1. (a)   The need for Quantum mechanics arose around the turn of the 20th century, spurred on by a number of experimental irregularities. For instance, on the basis of the understanding of nature at the time, the very existence of atoms was puzzling. It was known that atoms consisted of negatively charged electrons "orbiting" positively charged nuclei. From conventional electromagnetism, however, we know that any charged particle will emit radiation if it undergoes acceleration, and will therefore lose energy. Therefore, if an electron orbits a nucleus, by the very fact that it is moving around that nucleus, it must be undergoing acceleration. One would expect then that it would radiate, lose energy and spiral into the nucleus. This clearly doesn't happen, so a theory was needed to explain why not.

Another problem was the phenomenon of photoemission. This is the effect whereby electromagnetic radiation incident on a metal surface causes electrons to be ejected from it. It was observed that electrons are only ejected once the frequency of the light reaches some threshold value, their kinetic energy scales linearly with increasing frequency, and the number of ejected electrons increases as the light intensity increases. This is at odds with classical electromagnetism which predicts that (i) the frequency of light should not make any difference whatsoever, (ii) the energy of the ejected electrons should only depend on the intensity of light (The intensity, I is the energy per unit area per unit time, or $\frac{1}{2}\varepsilon_o E^2$). There is no reason classically for the number of ejected electrons to depend on the intensity. The explanation of this effect within the framework of quantum mechanics is that light consists of *parcels* of waves, called *photons*, which contain distinct amounts of energy, depending linearly on frequency. The proportionality constant is Planck's constant, h. Electrons are bound to the metal by an energy called the work function, f, and until the frequency of light is high enough, individual photons don't have enough enough energy to free them. Above this threshold frequency then, electrons are ejected and their excess energy is manifested as kinetic energy. The number of electrons ejected then scales linearly with the number of incident photons, which depends on the intensity.

Another turning point was the observation of electron diffraction – clear evidence that particles have a wave-like character. Further discussion could

focus on this, how electrons (and all other matter) have energy, and through the equivalence that energy = $\hbar^2 k^2/2m$, we can find the deBroglie wavelength associated with them, and find that it can be significantly less that that of visible light. Coupled with the fact that this wavelength is less than the atomic spacing in typical materials, and that electron beams can be focused, we find that it is possible to image (conducting) materials and observe diffraction patterns when an electron beam encounters a thin enough material film.

Other points include why some materials are conductors and others insulators, the origin of electrical resistance, the ultra-violet catastrophe, and the fact that atoms only absorb and emit light of specific colours.

(b)   Wavepackets are used to represent either particles (eg electrons) or short pulses of light.

To construct an electromagnetic wavepacket, add two sinewaves:

$E = E_o\{\cos(\omega t - kx) + \cos(\omega_2 t - k_2 x)\}$
$= 2E_o \sin(\omega_+ t - k_+ x)\cos(\omega_- t - k_- x)$

where $\omega_+ = (\omega + \omega_2)/2$, $\omega_- = (\omega - \omega_2)/2$ and similarly for k_+ and k_-..

This represents a sinewave of frequency ω_+ which has its amplitude modulated at the frequency ω_- to produce beats. This therefore exhibits localized areas of enhanced field which may be used to represent particles (assuming we have the appropriate relationship between ω and k). These localized areas travel at the group velocity $d\omega/dk$, which must be less than or equal to c. In the absence of dispersion (i.e. for free space) these wave-packets travel without spreading out. The phase velocity is $\omega/k = \omega_2/k_2 = c$. We could refine the wavepacket by summing together more sinewaves, with e.g. a Gaussian spectral function. This will increase the localization.

(c)   From lecture notes, the steps involved to derive an expression for a general electromagnetic wavepacket are as follows:

General wave-packet, $A = A_0 \int\limits_{-\infty}^{\infty} f(k)e^{i(kx-\omega t)}dk$

where $f(k)$ is the *spectral function* specifying the amplitude of the harmonic with wave-vector k, and frequency $\omega = ck$. As an example of this, consider a *Gaussian* spectral function, i.e.

$$f(k) = \frac{1}{\sqrt{2\pi\delta^2}} e^{-\frac{k^2}{2\delta^2}}$$

Using the fact that for electromagnetic waves, $\omega = ck$ ($\omega = 2\pi f$, $k = 2\pi/\lambda$), our integral becomes:

$$A = \frac{A_0}{\sqrt{2\pi\delta^2}} \int\limits_{-\infty}^{\infty} e^{-\frac{k^2}{2\delta^2}+ik(x-ct)} dk$$

This is known as a Gaussian Integral, and the easiest way to solve is to change the variables : let $x - ct \to x'$, and $k - (ib/2a) \to k'$, where $a = 1/2\delta^2$ and $b = x - ct$. This gives us

$$A = \frac{A_0}{\sqrt{2\pi\delta^2}} e^{\frac{-b^2}{4a}} \int\limits_{-\infty}^{\infty} e^{-ak^2} dk \quad\longrightarrow\quad A = A_0\delta e^{-\frac{\delta^2}{2}(x-ct)^2}$$

Wave-packet propagating in the +x direction

↑
This integral is just equal to $\sqrt{(\pi/a)}$

Plotting this, we see:

