1.59	64 Å	24 Å	0.823	1.933
Silicon	12	0.191	0.916	25 Å	15 Å	1.999	1.998

Thus the outlook was optimistic. If the linewidth is only 1 MHz at such a high doping level, and is due to exchange with neighboring electrons, then the linewidth would surely be much narrower at lower doping levels, and especially for one isolated electron. Indeed that was confirmed by Chiba and Hirai [23], who measured a $1/2\pi T_2$ linewidth of only ≈ 1 kHz at a doping of 10^{16} phosphorus ions per cm^3, by the very reliable spin-echo technique. The residual linewidth was interpreted as being due to spin diffusion via the nuclear spins. Indeed the linewidth was shown [24] to narrow further in isotopically purified, zero-spin, Si^{28}, making the T_2 dephasing even slower. The observed 1-kHz linewidth at $n=10^{16}/\mathrm{cm}^3$ is already narrow enough, in relation to the 9-GHz ESR frequency, to allow enough operations for fault tolerant computing [6].

In germanium the dominant mechanism for spin dephasing is quite different from the one in silicon. Theory [25,26] and experiment [27] confirmed that the dominant relaxation in germanium is through acoustic disturbances of the spin-orbit coupling. The g factor in germanium is much different from 2, the free-electron value, because of the relatively strong spin-orbit coupling. Germanium has four ellipsoidal conduction-band minima, which are aligned with the $\langle 111\rangle$ directions. In each minimum, the effective mass depends on the direction of electron motion, with a low effective mass (m_{xy}) in the transverse direction and a high effective mass in the longitudinal direction (m_z) (see Table I). The anisotropic effective mass results in an anisotropic g factor, with $g=g_\parallel$ for magnetic-field components in the $\langle 111\rangle$ direction, and $g=g_\perp$ for magnetic-field components perpendicular to this direction. For arbitrary angles ϕ between the magnetic field and the $\langle 111\rangle$ direction, the g factor is given by

$$g^2=g_\parallel^2\cos^2\phi+g_\perp^2\sin^2\phi. \tag{1}$$

The electronic ground state of the donor atom is an equal superposition (singlet) state of the four equivalent conduction-band minima, and therefore has an isotropic g factor: $g=g_\parallel/3+2g_\perp/3=1.563$. However, in the presence of lattice strain, the energies of the conduction-band minima shift with respect to each other. In the new donor ground state, probability is shifted among the four valleys, with some valleys more populated than others. This produces a shift Δg in the g factor, since each valley forms a different angle ϕ with the static magnetic field B. The corresponding

relative energy shift of the spin states is proportional to $(\Delta g)\mu B$, with μ the Bohr magneton. At finite temperatures, acoustic phonons cause time-varying strains with a finite power density at the spin transition energy, which induce spin-lattice relaxation.

At these temperatures it follows from this theory that the phase relaxation time is of the same magnitude as the population relaxation time $T_2 \approx T_1$. Experiments have shown that T_1 is around 10^{-3} sec for germanium at 1.2 K. We are not aware of direct measurements of T_2 by electron-spin-resonance experiments similar to those that were done in silicon. Unless there are other, as yet unknown, T_2 mechanisms in germanium, T_2 will be determined by acoustic vibrations and be of the order of 10^{-3} seconds, which is equal to the best measured T_2 in silicon, and is again sufficiently long to allow fault tolerant computing.

Several mechanisms could lead to a further improvement in the T_1 and T_2 caused by acoustic vibrations. First, working at lower temperatures will reduce the phonon energy density, which is proportional to T^4. Second, for the two orientations of germanium that we propose to use, $\langle 111 \rangle$ and $\langle 001 \rangle$, some special considerations can make the expected lifetimes longer. For germanium grown with strain in the $\langle 111 \rangle$ direction, the conduction-band minimum along the growth direction has a significantly lower energy than the other three minima. In the electronic ground state, virtually all population resides in this minimum, and there is little coupling to the three split-off valleys. In the theory by Roth [25] and Hasegawa [26], this effect is accounted for by a square dependence of T_1 on the energy splitting between the electronic ground state and excited states (singlet-triplet splitting). The grown-in strain increases this splitting from 2 to 200 meV, with a corresponding increase in lifetime of 10^4. For germanium grown with strain in the $\langle 001 \rangle$ direction and with the magnetic field aligned with that direction, a symmetry argument forbids a strain-induced g shift: the $\langle 001 \rangle$ direction makes equal angles with all conduction-band minima, and therefore a probability redistribution among these minima does not affect the g factor, as can be seen from Eq. (1). Thus further improvements in the already acceptable lifetimes appear possible.

The ESR of a bound donor in a semiconductor host provides many advantages: First, in a magnetic field of 2 T, the ESR frequency is ≈ 56 GHz, easily allowing qubit operations at up to ≈ 1 GHz. This is comparable to the clock speed of ordinary computers, and is consistent with the precision of electronic control signals that are likely to be available. Second, at temperatures well below 1 K, the electron spins are fully polarized, allowing a reproducible starting point for the computation. Finally, for electron spins isotopic purity is not compulsory, since the nuclear-spin inhomogeneity remains frozen at low temperatures.

III. SRT TRANSISTOR SIZE AND LITHOGRAPHIC CRITICAL DIMENSION

The Bohr radius of the bound carrier wave function regulates the size scale of spin-resonance transistors. In semiconductors the Bohr radius is much larger than in vacuum,

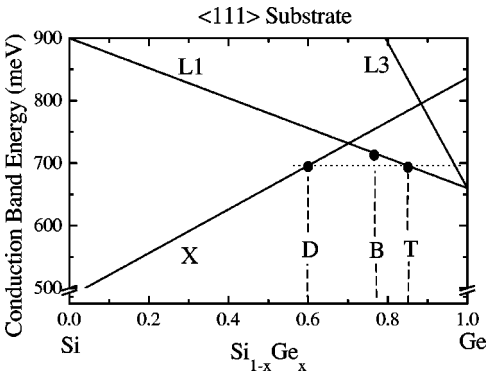

FIG. 1. The conduction-band energy in Si-Ge alloys, compositionally strained in the $\langle 111 \rangle$ direction, from neutral strain at 100% Ge. The X valley has six minima that remain degenerate. The L valley has four minima that are split between $L1$ and $L3$. The conduction band changes from X to $L1$ character at a composition of $Si_{0.3}Ge_{0.7}$. At this band transformation, the xy effective mass becomes relatively light, the Bohr radius increases, and the g factor drops from $g \approx 1.998$ to $g = g_{\parallel} \approx 0.823$. The fractional compositions D, T, and B will be used in our band-structure-engineered spin-resonance transistor.

