Light Emission from Silicon

S. S. Iyer and Y.-H. Xie

The possibility induction of light emission from silicon, an indirect bandgap material in which radiative transitions are unlikely, raises several interesting and technologically important possibilities, especially the fabrication of a truly integrated optoelectronic microchip. In this article, the natural considerations that constrain silicon from emitting light efficiently are examined, as are several engineered solutions to this limitation. These include intrinsic and alloy-induced luminescence; radiatively active impurities; quantum-confined structures, including zone folding and the recent developments in porous silicon; and a hybrid approach, the integration of direct bandgap materials onto silicon.

The electronic revolution that has made a very profound impact on every aspect of life today has been made possible by microminiaturization, a high level of functional integration with a simultaneous reduction of cost. The integration of optical components onto a Si-based chip allows further functional integration and consequently the realization of new concepts. To a limited extent, optoelectronic integration has been achieved within the confines of the technology of III-V materials and their hybridization with Si chips. More recently, there have been modest gains in the application of Si processing to the fabrication of several optoelectronic components and their limited integration into Si-based technology. These include detectors (*1*), wave guides (*2*), modulators (3), interferometers, and switches. These components are all optically passive in that there is no means to stimulate light emission from the silicon.

It can be argued that the inability of Si to emit light seriously compromises our ability to get true large-scale optoelectronic integration at reasonable cost. By semiconductor standards, Si is rather mediocre. The mobility of electrons in Si is low when compared with several compounds of III and V materials, and above all, its bandgap is indirect. These drawbacks, however, are compensated for by its easy processability, its superior structural properties, and the tremendous knowledge that we have gained about Si.

It is important to understand what makes Si an indirect bandgap material. Most semiconductors are direct gap materials. The band structure of a material is intimately dependent on several factors, including crystal structure, lattice constant, chemical species, bonding and bond lengths, electronegativity, stiffness, and elasticity. Many of these factors are lumped into the crystal potential, which is then used in the crystal Hamiltonian to solve for the crystal wave equation. Conventionally, the band structure of a semiconductor is represented by the dispersion relation $E_n(\mathbf{k})$, where E is the energy of an electron (or hole) at the band edge with a wave vector **k** in the first Brillouin zone (BZ) (Fig. 1) (*4*).

Crystal symmetry requires that $E(\mathbf{k})$ have extrema at the zone center and the zone boundary. However, these are not the only points at which extrema can occur. In the case of the essentially covalent group IV elements and compounds (C, SiC, Si, and Ge, for example), additional extrema occur in the lowest lying conduction bands away from the zone center. Note that the valence band structure for most semiconductors is similar, especially at $\mathbf{k} = 0$, the zone center. The energy gap is the difference between the valence band at $\mathbf{k} = 0$ and the lowest point in the conduction band. When this lowest point occurs at $\mathbf{k} = 0$, the semiconductor is direct gap because a transition can occur at the zone center with both the initial and final states having the same momentum vector $\mathbf{k} = 0$. A good example of such a case is GaAs.

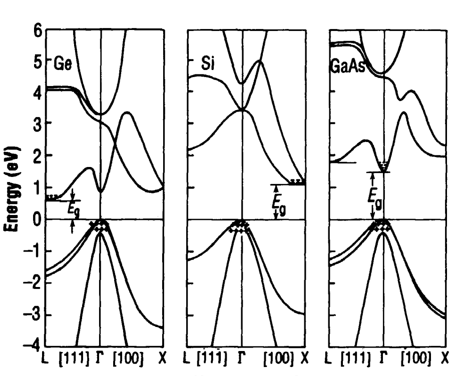

Fig. 1. Energy dispersion relations for common semiconductors Ge, Si, and GaAs. Shows energy gap (E_g): the direct gap of GaAs and indirect gap for Ge and Si. Electrons at the conduction band minimum (minus signs) in an indirect semiconductor need the assistance of a phonon to relax into the state filled with holes (plus signs) at the maximum of the valence band (*4*).

The case of Ge is very interesting in that a higher order band actually has a slightly lower energy in the X, that is, ⟨111⟩, direction of the BZ. Thus, Ge barely misses being direct gap. The minimum of its conduction band, which is the lowest occupied conduction band state, is called a satellite minimum. It is the nearest thing to a direct semiconductor that occurs naturally in the group IV systems. In Si, satellite minima in the L ⟨100⟩ direction impart a decisively indirect nature to the bandgap. Tin is an interesting group IV element in that its diamond allotrop, α-Sn, is a semimetal, that is, it has no energy gap at 0 K. We will discuss the possibilities of using alloys to bring about direct transitions later in this article and show how Sn can play a role.

In a direct semiconductor, radiative transitions can occur quite easily as an electron excited into the conduction band minimum at $\mathbf{k} = 0$ can spontaneously decay into the valence band state also at $\mathbf{k} = 0$, yielding a photon of energy equal to the bandgap (in semiconductor parlance, this is referred to as an electron hole radiative recombination). In an indirect semiconductor, a change of the electron momentum or wave vector is also needed. This can be accomplished by the transfer of momentum to the crystal through the creation of a phonon with a wave vector equal to that of the initial conduction-band state. This three-body event (electron, hole, and phonon) is significantly less likely to occur than a direct electron hole recombination. A phonon of the right **k**-vector and energy must participate, the phonon availability being governed in part by the phonon dispersion relation. This simply explains the inability of indirect gap semiconductors to emit light efficiently.

The indirect gap also affects absorption of incident radiation. Photoexcitation of carriers can occur both by the absorption and emission of phonons in indirect semiconductors. The absorption coefficient, α, has quadratic dependence on photon energy and has two branches. In direct semiconductors, no phonon assistance is needed, and α has a parabolic dependence on incident photon energy. The net result is that when band-to-band absorption occurs in direct semiconductors, the incident radiation is absorbed to much shallower depth than in an indirect semiconductor. A state

---

S. S. Iyer is with the IBM Research Division, T. J. Watson Research Center, Yorktown Heights, NY 10598. Y.-H. Xie is at AT&T Bell Laboratories, Murray Hill, NJ 07974.