Time-development of Gaussian wave-packet
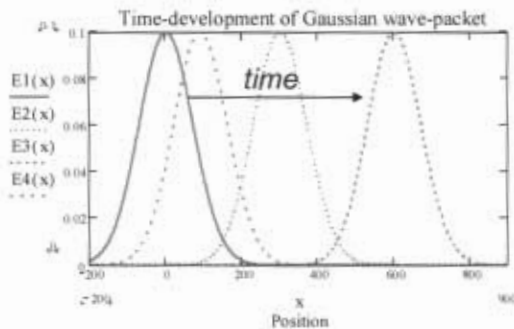
By using this weighted spectral function, we have produced a propagating wave-packet centered at $x = ct$, which moves with the velocity $c = \omega/k$. The spatial extent of this wave packet is constant at $\Delta x = 1/\delta$, and does not change with time. Remember, from the form of Gaussian we used, the width of $f(k) = \Delta k = \delta$. Therefore, we have the relationship:

$$\Delta x \Delta k = 1$$

This is equivalent to saying that the sharper the localization of the wave-packet in x-space, the wider is its spectrum in k-space. Our postulate from earlier said that due to the wave-particle duality particles have a wavelength, λ associated with them, which is related to their momentum by $\lambda = h/p$, and given that $k = 2\pi/\lambda$, we end up with the result:

$$\Delta x \Delta p = h$$

This is tantamount to saying that the more accurately we know the position of a particle, the less accurately we can know it's momentum, and vice versa. This is essentially *Heisenberg's uncertainty principle*, and is due to the very nature of waves. As an example of this, remember our initial wave, $E = E_0 e^{i(\omega t - kx)}$.

This describes a wave packet of a single wavelength, i.e. $\Delta k = 0$, so the spatial extent is infinite. The momentum is precisely $\hbar k$, and the position is completely indeterminate. (not useful for describing a particle)

For a wavepacket describing a matter particle, the key difference is that the energy is $E = \hbar^2 k^2/2m$, whereas for an electromagnetic wave, $E = \hbar kc$. The

consequence of this is that components with different k that are used to create the wavepacket all travel at different speeds, leading to spreading of the wavepacket, otherwise known as dispersion.

(d)   Heisneberg's uncertainty principle states that $\Delta x \Delta p \geq \hbar/2$. This means that we can never simultaneously know both the position and momentum of a particle, and the more accurately we know one, the less we can know about the other.

2.

(a)   The material is $Ga_{0.4}Al_{0.6}As$ and has a band gap of $1.45 + 0.6 \times 1.247 = 2.198$ eV. This is split equally between the conduction and valence bands, and as the band gap of GaAs is 1.45 eV, this means the difference in bandgaps, which is $(2.198 - 1.45)$ eV leads to quantum wells of depth $(2.198 - 1.45)/2 = 0.374$ eV. Therefore, in the absence of any band-bending, the heterostructure has a band profile as shown:
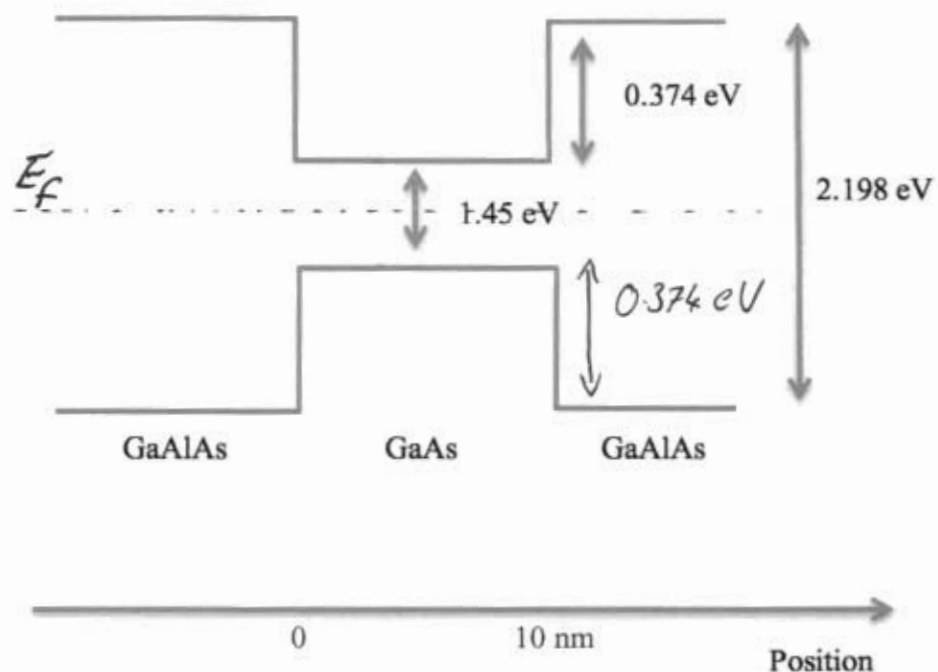
(b)   Assumptions:

(i)   The energy levels of interest are low enough that we can use the infinite well approximation. This is valid as long as they are less than 1/10 of the well depth.