since the Coulomb force is screened by the dielectric constant, and the effective mass is much smaller. Thus the bound carrier roams farther. The Bohr radius is $a_B = \epsilon(m_0/m^*)(\hbar^2/m_0 q^2)$ in the semiconductor, where m^*/m_0 is the effective mass relative to the free-electron mass, ϵ is the dielectric constant, $\epsilon = 16$ for Ge, $\epsilon = 12$ for Si, and the quantity in parentheses is the Bohr radius in vacuum.

It is common in Si-Ge alloys to have strain available as an engineering parameter. Strain engineering of valence-band masses has been very successful, and is used [28] in virtually all modern semiconductor lasers. As discussed above, in the conduction band, strain splits the multiple conduction-band valley energies, allowing one valley to become the dominant lowest-energy conduction band. If that valley also happens to be correctly aligned, the donor wave functions can have a low mass moving in the plane of the silicon wafer, and a high mass perpendicular to the wafer surface. That is exactly what we are looking for in spin-resonance transistors. We want large wave functions in the directions parallel to the wafer surface, in order to relax the lithographic precision that would have been demanded if the Bohr radius were small.

In Si-rich alloys there are six conduction band minima, in the six cubic directions, that are frequently labeled as the X directions. In Ge-rich alloys, there are four conduction minima located at the $\langle 111 \rangle$ faces of the Brillouin zone, labeled L. The Ge-rich case is particularly interesting, since it has a conduction-band mass of only $0.082 m_0$ in the transverse direction.

Under $\langle 111 \rangle$ strain the four conduction-band valleys split, so that one of them is lowest in energy and is labeled $L1$. The other three valleys remain degenerate and are labeled $L3$. Figure 1 shows the conduction-band structure in the Si-Ge alloys, grown compositionally strained in the $\langle 111 \rangle$ direction, with neutral strain at 100% Ge, as adapted from a more complete set of band structures from Ref. [29]. The hydrogenic Schrödinger equation for anisotropic effective

mass m_{xy} in the plane of the wafer, and m_z perpendicular to the plane of the wafer, was solved, for arbitrary values of m_{xy}/m_z by Schindlmayr [30].

The Bohr radius in the xy plane is influenced by both effective masses

$$a_{B,xy}=\frac{2\epsilon}{3\pi}\frac{2+(m_{xy}/m_z)^{1/3}}{m_{xy}}a_B^0, \tag{2}$$

with a_B^0 the Bohr radius of a free hydrogen atom and $m_{xy} \ll m_z$ assumed, as is appropriate for the z-oriented Si and Ge conduction-band ellipsoids. The Bohr radius in the heavy mass direction, $a_{B,z}$ is given by $a_{B,z}=(m_{xy}/m_z)^{1/3}a_{B,xy}$. Using the actual masses and the exact formula [30], we give the Bohr radii in Si and Ge for z-oriented conduction-band ellipsoids in Table I.

In Table I, special note should be taken of the Bohr radius of 64 Å for $\langle 111\rangle$ strained Ge-rich alloys in which the $L1$ band minimum forms the conduction band. At that orientation, the X-band minima in Si-rich alloys would have a Bohr radius of only ≈ 20 Å. Thus we achieve over a factor 3 increase in the transistor spacing by using a Ge-rich layer.

Given that the exchange interaction is a dominant influence among the donor spins, we make the point that Preskill's decoherence criterion can be redefined [31] as the on-off ratio of the spin-spin interaction, as induced by the transistor gates. The actual required transistor spacing is set by the need for the weakest possible exchange interaction when the two-qubit interaction is off, and a strong exchange interaction when two-qubit interactions are turned on. The exchange energy $4J$ between hydrogenic wave functions determines both time scales:

$$\frac{4J(r)}{h}\approx 1.6\frac{q^2}{\epsilon a_B}\left(\frac{r}{a_B}\right)^{5/2}\exp\frac{-2r}{a_B}. \tag{3}$$

If we require the exchange energy in the off state to be less than the measured [23], T_2 dephasing linewidth ≈ 1 kHz, then the donor ions would have to be about 29 Bohr radii apart, allowing a spacing of about 2000 Å. Such critical dimensions are well within the range that can be produced by electron-beam lithography.

Later we will show that by gate-controlled Stark distortion of the hydrogenic wave functions, the Bohr radius can be further increased, switching on the two-qubit interactions. Thus band-structure engineering allows us to use only one electrostatic gate to control both one- and two-qubit operations, rather than two separate A and J gates as required by Kane. This reduction of the number of gates by a factor of 2, though not essential for the operation of the our ESR, means that all lithographic dimensions are doubled, which significantly increases the manufacturability of the device.

IV. GATE-CONTROLLED SINGLE-QUBIT ROTATIONS IN THE SPIN-RESONANCE TRANSISTOR

The essence of a spin-resonance transistor qubit is that a gate electrode should control the spin-resonance frequency.

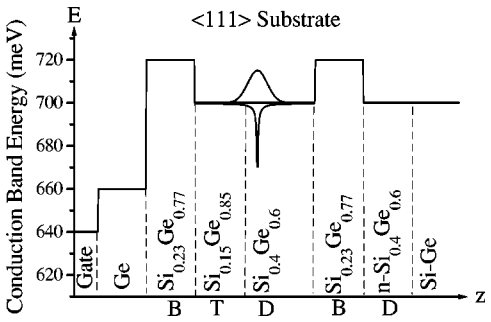

FIG. 2. The band-structure diagram for the proposed spin-resonance transistor, showing the Coulombic potential well of the donor ion in the $Si_{0.4}Ge_{0.6}$ D layer, where the conduction-band minimum is X like. The hydrogenic wave function partly overlaps the $Si_{0.15}Ge_{0.85}$ T layer, where the conduction-band minimum is L like. The donor electron is confined by the two $Si_{0.23}Ge_{0.77}$ B-barrier layers. The epilayer thicknesses are not to scale.

