Photon Emission from metals in the Scanning Tunnelling Microscope

Andrew Richard Downes

Gonville and Caius College and the Department of Engineering

University of Cambridge

Thesis submitted for the degree of Doctor of Philosophy, September 1996
Preface

This dissertation describes work carried out in the Department of Engineering between October 1992 and September 1996. It is original work of the author and not part of any collaboration, except where reference or acknowledgement is made in the text. It has not been submitted in whole or part for a degree from any other university. This dissertation contains approximately 26,000 words.

Acknowledgements

I would like to acknowledge a number of people for their help and support during this PhD. Firstly, I must thank my supervisor, Dr Mark Welland (a.k.a. ‘the boss’) for providing all the equipment I needed and for his wealth of scientific knowledge and his relaxed and positive attitude. Dr Mark Taylor was of great help with the theory, and Dr Sandy McKinnon helped turn me into an experimentalist. Dave Roberts is thanked for being a speedy and expert technician, as is Reto Schlittler for building a superb STM.

Thanks go to the rest of the lab: Armelle, Andy H., Claudia, Dina, Eric, ‘Louis’, Mark, and Sean, and to those who left along the way: Charles, Frank, Hiromi, Jack, Martin, Silke, and Tim, for all being very helpful and friendly, and for creating a good atmosphere. Special thanks go to Debs and Sean for regularly accompanying me to the ‘famous Cross Keys of Cambridge’ and generally making life much nicer. Steve and Lewis were also around to help me relax.
 Photon Emission from metals in the Scanning Tunnelling Microscope

Andrew Richard Downes

The Scanning Tunnelling Microscope is an established surface science tool, combining unprecedented resolution with real space mapping. One of its major drawbacks, however, is that it gives no chemical information, but Photon Emission is able to probe the inelastic channel, which for metals contains invaluable chemical information. It has already been shown that flat metal surfaces produce differing emission spectra.

Real surfaces may not be flat, so it is important to know how curvature changes the yield of photons, and the emission spectrum. For this reason, a two sphere model was developed. The mechanism for Photon Emission, that of excitation of localised surface plasmons followed by either radiation or dielectric loss, was split into separate problems so that the dependence of the overall photon emission on materials and curvature is clear.

Experiments were performed on small silver particles, and for what was later believed to be a silver tip it was found that the Photon Emission was approximately proportional to both the tip and particle radius. Light emission was observed from clusters containing as few as ~30 atoms, and the first chemically specific photon maps were presented, which distinguished silver particles from carbide deposits. Emission was also seen on the Si(111)-7x7 surface.

The new theory compared well with spectra from flat surfaces, and when extended to two spheres it also showed that the Photon Emission was approximately proportional to both the tip and particle radius. It was predicted that the onset of emission would occur at a lower tip bias for gold particles than for silver particles, and this was confirmed experimentally. Metal-specific spectra will be produced if the plasmon modes do not move in energy when curvature varies, staying close to the travelling surface plasmon energy. This should occur for small radius tips and particles, and the movement of modes should be reduced for tungsten tips. However, as these tips produce low count rates, a compromise of a silver tip with radius ~100Å was thought best for distinguishing metal surfaces by their spectra. These spectra also ought to be taken at a bias higher than the corresponding travelling surface plasmon energy.

It was then demonstrated that light is emitted from atomic size metal contacts and the variation of intensity with input power fits extremely well to a model of spontaneous emission from high temperature electrons. This relation was used to deduce the electron temperature as a function of applied voltage. It was also shown that below a critical size, very high currents densities can occur for these contacts because the electrons transfer far less energy to the lattice.
5.5 Photon Emission from flat surfaces 85
5.6 Summary of Photon Emission experiments 90

6. RESULTS OF THEORY AND COMPARISON WITH EXPERIMENT
   6.1 Photon Emission from flat surfaces 91
   6.2 Photon Emission from metal spheres 100
   6.3 Summary 106

7. LIGHT EMISSION FROM HOT ELECTRONS IN ATOMIC SIZE CONTACTS
   7.1 Introduction 107
   7.2 Results from contacts formed in the STM 111
   7.3 Analysis 120

8. CONCLUSIONS 124

BIBLIOGRAPHY 126

APPENDIX: FORTRAN PROGRAM 128
1.1 Scanning Tunnelling Microscopy

The invention of the Scanning Tunnelling Microscope (STM) was reported by Binnig and Rohrer in 1982. An extremely sharp metal tip is brought to within a few atom diameters of a conducting surface and a small bias voltage applied across the vacuum gap. There exists a small quantum mechanical probability for electrons with classically insufficient energy to tunnel through the barrier. The tunnel current, of order nanoamps, is exponentially dependent on the tip-surface separation, with a decay length of 1Å, and a lateral resolution of 3Å can be achieved when there is a single atom on the end of the tip. This is sufficient to produce atomic resolution images.

No other surface analysis technique is able to give real-space imaging on an atomic scale – they rely on indirect methods like diffraction and an interaction over a large area of sample to produce data which transforms into the spatial arrangement of an ordered region of a surface. Additionally, the interaction continues several layers down into the bulk. The STM is therefore uniquely able to study disordered regions, the spatial distribution of ordered regions and the interface and growth between order and disorder. Also, the low voltages used and lack of physical interaction means that the STM need not damage the sample, unlike in for example Scanning Electron Microscopy.

Topographic information about the surface can be obtained by scanning the tip in a raster fashion across the sample while keeping the tunnel current constant by displacing the tip normal to the surface using a servo-feedback loop, as drawn in figure 1.1. The tip is moved in three orthogonal directions by applying voltages to piezoelectric materials which expand or contract continuously down to picometre resolution. If a simplistic model of tunnelling is assumed, then the current only varies due to the separation, so in this ‘constant current’ mode the separation remains constant and the result will be a three dimensional map of outermost atomic positions.
Figure 1.1. Schematic showing the tip's path in following the surface. Note the slight difference in the two resulting from the curvature of the sample being on a similar lengthscale to that of the separation. Flatter samples will produce more accurate topographic data.

Since the invention of the STM, a number of related non-intrusive scanning probe techniques have appeared. Atomic Force Microscopy relies on the attractive or repulsive forces between tip and sample as its feedback mechanism, and can be performed on non-conducting surfaces. Near-Field Optical Microscopy usually employs some sort of force feedback, and a sharpened optical fibre to collect reflected or transmitted light. Although it has spectroscopic capabilities, its resolution is typically 50nm. Electron Spin Resonance in STM, Magnetic Force Microscopy and Scanning Thermal Microscopy demonstrate that STM has come a long way in the last decade or so.

Inelastic Electron Tunnelling Spectroscopy (IETS), where electrons tunnel from the tip to the surface by losing energy to create phonons ($\hbar \omega \sim 100\text{meV}$) relating to core level interactions, is almost impossible because the noisy and much larger elastic current is superimposed on the signal. So the STM is unable to give chemical information about surfaces because the electrons giving rise to tunnelling are predominantly 'free' electrons. The STM is however able to map work function variations (by a sinusoidal modulation of the separation) and probe the local density of states of a surface reconstruction. In the next section, a technique is presented which combines the high spatial resolution of the STM with chemically dependent photon emission.
1.2 Photon Emission from the STM

McCarthy & Lambe (1977) reported enhanced light emission from tunnel junctions, and Gimzewski et al. (1989) were first to see this effect from an STM (see figure 1.1). This inelastic 'channel' is now observable - the elastic current does not cause light emission. Thus, information is available on interaction with electrons localised to the atom, rather than the free electrons which dominate the elastic current. Although excitations such as luminescence can occur, when metals are considered it is thought that electrons excite surface plasmons ($\hbar \omega \sim 1-4\text{eV}$) which radiate photons of the same energy. Limited quantitative experimental data exists in this area, but what there is suggests a complicated interaction, and the need for a sufficiently detailed theoretical model. The nature of the surface plasmons (and hence the emitted photons) depends on the dielectric function, both real and imaginary parts, which vary from metal to metal. So the goal of STM Photon Emission is to give chemical information about metal surfaces combined with the atomic resolution topography.

![Figure 1.2. Fluorescence spectra of the emitted light between 1 and 4 eV with two peaks evident, whose positions seem to be independent of the bias voltage. Taken with an iridium tip on a polycrystalline silver surface [Gimzewski].](image)
Surface plasmons radiate at optical frequencies - the spectrum of light radiated from a 100Å size silver particle coated Al-Al₂O₃-Au solid state tunnel junction showed a peak at
Chapter 1 : Introduction

1.8eV, with a quantum efficiency of around $10^{-5}$ photons per tunnelling electron [McCarthy]. In the case of the STM, figure 1.2 shows the spectra obtained for an iridium tip above a polycrystalline silver film by Gimzewski et al. [1989]. The quantum efficiency here was around $10^{-3}$. Berndt and Gimzewski went on to perform spatial mapping of the photon count with sub-nanometre resolution on a variety of metal surfaces. It can be seen in figure 1.3 that an enhancement in Photon Emission occurs when the surface is curved. In figure 1.4 it is also clear that a change of tip will produce different spectra.

![Figure 1.4](image)

Figure 1.4. Effect of tip material on fluorescence spectra. (a) is a W tip and (b) is an Ag tip, both taken on Ag(111) surfaces with $V_t=-3V$. (c) is a W tip, (d) is an Au tip, both taken on Au(110) surfaces with $V_t=3V$. In either case, the use of a W tip reduces the emission [Berndt 1993b].

These experiments suggest that the photon emission is dependent on the type of metal and geometry of both tip and sample. More free-electron like metals like silver and gold radiate more than those such as tungsten and iridium - whether acting as tip or sample, and the curvature of the surface greatly enhances the emission rate. The proposed mechanism was that fluctuations in the ordinary elastic current caused localised surface plasmons (LSPs) to be excited by inelastically tunnelling electrons. These plasmons which are confined by the tip-sample curvature then radiate.
Figure 1.5. (a) grey scale topography of a polycrystalline silver film deposited onto silicon. Size 450Å x 450Å x 120Å height (b) Photon intensity map taken with a photomultiplier tube [grey scale 0 - 20000 counts per second] [McKinnon].
More complex effects can be seen in figure 1.5 [McKinnon]. The photon map correlates with the topography, but some grains radiate more than others. This may be due to the nature of the grains and how they are 'connected'. They developed a system by which photon counts at a series of voltages can be taken at each pixel in the scan, and these can be compared with the topography and work function measurements.

Figure 1.6. Au(110) surface covered with an annealed monolayer of C60. (A) topograph and (B) photon map are represented as a grey scale. Area = 65Å × 65Å, $V_t = -2.8V$, $I_t = 4.4nA$. Height scale in (A) = 2.5Å, intensity scale in (B) = 800cps. Experiments performed at 50K. [Berndt 1993b]
Molecular resolution was achieved by Berndt et al. [1993c] from C_{60} monolayers on Au(110). The total photon flux went down when moving from the gold to the C_{60} islands, but on these islands, resolution in the photon maps was superior to that of the topography. The mechanism for contrast is not known, but the authors tentatively proposed that electromagnetic confinement effects were responsible (see figure 1.6).

![Figure 1.7. Cross sections through a Au surface – (a) topography (b) corresponding section of the photon map – low pass filtered. V_t = 3.0V, I_t = 4.4nA. Experiments performed at 50K. [Berndt 1995]](image)

Atomic variations in the STM Photon Emission signal in figure 1.7 were seen by Berndt et al. [1995] on Au(110) surfaces but can be ascribed to a separation modulation. The tip follows the atomic scale corrugations of the surface, but the LSP extends around 50Å, so the separation for tunnelling remains constant, but the separation for the tip-sample LSP is decreased, causing enhancement of the emission between atoms, as drawn in figure 1.8. So the actual resolution of the technique must be equal to the extent of the plasmon on the surface, or the size of the grain or particle containing the plasmon.
Some problems with the technique were becoming apparent. The quantum cut-off in figure 1.2 means that the energy of the tunnelling electron must be at least equal to the energy of the photon produced, so bias voltages of more than say 2V must be used. Voltages higher than this can cause instabilities in the tip and sample. The spectral information in figure 1.2 can only be taken at one point and integrated photon maps take a long time to build up. The total photon flux will be at best $10^{-3}$ (highest quantum efficiency) $\times$ tunnelling electrons per second. A typical tunnel current of a few nanoamperes ($10^{10}$ electrons per second) would produce $10^7$ photons per second. A collection efficiency in Gimzewski's apparatus [1989] of order 1%, and a detection efficiency of around 10% for an optical multi-channel analyser would give $10^4$ counts per second, which requires photon counting systems. Even if a photomultiplier tube is used to count the total number of photons, a scan with a good signal to noise ratio will take at least 20 minutes - much longer for surfaces with a lower quantum efficiency. And increasing the current to 'create' more counts can only be taken so far, otherwise instabilities will creep in again. A lot of expense and effort is also required to build a photon counting system around an STM, in ultrahigh vacuum.
Figure 1.9. Schematic map of a surface distinguishing two dielectric materials, based on photon intensity or photon spectra.

However, this powerful technique could be chemical specific (see figure 1.9), and gives valuable inelastic information that cannot be acquired by any other means. Not intended as a technology, it probes the quantum processes occurring at surfaces, increasing our understanding of the interaction of tip and sample, and moving our idea from the tip as an unintrusive probe to one of an antenna-like cavity.
Chapter 2
Theory

2.1 Elastic tunnelling

Perhaps the most basic explanation of an STM is that of a needle maintaining a fixed distance above a surface, but this idea of the non-invasive probe is far from true. In order to measure quantitatively, we must first understand which parameters can affect the tunnelling current, and hence the 'distance' which is being kept constant.

A one-dimensional arrangement of tip and surface can be used as an approximation for the tunnelling barrier in the STM, since extension to three dimensions gives very similar results [Stoll] – 70% of the tunnelling current can originate from the last atom on the tip [Soethout]. The tip and sample are generally different metals, each with its own work function, $\phi$, and Fermi energy, $E_F$.

Figure 2.1. Energy diagram showing elastic electron tunnelling from filled states in the tip to empty states in the sample.
One analysis is to consider tunnelling as a scattering of electrons through a potential barrier (the Simmons method [1963]), but for comparison with inelastic tunnelling the Transfer Hamiltonian approach [Bardeen] is used here. This method considers tunnelling to be the result of transitions between unperturbed eigenstates of the electrodes, which have been assumed to be free-electron metals in this analysis, and thermal broadening is ignored.

Electrons in filled states in the tip can tunnel to empty states of the same energy in the surface through a vacuum barrier (figure 2.1). This is what is meant by elastic tunnelling. Upon application of a negative bias to the tip, the energy of the electrons increases with respect to those in the sample. States in the tip with energies up to eV below the Fermi level are able to tunnel elastically. An averaged barrier is used, indicated by the dashed line.

Electron wavefunctions on left and right are given by

\[
\chi_L = C \exp (-\alpha_L(z-s/2)) \quad \quad \quad \chi_R = C' \exp (-\alpha_R(z+s/2)).
\]  

L and R refer to left and right (tip and sample), \(\alpha_L\) refers to the inverse decay length of the evanescent decay of wavefunction amplitude for electrons tunnelling from left to right and C, C' are normalisation constants which cancel out later in the calculation. \(d\) is the width of the gap.

The matrix element for electron transfer is taken from Bardeen's paper,

\[
M = -\frac{\hbar^2}{2m} \int \Delta S (\nabla \chi_L^* \chi_R - \nabla \chi_R^* \chi_L)
\]  

The elastic current rate (electrons per second) \(\nu_{el}\) comes from Fermi's "Golden Rule", summed over states in the tip and corresponding states in the sample. The delta-function ensures that electrons conserve energy when tunnelling.

\[
\nu_{el} = 2 \frac{2\pi}{\hbar} \sum |M|^2 \delta(E_L - E_R) f_L(1-f_R)
\]  

for free electron metals leads to

\[
\nu_{el} = \frac{2A}{\pi^2 s m U_L U_R} \int_{E_F}^{E_{F-eV}} \sqrt{E_L E_R} \alpha_L^3 \exp(-2s\alpha_L)
\]
where \( \alpha = \frac{\sqrt{2m\phi_{\text{eff}}}}{\hbar} \) and \( U = E_F + \phi_{\text{eff}} \).

\( \alpha, U, E_F \) and \( \phi_{\text{eff}} \) have different values on the left and right of the barrier, hence the subscripts. \( m \) is the electron mass, \( A \) is the area of the tunnel junction and \( \phi_{\text{eff}} \) is the barrier height. The Fermi functions \( (f_L, f_R) \) can be assumed not to be thermally broadened when \( eV \gg kT \).

One improvement would be to treat the gap as a continuum of vanishingly narrow barriers with varying heights and integrate the tunnelling probability, so that the barrier is trapezoidal instead of rectangular (the WKB approximation, see Gasiorowicz [1974]). However this is analytically messy and would drastically increase the computing time required for modelling.

\[ \text{Figure 2.2. Field emission regime: } eV > \phi_{\text{sample}}. \]
Gundlach resonances [Gundlach] occur for $V > -\phi$ (see figure 2.2). Here, the barrier dips below the Fermi energy, so tunnelling does not take place in part of the gap. The electrons become wave-like again and undergo reflections at the mismatch in potential at the vacuum-sample interface. An average barrier approximation is therefore inappropriate for biases of similar size of the barrier, but is likely just to be numerically inaccurate for biases in the region of 1-4V. For specific widths of this wave-like region, or specific bias voltages, there are maxima and minima in the transmitted electron intensity when the wavefunctions interfere constructively and destructively.

The density of states of the free electron metals are proportional to $\sqrt{E_L}$ and $\sqrt{E_R}$ in (2.4). If, instead of inserting this form for free-electron metals, a general density of states is left in, an expression for the tunnel current would become

$$I_{el} = \frac{4\pi e}{\hbar} \int_{E_F-EV}^{E_F} dE_L |M|^2 (E) \rho_L(E) \rho_R(E + eV)$$

showing a dependence of the current on the densities of states of tip and sample. So, the STM can be used to map the product of densities of states of tip and sample at certain energies, determined by the bias voltage. (If two maps are taken at voltages $V_1$ and $V_1+\Delta V$, then the expression in the integral is almost constant, so the difference between currents will be proportional to the product of densities of states in the small voltage range, $\Delta V$). However, this spectroscopy rarely provides any chemical information. The surface density of states originates from the free electrons in a periodic potential, terminated at a surface, so any atom in that position will give rise to a similar density of states, unless chemical potentials ($\approx$Fermi energies) are sufficiently different.

The vertical resolution of the STM can be seen in the exponential dependence of the tunnel current on the separation in (2.4): $\exp(-2s\alpha_L)$. Also, the barrier height, $\phi_{eff}$, affects the tunnel current - for example, a lower barrier causes a higher probability of transition which increases the current, or in constant current mode, the tip retreats slightly. Only slightly, because the highly exponential dependence on distance, $s$, means that for every angstrom retreated, the current falls by approximately one order of magnitude. So a slight retreat (less than an angstrom) due to a barrier height change will probably be indistinguishable from topography changes.

To identify this barrier height change, the tip-sample distance is modulated at a frequency higher than the response of the feedback loop, and the modulation in the current at that frequency as found by a lock-in amplifier shows variations in the barrier height. If in (2.4), a small voltage is used, the integral can be taken as constant, and so the current becomes
\[ I_{el} = \frac{2A \hbar \pi^2}{sU_L U_R} eV \sqrt{E_L E_R} \alpha_L^3 \exp(-2s\alpha_L) \]  

(2.6)

and \( \frac{\partial I}{\partial s} = (-2\alpha_L) I_{el} - \frac{1}{s} I_{el} \)  

(2.7)

But if the current is kept constant (for frequencies well below that of the modulation), then \( \frac{\partial I}{\partial s} \) is effectively divided by \( I \). So,

\[ \frac{\partial I}{I \partial s} = (-2\alpha_L - \frac{1}{s}) \frac{-2\sqrt{2m\phi_{eff}(L)}}{h} - \frac{1}{s} \]  

(2.8)

Or, the map of \( \frac{\partial I}{\partial s} \) in constant current mode represents a map of \( \sqrt{\phi_{eff}(L)} = \sqrt{\phi_L + \phi_R - eV} \) which will only represent changes in the square root of the barrier height between the tip and sample (not just the work function, \( \phi_R \)) and then only if the tip does not change during the scan. Also there is an offset of \( \frac{1}{s} \) which is unknown and changes when the interaction area changes. A map of \( \frac{\partial I}{\partial s} \) will show up variations due to changes in sample work function, and will work better on flat surfaces where the interaction remains constant (i.e. where \( s \) and \( A \) don’t change).

In order to get a map of \( \frac{\partial I}{\partial s} \) for surfaces which are not flat it is necessary to normalise them from \( \frac{\partial I}{\partial z} \) which is what is actually measured, where \( s \) is the direction of tunnelling and \( z \) is in the direction of oscillation, in figure 2.3. So the measurements have to be divided by \( \cos(\theta) \), the angle between the surface normal and the direction of oscillation. This geometric factor dominates over the actual changes in the barrier height for granular surfaces and particles, but its removal is not common. Even when this removal is done, the difference between the barrier heights of most conductors is not great. Work functions tend to be in the range 4 - 4.5eV for noble and transition metals [Ashcroft and Mermin]. Silver and gold, for example, have the same value for their work function – (around 4.3eV, although this value may vary according to the method of measurement). The crystallographic face affects the work function and the classical dependence on particle size has been quoted as [Wood]:

\[ W(R) = W_\infty + \frac{3e^2}{8R} = W_\infty + \frac{5.40}{R(\text{Å})} eV \]  

(2.9)
It can be seen in (2.6) that at low biases, the current should be proportional to the voltage. Increasing the voltage at constant current will mean that the separation, s, will increase to compensate. Conversely, increasing the current will cause the separation to decrease. And because the lateral resolution is given by $\sqrt{2/\alpha^{-1}(R+s)}$ [Tersoff & Hamann], decreasing the gap by working at low biases and high currents will improve this resolution. Unfortunately, in photon emission high biases need to be used, which means there will be a loss of resolution (and the current cannot be too great or instabilities start to occur). This is only likely to be 2Å at worst when tunnelling at 3V as opposed to 400mV, for example.

2.2 Theories of STM Photon Emission

Photons can be generated by either of two mechanisms: plasmon-mediated inelastic electron tunnelling or decay of hot (elastic) electrons, illustrated in figure 2.4. In the first mechanism, the electrons tunnel and lose an amount of energy, $\hbar \omega$, to create a plasmon. These LSPs then decay to photons of the same energy and the electron is much closer to the Fermi level than it was. Phonon interactions later bring the electron down to the Fermi energy. In the second case, electrons tunnel elastically, and then lose energy by creating photons, so bringing themselves closer to the Fermi level. It was shown experimentally [Berndt and Gimzewski 1991a] that the photon emission in the STM was proportional to the density of states in the sample at $\hbar \omega$ below the Fermi level of the tip, indicating that inelastic tunnelling is responsible (at least predominantly).
Metals can be thought of as a plasma of free electrons in a fixed lattice of positive ions. Bulk plasmons are oscillations of charge density throughout the volume of the metal having energy \( \hbar \omega_p = \hbar \sqrt{n e^2 / \varepsilon_m} \), around UV frequencies, where \( n \) is the density of free electrons.

Maxwell's equations also allow electromagnetic waves to be present on metallic surfaces, with a range of energies from \( \hbar \omega = 0 \) to \( \hbar \omega_p / \sqrt{2} \). Their electric field amplitude decays exponentially into the surface, with a characteristic skin depth. These oscillations of surface charge density, surface plasmons, occur around optical frequencies. They cannot radiate from flat surfaces, as there is no net dipole moment in any direction (fig. 2.5a). A rough or curved geometry is essential for plasmons to radiate (fig. 2.5b).

\[ \text{Figure 2.4. Left: inelastic energy diagram; right: hot electron energy diagram.} \]

\[ \text{Figure 2.5. Charges on (a) a flat surface; (b) a curved surface.} \]
McCarthy and Lambe [1977] did not observe photon emission for flat surfaces in their Al-Al₂O₃-Ag tunnel junctions, proving that a direct electron–photon transition was not the responsible mechanism, although it is noted by Mills et al. [1982] that gold is able to emit photons *directly* in a one-step process with an efficiency comparable to that of the two-step process.

For very curved geometries, the surface plasmons are localised in the region of curvature, e.g. under the STM tip. This contrasts with the flatter surfaces encountered in solid state tunnel junctions, where travelling plasmons (which obey 'fast' or 'slow' mode dispersion curves) predominate. So in the STM geometry, these surface plasmons are localised and there is sufficient curvature for them to radiate even on a flat surface, as the inclusion of the tip has broken the translational symmetry (see figure 2.6).

![Figure 2.6](image-url)

*Figure 2.6. Tip above surface, inducing a net dipole. Compare with the flat surface in figure 2.5(a)*

Rendell and Scalapino [1978] presented a calculation of the frequency-dependent dipole moment and the radiation emission spectrum for a small particle located above a metal film excited in a solid state tunnel junction. They modelled the geometry as a sphere above a plane, and found analytical expressions for the potential, and therefore dipole moment as a function of frequency. They produced emission spectra for gold particles on an aluminium film, taking into account the absorptive nature of the gold in the imaginary part of its dielectric function, and the reflections from the aluminium. No retardation effects were included because the size of the system is much smaller than the wavelength of light produced. Their resonance-like structure at 1.9eV was in good agreement with spectra taken of gold particles on oxide-covered aluminium [Hansma].
Later [Rendell 1981], they extended their model to microstructures sitting on oxide. They investigated surface plasmon modes in smooth metal films above the oxide, and found approximate forms for surface plasmon frequencies for microstructures such as thin rods, cylindrical slabs and spherical particles. For thin rods, they found optical resonances for aspect ratios (height to diameter) of 3-7. For the spherical particles, they found that the excitation spectrum to be dominated by the dipole term in the multipole expansion, and that emission increased when the radius of particle increased or when the sphere-plane separation decreased. Also, the broadening of spectral modes was caused by the imaginary part of the dielectric function. In this model, they included a tunnelling excitation term - the LSPs must be excited somehow (for energy conservation) and noise in the tunnel current at optical frequencies 'drives' the plasmons.