(ii)   The electrons are non-interacting.
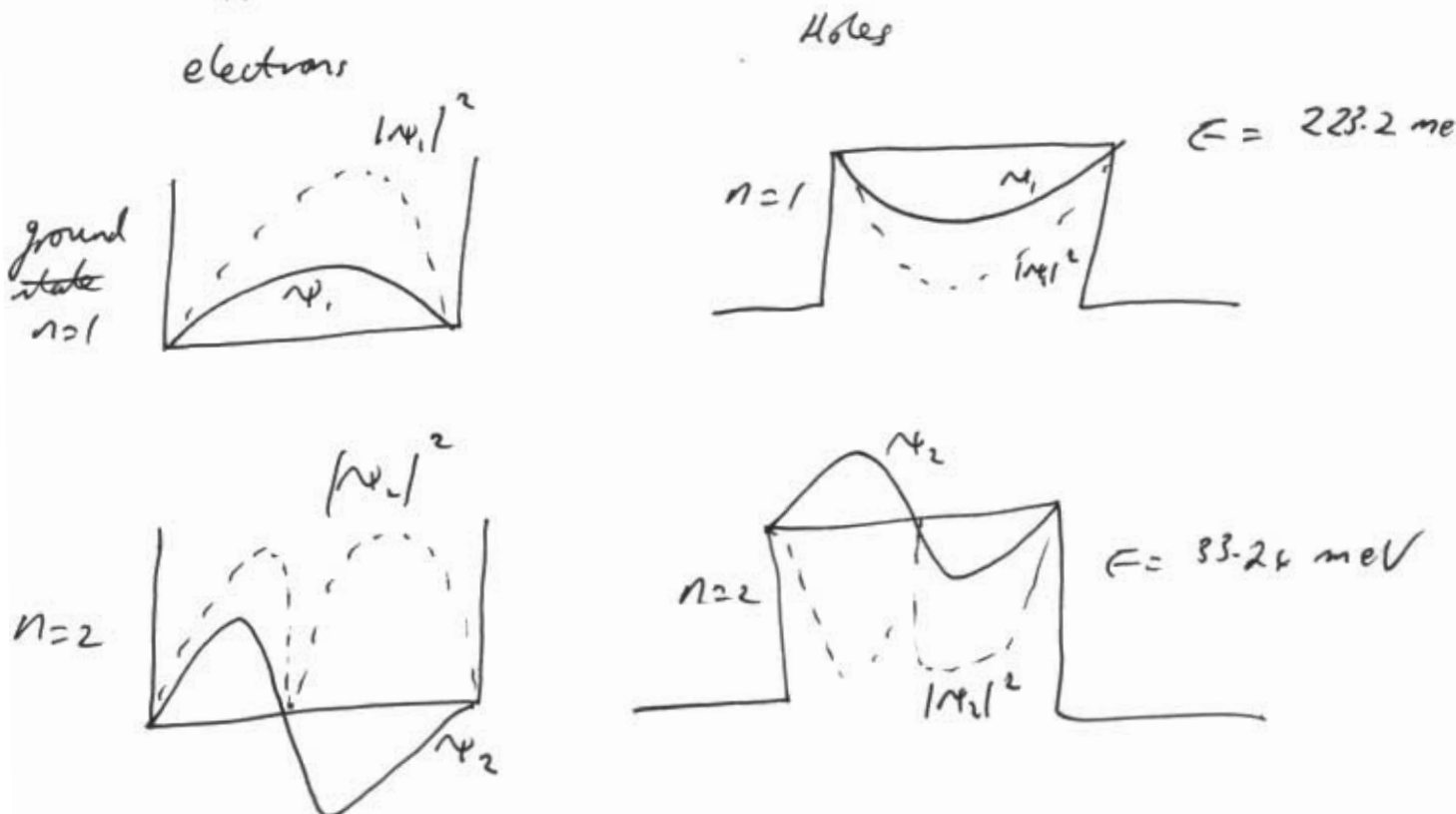
(iii)   There is no band-bending

The solution should include a derivation of the spectrum of energy levels associated with an electron of effective mass $m*$ in an infinitely deep well of width L:
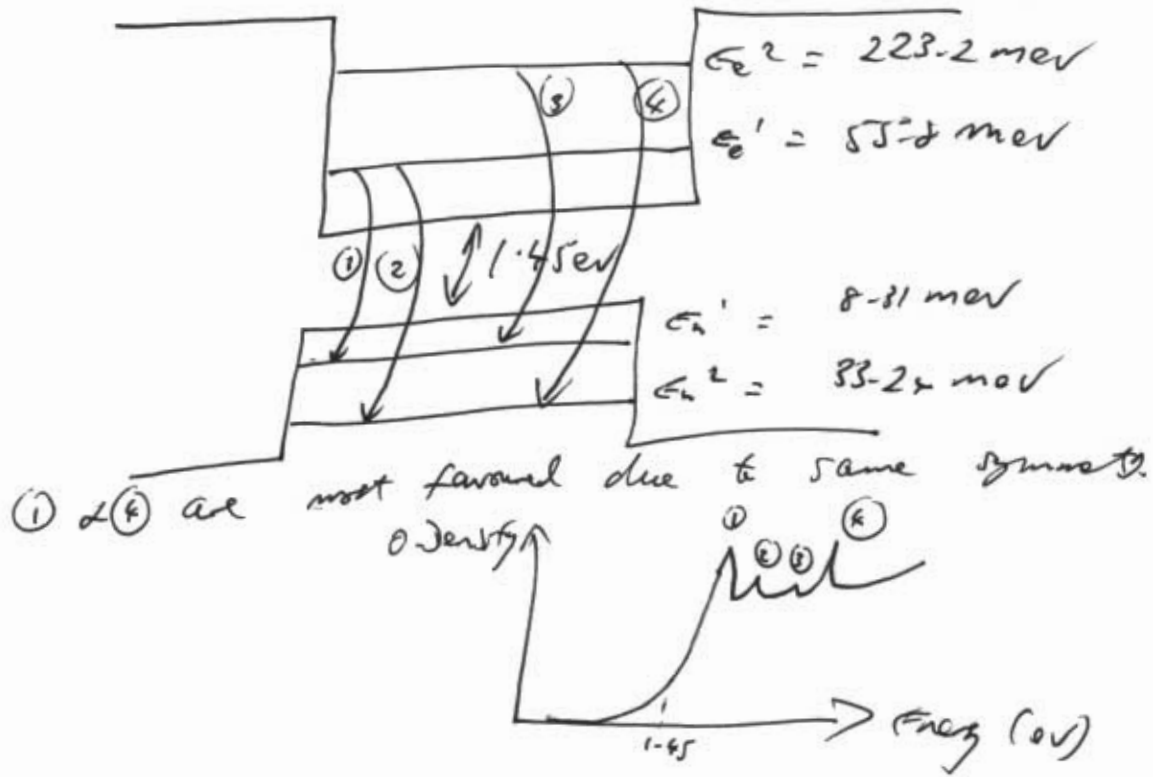
$$E_n = \frac{h^2 n^2}{8m^* L^2}$$