By tuning this frequency with respect to the frequency of a constant radiation field, that is always present while the computer is being operated, single-qubit rotations can be readily implemented on the electron spin. A band-structure diagram for the SRT is shown in Fig. 2.

We rely on the difference in the electronic g factor, $g=1.998$, for Si-rich alloys, and $g=g_{\|}=0.823$ for Ge-rich alloys, strained in the $\langle 111\rangle$ direction. Thus the electron-spin resonance transition can be readily tuned by an electrostatic gate on a compositionally modulated Si-Ge epilayer structure, such as shown in Fig. 3. In a study of the composition dependence of the g factor in Si-Ge alloys, Vollmer and Geist [32] showed that the g factor is most influenced by the band-structure crossover from X to $L1$ at a composition of $Si_{0.3}Ge_{0.7}$, and hardly at all by compositional changes away from that crossover. The ^{31}P dopant atoms are positioned in the $Si_{0.4}Ge_{0.6}$ D layer, a composition which is to the left of the crossover in Fig. 1. By electrostatically attracting the electron wave function into the $Si_{0.15}Ge_{0.85}$ T layer, the spin resonance can be tuned very substantially.

The two barrier layers of composition $Si_{0.23}Ge_{0.77}$, labeled B in Fig. 2, have a conduction-band structure as indi-

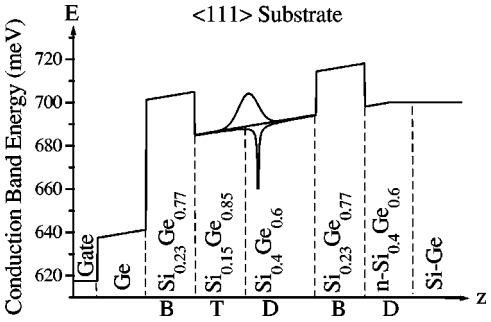

FIG. 3. The donor electron wave function is electrostatically attracted toward the $Si_{0.15}Ge_{0.85}$ T layer where the conduction band minimum is $L1$ like. There it will experience a smaller g factor, that is gate tunable. The actual g factor will be a weighted average between the D and T layers.

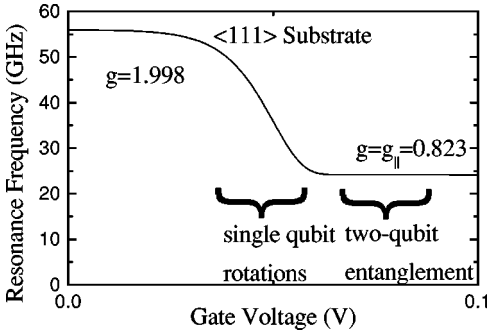

FIG. 4. A schematic of the dependence of the spin-resonance frequency on the transistor gate voltage. As the electrons are pulled toward the positive gate electrode and into the more Ge-rich alloy compositions, the heterobarrier *B* layer prevents the donors from becoming completely ionized. At intermediate gate voltages, the *g* factor can be tuned from $g=1.998$ to 0.823. The frequencies on the vertical axis correspond to a magnetic field of 2 T. The two-qubit tuning range will be explained in Sec. V.

cated in Fig. 1. They have an $L1$-like conduction-band minimum, to the right of X-$L1$ band-structure crossover, and thus have the same *g* factor as the $Si_{0.15}Ge_{0.85}$ *T* layer. The purpose of the *B* layers is to confine the donor electrons and prevent them from tunneling away and becoming lost. The energy height of the barrier need only be comparable to the donor binding energy, ≈ 20 meV to fulfill this task. On the other hand the $Si_{0.4}Ge_{0.6}$ *D* layer and the $Si_{0.15}Ge_{0.85}$ *T* layer should have no energy barrier between them, so that the *g* factor can be freely tuned. Thus the *D* layer and the *T* layer are selected at compositions straddling the X-$L1$ crossover in Figure 1, so that their respective conduction band energies E_D and E_T are the same. A schematic tuning curve for our proposed spin-resonance transistor is shown in Fig. 4. As the spin-resonance transistors are tuned in and out of resonance with the radio-frequency field, the electron spin can be flipped, or subjected to a phase change.

The wave-function distortion during tuning is shown for the left side transistor in Fig. 5. The confinement barriers of

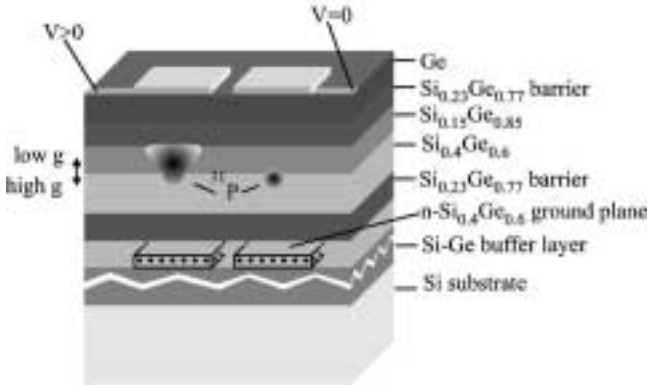

FIG. 5. The left transistor gate is biased $V>0$, producing single-qubit unitary transformations in the left SRT. The right gate is unbiased: $V=0$. The n-$Si_{0.4}Ge_{0.6}$ ground plane is counterelectrode to the gate, and it also acts as FET channel for sensing the spin.