Batson [1982] looked at energy losses in 100keV electrons due to the excitation of clusters of aluminium spheres, but incorrectly solved for the potentials in two spheres with variable radii and separation. However, his experimental results implied that a two-sphere plasmon has a lower energy than that of an isolated sphere.

![Figure 2.7. Sphere-plane geometry used to model an STM tip above a surface.](image)

Johansson et al. [1990] used the potentials found by Rendell and Scalapino for an iridium tip (sphere) above a silver film (plane) – modelled in figure 2.7 – and derived the intensity of the emitted light by using the reciprocity theorem in electrodynamics. However, their values for differential power were derived numerically and were wrapped up in a geometry factor for the sphere-plane system, so the mechanism for photon emission was not clear – what aspects of the mechanism would change for different metals, sizes, separation, bias voltage, etc. The results for the spectra were qualitatively good, if not quantitative – and the overall quantum efficiency of $10^{-4}$ (for a sphere on a plane) agrees more with Gimzewski's results for Ag(111) than his for polycrystalline silver film ($10^{-3}$). Johansson took experimental spectra from Berndt and Gimzewski [1993b] taken on Ag(111), Au(110) and Cu(111) surfaces and compared it to the results of his model in figure 2.8.
Figure 2.8. Johansson's theoretical predictions (c and d) and Berndt and Gimzewski's spectra (a and b) for flat samples and a variety of tips [Berndt 1993b].

Uehara et al. [1992] presented a theory similar to Johansson's, but with either a prism coupled geometry (by which travelling surface plasmons decay into photons and are collected by a prism beneath the thin film sample) or the tip side emission geometry as modelled by Johansson, and collected by Gimzewski and Berndt. They found in their model that these surface plasmon polaritons are not directly excited, but are the product of the decay of the localised surface plasmons. They also included a dependence of photon emission on surface roughness, but this effect was found to be less than 0.1% for typical surfaces.

Tsukada and co-workers [1992, Schimizu 1992] developed a microscopic theory of light emission. They claimed that the power spectrum is composed of a macroscopic part, representing the coarse-grain shape of the tip and surface, and a microscopic part, containing information on the electronic states of the surface. For a tungsten tip composed of ten atoms scanning above an Ag(111) surface, a 10% variation of the optical signal was predicted.

Persson and Baratoff [1992] considered photon emission in electron tunnelling to a small metal particle, comparing the efficiencies for emission via the plasmon mode and as a decay channel of hot electrons. The former process was found to dominate over the latter by a factor of $10^3$ for particles of a few hundred angstroms radius, but by a smaller factor as the particle size decreases.

These models have been able to give some insight into STM photon emission, giving spectra which have similarities to those taken at least on flat surfaces, and produce quantum efficiencies close to those found in experiment. Some factors affecting emission have also become clear. Coating a tungsten tip with silver will cause enhancement and reduce spectral
broadening of emission, the explanation given being the decreased losses due to the low imaginary part of the dielectric function of silver. It is accepted that the mechanism is one of excitation of localised surface plasmons, which decay into photons.

A goal of this field is to chemically assign metallic parts of a surface by their spectra or photon intensity. A good test would be to distinguish silver and gold, as their Fermi energies and work functions are extremely close. However, a number of factors are responsible for the form (shape and magnitude) of their spectra. Both tip and sample material will affect position of the peak(s) and the broadening of the spectra; tip-sample separation and tip and sample radii will also affect the spectra. For flat samples, e.g. distinguishing Ag(111) from Au(110) seems straightforward from figure 2.8, but one would not expect these two faces to occur on the same sample – real samples would contain grains, clusters or nanostructures of unknown metal(s). So it is important to decide whether sample curvature produces variations from the flat surface ‘signature’ which would render the metal(s) unable to be identified.

The majority of the rest of the thesis addresses this lack of understanding of how sample curvature affects photon emission. In the next chapter, a complete and detailed theory of photon emission is presented, where the tip and sample are modelled as spheres, so that the dependence of spectra and intensity on sample curvature may be investigated. Chapter 4 describes the apparatus required and the design of the optical collection system. Then in chapter 5, silver clusters are prepared on a silicon substrate, photon maps are recorded, and quantitative measurements are made on these particles. The results of the theory are shown in chapter 6, and compared with these experimental findings. Chapter 7 is a departure from the rest of the thesis, as light emission from an STM tip in contact with a metal surface is described in detail for the first time. A mechanism very different from that of plasmon excitation is found to be responsible.
Chapter 3
Theory for two sphere geometry

3.1 Bispherical co-ordinates

In order to describe a geometry for which surfaces may have curvature, a two-sphere geometry was chosen to simulate the tunnel junction in the STM. Figure 3.1 illustrates that the two spheres have variable radii and separation. The three Cartesian co-ordinates x,y,z are related to the bispherical co-ordinates [Dean] of $\mu$, $\eta$ and $\phi$ by the following transformations:

$$
\begin{align*}
    x &= \frac{a \sin \eta \cos \phi}{\cosh \mu - \cos \eta} \\
    y &= \frac{a \sin \eta \sin \phi}{\cosh \mu - \cos \eta} \\
    z &= \frac{a \sinh \mu}{\cosh \mu - \cos \eta}
\end{align*}
$$

(3.1)

Figure 3.1. Model of the STM geometry. Top and bottom spheres represent the tip and sample respectively. $\mu = \mu_1$ on the surface of the top sphere, and $\mu = \mu_2$ on the bottom sphere's surface.
The scaling parameter, $a$, is given by

$$a = \sqrt{\frac{q(q+4R_1R_2)}{2(s+R_1+R_2)}}$$

(3.2) where $q = s(s+2(R_1+R_2))$

A particular value of $\mu$ represents the surface of a sphere. The smaller the radius, the larger the value of $\mu$.

$$\mu_1 = \sinh^{-1} \left( \frac{a}{R_1} \right) \quad \mu_2 = -\sinh^{-1} \left( \frac{a}{R_2} \right)$$

(3.3)

The sphere-plane geometry can also be analysed as a subset of the two-sphere geometry by putting $\mu_2=0$ and approximating $R_2$ as infinite.

Now there exists a convenient geometry in whose co-ordinate system all the mathematics must be undertaken. The problem breaks down into a number of smaller tasks. Firstly, the potential of a plasmon in the spheres is found. This is done by solving Laplace’s equation in this geometry. Secondly, the radiation rate for an excited plasmon having this potential is determined. Thirdly, the rate of plasmon excitation by tunnelling electrons is calculated. Then the rate of plasmon de-excitation by interaction with electrons is found. When all these rates of excitation and de-excitation are known, they are combined in an equilibrium to determine the rate of photon emission. As the plasmon modes are damped by the de-excitation, they are spectrally broadened. The spectral distribution is found by a superposition of all the broadened plasmon modes.

### 3.2 Solving for potentials

Laplace’s equation ($\nabla^2 \psi = 0$; $\psi$ being the plasmon potential) must be satisfied in both the spheres and in the vacuum gap outside them if the wavelength of these charge oscillations is much larger than the size of the system - i.e. retardation effects are ignored. In the case of the STM, the tip and sample should have radii of no more than a few hundred angstroms and the measured wavelengths are several hundred nanometres, so it is a fair approximation to consider a non-retarded potential. There is cylindrical symmetry in this case, and a potential which satisfies Laplace’s equation in these co-ordinates with no $\phi$-dependence has been shown to be of the form

$$\psi = F \sum_{n} \left( A_n \exp\left(-\left(n + \frac{1}{2}\right)\mu\right) + B_n \exp\left(n + \frac{1}{2}\right)\mu\right) P_n(\cos \eta)$$

(3.4)
Chapter 3: Theory for two sphere geometry

[Rendell 1978]. However, in the top sphere the first term would become infinite as \( \mu \) heads to infinity, and similarly the second term is ignored for the bottom sphere as \( \mu \) becomes negative infinity. So the potential in all three regions (top sphere, gap, bottom sphere) is:

\[
\psi_1 = F \sum_n (A_n \exp(-(n + \frac{1}{2})\mu)) P_n(\cos \eta) \quad \mu \geq \mu_1 \quad (3.5)
\]

\[
\psi_2 = F \sum_n (A_n \exp(-(n + \frac{1}{2})\mu + B_n \exp(n + \frac{1}{2})\mu)) P_n(\cos \eta) \quad \mu_1 \geq \mu \geq \mu_2 \quad (3.6)
\]

\[
\psi_3 = F \sum_n (B_n \exp(n + \frac{1}{2})\mu) P_n(\cos \eta) \quad \mu \leq \mu_2 \quad (3.7)
\]

where \( F = \sqrt{\cosh \mu - \cos \eta} \) and \( P_n \) are Legendre Polynomials.

One set of boundary conditions is that the potentials must be matched at both surfaces, i.e.

\[
\psi_1(\mu_1) = \psi_2(\mu_1) \quad \text{which leads to} \quad A_n' = A_n + B_n \exp(2n + 1)\mu_1 \quad \text{and} \quad (3.8)
\]

\[
\psi_2(\mu_2) = \psi_3(\mu_2) \quad \text{which leads to} \quad B_n' = B_n + A_n \exp(-2n - 1)\mu_2 \quad . \quad (3.9)
\]

The other set of boundary conditions to apply is that the normal component of the electric displacement is continuous across both surfaces, i.e.

\[
D_n^+ (\mu_1) = D_n^+ (\mu_1) \quad \text{and} \quad D_n^+ (\mu_2) = D_n^+ (\mu_2) \quad (3.10)
\]

where \( D = \varepsilon_0 \varepsilon_r \mathbf{E} \) and \( \mathbf{E} = -\nabla \psi \)

\[
\nabla \psi = \frac{1}{h_\mu} \frac{\partial \psi}{\partial \mu} + \frac{1}{h_\eta} \frac{\partial \psi}{\partial \eta} + \frac{1}{h_\phi} \frac{\partial \psi}{\partial \phi} \quad (3.11)
\]

where \( h_\mu = h_\phi = \frac{a}{\cosh \mu - \cos \eta} \) and \( h_\eta = \frac{a \sin \eta}{\cosh \mu - \cos \eta} \). \quad (3.12)

The last two terms in (3.11) are zero in the direction normal to the surface. \( \varepsilon_r = \varepsilon_1 \) in the top sphere, \( \varepsilon_2 \) in the second sphere, and 1 in the vacuum gap.

The implementation of these two boundary conditions leads to two equations relating \( A_n \) and \( B_n \).
Chapter 3 : Theory for two sphere geometry

\( (1 - \varepsilon_1) (A_n + B_n \exp(2n + 1)\mu_1) \frac{\sinh \mu_1}{2} - \left( n + \frac{1}{2} \right) (1 - \varepsilon_1) (\cosh \mu_1 - \cos \eta) A_n \)

\( + \left( n + \frac{1}{2} \right) (1 + \varepsilon_1) (\cosh \mu_1 - \cos \eta) B_n \exp(2n + 1)\mu_1 \)

\( \times P_n(\cos \eta) = 0 \)  \hspace{1cm} (3.13)

\( (1 - \varepsilon_2) (A_n \exp(-(2n + 1)\mu_2) + B_n) \frac{\sinh \mu_2}{2} + \left( n + \frac{1}{2} \right) (1 - \varepsilon_2) (\cosh \mu_2 - \cos \eta) B_n \)

\( - \left( n + \frac{1}{2} \right) (1 + \varepsilon_2) (\cosh \mu_2 - \cos \eta) A_n \exp(-(2n + 1)\mu_2) \)

\( \times P_n(\cos \eta) = 0 \)  \hspace{1cm} (3.14)

However, as \( A_n \) and \( B_n \) cannot themselves be a function of \( \eta \), the terms in \( \cos \eta \) must be removed by the use of the recurrence relation for Legendre Polynomials,

\( (2n + 1) \cos \eta \quad P_n(\cos \eta) = nP_{n-1}(\cos \eta) + (n+1)P_{n+1}(\cos \eta). \)  \hspace{1cm} (3.15)

This gives rise to a recurrence relation with terms in \( n+1, n, \) and \( n-1 \) which is represented here in matrix form.

\[ P_{n-1}\begin{pmatrix} A_{n-1} \\ B_{n-1} \end{pmatrix} + Q_n\begin{pmatrix} A_n \\ B_n \end{pmatrix} + R_{n+1}\begin{pmatrix} A_{n+1} \\ B_{n+1} \end{pmatrix} = 0 \]  \hspace{1cm} (3.16)

where

\[ P_{n-1} = n\begin{pmatrix} -\chi_1 \exp(\mu_1) & \exp(2n\mu_1) \\ \exp(-2n\mu_2) & -\chi_2 \exp(-\mu_2) \end{pmatrix} \]  \hspace{1cm} (3.17)

\[ Q_n = \begin{pmatrix} \chi_1 \left( (2n + 1) \cosh \mu_1 - \sinh \mu_1 \right) & -(2n + 1) \cosh \mu_1 + \chi_1 \sinh \mu_1 \exp(2n + 1)\mu_1 \\ \chi_2 \left( 2n + 1 \cosh \mu_2 - \sinh \mu_2 \right) \exp(-(2n + 1)\mu_2) & \chi_2 \left( (2n + 1) \cosh \mu_2 + \sinh \mu_2 \right) \end{pmatrix} \]  \hspace{1cm} (3.18)

\[ R_{n+1} = (n + 1)\begin{pmatrix} -\chi_1 \exp(-\mu_1) & \exp((2n + 2)\mu_1) \\ \exp(-(2n + 2)\mu_2) & -\chi_2 \exp(\mu_2) \end{pmatrix} \]  \hspace{1cm} (3.19)

and \( \chi_i(\omega) = \frac{1 - \varepsilon_i(\omega)}{1 + \varepsilon_i(\omega)}. \)

Equation (3.16) can be rewritten as an infinite matrix which is equivalent to an infinite series of 2x2 matrix equations for all \( n \), or as one terminated after a large number of terms (up to \( n_{\text{max}} \)).
where vectors $v_n = \begin{pmatrix} A_n \\ B_n \end{pmatrix}$ and all terms off the diagonal are zero.

The matrix in equation (3.20) can be simplified to the product of a lower and an upper matrix,

\[
\begin{bmatrix}
I & I \\
\hat{P}_0' & I \\
\hat{P}_1' & I \\
\hat{P}_2' & I \\
\vdots \\
\hat{P}_{n_{\text{max}}-1}' & I \\
\end{bmatrix}
= \begin{bmatrix}
Q_0 & R_1 \\
Q_1 & R_2 \\
Q_2 & R_3 \\
Q_3 & R_3 \\
\vdots \\
Q_{n_{\text{max}}-1} & R_{n_{\text{max}}} \\
\end{bmatrix}
\begin{bmatrix}
\hat{v}_0 \\
\hat{v}_1 \\
\vdots \\
\hat{v}_{n_{\text{max}}-1} \\
\hat{v}_{n_{\text{max}}} \\
\end{bmatrix} = 0
\]

The determinant of the lower (left-hand) matrix is equal to 1, and the determinant of the upper (right-hand) matrix is equal to $\det(Q_1') \det(Q_2') \det(Q_3') \cdots \det(Q_{n_{\text{max}}}')$.

So whenever this product of determinants is equal to zero, we have found a frequency at which solutions to (3.16) exist – the only variables in (3.16) which are functions of frequency are the relative permittivities, $\varepsilon_1$ and $\varepsilon_2$. The equation

\[
\begin{pmatrix}
Q_{n_{\text{max}}}^{(1)} & Q_{n_{\text{max}}}^{(2)} & A_{n_{\text{max}}} \\
Q_{n_{\text{max}}}^{(3)} & Q_{n_{\text{max}}}^{(4)} & B_{n_{\text{max}}} \\
\end{pmatrix} = 0
\]

can be arranged as

\[
\frac{A_{n_{\text{max}}}}{B_{n_{\text{max}}}} = -\frac{Q_{n_{\text{max}}}^{(4)}}{Q_{n_{\text{max}}}^{(3)}} = -\frac{Q_{n_{\text{max}}}^{(2)}}{Q_{n_{\text{max}}}^{(1)}}.
\]
So, by putting $B_{n_{\text{max}}} = 1$, unnormalised values of $A_{n_{\text{max}}}$ and $B_{n_{\text{max}}}$ can be found, and by moving up one from the bottom line in the matrix in (3.20), i.e. to the equation

$$P_{n_{\text{max}} - 1} \nu_{n_{\text{max}} - 1} + Q_{n_{\text{max}}} \nu_{n_{\text{max}}} = 0,$$  \hspace{1cm} (3.21)$$

$A_{n_{\text{max}} - 1}$ and $B_{n_{\text{max}} - 1}$ can be evaluated. By climbing up the matrix in (3.20) line by line, all the values for $A_n$ and $B_n$ will be forthcoming.

It is convenient now to normalise these values of $A_n$ and $B_n$, as they represent the magnitude of the potential. This is done by equating the electrostatic energy to that of the ground state energy of the mode in question, i.e.

$$\frac{1}{2} \hbar \omega_{\text{mode}} = \frac{1}{2} \varepsilon_0 \int_{V} E^2 \, d\tau \quad \text{where $V$ constitutes all space; and using Green's first indentity,}$$

$$\frac{1}{2} \hbar \omega_{\text{mode}} = \frac{1}{2} \varepsilon_0 \left[ \int_{s_1} \psi_1 \left( \frac{\partial \psi_1^*}{\partial n} - \frac{\partial \psi_2^*}{\partial n} \right) ds_1 + \int_{s_2} \psi_3 \left( \frac{\partial \psi_2^*}{\partial n} - \frac{\partial \psi_3^*}{\partial n} \right) ds_2 \right],$$  \hspace{1cm} (3.22)$$

where $n$ is in the direction of the normal outwards from the spheres.

So $\frac{\partial \psi_1}{\partial n} = \frac{1}{h} \frac{\partial \psi_1}{\partial \mu}$ evaluated at $\mu = \mu_1$ on the surface of sphere 1, and $\frac{\partial \psi_1}{\partial n} = \frac{1}{h} \frac{\partial \psi_1}{\partial \mu}$ evaluated at $\mu = \mu_2$ on the surface of sphere 2.

After more algebra, we arrive at

$$\frac{1}{2} \hbar \omega_{\text{mode}} = \frac{1}{2} \varepsilon_0 a^4 \pi \left[ \sum_n B_n \left( A_n + B_n \exp(2n + 1) \mu_1 \right) + \sum_n A_n \left( B_n + A_n \exp-(2n + 1) \mu_2 \right) \right]$$  \hspace{1cm} (3.23)$$

which gives the normalisation factor to obtain the actual $A_n$ and $B_n$.

Now we have the potential of a particular mode in all regions, $\psi$, we can work out the power radiated from this oscillating dipole for this mode. Actually, $n_{\text{max}}$ modes are found as there are this many terms in the product of determinants.
Chapter 3 : Theory for two sphere geometry

3.3 Radiation rate

The cylindrically symmetric potentials in (3.4) have no \( \phi \)-dependence, meaning that the charge distributions on the spheres cannot give rise to any net dipole moment in the x-y plane. So there exists only a dipole moment in the z-direction. Starting from the now normalised potentials \( \psi \), the charge densities on spheres 1 and 2 respectively are

\[
\sigma_1 = \varepsilon_0 \left[ E_{1n} - E_{2n} \right]_{\mu=\mu_1} \\
\sigma_2 = \varepsilon_0 \left[ E_{2n} - E_{3n} \right]_{\mu=\mu_2}
\]

(3.24)

where \( E_{mn} = -\frac{1}{\hbar} \frac{\partial \psi_m}{\partial \mu} \) with \( \psi_m \) coming from (3.15),(3.16) or (3.17).

The dipole moments in the z-direction for spheres 1 and 2 are

\[
P_{z1} = \int_{s_1} \sigma_1 \, z \, ds_1 \quad \text{and} \quad P_{z2} = \int_{s_2} \sigma_2 \, z \, ds_2
\]

(3.25)

The total dipole is just the sum of the separate dipoles in both spheres.

\[
P_z = 4 \sqrt{\frac{\pi}{3}} \frac{a^2}{\varepsilon_0} \sum_n (2n + 1) (B_n - A_n)
\]

(3.26)

And from classical electrodynamics, the power radiated from these oscillating dipoles is

\[
\text{Power} = \frac{\omega^4}{32 \pi^2 \varepsilon_0 c^3} \int \left| P_z \right|^2 \, d\Omega \sin^2 \theta
\]

(3.27)

\[
= \frac{4 \omega^4 a^4 \varepsilon_0}{3 c^3} \left[ \sum_n (2n + 1) (B_n - A_n) \right]^2
\]

(3.28)

If there exists a metallic reflecting plane underneath the two spheres, as in figure 3.2, then (3.27) must be replaced by

\[
\text{Power} = \frac{\omega^4}{32 \pi^2 \varepsilon_0 c^3} \int \left| P_{z,\text{half}} \right|^2 \left( 1 + T_z \right)^2 \, d\Omega \sin^2 \theta
\]

(3.29)

where \( T_z \) is the Fresnel coefficient of a dipole perpendicular to a reflecting plane,
\[ T_z = \frac{\sqrt{\varepsilon_2 \cos \theta - 1 - \frac{\sin^2 \theta}{\varepsilon_2}}}{\sqrt{\varepsilon_2 \cos \theta + 1 - \frac{\sin^2 \theta}{\varepsilon_2}}}. \]  

(3.30)

Figure 3.2. Reflection of electromagnetic waves by the half-space below the sample sphere causes enhancement or cancellation of the radiation.

This value for the power is converted into a rate of photon emission – i.e. photons emitted per second – by dividing by \( \hbar \omega \) to give a rate which we will call \( \nu_{\text{rad}} \).

Both spheres should be charge neutral. This can be used to test the accuracy of the calculations leading to \( A_n \) and \( B_n \), as the expressions for the charges on the spheres are

\[ Q_1 = \int_{s_1} \sigma_1 \, dS \quad \text{and} \quad Q_2 = \int_{s_2} \sigma_2 \, dS. \]

Using (3.22) we arrive at

\[ Q_1 = 2 \pi \varepsilon_0 \delta \sum B_n \quad \text{and} \quad Q_2 = 2 \pi \varepsilon_0 \delta \sum A_n. \]
The errors in these sums (away from zero) are always found to be below $10^{-5}$ from the Fortran Program for each mode, when compared to $2\pi\varepsilon_0 a \sum |B_n|$ and $2\pi\varepsilon_0 a \sum |A_n|$.

### 3.4 Matrix element

The radiation rate now found assumes that there is a dipole excited, and that there are no loss processes due to the dielectric nature of the spheres. First, we must find the rate of excitation caused by inelastic tunnelling, which like the elastic tunnelling in section 2.1 is approximated as one dimensional with an average barrier.

**Figure 3.3. Inelastic energy diagram.** Electrons lose an energy $\hbar\omega_{\text{mode}}$ when tunnelling across the gap. $E_L$ and $E_R$ are the energies of a general pair of filled (starting) and empty (final) states. Fermi energies $E_{FL}$ and $E_{FR}$ are marked for left and right respectively.
The matrix element for electron tunnelling (considered to be from tip to sample for this discussion, but it may be chosen to be in the opposite direction), with the plasmon as the perturbation, from perturbation theory is

\[ M = \int dz \chi_L^* \psi_n(z) \chi_R \quad (3.32) \]

with the electron wavefunctions as those in (2.1) and the plasmon potential coming from (3.6). Here, tunnelling is considered to occur only due to overlapping wavefunctions in the gap, not the electrodes (spheres), which is found in practice to be an acceptable approximation.

\[ M = C^* C' \exp(\alpha_R z_1 - \alpha_L z_2) \]
\[ \times \int_{z_1}^{z_2} dz \exp(\alpha_L - \alpha_R)z \sum_n \left( A_n \exp\left(-\left(n + \frac{1}{2}\right)\mu + B_n \exp\left(n + \frac{1}{2}\right)\mu\right) P_n(\cos \eta) \right) \]

\[ M = C^* C' \exp(\alpha_R z_1 - \alpha_L z_2) \]
\[ \times \int_{z_1}^{z_2} dz \exp(\alpha_L - \alpha_R)z \sqrt{2} \sum_n (-1)^n \left( A_n \exp\left(-\left(n + \frac{1}{2}\right)^2\right) + B_n \exp\left(n + \frac{1}{2}\right)^2\right) \]

...if we approximate \( \eta = \pi \), and \( \mu = \frac{2z}{a} \) – i.e. tunnelling takes place at the very bottom of the top sphere, and as it does so over a few angstroms this is a fair approximation. The tunnelling gap must be much smaller than the radius of either sphere – also true. \( C \) and \( C' \) are normalisation constants.