From which we find that the electron energies are $E_n^e = 55.8\ n^2$ meV
and the hole energies are $E_n^h = 8.31\ n^2$ meV
The electron ground state energy ($n = 1$) is around 1/6 of the well depth, so our estimate will not be particularly accurate, whereas the hole ground state energy is around 1/45 of the well depth, and is therefore a good estimate.

(c)

electrons

Holes

$|\psi_1|^2$

ground state
$n=1$

ψ_1

$n=1$

ψ_1

$|\psi_1|^2$

$E = 223.2$ me

$|\psi_2|^2$

$n=2$

ψ_2

ψ_2

$n=2$

$|\psi_2|^2$

$E = 33.24$ meV

$\varepsilon_e^2 = 223.2 \text{ meV}$

$\varepsilon_e^1 = 55.8 \text{ meV}$

$\varepsilon_h^1 = 8.81 \text{ meV}$

$\varepsilon_h^2 = 33.24 \text{ meV}$

① & ⑥ are most favoured due to same symmetry.

3. (a) Any feature can be observed in an optical microscope, but the minimum size the image can have is $\frac{\lambda}{2NA}$ - this is known as the *diffraction limit*. The minimum spacing between objects that can be *resolved* is 0.8λ, which happens when the central dip in intensity is ≥ 20% of the maximum – this is the *Rayleigh criterion*.

i.e.
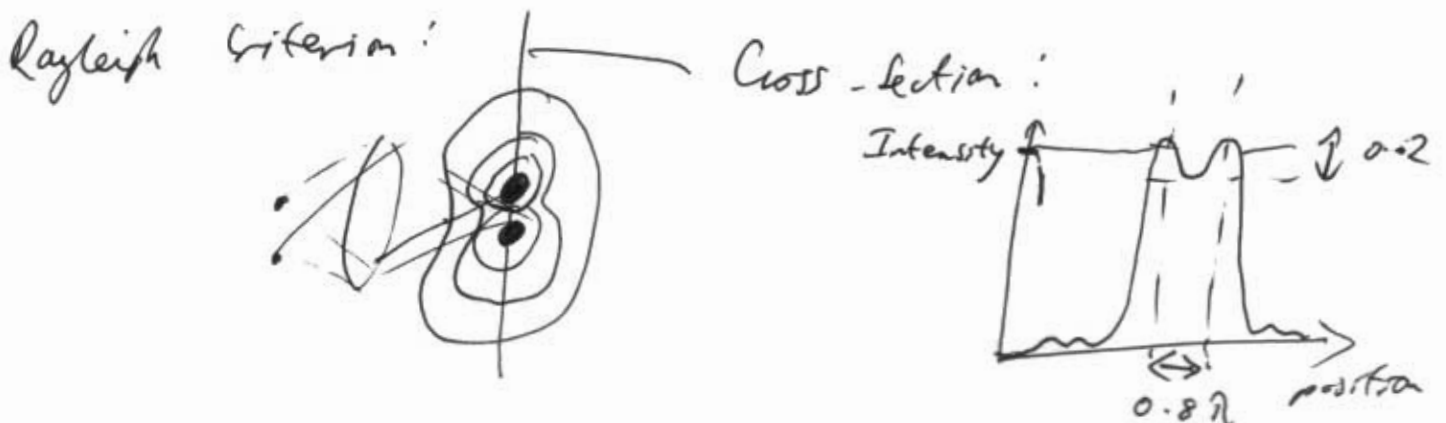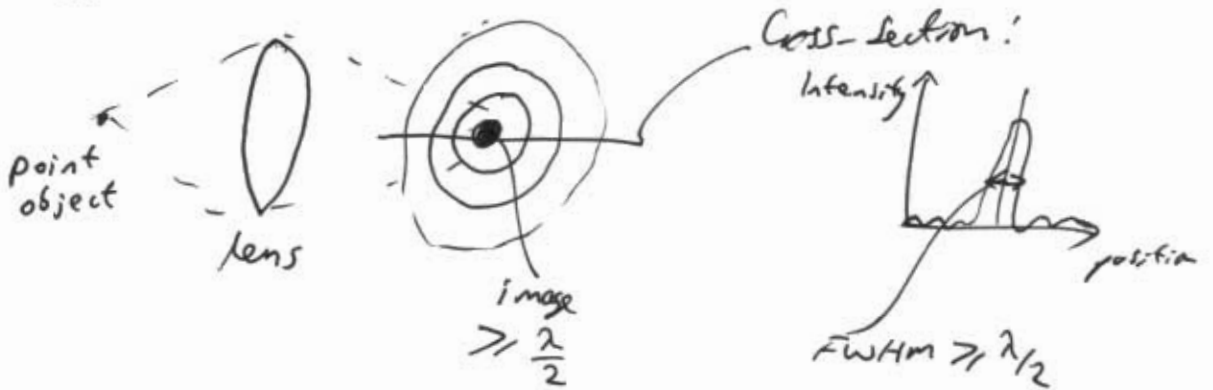
The Rayleigh limit can be overcome using evanescent waves. The reason for this is that the resolution depends on λ, via $k = 2\pi/\lambda$. If k can be increased somehow, the resolution will increase. As an example, if we want higher resolution in the x-direction, we need to increase k_x. As $k_x = \sqrt{k^2 - k_y^2 - k_z^2}$, if either k_y or k_z are complex (i.e. evanescent) then $k_x > k$, hence the resolution in x will be better than the diffraction limit. The type of microscope that utilizes this is the scanning Near Field Optical Microscope (SNOM).

