Preprint of D. A. B. Miller, “Optics for Digital Information Processing,” in *Semiconductor Quantum Optoelectronics*, eds. A. Miller, M. Ebrahimzadeh, and D. M. Finlayson, Proceedings of the Fiftieth Scottish Universities Summer School in Physics, St. Andrews, June 1998 (Institute of Physics, Bristol, 1999), pp 433-461

Optics for Digital Information Processing

David A. B. Miller

Ginzton Laboratory,

Stanford University

450 Via Palou

Stanford, CA 94305-4085

http://ee.stanford.edu/~dabm

These lecture notes summarize several related areas in the use of optics in information processing. There are three main sections on, respectively, the physics of optics and electronics for digital information processing, the physics of electroabsorption in semiconductors, and the potential application of both of these to dense optical interconnects.

1. Physics of optics and electronics for digital information processing

Optics is extensively used for long distance communication of information, and is playing an increasing role in networks over shorter distances. It is used too for information storage in optical disks of various kinds. Optics has, however, seen almost no use so far inside digital information processing machines. Despite this lack of use for information processing itself, there are several strong physical arguments why optics may ultimately play an important role in such applications. These physical arguments give reasons why research in such areas should be pursued, they show what directions may be important in such work and what devices and devices attributes may matter most, and suggest new opportunities where optics could fundamentally improve the performance of digital machines. Such discussions are particularly important because the physical reasons for use of optics in the information processing itself are often quite different from those that might be expected based only on experience from telecommunications. Many of the arguments presented here are discussed in greater detail in Ref. [1].

1.1 Differences and similarities between optics and electronics

Optics and electronics have several similarities and differences, which lie at the root of the advantages and disadvantages of each for use in information processing. Because optics is not yet extensively used within information processing machines, we have little practical experience as to its benefits and problems. Technologically, optics also does not yet meet the cost targets for use in high-volume applications. We can, however, use the basic physics of both kinds of approach to deduce where we might try to introduce optics, and why we should do that. If the physical arguments for introducing optics in some particular area are sufficiently compelling, that will give us the motivation to try to develop the necessary optical solutions despite the initial economic difficulties.

In both optics and electronics, essentially all communication is done through electromagnetic waves. It is actually very unusual for signals to be sent on electrons themselves. That may happen inside a cathode ray tube, for example, where the electron beam is modulated and carries the information to the surface of the screen. In general, however, even in electronic systems all of the information actually flows as waves. Sometimes these waves are dissipative, but the proof that the information is carried by the waves is given by the velocity of propagation of the information. This velocity is not generally the velocity of electrons, but is set by electromagnetic parameters such as inductance, capacitance and resistance (per unit length). Hence, the differences between optics and electronics do not stem from information in one case being carried on photons and in the other case being carried on electrons. In both cases they are carried by waves (which we could choose to describe as photons if we wished).

It is also true that essentially all of the interactions, in optics or an electronics, takes place through electrons. In electronics, currents are turned on and off by the action of voltages, which in the end are really due to the Coulomb repulsion of electrons. In the case of nonlinear optics, most nonlinear optical processes actually are a result of the nonlinear motion of the electrons in charge clouds.

Because of these similarities (that all signals are carried by waves and all interaction is through electrons), we need to look somewhat deeper to understand the real differences between optical and electronic approaches. One way of characterizing these differences is shown in Fig. 1.

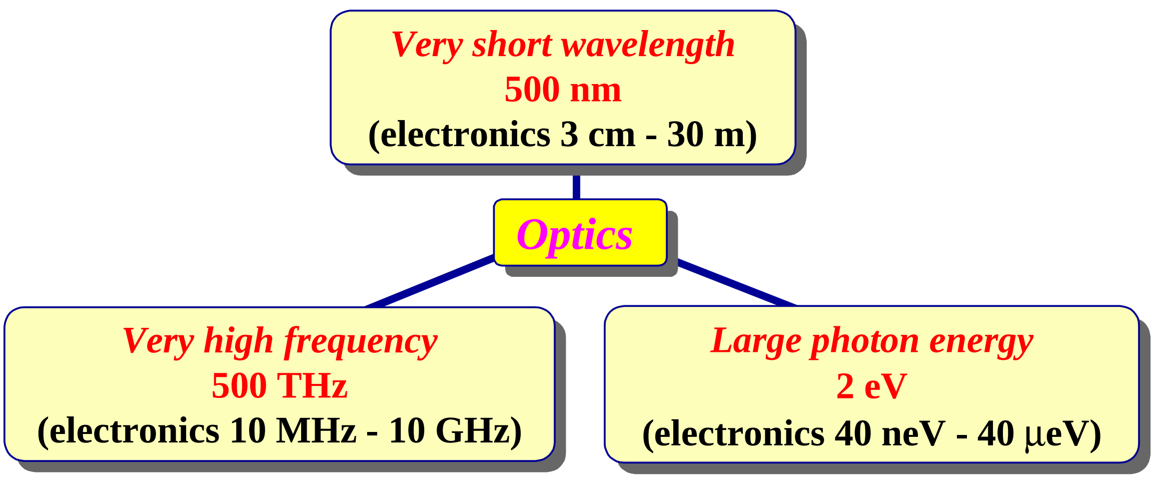

Fig. 1. Differences between optical and electronic physics.

The same fundamental difference between optics and electronics can be expressed in three equivalent ways. First, optics has a very short wavelength compared to the wavelengths of signals within electronic systems. Secondly, optics has a very high frequency compared to the frequencies of operation of electronic systems, and, thirdly, in optics we work with relatively large photon energies. In particular, in optics the photon energy is much larger than any thermal energies (for example, photon energies are of the order of electron volts, whereas thermal energies at room temperature are of the order of 25 meV.) In the electronic case, the opposite situation prevails, with the thermal energy being much larger than the photon energies (which may be in the range of micro eV). As a result, electronic systems are almost exclusively treated in terms of classical gasses of charged particles, whereas in optics we must treat the photons from a quantum mechanical point of view.

We can if we wish choose to categorize the consequences of the differences between optics and electronics for information processing in terms of these three attributes of shorter wavelength, larger photon energy, and higher frequency. Since these three attributes are merely different ways of expressing the same physical difference, this categorization is often arbitrary, but it gives us one way of trying to uncover the many benefits, some of which are by no means obvious at first sight.

1.2 Benefits of the short wavelength of light

Imaging optics

One obvious benefit of the short wavelength of light is the ability to use imaging optics. It is clear that with one lens we can image many points from one surface to another surface. For example, we could imagine that we have a number of optical outputs (for example, small lasers, or modulated light beams) on one surface, an imaging lens, and an array of photodetectors connected to electronic circuits on the another plane. With the imaging lens we could separately form the connections between the different points on the two planes, just us the eye is able to image many different points in the scene onto separate points on the retina. In this operation the information from each point in the scene or on the output chip crosses in space with the information from all of the other points, but the lens is able to sort all the information out and direct only the information from one point to a specific point on the photodetector chip. If we try to perform this imaging operation with electromagnetic waves at normal electronic frequencies, the wavelength would typically be much larger than the objects and we would only be able to resolve 1 separate channel in communicating from one chip to another. It is only because the wavelength is so small in optics that we can resolve so many points (one simple view allows us approximately one separate channel for every square wavelength of area on an input or output surface).

Low-loss interconnects with waveguides

The short wavelength of light has another, less obvious benefit. With light it is possible to guide waves using dielectrics, as for example in an optical fiber. This is only possible because the wavelength of light is significantly smaller than the dielectric waveguide. If we were to try to guide waves on dimensions small compared to a wavelength, we would need very large dielectric constants to give strong enough boundary conditions to bend the propagation of the waves over such small distances. In practice to do such manipulations we tend to need metallic conducting surfaces (which have "infinite" dielectric constants). This kind of situation is very common in electrical circuits, where it is the norm that the wavelength is much larger than the dimensions over which we wish to steer or guide the information. As a result, in electrical circuits we have to use metals to provide strong enough boundary conditions for this relatively extreme movement of information on scales short compared to a wavelength.

In fact, this necessity of guiding the information is the real reason for electrical wires in circuits. Though they do carry current, as we have deduced, the information is not really carried directly by the current, and hence the real function of the wires is to guide waves.

Metallic conductors are always lossy, but it is quite possible to make dielectrics that have very low loss. This difference in the loss mechanisms in the electric and metallic guides is responsible for the ability of optical fiber to carry information with very loss over extremely large businesses. Hence, the short wavelength of light makes it easier for us to make low loss propagation of information over long distances.

Impedance matching

Another of the less obvious advantages of optics, which stems from its short wavelength, relates to impedance matching. In optics it is relatively straightforward to make a simple anti-reflection coating that can essentially eliminate the reflection from a surface. Technically, such a coating is really a resonant, loss-less impedance transformer. It works by multiple reflections of waves within a quarter-wavelength-thick layer. Though it is also possible to make such a resonant impedance matching in electrical systems, such an approach will only work over relatively narrow bandwidth, and hence is not useful for systems that must work from zero frequency up to relatively high frequencies; in practice, therefore, in electronics we have to use terminating resistors to match impedances, leading necessarily to dissipation of energy in the impedance matching process.

The problem of impedance matching is encountered also in buses, where it must be possible to tap the line in the bus at multiple different points, for example to connect multiple different electrical boards. Simply connecting a new board to existing lines will automatically induce impedance matching problems. Typically we will simply put the new "load" in parallel with the wiring; this will change the impedance of the structure at that point since the effective impedance is a line impedance in parallel with the load impedance. Such a discontinuity in impedance will lead to wave reflections and undesired consequences.

The reason for this impedance mismatch is that we have attempted to divided the current between the line and the load without simultaneously dividing the voltage. In the case of an optical system, a simple beam splitter will simultaneously divided both the electric and magnetic fields in such a way that there is no impedance discontinuity and no necessity of back-reflection by the system. Hence optics may be able to solve some of the difficult impedance matching problems often encountered in electrical systems.

Wavelength-division multiplexing

A final benefit that could be attributed to the short wavelength of light is the ability to use wavelength-division multiplexing, that is, to send the information on one optical channel on several different wavelengths, thus allowing several information channels on the one physical channel. This possibility has recently become of great interest for telecommunications systems. It may also be possible to use it for interconnection within systems.

1.3 Benefits of the high frequency of light

Absence of frequency-dependent loss and cross-talk

The consequences of the very high frequency of light are in many ways not distinct from those of the short wavelength of light. Two major important consequences of the high frequency are the absence of frequency dependent “cross-talk” (the leakage of signals from one line to another) and of frequency dependent loss in optical interconnects. The reason for these absences in optics is very simple. In optical systems, all the information is carried by modulating a carrier frequency. Since this carrier frequency is extremely high (10^{15} Hz) the modulations that are imposed on the carrier frequency, even with gigahertz signals, impose negligible fractional additional bandwidth on the signal. As far as the propagation of the light beam is concerned, it is essentially unaffected by this relatively low frequency modulation. Only in propagating over very long distances can we even see any effects such modulation might have on the propagation.