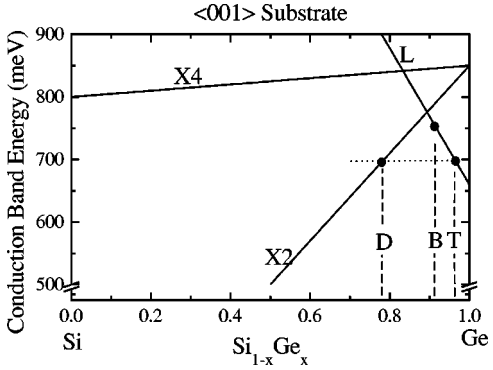

FIG. 6. The conduction-band energy in Si-Ge alloys, compositionally strained in the $\langle 001\rangle$ direction, from neutral strain at 100% Ge. The *L* valley has four minima that remain degenerate. The *X* valley has six minima along the cubic directions, that are split between $X4$ and $X2$. The compositions *D*, *T*, and *B* are much less strained than in the $\langle 111\rangle$ case, and allow for higher barrier heights to confine the dopant electron. For this crystal orientation, the *g* factor in the Ge-rich *T* and *B* layers is $g=1.563$.

composition *B* $Si_{0.23}Ge_{0.77}$, play an important role. They must confine the qubit donor electrons for long periods of time, or the carriers and their quantum information will be lost. For that purpose the *B*-barrier layers each need to be about 200 Å thick, for a carrier lifetime comparable to the ≈ 1-h T_1 spin-lattice relaxation for electron-spin flips. The two layers combined would total about 400 Å, well within the practical strain limit [33] of ≈ 1000 Å for growth of a 23% compositionally strained alloy. The *D* and *T* layers have thicknesses similar to the $a_{B,z}$ vertical Bohr radius, and contribute only slightly to the strain burden.

If one uses alloys grown in the $\langle 001\rangle$ direction instead, the numbers become even more favorable. Figure 6 shows the conduction-band structure in the Si-Ge alloys, grown in the $\langle 001\rangle$ direction [33], compositionally strained from neutral strain at 100% Ge. In this growth direction, the *L* band remains unsplit, and the *X* band splits up into a doubly degenerate $X2$ band and a quadruply degenerate $X4$ band. As can be seen, the conduction-band energy changes much more rapidly as a function of alloy composition for the $\langle 001\rangle$ growth direction. Moreover, the $X2$ and *L* bands cross over at approximately 90% Ge instead of 70% as in the Ge $\langle 111\rangle$ case. This allows us to select alloys with much lower strain, while obtaining a barrier height of 50 meV, more than twice the barrier height obtained in the $\langle 111\rangle$ direction. Consequently the layers can be made thinner, while still preventing tunneling of the dopant electron and the strain tolerance is significantly improved. The corresponding band-structure diagram for the $\langle 001\rangle$-oriented SRT is shown in Fig. 7. However, in the $\langle 100\rangle$ direction, the *g* factor is equal to the average value, $g=1.563$, so that the tuning range for the spin resonance frequency is less than in the $\langle 111\rangle$ case, as demonstrated in Fig. 8.

The use of the $\langle 001\rangle$ growth direction comes at the expense of an increased effective mass in the *xy* plane and a lighter mass in the *z* direction. The conduction-band ellipsoid pointing in the $\langle 111\rangle$ direction is 55° away from the $\langle 001\rangle$

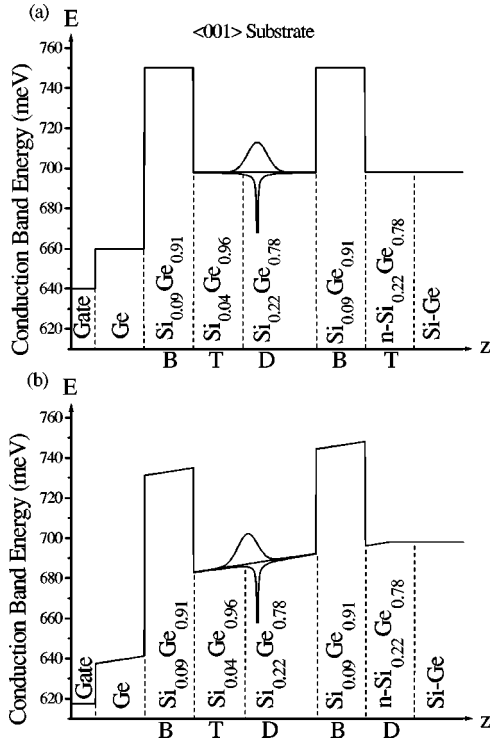

FIG. 7. The band-structure diagram for the spin-resonance transistor, with epilayers grown in the $\langle 001 \rangle$ direction. Both the unbiased (a) and the biased (b) cases are shown. The conduction-band energies allow the selection of layers with composition *D*, *T*, and *B*, such that the confining barrier height is increased to 50 meV, while the strain in the layers is reduced, compared to the $\langle 111 \rangle$ orientation. The epilayer thicknesses are not to scale.

direction, and thus the *z* direction no longer coincides with the heavy mass direction ($\langle 111 \rangle$). Some of the heavy mass is transferred into the *xy* plane, resulting in shorter Bohr radii. However, the lightest mass in Ge is equal to the heaviest mass in Si (see Table I). Therefore, the Ge-rich layer will always remain the layer with Bohr radii in the *xy* direction, which are at least as large as those in the Si-rich layer. Therefore Ge-rich layers will again perform the function of

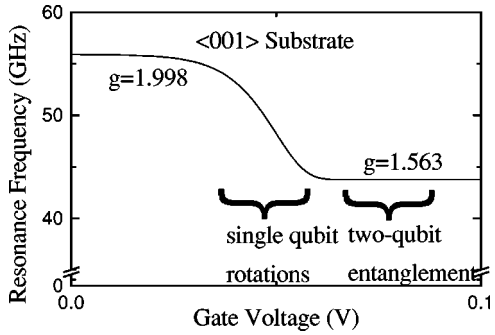

FIG. 8. A schematic of the dependence of the spin-resonance frequency on the transistor gate voltage for the case of a $\langle 001 \rangle$ substrate. The static magnetic field is in the $\langle 001 \rangle$ direction and has a strength of 2 T. The tuning range is reduced in this growth direction with respect to the $\langle 111 \rangle$ case, because the *g* factor in the Ge-rich layer is different: $g = 1.563$.

the tuning *T* layer, and the barrier *B* layer for structures grown in the $\langle 001 \rangle$ as they did for the $\langle 111 \rangle$ direction.