(b) SEM:
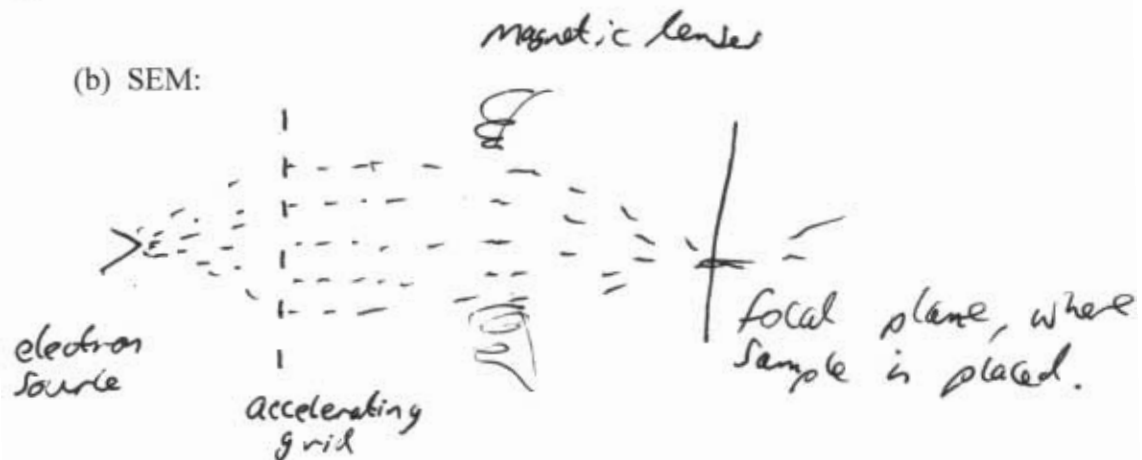
Basic principle: electrons have a wavelength, λ, determined by their kinetic energy – controlled using the voltage applied to the accelerating grid. The electron beam is focused using magnetic (or less commonly, electrostatic) lenses, and the minimum spot size $\sim \lambda/2$, as in optics. The difference however, is that the electron wavelength $\sim 10^{-12}$ m, so the resolution is far better than the optical microscope.

SEM Pros
- High resolution
- Large depth of field
- Can image large areas
- Several image contrast mechanisms

SEM cons
- Requires vacuum

- Requires a conductive coating, so cannot look at live samples or samples in vitro
- Expensive, both in the capital cost of the machine, and running costs

(c)   Basic operation of SPM: A sharp probe tip scans either in close proximity or in contact with the surface of interest. An interaction (different depending on the type of SPM) between this probe and the sample is monitored and is maintained at a constant value using a feedback loop that controls the tip-sample distance accordingly. The probe is raster scanned relative to the sample surface using actuators, usually piezoelectric.
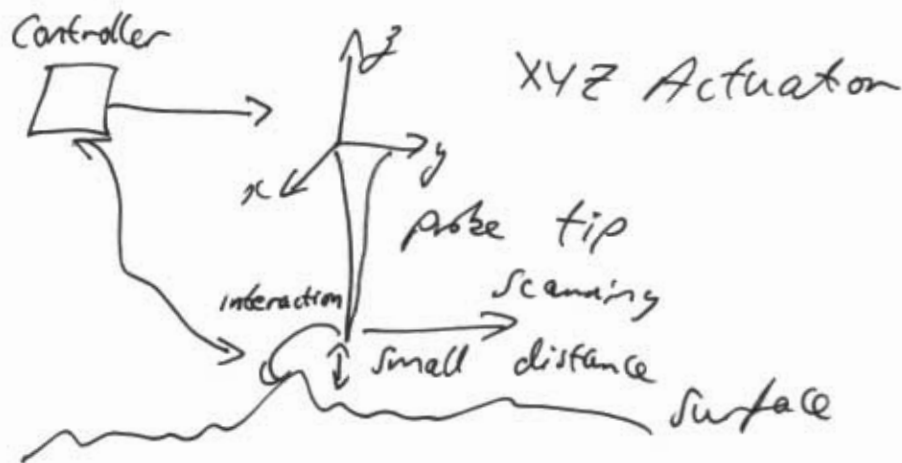
SPMs are used in R&D to measure surface roughness, feature sizes, for data storage, fault finding, materials analysis, routine quality control and as a way of exploring novel materials and devices.