Hence over the size scales that will be encountered inside an information processing machine, there is essentially no frequency dependence to the loss of optical systems, nor is there any frequency dependence of cross-talk. By contrast, in electrical systems the skin effect gives a very strong frequency dependence of loss, and capacitive coupling increases very greatly between wires at higher frequencies, giving strongly frequency-dependent cross-talk. Hence, an electrical system that is designed to work at, for example, 500 MHz, may well not work if we try to increase the operating frequency to even 600 MHz, because the loss and cross-talk (and also wave and impedance matching effects) are substantially different even at these modestly higher frequencies. By contrast, an optical system designed to run at 500 MHz is likely also to be able to run at 500 GHz, because, other than the propagation delay being a larger number of cycles, the optical system sees no difference between these two situations.

Short optical pulses

One ability of optics that is essentially missing in electrical systems is the ability to use and generate very short optical pulses. Mode-locked lasers can readily generate pulses in the picosecond regime or shorter, and, even in conventional optical systems, we can propagate and deliver such pulses with a high degree of synchronization over quite a large system. In electrical systems, it is quite difficult to generate such short pulses, and essentially impossible to propagate them in a large system. This ability to use short pulses opens some different opportunities for optics in systems. One is that we can contemplate delivering very precise clock signals over an entire system. Such deliver the is very difficult in electrical systems, and, in practice, the data and clock need to be re-synchronized in transporting information between different parts of an electrical system. Such re-synchronization can involve quite sophisticated and complex clock recovery circuits and significant buffering of data.

If we use optical modulators as the output devices in an interconnect system, we can also use short pulses to remove the signal skew from the different signals. Skew is the phenomenon that different signals acquire different amounts of delay in propagating through the logic gates and wiring of a system. Skew is obviously undesirable. If, however, we have electrical output signals being used to drive modulators then, even if the precise timing of the signals driving the different modulators is not exactly the same, reading out all of the modulators with synchronized short optical pulses during the time when all the modulator signals are "valid", we can remove the signal skew, generating perfectly synchronous optical outputs as the modulated, synchronized optical pulses. This is illustrated in Fig. 2.

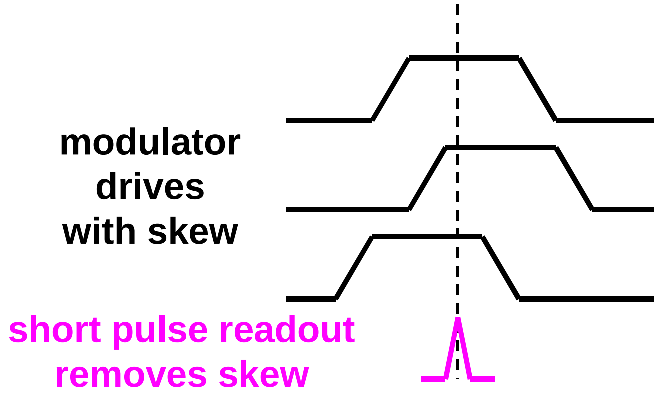

Fig. 2. Illustration of the use of synchronized short pulses to remove skew in the readout of modulators.

Optical logic gates

Yet another possibility with optics, which also relies on the high frequency of light, is the possibility of performing ultrafast logic. There are several ways of making very fast logic gates using optics, and this remains an interesting area of research. At the present time, however, such gates do not seem a practical way of performing any large amount of information processing. For one thing, we cannot contemplate making systems that compete with, for example, silicon technology in terms of complexity, size, or operating power. Optics is not likely to displace silicon as the dominant method of information processing at any time in the foreseeable future. Such high speed gates do, however, give promise for uses in special applications, and the idea of using optics for high speed multiplexing and demultiplexing of signals is one that could be feasible and could perform tasks that are beyond the capability of electronics. More work is, however, needed on such gates and on their integration with systems that are otherwise predominantly electronic.

There are also other problems with many of the proposals for optical logic gates. To be useful in any kind of logic system of even moderate complexity, logic gates must satisfy several criteria. First they must be cascadable, which means it must be possible to drive the next gate with the output of the previous one. Secondly, the gates must have complete logical functionality, that is, it must be possible to make an arbitrary logic system based on the proposed gate; a two-input NOR gate will suffice to give logical completeness, and this criterion is seldom the basic problem for optical logic. Thirdly, the gate must have fan-out; that is, it must be capable of driving at least two other gates. This means that the gate must have signal gain of at least two, and many proposals for logic gates fail on this criterion. Fourthly, a logic gate must have logic level restoration; even though the input may fluctuate somewhat within allowed ranges for logic zero and logic 1, the output logic levels must be much better defined. This fourth criterion is one on which many proposals for logic gates fail.

Among the specific problems for optical logic gates in meeting this these criteria is that, firstly, cascadability must apply to the wavelengths and the beam or pulse shapes; many proposed gates do not satisfy these requirements. In considering restoration of logic levels in optics, it must be remembered that the beam quality has to be restored in the course of the restoration; this means, for example, that simple amplifiers do not necessarily result in proper restoration of the logic level. The same issue of having better quality of outputs than inputs also applies, for example, to pulse shapes. Again, many proposed optical schemes do not satisfy these restoration requirements.

Many proposed optical logic schemes require that some parameter be critically set in order to get the gate to show the necessary functionality or gain; for example, we might set some system just near to the point when it trips over into another state, so that a small change in input gives a large change in output. Such "critical biasing" is unacceptable for any gate the much must work in even modestly complex systems, because it is not in practice reasonable to imagine that we will critically adjust each of several gates in the system; we must have gates that can be used without such adjustments.

Another subtle issue with many proposals for optical logic gates is that they do not have input-output isolation; feeding small to changes in the output back into the system can often upset the system, and such changes fed back into the output of the system may be effectively interpreted by the logic gate as inputs, and hence amplified. Such lack of isolation between output and input is highly undesirable in logic gates; many optical schemes are even reversible, and hence have essentially no input-output isolation. In electronics, one of the main benefits of the transistor is that changes in the output are not fed back to the input of the device. Many electronic logic devices have failed precisely because of their lack of input-output isolation (for example, tunnel diode logic and Josephson devices have both had this problem, which has contributed to their demise).

Hence, the challenge for useful optical logic gates is quite substantial. Not only must we have high speed, low energy operation, but we must satisfy a large number of qualitative constraints to make a device that is a viable option for any system of even moderate complexity. These issues are discussed at greater length in Ref. [2].

1.4 Benefits of the large photon energy of light

The large photon energy of light is in some ways qualitatively the largest difference between optical and electrical systems. One immediate consequence is that essentially all generation and detection of light is done through purely quantum mechanisms (absorption and emission of photons), whereas, in electrical systems, both the generation and detection of voltages and currents is predominantly a classical phenomenon, and can be treated as such for the purposes of nearly all design and modeling.

Voltage isolation

The fact that the we generate and detect light quantum mechanically has several straightforward practical consequences. First of all, any optical connection automatically provides voltage isolation. This point is illustrated in Fig. 3. Here we have a light-emitting diode or laser diode, which can emit one photon for every electron in the ideal case. The light from this emitter then shines onto a detector that generates a current proportional to the number of photons landing on the detector. The relative voltage between the light source and the detector is completely irrelevant. The detector does not work by measuring classical voltages, but rather counts photons. This effect is exploited routinely in devices such as optical isolators that contain a light emitting diode and photodetector in a single package. Any optical connection (including those based on modulators rather than light emitters) will also have this isolating property, and this could turn out to be quite useful in large electrical systems, where it is difficult to maintain uniform ground potentials throughout the entire system.

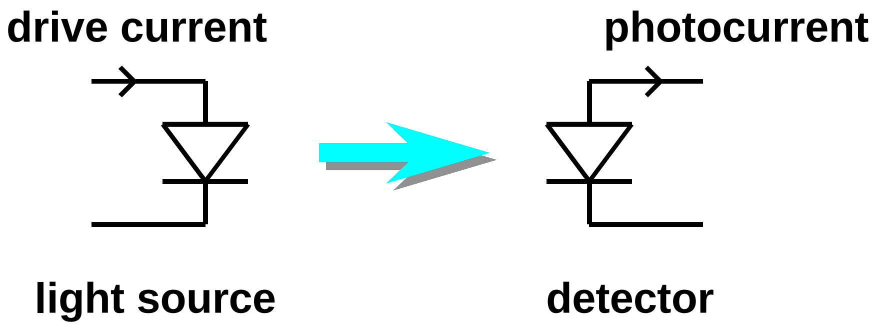

Fig. 3. Illustration of voltage isolation through the use of a light emitter and the photodetector.

Quantum impedance conversion

A second, and more subtle, benefit of the quantum nature of an emission and detection in optics is the process of quantum impedance convention [3]. The individual devices on, for example, a silicon chip are generally small, high-impedance devices. The line that connects one device to another is, however, a line that has either low impedance or a high capacitance per unit length. The low impedance and high capacitance per unit length of electrical lines are essentially practically unavoidable; both of these quantities only vary approximately logarithmically as we change the ratio of the size of conductors to the separation between them (this logarithmic relation is exact, for example, in coaxial lines); as a result, all well-designed lines have essentially the same impedance (~ 50 Ω) and the same capacitance per unit length (~ 2 pF/cm). Line designs with higher impedance or lower capacitance per unit length lead to either high loss in the line (because one conductor becomes very small) or large effective cross-sectional areas in the lines (because the conductor separation becomes very large), both of which are unacceptable in practice. Therefore, there is a basic problem with impedance matching in electrical systems. Though the individual devices want to be small, we typically have to design line drivers that are large so that we can handle the higher currents needed in the high capacitance or low impedance lines.

Optics, however, completely avoids this impedance problem. This is most easily seen by considering an extreme example with a photodetector. Suppose that we take a photodiode and connect it to a 1 GΩ resistor. As we shine light on the diode, we will generate a photocurrent, which in turn will give us a voltage across the resistor. Suppose now that we shine a one nanowatt light beam onto the photodetector. If the photon energy is of the order of one electron-volt, then in an efficient photodiode we will generate approximately one nanoamp of current. This one nanoamp of current could give us one volt across the 1 GΩ resistor. The one nanowatt light beam propagating in free space was essentially one nanowatt in 377 Ω impedance, corresponding to approximately 600 μV of classical voltage in the light beam. The photodiode has essentially transformed 600 μV in 377 Ω into one volt in 1 GΩ. The reason why it does this is because of the photoelectric effect, fundamentally a quantum process. Hence a photodetector can be a quantum impedance converter, and can be relatively efficient at this task.