Figure 3.4. The lateral size of the tunnelling junction is small compared to the radius of the sphere, so for low \( n \) (the most important components) the approximation of \( \eta = \pi \) should be valid.
\[ M = \sqrt{2} C^* C' \exp(\alpha_R z_1 - \alpha_L z_2) \sum_n (-1)^n \left[ A_n \int_{z_1}^{z_2} dz \exp(\alpha_L - \alpha_R - \frac{2n+1}{a}) z \right. \\
+ B_n \int_{z_1}^{z_2} dz \exp(\alpha_L - \alpha_R + \frac{2n+1}{a}) z \left. \right] \]  
(3.35)

\[ M = \sqrt{2} \sqrt{\frac{4E_L E_R}{D_L D_R U_L U_R}} \exp(\alpha_R z_1 - \alpha_L z_2) \]

\[ \times \sum_n (-1)^n \left[ A_n \left[ \exp(\alpha_L - \alpha_R - \frac{2n+1}{a}) z_1 - \exp(\alpha_L - \alpha_R - \frac{2n+1}{a}) z_2 \right] \\
- \exp(\alpha_L - \alpha_R + \frac{2n+1}{a}) z_2 - \exp(\alpha_L - \alpha_R + \frac{2n+1}{a}) z_1 \right] \left. \right] \]  
(3.36)

by inserting \( C^* C' \) from the normalisation of the wavefunctions to the size of the electrode, \( D \).

The rate of inelastic tunnelling from Fermi’s Golden Rule will then be

\[ \nu_{\text{inel}} = \sum_L \sum_R \frac{2\pi}{\hbar} |M|^2 \delta(E_L - E_R - \hbar\omega) f_L (1 - f_R) \]  
(3.37)

where \( f_L \) and \( f_R \) are the Fermi factors (=1 when an energy level is occupied – not corrected for thermal broadening) and the factor of 2 is due to including both possible spin states.

\[ \nu_{\text{inel}} = \sum_{k_L} \sum_{k_R} \frac{4\pi}{\hbar^2} |M|^2 \delta_{k_L k_R} \delta(E_L - E_R - \hbar\omega) f_L (1 - f_R) \]  
(3.38)

ensuring momentum conservation in the ‘parallel’ direction (normal to \( z \)-tunnelling), and energy conservation by creating a photon of energy \( \hbar\omega \). Converting the sum to an integral yields

\[ \nu_{\text{inel}} = \frac{4\pi e^2}{\hbar} \int_{0}^{\infty} dk || \frac{A}{4\pi^2} \int_{0}^{2\pi} d\phi \int dE_L \int_{0}^{D_L} dk_L \int_{0}^{D_R} dk_R \frac{|M|^2 \delta(E_L - E_R - \hbar\omega) f_L (1 - f_R)}{\pi} \]  
(3.39)

\[ \nu_{\text{inel}} = \frac{A e^2 D_L D_R}{\pi^2 \hbar} \int_{0}^{\infty} dk || \int_{0}^{2\pi} d\phi \int_{0}^{h^2/2k_L} dE_L \int_{0}^{h^2/2k_R} dE_R \frac{|M|^2 \delta(E_L - E_R - \hbar\omega) f_L (1 - f_R)}{\hbar^2 k_L \hbar^2 k_R} \]  
(3.40)
as $E = \frac{\hbar^2 k^2}{2m}$. Then,

$$v_{\text{inel}} = \frac{\Lambda e^2 D_L D_R m}{\pi^2 \hbar^3} \int_0^\infty \frac{dE_L}{E_L + \hbar \omega - eV} \left[\frac{dE_L}{\sqrt{E_L E_R}}|M|^2 \frac{1}{\left(\frac{1}{\alpha_L} + \frac{1}{\alpha_R}\right)}\right]$$

(3.41)

$$v_{\text{inel}} = \frac{\Lambda e^2 D_L D_R m}{\pi^2 \hbar^3} \int_{E_{FL}}^{E_{FL} + \hbar \omega - eV} \frac{dE_L}{\sqrt{E_L E_R}} |M|^2 \frac{1}{\left(\frac{1}{\alpha_L} + \frac{1}{\alpha_R}\right)}$$

(3.42)

$$v_{\text{inel}} = \frac{8\Lambda e^2 m}{U_L U_R \pi^2 \hbar^3} \int_{E_{FL} + \hbar \omega - eV}^{E_{FL}} \frac{dE_L}{\sqrt{E_L E_R}} \exp\left(\alpha_R z_1 - \alpha_L z_2\right) \times$$

$$\sum_n (-1)^n A_n \left[ \begin{array}{c}
\exp\left(\alpha_L - \alpha_R\right) - \left(\frac{2n + 1}{a}\right)z_1 - \exp\left(\alpha_L - \alpha_R\right) - \left(\frac{2n + 1}{a}\right)z_2 \\
\alpha_R - \alpha_L + \left(\frac{2n + 1}{a}\right)
\end{array} \right]$$

$$\begin{bmatrix}
\exp\left(\alpha_L - \alpha_R\right) + \left(\frac{2n + 1}{a}\right)z_2 - \exp\left(\alpha_L - \alpha_R\right) + \left(\frac{2n + 1}{a}\right)z_1
\end{bmatrix}^2 $$

(3.43)

where $\alpha_L = \frac{\sqrt{2m(U_L - E_{FL})}}{\hbar}$. The integral is performed numerically.

So we now have a value for the rate of inelastic tunnelling for a given mode.

### 3.5 Dielectric loss rate

In order to consider a steady-state in which there is excitation by tunnelling, and deexcitation by both radiation and loss processes, we must evaluate a loss rate which relies on there being an imaginary part of the dielectric function.

The power lost per unit volume from the oscillation,

$$\text{Power} = -\frac{1}{2} \text{Re} \left\{ \mathbf{E}^\ast \cdot \mathbf{E} \right\}$$

(3.44)
where $\mathbf{D} = -(i\omega)\varepsilon_r \varepsilon_0 E$ and $\varepsilon_r = \varepsilon_{\text{real}} + i\varepsilon_{\text{im}}$

$$\text{Power} = \frac{1}{2} \omega \varepsilon_0 \varepsilon_{\text{im}} |E|^2$$  \hspace{1cm} (3.45)

Integrating this power density over a volume gives

$$\text{Power} = \frac{1}{2} \omega \varepsilon_0 \varepsilon_{\text{im}} \int \mathbf{E}^* \cdot \mathbf{E} \, d\text{Vol}$$  \hspace{1cm} (3.46)

$$\text{Power} = \frac{1}{2} \omega \varepsilon_0 \varepsilon_{\text{im}} \left[ \varepsilon_{\text{im}}^{(1)} \int_{s_1} \frac{\partial \psi_1^*}{\partial n} \psi_1 \, ds_1 + \varepsilon_{\text{im}}^{(2)} \int_{s_2} \frac{\partial \psi_3^*}{\partial n} \psi_3 \, ds_2 \right]$$  \hspace{1cm} (3.47)

where $\varepsilon_{\text{im}}^{(1)}$ and $\varepsilon_{\text{im}}^{(2)}$ are the imaginary parts of the dielectric function for top and bottom spheres respectively. $\psi_1$ and $\psi_3$ come from (3.5) and (3.7).

$$= \frac{1}{2} \omega \varepsilon_0 \left[ \varepsilon_{\text{im}}^{(1)} \int_{s_1} \sum_n \frac{\sinh \mu_1}{2 \cosh \mu_1 - \cos \eta} \left( n + \frac{1}{2} \right) \sqrt{\cosh \mu_1 - \cos \eta} \left[ A_n \exp - (n + \frac{1}{2}) \mu_1 P_n(\cos \eta) \right] \right]$$

$$\times \frac{\sinh \mu_1}{2 \cosh \mu_1 - \cos \eta} \sum_m A_m \exp - (m + \frac{1}{2}) \mu_1 P_m(\cos \eta) \, ds_1$$

$$+ \varepsilon_{\text{im}}^{(2)} \int_{s_2} \sum_n \frac{\sinh \mu_2}{2 \cosh \mu_2 - \cos \eta} \left( n + \frac{1}{2} \right) \sqrt{\cosh \mu_2 - \cos \eta} \left[ B_n \exp (n + \frac{1}{2}) \mu_2 P_n(\cos \eta) \right]$$

$$\times \frac{\sinh \mu_2}{2 \cosh \mu_2 - \cos \eta} \sum_m B_m \exp (m + \frac{1}{2}) \mu_2 P_m(\cos \eta) \, ds_2$$

$$= \frac{1}{2} \omega \varepsilon_0 \left[ \varepsilon_{\text{im}}^{(1)} \int_{0}^{2\pi} \int_{0}^{\pi} \frac{a^2 \sin \eta}{(\cosh \mu_1 - \cos \eta)^2} \sum_n \left[ (2n + 1) \cosh \mu_1 - \cos \eta - \sinh \mu_1 \right] ight]$$

$$\times \left[ A_n \exp - (n + \frac{1}{2}) \mu_1 P_n(\cos \eta) \right] \left( \cosh \mu_1 - \cos \eta \right) \sum_m A_m \exp - (m + \frac{1}{2}) \mu_1 P_m(\cos \eta)$$

$$+ \varepsilon_{\text{im}}^{(2)} \int_{0}^{2\pi} \int_{0}^{\pi} \frac{a^2 \sin \eta}{(\cosh \mu_2 - \cos \eta)^2} \sum_n \left[ (2n + 1) \cosh \mu_2 - \cos \eta + \sinh \mu_2 \right]$$

$$\times \left[ B_n \exp (n + \frac{1}{2}) \mu_2 P_n(\cos \eta) \right] \left( \cosh \mu_2 - \cos \eta \right) \sum_m B_m \exp (m + \frac{1}{2}) \mu_2 P_m(\cos \eta)$$

$$= \frac{1}{2} \omega \varepsilon_0 \left[ \varepsilon_{\text{im}}^{(1)} \int_{0}^{2\pi} \int_{0}^{\pi} \frac{a^2 \sin \eta}{(\cosh \mu_1 - \cos \eta)^2} \sum_n \left[ (2n + 1) \cosh \mu_1 - \cos \eta - \sinh \mu_1 \right] ight]$$

$$\times \left[ A_n \exp - (n + \frac{1}{2}) \mu_1 P_n(\cos \eta) \right] \left( \cosh \mu_1 - \cos \eta \right) \sum_m A_m \exp - (m + \frac{1}{2}) \mu_1 P_m(\cos \eta)$$

$$+ \varepsilon_{\text{im}}^{(2)} \int_{0}^{2\pi} \int_{0}^{\pi} \frac{a^2 \sin \eta}{(\cosh \mu_2 - \cos \eta)^2} \sum_n \left[ (2n + 1) \cosh \mu_2 - \cos \eta + \sinh \mu_2 \right]$$

$$\times \left[ B_n \exp (n + \frac{1}{2}) \mu_2 P_n(\cos \eta) \right] \left( \cosh \mu_2 - \cos \eta \right) \sum_m B_m \exp (m + \frac{1}{2}) \mu_2 P_m(\cos \eta)$$

(3.49)
when \( ds \) is converted to \( d\phi \ d\eta \ h_\phi \ h_\eta \). Both the \( \varepsilon^{(1)}_{im} \) and \( \varepsilon^{(2)}_{im} \) parts in (3.49) have two integrals:

the first,

\[
\int_{0}^{\pi} d\eta \ P_n(\cos \eta) P_m(\cos \eta) \sin \eta \ \text{which} \ = \ \frac{2}{2n+1} \delta_{nm} \quad (3.50)
\]

and the second,

\[
\int_{0}^{\pi} d\eta \ P_n(\cos \eta) P_m(\cos \eta) \frac{\sin \eta}{\cosh \mu_x - \cos \eta} =
\]

\[
\int_{0}^{\pi} d\eta \ \sin \eta \ P_n(\cos \eta) P_m(\cos \eta) \left[ \sum_{l} (2l+1)P_l(\cos \eta)Q_l(\cosh \mu_x) \right] \quad (3.51)
\]

where the polynomial \( Q_0(x) = \frac{1}{2} \frac{(x+1)}{(x-1)} \) and \( n Q_n = (2n-1) x Q_{n-1} - (n-1)Q_{n-2} \ . (3.52) \)

So the overall power loss is

\[
\begin{align*}
\frac{a\omega \ v_{\text{die}} \pi}{2} & \left[ \varepsilon^{(1)}_{im} \left\{ \sum_{n} (2A_n')^2 \exp - (2n+1)\mu_1 - \sinh \mu_1 \sum_{n} A_n' \exp - (n+\frac{1}{2})\mu_1 \sum_{m} A_m' \exp - (m+\frac{1}{2})\mu_1 \\
& \sum_{l} (2l+1)Q_l(\cosh \mu_1) \int_{0}^{\pi} d\eta \ P_l(\cos \eta) P_n(\cos \eta) P_m(\cos \eta) \sin \eta \right\} \\
+ \varepsilon^{(2)}_{im} \left\{ \sum_{n} (2B_n')^2 \exp (2n+1)\mu_2 + \sinh \mu_2 \sum_{n} B_n' \exp (n+\frac{1}{2})\mu_2 \sum_{m} B_m' \exp (m+\frac{1}{2})\mu_2 \\
& \sum_{l} (2l+1)Q_l(\cosh \mu_2) \int_{0}^{\pi} d\eta \ P_l(\cos \eta) P_n(\cos \eta) P_m(\cos \eta) \sin \eta \right\} \right] \\
\end{align*}
\]

(3.53)

The integral \( \int_{0}^{\pi} d\eta \ P_1(\cos \eta) P_n(\cos \eta) P_m(\cos \eta) \sin \eta \)

is a triple sum of combinations of \( n, m \) and \( l \) - just a numerical integration.

The rate of dielectric loss, \( v_{\text{die}} \) is then just the power loss divided by the energy of the mode, \( h\omega_{\text{mode}} \).
3.6 Rate equation

In order to obtain a steady state we must construct a rate equation balancing excitation, radiation and dielectric loss.

\[(1 + n_{occ})v_{inel} = n_{occ}v_{rad} + n_{occ}v_{diel}\]  \hspace{1cm} (3.54)

or \[\Omega_{inel} = \Omega_{rad} + \Omega_{diel}\]  \hspace{1cm} (3.55)

where \(n_{occ}\) is an averaged occupation number for the plasmons. The enhancement factor for bosons (i.e. the plasmons) is present on the left hand side of (3.54). \(n_{occ}\) is found to be much smaller than 1 here.

So the rate of inelastic events will be \((1 + n_{occ})v_{inel}\), that of radiation will be \(n_{occ}v_{rad}\)

and rearranging (3.54) gives \(n_{occ} = \frac{v_{inel}}{v_{rad} + v_{diel} - v_{inel}}\). \hspace{1cm} (3.56)

So the rate of inelastic events, \(\Omega_{inel}\) is \[\frac{v_{inel} (v_{rad} + v_{diel})}{v_{rad} + v_{diel} - v_{inel}}\] \hspace{1cm} (3.57)

and the rate of photon emission events, \(\Omega_{rad}\) is \[\frac{v_{inel} v_{rad}}{v_{rad} + v_{diel} - v_{inel}}\]. \hspace{1cm} (3.58)

All these rates are for separate modes. The total rates can be found simply by adding the rates for all modes found, as they are independent.

Due to the uncertainty principle, each mode can radiate with a different frequency to that of the plasmon oscillation. Experimental results show broadened modes – see for example figure 1.4 – and often they are so broadened they are indistinguishable as separate peaks. The finite lifetime originates from the presence of the dielectric loss and radiative processes which damp the oscillation. So the width of the Lorentzian,

\[\Delta E_{mode} = \hbar v_{diel} + \hbar v_{rad}\] \hspace{1cm} (3.59)

The response away from the eigenfrequency means that the rate equation (3.54) should be replaced by

\[g(\omega - \omega_{mode})(1 + n_{occ})v_{inel}(\omega) = g(\omega - \omega_{mode})n_{occ}v_{rad}(\omega) + g(\omega - \omega_{mode})n_{occ}v_{diel}(\omega)\] \hspace{1cm} (3.60)
where $g$ is a Lorentzian with a width of $\Delta E_{\text{mode}}$ (fwhm). All the rates, $\nu_{\text{inel}}, \nu_{\text{rad}}$ and $\nu_{\text{diel}}$ must be evaluated at all frequencies, $\omega$.

Figure 3.5. Schematic spectrum. The broadening of each mode, $\Delta E$, is shown. The total spectrum will be the sum of the broadened individual modes.

An improvement on the Lorentzian broadening can be made by using an energy-dependent or 'variable' broadening. This is necessary because of the extremely high damping from the dielectric processes. The normal Lorentzian lineshape is just the power spectrum of a damped oscillator,

$$x(t) = x(0)e^{-\gamma t/2}e^{i\omega_0 t}. \quad (3.61)$$

The ratio of off-peak response to that at the eigenfrequency,

$$g = \frac{I(\omega)}{I(\omega_0)} = \left| \frac{\tilde{x}(\omega)}{\tilde{x}(\omega_0)} \right|^2 = \frac{1}{1 + (2(\omega - \omega_0)/\gamma)^2} \quad (3.62)$$

when the damping factor, $\gamma$, is a constant. If this factor is a function of frequency the expression becomes

$$g = \frac{I(\omega)}{I(\omega_0)} = \left| \frac{\tilde{x}(\omega)}{\tilde{x}(\omega_0)} \right|^2 = \frac{\gamma^2 / 4}{\gamma^2 / 4 + (\omega - \omega_0)^2} \quad (3.63)$$

as $\tilde{x}(\omega) = \int x(t) \, dt \, e^{-i\omega t} = x(0) \int_0^\infty e^{(i\gamma/2 + i\omega_0 - i\omega) t} \, dt = \frac{x(0)}{\gamma / 2 + i(\omega - \omega_0)}. \quad (3.64)$
Converting frequency to energy,\[ g = \frac{[\Delta E(E_0)]^2}{[\Delta E(E)]^2 + 4[E - E_0]^2} \]  
(3.65)

where $\Delta E$ is the fwhm linewidth from (3.59) which is a function of energy. This dependence in the model is apparent from (3.58) - both terms are functions of energy due to the frequency-dependent dielectric functions.

Figure 3.6. Schematic spectrum. Each mode now has 'variable' broadening, and is pinned to zero when $\varepsilon = -2$. Above this energy, then photon intensity is taken to be zero.

A bulk plasmon occurs at a frequency (or energy) where (the real part of) $\varepsilon = 0$, and a surface plasmon when $\varepsilon = -1$. A plasmon on a single sphere occurs at $\varepsilon = -2$, so we can expect the spheres to be decoupled for energies above that at which $\varepsilon = -2$. This is confirmed by the drop in the response function of Johansson [1990]. However, because this use of broadening would still give a response above this energy, it is necessary to force the spectra to zero at and above this energy. This is achieved by the use of an image peak symmetrically above the decoupling energy, as shown in figure 3.6. Both this modification, and the broadening
approximations effectively make $A_n$ and $B_n$ frequency dependent, as they are in Johansson’s model.

### 3.7 Dielectric functions of small particles

Measurements of dielectric functions for metals are performed in high vacuum or better on flat films [Johnson; Hagemann], however, the dielectric properties of nanometre-sized particles are not the same as those of bulk materials [Perenboom; Schimmel]. So it is important to investigate size effects because the dielectric functions are critical to the form of the spectra.

The dielectric function, $\varepsilon(\omega)$, relates the internal field ($D$) in the metal to the external field ($E$). The real part of the dielectric function, $\varepsilon_{\text{real}} = 1 + \chi$, where $\chi$ is the polarizability of the medium originating from the shielding effect of the electrons. The imaginary part originates from the time constant of the electrons contributing to the shielding (at least, at optical frequencies) – a phase difference between the $E$ and $D$ fields causes damping. Another contribution to the imaginary part comes from interband transitions. For gold these occur above 2.4eV, for copper 2.1eV and for silver about 4eV. Electronic transitions are excited by photons or plasmons with energy above that of the gap, and the electrons relax back down to their ground states predominately via non-radiative routes.

In the free-electron region – below where the interband transitions occur – the real and imaginary parts of the dielectric functions follow the Drude theory for the noble metals under consideration: gold, silver and copper. Here,

$$\varepsilon_{\text{real}} = 1 - \frac{\omega_p^2 / \omega^2}{1 + 1 / \omega^2 \tau^2} \approx 1 - \frac{\omega_p^2}{\omega^2}$$

and

$$\varepsilon_{\text{im}} = \frac{\omega_p^2}{\omega^2} \left( \frac{1}{\omega^2 \tau^2} \right) \approx \frac{\omega_p^2}{\omega^3 \tau}$$

as $\omega \tau \gg 1$

where $\tau$ is the relaxation time (of the order of $10^{-14}$ s). Figure 3.7 shows experimental data for silver.

So there exists a energy ‘window’ for each of these three metals where $\varepsilon_{\text{im}}$ is lowest which will restrict the spectra because the damping rate is proportional to $\varepsilon_{\text{im}}$ in (3.53); and the photon emission rate is inversely proportional to the damping rate, $\nu_{\text{dil}}$, as it dominates the denominator of (3.58).
When considering small particles, the contribution to the dielectric function from interband transitions is similar to that for the bulk. The free-electron contribution, however, is altered due to the presence of the surface – the mean free path will be limited to $d/2$. So a surface limited scattering time must be substituted into (3.66):

$$\frac{1}{\tau} = \frac{1}{\tau_0} + \frac{2v_F}{d}$$  \hspace{1cm} (3.67)

where $\tau_0$ is the scattering time in the bulk, and $v_F$ is the Fermi velocity.

The contribution from surface scattering becomes important in the range below 100nm for these noble metals. Unless $d$ becomes very small, $\varepsilon_{\text{real}}$ will not be measurably affected, however this is not true for $\varepsilon_{\text{im}}$:

$$\varepsilon_{\text{im}} \approx \frac{\omega_p^2}{\omega} \left( \frac{1}{\tau_0} + \frac{2v_F}{d} \right).$$ \hspace{1cm} (3.68)

So smaller particles should give rise to broader peaks in the spectra, as the broadening is proportional to $\varepsilon_{\text{im}}$ in (3.53), although the positions of the modes will not change as $\varepsilon_{\text{real}}$ doesn’t change.
When comparing figure 3.7 with spectra – for example figure 1.4c, it can be deduced that the lowest energy peak of 2eV corresponds to $\varepsilon_{\text{real}} \approx -10$. This is a much lower (larger negative) figure than for plasmons occuring on isolated spheres and as most metals have $\varepsilon_{\text{real}}$ in the range -2 to -10 within the optical range, it would be fair to assume that photon emission would occur for most metals. The only problem would seem to be that $\varepsilon_{\text{imag}}$ is usually large at these frequencies, implying that emission would be reduced compared to the free-electron metals and the peaks would have a width of many eV meaning that the shape of the spectra would be governed almost entirely by $\varepsilon_{\text{imag}}$. This is confirmed by Berndt et al. [1992] who observed mainly featureless spectra for Ti and Fe with light intensities lower than for noble metals by a factor of 20.

It was considered in the previous section that we can expect the spheres to be decoupled for energies above that at which $\varepsilon = -2$. However, if the particles are non-spherical this figure will be different. For prolate spheroids this decoupling will occur below -2, and for oblate spheroids between -1 and -2. So the cut-off in spectra is expected to be at higher energies (shorter wavelengths) for flat grains or blunt tips and at lower energies for sharp tips or high aspect structures.

### 3.8 Comparison with existing sphere-plane model

Johansson’s model [1990, 1991], briefly described in section 2.2, should be returned to so that it may be may compared to the model in this chapter. It starts with the reciprocity theorem of electrodynamics, to relate the electric field some place far away from the tip to the inelastic tunnel current under the tip. It achieves this by swapping the ‘detector’ (far from the tip) and ‘source’ (under the tip) in the same way that the radiation pattern of an antenna can be swapped for the response of a receiver. Like an antenna, an enhancement in the field occurs under the tip because of the favourable geometry and materials. This response function, or enhancement factor, is solved from potentials of the same form as (3.5), (3.6) and (3.7), and the recurrence relations for the sphere-sphere case in (3.13) and (3.14) effectively reduce to those quoted by Johansson for the sphere-plane case by substituting $\mu_2 = 0$.

The co-efficients corresponding to $A_n$ and $B_n$ can be calculated at any energy in Johansson’s model. It was quoted that at lower energies, $A_0$ and $B_0$ were the dominant ‘modes’ and at higher energies higher numbered ‘modes’ dominated, although in the model presented in this chapter these ‘modes’ are merely components of each mode. A similar finding is also made.
for this model, that for the lowest energy mode the dominant components are $A_0$ and $B_0$, and for example the fourth lowest energy mode will be dominated by $A_3$ and $B_3$.

The inelastic current matrix is calculated for a trapezoidal barrier by Johansson, however the inelastic matrix uses a response function which is constant throughout the gap, whereas in reality it often inverts. For the new model, this perturbation is taken as varying in the gap and an approximate contribution from the electrodes is also included in the matrix.

In Johansson’s model, at any frequency the electric field and subsequently the photon emission are calculated at the detector. In the new two-sphere model, modes at a variety of frequencies are responsible for the photon emission, and their potentials and energies are calculated explicitly. Other substantial advantages of this model are its two-sphere capability, and the separate calculation of the three rates of excitation and de-excitation. Also, all these rates and other physical parameters such as dipoles and charge densities can be evaluated for each mode and each sphere (or plane) if required. Furthermore, the explicit nature of the calculations involving the potentials should be easier to understand than the use of the somewhat mysterious enhancement factor, which lumps together all the geometric and material properties.