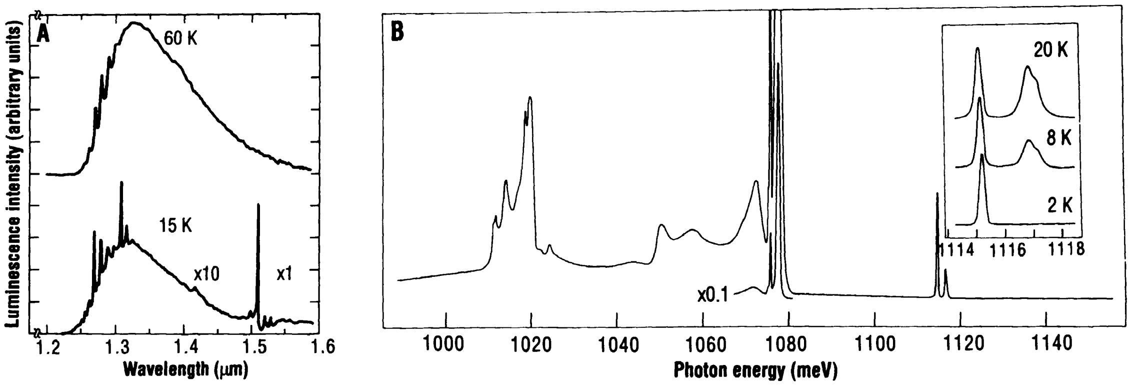

Fig. 2. (**A**) Photoluminescence (PL) spectra (corrected for the response of the optical system) of Si-S at 60 and 15 K (*58*). (**B**) Uncorrected PL spectrum of Si-Be at 7 K. (**Inset**) Thermalization behavior of the higher energy luminescence system (*10*).

in the forbidden gap can also mediate an electron-hole recombination. In an indirect semiconductor, phonon assistance may still be required, making gap state–mediated radiative recombination potentially less efficient than a direct transition. These concepts are developed more rigorously in some of the classic textbooks (5).

One more introductory point concerning excitonic luminescence is in order. Excitons are electron-hole pairs that are bound to each other and may, in addition, be either free or bound to impurities. When excitons decay radiatively, light is emitted. Excitonic luminescence is a good indicator of material quality, but unfortunately, in general, excitons have lifetimes at room temperature that are too short to be significant. However, recent work by a few groups (6) has shown that the confinement of the exciton in a quantum well can be used to tune the emission wavelength as well as to increase the excitonic lifetime. At present, however, we believe that excitonic luminescence is more appropriate as a diagnostic tool.

Engineering Luminescent Transitions in Silicon

A number of avenues are available for the engineering of luminescing transitions in an otherwise indirect material. Luminescence is a result of significant overlap (in direct and reciprocal space) in the electron and hole wave functions. Whenever there is such overlap, luminescence is possible. However, the strength of the luminescence, that is, the emission rate and quantum efficiency, depends on the extent of this overlap and the transition probability. Engineering solutions seek to increase this overlap and can be split into four classes.

1) In impurity-mediated luminescence, an impurity that has an energy level in the gap of the semiconductor is used as an intermediary state through which the electron can recombine with the hole. The impurity level must be dispersed in **k**-space to be efficient.

2) Band structure engineering makes use of alloys. Two or more group IV elements are alloyed to shift the energy bands a little so that a direct transition is allowed.

3) Quantum confinement can also increase the probability of a direct transition. We consider two related phenomena: zone folding—where an ultra-short–period superlattice with periodicity comparable with the lattice constant induces greater overlap of the conduction band states at the zone edge with the valence band states at the zone center, increasing the possibility of a direct transition—and other strict quantum confinement in one or zero dimensions, such as in quantum wires and dots. We discuss this with special reference to recent results from porous Si.

4) Strictly hybrid approaches, in which a direct bandgap material is grown on or joined to Si.

Luminescence Through Impurities in Silicon

Electron-hole pairs injected either electronically or optically can recombine through impurity centers with enhanced recombination rates compared with those of the pure Si crystal, in which recombination is intrinsically very slow (up to several milliseconds). The enhancement, in a simplified manner, can be considered as the consequence of the relaxed **k**-selection (momentum conservation) requirement that is caused by the localization of the electron-hole pairs near impurity centers. The detailed processes depend on the properties of the specific impurity center, such as the symmetry of the impurity state with respect to that of the wave functions of the electrons and holes, the degree of localization, and so on. The energy from the recombination of an electron-hole pair can be released by the generation of a photon, a phonon or phonons, or through a variety of other channels. Each of these channels has a certain probability, and the one that has the highest probability dominates the nature of a given impurity center.

For example, the centers for which the dominant energy release channel is the emission of a photon are customarily termed "radiative impurity centers," whereas those for which the dominant energy release channel is phonon emission, the Auger process and so forth, are termed "nonradiative impurity centers." In Si, the recombination of electron-hole pairs through impurity centers is, in general, nonradiative. Of the limited number of known radiative impurity centers in Si, only isoelectronic centers (*7*) and rare-earth centers (8) will be discussed here because of their relative technological importance.

The term isoelectronic centers stems from the fact that these impurities are isovalent with the Si in the lattice; in other words, they are electrically neutral. One such example is the S-doped Si (Si-S). Samples are usually fabricated by ion implantation or diffusion of S into Si followed by a high-temperature anneal (≈1200°C) and a rapid quench (9). The photoluminescence (PL) spectrum of Si-S consists of two peaks: one broad peak at a wavelength of 1.32 μm and one sharp peak near 1.50 μm (Fig. 2A). Although the 1.32-μm emission does persist to a temperature of T = 170 K, the 1.50-μm emission decays much more rapidly as sample temperature rises. The thermal dissociation energy for the 1.32-μm bound exciton emission is experimentally determined to be ≈139 meV. The quantum efficiency at low temperatures is 2 to 5% and is dependent on PL excitation power.

Another example of radiative isoelectronic centers in Si is the Be-Be pair (*10*). The PL spectrum of Si-Be is dominated by a cluster of peaks around 1.08 eV (Fig. 2B). Recently, Modavis and his co-workers (*11*) have fabricated GeSi-Si quantum wells doped with Be. The quantum wells are believed to provide confinement for excitons and thus raise their dissociation temperature (*12*). For Si-Be, the PL intensity decreases with increasing sample temperature (for $T >$ 34 K) and disappears at about 60 K.