The same kind of impedance convention can operate at the transmission end. We could, for example, have a very efficient laser diode that gave us one photon out for every electron of current in. Such a device would also be able to transform from a high impedance drive into the low impedance of free space. Such light emitters are becoming more feasible, though at the moment it is still quite difficult to get laser diodes to be very efficient at low currents (and light-emitting diodes are never very efficient at converting electrons of current into usable photons). Another attractive option, however, is to use modulators. For example, the quantum well modulator (discussed below) usually requires that we pass one electron of current for each photon to be absorbed, and so it also performs the same kind of impedance transformation; such modulators can operate efficiently to arbitrarily low optical powers. Hence with optoelectronic devices we can solve the problem of the low impedance and high capacitance of electrical lines. In practice, to achieve an advantage this way requires that the optoelectronic devices are integrated with low capacitance into the electrical circuit. In the end, the amount of energy we need to transmit is determined by the number of electrons we need to generate in the photodetector to discharge its internal capacitance (and that of the associated input circuit) to the required voltage swings. Hence, very good integration technology is required to take advantage of this impedance transforming benefit of optics. This impedance conversion does, however, give us a distinct possibility for reducing the power consumption of interconnects through use of optical and optoelectronic technology.

2. Physics of electroabsorption in semiconductors

There are many reasons for being interested in the use of optics in information processing, as summarized above. There is, however, a need for mechanisms and devices that can exploit the advantages of optics. Historically, the absence of microscopic physical mechanisms that were strong enough (e.g., required only small energies to operate), and additionally were compatible with interfacing with silicon electronics, has held back the use of optics in mainstream information processing.

One of the more promising mechanisms to emerge in recent years has been electroabsorption in quantum well semiconductors. Such effects are strong, and the devices are compatible with electronics, allowing potentially very large numbers of them to be integrated with electronics. In this section, we summarize the physics of electroabsorption in semiconductors in general, leading up to the description of the effects in quantum wells. An extensive discussion of the physics of interband optical properties of quantum wells can be found in Ref. [4], including the electroabsorption mechanisms discussed here.

Electroabsorption generally is some effect whereby absorption of a material changes in response to an electric field. The first such mechanism to be proposed for semiconductors is the so-called Franz-Keldysh effect [5][6][7][8]. With the advent of quantum wells, this subject received renewed interest because quantum wells show especially strong electroabsorption effects, in particular the quantum-confined Stark effect (QCSE) [9].

Such electroabsorption mechanisms are practically interesting since they enable electrically driven optical modulators in semiconductors. They only require electric fields to be applied to the semiconductor, and so they are intrinsically very fast mechanisms (no carrier injection is required, and hence the device is not limited by the time constants associated with carrier population changes). The necessary fields are usually applied by reverse-biasing a diode structure containing the semiconductor material.

The Franz-Keldysh effect in bulk materials allows quite effective waveguide modulators to be made, though it is not a strong enough mechanism for simple high-contrast modulators for light perpendicular to the wafer surface; in practical diode structures, fields can only be applied over microns of thickness, and so the vertical diodes that can be conveniently fabricated are not thick enough given the absorption changes possible in the Franz-Keldysh effect. The QCSE does allow modulators that can work very effectively in only microns of thickness, and opens up many device possibilities. In addition, the QCSE makes good waveguide modulators, and also permits modulators made using the same layer structure as modern quantum well laser diodes, thus allowing integrated laser/modulator structures, for example.

The theory and mechanisms of electroabsorption in direct gap semiconductors will be covered below. Briefly, the history is as follows. The Franz-Keldysh mechanism was proposed, based on a non-excitonic model of optical absorption in direct gap semiconductors (the excitonic absorption models were still being developed at that time). This led to a simple picture that predicted that, with applied field, additional absorption would appear for photon energies just below the bandgap energy, and so-called "Franz-Keldysh" oscillations would appear in the absorption above the band gap. Both these phenomena are observed qualitatively in experiments. Actual measurements of electroabsorption also show, however, that there is an additional mechanism that results in the broadening of the exciton peak with applied field. Later theories [10][11] properly modeled these excitonic effects, and show that, in fact, the exciton broadening phenomena strongly dominate the actual induced absorption just below the band gap energy. The emergence of quantum wells, with their strong excitonic peaks at room temperature, led to empirical observations of strong electroabsorption [9], and the clear observation of very different electroabsorptions for the distinct cases of electric field parallel to the layers (which shows electroabsorption qualitatively similar to bulk excitonic electroabsorption) and perpendicular to the layers. The case of field perpendicular to the layers (QCSE) [9] gives rise to strong shifting of the absorption edge, *without* strong exciton broadening. This QCSE effect was subsequently related to the earlier Franz-Keldysh models [8], and shown to be the quantized version of these bulk mechanisms.

2.1 Formalisms for optical absorption in direct gap semiconductors

Before proceeding to examine electroabsorption, it is useful to remind ourselves of the formalism for calculating absorption spectra in semiconductors. There is not space here to derive the various expressions, but we will attempt to sketch the formalism. To work out optical absorption, we can write an expression of the form

$$W_{TOT} = \frac{2\pi}{\hbar} \sum_i \sum_f \left| \left\langle f \left| H'(\mathbf{r}) \right| i \right\rangle \right|^2 \delta\left(E_f - E_i - \hbar\omega\right) \tag{1}$$

for the total transition rate (rate of absorption of photons), W_{TOT}, for photons of energy $\hbar\omega$. Here we are considering each possible initial state (i) of the system and each possible final state (f). $H'(\mathbf{r})$ is the "perturbing Hamiltonian" corresponding to a monochromatic electromagnetic field interacting with the semiconductor. To make a transition from an initial state to final state, we are requiring that the energy difference between these states matches the photon energy $\hbar\omega$, hence the δ-function. We also find (from time-dependent perturbation theory) that the strength of any such transition is proportional to the squared "matrix element" $\left|\left\langle f\left|H'(\mathbf{r})\right|i\right\rangle\right|^2$.

The most common way to consider the initial state i is to say that it corresponds to an electron in a specific state in the valence band, and similarly, the final state f corresponds to an electron in a specific state in the conduction band. Such an approach can be used when we neglect any interaction (Coulomb attraction) between the electron in the conduction band and the "hole" (absence of electron) it leaves behind in the valence band. It is the approach used to calculate optical absorption in semiconductors when we neglect all excitonic effects (effects from the Coulomb attraction of the electron and hole). In this picture, optical absorption between two states can be reduced to considering the overlap integrals of the electron ($\phi_e(\mathbf{r})$) and hole ($\phi_h(\mathbf{r})$) "envelope" wavefunctions . The unit cell parts of the wavefunction take care of the other aspects of the matrix element, including the usual requirement of opposite parities for wavefunctions for allowed dipole transitions, giving the constant $\left|p_{cv}\right|^2$ that now appears outside the summation. Hence the optical absorption takes on the form, from Eq. (1),

$$W_{TOT} \propto \left|p_{cv}\right|^2 \sum_e \sum_h \left|\left\langle \phi_e \middle| \phi_h \right\rangle\right|^2 \delta\left(E_e - E_h - \hbar\omega\right) \tag{2}$$

where e refers to electron states in the conduction band, and h refers to states in the valence band. The fact that the electron and hole wavefunctions are just plane waves in an unbiased semiconductor if we neglect electron-hole Coulomb attraction leads to the simple "momentum conservation" requirement (both electron and hole plane waves must have essentially the same wavevector), and the usual "$E^{1/2}$"optical absorption edge for direct gap semiconductors (neglecting excitons) once we sum over the states.

In the case of quantum wells, neglecting excitons, we can conveniently sum over the states corresponding to motion in the plane of the layers (the x and y directions); such in-plane motion states turn out to have uniform densities of states (and momentum conservation also applies to such in-plane motion), and so we obtain absorption that is a set of steps, which can formally be written as

$$W_{TOT} \propto \left|p_{cv}\right|^2 \sum_{nc,nv} \left|\left\langle \psi_{nc} \middle| \psi_{nv} \right\rangle\right|^2 \Theta\left(\hbar\omega - E_{nc} - E_{nv} - E_g\right) \tag{3}$$

where $\Theta(x)$ is the step function, nc and nv index the conduction and valence subbands respectively, ψ_{nc} and ψ_{nv} are the "quantum-confined" wavefunctions in the direction perpendicular to the layers (the z direction) for each of these subbands, and E_{nc} and E_{nv} are the energies of the edges of these subbands. This formalism gives a quantum well absorption that is a series of steps, the height of the steps being proportional to the overlap integrals (squared) for each valence and conduction subband of interest ($\left|\langle \psi_{nc} | \psi_{nv} \rangle\right|^2$).

To understand excitonic effects, we use another way of characterizing the states in which we consider the states of the electron-hole pair as a unit. (By a mathematical accident, this method also turns out to be useful for analyzing electroabsorption in bulk semiconductors even neglecting excitonic effects, as we will see below.) In this way of looking at the problem, there is only one initial state of interest, which is the state of the crystal in which there are no electron-hole pairs. The excited ("final") states of interest correspond to states of the crystal in which there is one electron-hole pair, and so the sum over excited states then becomes the sum over all possible states of one electron hole pair. Now, instead of considering electron wavefunctions $\phi_e(\mathbf{r}_e)$ and hole wavefunctions $\phi_h(\mathbf{r}_h)$, we instead consider the "relative motion" wavefunction $\phi_{eh}(\mathbf{r}_{eh})$ and a "center of mass" motion wavefunction. This kind of approach works mathematically any time we have a potential energy that only depends on the relative separation of the electron and hole, which is the case of excitons (where we include the Coulomb attraction of the electron and hole) and also for the particular case we consider first in the next section, which is that of electroabsorption neglecting the Coulomb attraction, because the potential energy term we introduce to describe the effect of electric field on the (non-interacting) electron and hole happens to have the same mathematical characteristic of depending only on the separation of the electron and hole.

For example, the Schroedinger equation for a single electron-hole pair with Coulomb attraction, which can be written

$$\left[-\frac{\hbar^2}{2m_e}\nabla_e^2-\frac{\hbar^2}{2m_h}\nabla_h^2-\frac{e^2}{4\pi\varepsilon_r\varepsilon_o\left|\mathbf{r}_e-\mathbf{r}_h\right|}\right]\mathbf{\Phi}_{eh}(\mathbf{r}_e,\mathbf{r}_h)=E_{eh}\mathbf{\Phi}_{eh}(\mathbf{r}_e,\mathbf{r}_h), \tag{4}$$

where the subscripts e and h refer to electron and hole coordinates respectively, can be separated using "center of mass" coordinates in the usual manner to give a relative motion Schroedinger equation

$$\left[-\frac{\hbar^2}{2\mu_{eff}}\nabla_{\mathbf{r}}^2-\frac{e^2}{4\pi\varepsilon_r\varepsilon_o\left|\mathbf{r}_e-\mathbf{r}_h\right|}\right]\phi_{eh}(\mathbf{r})=E_{rel}\phi_{eh}(\mathbf{r}) \tag{5}$$

where μ_{eff} is the reduced effective mass, as well as a center of mass equation whose solutions are plane waves (corresponding to bodily movement of the entire electron hole pair). This equation gives rise to the hydrogenic "excitonic" states (including discrete states below the band edge, and a continuum of strongly correlated states (with enhanced optical absorption) above the band edge).