### 3.9 Summary

Bispherical co-ordinates have been used here to describe the geometry found in the STM: either a two-sphere arrangement, where a curved surface or particle lying on a flat surface is represented by the bottom sphere and the tip represents the top sphere; or a sphere-plane arrangement where the tip is treated as a sphere, and the sample as a plane. The use of Laplace’s equation and electrostatic boundary conditions give solutions for the plasmon potential in these co-ordinates, and from this potential a dipole radiation rate, an inelastic tunnelling rate and a dielectric loss rate are all calculated.

These rates are all for a specific mode, which constitutes a solution at a specific frequency. For each mode, these rates must be balanced in a steady-state rate equation, from which the inelastic current and the photon emission rate are calculated. The spectral broadening of these peaks arises from the decay processes of the plasmon.
Chapter 4  
Apparatus

4.1 Introduction

STM Photon Emission can be performed in air [Sivel; Smolyaninov], however, the data in an Ultra-High Vacuum (UHV) environment is much more reliable. The reasons for this are as follows: Firstly, although STMs can be used in air to take topographic data which is accurate, it is probably not the case that true tunnelling is occurring, as a meniscus of water is present underneath the tip as it scans over the sample. In air, surfaces are generally covered by a ~5Å film of water [Hu], unless they are hydrophobic. Secondly, surfaces which are not prepared in vacuum will usually oxidise to a certain extent, leaving a possible metal-oxide-water-oxide-metal barrier – not an ideal barrier to study and model photon emission. And because even at $10^{-9}$ mbar, one monolayer of air molecules impacts onto the surface every 20 minutes, in air the oxidation or contamination will be a factor of $10^{12}$ worse. Oxides and contaminants will damp out the plasmons. Thirdly, the stability in air is much different to that in UHV – it is more difficult to scan at the higher voltages required on metals (around 3V) in air, because of tip rearrangement and other effects occurring due to higher electric fields, inducing surface oxidation or modification, and ionisation of contaminants in the gap.

As a result of non-vacuum tunnelling, the voltage-distance characteristics of the junction in air are much different to those in UHV. Whereas in UHV, a graph of light intensity against bias voltage will show little structure, in air the same graph will show discernible peaks [Smolyaninov]. So it is best to consider the more ideal case involving two clean surfaces before modifying this study to apply to Photon Emission in air.

4.2 Chamber design

Any UHV chamber must have a series of pumps to bring the pressure inside the chamber from atmospheric pressure ($10^{+3}$ mbar) to below $10^{-10}$ mbar. In the chamber used for the experiments in this thesis, the pumps were located in two areas: the ion pump [#13] and Titanium sublimation pump [#34] below the “scanning” chamber – in the top half of figure 4.1 – and the turbomolecular-rotary combination [#12] below the “sample preparation” chamber, separated by a valve [#7].
Figure 4.1. Diagram of UHV system. Actual size of table: 6 feet x 4 feet (1.8m x 1.2m). For a key to numbers, see next page.
### Chapter 4: Apparatus

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gas bottles (99.999% pure Argon and 99.99% pure Neon)</td>
</tr>
<tr>
<td>2</td>
<td>Magnetic linear arm for transfer from loadlock to sample preparation chamber</td>
</tr>
<tr>
<td>3</td>
<td>Magnetic linear arm for transfer from sample preparation chamber to scan chamber carousel</td>
</tr>
<tr>
<td>4</td>
<td>Magnetic linear arm for transfer from carousel to STM</td>
</tr>
<tr>
<td>5</td>
<td>Gate valve between loadlock and sample preparation chamber</td>
</tr>
<tr>
<td>6</td>
<td>Gate valve between sample preparation chamber and scan chamber</td>
</tr>
<tr>
<td>7</td>
<td>Gate valve between sample preparation chamber and turbomolecular pump</td>
</tr>
<tr>
<td>8</td>
<td>Fine control leak valve (controllable leak rates down to $\approx 10^{-11}$ mbar s$^{-1}$; may be opened fully)</td>
</tr>
<tr>
<td>9</td>
<td>Admittance valve to nitrogen or air</td>
</tr>
<tr>
<td>10</td>
<td>Valve to isolate gas bottles (normally closed; open for roughing gas lines)</td>
</tr>
<tr>
<td>11</td>
<td>Elbow valve to turbomolecular pump</td>
</tr>
<tr>
<td>12</td>
<td>Turbomolecular pump (pressure range $\approx 10^{-1}$ to $10^{-9}$ mbar) backed by rotary pump</td>
</tr>
<tr>
<td>13</td>
<td>Ion pump (pressure range $\approx 10^{-5}$ to $10^{-10}$ mbar)</td>
</tr>
<tr>
<td>14</td>
<td>Viton sealed quick-access sample entry door</td>
</tr>
<tr>
<td>15</td>
<td>Pirani gauge (pressure range $10^{-13}$ to $10^{-4}$ mbar)</td>
</tr>
<tr>
<td>16</td>
<td>Ion gauge (pressure range $10^{-3}$ to $10^{-10}$ mbar)</td>
</tr>
<tr>
<td>17</td>
<td>Quadrupole Mass Spectrometer (pressure range $10^{-4}$ to $10^{-13}$ mbar; mass range 1 to 120)</td>
</tr>
<tr>
<td>18</td>
<td>3-axis micrometer sample heating stage</td>
</tr>
<tr>
<td>19</td>
<td>6-position carousel</td>
</tr>
<tr>
<td>20</td>
<td>Sputter ion gun directed up towards sample. 0.5kV to 2.0kV energy; up to 30mA emission current; focus from $\approx 2$ to 5mm</td>
</tr>
<tr>
<td>21</td>
<td>Micro-evaporator directed towards sample. Consists of two alumina baskets with tungsten wire windings (for temperatures up to $&gt;1100^\circ$C) and one molybdenum boat for metal evaporation (for temperatures up to $&gt;1200^\circ$C)</td>
</tr>
<tr>
<td>22</td>
<td>Electrical feedthrough: two wires for thermal heating of samples on stage 18, and one for the thickness monitor, also located on stage 18</td>
</tr>
<tr>
<td>23</td>
<td>Electrical feedthrough for thermal heating of samples docked in position on the carousel (19)</td>
</tr>
<tr>
<td>24-28</td>
<td>Viewports</td>
</tr>
<tr>
<td>29,30</td>
<td>Alignment bellows for angular positioning of linear transfer arms</td>
</tr>
<tr>
<td>31</td>
<td>Magnetic linear and angular motion drive fitted with hex key for locking parts of the STM</td>
</tr>
<tr>
<td>32</td>
<td>Rotary wobble-stick for removal of tip and sample from STM</td>
</tr>
<tr>
<td>33</td>
<td>STM flange</td>
</tr>
<tr>
<td>34</td>
<td>Titanium sublimation pump (pressure range $\approx 10^{-6}$ to $10^{-11}$ mbar)</td>
</tr>
</tbody>
</table>
These two sub-chambers may be isolated from each other by a valve [#6] so that the best pressure may be achieved in the scanning chamber, which has fewest parts, and the sample preparation chamber may be operated at slightly inferior pressures and may be isolated from the scanning chamber when performing contaminating preparations. A third sub-chamber, the “loadlock” is present, separated from the sample preparation chamber by a valve [#5] and to the turbomolecular/rotary pumps by a controllable elbow valve [#11].

In addition, there ought to be pressure gauges capable of reading down through the whole range of pressures encountered. In this system, there are three gauges - the pirani gauge [#15] for high pressures in the loadlock (10^{-3} to 10^{-4} mbar), the ion gauge [#16] in the scanning chamber (10^{-3} to 10^{-10} mbar) and the mass spectrometer [#17] in the preparation chamber for leak testing and analysis of vacuum (10^{-4} to 10^{-13} mbar).

For resistive heating of samples (the sample holder is a two-part design with isolated clips on either end of the sample), there should be an electrical heating stage. In this system there are two: the first is on the heating stage [#18] in the preparation chamber, and the second is on one position of the carousel. This gives flexibility in sample preparation. Some surfaces are best evaporated onto whilst heated, so this can be done in the preparation chamber. For samples which will be heated, it is essential that they are degassed first. This liberates high pressures of contaminants, and is best performed in the preparation chamber. For silicon processing, e.g. the formation of Si(111)-7x7, the sample will need to be heated to high temperatures (over 1200°C).
while at a low pressure (in the $10^{-10}$ mbar range). This is best performed in the scanning chamber.

In order to keep a number of samples or tips in the chamber at any time, there should be a carousel. This system has a six-position carousel [#19] which is rotated by the pushing the pins with the transfer arm.

To move the sample from the entry point to the STM and on and off the carousel, there must be some sort of in-vacuum transfer facility. For large amounts of travel, this must be done by using magnetic linear motion arms [#2, 3, 4]. The sample is moved from the entry door [#14] onto the heater stage [#18], then it is taken off by the long arm [#3] and put onto the carousel [#19] where it is pushed around and taken off by the short arm [#4] and brought towards the STM [#33]. From there, it is taken off and put onto the STM by the rotary wobble-stick [#32].

To deposit molecules and metals onto surfaces, there must be an evaporation source. Here [#21], this is a home-made tungsten wire dual basket design for metals or molecules with a thermocouple, and a separate molybdenum boat for metals. In order to judge how thick a layer has been deposited, a quartz crystal thickness monitor with a resolution of 0.1Å is fixed onto the heater stage [#18] only 10mm from the centre of the sample.

To clean surfaces in vacuum, a sputter ion gun and a gas inlet for Argon or Neon is effective. Here, a sputter ion gun with energies up to 2kV is used in conjunction with a gas admittance system. The gas bottles [#1] are high quality Neon and Argon, and after the lines have been well pumped by the turbomolecular pump with valve #10 open, around $10^{-5}$ mbar is let into the preparation chamber by the high precision valve [#8]. The amount of sputtering can either be deduced from tables, or directly from the thickness monitor.

A sample can be pumped down in the loadlock to acceptable pressures in about 5 minutes, and the turbomolecular pump can achieve pressures in the preparation chamber of around $10^{-7}$ mbar. With the ion pump/sublimation pump combination, a base pressure of around $10^{-8}$ mbar is reached about one day after exposure to air. To do better, the system must be baked to around 130°C for 1-2 days to remove contaminants, especially water. Then the pressure should fall below $10^{-10}$ mbar when the system is cooled. All components are bakeable to this temperature (with the exception of the Pirani gauge [#15] which is removed). The molecular species responsible for the pressure in the system can be checked with the mass spectrometer [#17].
4.3 The STM

The STM drawn in figure 4.3 used in UHV was built at IBM Rüschlikon, Switzerland. All probe microscopes need to have coarse movement as well as fine movement. The fine movement here is given by the use of the piezoelectric material, Lead Zirconate Titanate (PZT). When an electric field is applied across the material, it expands or contracts smoothly to picometre resolution. To enable three-axis movement, orthogonal extensions are required – either in some tripod formation, or more easily with the use of segmented tubes (in figure 4.4).

Figure 4.3. Top view of the STM used. Actual size. [Built by Reto Schlittler at IBM Research Lab., Saümerstrasse 4, CH-8803, Rüschlikon, Switzerland]

The STM drawn in figure 4.3 used in UHV was built at IBM Rüschlikon, Switzerland. All probe microscopes need to have coarse movement as well as fine movement. The fine movement here is given by the use of the piezoelectric material, Lead Zirconate Titanate (PZT). When an electric field is applied across the material, it expands or contracts smoothly to picometre resolution. To enable three-axis movement, orthogonal extensions are required – either in some tripod formation, or more easily with the use of segmented tubes (in figure 4.4).

Figure 4.4. Diagram of piezoelectric tube scanners: side view (left) and top view (right).
The sensitivity of the tube (Angstroms per Volt) in the x and y directions is given by

\[ \delta_x = \frac{\sqrt{2} d_{31} L^2}{\pi WD} \]  

(4.1)

and in the z-direction by

\[ \delta_z = \frac{d_{31} L}{W} \]  

(4.2)

where \( d_{31} \) is the piezoelectric charge constant of the tube. For the parameters of this tube, the x and y sensitivities are 100 Å/V and in the z-direction, 25 Å/V. The breakdown strength of the material limits the use of high voltages to ±75V in x, y and z, and so the total accessible scan area is 1.5 microns square and the total extension in the z-direction is 375nm. This enables both atomic resolution (using 16-bit electronics, one bit in the x-y direction is 0.2Å, and in the z direction is 0.06Å – or 0.006Å with an extra gain stage) and the study of nanostructures (of size ≈10-100nm). The sample must be brought within this 375nm range of tip excursion, and this is done by having three tubes in a slip-stick stage. The tubes extend rapidly [slip] to overcome the friction of a sapphire - silicon nitride contact, and pull the stage back slowly when the two are in contact [stick]. This can be seen in figure 4.5.

\[ \]
In this STM there are three tubes each of whose outer electrodes are segmented into two halves. When these two halves are connected together, the motion with a voltage ramp (relative to the central electrode) is in the z-direction; when the halves are separately ramped, the motion is left-right. This coarse left-right movement is essential to view other parts of a sample so as to establish uniformity. All three tubes are connected together, so perform the same motion.

In order to isolate the microscope from vibration, the STM lies on a multilayer of viton tubes and metal plates. The chamber itself is fixed to the aluminium table which in turn is allowed to float on isolating air-filled legs. A good rms noise figure taken with this STM over an image will be around 0.1 Å.

### 4.4 Optical collection

It was discussed in section 1.2 that the low photon production rates from these experiments make it important to collect as high a photon fraction as possible. However, another point to consider is that a point source is required for spectroscopy. The spectrometer (to be added at a later date) accepts a point source of light diverging at up to 14°. It then becomes split into a range of wavelengths, and the larger the size of this “point”, the more blurred the spectrum of wavelengths becomes.

So any system of optical collection involving mirrors or other components which would produce multiple sources of light may not be used. One design would be to bring an optic fibre up close to the tip position (say, a 100micron diameter fibre placed 100microns away). However, this was not used because of difficulty of positioning in UHV – another coarse positioning device would have to be used, and a good optical microscope to see how far away the fibre would be from the surface.

Another idea would be to focus the light onto a fibre in vacuum, which would be glued through the flange. This was rejected for two reasons. Firstly, there would be a need for some positioning of the fibre inside the chamber, as there will be an uncertainty in the position of the source due to changing tips and samples of ≈1mm. Secondly, it was thought that a mechanical link through the chamber was a bad idea – especially a fragile fibre, as it may have to replaced.

The idea that became used was to have a fixed lens system which focusses the light in the chamber down to a point again, outside the chamber. So, a low f/ number lens – f=26.5mm, diameter=30mm – was chosen to collimate the light, so is located one focal length away from the tip. Then a second lens focusses this down to a point several centimetres outside the
chamber. This way, the uncertainty of tip position manifests itself as an uncertainty in the image point outside the chamber. The angular collection on the first lens is about ±(25-35)°, giving about 0.7 steradians or just over 10% of the $2\pi$ steradians being emitted. This compares well with the system used by Gimzewski [1989] with a collection of 0.1 steradians.

Standard singlet lenses have high aberrations for non-centre rays. Spherical aberration results from the convex lens surface being spherical, and not the right shape to focus all axial rays. This is shown in figure 4.6. Spherical aberration for a biconvex lens with $f=30\text{mm}$, diameter $= 26.5\text{mm}$ will produce a spot size (circle of least confusion) of around 5mm. Also, the wavelength-dependent refractive index of the glass will give different focal points for different wavelengths of light, differing generally by about 1-2%. The use of an achromatic lens, consisting of a sandwich of two lenses, removes both of these types of aberrations. However, the lowest $f$/aperture ratio for commercial lenses of this type is about 2, compared to 0.88 for an aspheric lens. An aspheric lens gets rid of the spherical aberration, but does not account for the chromatic aberration. However, the reduction in collection efficiency would be about 80% in order to maintain no chromatic aberration.

![Figure 4.6. Ray diagram for a biconvex lens, showing the poor focusing of marginal rays (those close to the edge of the lens).](image)

The final design in figure 4.7 consists of an aspheric lens ($f=30\text{mm}$, diameter $= 26.5\text{mm}$) to collect (collimate) light from the tip at 30° to the surface. An achromatic lens ($f=84\text{mm}$, diameter=40mm) then focuses this down to a point outside the chamber. The longer focal length of the second lens means that the angle of the cone of light has been reduced from 35° to 12°, although this does give magnification of the spot size to around 1mm.
Unfortunately, due to the space available around the STM, it is necessary to reflect the collimated light before it is refocussed. So a mirrored surface which is a good reflector at 60° from the normal, and between 400-900nm wavelength was required. A bare silver surface is the best reflector, but as this tarnishes easily in air to decrease the reflectivity, a bare aluminium coating was chosen, which only degrades by oxidation in the near Ultra-Violet. The aluminium was evaporated onto optically flat glass in a vacuum of $2 \times 10^{-7}$ mbar to a thickness of 400Å.

![Diagram of internal optical collection](image)

**Figure 4.7.** Diagram of internal optical collection. The two lenses are fitted to holders which can be tilted and have 1-2mm of lateral movement. Scale: approximately actual size.

Another benefit of this lens system is that the reverse process is possible, i.e. light-assisted tunnelling, by focusing a laser spot at the position of the external focus, which will be focused back to the sample. Also, as an image of the end of the tip is produced outside the vacuum with a 3x magnification, the coarse approach of the sample towards the tip can be easily monitored.

Throughout all the debates of which collection system to use, there was an awareness of the range of wavelengths that are most important in Photon Emission. Figure 1.4 clearly shows that it is only necessary to collect photons in the range 400-900nm for those metals. This can be done with ordinary glass (no special ultraviolet-enhanced glasses required) and the reflectance from the aluminium mirror will be very good in this range – evident in figure 4.8. The spectrometer-CCD combination (to be added later) responds to photons in the range 400-900nm, and the photomultiplier tubes to the ranges 300-650nm and 300-750nm.
Figure 4.8. Percentage of light remaining after losses due to (1) internal transmission through all glass components (lenses and viewport), (2) dielectric mismatch from vacuum-lens interfaces with MgF₂ coating and (3) air-viewport and viewport-vacuum interfaces, (4) reflection from bare aluminium mirror. Figures are taken from manufacturers’ catalogues and other literature [Kingslake].

Emitted photons can either be detected by counting numbers of photons or by taking a spectrum of photon energy (wavelength), in a given time period. The main way of collecting data is by counting photons for a given length of time at every pixel (imaging point of the topographic image) to produce a photon map. A calculation for tunnelling onto silver particles would proceed as follows: 10nA tunnelling current is around 6x10¹⁰ electrons per second. About 10⁻⁵ of these would produce photons, and 1% of would be counted (figure 4.9). This means 6,000 counts per second. The dark counts of the photomultipliers used are around 100 cps, so the signal to noise ratio is not a problem. However, a statistically valid map of 256 x 256 pixels ought to have around 100 counts at each pixel. The software only takes data from left to right on each scan line, so a full scan will take 2 x 1/60 sec x 65536 pixels ≈ 10 minutes. This is also not a problem, unless atomic scale maps need to be taken, as the drift for the STM is of the order 10Å/minute.
Figure 4.9. Expected spectral response of two collection systems: (1) Bialkali Photomultiplier tube, (2) S20 Photomultiplier tube. Total collection efficiency means the ratio of counted photons to the number produced. Calculated from figures 4.8 and 4.9, together with an estimated collection efficiency of the first lens of 10%.

4.5 Data acquisition

The STM is controlled via a PC based computer program written in the Engineering Department [Wong]. The computer program interfaces with a Digital Signal Processing (DSP) chip (50Mhz AT&T DSP32C), which outputs the variables to the STM – the tip voltage and x and y scan co-ordinates, and monitors the tunnel current which provides the feedback for the z co-ordinate. So the computer records the z co-ordinate at each pixel (x-y co-ordinate), building up a topography line by line.

The voltages applied to the x, y and z piezos are through 16-bit Digital to Analogue Converters (DACs), giving resolutions quoted in section 4.3 of 0.2Å in x and y directions and down to 0.006Å in the z direction. In practice, these resolution limits are far exceeded by the resolution of tunnelling of at least 3Å in x and y, and mechanical noise in z. The small tunnel
current (typically 0.1-10nA) is first amplified (strictly, converted to a voltage) with a gain of $10^8$ V/A, then this voltage is passed through an Analogue to Digital Converter (ADC) in order to be monitored by the DSP. The sampling frequency, and thus the control bandwidth is 70kHz, which is much higher than the resonant frequency of the piezo tube (around 4 kHz) so controlling against the mechanical instabilities and impulses is limited by the piezo (“scan”) tube, not the electronics. The DSP can also record any other variables simultaneously, by using other ADCs, or in the case of photons, by a separate counter.

When a photon impacts on the cathode of the photomultiplier tube, a cascade of electrons is produced, which is amplified to a small current. These increases in current are sensed by a discriminator, and the final output is a square pulse of 5V. The counter, which is able to decode these pulses, is linked to the DSP and is run at each pixel for a time interval set by the program. The feedback loop still operates in this time interval so as to aid stability. This is the standard mode of data acquisition for photon maps, producing a topographic map and a corresponding photon map and current map (which differs from just the set current because of instabilities).

As was mentioned in Chapter 1, McKinnon [1993] developed a method of data acquisition called PITS (Photon Imaging Tunnelling Spectroscopy) whereby the feedback loop is held at each pixel, and a new tip voltage is used for a period of time while recording a photon count. Further voltages can be used, and a number of photon maps at these voltages can be recorded. There are problems with this method, however. Holding the feedback loop for a period of time long enough to produce a high enough number of counts is detrimental to the stability of the tunnel gap, and using higher tip voltages (>3V) leads to high fields which can cause mechanical instabilities in both the tip and the sample.

To obtain current-voltage and photon-voltage characteristics, the feedback loop is held and the voltage ramped linearly from $V_{\text{start}}$ to $V_{\text{end}}$. Also, the $z$ displacement can be ramped from $z_{\text{start}}$ to $z_{\text{end}}$. The advantage of this method for obtaining photon-voltage curves against that of maintaining the feedback (used by Berndt and Gimzewski) is that the tip-sample separation remains constant – decreasing the separation should cause an increase in the efficiency for photon emission. The disadvantage is that of the mechanical instability of the tip-sample gap adding noise to the data, or drift to the separation. This should not be a problem under stable tunnelling conditions – mechanical noise will average out at each data point, so when a steady thermal equilibrium has been reached the feedback loop can be disabled for several seconds.
Chapter 5
Photon Emission from silver clusters

5.1 Sample preparation

When silver is evaporated onto surfaces it forms polycrystalline films, or if the surface is flat or heated, the growth is more epitaxial - i.e. ordered (lattice-like) and layer by layer. In this study of photon emission, however, we are interested in spherical particles, preferably well separated on a flat surface. As we were unable to produce mass-selected clusters of metals without the appropriate apparatus in situ, we looked for an alternative way of producing small silver or gold clusters on surfaces.

Colloids of gold are commercially available with well defined diameters of, for example 5nm and 20nm. However, these are covered in a surfactant and are present in a solution. To coat a surface, one would simply add a small droplet and spin the sample rapidly for a given time until the desired coverage had been reached. But the removal of the solution molecules and the removal of the surfactant while the colloids are on the surface would be perhaps an impossible task. So, because a clean surface with clean particles is required, this method of particle production was discarded.

Pfund [1933] produced black films of gold, silver, copper and other metals by evaporation of small amounts of metal in 1-50 Torr of air. Experiments conducted in this department [Wang] showed that nanometre scale spherical particles of gold were formed in 0.5 mbar of Argon. Higher pressures formed larger and less spherical particles, and a separation of 5cm between evaporation source and surface was found to improve the uniformity of size and produce smaller particles than for one of 10cm. The nucleation and growth of these clusters occurs between the source and substrate, and is helped greatly by the lowering of particle energy due to collisions between metal and gas atoms. The frequency of collision will be proportional to the Argon pressure; increasing this pressure will lower the energy of the cluster to increase the flight time and therefore size. Increasing the separation will produce larger particles, and it was also found that increasing the rate of evaporation caused particle size to increase - i.e. a higher density of metal atoms in the gap means a higher rate of particle growth.
Chapter 5: Photon Emission from silver clusters

Figure 5.1. Scanning Electron Microscope (SEM) image of silver particles deposited onto an oxidised Silicon wafer by evaporation in Argon. The black scale bar represents 100nm.

Figure 5.2. STM image taken in air. Size: 3000Å (lateral) by 111Å (height range – black to white), conditions: tunnel current 0.4nA, tip bias +0.34V. The tip has moved a particle upwards and scans over it on every line, until finally it touches it and smears it out. It then continues to scan normally.
Figure 5.3. Enhanced low resolution SEM image of silver particles deposited onto graphite by evaporation in Argon. The black scale bar represents 1μm.

Figure 5.4. STM images taken in air of a silver cluster on graphite. (a) first scan (particle height 17Å), scan size: 300Å, height range: 42Å, conditions: 0.2nA current, +1.0V bias. (b) fourth scan (particle height 4Å, just visible over the noise), scan size: 100Å, height range: 15Å, conditions: 0.2nA current, +0.46V bias.
Figure 5.1 shows a fairly low density of particles on atomically flat silicon oxide which was prepared in an atmosphere of 0.5mbar Argon with a separation of 5cm and a relatively high evaporation rate. Particle diameters are in the range ~7-30 nm, but many of the particles are non-spherical. This preparation was performed in a different evaporator than the apparatus shown in section 4.2.