(d)   The distance between a point on the tip or radius R that is a lateral distance, x, away from the apex (which is a distance d from the surface) is $d + \frac{x^2}{2R}$ away from the surface. As the current drops by a factor of ~ 10 for every 0.1 nm extra distance from the surface, we can estimate the distance of x for which $d + \frac{x^2}{2R} = d + 0.1$ nm as being the resolution. i.e. $x^2 = 0.2R => x \sim 1.3$ nm.
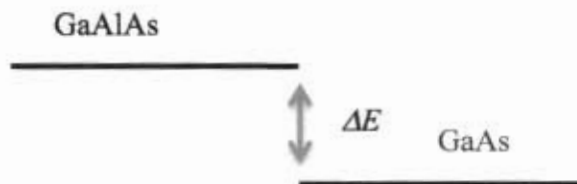
4.

(a)
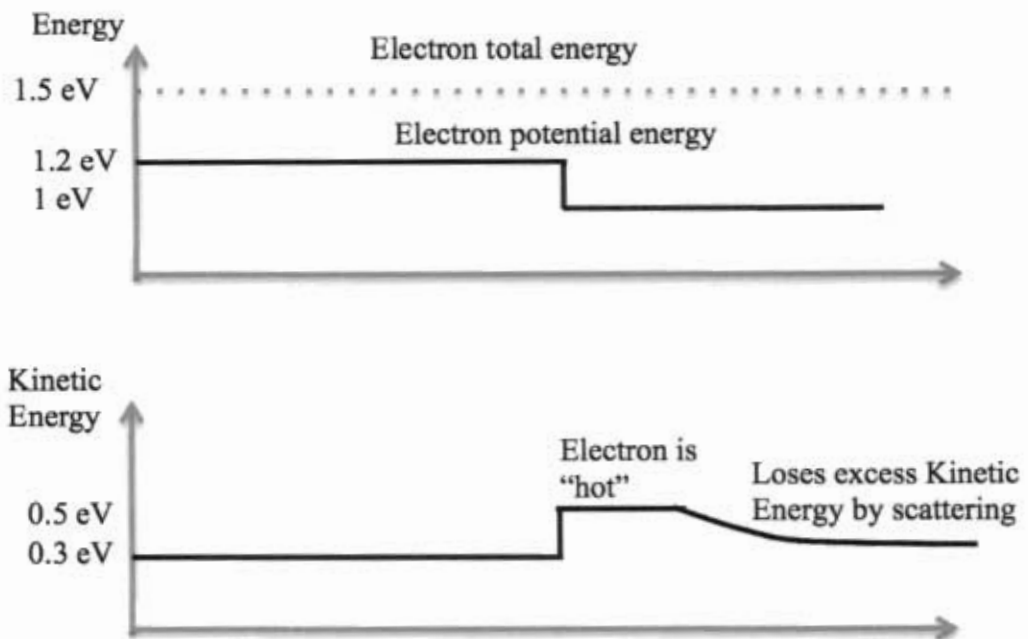
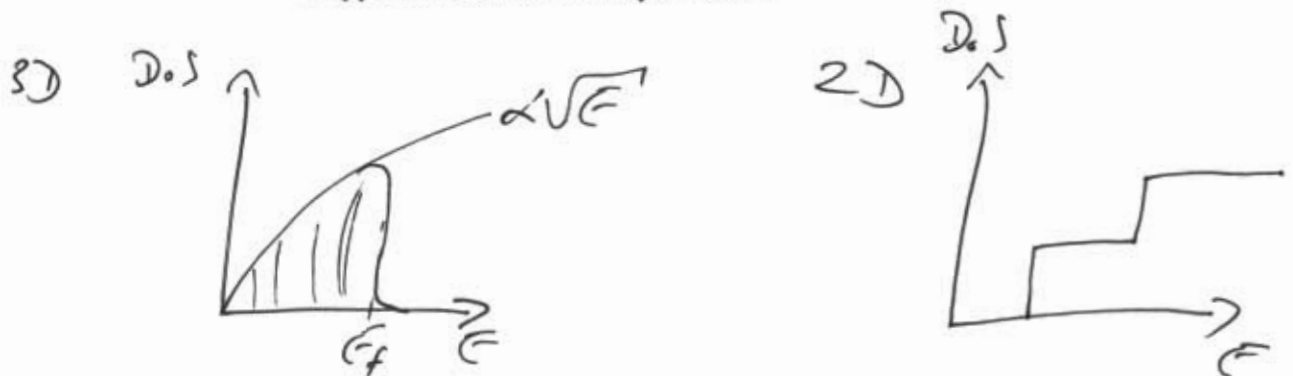(b)   Consider the two materials independently: they have different band-gaps, where the difference between them both is the band offset, ΔE.