V. TWO-QUBIT INTERACTIONS

The spin-resonance transistors must be spaced far enough apart, that they will not produce phase errors in one another. At the same time it is necessary to allow wave-function overlap for the exchange interaction to activate the two-qubit interactions. These are needed to produce for example a ''controlled NOT'' (CNOT) gate, which is required to build a universal set of quantum logic gates. To achieve this we rely on our ability to tune the Bohr radius of the donors in the *xy* direction parallel to the semiconductor surface.

The Bohr radius a_B of a hydrogenlike donor increases with decreasing binding energy. A famous example is excitons confined in a two-dimensional (2D) flat quantum well: The excitonic binding energy is four times greater [34] than it would be in three dimensions. The reason is that spatial confinement forces the electron to spend more time near the positive charge, and it experiences tighter binding. Accordingly the Bohr radius is diminished. For the same reason, confinement by heavy mass in the *z* direction reduces the Bohr radius in the *xy* plane, as can be seen from Eq. (4). Without this reduction the effective mass in the *xy* direction in strained $\langle 111 \rangle$ Ge would even be higher.

Our technique for two-qubit interactions does not require any *J* gates. By increasing the gate voltage, we pull the electron wave function away from the positive ion, to reduce the binding energy, and increase the wave-function overlap between electrons bound to neighboring dopant ions. As shown in Fig. 3, the electrons can be electrostatically attracted to one of the barriers formed by the $Si_{0.23}Ge_{0.77}$ *B*-composition layer, forming a type of modulation-doped channel in the *xy* plane. The binding energy to the positive ions is greatly weakened, since the electrons are spending most of their time near the $Si_{0.23}Ge_{0.77}$ *B* barrier. Consequently the Coulomb potential becomes weakened to the form

$$V = -\frac{1}{4\pi\epsilon_0\epsilon}\frac{q}{\sqrt{r^2+d^2}}, \tag{4}$$

where $r^2 = x^2 + y^2$ is the horizontal distance from the donor ion, squared, and *d* is the vertical spacing from the barrier to the donor ion, and *q* is the electronic charge. Thus by adjusting the vertical depth of the ion, *d*, the Coulomb potential can be made as weak as desired. The weak Coulomb binding energy implies a large Bohr radius. The large radius permits a substantial wave-function overlap in the *xy* plane along the *B*-barrier layer, and a substantial two-qubit exchange interaction. It should be possible to tune from negligible exchange interaction, all the way to a conducting metallic 2D electron gas, by adjusting the vertical spacing *d*. As the electrons overlap, they will interact through the exchange interaction. It was already shown by DiVincenzo [35], that the exchange interaction can produce CNOT quantum gates.

The gate bias voltage range for two-qubit entanglement is indicated by the second curly bracket in Fig. 4. That voltage range attracts the electrons away from the positive ions and

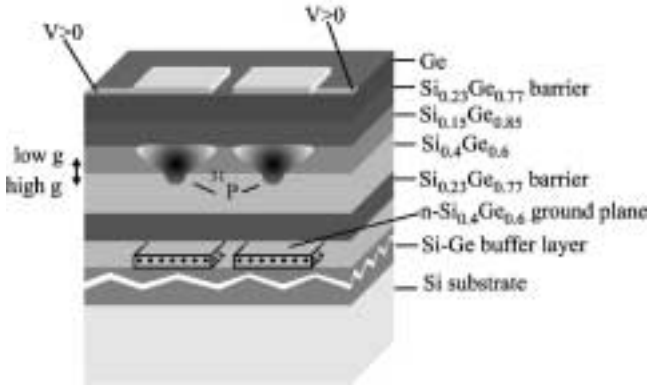

FIG. 9. Attracting the electrons to the $Si_{0.23}Ge_{0.77}$ *B* barrier reduces their Coulomb binding energy and increases their wave function overlap, allowing two-qubit interaction.

toward the $Si_{0.23}Ge_{0.77}$ *B* barrier, thus increasing their wavefunction overlap. In the midvoltage range, the first curly bracket in Fig. 4, one-qubit rotations take place. Thus both one- and two-qubit interactions can be controlled by a single gate. Gate tuning of a two-qubit exchange interaction is illustrated in Fig. 9.

VI. DETECTION OF SPIN RESONANCE BY A FET TRANSISTOR

It is a truism of semiconductor electronics that we need crystals of high perfection and extraordinary purity. Semiconductor devices are very sensitive to the presence of chemical and crystallographic faults down to the level of 10^{11} defects/cm3 in the volume, and 10^{8} defects/cm2 on the surface. Such defect concentrations are far below the level of sensitivity of even the most advanced chemical analytical instruments. These imperfections influence the electrical characteristics of semiconductor devices, as they vary their charge states. Thus conventional electronic devices are sensitive to very low concentrations of defects.

The detection sensitivity becomes particularly striking when the electronic devices are very tiny, as they are today. If electronic devices are small enough, then there is a good probability that not even one single defect might be present in, or on, the device. That helps define the potential yield of essentially perfect devices. But if a defect were to be present, it would have an immediate effect on the current-voltage (*I*-*V*) characteristics of that device. Therefore, the new world of small transistors is making it relatively easy to detect single defects, as their charge states directly influence the *I*-*V* curves.

As Kane pointed out, the essential point for us is to detect spin, not by its miniscule magnetic moment, but by virtue [16] of the Pauli exclusion principle. A donor defect can bind [36] a second electron by 1 meV, provided that second electron has opposite spin to the first electron. Thus spin detection becomes electric charge detection, the essential idea [35] behind spin-resonance transistors. In a small transistor, even a single charge can be relatively easily monitored.