Figure 5.2 is an STM image (taken in air) of silver particles on a Highly Oriented Pyrolytic Graphite (HOPG) surface which was prepared in the same way as that in figure 5.1, but with a smaller surface density. Due to the weak van der Waals bonding between the particle and the surface, the STM tip was able to push the particles around the surface until they hit a step. A lower resolution SEM image (figure 5.3) shows that the particles combine together along steps even without the presence of an STM tip. One small particle was imaged in the STM in figure 5.4, but it became smaller with successive scans. Ohto et al. [1995] found that holes were formed by the presence of the tip in thin Au, Ag, Cu and Sn films deposited on graphite, silver being the material which moved most easily, thermomigration and electromigration being the mechanisms which explain this. A different method, that of metal evaporation onto heated graphite without the need of an inert gas can produce spherical Sn particles of size 10-40nm or mixtures of Au islands and particles, and the Au evaporation was tried in UHV but the same problems of migration occurred at higher bias and current.

As the STM needs to tunnel to a conducting surface, and a clean and flat surface is required to make the particles visible, the well-characterised Silicon (111)-7x7 reconstruction was chosen as the substrate. Its preparation involves cycles of heating the sample to 1200°C in a pressure below 10^{-9} mbar, followed by annealing at 800°C; then it is cooled to room temperature. This is all performed in the scan chamber described in section 4.2. Figure 5.5 is an STM image of the 7x7 reconstruction.

Figure 5.6 shows an STM image of a silver particle evaporation performed in the sample preparation chamber. The coverage is lower than that in figure 5.1, and the silicon surface is coloured red to illustrate that the particles are separated. The diameters range from 4Å to 30Å, and the evaporation parameters were as follows. Argon pressure: 0.5mbar, evaporation rate: ≈ 0.05 Å/sec, film thickness: ≈ 0.5Å, exposure time to Argon: 10 minutes, evaporation distance ≈ 5cm. A problem of contamination is evident – several minutes of exposure to 0.5mbar Argon renders the 7x7 reconstruction unidentifiable. However, the impurity of the Argon is 1 part in 10^5, and the Argon itself is unreactive and is very quickly pumped from the chamber, as it does not stick (bond) to metals well. So contamination, although a worry, only occurs at a monolayer level on silicon, and far less on the noble metal particles.
Figure 5.5. STM image of the Silicon(111)-7x7 reconstruction, showing adsorbates (white, diffuse objects) and missing atoms. Scan size: 200Å, height range: 3.3Å, conditions: 1.2nA current, -1.5V bias.

Figure 5.6. STM image taken in UHV of silver particles evaporated onto Si(111). Scan size: 1000Å, height range: 37Å, conditions: 2nA current, -2.0V bias. The first 2Å of the height are coloured red.
The image in figure 5.6 was taken with a lower bias voltage and tunnel current than is necessary for photon emission. When 10nA and 3V were used, scanning became far more unstable, and the surface was modified when imaged subsequently (with the lower bias and current). Figures 5.7(a) and (b) were taken before and after voltage ramps were performed with the feedback loop held. Two particles look to have been removed, although this may well be a “double-tip” image effect judging by the non-spherical looking particles – meaning that only one particle may have been removed. The most likely explanation is that the particle is not well bonded to the surface, and when the electrostatic force from the tip exceeds the force bonding the particle to the surface, it is accelerated towards the tip. Surges in the tunnel current, which were observed during unstable scanning, corroborate this. It is most likely that the particle does not keep its shape, but instead “wets” the end of the tip. To combat these instabilities, the sample was resistively heated to 500-600ºC, removing contaminants and providing enough energy to cause stronger bonding (chemisorption, as opposed to physisorption) to occur. Scanning at the higher current and voltage was subsequently far more stable. Another problem was the inability to choose particular sizes with the required coverage. As both the size of the particles and the coverage are proportional to the evaporation rate, large particles cannot be produced in small quantities. This is simply due to the inability to rotate the transfer arm into the beam and back in under a second. Thus the study of isolated particles made under these conditions is limited to those with diameters less than ~100Å. Although the preparation of small spherical particles in this complicated way may not resemble real surfaces, the preparation of well separated spheres is essential to measure the properties of a simple system and to compare with the two-sphere theory, and only then can the properties of real surfaces be understood.

5.2 Tip preparation

The STM tip acts as the other sphere in the two-sphere model. Actually, the shape is more paraboloidal, but the approximation is valid to within a few percent for the bottom 90º of the ‘circle’. The standard material used for tips in the last ten years has been tungsten, as it can be electrochemically etched to a radius of ~10nm. This is done by submerging a tungsten wire in 2.0M Sodium Hydroxide and applying a positive bias with respect to another electrode in the solution. The wire etches preferentially at the meniscus and the rest of the wire drops off after a few minutes, leaving a sharp tip. However, as it is still in contact with the solution, the etch continues until the wire is removed, and a delay of a few seconds can result in a radius of a few hundred nanometres. To avoid this, the bias voltage is switched off automatically within a few milliseconds of the current dropping below a set threshold - i.e. when the rest of the wire drops. The tip is then cleaned in cycles of water and methanol, and placed into vacuum. An SEM image of an average tip is shown in figure 5.8, but since the resolution of the microscope is limited to around 40nm, all that can be said is that the radius is less than 40nm.
Figure 5.7. (a) Image before and (b) image after voltage ramps. Those particles marked with a cross were removed. Scan size: 1000Å, height range: 37Å, conditions: 2nA current, -2V bias. Voltage ramped slowly 10 times from -2V to -3.5V.

Figure 5.8. Low resolution Scanning Electron Microscope image of a newly etched tip. The scale bar is 400nm long, and the resolution of the microscope is around 40nm.
In order to be sure than the end of the tip is metallic with no oxide, water, methanol or remnants of chemical etchant, a second cleaning and sharpening procedure is applied in vacuum – that of Neon ion bombardment. A schematic is shown in figure 5.9. The tip is loaded into the STM and a tantalum sample, previously cleaned with Argon ions in the sputter ion gun, is approached. A high negative potential (~1000V) is applied to the tip, and the sample is grounded. When the separation is around 0.1-2mm, a field emission current from the tip to the tantalum sample (limited to 10µAmps) is observed. Then, Neon is introduced into the chamber at a pressure of 10^{-5} mbar. The high energy (~1000eV) electrons ionise the Ne atoms to Ne^+ which are accelerated towards the negative tip. The tungsten is sputtered away by these high-energy ions and the tip is naturally sharpened, to a final radius of <10nm, and because the current is enhanced when the radius of the emitter (here, the tip) decreases it can be easily judged when the tip has become sharper.

Figure 5.9. Electrons are emitted from the negatively biased tip towards the sample (far to the left, not shown). The current of ionised Neon represents less than 10^{-3} of the electron current [Cavaille and Drechsler].

It is important not to touch the sample with the tip. Figure 5.10 shows a tip after a poor coarse approach, when the sample was allowed to touch the tip. Although there may be ‘microtips’ of a few atoms at the very end which make tunnelling acceptable, even with atomic resolution, these tips will be no good for photon emission as the radius is too large, and whatever light emission occurs will be blocked by the presence of the tip whose dimensions are larger than the wavelength of light. Additionally, multiple images are likely to occur.

It must be said that even a careful preparation of a tip will not necessarily yield good photon emission. Sometimes during scanning, applying voltage pulses to the tip would make the photon emission come and go, but by and large there were tips that would and tips that would not emit photons, the ratio of good ones being ~10-30%. The reasons for this come down to speculation – contamination, peculiar (non-paraboloidal) shape or too large or small end radius being possible explanations.
5.3 Image analysis

Having successfully prepared silver spheres on a flat substrate and sharp, clean tungsten tips, attention can turn towards the experimental results. However, there should first be some discussion of how quantitative results can be extracted from the images.

The topography of flat surfaces or those with low curvature will be very similar to that measured by the STM, however the interaction of tip and sample geometries will seem to broaden spheres or surfaces where the curvature is comparable with that of the tip (i.e. when sample radius \( \leq \) tip radius). The image cannot be deconvolved back to the original topography, even if the tip shape is known exactly. A schematic is shown in figure 5.11: if the tip were of much smaller radius the tip path would resemble more closely the real surface, and the smaller, centre particle would be resolved. The simulated tip path is the geometric interaction of the tip (plus tunnelling separation) with the surface, neglecting variations in the separation caused by variations in material – present in the work function and the electron density of states, and assuming curvature on a larger scale than the tunnelling separation – as the current is proportional to the area (even if the area doubles, however, the tip will pull back only \( \approx 0.3\AA \)).
Figure 5.11. Schematic demonstration of the convolution of tip and sample shape. Actual surface including spheres: shaded. Extent of tip: dashed line. ‘Effective’ tip radius, including tunnelling separation: thin line. Simulated tip path: thick line.

We can, however, find the tip and sample radius from the apparent height and diameter of isolated particles in the image. Figure 5.12 tells us that the apparent height will be just equal to the diameter of a particle on a flat surface, i.e. \( R_S = H/2 \). If the particle is assumed spherical, then simple geometry in the right angled triangle in figure 5.12 leads us to

\[
\left( \frac{D}{2} \right)^2 = (R_t + s + R_s)^2 - (R_t - s + R_s)^2
\]

\[
\left( \frac{D}{2} \right)^2 = 2(R_t + s)(2R_t)
\]

and the relation,

\[
R_t + s = \frac{D^2}{16R_s}
\]  

(5.1)
Figure 5.12. Determination of tip radius ($R_t$), and sample radius ($R_s$) from apparent height ($H$) and diameter ($D$) in the image.
Another method of estimation of tip radius, applicable to flatter surfaces, is by that of step broadening, shown in figure 5.13. From the right angled triangle we get

\[
(R_t + s)^2 = (R_t + s - h)^2 + D^2
\]

\[
R_t + s = \frac{h^2 + D^2}{2h}
\]

(5.2)

Figure 5.13.Determination of tip radius from apparent lateral extent of an abrupt step.

All the photon maps presented in this thesis are processed by dividing by the current which is also simultaneously recorded. The current is usually noisy when tunnelling at high biases and currents, so the photon counts, which are proportional to the current, are similarly noisy. So as the current has a very strong dependence on separation, whereas the quantum efficiency has a weak dependence, it is acceptable to renormalise the photon counts to what they would be at the set current, thereby removing the unwanted effects of current variations. I.e.

\[
\text{photons} = \text{recorded counts} \times \frac{\text{set current}}{\text{recorded current}}.
\]

The dark counts of the bialkali photomultiplier tube were subtracted from the photon maps, and in order to make all the range of counts visible, the contrast of each was increased. The dark counts of this tube were a few thousand per second, caused by photons coming from the ion pump. This noise converts to a quantum efficiency of \(\sim 10^{-5}\) at the currents used, but it was constant and subtracted from the signal, so actually the quantum efficiency is limited by drift, its observable lower limit being more like \(\sim 10^{-6}\).
5.4 Photon Emission from silver clusters on Si(111)-7x7

Figure 5.14. (a) Topography of silver particles fabricated as described in section 5.1. Image size: 2000Å, height range: 40Å. Tunnel current: 5nA, tip bias: -3.1V. (b) Photon map: 0-193 counts, or quantum efficiency 0 - 2.9x10^-5. Tip radius ≈70Å, particle heights: up to 20Å.
A typical topography and corresponding photon map are shown in figure 5.14. The topography image is not as ideal as figure 5.6 because it is taken at a high bias and current, and it took 20 minutes to collect enough counts to produce a good photon map – about 10 times longer than for figure 5.6. Immediately, a high similarity between the two images can be seen. The grey scale for the topography (black: low, white: high) is reproduced by the photon map (black: no counts, white: maximum counts). This means that no photons are being produced on the silicon, and that larger particles produce more photons than smaller ones. Also, each particle has a maximum in the emission at its centre. There is a $\cos^2 \theta$ dependence on the emission of electromagnetic radiation from a dipole, but in figure 5.15 it can be seen that for a non-spherical tip, the highest curvature will occur at the bottom of the tip or what appears to be the top of the particle in the image. Figure 5.16 confirms that a $\cos^2 \theta$ dependence alone cannot account for the experimental data, and shows the photon signal is slightly sharper than the topography.

**Figure 5.15.** Schematics of (left:) spherical tip and (right:) parabaloidal tip scanning over a particle. For the spherical tip, the curvature remains the same away from the centre of the sphere, but for a parabaloidal tip something closer to a sphere-plane arrangement occurs.

**Figure 5.16.** Linecut through four particles in figure 5.17. Thick line: topography. Bottom trace: quantum efficiency. Middle trace: rescaled $\cos^2 \theta$ variation (right hand scale).
Figure 5.17. (a) Topography of silver particles on silicon. Image size: 2000Å, height range: 57Å. Tunnelling current: 10nA, tip voltage: -3.1V. (b) Photon map: 0-140 counts, or quantum efficiency 0 - 2.7x10^{-5}. Tip radius ≈ 73Å, particle heights: up to 17Å.
The tip is stable in the top half of figures 5.17(a) and (b), so we can analyse the variation of the photon emission with sample radius in the knowledge that the tip radius has not changed. This tip radius, calculated using equation (5.1) as 73Å, did not change in this region. A graph of quantum efficiency versus sample diameter is shown in figure 5.19, and an approximately linear dependence of the photon emission with measured particle size is found. As the counts are in the range ~15-100, just statistics would explain the observed ‘noise’. The particles are well separated, as can be seen in figure 5.18.

Figure 5.18. Linecuts through the centre of four particles in figure 5.17. Bottom trace: quantum efficiency, defined as the number of emitted photons per tunnelling electron. Top trace: topography. Particles are drawn in, and the substrate is added as a straight line. Also, tips are drawn as the inner semi-circle, and the tunnelling distance added to make the outer semi-circle.
Figure 5.19. Scatter graph of quantum efficiency, measured at the centre of the particle, against diameter of the particle for twenty particles in the top half of figure 5.17. Statistical errors are shown, errors in measuring diameters are not. The straight line is not a best fit, rather a guide to the eye.

In figure 5.20 the tip radius has changed several times. The particles seem to be of a smaller lateral size in the bottom half of 5.20(a) than in the top half. Equation (5.1) shows that this means the tip radius has changed, as confirmed by the tip jumps producing horizontal lines in the topography – the tip is probably picking up silver particles. In order to determine the relationship between the quantum efficiency and tip radius, the photon counts must be divided by the size of the particle to remove the dependence on sample radius (shown in figure 5.19 to be approximately linear, and the variation in sizes is only ~25%, so this is acceptable). The tip radii are determined using equation (5.1) and the results are plotted in figure 5.21. Unfortunately, little more can be added to what can be seen by eye – that a larger tip radius produces more photons. The large deviations from a clear relationship are due to the statistical noise and the difficulty of reliably measuring the tip radius in each case due to unstable scanning. Also the chemical composition of the tip may well change by gaining or losing silver from the surface.
Figure 5.20. (a) Topography of silver particles on silicon. Image size: 1500Å, height range: 37Å. Tunnelling current: 5nA, tip voltage: -3.3V. (b) Photon map: 0-139 counts, or quantum efficiency 0 - 5.1x10^{-5}. Tip radius ≈50-120Å, particle heights: up to 12Å.
It is also possible to measure the variation of quantum efficiency between images. The tip radius changes between images as the tip often touches the surface, and the tip bias is also pulsed in an attempt to improve the cleanliness of the tunnelling gap when good photon maps are not being taken, so changing the end radius. In figure 5.22 it is found approximately that

\[
\text{quantum efficiency} \propto R_{\text{tip}} R_{\text{sample}}. \tag{5.3}
\]

The tip radius is deduced in the usual way (from equation 5.1) but the quantum efficiency needs to be corrected slightly because tip biases of -3.1V, and -3.3V, and -3.4V were used. It was found that introducing other tips onto the same graph interfered with this straight line. Some points had a very low quantum efficiency - probably due to the tip being tungsten and what is shown here is silver coated, and others were probably partially coated in silver.
Figure 5.22. Graph of quantum efficiency against the product of tip radius and sample radius. Tip radii are in the range 65-130Å and sample radii are 4-10Å. Statistical errors are shown. The straight line is not a best fit, rather a guide to the eye.

Figure 5.23 was taken on a sample with high coverage. It would seem that there is some peak in efficiency at the centre of many of the particles. On closer inspection, there are many particles which are touching, and the peak in efficiency occurs ‘between’ pairs of particles. Figure 5.24 is a linecut between two particles with the tip and spheres added. This peak in efficiency is not likely to be just the dependence on tip curvature demonstrated in figure 5.15, instead it is some kind of three-sphere plasmon. Efficiencies are around 5 times higher than for single particles, and the fact that the dipole has increased because two particles are being tunnelled into at the same time is not an explanation – figure 5.18 shows no enhancement when two particles are being tunnelled into, but are spaced apart. Another possibility is that of a ‘double tip’ – where there are two protrusions at the end of a tip. However, a tip reconstruction program, which is capable of showing up such double tips by analysing the topography for repeat structures instead shows a single tip in figure 5.25. So some sort of further enhancement, comparable with the difference between a sphere-plane and sphere-sphere plasmon, is occurring when these three spheres are brought together.
Figure 5.23. (a) Topography of silver particles on silicon. Image size: 1500Å, height range: 38Å. Tunnelling current: 10nA, tip voltage: -3.1V. (b) Photon map: 0-241 counts, or quantum efficiency $0 - 3.0 \times 10^{-5}$. Tip radius $\approx 40\text{Å}$, particle heights: up to 30Å.
Chapter 5 : Photon Emission from silver clusters

Figure 5.24. Linecut through two adjacent particles in figure 5.23. Bottom trace: Quantum efficiency. Top trace: topography, with spheres and tip added.

Figure 5.25. Three dimensional view of a tip reconstructed from the topography in figure 5.23, by using only the topographic data [Williams]. The surface represents the shape of the tip required to form the image. The tip is probably not this sharp – this is likely to be an artefact from the noise in the image, and the real tip will extend back beyond the flat plane from which it seems to protrude. Vertical range: 34Å, width 205Å.
Figure 5.26. (a) Topography of silver particles on silicon, with a number of tip deposits. Image size: 1000Å, height range: 116Å. Tunnelling current: 5nA, tip voltage: -3.4V. (b) Photon map: 0-606 counts, or quantum efficiency 0 - 2.2x10⁻⁴. Tip radius ≈65Å, particle heights: up to 8Å.
The STM is capable of making structures on the surface by pulsing the tip with a higher voltage. A number of such pulses were performed at -7.5V, and then the result imaged in figure 5.26. The three ‘tip deposits’ visible are 30Å, 30Å, and 100Å in height and are likely to be silver, as tungsten would give far lower emission. The shape of the deposits, responsible for the maxima in photon emission, is hidden in the topography but is evident in the photon map.

The photon vs. voltage characteristics were measured above the largest particle, and above one of the small silver particles of size ~10Å. The results are shown in figure 5.27 and 5.28. The loop on both occasions was held while the tip bias was varied and the current and photon counts logged at each sampling point. A peak in quantum efficiency is seen in figure 5.27 at -5.2V, and was around the same for a positive bias (i.e. ≈+5.2V). The noisier signal in figure 5.28 prevents the assignment of a peak, but it can be said that the two curves are similar.

These spectra are different to the ones seen in Gimzewski’s paper, when the current is kept constant and the voltage is varied. The feedback loop then responds to the change in current caused by a change in voltage and the tip moves back as the voltage increases, to compensate. As the tunnelling separation increases the efficiency goes down, so their peaks should occur at lower voltages; and indeed their peaks are seen at 3.5V. The method used here removes the dependence of the photon emission on the tunnelling separation from the spectroscopy.
If during its preparation, the Silicon surface is not heated for many seconds at 1200°C there will remain some silicon carbide on the surface, in the form of ~100Å high islands. The steps are pinned by the carbide during the step flow at high temperatures (>800°C). Figure 5.29 shows images of a sample which was incompletely flashed, then onto which silver particles were deposited at a high rate so as to make particles comparable in size to the carbide. The actual sizes of the silver were up to ≈80Å, and the carbide was up to ≈240Å in height.

Although the photon map is noisy due to a small tip radius and a purely tungsten tip – no contact with the particles on the surface was observed – it can be seen that some of the particles emit photons (silver) and others do not (carbide). This has shown for the first time a chemically specific photon map. Confirmation of which substance is which, comes from the fact that the positions of the non-emitting (carbide) structures are located around kinks in steps. Some silver particles are seen to be close to one of the carbide islands, others are seen on a flat plane at the middle on left of the image. A three dimensional overlay can be seen in figure 5.32.
Figure 5.29. (a) Topography of silver particles on silicon. Image size: 5000Å, height range: 392Å. Tunnelling current: 10nA, tip voltage: -3.1V. (b) Photon map: 0-64 counts, or quantum efficiency 0 - 1.2x10^{-5}. Tip radius ~90Å, particle heights: up to 75Å.
Figure 5.30. (a) Topography of silver particles on silicon. Image size: 3000Å, height range: 406Å. Tunnelling current: 10nA, tip voltage: -3.3V. (b) Photon map: 0-632 counts, or quantum efficiency 0 - 1.2x10^{-4}. Tip radius ≈40Å, particle heights: up to 60Å.
Another sample with carbide and silver on the same surface is imaged in figure 5.30. Here the tip radius is larger, but unfortunately the size variation between the two species is much greater. There are a few cases of emission seeming to come from the larger (carbide) structures. These spots of emission are just silver particles – they are deposited uniformly on the surface, so a proportion will land on top of the carbide. But it should be asked why it is that there are not more emitting features on top of the carbide islands, i.e. why the photon map appears dark for the carbide, instead of littered with silver particles like the rest of the surface. The answer comes from considering the tip path over the surface, demonstrated in figure 5.31. A three dimensional overlay of another pair of images of this sample can be seen in figure 5.32.

![Figure 5.31. Schematic tip path over a surface with carbide structures and silver particles. The side of the tip will tunnel to the side of the large carbide structure, so there will be no emission for the majority of the carbide particle, even though there are unimaged silver particles beside it.](image-url)

It was said that an enhancement of the photon emission occurred for curved surfaces, so it may be confusing that these small particles have quantum efficiencies around $10^{-5}$, whereas similar flat surfaces have efficiencies around $10^{-4}$ to $10^{-3}$ for the same radius tip (in the next section, gold is found to have an efficiency of $6 \times 10^{-4}$ at -3.0V). Also, figure 1.3 shows a 50Å high protrusion formed on a copper surface, and an enhancement of an order of magnitude is seen in the efficiency. The curved surfaces emit more photons than the small particles because the plasmon is limited to the particle size, when on flat surfaces it would extend ~50Å (the geometric mean of the tip radius and the tip-sample separation). Then the dipoles would be smaller, and hence the emission would be reduced.
Figure 5.32. Top: three-dimensional depiction of figure 5.29, with the colour (blue -> red) representing the amount of photon counts, superimposed on the topography. Bottom: a mixture of carbide and silver particles, similar to the data in figure 5.31. Photons (red -> blue): 0-697 counts or quantum efficiency 0-1.2x10^{-4}; topography: 3000Å wide, 305Å high; conditions: 10nA, -3.3V.
5.5 Photon Emission from flat surfaces

When tunnelling on flat metals, photon maps should be featureless, as there are no material or topographic changes. This was found to be true, except at step edges, where the photon counts dropped. This has been reported elsewhere [Berndt 1993a], where a $\cos^2 \theta$ variation was found. Here, on the other hand, a far larger drop in emission was seen. A similar explanation as that given in figure 5.15 is used here, that the part of the tip tunnelling to the step has a lower curvature. Figure 5.33 is a linecut of an atomically flat gold surface, for which the average quantum efficiency is $6 \times 10^{-4}$. Also, three adsorbates are visible (at x=310, 530 and 550Å) and the emission is noticeably reduced. This may be due to changes in the local density of states, a decrease in efficiency as the separation between the tip and gold surface increases, or a combination of these two effects. The tip radius is calculated to be $\approx 70\text{Å}$ from equation (5.2). Photon-voltage characteristics for this surface were similar to those in figure 5.27, however the data was too noisy to present.

![Figure 5.33](image-url)

*Figure 5.33. Top: linecut of topography of a flat gold sample. Middle: rescaled $\cos^2 \theta$ graph, derived from the topography (right hand scale). Bottom: quantum efficiency. Conditions: 2nA, -3.0V.*
Figure 5.34. (a) Topography of Si(111)-7x7 with carbide structures. Image size: 5000Å, height range: 133Å. Tunnelling current: 10nA, tip voltage: -3.6V. (b) Photon map: 0-329 counts, or quantum efficiency 0-6.2x10⁻⁵. Tip radius ≈30Å.
Figure 5.34 shows the topography and corresponding photon map for Si(111)-7x7, with carbide but no silver particles. The sample was treated at a pressure of $<10^{-9}$ mbar, so the silicon is bare and has not reacted with any gases. It can be seen that the silicon emits photons but the carbide does not, and that there are some triangular features on the silicon which show reduced emission. There are also lines of reduced emission corresponding to steps.

It is not unexpected that Si(111)-7x7 should emit photons – it is metallic-like in its band structure, with a continuum and high concentration of free electrons. This means it can sustain a plasmon, and the surface of silicon has a relatively low imaginary part of its dielectric function ($\sim 10$ in the optical range [Hammadi]). The smaller tip radius, lower free electron density and higher dielectric losses explain why the efficiency is an order of magnitude lower than for gold, even using a higher bias. No photons were observed on the silicon substrate on which the silver particles lay in the previous section. This is most likely due to the silicon reacting with impurities in the argon gas and the molecules leaving the walls of the sample preparation chamber and gas lines, as the pumps were valved off for a minute while the evaporation took place. Only $5 \times 10^{-6}$ mbar sec of oxygen will cause the surface to saturate [Dai].