There are several common characteristics of luminescence through isoelectronic centers in Si. (i) It is vanishingly weak at 300 K. The luminescence intensity together with the quantum efficiency increase with decreasing temperature as the result of reduced nonradiative recombination rates at low temperatures, among other reasons. (ii) The quantum efficiency is a few percent at low temperatures. (iii) Because of the long spontaneous emission lifetime (milliseconds), the radiative process is vulnerable to the competing nonradiative recombination processes. These nonradiative processes determine the minority carrier lifetime at high temperatures.

The luminescence mechanism in rare-earth doped Si is fundamentally different from that of the isoelectronic centers. For simplicity, only Er-doped Si (Si-Er) will be discussed. The characteristic feature of the Si-Er luminescence process is that it is an internal transition of the 4*f* states of Er^{3+} (*13*) (analogous to gas ions in a gas laser). The luminescence has one very sharp peak at 1.54 μm, a wavelength that is important for optical fiber communications. However, the advantage of the emission wavelength is offset by the fact that the 4*f*-4*f* transitions are forbidden in the case of free Er^{3+} by the parity selection rule and are possible in Si crystal only because of the crystal field. The crystal field intermixes states of opposite parities (*14*) and splits each single state of a free Er^{3+} into a manifold of substates with slightly different energies. The forbidden origin of these transitions makes the transition probability fairly low, which is manifested as the long spontaneous emission lifetime (milliseconds) and the small interaction cross section with electromagnetic waves.

For optoelectronic applications, the interest is in luminescence as a result of the electronic current injection, known as electroluminescence (EL). To date, EL from the Er impurity has been observed in many Er-doped semiconductors, including Si (*15*), although the quantum efficiency and the overall photon intensity have been low. The detailed mechanism of energy transfer from the recombination of electrons and holes to Er^{3+} is not understood. The presumed process of EL is illustrated in Fig. 3, where the atomic transitions within the Er^{3+} are shown in relation to the Si conduction and valence band. Some processes and states are not known. Study of the properties of the Er defect state as it appears to the injected carriers in the conduction and valence bands and the detailed mechanism of energy transfer are two of the most challenging areas in the research of Er-doped semiconductors. Another technologically important factor is the low solid solubility of Er in Si, as low as $\approx 10^{18}$ cm^3. The long spontaneous emission lifetime, small interaction cross section, and the low solid solubility are serious hurdles for the use of Er-doped Si in optoelectronic applications (*16*).

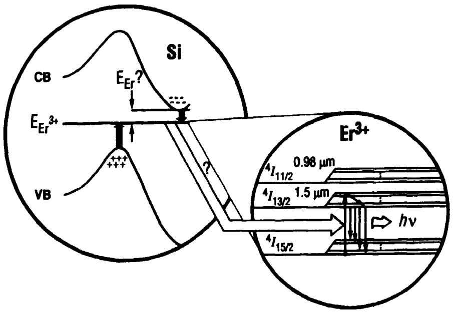

Fig. 3. Illustrative drawing of the electroluminescence (EL) process of Er^{3+} in Si in relation to Si conduction band (CB) and valence band (VB). Detail indicates energy levels of Er^{3+}. Question mark indicates unknown energy separation or energy transfer path.

Although the rare-earth elements and isoelectronic centers in Si are fundamentally different, there are some common features. Three of the most important common features are listed below.

1) The localization of charged carriers by the short-range potential associated with the radiative impurities in Si enhances the zero-phonon transition probability for the injected electron-hole pairs. The enhancement is generally moderate, however, which leaves the luminescence vulnerable to nonradiative shunt paths from unwanted but ever-present deep-level defects. Consequently, the samples may luminesce with acceptable quantum efficiency at low temperatures, but all of them degrade significantly as the temperature is raised. To date, there is no impurity doping in Si that can result in a 300 K quantum efficiency close to that of GaP-N ($\approx$0.1%) (*17*), the material in commercial light-emitting diodes (LEDs). Research for efficient radiative impurities in Si is an ongoing effort.

2) The maximum concentration of a certain type of impurity in Si is limited by the solid solubility. This results in the saturation of the output optical intensity with pumping intensity.

3) The understanding of the luminescence mechanisms is incomplete for both isoelectronic centers and Si-Er. In the former case, questions remain as to the details of the microstructures of the impurity complexes and their associated short-range potential. In the case of Si-Er, the excitation mechanism is still in question.

We take a brief look at material that is used for commercial blue LEDs, namely, SiC. Crystallographically, SiC has a polytype structure with a variety of packing orders between Si and C (*18*). It also has an indirect energy band structure very similar to that of Si. The values of the energy bandgap range from 2.2 to 3.3 eV for the different polytypes. The mechanism for the observed blue luminescence at 2.65 eV is believed to be associated with the Al doping in 6H-SiC (α-SiC) (*19*). It is the wide bandgap and the know-how in the fabrication of good *pn* junctions that gave SiC commercial viability. The external quantum efficiency, on the other hand, is fairly low (typical values range from 9 × 10^{-6} for the 2.58-eV luminescence to 5 × 10^{-5} for the 2.15-eV luminescence). This low quantum efficiency once again bears out the common characteristics of the impurity-mediated luminescence from indirect semiconductors. Practically, this means that the currents needed to obtain EL are high, a disadvantage in many applications.

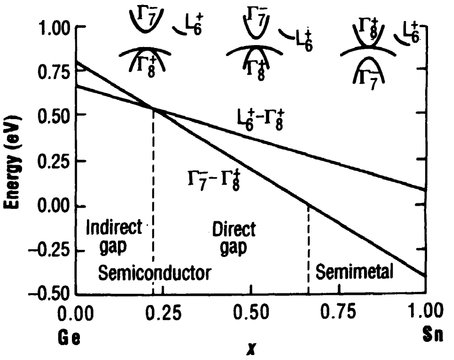

Fig. 4. Band structure for Ge-Sn alloy. Shows transition from semimetal, direct gap, and indirect gap regimes. Schematic representations of the relevant conduction and valence band dispersion relations are also shown (*20*).