When the materials are placed together, electrons will diffuse down the potential step from the GaAlAs where there is a higher concentration of electrons, and go to the GaAs where there are fewer electrons. Likewise, holes will diffuse to the left. The net effect of this is that a dipole layer of charge is formed just at the interface, which causes the bands to electrostatically bend, resulting in the following band structure:

GaAlAs

holes $\longrightarrow$ electrons diffuse

diffuse

GaAs

$\longrightarrow$ 2DEG in GaAs
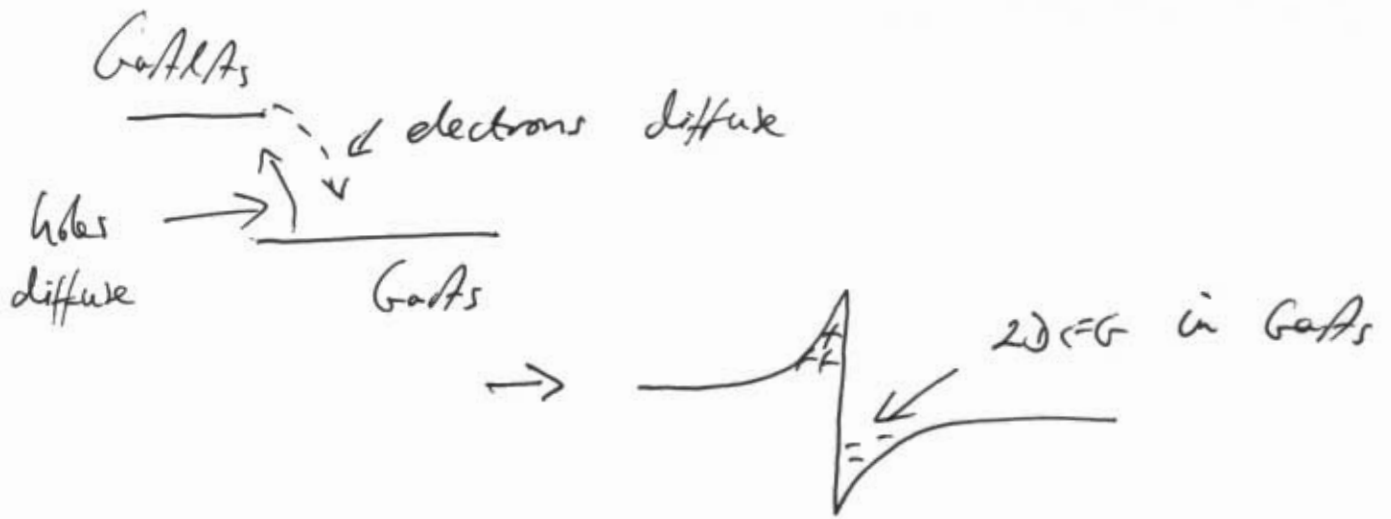
3D vs 2D is apparent from the density of states:

3D D.S $\propto \sqrt{E}$

E_f E

2D D.S

E

Also, in 2D, the electrons are in sub-bands, i.e.

3D

$\rightarrow k$

2D E $n=3$ $n=2$ $n=1$

$\rightarrow \epsilon$
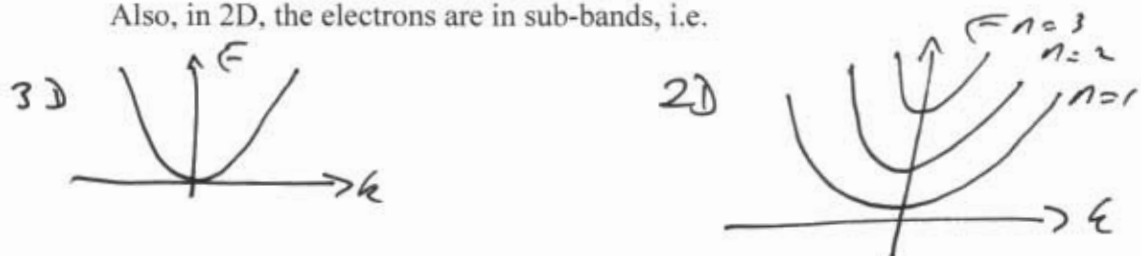
$2D: E = E_n + \dfrac{\hbar^2 k_x^2}{2m} + \dfrac{\hbar^2 k_y^2}{2m}$ where $E_n =$ Quantized levels of 2DEG.

The electrons are confined in a triangular well, which to a reasonable degree of accuracy can be approximated as an infinite square well, to get the basic behavior in principle.