A fairly conventional, small FET is very capable of measuring single charges, and therefore single spins as well. A single electronic charge, in the gate insulator, can have a profound effect on a low-temperature FET. At more elevated temperatures for example, the motion of such individual charges produces telegraph noise in the FET channel current. An illustration of such single charge detection [37] is in Fig. 10. A single electrostatic charge can add one additional carrier to the few hundred electrons in a FET channel. However the 2-nA change in channel current seen in Fig. 10 represents a few percent change, and is caused by long-range Coulomb scattering influencing the resistance seen by all the electrons. At low FET operating temperatures, ≈ 1 K, the random flip-flops disappear, but the sensitivity to single charges remains [38].

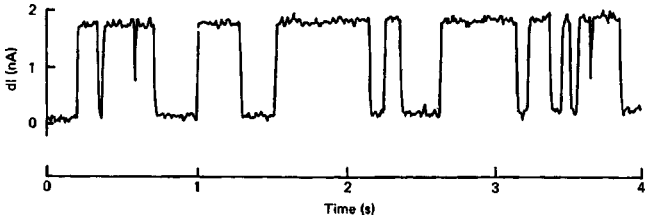

FIG. 10. The current noise in a small FET at 83 K from Ref. [37]. At this temperature the channel current fluctuates between two states, caused by a single trap being filled and emptied by a single charge. The change in channel current is ≈ 2 nA, which represents a few percent of total channel current, and is easily measured.

In our spin-resonance transistor design, shown in Figs. 5 and 9, the FET channel is labeled as the *n*-$Si_{0.4}Ge_{0.6}$ ground plane counterelectrode. It is located under the ^{31}P qubit donor, and in turn, the donor is under the top surface gate electrode. Thus the spin qubit is sandwiched between two electrodes. As in a normal FET the gate electrode modulates the *n*-$Si_{0.4}Ge_{0.6}$ channel current. The qubit electron donor is positioned in the gate insulator region where its charge state can have a strong influence on the channel current. Thus the successive charge states: ionized donor, neutral donor, and doubly occupied donor (D^- state) are readily sensed by measuring the channel current.

In Figs. 5 and 9, the two transistors have separate sensing channels under each transistor, so that they can be separately monitored, or indeed monitored differentially. By adjusting the gate electrodes, both qubit donor electrons can be attracted to the same donor. If they are in the singlet state they can join together forming the *D* state on one of the two dopant ions, but in the triplet state they could never occupy the same site.

Since the D^- state forms on one transistor, and an ionized donor D^+, on the other transistor, there would be a substantial change in differential channel current to identify the singlet state. For the triplet state, both donors remain neutral, and differential channel current would be constant. As indicated by the caption to Fig. 10, we can anticipate a few percent change in FET current associated with the singlet spin state, making spin readily detectable.

VII. SMALL-SCALE DEMONSTRATION

A possible two-qubit demonstration device is shown in Fig. 11. The differential current between the two FET's

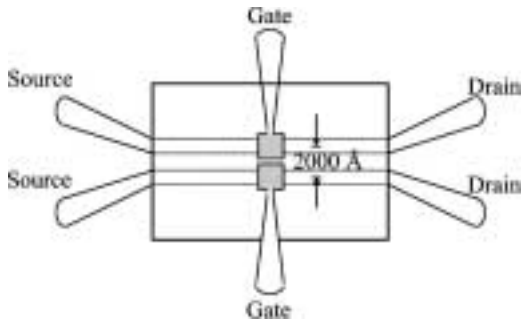

FIG. 11. Top view of the proposed device to demonstrate a CNOT gate. A perspective view (not including the source and drain) is shown in Fig. 5. Fluctuations in the current that flows from source to drain signal the charge state of the dopant ion under each electrode.

channels in Fig. 9 would monitor the electron-spin resonance. In practice a large number of transistor pairs would be arrayed along the two FET channels in Fig. 11, to allow for a finite yield in getting successful pairs. A good pair can be sensed using the same technique used in Sec. VI for the detection (measurement) process.

There are two levels of doping in our proposed device: The first level of doping is the conducting FET channel doping, that needs to be at a heavy concentration to overcome freeze-out at low temperatures. This is a standard design technique in low-temperature electronics. The second level of doping is in the qubit layer, that allows only one donor ion per transistor. Both doped regions need to be spatially patterned. The doped layers can be implemented by conventional ion implantation through a patterned mask, possibly with an intermediate epitaxial growth step to minimize ion straggle. Conventional annealing can be used to remove ion damage.

The ion-implantation dose for the qubit layer would be adjusted so that, on average, only one phosphorus ion would fall into each opening in the photoresist layer of Fig. 12. By

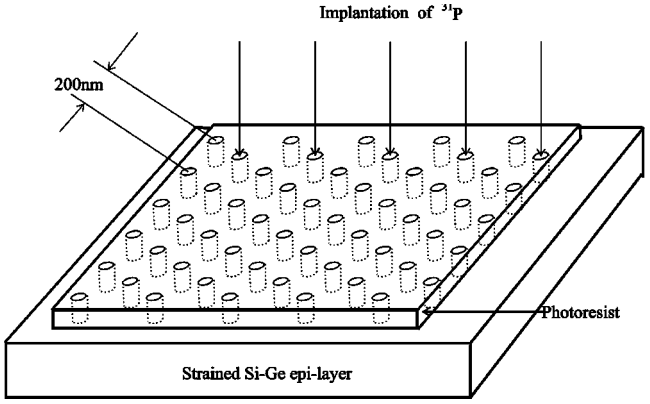

FIG. 12. The ion implantation step for inserting an array of qubit donor ions. The buried FET channels, that act as counterelectrodes to the gates and sense the spin-charge state, would be produced the same way. In a small-scale demonstration, the array would consist of only two rows, aligned with the FET channels of Fig. 11. This should provide an adequate yield of good qubit pairs.