Band Structure Engineering Through Alloying

The achievement of compositional as well as strain modulation with alloys has produced remarkable results in the engineering of the electronic properties of semiconductors in, for example, $Si_{1-x}Ge_x$ alloys. Although alloys of Si and Ge and of Si and C do not cause any increase in the direct transition probability, the case of Ge and Sn is an interesting one. Addition of about 23 atomic percent Sn to Ge is predicted to induce a direct transition (Fig. 4) (*20*). Furthermore, this direct gap is tunable from 0 to 0.5 eV as the Sn concentration is varied from 23 to 65%. There have been many attempts to prove this theoretical prediction experimentally, but results have been mixed at best (*21*). Two problems plague the Ge-Sn system: the huge lattice mismatch to both Si and Ge, which no doubt introduces dislocations that in turn are very likely efficient nonradiative paths, and the extremely limited solubility and phase instability of the alloy. The Si-Sn system may exhibit similar band structure variations but is also plagued with similar problems. Attempts to eliminate the lattice-mismatch effect with the growth of Ge-Sn on lattice-matched substrates, such as CdTe, have not yet yielded convincing results of a direct gap alloy.

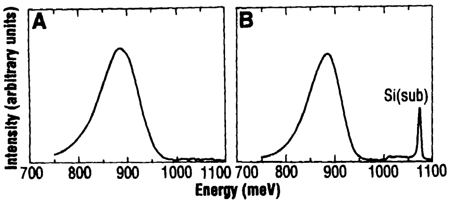

Fig. 5. (**A**) EL and (**B**) PL spectra show a typical broad feature in $Si_{1-x}Ge_x$ alloys (here $x = 0.18$) at about 0.85 eV (*22*). The Si peak from the substrate is also shown. The spectra are taken at 4.2 K and persist until ~80 K. Conditions for analysis: (A) 5 mA, 11.9 V; (B) 120 mW, 514.5 nm.

Alloys of SiGe show an interesting luminescence feature at about 0.85 eV (*22*) (Fig. 5). The feature is broad and strong, and EL has been observed from this feature at temperatures as high as 80 K. The peak position tracks the strained bandgap of $Si_{1-x}Ge_x$ lying below it by about 120 meV. Our present understanding of this feature is incomplete, but it is believed to be a result of compositional fluctuations, which lead to quantum-confined enhancement of either band edge or excitonic luminescence. The broadening observed in these spectra may be caused by size fluctuations of the luminescing entities. Once again, we emphasize that these are as yet relatively weak features.

Zone Folding

The concept of zone folding was proposed by Gnutzmann and Clausecker (*23*) and is most easily explained with reference to Fig. 6, where we show schematically the valence and conduction bands of Si along the X ⟨100⟩ direction of the BZ. The zone boundary is at $2\pi/a$, where a is the lattice constant. The concept of zone folding rests with an increase in the coupling of the conduction band minima at 0.8 of the zone boundary with the valence band maxima at zone center. This is done conceptually by repeated folding of the BZ on itself in submultiples, chosen in this case to be 1/5. This defines the new reduced zone boundary about which the bands should be folded. Additionally, the requirement of extrema in the band structure (the Kramer conditions) at the minizone boundary and zone center needs to be satisfied, resulting in the depicted zone-folded band structure. The vertical transition is referred to as a quasi-direct transition. In real space, this transformation is effected by a decrease of the periodicity in the ⟨100⟩ direction by a factor of 5 (the superlattice is one-dimensional and transforms to reciprocal space as π/d, where d is the period). This simple explanation assumes that the constituents of the superlattice have similar band structures. How do we realize this in practice? One possible method is to make a superlattice of Si and a Si-like $Si_{1-x}Ge_x$ alloy with the new period. However, one now has to account for the perturbations introduced by the interface and the fact that the $Si_{1-x}Ge_x$ layer is strained (this itself causes shifts in the band position), and the simple multiple folding that is proposed is found not to yield the desired quasi-direct transition.

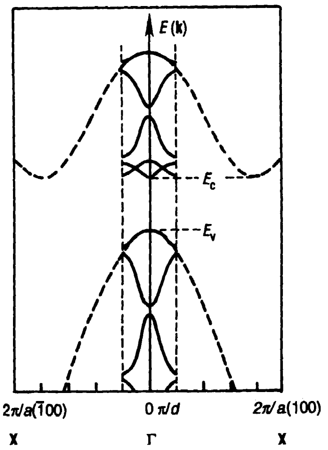

Fig. 6. Illustration of zone folding in the case of Si(100) (*59*). By folding in submultiples of one-fifth of the zone boundary, one can fold the conduction-band minimum, $0.2\pi/a$ from the zone boundary, to zone center. This transition is now called quasi-direct.

The problem has been analyzed by several groups both theoretically (*24, 25*) and experimentally (*26, 27*). To summarize the result, one must account for the individual strains in the two superlattice constituents, as well as band offsets. Moreover, the lowest conduction band states in the unfolded structure need not fold to the lowest state in the folded structure; that is, the lowest conduction band state in the superlattice is indirect. In fact, this is the case for a Si_mGe_n superlattice grown on Si (that is, Ge layers are strained) with $m = n$.

A more elaborate analysis shows that quasi-direct transitions are more likely in structures in which the Ge layers are unstrained. However, in this and other cases, one must also be concerned with the dispersion of the lowest lying states. If the dispersion is negative, the level that is zone folded is not the lowest level in that band. This negative dispersion is reduced when the period is increased, namely, when the individual layer thickness is increased.

This brings up another important point. The character of the conduction band states is such that it is localized in the Si-like layers, while the valence band states are localized more in the Ge-like layer (this depends on the band offsets). Thus, although the transitions are "direct" in reciprocal space, they are "indirect" in real space because the electrons and holes are localized in different layers, Si and Ge, respectively (hence the term quasi-direct). Thus, the optical transition matrix element that indicates the probability of the transition decreases as the period is made larger. All these considerations make the practical realization of a direct gap zone-folded structure quite difficult. In addition, if one requires a large number of periods so that the radiating volume is appreciable, structural stability of these strained layers is also a concern. There have been reports of low-temperature luminescence in Si_6Ge_4 superlattices grown on a $Si_{1-x}Ge_x$ ($x = 0.4$) substrate at about 0.85 eV. More recently, Presting *et al.* (*28*) reported 5 K PL and EL of comparable intensity in Si_5Ge_5 structures (Fig. 7). These results are encouraging, but we must be concerned with alloy-related luminescence that has been seen in buffer layers as discussed earlier and with the fact that the luminescence is relatively weak and is negligible at 300 K.