![Figure 5.35. Three dimensional view of the topography in figure 5.34(a). An oblique light helps to visualise steps and other small topographic effects.](image)
The origin of the triangular areas of reduced quantum efficiency (from ≈ 6x10^{-6} to ≈ 4x10^{-6}) is more difficult to explain. There is a slight topographic difference between the regions visible in figure 5.35 – the regions of reduced emission are higher by approximately half an Angstrom. A number of possibilities arise as explanations. Surface contamination would damp out the plasmons and reduce the emission (or alter the modes); a different phase would have a different electronic surface states and hence different plasmon modes; and subsurface contamination would cause the damping to increase. It is clear that these structures are bound by steps and are not necessarily adjacent to the carbide, and the [110] symmetry would favour submonolayer metal contamination. For example, tungsten atoms on this surface form islands bordered by [110] directions when annealed for long enough [Gray]. Also, the topography change of half an Angstrom would suggest metal contaminants. No atomic resolution images were taken as it was only clear after processing the data that these structures existed.

The photon-voltage characteristics of the surface are shown in figure 5.36. As for other metal particles, the quantum efficiency rises up to about 4-5V, then levels off. In figure 5.37, feedback is maintained as the voltage is changed, and the peak occurs at a higher voltage. This would suggest that the efficiency is higher when the tip is further from the surface – contrary to noble metals. This is confirmed by figure 5.38.

![Figure 5.36. Quantum efficiency as a function of voltage for Si(111)-7x7. The loop was held with a tunnel current of 5nA and a tip voltage of -2.0V. Below ~-3.5V the data becomes too noisy to be meaningful.](image_url)
Figure 5.37. Quantum efficiencies at 10nA for four different biases. The loop is not held between each reading, instead the tip retreats slightly as the voltage is increased, to maintain the current. Errors are less than a few percent.

Figure 5.38. Quantum efficiency as a function of distance moved towards the silicon sample. The loop was held with a tunnel current of 1nA and a tip bias of -5.0V.
5.6 Summary of photon emission experiments

- Flat surfaces of Si(111)-7x7 were prepared with well separated silver particles of size 5-75Å, and clean tips with radii in the range 40-130Å were used to image them. Photon emission with an efficiency of ~10^{-5} photons per tunnelling electron was observed (at a bias of -3.1V). These particles are the smallest recorded to emit photons.

- It was found that the efficiency fitted the relations $\eta \propto R_{\text{sample}}$ and $\eta \propto R_{\text{tip}} R_{\text{sample}}$, approximately.

- An enhancement in photon emission by a factor of ~5 was observed when the tip tunnelled to two adjacent particles.

- Efficiency-Voltage curves peaked at -5.2V for silver deposited from the tip, and this peak was similar for any size of particle.

- On the same image, silver particles showed photon emission and carbide of similar size showed no emission. This is the first example of a chemically specific photon map, and cannot be confused with topographic effects. A second experiment had a far better signal to noise ratio, although particle sizes were very different.

- On flat gold surfaces, drops in emission were seen for adsorbates and across steps. The drop across the step showed a dip related to a change in curvature across the tip. An efficiency of 6x10^{-4} was recorded at -3.0V.

- Photon emission from clean Si(111)-7x7 was observed with an efficiency of 6x10^{-6} at -3.6V. Some chemical influence in triangular regions, probably a monolayer of metal, caused a drop in emission by approximately one third.

- Efficiency-Voltage characteristics for silicon were different from that of gold, as the efficiency decreased as the tip moved towards the surface.
Chapter 6
Results of theory and comparison with experiment

6.1 Photon Emission from flat surfaces

A Fortran program was written to numerically calculate results for the model described in chapter 3. Any material can be used for tip or sample, if the dielectric function, Fermi energy and work function are known. Emission spectra, integrated photon counts, inelastic and elastic currents and mode energies are recorded for given materials and geometry. Other important results, such as dipoles, potentials and broadening of modes can be monitored, and the contribution from each sphere can be isolated.

The first test of the two-sphere model is to reduce the second sphere to a plane by putting $\mu_2 = 0$ and compare this subset of the new two-sphere model with Johansson’s sphere-plane model [1990-91]. As our (sphere-plane) models are only the same in that we start by solving Maxwell’s equations in bispherical co-ordinates, it is essential to see that they produce similar results, then we can be assured that both models are accurate, at least from their starting assumption of modelling the tip as a sphere. At the same time, we can compare with the variety of experimental data obtained by Berndt and Gimzewski and check on the limitations of the models.

Figures 6.1(a) and (b) show experimental spectra for W and Ag tips on an Ag(111) surface, beside which in (c) and (d) are predictions from Johansson’s model. Predictions from the new model are presented in figure 6.2(a) and (b). There is a good similarity between the two theories – peaks occur at 460nm for W-Ag and 560nm and 425nm for Ag-Ag. Quantum efficiencies were calculated to be $2.6 \times 10^{-4}$ (W-Ag) and $6.5 \times 10^{-4}$ (Ag-Ag) in the new model.

For the Ag-Ag case, the excited modes have energies corresponding to the peaks, but for W-Ag, the peak at 460nm does not actually correspond to the first excited mode of 3.29eV (=377nm). This is above the maximum energy of the tunnelling electrons ($e \times 3V = 3eV$), but these modes can still be excited because the f.w.h.m. broadening of the mode is very large.
(3.3eV). This peak shifts down in energy (up in wavelength) as the tip radius increases or separation decreases, because the modes decrease in energy. This red shifting of modes is true for the Ag-Ag or any other metal system tested. The modes occur at energies up to where $\varepsilon_{\text{real}} = -1$ for the sample (3.68eV for Ag, 3.16eV for Au) and when the tip is W or Ir, the modes occur close to this energy. For Ag or Au tips, the modes start at lower energies (2.26eV for Ag-Ag, compared to 3.29eV for W-Ag in figure 6.2).

![Figure 6.1](image)

**Figure 6.1.** Experimental photon emission spectra for (a) tungsten and (b) silver-coated tips on a flat Ag(111) surface. Tip biases are -3V in both cases, but currents were not given. SEM images suggested radii of $\approx 300\text{Å}$. (c) and (d) are Johansson’s predictions for the same systems using tip radii of 300Å, but separations were not given [Berndt 1993b].

![Figure 6.2](image)

**Figure 6.2.** Predictions for (a) W-Ag and (b) Ag-Ag systems of figure 6.1 from the model described in chapter 3 (“new” model). A tip radius of 300Å was used in both cases, and separations were (a) 5Å and (b) 7Å.
Chapter 6: Results of theory and comparison with experiment

Figure 6.3. Experimental photon emission spectra for (a) tungsten and (b) gold-coated tips on a flat Au(110) surface. Tip biases are +3V in both cases, but currents were not given. SEM images suggested radii of \( \approx 300\,\text{Å} \). (c) and (d) are Johansson’s predictions for the same systems using tip radii of 300Å, but separations were not given [Berndt 1993b].

![Figure 6.3 Experimental photon emission spectra](image)

Figure 6.4. Predictions for W-Au and Au-Au systems in figure 6.3 from the new model. A tip radius of 300Å was used in both cases, and separations were both 5Å.

![Figure 6.4 Predictions for W-Au and Au-Au systems](image)

Figure 6.3, like figure 6.1, shows experimental spectra and corresponding predictions from Johansson’s model, this time for W-Au and Au-Au junctions. Spectra from the new theory are given in figure 6.4, and the two theories are again similar, although less so for the Au-Au case. All the experimental and theoretical spectra stop at \( \approx 2.5\text{eV} \) (500nm) because \( \varepsilon_{\text{real}} = -2 \) for gold here, and as was discussed in chapter 3, the plasmon modes decouple above this energy. For the W-Au case, modes were found at 505nm and 495nm, and for Au-Au at 680nm, 535nm,
515 nm, and 500 nm. The gold tips were made by coating \(\approx1000\) Å of gold onto a \(\approx 300\) Å W tip, so in figure 6.5 a prediction is made for a radius of 1300 Å which matches the experiment better. Quantum efficiencies are \(3.0 \times 10^{-4}\) (W-Au), \(5.6 \times 10^{-4}\) (Au-Au, 300 Å) and \(1.8 \times 10^{-2}\) (Au-Au, 1300 Å), so it may be this increased tip radius by film deposition which causes the observed enhancement in photon emission of an order of magnitude [Berndt 1993b].

![Figure 6.5](image)

**Figure 6.5.** Prediction for Au-Au system with a tip radius of 1300 Å and separation of 7 Å.

Results for a copper surface are shown in figure 6.6, and all spectra show some quantitative similarities. The new model predicts a quantum efficiency of \(7.3 \times 10^{-5}\), and there is just one mode excited, at 385 nm. The features in the spectrum come from the variations in the imaginary part of the dielectric function, responsible for the dielectric losses. It is found that it is the dielectric function of the sample which dominates these losses, and functions for Ag, Au and Cu are shown in figure 6.7. As was mentioned in section 3.7, excitation of electronic interband
transitions cause jumps in the functions at 310nm for Ag, 500nm for Au and 600nm for Cu. However, only the jump for copper occurs below $\varepsilon_{\text{real}} = -2$ so is visible in the spectrum, and all the spectra with a tungsten tip (when the modes are at a higher energy) are highly dependent on the imaginary dielectric function of the surface. The imaginary dielectric functions of tungsten and iridium, although not critical to the spectra, are shown in figure 6.8. Tungsten is the most neutral over the whole spectrum, so is most likely to reveal the true dielectric function (both imaginary and real) of the surface in the photon emission spectrum.

**Figure 6.7.** Imaginary part of the dielectric function used in the new model for the noble metals Ag, Au and Cu [Lynch and Hunter].

**Figure 6.8.** Imaginary part of the dielectric function used in the new model for the commonly used tip metals W and Ir [Lynch and Hunter].
The spectrum is also dependent on the density of states for tunnelling, which ties the spectrum to zero at the bias voltage, because the photon emission \( \approx \frac{\nu_{\text{inel}}}{\nu_{\text{rad}}} \nu_{\text{rad}} \) from (3.58). This is demonstrated experimentally for tungsten on Au(110) in figure 6.9(a) and by the new model in 6.9(b). For 2V bias, the spectrum is forced to zero to 620nm and the peak occurs at \(~700\)nm. At higher voltages the peak heads towards 600nm, because the influence of the inelastic tunnelling rate disappears from the spectrum, leaving the influence coming from the dielectric function of the sample in \( \nu_{\text{die}} \).

![Figure 6.9. Series of spectra for W-Au, (a) experimentally [Berndt 1991b] and (b) from the new model where the separation is 5Å and tip radius is 300Å.](image)

A similar set of spectra for W-Cu in figure 6.10(a) shows the spectra stabilising at higher voltages. As for the W-Au case, they are slightly different from the experimental ones in 6.10(b) probably because the modes are predicted to be at shorter wavelengths than will have occurred experimentally, thus skewing the spectrum to shorter wavelengths. I.e. the tip radius in the experiment may well have been greater than 300Å.
Chapter 6: Results of theory and comparison with experiment

Figure 6.10. Series of spectra for W-Cu, (a) from the new model, where the separation is 5Å and tip radius is 300Å, and (b) experimentally [Berndt 1991b].

Figure 6.11. Quantum efficiency for W-Au calculated for a series of tip radii. The separation is 5Å and tip bias is +3.0V.
When considering the integrated photon counts in the whole energy range, the dependency of quantum efficiency on tip radius is found to be $\eta \propto R^{2.4}$ for the W-Au junction in figure 6.11, and the value of this exponent was in the range 2.0-2.5 for all the sphere-plane junctions considered above. For the dominant mode, the dependence of the individual rates on tip radius was as follows: $\nu_{\text{inel}} \propto R^{-0.4}$, $\nu_{\text{rad}} \propto R^{2.4}$, $\nu_{\text{dieu}} \propto R^{0.4}$. Combining these together gives $\eta \propto R^{1.6}$ for this dominant mode, which differs with the $\eta \propto R^{2.4}$ found for the overall emission. However, as the radius increases, the energy of the dominant mode decreases, so the proportion of the broadened mode able to emit light (i.e. is below the cut-off energy where $\varepsilon_{\text{real}} = -2$) increases. Also, more modes are being excited, and a combination of these two factors must provide the missing $R^{0.8}$.

The relative strengths of these individual rates can be analysed, the example used being the W-Ag junction of figure 6.2a. These rates are dependent on frequency, but at the peak of ~2.5eV they are:

- $\nu_{\text{inel}} = 7\times10^{10} \text{s}^{-1}$
- $\nu_{\text{rad}} = 3\times10^{13} \text{s}^{-1}$
- $\nu_{\text{dieu}} = 5\times10^{15} \text{s}^{-1}$

These results are for the dominant mode, but are similar for the other modes too. This results in a broadening of $\hbar(\nu_{\text{dieu}} + \nu_{\text{rad}}) = 3.3\text{eV}$, dominated by the dielectric loss, not the natural (radiative) lifetime. For the Ag-Ag case in figure 6.2b, the first mode has a far smaller linewidth of 0.06eV, and the second $\approx 0.2\text{eV}$.

The overall inelastic current for W-Ag is calculated as $3.6\times10^{10}\text{electrons s}^{-1}$ and the elastic current as $3.2\times10^{11}\text{electrons s}^{-1}$ (50nA). The whole mechanism of photon emission can be drawn as an event tree, as in figure 6.12. For the Ag-Ag junction in figure 6.2b, the proportion of inelastic current is only 0.2%, but the radiation accounts for 30% of the loss processes.

![Figure 6.12. Event tree for the W-Ag system in figure 6.2a for the whole range of photons.](image)
When calculating the voltage dependence of the integrated photon yield, it was not possible to find a peak where it had been seen on flat Au at ~5V (mentioned in section 5.5). Figure 6.13 shows the photon-voltage yield for W-Au. At these high voltages (>~4V) it is difficult to make predictions with such a simple model for tunnelling - that of an average barrier, but at low voltages it should be more accurate. It is likely to be for this reason that a peak is not observed.

![Graph](image)

**Figure 6.13. Variation of quantum efficiency with voltage for a W-Au junction with 300Å radius tip and a constant 8Å separation.**

The quantum efficiency for a W-Au junction with a tip radius 70Å, separation 5Å and tip bias -3.0V was found to be $4 \times 10^{-5}$. This figure is some way from the one found experimentally of $6 \times 10^{-4}$ (in figure 5.33). There are a number of possibilities why this discrepancy may occur. Firstly, the tip was estimated to be of this radius by the apparent broadening of steps, but this only gives the radius of the few Angstroms closest to the surface anyway, so is a poor guide, and a radius of 220Å would give the right efficiency anyway. Secondly, the approximation of the tip as a sphere instead of a parabola may increase the mode energies so they enter less into the spectrum. Thirdly, the inelastic and elastic currents are calculated separately and by different means, so as they are extremely sensitive (for example to changes in barrier shape and separation) it may be optimistic to expect good agreement between two such simple tunnelling approximations.
6.2 Photon Emission from metal spheres

It has been shown that the new model compares well with the available data for the sphere-plane case, so the next task is to compare it with the data available for particles. However, the only data exists in this thesis (chapter 5) where there is no spectral information. The comparisons with experiment must instead be:

i) The dependence of the quantum efficiency with tip radius and sample radius (in chapter 5 it was shown that $\eta \propto R_{\text{tip}} R_{\text{sample}}$),

ii) The efficiency-voltage characteristics for a particle and tip deposit (to compare with figures 5.27 and 5.28).

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig6_14.png}
\caption{Dependence of quantum efficiency on sample radius, for (a) W-Ag and (b) Ag-Ag junctions. Parameters are in both cases: tip radius 73Å, separation 5Å, tip bias -3.1V.}
\end{figure}
Chapter 6: Results of theory and comparison with experiment

Figure 6.14 shows the relationship between efficiency and sample radius for both W-Ag and Ag-Ag junctions, but when modelling very small particles (of radius less than ~10Å) the results start to become numerically unstable. This is due to the denominators in equation 3.43 heading through zero, or in more physical terms, that the sphere is comparable in size to the gap so the approximations made in going from (3.33) to (3.34) are no longer true. Nevertheless it is seen that for the silver tip, the relation is approximately linear, but for the tungsten tip it is found that \( \eta \propto (R_{\text{sample}})^{3.5} \). Some of this difference comes from the fact that for tungsten more of the mode is entering the spectrum when the radius increases.

In figure 6.15, it is found (approximately) that \( \eta \propto R_{\text{tip}} \) for a silver tip, whilst a tungsten tip has a variation of \( \eta \propto (R_{\text{tip}})^{0.5} \). As for the sphere-plane case, it is the imaginary dielectric function of the sample which dominates the losses for a W-Ag or W-Au system, so changing the tip radius will have little effect. But for an Ag-Ag or Au-Au system, it is the tip which dominates the losses, so there is a stronger dependence on tip radius.

![Figure 6.15. Dependence of quantum efficiency on tip radius, for (a) W-Ag and (b) Ag-Ag junctions. Parameters are in both cases: sample radius 10Å, separation 5Å, tip bias -3.1V.](image)
When combining the variations of efficiency with tip and sample radius, it can be seen that they compare well to the experimental data in chapter 5 ($\eta \propto \frac{R_{\text{tip}}}{R_{\text{sample}}}$), if a silver tip is employed. This would confirm the suspicions that the tip had indeed become coated with silver during the experiments, and may explain why some tips ‘worked’ (those that became covered in silver) and others didn’t (clean W tips). However, the predicted efficiencies for these silver tips are still much lower than those observed ($10^{-5}$).

In chapter 5 it was found that the silver particles have a much lower efficiency than a plane, and we are now in a position to predict for what radius of sphere will the efficiency be comparable to a plane. For a W-Au junction with tip radius 70Å and separation 5Å at -3.0V, a sphere of radius $\approx 85$Å is predicted to match the plane’s efficiency of $4\times10^{-5}$.

Now the second comparison with experiment must be made – that of the efficiency-voltage characteristics. First, the tip deposit imaged in figure 5.26 is modelled in figure 6.16. Here, an onset in the efficiency can be seen at about 3–3.5V, and in the experimental data in figure 5.27 it is more like 3.5–4V. The characteristics of both tips (W, Ag) are similar to the experiments, although there is still the inability to model the peak in efficiency at high voltages. When modelling a small particle in figure 6.17, again the onsets are around 3–3.5V but the efficiency for the silver tip is more than an order of magnitude higher. This is again due to more of the mode being in the emission spectrum. Either of these shapes compare well with figure 5.28, but the magnitude of the efficiency suggests a silver tip.

![Figure 6.16. Dependence of quantum efficiency on tip bias, for (a) W-Ag and (b) Ag-Ag junctions. Parameters are in both cases: tip radius 65Å, sample radius 50Å, separation 5Å, tip bias -3.1V.](image-url)
Chapter 6: Results of theory and comparison with experiment

Figure 6.17. Dependence of quantum efficiency on tip radius, for (a) W-Ag junction (left hand scale) and (b) Ag-Ag junction (right hand scale). Parameters are in both cases: tip radius 65Å, sample radius 10Å, separation 5Å, tip bias -3.1V.

Figure 6.18. Predictions of efficiency-voltage characteristics for (a) W-Ag and W-Au junctions and (b) Ag-Ag and Ag-Au junctions. Parameters are in all cases are: tip radius 73Å, sample radius 10Å and separation 5Å.
Chapter 6: Results of theory and comparison with experiment

Having made good comparisons with chapter 5, it is now safe to make predictions for other sample materials and for spectra. Figure 6.18 shows predictions for the voltage dependence of the efficiency for gold and silver particles. Both W and Ag tips give similar findings: that the two metals would show similar emission at \(\sim 3–4\) V, but their turn-ons are rather different. As has just been shown, for Ag this occurs at \(\sim 3–3.5\) V, but for Au particles it occurs at \(\sim 2\) V. These turn-ons were shown to be independent of particle size, which means that gold and silver particles can be distinguished solely on their efficiency-voltage characteristics.

Data not presented in chapter 5 confirms this. Gold particles of size 5-20Å were prepared in the same way as the silver particles, and a typical efficiency-voltage curve is shown in figure 6.19. This demonstrates the aim of the thesis, which is to distinguish gold and silver particles by their photon emission.

\[\text{Figure 6.19. Quantum efficiency as a function of bias voltage taken above a gold particle. The loop was held while the current was 4nA and the tip voltage was +1.6V.}\]

The predicted spectra for Ag and Au particles at -4V bias are shown in figure 6.20. For either a W or Ag tip it can be seen that the spectra are easily separable, and they are quite insensitive to changes in sample radius. The form of the spectra are similar to those for the sphere-plane case with the same radius tip (not shown), but in this case the modes are of slightly lower energy. Even without access to a spectrometer, gold and silver particles could be distinguished by using a filter admitting light either above or below \(\approx 500\) nm.
Chapter 6: Results of theory and comparison with experiment

Figure 6.20. Predictions of emission spectra for (a) W-Ag and W-Au junctions and (b) Ag-Ag and Ag-Au junctions. Parameters are in all cases are: tip radius 73 Å, sample radius 10 Å and separation 5 Å. Overall efficiencies are: W-Ag $2.6 \times 10^{-7}$, W-Au $3.0 \times 10^{-7}$, Ag-Ag $4.9 \times 10^{-6}$, Ag-Au $6.5 \times 10^{-7}$.

Larger radius tips, such as the 300 Å tips used in figures 6.1 and 6.3 cause the spectra to be shifted up in wavelength, so that when the sample radius changes the spectra will also change, maybe overlapping. When distinguishing particles, the position of peaks in the spectra should ideally not change, instead each metal having its own 'fingerprint'. So there must be a compromise between this advantage of smaller tips, and the advantage of larger tips having higher efficiencies. The ideal tip would seem to be a ~100 Å radius silver tip when the spectra remain fairly independent of particle size, but the counts are high enough due to the silver in the tip. This tip would also have good enough topographic resolution.
Chapter 6 : Results of theory and comparison with experiment

6.3 Summary

**sphere-plane**

- New model compares quite well with Johansson’s sphere-plane model and Berndt & Gimzewski’s spectra of flat surfaces.

- Peaks for W-Ag, W-Au are not excited modes, rather tails of broadened modes.

- As tip radius or separation increases, modes are shifted up in wavelength.

- Spectra for Au surfaces stop at \(\approx 500\text{nm}, \) where \(\varepsilon_{\text{real}} = -2\).

- Spectra for Cu surfaces show characteristic drops in emission below \(\approx 600\text{nm},\) due to a rise in the imaginary dielectric function.

- W tips give spectra most dependent on the dielectric function of the surface.

- Higher biases reveal the best information about the sample, unaffected by quantum cut-off.

- \(\eta \propto (R_{\text{tip}})^x\) where \(x=2.0-2.5\) for the junctions studied.

- No peak predicted in efficiency-voltage curves – shows limitations of model at high bias.

**sphere-sphere**

- \(\eta \propto R_{\text{tip}} R_{\text{sample}}\) (approximately) for Ag-Ag in the range corresponding to results from chapter 5.

- Onset of emission predicted at \(~3–3.5V\) for Ag particles. Compares with experiments.

- Onset of emission predicted at \(~2\text{V}\) for Au particles, and confirmed by extra experiment. So **have distinguished Ag and Au particles experimentally**.

- Another way of distinguishing Ag and Au particles will be by their spectra – and changing the radius will not greatly affect the spectra of W-Ag and W-Au, as the modes do not move very far in energy (bunched close to \(\varepsilon_{\text{real}} = -1\)). Ag/Au-Ag/Au spectra will be more affected, as will those of W-Ag and W-Au with larger tip radii. Ideal tip should be 100Å radius silver tip.
Chapter 7
Light emission from hot electrons in atomic size contacts

7.1 Introduction

In ordinary metal wires, current densities of $\sim 10^{11}$ Amps m$^{-2}$ will cause melting as the metal will not be able to lose enough power by conduction or radiation [Morris]. The electrons which have gained energy lose it to the lattice instantly, so the lattice temperature is equal to the electron temperature. This means that as wires in electronic devices shrink in size, the voltages used must drop or it will fail. However, in nanometre scale contacts the major energy loss mechanism from electrons to the lattice vanishes, allowing $\sim 10^{15}$ Amps m$^{-2}$ to be carried. Visible light (1.5-3.5eV) was emitted even for voltages of $\sim 1$V across a contact, as the electron temperature had risen sufficiently.

The mechanism of energy transfer between electrons and the lattice in the bulk is that of the Cerenkov generation of longitudinal acoustic waves (phonons) by travelling electrons. In the bulk, electrons can be shown to have harmonics of all frequencies, but in small particles of size L, the electron wavelength is limited to below 2L (when one half-wavelength is accommodated). Their harmonics must be of frequency $> \frac{v}{2L}$, v being the electron velocity. Travelling electrons have energy greater than the Fermi energy (only electrons with energy greater than the Fermi energy can conduct), so electrons generating longitudinal phonons in particles are made up of harmonics, the lowest frequency being $\frac{v_F}{2L}$. I.e. their lowest energy component is $\frac{h v_F}{2L}$, h being Planck’s constant.

Phonons can have energies up to the Debye energy, the lattice equivalent of the Fermi energy. When the lower electron energy limit rises above this upper phonon energy limit – i.e.
when \( \frac{h v_c}{2L} > \hbar \omega_0 \) – resonant excitation of longitudinal phonons becomes impossible. This condition is equivalent to \( L < L_0 \) where \( L_0 = \frac{\pi v_F}{\omega_D} \), \( \omega_D \) being the Debye frequency. \( L_0 \) is calculated to be 200nm for gold (100nm for W, 120nm for Al, 500nm for Pb) [Ashcroft and Mermin 2]. For particles below this size, the dominant loss mechanism is then that of excitation of surface modes, which is at least 2 orders of magnitude less efficient [Gorban].