In the approach where we consider the electron-hole pair as a unit, the strength of the optical absorption that "creates" this electron hole pair turns out to be proportional to $\left|\phi_{eh}(0)\right|^2$, i.e., the probability of finding the electron and hole in the same place. (Note the terminology now is that we are creating the electron-hole pair – initially, there were no electron-hole pairs, and now, after absorbing a photon, there is one electron-hole pair.)

This electron-hole pair picture substitutes $\left|\phi_{eh}(0)\right|^2$ for $\left|\langle\phi_e|\phi_h\rangle\right|^2$; this electron-hole pair picture gives the same answers as the simple separate electron and hole picture in cases where there is no electron-hole interaction, but goes on to be valid even when there is electron hole interaction (as in the case of excitons), or in the presence of potentials that only depend on the relative position of electron and hole (as we will see below for the case of the Franz-Keldysh effect). In this approach, therefore, we have an optical absorption of the form

$$W_{TOT} \propto \left|p_{cv}\right|^2 \sum_f \left|\phi_{eh}(0)\right|^2 \delta\left(E_f - E_i - \hbar\omega\right) \tag{6}$$

In the case of the quantum well, we can often use what we can call a "strong confinement" approximation, in which we presume the wavefunctions of the electron and hole perpendicular to the layers are not strongly affected by the Coulomb attraction. Hence the wavefunction of interest becomes one that has relative electron and hole coordinates in the plane of the layers but absolute electron and hole coordinates perpendicular to the plane of the layers. The transition rate in optical absorption therefore becomes of the form

$$W_{TOT} \propto \left|p_{cv}\right|^2 \sum_{nc,nv,nxy} \left|\langle\psi_{nc}|\psi_{nv}\rangle\right|^2 \left|\phi_{nxy}(0)\right|^2 \delta\left(\hbar\omega - E_{nxy} - E_{nc} - E_{nv} - E_g\right) \tag{7}$$

which is a kind of "hybrid" expression between the two previous approaches. Here, $\phi_{nxy}\left(\mathbf{r}_{xy}\right)$ is the relative motion wavefunction of the electron-hole pair in the plane of the quantum well layers in some state labeled by an integer nxy, and will in practice be some "excitonic" orbital in the plane.

2.2 Electroabsorption mechanisms neglecting excitons

Franz-Keldysh effect

There are two ways to explain the Franz-Keldysh effect (both of which are ultimately exactly equivalent). In both pictures, we make use of the fact that the conduction and valence bands are "tilted" by the presence of the electric field, **E** (see Fig. 4.) In the first explanation, (which is essentially the explanation proposed by Franz and Keldysh), the process is described in a "time-dependent" manner. The time-dependent picture has a relatively simple "intuitive" explanation; the photon can be viewed as raising the electron from the valence band into the bandgap region below the conduction band, from where it can "tunnel" into the conduction band because the conduction band is tilted. The "time-dependent" tunneling picture, though intuitively appealing, is less simple to calculate. The second picture [7] is an "eigenstate" picture that is relatively simple to calculate, reducing to a simple calculation of overlaps of initial and final states.

Let us examine the problem of an electron (or a hole) in a bulk semiconductor in the presence of an electric field in the z direction. Adding the electrostatic potential from the field in the z direction gives the Schroedinger equation for the electrons (in the z direction)

$$\left[-\frac{\hbar^2}{2m_e}\frac{\partial^2}{\partial z_e^2}+V_e-e\mathbf{E}z\right]\phi_e(z_e)=E_c\phi_e(z_e) \tag{8}$$

Here V_e is the energy of the edge of the conduction band in the absence of electric field. We know the solutions to this equation are Airy functions, obtaining the "*Ai*" "oscillatory" function for positions z where the eigen energy $E_c > V_e - e\mathbf{E}z$ (i.e., to the left of the point where the dashed line crosses the conduction band edge in Fig. 4), and the "*Bi*" decaying functions otherwise, as sketched in Fig. 4. We will not go through the details here of the Airy functions, though this is straightforward mathematically.

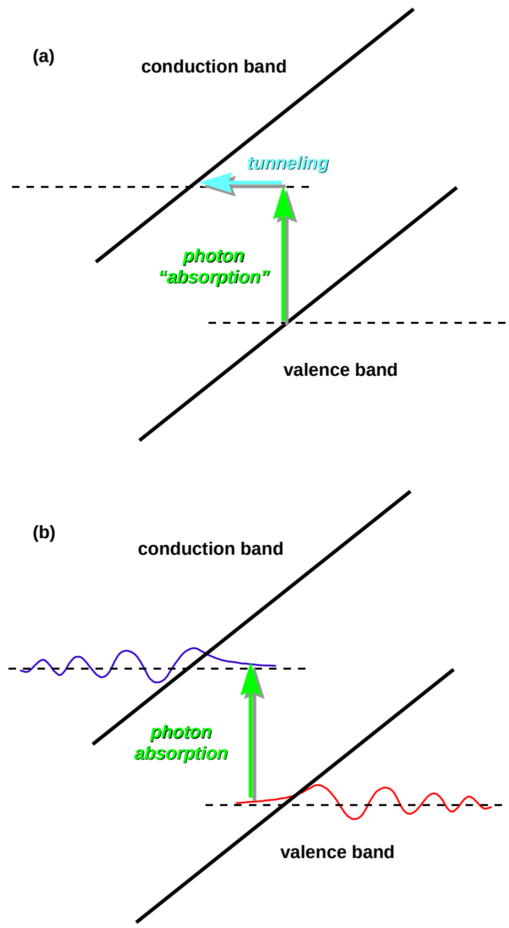

Fig. 4. Illustration of the model pictures of the Franz-Keldysh effect. (a) "photon-assisted tunneling" picture; (b) absorption between eigenstates.

Note also that all of the eigenfunctions for any energy E_c have exactly the same form; they are merely shifted sideways from one another by an amount corresponding the electrostatic potential drop, i.e., the eigenfunction for energy E_{c1} is the same as that for energy E_{c2} except that it is shifted in the z direction by an amount $z_{shift} = (E_{c1} - E_{c2})/e\mathbf{E}$.

The behavior of holes is exactly analogous, though of course the hole wavefunctions go in the other direction. Hence we would have the equation for the holes

$$\left[-\frac{\hbar^2}{2m_{hz}}\frac{\partial^2}{\partial z_h^2}+V_h+e\mathbf{E}z\right]\phi_h(z_h)=E_v\phi_h(z_h) \tag{9}$$

Having worked out allowed states and wavefunctions, we now want to evaluate optical absorption.

For our Franz-Keldysh problem, we could take the simple approach of evaluating all of the overlap integrals between electron and hole states. We would end up with integrals over Airy functions, which is a little inconvenient to do, though can be done.

If, however, we write the Schroedinger equation for the electron-hole pair in the presence of the field, we obtain, adding the Hamiltonians from the separate electron and hole problems,

$$\left[-\frac{\hbar^2}{2m_e}\frac{\partial^2}{\partial z_e^2}-\frac{\hbar^2}{2m_h}\frac{\partial^2}{\partial z_h^2}+E_g-e\mathbf{E}(z_e-z_h)\right]\mathbf{\Phi}_{eh}(z_e,z_h)=E_{eh}\mathbf{\Phi}_{eh}(z_e,z_h) \tag{10}$$

Here we have taken the energy origin at the top of the valence band. Note that the "potential" term in this equation only depends on the relative position of the electron and hole, z_e - z_h. Hence we can formally separate into center of mass coordinates to get a plane wave motion for the center of mass motion, which we need consider no further, and an equation for the relative motion wavefunction $\phi_{eh}(z_{rel}$ (= z_e - z_h)), of the electron and hole

$$\left[-\frac{\hbar^2}{2\mu_{eff}}\frac{\partial^2}{\partial z_{rel}^2}+E_g-e\mathbf{E}z_{rel}\right]\phi_{eh}(z_{rel})=E_{FK}\phi_{eh}(z_{rel}) \tag{11}$$

This equation is mathematically essentially identical to the equations (8) and (9), and has solutions that are simply Airy functions. Hence the relative motion wavefunction of the electron-hole pair is also an Airy function (see Ref. [7]). The optical absorption spectrum is now relatively simply calculated using the expression (6).

The Franz-Keldysh model predicts two interesting consequences, both of which can be seen in Fig. 5. First (and most usefully for devices), it predicts that by turning on an electric field we can induce an absorption below the bandgap energy in a semiconductor, turning it from being transparent to being absorbing, a fact that we could use for optical modulators. Second, it predicts that there are slight oscillations in the optical absorption above the bandgap energy. These come about because, as we perform the appropriate overlap integrals for different energies, we are "sliding" the electron and hole wavefunctions past one another as we change energy. For energies above the bandgap energy, the oscillatory parts of the wavefunctions overlap one another, and so the overlap integral tends to oscillate as the energy changes.