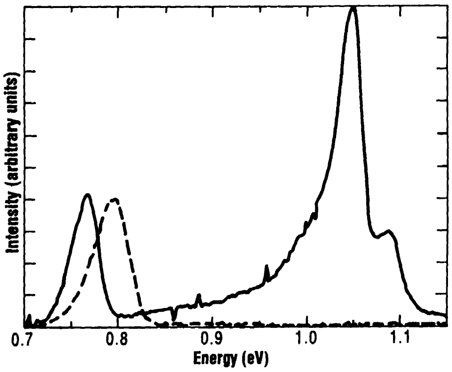

Fig. 7. The PL (solid) and EL (dashed) spectra for a Si_5Ge_5 mesa diode at 5 K (*28*).

Quantum Wires and Dots

Quantum confinement in thin (few nanometers) wires or dots provides another possible approach to the engineering of a direct transition. In many ways, the arguments are similar to those of zone folding. As one makes the physical structure of the semiconductor into a fine quantum wire, the values of the allowed energy levels increase. This is analogous to the increasing energy levels for a particle in a quantum mechanical box. As a result, the bandgap of a semiconductor that is a quantum wire or dot increases (Fig. 8). Qualitatively, the confinement of the carriers in real space causes their wave functions to spread out in momentum space, which increases the likelihood of strongly radiative transitions. In addition, scattering at the wire or dot boundaries, a very likely event, can supply the needed momentum more readily in a confined structure.

Clearly, an exact estimation of the bandgap of a Si wire will be quite complex, but our qualitative arguments show that it is possible that such a structure could yield direct, by which we mean strongly radiative, transitions. In fact, first principles calculations for a hydrogen-terminated Si filament 1.14 nm wide by Buda *et al.* (*29*), using density functional theory, gives a direct transition at $\mathbf{k} = 0$ with an energy gap for the wire of 1.84 eV. Although the calculation is known to underestimate the gap, it shows that conceptually direct gaps can occur in Si filaments of sufficiently narrow diameter. Qualitatively similar results have also been reported independently by Read *et al.* (*30*).

Porous Si—A Special Case

The quantum wire approach to the creation of direct gaps in Si received a sharp boost from the initial discovery that porous Si with very fine filaments, a few nanometers across, can luminesce intensely in the visible at room temperature. This remarkable discovery was made by Canham (*31*) at the Royal Signal and Radar Establishment (United Kingdom). Earlier, Lehman and Gösele (*32*) at Duke University had observed that the absorption edge of porous Si was at a higher energy, by about 0.5 eV, than that of bulk Si. The quick and easy reproduction of these results in laboratories around the world gave them credibility and kindled intense investigations into this phenomenon.

Porous Si is itself fabricated by an anodic dissolution process that is usually done in a HF-based electrolyte. Although the microscopic details of pore formation are complex and as yet incompletely understood, a labyrinth of interconnected Si is created. The microstructure of porous Si depends on the composition of the electrolyte, the type and resistivity of the initial Si substrate, the current density during anodization, and the post-anodization treatment. Under the right conditions, a large pore structure is created, in which as much as 85 to 90% of the Si is etched off leaving filaments as thin as a few nanometers across. The principal feature of porous Si is extremely fine structures, either wires or dots, which are small enough to exhibit some quantum confinement effects. Structural analysis of this aptly called "quantum sponge" (*33*) is quite difficult, but several workers, using electron microscopy, x-ray diffraction, and Raman spectroscopy, have confirmed that the filaments are crystalline. The material is fragile and by nature has a very large surface area. Immediately after anodization, this surface is usually hydrogen-passivated with little evidence of oxygen or other impurities. One can widen the pores or, equivalently, *make the filaments* thinner, by leaching in HF solution. The structures show intense luminescence in the red region of the spectrum when excited with moderate intensity ultraviolet light at room temperature.

We point out some important features in the typical PL spectra from porous Si (Fig. 9): The line shape is broad compared with, say, the PL from GaAs or a defect level in Si. Although the peak intensity is smaller than that of a direct-gap semiconductor such as GaAs, the integrated intensity is comparable. Furthermore, by changing the anodization conditions to obtain a finer filamentary network, one can shift the peak to shorter wavelengths, which suggests that quantum confinement plays a significant role in the light generation mechanism. The broadening of the peak can be ascribed to quantum-size fluctuations in the porous Si. However, this does not by any means prove that the material is direct gap. In a direct gap semiconductor, the luminescent decay time is typically not more than a few nanoseconds. In porous Si, however, this lifetime can be several tens of microseconds, which raises the possibility that this is an indirect transition.

In fact, some notable objections to the direct gap hypothesis have been proposed by workers at the Max Planck Institute in Stuttgart (*34*). They have correlated the spectral properties of porous Si to that of siloxene, a Si-based polymer, the inference being that a siloxene-like compound is formed on the large surface presented by porous Si during its formation and causes the luminescence observed. This argument is somewhat controversial but worth considering. Using infrared spectroscopy (*35*), we can argue that the bond structure of freshly prepared porous Si does not contain Si–O or O– bonds, definitive features of siloxene and its derivatives. In fact, when such signatures are seen in porous Si, for example, after laser excitation in air, it is usually accompanied by a dramatic quenching of the luminescence intensity. Others (*36*) have argued that the hydrogenated surface of porous Si is similar to hydrogenated amorphous Si, which can luminesce in the visible. Hydrogenated amorphous Si has a gap that is significantly larger than that of porous Si, and the luminescence is attributed to impurity states. Perhaps, it can be argued, analogous impurity states exist in porous Si. Note, this is not the same as the assertion that luminescence from porous Si is caused by amorphous Si-H, a subtle but important distinction.