The equations governing the rates of energy transfer for electrons and phonons are [Belotskii],

\[
\frac{\partial}{\partial t}(C_{el} T_{el}) = \nabla \cdot (K_{el} \nabla T_{el}) - \beta (T_{el} - T) + Q \quad (7.1)
\]

\[
\frac{\partial}{\partial t}(C T) = \nabla \cdot (K \nabla T) + \beta (T_{el} - T) \quad (7.2)
\]

where \( C \) is the heat capacity per unit volume, \( T \) is the absolute temperature, \( K \) is the thermal conductivity, \( Q \) is the power input per unit volume and \( \beta \) is a coefficient of heat transfer between electrons and phonons. The variables with subscript ‘el’ refer to the electrons and those without refer to the lattice.

At steady state and in this contact,

\[
0 = \nabla \cdot (K_{el} \nabla T_{el}) - \beta ' (T_{el} - T) + \frac{IV}{\text{volume}} \quad (7.3)
\]

\[
0 = \nabla \cdot (K \nabla T) + \beta ' (T_{el} - T) \quad (7.4)
\]

where \( \beta ' \) is the coefficient of energy transfer from electrons to surface oscillations, and \( V \) and \( I \) are the applied bias and current through the volume of the contact respectively. Both terms in equation (7.4) are negligible compared to the other two terms in equation (7.3). Energy transfer to the lattice is reduced in the contact and the electrical power (IV) is dissipated by the flow of energy between electrons away from the source. Away from the contact, the bulk regime applies again, and energy from the electrons will quickly be lost to phonons. So the picture of energy transfer here is that electrons acquire energy and become hot in the contact, this energy is carried...
some distance away by electrons and then lost to the lattice, but the power density has become
diluted so the surrounding lattice does not melt.

In a film of small particles resistively heated [Belotskii], there is thermal homogeneity, so
the term \( \nabla \cdot (K_{el} \nabla T_{el}) \) in equation (3) vanishes, leaving a different equilibrium than the one for
contacts. Then,
\[ \beta'(T_{el} - T) = IV / \text{volume} . \]

A convenient water analogy has been drawn from this equilibrium. Figure 7.1 shows an inner
cylinder, whose height represents the electron temperature, and an outer cylinder whose height
represents the lattice temperature. The width of the slit is given by \( \beta' \), and the melting
temperature, \( T_m \), is also marked. The smaller the value of \( \beta' \), the higher the electron temperature.

![Water analogy diagram for particles](Image)

*Figure 7.1. Water analogy diagram for particles [reproduced from Belotskii]. For an
explanation, see above.*

At sufficiently high electron temperatures, and therefore energies, we should see
spontaneous photon emission with a spectrum of \( W(\nu) \sim \exp \left( \frac{-\nu}{kT_{el}} \right) \), and electron emission
current of \( I_e \sim \exp\left(\frac{-\phi}{kT_{el}}\right) \). \( k \) is the Boltzmann constant, \( \nu \) is the photon frequency and \( \phi \) is the work function. Tomchuk and Fedorovich [1966] showed that \( (kT_{el})^2 = (kT)^2 + \alpha IV \) when comparing energy losses from electrons to the lattice, \( \alpha \) being an empirical constant. If \( T_{el} >> T \), this relation reduces to \( kT_{el} = \sqrt{\alpha IV} \), so

\[
W(\nu) \sim \exp\left(\frac{-h\nu}{\sqrt{\alpha IV}}\right).
\]

Both this relation and the relation \( W(\nu) \sim \exp\left(\frac{-h\nu}{kT_{el}}\right) \) have been confirmed experimentally for films of small particles in figure 7.2 [Bischoff]. The electron emission also follows the appropriate variation with \( IV \).

Figure 7.2. Left: Intensity of electroluminescence (EL) vs. reciprocal bias voltage measured at three different photon energies, \( h\nu \). Because the resistance is constant, the denominator \( \sqrt{IV} = \sqrt{V^2 / R} \propto V \). Right: Luminescence spectra from several devices of Au islands on various substrates [Bischoff].
Chapter 7: Light emission from hot electrons in atomic size contacts

7.2 Results from contacts formed in the STM

We used a UHV STM to make our contacts because emitted electrons can optically excite, or photons can be absorbed by, contamination on the surface [Dan’ko]. Contacts were formed by the controlled movement of a sharp tungsten STM tip by a few angstroms into an atomically flat gold surface. Drawings of likely mechanisms are shown in figure 7.3.

![Diagram of atomic neck formation](image)

**Figure 7.3. Stages in the formation of atomic necks.** The tip is moved from tunnelling (1) towards the surface until contact is made (2). The tip may be dug further into the surface (3). Upon withdrawal, the diameter of the neck decreases atom by atom (4, 5) leading to a clean break, or the withdrawal may be disordered (6). In these simple diagrams, the forces between the body of the tip (not shown) and the sample are ignored.

The gold sample used was deposited epitaxially onto mica heated to ~500°C in another UHV evaporator, and STM images showed flat-topped islands. The surface was then cleaned by sputtering in the sample preparation chamber and then heated for over 20 hours, again at ~500°C to produce a surface with an r.m.s. height variation of 6Å over a 5000Å square. Its topography is shown in figure 7.6a.
We measured the current through the contact with a logarithmic current to voltage converter ("amplifier"), while simultaneously measuring the light emitted with the S20 photomultiplier tube whose spectral response is plotted in figure 4.9. The amplifier has a fairly linear response of log(Input current) vs Output Voltage, and works from ~0.1nA to ~20mA. The currents were calculated from the logged voltages, and are prone to no more than a few percent error.

The tip could be pushed into the surface (from its normal tunnelling separation of ~6Å) at a constant rate, or inserted then pulled back steadily, or just inserted into the surface by a set displacement. The tip voltage may be ramped or be left constant. However, due to the high van der Waals and electrostatic forces present, the tip would often ‘snap’ into the surface meaning that the position of the tip cannot be guaranteed. Some evidence of steps in the conductance as the tip is pulled out could be seen in figure 7.4. Each step is of size $2e^2/h$, corresponding to one atom in the area of contact. The ‘quantum of conductance’ arises from the elastic scattering of an obstacle.

![Figure 7.4](image.png)

**Figure 7.4.** Tip inserted 20Å from the tunnelling position, 0Å, into an Au surface and pulled back by 70Å (left to right). Steps in the conductance are evident, due to changes in the number of atoms in the diameter of the neck. Non-integral steps may represent disorder in the contact.
Figure 7.5. W tip with applied bias of 1.5V inserted 10Å into an Au surface from tunnelling (0Å) and withdrawn at a constant rate by a total of 30Å. Bottom trace: fit to raw photon emission data. Dark counts are included. Top trace: current, which continues falling below 10^{-7} Amps.

Figure 7.6a. STM image, lateral size 5000Å and peak to peak height 49Å, taken at 1.9nA tunnel current and 1.5V tip bias immediately before the contact of figure 7.5 was made.

Figure 7.6b. STM image, lateral size 5000Å and peak to peak height 110Å, taken at 1.9nA tunnel current and 1.5V tip bias immediately after the contact of figure 7.5 was broken.
Figure 7.7. W tip pressed into atomically flat Au and 1.5V bias on the tip. Open circles are current plotted on a logarithmic scale, saturating the amplifier at 14mA. Filled circles are photons per second detected.

Figure 7.8. STM image, lateral size 4000Å and peak to peak height of 316 Å, taken at 3.4nA tunnel current and 0.3V tip bias showing a crater formed immediately after the contact of Figure 7.7 was broken. The image is edge-enhanced.
Upon application of 1.5V across the contact we were able to observe photon emission. Figure 7.5 shows a small signal superimposed on the dark counts of the tube, and figures 7.6a and 7.6b show STM images of the surface before the contact was made and after the contact was broken. In figure 7.7, a stronger light emission signal is seen, until the current saturates because the size of the contact has become too large, due to forces pulling the tip towards the surface. This and other results show a size restriction to the light emission – for 1.5V bias across the contact, currents of ~1mA show the strongest light emission. This corresponds to a contact size of a few atoms in diameter. At larger sizes, the electron temperature is reduced because the energy is lost more effectively to the lattice surrounding the contact. At even larger sizes (> $L_0$), after saturation of the amplifier, melting will occur as is evident in the STM image, figure 7.8. The contact has reached size $L_0$ in one direction. At this size or larger, losses to the bulk must be high enough to cause melting for voltages of this order.

![Figure 7.5](image)

**Fig 7.9.** Photons collected from a Pt wire being drawn apart in air. Top trace: current, bottom trace: photons. Data courtesy of Prof. N. Garcia, Universidad Autónoma de Madrid.

The phenomenon of light emission is not limited to UHV conditions or to any specific metal. Other data from necks in wires drawn apart in air shows similar light emission for all metals used (Al, Au, Pt, W). Figure 7.9 shows unpublished data for a Pt wire.
Figure 7.10. Data taken for a varying sample voltage with a single atom contact. Upper line, current; dashed line, current relating to one quantum. Lower line: quantum efficiency.

Figure 7.11. Logarithmic plot of the data in figure 7.9. Slope = -0.057, so $\alpha = 3.8\times10^{-35}$ using $h\nu = 2.2eV$. 
Figure 7.10 shows a voltage ramp from 2.0 to 1.2 Volts. The current corresponds to one quantum of conductance, \(2e^2/h = 1/(12.9k\Omega)\), or a single atom contact. The quantum efficiency (defined as detectable photons emitted per travelling electron) falls from \(\approx 1.5 \times 10^{-7}\) towards zero, because the electron temperature drops. In figure 7.11, a logarithmic plot of the optical power with respect to \(1/\sqrt{IV}\) gives a straight line, as \(\ln(W) = -\frac{h\nu}{\alpha IV} + \text{constant}\) from (7.5). It is fortunate that the photomultiplier is actually measuring photons with energies \(\approx 2.2 \pm 0.5\text{eV}\) during this ramp, hence the straight line as \(\nu\) is almost constant. So we can deduce \(\alpha\) from the slope and plot the approximate electron temperature from the relation \(kT_{el} = \sqrt{\alpha IV}\), in figure 7.12.

![Figure 7.12. Electron temperature for the data in figure 7.9 using \(\alpha=3.8 \times 10^{-35}\) and \(h\nu=2.2\text{eV}\).](image)

To be more exact in this experiment, we placed a filter centred at 2.5eV in front of the photomultiplier tube, and ramped the voltage. In this trial, a conductance of \(\approx 0.4\) quanta was measured, probably meaning that there was some disorder of the atoms in the contact [Pascual], but it is of atomic dimensions nevertheless. A log plot is shown in figure 7.13, and a value for \(\alpha\) is deduced from the slope. An electron temperature is plotted in figure 7.14. Temperature plots derived in the same way from other contacts are shown in figure 7.15.
Figure 7.13. Log plot of detected photons to determine the electron temperature. The range of counts per point is 12–94, and the slope is -0.079, so $\alpha = 2.5 \times 10^{-35}$.

Figure 7.14. Electron temperature plotted as a function of bias, using the relation $kT_{el} \approx \sqrt{\alpha IV}$. 
Figure 7.15. Plots of electron temperature for various contacts. (a) $\alpha = 4.0 \times 10^{-35}$, (b) $\alpha = 4.9 \times 10^{-35}$, (c) $\alpha = 5.9 \times 10^{-35}$. Conductances in quanta are (a) $\approx 0.65$, (b) $\approx 1$, (c) $\approx 0.15$. 
Chapter 7 : Light emission from hot electrons in atomic size contacts

7.3 Analysis

The mechanism responsible for photon emission is given as spontaneous emission from hot electrons, but a number of other possible mechanisms should also be considered. The first is ohmic heating of the lattice, as it would give the same spectrum as for electron emission. However, the temperatures measured are far higher than the melting point of Gold (1336K) and most of the contacts were stable for seconds. Also figure 7.6b, where the surface has a column of height ~50Å and diameter <100Å, shows that melting has not occurred – the surrounding lattice is intact. When the contact size reaches $L_0$ (200nm), much greater damage occurs. Figure 7.8 shows a deep crater, most probably formed by melting.

The second potential mechanism would be some form of photon emission from de-excitation of electrons. Hot electrons can produce photons, but this is impossible because the energy of the emitted photons is limited to the energy of the electrons above the Fermi energy of the substrate – i.e. just the voltage across the junction. Some of the photons produced had energies in excess of the electrons providing the excitation – 2.5eV light was emitted by electrons with energy ~1eV (i.e. ~1Volt across the junction). Another form of electron de-excitation would be luminescence from one of the materials (tip, sample, or contamination). However, any such excitation of electrons from one energy level to another would occur at a specific energy, so would not result in the spectrum of (7.5), the energy of the photons would also be limited by the voltage.

A third mechanism, that of surface plasmon de-excitation, would also have a spectrum not of the form required to give a straight line in figure 7.13. Photon emission excited by electron tunnelling gives spectra like those in figure 1.4, and surface plasmon excitation by ballistic electrons would occur at around 3.2eV for gold [from Lynch and Hunter], so the photons would have the same energy. It can therefore only be a mechanism from high temperature electrons which is responsible.

Although the vanishing of the electron-phonon interaction was highlighted for particles, it clearly happens for contacts or ‘necks’ in wires. This is understandable if the length of the contact as well as the diameter is $<L_0$, shown in figure 7.16. At some point further up the tip, the diameter is $>L_0$, but current densities are only $\sim 10^9$ Amps m$^{-2}$ here.
Chapter 7 : Light emission from hot electrons in atomic size contacts

Figure 7.16. Left: idealised contact with width $L$ and a known volume. Right: More likely contact with area $A$, formed by an STM tip. The effective volume and surface area are not known, as the hot electrons and photon emission will be localised close to the contact.

We can investigate the thermal transport from the contact, to some extent, by considering equations (7.3) and (7.4).

(7.3) leads to: $K_{el} \frac{\partial^2 T_{el}}{\partial r^2} \approx IV / \text{volume}$

and assuming a temperature distribution of $T_{el}(r) = T_{el}(0) \exp(-r / r_{el})$

where $r$ is the distance from the source and $r_{el}$ is a characteristic thermal diffusion length for electrons, we can write

$K_{el} \frac{T_{el}(0)}{r_{el}^2} \approx IV / \text{volume}$

For the contact in figure 7.14, which is likely to consist of the order of 10 atoms, and $T_{el}(0) \approx 3000K$, $r_{el} \approx 10\text{Å}$.

To make an analysis for the lattice equivalent, we must first estimate $\beta'$ in equation (7.4). An analysis of electron-surface phonon interactions gives [Belotskii]
\[ \beta' \approx n \frac{v_F}{L} \frac{m}{M} \lambda_i^2 k (n_a)^{2/3} \] 

(7.6)

where \( n \) is the electron concentration, \( v_F \) is the Fermi velocity, \( L \) is the particle size, \( m \) is the electron mass, \( M \) is the atom mass, \( \lambda_i \) is the Broglie wavelength of an electron, \( n_a \) is the atom concentration and \( k \) is the Boltzmann constant. So a value of \( \beta' \) for a contact of ~10 atoms would be \( \approx 10^{15} - 10^{16} \text{ J m}^{-3} \text{ s}^{-1} \text{ K}^{-1} \). This compares with a figure one order of magnitude higher for \( \beta \) in the bulk [Anisimov]. The smaller the size of particle or contact, \( L \) in equation (7.6), the higher the losses to the surface phonons, but even this value of \( \beta' \) when inserted into equation (7.3) represents less than \( 10^{-4} \) of the power dissipation - the rest is by electron transport.

It is not immediately clear why some contacts have higher electron temperatures than others for similar voltages (figures 7.14 and 7.15), i.e. why \( \alpha \) changes. If the contact area is low, losses by electron transport will be reduced. The photon emission may be written as

\[ W = C \exp \left( -\frac{h\nu}{\sqrt{\alpha IV}} \right) \]

The parameter \( C \) should be proportional to the outer surface area of the contact, as it is at this interface where the electrons lose energy to create photons.

Parameters for the four contacts are tabulated in figure 7.17. There appears to be some relation between \( \alpha \) and \( C \) – as \( \alpha \) rises (i.e. the temperature rises as \( kT = \sqrt{\alpha IV} \), and so the losses must be getting smaller), \( C \) (and hence the outer surface area) becomes smaller. No obvious relation is apparent between \( \alpha \) and \( G \), the conductance in quanta. This would either lead us to believe that the loss mechanism is predominantly that of surface phonon excitation as opposed to electron transport, or that the parameter \( C \) may be a better guide to the contact area than the conductance. As we know that the former is not true, it must be the latter.

<table>
<thead>
<tr>
<th>Contact</th>
<th>( \alpha/10^{-35} )</th>
<th>( C )</th>
<th>( G )</th>
</tr>
</thead>
<tbody>
<tr>
<td>fig. 7.14</td>
<td>2.5</td>
<td>1.5x10^7</td>
<td>0.4</td>
</tr>
<tr>
<td>fig. 7.15a</td>
<td>4.1</td>
<td>5.4x10^5</td>
<td>0.65</td>
</tr>
<tr>
<td>fig. 7.15b</td>
<td>4.9</td>
<td>2.3x10^4</td>
<td>1</td>
</tr>
<tr>
<td>fig. 7.15c</td>
<td>7.5</td>
<td>1.4x10^4</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Figure 7.17. Table of parameters relating to four contacts to determine whether there exists any systematic relation between them. Units are not important.
Returning to the determination of the lattice thermal diffusion length, we have estimated \( \beta' \) as \( \approx 10^{15}-10^{16} \) \( \text{J m}^{-3} \text{s}^{-1} \text{K}^{-1} \), and using equation (7.3) we are able to estimate a length of \( \approx 100\angstrom \). Because the electron length (10Å) is low, the electrons’ energy can be dispersed before bulk phonons can be excited. Also, the high lattice length means that any lattice energy can be dispersed a long way from its source at the surface of the contact.

What is not clear is whether a long wire with diameter \(< L_0 \) will be more particle-like or bulk-like in its electron-phonon interactions. Along the length of the wire, any electron wavelength can occur, but if the diameter is \( < L_0 \) then the prevalence of conducting electrons with harmonics of wavelength \( \geq L_0 \) will be low, as is demonstrated in figure 7.17. So there should be a regime somewhere between particles and the bulk for this one-dimensional system, where electrons are able to have temperatures above the melting temperature, and maximum current densities are higher than for the bulk provided there is enough thermal contact between the wire and its substrate. The thinner the wire, the more particle-like it should be. This result will be of major importance to the electronics industry as diameters of connecting wires will fall below \( L_0 \).

Figure 7.17. Schematic of electron paths through a 1D wire with diameter, \( L << L_0 \). The narrower the wire, the smaller the proportion of electrons harmonics with wavelengths \( \geq L_0 \).
Chapter 8
Conclusion

The STM is an established surface science tool, combining unprecedented resolution with real space mapping. There are two major drawbacks with the technique: that it must be performed on conducting surfaces, and that it gives no chemical information. Photon emission, however, is able to probe the inelastic channel, which for metals contains invaluable chemical information. It has been shown by Berndt and Gimzewski that flat metal surfaces produce different emission spectra.

Real surfaces may not be flat, so it is important to know how curvature changes the yield of photons, and the emission spectrum. For this reason, a two sphere model was developed. The mechanism for photon emission, that of plasmon excitation followed by either radiation or dielectric loss, was split into a number of separate problems so that the dependence of the overall photon emission on materials and curvature is clear.

Experiments were performed on small silver particles, and for what was later believed to be a silver tip it was found approximately that the efficiency, \( \eta \propto R_{\text{tip}} R_{\text{sample}} \). Light emission was observed from clusters consisting of as few as \( \sim 30 \) atoms, and the first chemically specific photon maps were presented, which distinguished silver particles from carbide deposits. Emission was also seen on the metallic-like Si(111)-7x7 surface.

The new theory compared well with spectra taken on flat surfaces, then when extended to two spheres it corroborated the result \( \eta \propto R_{\text{tip}} R_{\text{sample}} \). It was predicted that the onset of emission would occur at a lower bias for gold particles than for silver particles, and this was confirmed experimentally. So this model is accurate enough to make predictions about how sample curvature will affect spectra.

Spectra characteristic to a particular metal will be produced if the modes do not move noticeably when curvature varies, i.e. when the modes do not move far away from \( \varepsilon_{\text{real}} = -1 \). It was predicted that this should occur for small radius tips and samples, and the movement of
modes should be less pronounced for tungsten tips. However, as these tips may produce low count rates, a compromise of a silver tip with radius ~100Å was thought to be best for distinguishing metal surfaces on their spectra. These spectra should also be taken at a sufficiently high bias to remove the effect of the quantum cut-off, where the photons are restricted to be of less energy than the tunnelling electrons.

It was then demonstrated that light is emitted from atomic size metal contacts and the variation of photon counts with input power fits extremely well with a model of spontaneous emission from high temperature electrons. This relation was used to deduce the electron temperature as a function of applied voltage. It was also shown that below a critical size, very high currents densities can occur for these contacts because the electrons transfer far less energy to the lattice.

Both the light emission from the heated contacts, and that from metals by tunnelling demonstrate the importance of this extra channel of information. Light emission from the STM offers an unintrusive chemical probe with potential in a number of areas in nanoscale physics and technology, and the increase in current densities in small metallic contacts may prove to be of significance to the semiconductor industry as it heads towards nanometre scale structures and devices.
Bibliography


Appendix: FORTRAN program

c **************************************************************************
c MAIN PROGRAM

choose real 8 (double precision) for all variables beginning with A-H and O-Z, all others (J-N) are integers
IMPLICIT REAL*8 (A-H,O-Z)
COMMON /C1/F(100),E1(100),E2(100)
COMMON /C2/Q(0:200),Q2(0:200),Q3(0:200),Q4(0:200),R1(0:200)
COMMON /C3/R(0:200),R3(0:200),R4(0:200),P1(0:200),P2(0:200),P3(0:200)
COMMON /C4/P(0:200),QP1(0:200),QP2(0:200),QP3(0:200),QP4(0:200)
COMMON /C5/EIM(100),E1IM(100)

EXTERNAL BISECT

DIMENSION VRAD(0:200,0:200),VINEL(0:200,0:200),VDIEL(0:200,0:200)
DIMENSION XA(4),YA(4),YB(4),SPECTRUM(0:200),CURR(0:200),BROAD(0:200)
REAL*8 MU1,MU2,I1,I2,LEFT,NUMPHOTONX,NUMPHOTONZ
REAL*8 E(4),E1(4),E2(4),SPECTRUM(0:200),CURR(0:200),BROAD(0:200)

INTEGER TRUTH

SMIN=5.0D0
SMAX=5.01D0
SSTEP=1.0D0

VSTART=-1.6D0
VEND=-4.001D0
VSTEP=-0.2D0

E=1.6021892D-19
EPS0=8.854187818D-12
PI=3.141592653589793D0
HBAR=1.054589D-34

OPEN(UNIT=28,FILE='/usr/people/andyd/WSss.240.txt')
OPEN(UNIT=29,FILE='/usr/people/andyd/WSsp.240.txt')
OPEN(UNIT=30,FILE='/usr/people/andyd/WSsc.240.txt')
OPEN(UNIT=31,FILE='/usr/people/andyd/WSse.240.txt')

READ(UNIT=19,*)K
READ(UNIT=19,*)F(I)
READ(UNIT=19,*)E1(I)
READ(UNIT=19,*)E2(I)
READ(UNIT=19,*)E1IM(I)
READ(UNIT=19,*)E2IM(I)

11 CONTINUE

NMAX is maximum mode number - the cut-off in the large matrix
NMAX=100

travelling surface plasmon energy, where epsilon=-1 (energy of 'NMAX=infinity' mode); energy where epsilon=-2
EPSMINUS1=3.68172D0
EPSMINUS2=3.51D0
DO 800 S=SMIN,SMAX,SSTEP

DO 809 RAD1=20.0D0,320.1D0,20.0D0
DO 808 RAD2=10.0D0,160.1D0,10.0D0
CALL WORKOUTMU(S,RAD1,RAD2,MU1,MU2,A)

c 'old' energy: start at low energy (here 1.5eV) and increase slowly to find where product of determinants goes through zero
GOLD=1.5D0

new' energy is 0.1% of the way to EPSMINUS1; step should be small so as not to miss out two modes close together
G=(EPSMINUS1-GOLD)*1.0D-3+GOLD
PRODUCTOLD=0.0D0
CALL PRODUCTOFDETS(MU1,MU2,G,PRODUCT,NMAX)

solve matrix for MODE=0.1.2...
DO 101 MODE=0,NMAX

TRUTH being zero means don't find two points between which PRODUCT goes through zero
TRUTH=0