(c)   High mobility: 2DEGs can be used to create very high-speed devices. This happens as the 2DEG is created within an *intrinsic* semiconductor, i.e. one without any dopants, so there is very little scattering of the electrons as they traverse the material (apart from scattering from surfaces, phonons and other electrons). An example of the implementation is in the modified bipolar transistor as shown below:
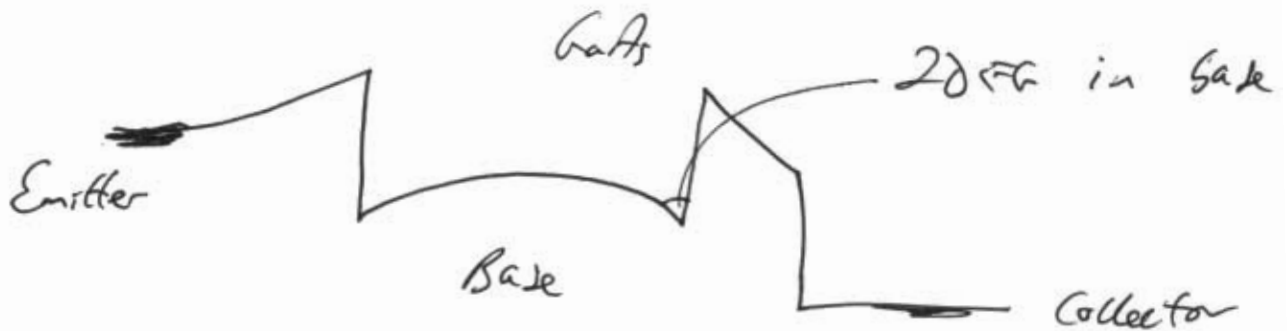
Semiconductor heterostructures are widely used to create lasers, e.g. the laser in a blu-ray player. They are also used to create high-speed devices, e.g. HEMTs and resonant tunneling structures. They are generally more complex and therefore more expensive to make than simple doped Si devices, so are typically only used in niche applications.

5.(a) Moore's law: the number of transistors per die increases by a factor of 2 roughly every 16-18 months. This is an economics-driven law and has been enabled by better fabrication techniques, better quality materials, reduced wavelength (Extreme UltraViolet, EUV) photolithography, better lithography masks, etc. Smaller generally means faster, more power efficient, and better functionality and interconnectivity.

(b)   Quantum: e.g. (i) Tunneling across the gate oxide in FETs – this reduces the gain of transistors at high frequency, and can be overcome through the use of high-k dielectrics. Intel has recently introduced the use of HfO_2 as the gate oxide in their transistors to help overcome this. (ii) Coherence and resonant tunneling, electron interference –there is no way around this apart from somehow randomizing electron phase, which can be orchestrated through

scattering, but then this would result in slow, noisy devices, and it gets more noticeable as dimensions shrink. It is manifested as oscillations in current-voltage characteristics. The best option is to actively design devices that take advantage of the electron coherence.

Classical: Electromigration – as a current flows, the electrons impart some momentum to the atoms in a wire and can dislodge them. This leads to a net atomic flux along the direction of electron flow, which ultimately leads to wire failure. This is the primary reason why electronic components eventually fail. It is particularly problematic in wires with cross-sectional dimensions below around 100 nm.

(c) The need for alternative architectures is driven by the lack of control of existing fabrication techniques, and the increasing complexity associated with smaller devices, e.g. tri-gate or wrapped-gate devices. RTDs based on highly doped p-n junctions are highly reliable and simple to fabricate, whereas those based on heterojunction band engineering structures are more complex to fabricate, in that there needs to be atomic-level control over the layers and the interfaces between materials. As a result, there are issues over reliability and reproducibility of such devices. However, with advances in thin-film deposition techniques, this is improving.

Resonant tunneling:

Resonant tunneling diode
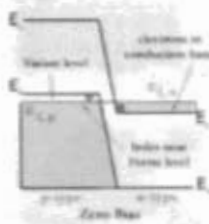
Zero Bias

Operating under zero bias

The electrons in the conduction band of the n-type side can tunnel into the p-type valence band if there are any holes just below E_{V}. Also electrons in valence band of p-type can tunnel into conduction band of n-type if there are holes just below the E_{C}. So, there is zero net current under zero bias.
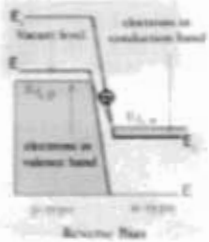

Reverse Bias

Operating under reverse bias

Under reverse bias the bands are raised on the p-type side and lowered on the n-type side. This results in a large electron current flowing easily from p to n by tunnelling across the depletion region. Therefore, under reverse bias, there will be a net conventional current flow from n-type into p-type.

Forward Bias

Operating under forward bias (1) small forward bias

Under low forward bias the bands are lowered on the p-type side and raised on the n-type side. The bias is so low (~ 0.6 V) that there is no substantial thermal diffusion current.
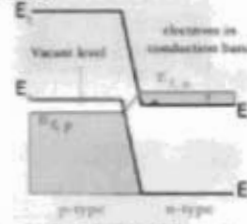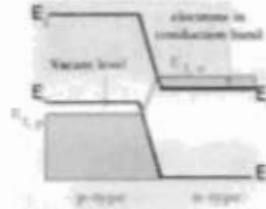
Forward Bias (increased)

Operating under forward bias (2) increased forward bias

As the forward-bias voltage increases, more of the n-type conduction band electrons are facing the forbidden energy gap of the p-type side. So this reduces the conduction to valence band current.
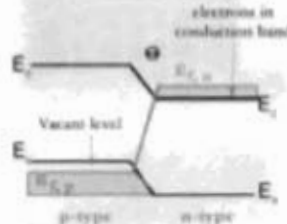
Forward Bias (at valley)

Operating under forward bias = valley voltage

As the forward bias voltage continues to increase, it will reach a value where forward current has decreased to its minimum value (no tunnelling is possible as there are no vacant states into which the electrons can tunnel).

Forward Bias (Thermal Diffusion)

Operating under forward bias > valley voltage

When the height of the barrier is reduced to a level that permits conventional (thermal diffusion) current to flow over the barrier, the forward current will increase as the thermal diffusion takes over as the main current carrying mechanism.