It is quite possible that the luminescence we observe is defect-related and, in that case, likely to be surface defect–mediated. Even so, quantum confinement will certainly influence the energy level and radiative lifetime. Some insight into this can be obtained from the temperature evolution of luminescence. High temperatures cause the desorption of the hydrogen termination, and we observed that the emitted visible light is severely quenched when this happens. However, the loss can be reversed by rehydrogenation through, for example, exposure to HF. Quenching can be prevented by the encapsulation of the porous Si with amorphous Si, for example. We can argue that the hydrogen termination is essential to reduce the surface recombination velocity and keep the radiative lifetime long compared with the nonradiative lifetime. The low-temperature behavior, however, is very interesting. The luminescence goes through a maximum at about 200 K and is completely quenched at temperatures below a few kelvin (Fig. 10) (*37*). This result is quite surprising and may indicate the quenching of a defect state as the temperature is reduced. It might also be argued that there is a complex interplay of radiative and nonradiative lifetimes, and the latter become dominant at low temperatures. It is more likely, however, that the radiative

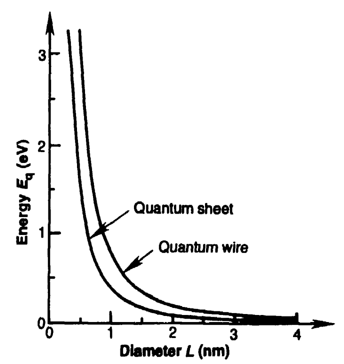

Fig. 8. The energy of an allowed quantum-confined electron state plotted as a function of the confining dimension for a quantum sheet (two-dimensional confinement) and a quantum wire (one-dimensional confinement). The increase in the allowed energy level can explain the visible luminescence from Si quantum wires (or dots) at energies above the bulk bandgap (*33*).

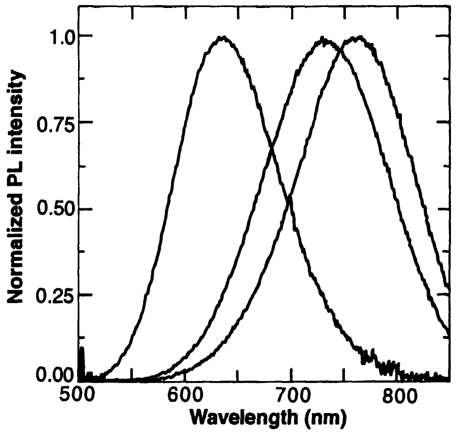

Fig. 9. Typical PL spectra for porous Si etched under different conditions. The blue shift may be a consequence of the greater confinement designers achieve by etching a finer structure (*37*). The incident power of the laser was 3 mW at 488 nm; experiment done in N_2.

transition is assisted by phonons; the phonon population decreases with temperature. Additionally, polarization dependence, transport, and magneto-optic measurements, conventional methods for the identification of quantum confinement phenomena, have been inconclusive on application to porous Si, presumably because of the isotropic nature of the quantum sponge.

There have been reports of electroluminescent devices made with porous Si. One such structure, fabricated at IBM (*37*), consists of a *pn* junction (Fig. 11). The fabrication sequence included the definition of a mesa of highly *n*-type doped material formed by implantation into a lightly *p*-type epitaxial layer grown on a highly *p*-type substrate and the lateral preferential anodization of the *p*-layer, as shown, with a finite but small excursion of the porous layer into the *n*+ overlayer. A relatively good junction was obtained, and the device showed weak EL with a spectral response close to the PL produced in the same porous Si layer. Because porous Si has a larger gap than nonporous Si, a large-gap heterostructure may be required to inject carriers into the porous Si. This approach has also been reported with encouraging results for indium tin oxide (*38*).

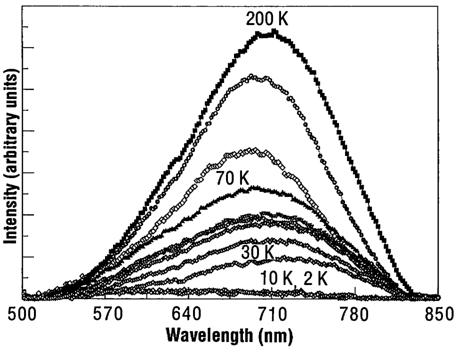

Fig. 10. Degradation of PL as a function of temperature. The PL spectrum of porous Si has a maximum around 200 K and is quenched at lower temperatures. This, combined with the long PL lifetime, suggests that impurity effects may play a major role (*37*).

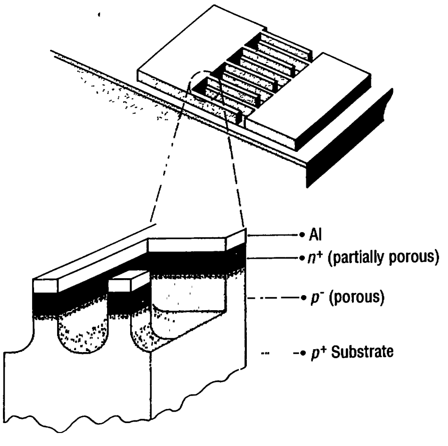

Fig. 11. A *pn* junction that has been fabricated in porous Si. The structure exhibited weak EL (*37*). Others have also reported EL in porous-Si structures.

To sum up, the mechanism for light emission from porous Si remains somewhat controversial, but no one will argue with the fact that the emission is easily reproduced, is intense, can be tuned, and has an intensity that can be stabilized after an initial decrease with proper control of the ambience. Even so, porous Si remains a complex, microscopically variable, and unstable material. Attempts to produce controlled filaments of nanometer size are in progress in several laboratories around the world, but no definitive results are yet available.

Hybrid Approaches

An obvious way to get light from a Si-based chip is to integrate a III-V material on Si. The issues involved in this type of integration are significantly different from those of the other approaches discussed so far in that those main objectives were to increase the luminescence efficiency and to understand the luminescence mechanism. The emphasis here is the reduction of material defects to improve the device reliability. This is because the band-to-band recombination in direct bandgap semiconductors is well understood, and the quantum efficiency of the luminescence processes is, in general, high (tens of percent). Among the various approaches to fabrication of these hybrids, two will be discussed in detail: heteroepitaxy of GaAs on Si and epitaxial lift-off (*39*). The substrate orientation is assumed to be (100) throughout the discussion because of the overwhelming practical importance of the Si(100) surface.