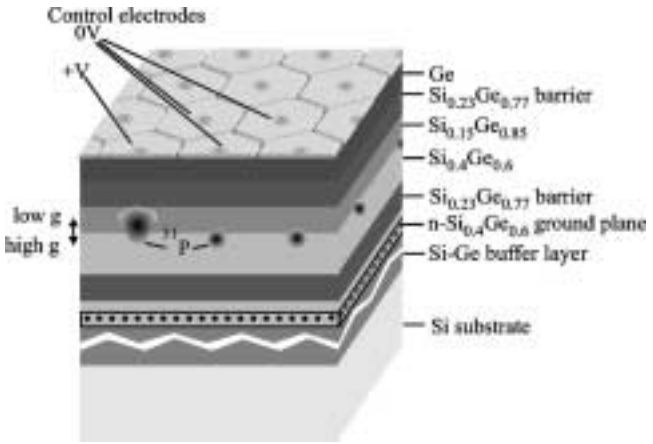

FIG. 13. In the future, we can expect arrays of Si-Ge SRT transistors. The center-to-center spacing would be ≈2000 Å. The gate electrodes on top will perform both single- and two-qubit operations, and can be used for data and instruction readin.

Poissonian statistics, the probability of getting exactly one phosphorus ion is 36.7%. Thus the probability of getting two adjacent gates to work would be 13.5%. That is adequate yield for a small-scale two-qubit demonstration device. To improve the yield for scale-up, there are many options. For example, the dopant could be sensed by its electric charge, and reimplanted if it were absent. Sensing an individual dopant is not difficult. It can be done, for instance, by monitoring the *I-V* curve at each site. By changing the voltage on a particular *A* gate, the electrons can be stripped off the donor. As result one can see no change, a single change, or a double change of the current depending on whether there is no donor, one donor, or two donors (etc.) in that site.

VIII. SCALING UP

There are a number of potential problems in scaling to a large computer. The future usefulness of electron spins will depend heavily on the favorable homogeneous T_2 spin-echo linewidth [23] in silicon, only 10^3 Hz. The T_2 lifetime in Si-Ge alloys has not been measured, and it will have to be demonstrated that it is as favorable as in pure silicon. On the other hand, there also appear to be methods such as isotopic purification, whereby this linewidth can be improved, particularly for well-isolated electrons.

In very large arrays, there are problems associated with the implantation yield of qubit donors. Poissonian statistics gives a yield of 36.7%, while a yield of 50% will required for percolation, or quantum connectivity, through the two-dimensional triangular array. There have been numerous non-Poissonian doping schemes proposed including sense-reimplant, self-assembly of molecular dopants, and scanning probe writing. Innovative doping methods have a long history, and we should anticipate that a suitable method will be optimized in time for scale-up to large quantum computers.

For instance, the sense-reimplant method (in which empty sites are sensed, and reimplanted with doping probability p_n in the nth implant) yields $pe^{-p}(1-e^{-np})/(1-e^{-p})$ good sites when $p_i=p$ is chosen. With this formula, $n=2$ (only

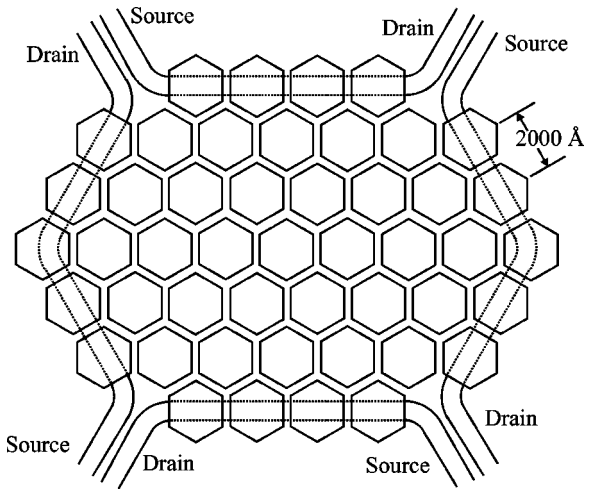

FIG. 14. In a large array, the readout qubits would be located around the periphery. Buried FET channels would sense the spin-charge state of a selected qubit. The channel current can change by a few percent in response to a single electronic charge.

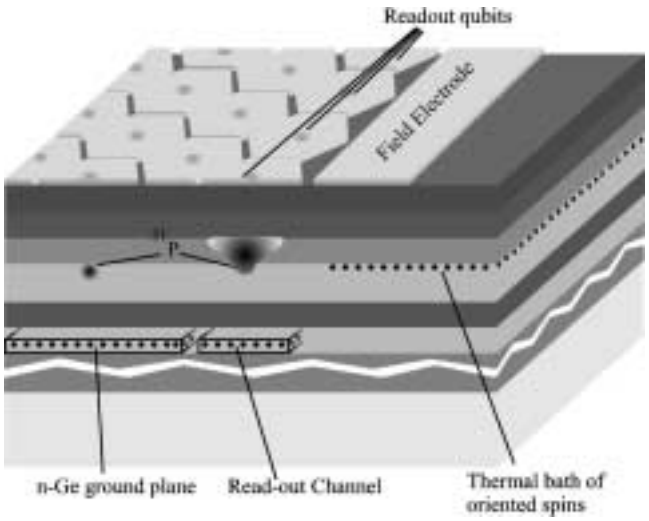

FIG. 15. A perspective view of Fig. 14, gives more details of the readout architecture for the peripheral qubits. The field electrode allows the readout qubits to interact with the heat bath of oriented electron spins.

one additional implant) already passes the percolation limit to yield 52.16%, while more implants, $n=3$, 5, 9, and 24, yield more than 60%, 70%, 80%, and 90% good sites, respectively. With $n=2$ an optimization of the doping probability in each implant (to be $p_1=0.632$ and $p_2=1$) provides the optimal yield of 53.15%.

The other scale-up issue revolves around the fact that each transistor will not be identical. As Kane noted, the transistors will have to be checked and calibrated repeatedly for use in a full-fledged quantum computer. The reason is that the nuclear spins, although almost static, will be different for each transistor. In addition the local alloy structure is different near every donor. We should not be discouraged by this checking and calibration requirement. In manufacturing classical integrated circuits, testing and repair are the largest expense. It is common to have only a finite yield of good devices, and to reroute wiring around bad transistors. This is probably inherent in the manufacture of any large-scale system.

The size of spin-resonance transistors, the required defect density, the increasing use of Si-Ge alloys, are all near to the present state of technology. If the spin-resonance transistor is successfully developed, we can anticipate arrays of qubits appearing much as in Fig. 13.