Quantum well electroabsorption neglecting excitons

In a quantum well with an electric field, we need to add the appropriate electrostatic potential to the electron and hole Schroedinger equations. Specifically, if the field perpendicular to the layers is **E**, the resulting equations become, for the electron

$$\left[-\frac{\hbar^2}{2m_e}\frac{\partial^2}{\partial z_e^2}+V_e(z_e)-e\mathbf{E}z\right]\phi_e(z_e)=E_{nc}\phi_e(z_e) \tag{12}$$

and for the hole

$$\left[-\frac{\hbar^2}{2m_{hz}}\frac{\partial^2}{\partial z_h^2}+V_h(z_h)+e\mathbf{E}z\right]\phi_h(z_h)=E_{nv}\phi_h(z_h) \tag{13}$$

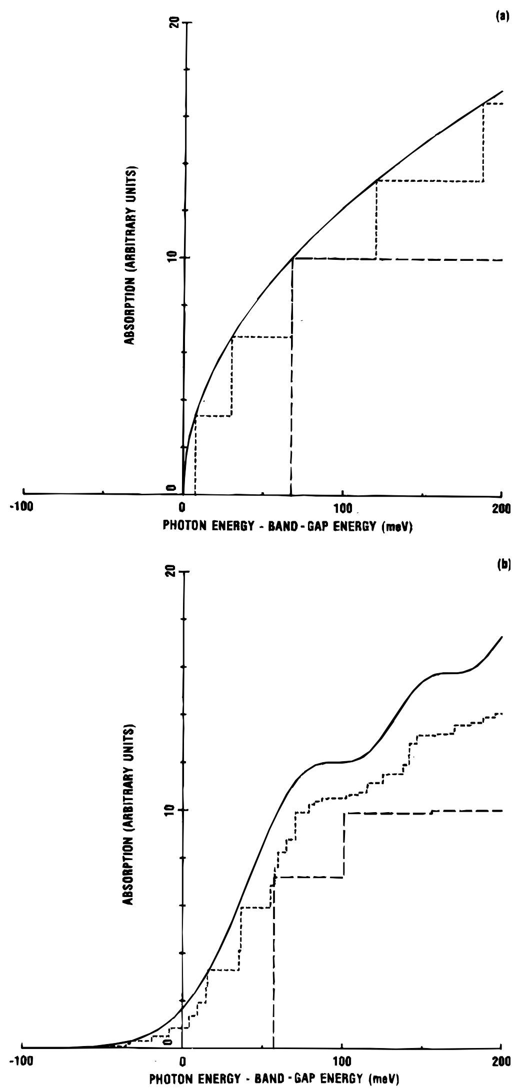

Fig. 5. Form of the theoretical Franz-Keldysh optical absorption (smooth curve) and the quantum well absorption (neglecting excitons) (steps) at 10^5 V/cm [8]. The quantum well calculations are for GaAs-like infinite quantum well. For the upper graph, the quantum well transitions are labeled (heavy hole subband number, electron subband number), and the quantum well is 15 nm thick. For the lower graph, the short dashed line is for a 30 nm thick quantum well, and the long dashed line is for a 10 nm thick quantum well.

These differential equations are mathematically essentially identical to Eqs. (8), (9), or(11), except that we have the additional "quantum well" potentials, $V_e(z_e)$ or $V_h(z_h)$. When these potentials are a set of steps (i.e., mathematically, the potentials are "piece-wise constant"), as in the case of a simple quantum well, we therefore find the same kind of mathematical solutions as obtained above in the bulk case. Those solutions are Airy functions [9]. The principal difference in the quantum well case is that, within the well, we have both "forward" and "backward" Airy functions because of the discrete reflections off of the walls of the quantum well. Fig. 6 illustrates solutions for the various quantum well levels in the valence and conduction bands for a GaAs-like infinite quantum well.

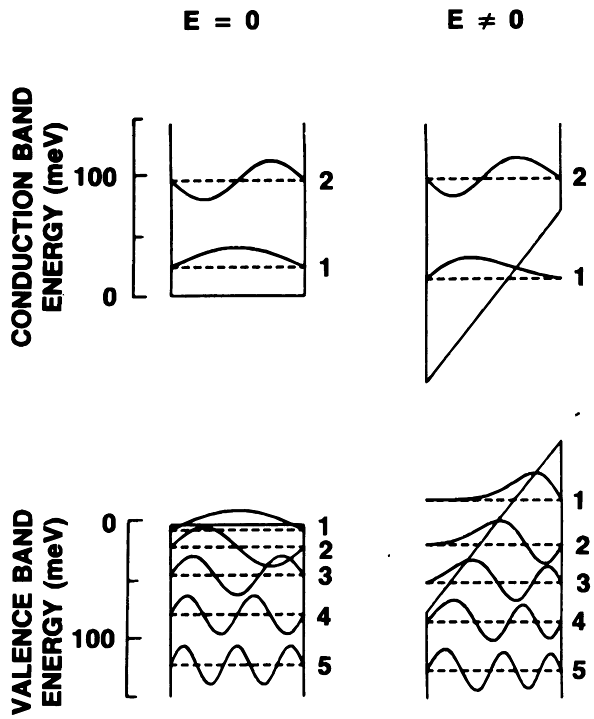

Fig. 6. Energy levels and wavefunctions for valence and conduction states in a quantum well with an electric field applied perpendicular to the layers [8]

Note that the electron and hole, at least in the lowest subband, are pulled in opposite directions by the field. Note also that the electron and hole quantum well wavefunctions are no longer in general orthogonal. Hence many previously forbidden transitions now become allowed.

In practice, we will likely not use the Airy functions, since direct numerical calculations (e.g., by the tunneling resonance method) are usually easier for actual quantum well structures, but the formal connection to the bulk (non-excitonic) electroabsorption is interesting.

Neglecting excitonic effects, the optical absorption spectrum would still be described by the same general formula as before, i.e., Eq. (2), with the new wavefunctions and energies as in Fig. 6. Now, however, the energy of the transitions will be shifted and we will have many more possible, previously "forbidden" transitions. Fig. 5 shows the results of a quantum well calculation (for an infinite well) as the thickness of the well is changed from 10 nm, through 15 nm, to 30 nm. The 10 nm case shows very strongly "quantized" behavior that is qualitatively little like the Franz-Keldysh absorption. In the 30 nm case, the sum of all of the large number of possible transitions leads to an absorption that is qualitatively almost identical to the Franz-Keldysh result, including the "Franz-Keldysh oscillations" for photon energies above the bandgap energy. The 15 nm case is intermediate, with the strongly quantized behavior still evident and the trends of the Franz-Keldysh effect also emerging.

In fact, the quantum well problem is actually equivalent to the Franz-Keldysh problem if we solve it for a very thick quantum well; the exact equivalence has been proven [8]. As we make the quantum well thicker therefore, its absorption, in this non-excitonic model, limits to the Franz-Keldysh effect; hence we could choose to refer to the quantum well electroabsorption in this limit as the quantum-confined Franz-Keldysh effect, though this term may be of only academic interest since the quantum-well electroabsorption with excitons (which is the real mechanism) is better known as the quantum-confined Stark effect, as we will explain below. In the calculations for the 10 nm quantum well in Fig. 5, we can already see one of the particularly desirable effects in quantum well electroabsorption, which is the movement of a discrete, abrupt absorption edge to lower photon energies with perpendicular field. To understand the actual electroabsorption in quantum wells, however, we have to understand the behavior of the excitons, which dominate the absorption near the band gap, and we return to this below.

2.3 Excitonic electroabsorption

Excitonic electroabsorption in bulk semiconductors

The models of electroabsorption discussed so far have neglected excitons. Excitonic effects turn out to be very important in electroabsorption, and also lead to qualitatively different behavior in quantum wells and bulk semiconductors.

An exciton is physically like a hydrogen atom, and so we might expect the same kind of effects on exciton levels as we would see in a hydrogen atom. In particular, we might expect Stark shifts of the levels (Stark shifts being the term applied to levels shifts with electric field in atoms). Such an effect can be seen with excitons. Usually we are looking at the energy of the ground state of the exciton, since the 1*S* excitonic peak is the dominant feature we can examine for shifts. Fig. 7 illustrates in a somewhat simplistic way the distortion of the exciton potential and classical orbit as we apply a field to it. The exciton becomes polarized by the field, with the electron on the average moving in one direction and the hole in the other. Both the electron and the hole have "run downhill" towards the respective electrodes, reducing their potential energy; the reduction may be partially offset by some increases of kinetic energy because of the distortion of the wavefunction, but the net result is a slight decrease in the energy of the exciton. This decrease in energy is approximately quadratic in the electric field; it cannot be linear because the 1*S* exciton is spherically symmetric to start with. As a result, when we look at the absorption spectrum that leads to the creation of the 1*S* exciton, we might expect to see a quadratic decrease in energy with field since the exciton is created in this polarized state with lower energy.

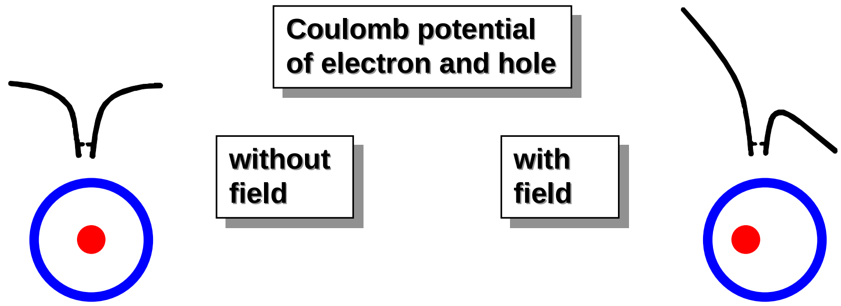

Fig. 7. Illustration of the electron-hole Coulomb potential and the classical orbit of the electron and hole as electric field is applied to the exciton.

This quadratic Stark shift does exist, but it is not a very large effect in the optical absorption, and it is not the dominant excitonic electroabsorptive effect at the fields we normally apply in electroabsorptive devices. In a hydrogenic system, the Stark shift of the ground state is limited to perhaps 10% of the binding energy. Excitons have binding energies of ~ 4 meV in bulk GaAs, thereby limiting the shift to < 1 meV, an amount too small to be useful for a device.

There is, however, a relatively strong electroabsorption associated with the exciton in bulk materials. This effect results essentially from the lifetime broadening of the exciton that results from its rapid ionization by the electric field. We can see from the form of the potential sketched in Fig. 7 that, when the potential is "tilted" by the field, it is possible for the electron and hole to tunnel through the potential barrier that "binds" them. If this tunneling is very rapid (e.g., if it is short compared to a classical orbit time), the broadening can be of the order of the binding energy, and that broadening can give substantial change of absorption in situations where the exciton absorption peak was initially clear (e.g., low temperature in bulk materials).

In a typical semiconductor situation, it is straightforward to apply electric fields of 10^4 - 10^5 V/cm. Such fields correspond to 1 - 10 V/μm, the kind of fields readily encountered in semiconductor diodes, for example. 1 V/μm corresponds to 10 meV across 10 nm, i.e., substantially more than one binding energy across one exciton diameter. Such fields are therefore huge as far as the exciton is concerned, and it is easy to believe that the exciton could be field ionized very rapidly in such fields. In fact, in the region below the bandgap energy, the excitonic effects dominate most of the electroabsorption, and one very strong effect is the broadening of excitonic absorption peaks with field. In fact, it is arguable whether the original Franz-Keldysh calculation has any validity at all in this region. (The Franz-Keldysh oscillations above the bandgap are preserved in the excitonic calculation, however.) Despite this inaccuracy of the Franz-Keldysh model, the electroabsorption below the bandgap energy is still typically referred to as the Franz-Keldysh effect, perhaps because the excitonic electroabsorption effect has no particular name. We will return to this line-broadening electroabsorption below in discussing quantum well electroabsorption.