Heteroepitaxy of GaAs on Si. There are two fundamental problems for this approach. One is the mismatch of the lattice constants. Among the common semiconductors, Si has one of the smallest lattice constants (5.4310 Å) (*40*). Consequently, all the practical GaAs-Si devices are fabricated with relaxed GaAs films. Dislocations, in this case, become the major concern for device reliability. Dislocations are known to function as fast diffusion paths for impurities (*41*) and as short circuits in *pn* junctions (*42*) and are known to be the points of origin for the catastrophic breakdown of semiconductor lasers (*43*).

A variety of methods have been used to reduce the threading dislocation density (ρ_t). The most common approach is the use of the strained layer superlattices (SLS) to filter dislocations (*44*). Another approach is the use of a compositionally graded buffer layer (*45*). A value of ρ_t as low as 2.9×10^4 cm^{-2} has been reported with the use of the SLS approach and has been characterized by a count of the etch-pit density (*46*). However, it is important to realize that investigators will have to use different characterization techniques to measure widely differing densities of dislocations to ensure an accurate count, as cautioned by Stirland (*47*).

The second problem in the heteroepitaxy approach is the polar-nonpolar mismatch. This mismatch comes from the fact that the group III element site in a III-V semiconductor crystal is distinguishable from the group V element site, whereas all sites in a Si crystal are identical. The obvious consequence of such a mismatch is the problem of antiphase boundaries (APBs) (*48*). A monolayer (versus a bilayer) step on the initial Si(100) surface results in a phase difference in the stacking sequence of the compound semiconductor between domains on opposite sides of the step. The domain walls are termed APBs. Across an APB, the bonds are Ga–Ga or As–As instead of the normal Ga–As bonds in bulk GaAs, resulting in electronic potential fluctuations that inevitably affect the electronic properties of the crystal. As an example, the presence of an APB in the depletion region of a semiconductor *pn* junction can result in a high reverse leakage current through the junction. For vicinal growth planes to (100), Alerhand *et al.* (*49*) calculated the finite temperature phase diagram and showed that the double-layer steps are the stable phase above a critical angle of the tilt for a given substrate temperature. Experimental results, largely from scanning tunneling microscopy, support the calculated phase diagram (*50*). Therefore, GaAs grown on the double-stepped surfaces can be free of APBs (*51*).

In addition to the APB problem, a less obvious consequence of this mismatch is the interface charge problem. An abruptly terminated perfect interface between a polar semiconductor and a nonpolar semiconductor necessarily results in a heavily charged plane (*52*). Interface mixing, a fraction of Si atoms replaced by Ga atoms, for example, should reduce the charge density; but, in reality, kinetic limitations prevent the total elimination of this interface charge.

Epitaxial lift-off. In contrast to the approach of the heteroepitaxy of GaAs on Si, the epitaxial lift-off approach consists of wet chemical etching and the bonding of the film onto an arbitrary substrate by the van der Waals force. The only epitaxial step is the growth of Al$_x$Ga$_{1-x}$As-GaAs ($x < 0.4$) on GaAs substrates with a thin ($\approx$50 Å) AlAs release layer. The previously discussed obstacles encountered in the heteroepitaxy of GaAs on Si do not exist for this approach. The key for the success of this technique is the high etching selectivity ($>10^7$) between $Al_xGa_{1-x}As$ ($x < 0.4$) and AlAs in HF acid, which was first observed by Konagai *et al.* (*53*). Yablonovitch and co-workers (*39*) used this high selectivity and developed the technique of epitaxial lift-off. The process involves the coating of the sample with wax to exert tension on the film being lifted off. The sample is then soaked in HF acid, which selectively attacks the AlAs release layer. The lifted-off film will then be bonded to the substrate of choice by van der Waals force. It is crucial for the substrate surface to be free of dirt and particles to achieve uniform bonding.

A variety of GaAs transistors (*54, 55*) and lasers (*56*), among other devices (*57*), have been successfully fabricated on Si substrates with these hybrid approaches. The hurdles that remain before the GaAs-Si technology can be commercialized are reliability and manufacturability. In other words, the dislocation density needs to be further reduced for the heteroepitaxy approach, and techniques for the mass production of the films need to be established for the epitaxial lift-off approach. There is one potential problem that is common to all the GaAs-Si structures: mismatch of the thermal-expansion coefficients. This means that the maximum temperature cycle on these structures will have to be limited to prevent catastrophic degradation, such as cracking. For this reason, we expect hybrid integration to be associated with a low-temperature process.

Conclusions

The achievement of light emission from Si is a challenge with a potentially high payoff. There are two issues that must be stressed here: one is overcoming the physics that makes light emission from Si an almost intractable problem; the other is the practicality of the solution. Although impurities, zone folding, and alloying do allow light emission, they are not necessarily practical, and it may not be possible to use these effects in the production of practical devices, such as LEDs and lasers, especially at room temperature. Porous Si, especially if it leads to quantum confinement, suggests that we might be able to engineer fine light-emitting structures for both the study of the physics as well as its practical utilization. We foresee such developments, at the experimental level at least, in the next few years. This is an exciting and challenging task, but our best short- to medium-term approach remains a hybrid approach.