The readout of data requires that the buried counterelectrode, opposite the gate, should also function as FET channel. In a quantum computer, the result of the quantum computation is usually displayed on a small subarray of all the qubits. Hence the readout qubits can be located at the edge of the array. Figure 14 shows a qubit array, with readout FET channels (counterelectrodes) buried under the peripheral qubits of the array. A single buried FET readout channel can serve many qubits, since a chosen qubit can be selected for readout by its gate electrode.

The readout operation can be expedited if there is a thermal reservoir of donors surrounding the peripheral qubits as shown in Fig. 15. These can be attracted by a field electrode to the $Si_0.23Ge_{0.77}$ B barrier under the electrode, forming in effect a modulation-doped layer. Since the operating temperature of the computer is such that $kT \ll E_z$ with E_z the Zeeman energy of the electron spins, these qubits would be oriented by the magnetic field, and would act as a spin heat bath of known orientation. By attracting those bath spins to a peripheral readout qubit gate electrode, a singlet state could be formed, sensing that the readout qubit had been flipped. The current in the FET channel would then change, completing the readout operation.

After readout, the gate voltage could be made even more positive, and the readout qubit could thermalize with the surrounding heat bath. In effect, this resets the initial state of that peripheral qubit, which could then be swapped into the interior qubits for reuse as fault-correcting ancilla qubits.

Without a doubt there will be many other issues regarding scale-up. Semiconductors, particularly silicon, provide a track record of being tractable, engineerable materials in which many difficult accomplishments have become routine.

ACKNOWLEDGMENTS

This work was sponsored by the Defense Advanced Research Project Agency. The content of this paper does not necessarily reflect the position or policy of the Government, and no official endorsement should be inferred. The work of V.R. and T.M. was supported in part by grants from the Revolutionary Computing group at JPL (Contract No. 961360).

[1] P.W. Shor, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science, Santa Fe, NM*, edited by S. Goldwasser (IEEE Computer Society Press, Los Alamitos, 1994), p. 124.
[2] L.K. Grover, Phys. Rev. Lett. **79**, 325 (1997).
[3] R. Feynman, Int. J. Theor. Phys. **21**, 467 (1982).
[4] W.G. Unruh, Phys. Rev. A **51**, 992 (1995).
[5] P.W. Shor, Phys. Rev. A **52**, R2493 (1995).
[6] J. Preskill, Proc. R. Soc. London, Ser. A **454**, 385 (1998).
[7] J.I. Cirac and P. Zoller, Phys. Rev. Lett. **74**, 4091 (1995).
[8] N.A. Gershenfeld and I.L. Chuang, Science **275**, 350 (1997).
[9] D.G. Cory, M.D. Price, and T.F. Havel, Physica D **120**, 82 (1998).
[10] Q.A. Turchette *et al.*, Phys. Rev. Lett. **75**, 4710 (1995).
[11] G.J. Milburn, Phys. Rev. Lett. **62**, 2124 (1989).
[12] A. Shnirman, G. Schön, and Z. Hermon, Phys. Rev. Lett. **79**, 2371 (1997).
[13] A. Barenco, D. Deutsch, A. Ekert, and R. Jozsa, Phys. Rev. Lett. **74**, 4083 (1995).
[14] C. Monroe *et al.*, Phys. Rev. Lett. **75**, 4714 (1995).
[15] J.A. Jones and M. Mosca, LANL preprint, quant-ph/980127.
[16] B. Kane, Nature (London) **393**, 133 (1998).
[17] D.P. DiVincenzo, in *Mesoscopic Electron Transport*, Vol. 345 of *NATO Advanced Study Institute, Series E: Applied Sciences* , edited by L. Sohn, L. Kouwenhoven, and G. Schoen (Kluwer Academic, Dordrecht, 1997), p. 657.
[18] S. Bandyopadhyay, A. Balandin, V.P. Roychowdhury, and F. Vatan, Superlattices Microstruct. **23**, 445 (1998).
[19] G. Feher, Phys. Rev. **114**, 1219 (1959).
[20] G. Feher and E.A. Gere, Phys. Rev. **114**, 1245 (1959).
[21] D.K. Wilson and G. Feher, Phys. Rev. **124**, 1068 (1961).
[22] G. Feher, in *Paramagnetic Resonance*, edited by W. Low (Academic Press, New York 1963), Vol. II p. 725.
[23] M. Chiba and A. Hirai, J. Phys. Soc. Jpn. **33**, 730 (1972).
[24] J.P. Gordon and K.D. Bowers, Phys. Rev. Lett. **1**, 368 (1958).
[25] L.M. Roth, Phys. Rev. **118**, 1534 (1960).
[26] H. Hasegawa, Phys. Rev. **118**, 1523 (1960).
[27] D.K. Wilson, Phys. Rev. **134**, A265 (1964).
[28] E. Yablonovitch and E.O. Kane, J. Lightwave Technol. **6**, 1292 (1988).
[29] K. L. Wang and R. P. G. Karunasiri, in *Semiconductor Quantum Wells for Long Wavelength Infrared Detectors*, edited by M. O. Manasreh (Artech House, Norwood, MA, 1993).
[30] A. Schindlmayr, Eur. J. Phys. **18**, 374 (1997).
[31] J. Gea-Banacloche, Phys. Rev. A **57**, R1 (1998).
[32] H. Vollmer and D. Geist, Phys. Status Solidi B **62**, 367 (1974).
[33] K.L. Wang, S.G. Thomas, and M.O. Tanner, J. Mater. Sci. **6**, 311 (1995).
[34] G. Bastard, Phys. Rev. B **24**, 4714 (1981).
[35] D. Loss and D.P. DiVincenzo, Phys. Rev. B **57**, 120 (1998).
[36] M. Taniguchi and S. Narita, J. Phys. Soc. Jpn. **43**, 1262 (1977).
[37] M.J. Kurten and M.J. Uren, Adv. Phys. **38**, 367 (1989).
[38] It should not be surprising that single-electron charges are so easy to detect. That task was already accomplished many years ago in the Millikan oil-drop experiment of 1910.