The model including excitons is not particularly difficult in principle, though unfortunately it does not have any very good closed-form results. Formally, we take the electron-hole pair Schroedinger equation, Eq. (5), and add in the potential energy of the electron and the hole from the electric field in the z direction. As we noticed above in examining the formal theory of the Franz-Keldysh effect, this additional term, $-e\mathbf{E}(z_e - z_h)$, is only a function of the relative position of the electron and the hole. Hence, we can still formally separate the electron-hole Schroedinger equation into center of mass coordinates and relative motion coordinates. The center of mass motion is still simply a plane wave (the electric field does not influence the center of mass motion since there is no net force on the electron-hole pair since overall it is neutral). Adding potential energy $-e\mathbf{E}(z_e - z_h)$ to the relative motion Schroedinger equation (5), we obtain

$$\left[-\frac{\hbar^2}{2\mu_{eff}}\nabla_{\mathbf{r}}^2 - \frac{e^2}{4\pi\varepsilon_r\varepsilon_o\left|\mathbf{r}_e - \mathbf{r}_h\right|} - e\mathbf{E}\left(z_e - z_h\right)\right]\phi_{eh}\left(\mathbf{r}\right) = E_{rel}\phi_{eh}\left(\mathbf{r}\right) \tag{14}$$

To work out the excitonic electroabsorption in the bulk, therefore, we have formally to solve for the eigenstates of this equation. In principle, knowing all of the allowed eigenenergies and all of the associated relative motion wavefunctions allows us to work out the optical absorption using the fact that the absorption strength for creation of an electron-hole pair in a given state is proportional to $\left|\phi_{eh}\left(0\right)\right|^2$ (Eq. (6)). This bulk electroabsorption has been calculated and discussed by, e.g., Dow and Redfield [10] and Merkulov and Perel [11].

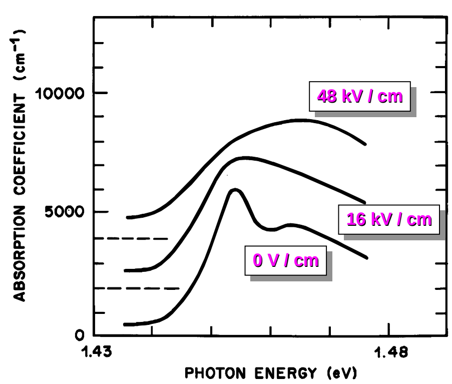

Fig. 8. Absorption of a GaAs/AlGaAs quantum well sample at room temperature for various values of electric field applied in the plane of the quantum well layers (i.e., parallel to the quantum wells). The spectra are shifted vertically for clarity [9].

Excitonic electroabsorption in quantum wells for field parallel to the layers

The situation for electroabsorption in quantum wells for electric fields parallel to the layers is qualitatively similar to bulk excitonic electroabsorption. The main difference in the quantum well case is that the exciton absorption peak is visible at room temperature, and hence the electroabsorption is conspicuously dominated by exciton lifetime broadening, a fact that is not usually so clear in bulk materials at room temperature. Fig. 8 shows absorption spectra of a room temperature GaAs/AlGaAs quantum well sample. With increasing field, the exciton peaks broaden and disappear. The slight Stark shift that one might expect to see (which would be a shift to *lower* energy of the peak) is not obviously resolved here, and the peak appears to shift to higher energy. This higher energy shift is in a sense only an apparent shift; the exciton has long since ceased to exist as a meaningful bound state.

We will not go through the theory of this electroabsorption formally. To do the theory we would likely start by taking the strong confinement approximation that would allow us to separate the wavefunction. We would then be left to deal with a "quasi-2D" problem for the *x-y* wavefunction of the "exciton" in the presence of a field in the plane of the quantum wells. This problem is related to the problem of the 2D exciton in the presence of an in-plane field, which has been treated by Lederman and Dow [12]. It is qualitatively similar to the 3D excitonic electroabsorption problem, showing exciton broadening and the appearance of Franz-Keldysh oscillations.

Excitonic electroabsorption in quantum wells for field perpendicular to the layers - quantum-confined Stark effect

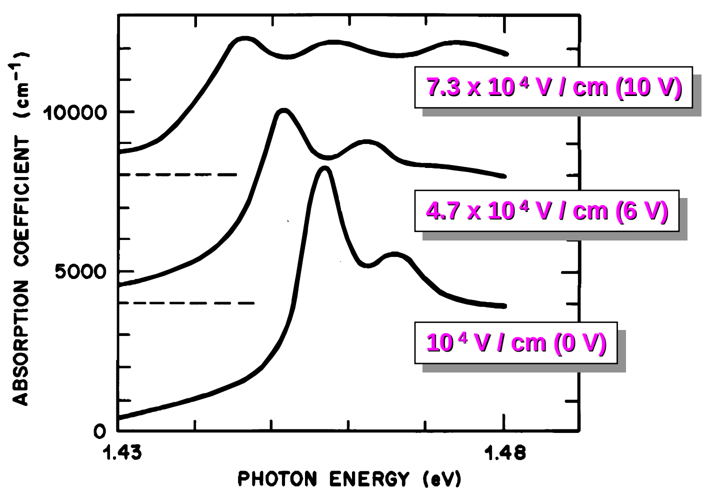

Fig. 9. Optical absorption spectra for a GaAs/AlGaAs quantum well sample at room temperature for various electric fields applied perpendicular to the quantum well layers. The spectra are shifted vertically for clarity [9].

When we apply electric fields perpendicular to quantum well layers we see an effect that is essentially unique to quantum wells, and which enables a different class of devices. This effect is known as the quantum-confined Stark effect (QCSE) [9]. A typical set of absorption spectra is shown in Fig. 9. To see spectra like this, we can construct a "quantum well diode" as shown in Fig. 10.

The distinguishing characteristics of this electroabsorption compared to bulk (excitonic) electroabsorption are

(i) the absorption edge shifts with field,

(ii) the exciton absorption peaks remain resolved.

The qualitative explanation for this effect is quite simple. Consider the lowest exciton state. When we apply an electric field perpendicular to the layers, we pull the electron in the exciton in one direction, and the hole in the other. However, the walls of the well stop us from pulling the electron and hole very far apart, stopping the field ionization. Hence the exciton is polarized by the field, but *not* field ionized. The electron and hole can still orbit round about one another, albeit in somewhat displaced orbits. Hence the exciton still exists with essentially as long a lifetime as it had before, and the absorption line does not substantially broaden with field (at least, not from field ionization). We could view this as a polarization, **P**, of the exciton, in the presence of a field **E**, which leads to an energy shift -½ **P.E**, a Stark shift. (This is somewhat counteracted by some increase in kinetic energy of the electrons and holes as they are squeezed towards the walls of the wells.)

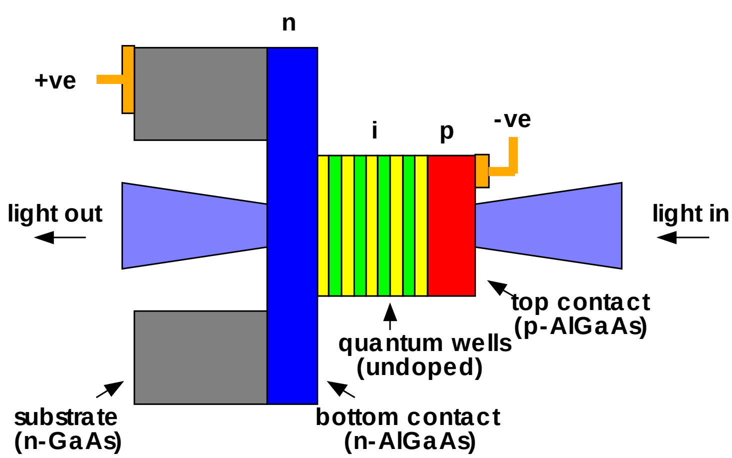

Fig. 10. Structure of a simple quantum well diode. The GaAs/AlGaAs quantum wells are contained in the undoped (*i*) region of a *p*-*i*-*n* diode. The *p* and *n* regions are made of AlGaAs, which is transparent at the operating wavelength. Light is shone through the structure, and the diode is reverse biased to apply the electric field perpendicular to the quantum well layers [9].

Of course, if the well was substantially larger than the exciton diameter, the exciton could be effectively field ionized by pulling the electron and hole to the extreme sides of the well. Hence it is important that the exciton is substantially confined by the walls of the well. As a result, this effect is called the quantum-confined Stark effect (QCSE)[9]. The energy shifts possible with the QCSE are very large. It is possible to shift the exciton peaks by many times the binding energy. This is not unphysical because it is a reduction in the energy of the particle, not an increase. Fig. 11 shows measured and calculated positions of the heavy and light hole exciton peaks in a GaAs/AlGaAs quantum well [9].

As the electron and hole are pulled to the sides of the well, they are on the average somewhat farther apart. This reduces the excitonic absorption strength in two ways. First, obviously the electron and hole are further apart in the *z* direction (i.e., perpendicular to the layers). Second, because the electron and the hole are on the average farther apart, there is a consequent reduction of the Coulomb attraction between them, resulting in a slight increase in the size of the orbit in the quantum well plane. Both of these effects together contrive to give a reduction in the observed height of the exciton peak as the field is increased. Despite this reduction, in practice the exciton peak still results in strong absorption, even for substantial shifts.

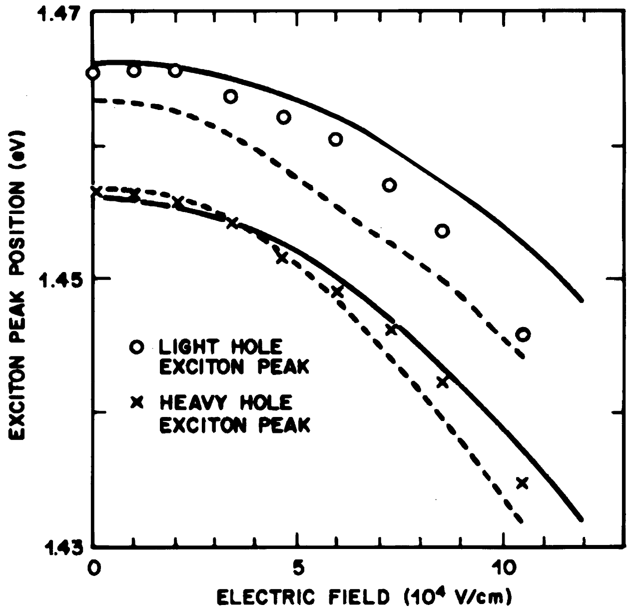

Fig. 11. Measured (points) and calculated (lines) energies of the heavy and light hole exciton peaks in a GaAs/AlGaAs multiple quantum well structure as a function of electric field applied perpendicular to the quantum well layers [9]. The solid and dashed lines are for different values of the band discontinuity offset ratios.