REFERENCES AND NOTES

1. E. Bassous *et al.*, *Tech. Dig.* **1991**, 187 (1991).
2. S. F. Pesarcik *et al.*, *Electron. Lett.* **28**, 159 (1991).
3. G. V. Treyz, P. G. May, J.-M. Halbout, *Appl. Phys. Lett.* **59**, 771 (1991).
4. M. L. Cohen and T. L. Bergstresser, *Phys. Rev.* **141**, 789 (1966).
5. For a good introduction, see J. I. Pankove, *Optical Processes in Semiconductors* (Dover, New York, 1971).
6. L. C. Lenchyshyn *et al.*, *Appl. Phys. Lett.* **60**, 3174 (1992).
7. R. A. Faulkner, *Phys. Rev.* **175**, 991 (1968).
8. H. Ennen, J. Schneider, G. Pomerenke, A. Axmann, *Appl. Phys. Lett.* **43**, 943 (1983).
9. T. G. Brown and D. G. Hall, *ibid.* **49**, 245 (1986).
10. M. O. Henry, E. C. Lightowlers, N. Killoran, D. J. Dunstan, B. C. Cavenett, *J. Phys. C.* **14**, L225 (1981).
11. R. A. Modavis *et al.*, *Appl. Phys. Lett.* **57**, 954 (1990).
12. R. A. Modavis, D. G. Hall, J. Bevk, B. S. Freer, *ibid.* **59**, 1230 (1991).
13. S. Hüfner, *Optical Spectra of Transparent Rare-Earth Compounds* (Academic Press, New York, 1978).
14. J. H. van Vleck, *J. Phys. Chem.* **41**, 67 (1937).
15. H. Ennen *et al.*, *Appl. Phys. Lett.* **46**, 381 (1985).
16. Y. H. Xie, E. A. Fitzgerald, Y. J. Mii, *J. Appl. Phys.* **70**, 3223 (1991).
17. D. R. Wight, *J. Phys. D* **10**, 431 (1977).
18. K. Koga and T. Yamaguchi, *Prog. Cryst. Growth Charact. Mater.* **23**, 127 (1991).
19. M. Ikeda, T. Hayakawa, S. Yamagiwa, H. Matsunami, T. Tanaka, *J. Appl. Phys.* **50**, 8215 (1979).
20. S. Oguz *et al.*, *Appl. Phys. Lett.* **43**, 848 (1983).
21. P. R. Pukite, A. Harwit, S. S. Iyer, *ibid.* **54**, 2142 (1989).
22. N. L. Rowell *et al.*, *ibid.* **58**, 957 (1991).
23. U. Gnutzmann and K. Clausecker, *Appl. Phys.* **3**, 9 (1974).
24. S. Froyen *et al.*, *Phys. Rev. B* **36**, 4547 (1987).
25. M. S. Hybertsen and M. Schluter, *ibid.*, p. 9683.
26. T. P. Pearsall *et al.*, *Phys. Rev. Lett.* **58**, 729 (1987).
27. G. Arbstreiter, *Thin Solid Films* **183**, 1 (1989).
28. H. Presting *et al.*, *Mater. Res. Soc. Symp. Proc.* **256**, 83 (1992).
29. F. Buda *et al.*, *Phys. Rev. Lett.* **69**, 1272 (1992).
30. A. J. Read *et al.*, *ibid.*, p. 1232.
31. L. T. Canham, *Appl. Phys. Lett.* **57**, 1046 (1990).
32. V. Lehman and U. Gösele, *ibid.* **58**, 56 (1991).
33. U. Gösele, V. Lehman, A. Ackaert, L. Buydens, P. van Daele, *Extended Abstracts of the 1992* **1992**, 469 (1992).
34. H. D. Fuchs *et al.*, *Mater. Res. Soc. Symp. Proc.* **256**, 159 (1992).
35. M. Tishler, R. T. Collins, J. H. Stathis, J. C. Tsang, *Appl. Phys. Lett.* **60**, 639 (1992).
36. S. M. Prokes *et al.*, *ibid.* **61**, 1447 (1992).
37. E. Bassous *et al.*, *Mater. Res. Soc. Symp. Proc.* **256**, 23 (1992).
38. N. Khalkoran *et al.*, *ibid.*, p. 89.
39. E. Yablonovitch, D. M. Hwang, T. J. Gmitter, L. T. Florez, J. P. Harbison, *Appl. Phys. Lett.* **56**, 2419 (1990).
40. S. M. Sze, *Physics of Semiconductor Devices* (Wiley, New York, 1981).
41. W. C. Dash, *J. Appl. Phys.* **27**, 1193 (1956).
42. I. V. Klyatskina, M. L. Kozhukh, S. M. Ryvkin, V. A. Trunov, I. S. Shlimak, *Sov. Phys. Semicond.* **13**, 638 (1979).
43. P. M. Petroff and R. L. Hartman, *J. Appl. Phys.* **45**, 3899 (1974).
44. M. Yamaguchi, T. Nishioka, M. Sugo, *Appl. Phys. Lett.* **54**, 24 (1989).
45. E. A. Fitzgerald *et al.*, *ibid.* **59**, 811 (1991).
46. K. Nozawa and Y. Horikoshi, *Jpn. J. Appl. Phys.* **30**, L668 (1991).
47. D. J. Stirland, *Appl. Phys. Lett.* **53**, 2432 (1988).
48. H. Kroemer, *Mater. Res. Soc. Symp. Proc.* **67**, 3 (1986).
49. O. L. Alerhand *et al.*, *Phys. Rev. Lett.* **64**, 2406 (1990).
50. M. G. Lagally, Y. W. Mo, R. Kariotis, B. S. Swartzentruber, M. B. Webb, *Kinetics of Ordering and Growth at Surfaces* (Plenum, New York, 1990).
51. R. J. Fischer *et al.*, *Appl. Phys. Lett.* **47**, 397 (1985).
52. W. A. Harrison, E. A. Kraut, J. R. Waldrop, R. W. Grant, *Phys. Rev. B* **18**, 4402 (1978).
53. M. Konagai, M. Sugimoto, K. Takahashi, *J. Cryst. Growth* **45**, 277 (1978).
54. F. Ren *et al.*, *Electron. Lett.* **24**, 1037 (1987).
55. C. van Hoof, W. D. Raedt, M. van Rossum, G. Borghs, *ibid.* **25**, 136 (1988).
56. E. Yablonovitch, E. Kapon, T. J. Gmitter, C. P. Yun, R. Bhat, *IEEE Photonics Technol. Lett.* **1**, 41 (1989).
57. I. Pollentier *et al.*, *Electron. Lett.* **26**, 193 (1989).
58. T. G. Brown, P. L. Bradfield, D. G. Hall, *Appl. Phys. Lett.* **51**, 1585 (1987).
59. E. Kasper and F. Schaffler, in *Strained-Layer Superlattices: Material Science and Technology*, vol. 33 of *Semiconductors and Semimetals*, T. P. Pearsall, Ed. (Academic Press, New York, 1990), chap. 4.

60. We thank our colleagues at IBM—especially T. P. Smith, J.-M. Halbout, E. Bassous, V. P. Kesan, P. Munguia, S. F. Pesarcik, B. L. Williams, J. C. Tsang, M. A. Tishler, and R. T. Collins—and AT&T—G. A. Baraff, M. S. Hybertsen, J. Bevk, D. G. Hall, E. A. Fitzgerald, P. J. Silverman, and L. C. Feldman—for their collaboration and discussions.