From the point of view of devices, one of the most important aspects of the QCSE is that the absorption changes that can be made are large enough to allow absorption modulators that can work for light propagating perpendicular to the surface of the device ("surface-normal" modulators). The mechanisms in bulk semiconductors are not really large enough to do this effectively. Having surface normal modulators allows two dimensional arrays of devices, which in turn allows possibilities such as imaging whole arrays of thousands of optical interconnects on and off of the surface of a silicon chip with quantum well diodes bonded to it [13][14]. Quantum well modulators can also be made in waveguides, where they are useful for high speed modulation, for example, for telecommunications. It is also possible to use many different kinds of quantum well structures, such as graded or stepped wells, or coupled wells, though space does not allow a discussion of these here.

The formal theory to calculate QCSE electroabsorption is little different from the theory mentioned above for excitonic absorption in quantum wells. We do not have the problem that we have to deal with excitonic ionization. The steps in the calculation become as follows.

(i) We take a wavefunction that is a product of electron and hole *z* wavefunctions, and a relative motion wavefunction in the *x*-*y* plane (and a plane wave for the center of mass motion that will be of no further interest to us).

(ii) Calculate the individual electron and hole wavefunctions by solving the quantum well problems, using the Schroedinger equations with field, Eqs. (12) and (13).

(iii) Use a variational wavefunction for the x-y relative motion wavefunction, corresponding to a 1S orbital with an adjustable radius.

(iv) Finally, we have our estimate of the exciton energy

$$E_{ex} = E_g + E_{nc} + E_{nv} + E_B \tag{15}$$

where the "quantum well" energies, E_{nc} and E_{nv}, are those calculated for the "sloping" quantum wells in the presence of the field, and E_B is the result of our variational exciton calculation in the presence of the "skewed" quantum well z wavefunctions for electrons and holes.

The primary changes in energy are from the "quantum well" energies, E_{nc} and E_{nv}. E_B does change with field; the exciton becomes less tightly Coulomb bound, so E_B actually decreases in magnitude a little (usually only a few meV), and the exciton does become somewhat larger in diameter [9].

For an extended review of the physics of electroabsorption and other optical effects in quantum wells, see Ref. [4].

3. Dense optical interconnects

So far, we have discussed some of the basic physical reasons why optics is or is not interesting for use in information than if processing, and also we have summarized some of the physics of electroabsorption in quantum wells. These two areas come together when we discuss the possibilities of optics for use in interconnections inside computers.

3.1 Optics for interconnects in silicon systems

Many of the physical features that we discussed above show that, though the issue of optical logic may remain for future discussions, the use of optics for interconnects as many features in its favor now. In addition to these general physical arguments, it is arguably becoming the case that electronic systems are in practice becoming increasingly limited by the difficulty of performing interconnects. Indeed, this is a looming problem at all levels of computing systems. It is encountered first at longer distances, but scaling analysis suggests that it may even become a problem on silicon chips themselves. It is important to note that, in discussing interconnects, optics is not competing against silicon, but against copper and aluminum. We are not trying to replace the highly successful silicon logic technology, which enables us to make systems of a complexity well beyond that of any other technology we currently control. We are, however, considering replacing wiring with optical connections. If we can successfully integrate the optical connections with silicon, then we are augmenting the capabilities of silicon rather than competing with them.

Silicon integrated circuit technology is improving at a rate of approximately a factor of two every 18 months, for example, if we look at the number of transistors on a chip. The United States Semiconductor Industry Association has predicted that by the year 2006 transistors will be made in 0.1 micron line width technology, the clock rate on chip will be approximately 3.5 GHz, and chips may have three thousand pins. Beyond that point, their predictions are that interconnects will be a severe limit on the performance of chips, even considering the interconnections on the chip. It is arguably the case that, even today, electrical interconnects restrict the performance of systems when data has to flow on and off chips or on and off boards. The competitive nature of the integrated circuits business suggests that this point of substantial difficulty in interconnections on chips may be achieved even sooner than the published industry projections suggest.

There are several solutions to the problems of interconnects. We might firstly attempt to improve design tools to emphasize interconnect more than logic layout. Secondly, we might try to design architectures that need less interconnect (though such an approach can likely only be applied efficiently to certain classes of problems). Thirdly, we might attempt to improve the signaling on electrical interconnects lines, for example, by equalization.[15][16] Finally, we might attempt to find another physical solution to the problem of interconnect. Arguably the only viable physical solution, that is, one that does not require a fundamental breakthrough or the demonstration of a radically new technology beyond anything we can currently construct in the laboratory, is optical interconnects.

3.2 "Aspect ratio" limit on electrical interconnects

One practical example of the limitations of electrical interconnects as we try to scale to larger capacities, larger numbers of lines and higher clock frequencies is the so-called "aspect ratio" limit of electrical interconnects [15], a limit that the derives entirely from resistive loss on electrical lines. This limit says that the capacity of unequalized electrical lines is approximately $10^{15} A/l^2$ bits/s for lines off chip, where A is the total cross-sectional area of the electrical lines, and l is the length of a line. For lines on chip, the capacity may arise somewhat to approximately $10^{16} A/l^2$, and equalization may increase this to about $10^{17} A/l^2$. For example, if we worked with lines that are about 30 times as long as they are wide, then the ratio A/l^2 becomes of the order of 10^{-3}. It is difficult to imagine making any moderately complicated interconnect with wires that are much "stubbier" than this. Hence, we would start to run into basic problems with electrical interconnects for aggregate bit rates of the order of 1 Tb/s.

It is important to realize that such problems with interconnects cannot be solved either by making the system smaller (miniaturization) or larger. The ratio A/l^2 is scale-invariant. Thus, when a system becomes sufficiently complex or sufficiently fast to run into this limit, it is necessary to seek other physical solutions to interconnect. Equalization will help, though at the expense of increased circuit complexity and system set-up, but it also does not remove the underlying scaling problem. Put simply, we cannot take an electrical machine and make it faster and faster while retaining the same architecture if we stick with electrical interconnects. Optical interconnects completely avoid this problem because the same physical loss mechanism is simply not present. Optical systems can exceed these limits by as much as 12 orders of magnitude.

3.3 Design benefits of optical interconnects

From a practical point of view, optical interconnects can avoid various design problems in interconnect, all because of the fundamental physical reasons stated above. For example, we can avoid having to put electrostatic discharge protection on inputs, we can reduce the area of off chip drivers and pads (because of quantum impedance conversion), we can avoid pin inductance on chips (and the consequent necessity of dedicating many pins to power and ground connections), we can increase the number of interconnections on and off chip, we can avoid impedance matching, wave reflections, and line termination problems of electrical buses, and can also avoid such problems as the delay variation on electrical lines that results from the temperature dependence of resistivity of metals.

3.4 Technologies for optical interconnects

For optical interconnects to be practically viable, we require both a viable approach for the optics that must direct the signals into and out of the desired places in the system, and also the optoelectronic devices that will interface between the electronic signals and the optical ones. The optics does not appear to represent any fundamental problem, though it would need significant practical development effort to make optical systems of sufficient complexity and sufficiently low cost. There are certainly significant issues if we choose to try to integrate waveguides, for example, on to silicon circuits, though it is by no means necessary to operate with guided wave systems; "free-space" optics, using imaging lenses and similar "bulk" optics can be very effective for transporting up to thousands of light beams inside rigid modules over short (for example, centimeter) distances.

The optoelectronic devices have, however, been a significant basic issue, at least until recently. There are two problems for the optoelectronic devices. First, we require devices that can operate at sufficiently low power consumption and can be made in sufficiently large numbers and low enough cost for use in interconnects. Historically, the problem is greatest for the output devices. It has been difficult to make either lasers or modulators with sufficiently good performance. A second problem with optoelectronics is that, for use in dense optical interconnects, it must be possible to integrate the optoelectronics efficiently and in large numbers with silicon electronics. As is well known, silicon itself, though capable of making quite good photodetectors, cannot at present make either efficient light emitters or light modulators that are good enough for use in high-speed interconnects. It is also relatively difficult to make efficient high-speed silicon photodetectors in the CMOS (complementary metal-oxide-semiconductor) silicon technology.

At present, the only viable solutions for optical output devices appear to be either vertical cavity surface emitting lasers (VCSELs) or quantum well modulators. Both of these devices can be made in two dimensional arrays. Quantum well modulators have demonstrated rather large arrays (for example, more than four thousand devices integrated to a silicon chip) [14]. Large arrays of modulators have been investigated for some time, having been considered as potential optoelectronic logic devices (self-electrooptic-effect devices or SEEDS) [17][18]." The only significant potential disadvantage of the modulators is that they require external beams to be generated. This can be turned to advantage because one laser source can be used together with relatively simple diffractive optics to generate large arrays of beams, and then it is only necessary to control the wavelength and modal properties of one, centralized laser. Additionally, this centralized laser can be clocked, thereby allowing the outputs of all of the modulators to be clocked, or possibly simultaneously allowing clock distribution using the same set of laser beams. Modulators also have the feature that they can be used with short pulse lasers, thereby allowing the signal retiming discussed above, as well as delivering the sharpest possible signal to the receiver circuits, a feature that can improve the performance of the receiver system overall. Disadvantages of the lasers include the fact that they are not yet as advanced as modulators in terms of demonstrations in large systems, and in systems integrated to silicon, the necessary low-threshold lasers are still at an experimental stage, All lasers suffer from a problem of turn-on delay that makes them increasingly difficult to use at very high clock rates. Modulators, by contrast, have no internal speed limitations other than the usual resistive-capacitive times, and they do not show any analogous "turn-on" delay phenomena. These various features have been discussed at greater length in Ref. [19], which also gives a critical comparison of modulator and laser features.

Light-emitting diodes may be usable for dense optical interconnects, but they have several difficulties. Perhaps the main issue is that it is not possible to couple efficiently all of the light from a light emitting diode into the receiving detector. This is because the light emitting diode emits over a very large angular range, and practical optics cannot collect all of the light. This low efficiency means that the power dissipation required for use of light emitting diodes may be higher; since dense systems may well be limited by power dissipation [20], this could be a significant disadvantage. A second issue is that efficient light emitting diodes are not normally very fast because the are usually limited by the spontaneous emission time.

In the addition to having viable optoelectronic devices in terms of speed, power, and numbers, it is also necessary to be able to integrate these with mainstream silicon integrated circuits. The ideal goal might be to be able to integrate such optoelectronic devices monolithically with silicon integrated circuits. It is still not possible to integrate lasers monolithically with silicon and obtain devices with good enough lifetimes for practical use. The key problem is that the III-V materials are not lattice matched to the silicon, and consequently a large number of crystal defects are formed at the interface. Light-emitting devices tend to be very sensitive to these defects, perhaps for a least two reasons; first, defects can act as recombination centers, inhibiting efficient radiative recombination, and second, the defects tend to propagate and grow under the high forward bias current densities in light emitters. Modulators have been demonstrated monolithically integrated to silicon [21], with apparently good lifetimes and performance. Even when the devices can be successfully integrated on silicon substrates, that does not mean that in practice they can be integrated with silicon integrated circuits. It is particularly important, especially for the introduction of optical interconnects, that such optoelectronic devices can work with silicon integrated circuits without requiring modification of the silicon integrated circuit fabrication processes. Monolithic growth of III-V devices on silicon is not very compatible with existing silicon processes. For example, Ga makes silicon dioxide conducting, and hence is a highly undesirable material in any silicon process. Also, III-V devices tend to use gold metalization, whereas silicon tends to use aluminum metalization. The interface between gold and aluminum is known to be a troublesome one that can lead to the opening of voids and the breakdown of conduction across the interface between these metals. There likely are technical solutions to these metalization and processing issues, but these would have to be resolved before monolithic integration would be practical.

Because of these process compatibility issues, a more promising approach for the introduction of dense optoelectronics with silicon is to use some hybrid technique in which the silicon circuit can be fabricated as usual with the optoelectronic devices to be attached after the final silicon circuit fabrication. Various such hybrid techniques have been reviewed in Ref. [19]. One of the most promising and flexible techniques for initial introduction of dense optoelectronics with silicon is solder bonding. In this technique, solder is deposited on one or both of the chips to be bonded, and the two chips are brought together under careful alignment with controlled temperature and pressure. The technique used most extensively with quantum well modulators is illustrated in Fig. 12 [22]. In this technique an array of modulator/photodetector devices is made on a gallium arsenide substrate. This entire array is then bonded onto matching contact pads on a silicon integrated circuit. An epoxy is flowed into the remaining space between the two chips, and the entire gallium arsenide substrate is removed by a selective chemical etch. An anti-reflection coating is applied to the top of the modulators. The net result is to form an array of isolated modulator/detector devices each connected to the silicon circuit. This process is the one that has been successfully used to attach thousands of devices to silicon circuits.[13][14]

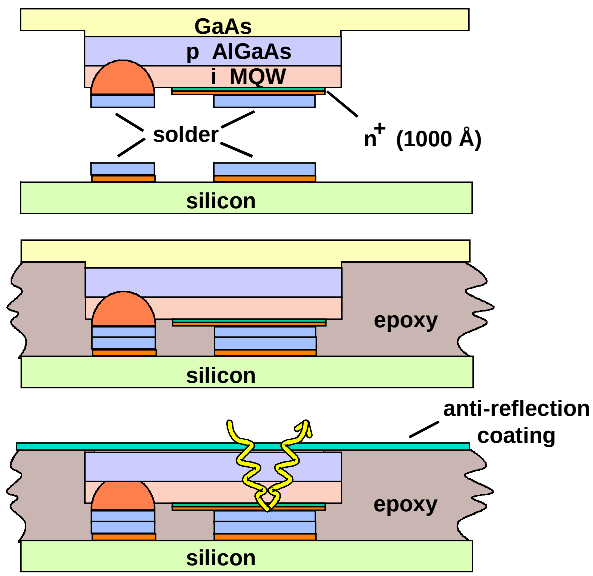

Fig. 12. Illustration of the solder bonding process used for quantum well diodes and silicon integrated circuits.

The idea of using optical interconnects to substitute for some of the wiring in computers, especially over shorter distances of centimeters to meters, is a radical one. It is important to understand whether such an idea can indeed solve the kinds of scaling problems that are anticipated to a current as silicon circuits become increasingly capable. This problem has been analyzed by Krishnamoorthy and Miller [20] for example. In this analysis, it turns out that the power dissipation of the optical receiver and transmitter circuits is likely one of the major limitations to the density of optical interconnects, and they conclude that, in a system based on the use of modulators [20], the receiver power dissipation will dominate in a well-designed system. The reason why the receiver power dissipation is relatively high is because the input stage of these receivers must be biased at an operating point that allows small signal amplification; such biasing tends to require a continuous current to be passing through the input stage. The receiver design is therefore very important for these systems. Such receivers must be kept as simple as possible, and it is very important to minimize the capacitance of the input photodetectors and the associated connection to the receiver, which argues for very good integration techniques. One very important conclusion out of this analysis, however, is that, unlike electrical interconnections, it appears that the ability of the optics to interconnect on and off the chip will scale with the ability of the chip to perform logic operations. This conclusion is in contrast with the behavior of electrical interconnects, which do not scale to keep up with the ability of the chip to perform logic operations. Hence, there is good reason to believe that optics would be a worthwhile technical solution in which to invest because of its future ability to scale.

4. Conclusions

At the time of writing this article, there are many physical reasons why optics should be useful in digital information processing, particularly for use in interconnecting information within digital processors. Optical logic gates remain a possibility, but for the foreseeable future, they are likely only to be used in small numbers in very particular situations where there potential high speed could be important. The use of optics for interconnects does, however, appear to be an important and viable possibility, because there is a real problem with scaling electrical wiring to keep up with the advance of transistor technology, and because optics appears to have both the necessary physical features and the necessary devices to interface in and out of the electronic processing. Quantum well modulators currently appear to be arguably the most viable device for dense optical interconnects, and represent an exciting opportunity in this application.

5. Further Reading

The discussion of the various topics in these notes are necessarily brief. More extensive information can be found in the following references.

For a discussion of the physical reasons for the use of optics in digital information processing, including optical interconnections, see Ref. [1].

The physics of interband optical properties of quantum wells, including electroabsorption effects, is discussed at length in Ref. [4].

Quantum well electroabsorptive device arrays are discussed in Refs. [17] and [18]. Integration of optoelectronic devices with silicon circuitry is described in Refs. [14] and [19.

6. References

[1] D. A. B. Miller, “Physical Reasons for Optical Interconnection,” Special Issue on Smart Pixels, Int’l J. Optoelectronics **11** (3), 155-168 (1997)

[2] D. A. B. Miller, “Optical Switching Devices: Some Basic Concepts,” in *Optical Computing*, Prod. Thirty-Fourth Scottish Universities Summer School in Physics, B. S. Wherrett, and F. A. P. Tooley, eds., (Adam Hilger, Bristol, 1989), pp 56-70

[3] D. A. B. Miller, "Optics for low-energy communication inside digital processors: quantum detectors, sources, and modulators as efficient impedance converters,” Optics Letters, **14**, 146-148, (1989).

[4] S. Schmitt-Rink, D. S. Chemla, and D. A. B. Miller, "Linear and Nonlinear Optical Properties of Semiconductor Quantum Wells," Adv. Phys. 38, 89-188, (1989).

[5] W. Franz, Z. Naturforsch. **13a**, 484 (1958).

[6] L. V. Keldysh, Zh. Eksp, Teor. Fiz. **34**, 1138 (1958) [Sov. Phys. JETP **7**, 788 (1958).

[7] K. Tharmalingham, Phys. Rev. **130**, 2204-2206 (1963).

[8] D. A. B. Miller, D. S. Chemla and S. Schmitt-Rink, "Relation Between Electroabsorption in Bulk Semiconductors and in Quantum Wells: The Quantum-Confined Franz-Keldysh Effect," Phys. Rev. **B33**, 6976-6982 (1986).

[9] D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood and C. A. Burrus, "Electric Field Dependence of Optical Absorption near the Bandgap of Quantum Well Structures," Phys. Rev. **B32**, 1043-1060 (1985).

[10] J. Dow and D. Redfield, Phys. Rev. B **1**, 3358 (1970); Phys. Rev. B **5**, 594 (1972).

[11] I. A. Merkulov and V. I. Perel, Phys. Lett. **45A**, 83-84 (1973).

[12] F. L. Lederman and J. D. Dow, Phys. Rev. **B13**, 1633 (1976).

[13] A. L. Lentine, K. W. Goossen, J. A. Walker, L. M. F. Chirovsky, L. A. D'Asaro, S. P. Hui, B. T. Tseng, R. E. Leibenguth, D. P. Kossives, D. W. Dahringer, D. D. Bacon, T. K. Woodward, D. A. B. Miller, "Arrays of Optoelectronic Switching Nodes Comprised of Flip-Chip-Bonded MQW Modulators and Detectors on Silicon CMOS Circuitry" IEEE Photonics Technology Letters, **8**, 221-223 (1996).

[14] A. V. Krishnamoorthy and K. W. Goossen, "Progress in optoelectronic-VLSI smart pixel technology based on GaAs/AlGaAs MQW modulators," Int. J. Optoelectronics **11**, 181-198 (1997)

[15] D. A. B. Miller and H. M. Ozaktas, "Limit to the Bit-Rate Capacity of Electrical Interconnects from the Aspect Ratio of the System Architecture," J. Parallel and Distributed Computing **41**, 42-52 (1997)

[16] W. J. Dally and J. Poulton, IEEE Micro, Jan/Feb. 1997, pp 48 - 56

[17] D. A. B. Miller "Quantum-well self-electro-optic effect devices," Optical and Quantum Electronics, **22**, S61-S98, (1990).

[18] A. L. Lentine, D. A. B. Miller "Evolution of the SEED technology: bistable logic gates to optoelectronic smart pixels" IEEE J. of Quantum Electronics, **29**, 655-669 (1993)

[19] D. A. B. Miller, "Dense Two-Dimensional Integration of Optoelectronics and Electronics for Interconnections," in *Heterogeneous Integration: Systems on a Chip*, Anis Husain and Mahmoud Fallahi, Eds., SPIE Critical Reviews of Optical Engineering, Vol. CR70 (SPIE, Bellingham, 1998), 80-109.

[20] A. V. Krishnamoorthy and D. A. B. Miller, "Scaling Optoelectronic-VLSI Circuits into the 21st Century: A Technology Roadmap," IEEE J. Selected Topics in Quantum Electronics **2,** 55-76 (1996).

[21] K. W. Goossen, G. D. Boyd, J. E. Cunningham, W. Y. Jan, D. A. B. Miller, D. S. Chemla, R. M. Lum, "GaAs-AlGaAs Multiquantum Well Reflection Modulators Grown on GaAs and Silicon Substrates" IEEE Photonics Tech. Lett., **1**, 304-306, (1989).

[22] K. W. Goossen, J. A. Walker, L. A. D'Asaro, B. Tseng, R. Leibenguth, D. Kossives, D. D. Bacon, D. Dahringer, L. M. F. Chirovsky, A. L. Lentine, D. A. B. Miller "GaAs MQW Modulators Integrated with Silicon CMOS" IEEE Photonics Technology Letters, **7**, 360-362, (1995)