DO 401 WHILE(TRUTH.EQ.0)

if it's the first MODE...
IF (PRODUCTOLD.EQ.0.0D0)THEN
CALL PRODUCTOFDETS(MU1,MU2,G,PRODUCT,NMAX)
GSTORED=G
G=G+(G-GOLD)
GOLD=GSTORED
PRODUCTOLD=PRODUCT
CALL PRODUCTOFDETS(MU1,MU2,G,PRODUCT,NMAX)

else if (PRODUCT/PRODUCTOLD).LT.0.0D0)THEN
LEFT=GOLD
RIGHT=G
GOLD=G
PRODUCTOLD=PRODUCT

this calls the 'BISECT' routine
GZERO=BISECT(MU1,MU2,LEFT,RIGHT,1.0D-16,NMAX)

write mode number (0,1,...) and energy, and product should alternate +1,-1,...
WRITE(6,*)MODE,GZERO,INT(PRODUCT)

now move a little past the zero so that the next mode can be found
GOLD=GZERO+ABS(EPSMINUS1-GZERO)*1.0D-6+1.0D-8
G=ABS(EPSMINUS1-GOLD)*1.0D-4+GOLD+1.0D-8
CALL PRODUCTOFDETS(MU1,MU2,GOLD,PRODUCT,NMAX)
PRODUCTOLD=PRODUCT
CALL PRODUCTOFDETS(MU1,MU2,G,PRODUCT,NMAX)
TRUTH=1

ELSE
increment another small step towards EPSMINUS1
GSTORED=G
G=G+(G-GOLD)
GOLD=GSTORED
PRODUCTOLD=PRODUCT
CALL PRODUCTOFDETS(MU1,MU2,G,PRODUCT,NMAX)

(i.e. go and work out PRODUCT again)

ENDIF
CALL PRODUCTOFSUMS(MU1,MU2,GZERO,UNIMPORTANT, NMAX)

VARRAY(2*NMAX, MODE) = 1.0D0
VARRAY(2*NMAX + 1, MODE) = -QP1(NMAX)/QP2(NMAX)

DO 201 I = NMAX - 1, 0, -1
DET = QP1(I)*QP4(I) - QP2(I)*QP3(I)
VARRAY(2*I, MODE) = (VARRAY(2*I + 2, MODE)*(QP2(I)*R3(I) - QP4(I)*R1(I)) + 
VARRAY(2*I + 3, MODE)*(QP2(I)*R4(I) - QP4(I)*R2(I)))/DET
VARRAY(2*I + 1, MODE) = (VARRAY(2*I + 2, MODE)*(QP3(I)*R1(I) - QP1(I)*R3(I)) + 
2VARRAY(2*I + 3, MODE)*(QP3(I)*R2(I) - QP1(I)*R4(I)))/DET

201 CONTINUE

GSET(MODE) = GZERO
HW = GSET(MODE)

DO 615 I = 0, NMAX
AN = VARRAY(2*I, MODE)
BN = VARRAY(2*I + 1, MODE)
SUM = SUM + BN*(AN + BN*EXP((2.0D0*I + 1.0D0)*MU1)) + 
AN*(BN + AN*EXP(-MU2*(2.0D0*I + 1.0D0)))
IF((I.GE.1).THEN
IF(((SUM - SUMOLD)/SUM).LE.(1.0D-3))GO TO 614
ENDIF
SUMOLD = SUM

615 CONTINUE

ERROR1 = CHECK1/CHECK3*100.0D0
ERROR2 = CHECK2/CHECK4*100.0D0
write out errors if they're bigger than 0.1%
IF(ERROR1.GE.0.1D0)WRITE(6,*)'ERROR1(%)= ',ERROR1
IF(ERROR2.GE.0.1D0)WRITE(6,*)'ERROR2(%)= ',ERROR2

factor to normalise An and Bn to ground state energy
FACTOR=2.5D9/(PI*A*EPS0)*HWJ/SUM

VARRAYNORM is array of normalised An and Bn
DO 1199 I=0,2*NMAX+1
    VARRAYNORM(I,MODE)=VARRAY(I,MODE)*SQRT(FACTOR)
1199 CONTINUE

SUM01=0.0D0
SUM02=0.0D0
SUM04=0.0D0
PHI1=0.0D0
PHI2=0.0D0
DO 378 I=0,NMAX
    PHI1=PHI1+VARRAYNORM(2*I,MODE)+VARRAYNORM(2*I+1,MODE)
    PHI2=PHI2+(DEXP(-(I+0.5D0)*MU1)*VARRAYNORM(2*I,MODE))+(DEXP((I+0.5D0)*MU1)*VARRAYNORM(2*I+1,MODE))
SUM01=SUM01-(2.0D0*I+1.0D0)*VARRAYNORM(2*I,MODE)
SUM02=SUM02+(2.0D0*I+1.0D0)*VARRAYNORM(2*I+1,MODE)
378 CONTINUE

phi should the dipoles need to be written on screen...
WRITE(6,*)'DIPOLES ',SUM01,SUM02

phi should the potentials need to be written on screen...
WRITE(6,*)'PHI1 ',PHI1*DSQRT(2.0D0),' PHI2 ',PHI2*DSQRT(2.0D0)

now find the radiation rate
CALL DIPOLEPOWER(VARRAYNORM,MU1,MU2,A,NMAX,VRAD)

then the dielectric loss rate
CALL DIELECTRICLOSS(VARRAYNORM,MU1,MU2,A,GSET,NMAX,VDIEL,BROAD)

for the appropriate bias voltage...
DO 1491 V=VSTART,VEND,VSTEP

...find the inelastic rate
CALL INELASTIC(GSET,V,VARRAYNORM,S,MU1,MU2,A,NMAX,VINEL,ELASTIC)
WRITE(31,*)'V,ELASTIC

"Image" peak above travelling surface plasmon frequency, EPSMINUS1 (where eps=-1)
PEAK=EPSMINUS2+EPSMINUS2-GSET(MODE)

if mode is above epsilon=-2, then it decouples: ignore it
IF(GSET(MODE).GE.EPSMINUS2)GO TO 671

DE is broadening at mode energy; DEE is variable broadening: the broadening at this frequency
DE=(VRAD(MODE,IF)*((GSET(MODE)/FREQ)**3)+BROAD(MODE))*HBAR/E
DEE=(VRAD(MODE,IF)+VDIEL(MODE,IF))*HBAR/E

if there’s no current at this frequency
IF((VINEL(MODE,IF).EQ.0.0D0).AND.(VRAD(MODE,IF).EQ.0.0D0))GO TO 671

FIRST is contribution to photon spectrum, SECOND is contribution to current spectrum
FIRST=VRAD(MODE,IF)*VINEL(MODE,IF)/(VRAD(MODE,IF)+VDIEL(MODE,IF)-VINEL(MODE,IF))
SECOND=(VRAD(MODE,IF)+VDIEL(MODE,IF))/(VRAD(MODE,IF)+VDIEL(MODE,IF)-VINEL(MODE,IF))*VINEL(MODE,IF)

Lorentzian factor
GLOR=1.0D0/(2.0D0*ATAN((PEAK-GSET(MODE))/DE)+ATAN(2.0D0*GSET(MODE)/DE)-ATAN(2.0D0*PEAK/DE))
GLOR=DE*DE/(DEE*DEE+4.0D0*(FREQ-GSET(MODE))*(FREQ-GSET(MODE)))*DEE*DEE+4.0D0*(FREQ-PEAK)/(FREQ-PEAK))

safeguard:
IF (GLOR.LE.0.0D0)GLOR=0.0D0

add components from each mode onto spectra
SPECTRUM(IF)=SPECTRUM(IF)+FIRST*GLOR
CURR(IF)=CURR(IF)+SECOND*GLOR

if required, write three rates on screen
WRITE(6,*)VINEL(MODE,IF),VDIEL(MODE,IF),VRAD(MODE,IF)

write total photons at this frequency (i.e. spectrum) into file
WRITE(28,*)IF*0.05D0,SPECTRUM(IF)

add spectral components onto total photons, total current
PHOTONS=PHOTONS+SPECTRUM(IF)*0.05D0
CURRENT=CURRENT+CURR(IF)*0.05D0

write out total photons, total current to file. can include other parameters
WRITE(29,*)V,PHOTONS
WRITE(30,*)V,CURRENT

write bias, separation, radii, elastic & inelastic currents, total photons on screen
WRITE(6,*)VOLTS;'V='S
WRITE(6,*)R1 'R1' R2 'R2' RAD1
WRITE(6,*)ELASTIC='ELASTIC
WRITE(6,*)INELASTIC='INELASTIC='PHOTONS='PHOTONS
WRITE(6,*)

END

the end of the main program. The rest is the subroutines....
this subroutine calculates \( \mu_1, \mu_2, \) & \( a \) from radii and separation

**SUBROUTINE WORKOUTMU**

\[ \text{IMPLICIT REAL*8 (A-H,O-Z)} \]
\[ \text{REAL*8 MU1,MU2,LG} \]

\[ \text{if RAD2 is put } = 0.0 \text{D0 in main program, it is understood that this refers to a plane. then use alternative calculations} \]
\[ \text{IF (RAD2.LT.0.1D0) THEN} \]
\[ \quad A = \text{DSQRT}(S*(S+RAD1+RAD1)) \]
\[ \quad \text{ELSE} \]
\[ \quad \text{else for sphere-sphere...} \]
\[ \quad \text{SIMP} = S*(S+2.0D0*(RAD1+RAD2)) \]
\[ \quad A = \text{DSQRT}(\text{SIMP}*(\text{SIMP}+4.0D0*RAD1*RAD2)))/(2.0D0*(S+RAD1+RAD2)) \]
\[ \text{ENDIF} \]

\[ \text{IF (RAD2.LT.0.1D0) THEN} \]
\[ \quad \text{approximate \( \mu_2 \) as } 1e(-7) \text{ for plane, as this has better numerical stability than \( \mu_2 = 0 \)} \]
\[ \quad \text{MU2} = 1.0D-7 \]
\[ \quad \text{ELSE} \]
\[ \quad \quad \text{MU2} = -\text{DLOG}((A+\text{SQRT}(\text{RAD2}^2+\text{RAD2}^2)))/\text{RAD2} \]
\[ \quad \text{ENDIF} \]

\[ \text{MU1} = \text{DLOG}((A+\text{SQRT}(\text{RAD1}^2+\text{RAD1}^2)))/\text{RAD1} \]

\[ \text{write(6,'(*A= ',A,'; \mu_1= ',MU1,'; \mu_2= ',MU2 \text{ RETURN} \}

**END**

**SUBROUTINE POLINT**

\[ \text{IMPLICIT REAL*8 (A-H,O-Z)} \]
\[ \text{DIMENSION C(10),D(10),XA(L),YA(L)} \]
\[ \text{INTEGER Q} \]

\[ \text{NS} = 1 \]
\[ \text{DIF} = \text{ABS}(X-XA(1)) \]
\[ \text{DO 2222 I = 1,4} \]
\[ \quad \text{DIFT} = \text{ABS}(X-XA(I)) \]
\[ \quad \text{IF (DIFT.LT.DIF) THEN} \]
\[ \quad \quad \text{NS} = I \]
\[ \quad \text{DIF} = \text{DIFT} \]
\[ \quad \text{ENDIF} \]
\[ \quad C(I) = YA(I) \]
\[ \quad D(I) = YA(I) \]
\[ 2222 \text{ CONTINUE} \]
\[ \text{Y} = YA(\text{NS}) \]
\[ \text{NS} = \text{NS} - 1 \]
\[ \text{DO 3333 Q = 1,3} \]
\[ \text{DO 4444 I = 1,4-Q} \]
\[ \quad HO = XA(I) - X \]
\[ \quad HP = XA(I+Q) - X \]
\[ \quad W = C(I+1) - D(I) \]
\[ \quad \text{DEN} = HO*HP \]
\[ \quad \text{IF (DEN.EQ.0.0D0) THEN} \]
\[ \quad \quad \text{WRITE(*,'(\*A= \*,A; \mu_1= \*,MU1; \mu_2= \*,MU2 \text{ RETURN)"} \]
\[ \quad \text{ENDIF} \]
\[ \quad C(I) = (HO*W)/DEN \]
\[ \quad D(I) = (HP*W)/DEN \]
\[ 4444 \text{ CONTINUE} \]
IF (2*NS.LT.4-Q) THEN
    DY=C(NS+1)
ELSE
    DY=D(NS)
    NS=NS-1
ENDIF
Y=Y+DY
3333 CONTINUE
RETURN
END

c*****************************************************************************
cthis finds the energy where PRODUCT is zero, by continually bisecting the energy axis until it exceeds accuracy. starts with
cone energy where PRODUCT is negative, and a close second energy where it is positive. modified from Numerical recipes

FUNCTION BISECT(MU1,MU2,XLEFT,XRIGHT,XACC,NMAX)
IMPLICIT REAL*8 (A-H,O-Z)
COMMON /C1/F(100),E1(100),E2(100)
PARAMETER(JMAX=100)

CALL PRODUCTOFDETS(MU1,MU2,XRIGHT,PRODUCT,NMAX)
FMID=PRODUCT
CALL PRODUCTOFDETS(MU1,MU2,XLEFT,PRODUCT,NMAX)
FF=PRODUCT

IF(FF/FMID.GE.0.0D0) THEN
    WRITE(6,*)'ROOT MUST BE BRACKETED FOR BISECTION',FMID,FF
ENDIF

IF(FF.LT.0.0D0) THEN
    BISECT=XLEFT
    DX=XRIGHT-XLEFT
ELSE
    BISECT=XRIGHT
    DX=XLEFT-XRIGHT
ENDIF

DO 1111 J=1,JMAX
    DX=DX*.5D0
    XMID=BISECT+DX
    CALL PRODUCTOFDETS(MU1,MU2,XMID,PRODUCT,NMAX)
    FMID=PRODUCT
    IF(FMID.LE.0.0D0) BISECT=XMID
    IF(ABS(DX).LT.XACC .OR. FMID.EQ.0.0D0) THEN
        RETURN
    ENDIF
1111 CONTINUE
WRITE(6,*)'TOO MANY BISECTIONS'
END

*****************************************************************************
ccalculates the radiation rate, VRAD

SUBROUTINE DIPOLEPOWER(VARRAYNORM,MU1,MU2,A,NMAX,VRAD)
IMPLICIT REAL*8 (A-H,O-Z)
COMMON /C1/F(100),E1(100),E2(100)
COMMON /C5/E2IM(100),E1IM(100)
REAL*8 MU1,MU2,I1,I2,INTEGRALZ(100)
DIMENSION VARRAYNORM(0:400,0:400),GSET(0:200),XA(4),YA(4),YB(4)
DIMENSION VRAD(0:200,0:200)
INTEGER B, STEP, TWOSTEP  
REAL SINTH, SIN, COS  
COMPLEX COMPE, TZ(100), TR(100)

C physical constants, etc.  
B=0  
EPS0=8.854187818D-12  
E=1.6021892D-19  
Pl=3.141592653589793D0  
HBAR=1.054589734D-34  
C=2.99792458D8  
AA=A*1.0D-10

DO 650 MODE=0, NMAX  
SUM1=0.0D0  
SUM2=0.0D0  
DO 651 N=0, NMAX  
SUM1=SUM1+(2.0D0*N+1.0D0)*(VARRAYNORM(2*N+1, MODE)-VARRAYNORM(2*N, MODE))  
651 CONTINUE

C reflection from a plane: interpolate epsilon of sample (real and imaginary)

DO 911 IFREQ=11,120  
FREQ=IFREQ*0.05D0  
DO 102 N=1,87  
IF ((F(N).LE.FREQ).AND.(F(N+1).GE.FREQ)) B=N  
102 CONTINUE  
IF (B.EQ.0) PAUSE 'B=0'

DO 559 M=1,4  
XA(M)=F(B-2+M)  
YA(M)=E2IM(B-2+M)  
YB(M)=E2(B-2+M)  
559 CONTINUE

L=4  
CALL POLINT(XA, YA, L, FREQ, Y, DY)  
EPSIM2=Y  
CALL POLINT(XA, YB, L, FREQ, Y, DY)  
EPS2=Y  
I1=0.0D0  
I2=0.0D0  
COMPE=CMPLX(EPS2, EPSIM2)

C integrate from 0 to 90 degrees  
DO 888 STEP=1,91

DEGZ=REAL(STEP-1)  
THETA=(DEGZ)*PI/180.0  
SINTH=SIN(THETA)  
COSTH=COS(THETA)

TZ(STEP)=(CSQRT(COMPE)*COSTH-CSQRT(1.0-SINTH*SINTH/COMPE))/9(CSQRT(COMPE)*COSTH+CSQRT(1.0-SINTH*SINTH/COMPE))  
TR(STEP)=(COSTH-CSQRT(COMPE)*CSQRT(1.0-SINTH*SINTH/COMPE))/9(COSTH+CSQRT(COMPE)*CSQRT(1.0-SINTH*SINTH/COMPE))

C in this case, have no reflection from plane  
INTEGRALZ(STEP)=SINTH*SINTH*SINTH
I1=I1+INTEGRALZ(1)+INTEGRALZ(91)
I1=I1+4.0D0*INTEGRALZ(90)
DO 777 TWOSTEP=1,44
    I1=I1+2.0D0*INTEGRALZ(2*TWOSTEP+1)+4.0D0*INTEGRALZ(2*TWOSTEP)
777 CONTINUE
I1=I1/3.0D0*PI/180.0D0*2.0D0*PI

c omega, the frequency
W=FREQ/HBAR*E
VRAD(MODE,IFREQ)=AA*AA*AA*AA*EPS0/(HBAR*C*C*C)*SUM1*SUM1*I1*W*W*W

***********************************************************************
SUBROUTINE PRODUCTOFDETS(MU1,MU2,G,PRODUCT,NMAX)
IMPLICIT REAL*8 (A-H,O-Z)
EXTERNAL POLINT
COMMON /C1/F(100),E1(100),E2(100)
COMMON /C2/Q1(0:200),Q2(0:200),Q3(0:200),Q4(0:200),R1(0:200)
COMMON /C3/R2(0:200),R3(0:200),R4(0:200),P1(0:200),P2(0:200),P3(0:200)
COMMON /C4/P4(0:200),QP1(0:200),QP2(0:200),QP3(0:200),QP4(0:200)
DIMENSION XA(4),YA(4),YB(4)
REAL*8 MU1,MU2
INTEGER B

IF ((F(I).LE.G).AND.(F(I+1).GE.G))B=I
102 CONTINUE

IF (B.EQ.0)PAUSE'B=0'
DO 555 M=1,4
    XA(M)=F(B-2+M)
    YA(M)=E1(B-2+M)
    YB(M)=E2(B-2+M)
555 CONTINUE

CALL POLINT(XA,YA,L,G,Y,DY)
EPS1=Y
CALL POLINT(XA,YB,L,G,Y,DY)
EPS2=Y
CHI1=(1.0D0-EPS1)/(1.0D0+EPS1)
CHI2=(1.0D0-EPS2)/(1.0D0+EPS2)

HCOS1=0.50D0*(DEXP(MU1)+DEXP(-MU1))
HCOS2=0.50D0*(DEXP(MU2)+DEXP(-MU2))
HSIN1=0.50D0*(DEXP(MU1)-DEXP(-MU1))
HSIN2=0.5D0*(DEXP(MU2)-DEXP(-MU2))

DO 107 N=0,NMAX+1

 Abbreviated expressions in matrix
FN1=DEXP(-2.0D0*N+1.0D0)*MUI1
FN2=DEXP((2.0D0*N+1.0D0)*MUI2)

Q, P, and R: the smaller 2x2 matrices making up the larger matrix
Q1(N)=CHI1*((2.0D0*N+1.0D0)*HCOS1-HSIN1)*FN1
Q2(N)=-(CHI1*HSIN1+(2.0D0*N+1.0D0)*HCOS1)
Q3(N)=(CHI2*HSIN2-(2.0D0*N+1.0D0)*HCOS2)
Q4(N)=CHI2*((2.0D0*N+1.0D0)*HCOS2+HSIN2)*FN2
R1(N)=-(N+1.0D0)*CHI1*DEXP(-MUI1)*FN1
R2(N)=(N+1.0D0)*DEXP(MUI1)
R3(N)=(N+1.0D0)*DEXP(MUI2)
R4(N)=-(N+1.0D0)*CHI2*FN2*DEXP(MU2)

P1(N)=-N*CHI1*FN1*DEXP(MUI1)
P2(N)=N*DEXP(-MUI1)
P3(N)=N*DEXP(MUI2)
P4(N)=N*CHI2*FN2*DEXP(MUI2)

107 CONTINUE

Q'  
QP1(0)=Q1(0)
QP2(0)=Q2(0)
QP3(0)=Q3(0)
QP4(0)=Q4(0)

PRODUCT=1.0D0

DO 109 I=1,NMAX+1

DETQP=QP1(I-1)*QP4(I-1)-QP2(I-1)*QP3(I-1)

QP1(I)=Q1(I)-(P1(I)*(QP4(I-1)*R1(I-1)-QP2(I-1)*R3(I-1))+P2(I)*(QP1(I-1)*R3(I-1)-QP3(I-1)*R1(I-1)))/DETQP
QP2(I)=Q2(I)-(P1(I)*(QP4(I-1)*R2(I-1)-QP2(I-1)*R4(I-1))+P2(I)*(QP1(I-1)*R4(I-1)-QP3(I-1)*R2(I-1)))/DETQP
QP3(I)=Q3(I)-(P3(I)*(QP4(I-1)*R1(I-1)-QP2(I-1)*R3(I-1))+P4(I)*(QP1(I-1)*R3(I-1)-QP3(I-1)*R1(I-1)))/DETQP
QP4(I)=Q4(I)-(P3(I)*(QP4(I-1)*R2(I-1)-QP2(I-1)*R4(I-1))+P4(I)*(QP1(I-1)*R4(I-1)-QP3(I-1)*R2(I-1)))/DETQP

PRODUCT=PRODUCT*ABS(DETQP)/DETQP

109 CONTINUE

RETURN

END

***************************************************************************
finds the inelastic tunnelling rate. also elastic rate

SUBROUTINE INELASTIC(GSET,V,VARRAYNORM,S,MU1,MU2,A,NMAX,VINEL,ELASTIC)
IMPLICIT REAL*8 (A-H,O-Z)
REAL*8 M,IGRAL,MU1,MU2,IGRAND(0:200)
REAL*8 IGRALE,INTEGRAND(0:200)
DIMENSION GSET(0:200),VARRAYNORM(0:400,0:400)
DIMENSION VINEL(0:200,0:200)
DIMENSION VINE(0:200,0:200)

convert the separation, S, to the same in metres as X
X=S*1.0D-10
c physical constants
EPS0=8.854187818D-12
E=1.6021892D-19
HBAR=1.054589D-34
M=9.1095D-31
PI=3.141592653589793D0
Z2=A*1.0D-10*(DEXP(MU1)-DEXP(-MU1))/(2.0D0+DEXP(MU1)+DEXP(-MU1))
Z1=A*1.0D-10*(DEXP(MU2)-DEXP(-MU2))/(2.0D0+DEXP(MU2)+DEXP(-MU2))
WRITE(6,'Z1, Z2 ',Z1,Z2)
c fermi energy of tip (L) & sample (R) in eV
EFR=5.5D0
EFL=4.9D0
c work functions of tip (L) & sample (R) in eV, then average barrier height
PHIL=4.55D0
PHIR=4.3D0
PHIEFF=0.50D0*(PHIL+PHIR-ABS(V))
UL=(EFR+PHIEFF)*E
UR=(EFL+PHIEFF-V)*E
IF (V.GE.0.0D0) THEN
UL=(EFR+PHIEFF+V)*E
ENDIF
c approximation of area to take into account varying separation
AREA=(2.0D0+S)*1.5D-19
DO 1144 N=0,20
DO 944 IFREQ=11,120
FREQ=0.05D0*IFREQ
HWJ=FREQ*E
IGRAL=0.0D0
IF (ABS(V).LE.FREQ) THEN
VINEL(N,IFREQ)=0.0D0
GO TO 944
ENDIF
c if bias voltage cannot overcome plasmon energy then integral=0
IF(ABS(V).LE.FREQ) THEN
VINEL(N,IFREQ)=0.0D0
ENDIF
c getting the correct starting values for EL and ER: energies on left and right, different for positive and negative biases
IF (V.LE.0.0D0) THEN
EL=EFL*E+HWJ+(E*V)
EL=EL-0.0D0
IF(V.GE.0.0D0) THEN
ER=EFL*E
ER=ER-0.0D0
ENDIF
DE=1.0D-2*E*(ABS(V)-FREQ)
DO 1122 ISTEP=0,100
ALPHAL=SQRT(2.0D0*M*(UL-EL))/HBAR
ALPHAR=SQRT(2.0D0*M*(UR-ER))/HBAR
SUMELEMENT=0.0D0
DO 1133 I=0,20
AN=VARRAYNORM(2*I,N)
BN=VARRAYNORM(2*I+1,N)
ANP=AN+BN*EXP((2.0D0*I+1.0D0)*MU1)
BNP=BN+AN*EXP(-MU2*(2.0D0*I+1.0D0))

C abbreviate...
B=(2.0D0*I+1.0D0)/A*1.0D10

c element of tunnelling in the gap
ELEMENT=AN*(DEXP(-((ALPHAR-ALPHAL+B)*Z1)))*DEXP((ALPHAR-ALPHAL+B)*Z2))/((ALPHAR-ALPHAL+B)+BN*(DEXP(2.0D0*I+1.0D0)=(2.0D0*I+1.0D0)/A*1.0D10)

C elements for tunnelling in electrodes
ELEMENT1=ANP*DEXP(-((ALPHAR+B)*Z2))
ELEMENT2=BNP*DEXP((ALPHAL+B)*Z1)

C (-1) to the power of I
IPOWER=4*(INT(I*0.5)-0.5*I)+1
SUMERATE=SUMERATE+IPOWER*(ELEMENT*DEXP(ALPHAR*Z1)+DEXP(ALPHAL+B)+ELEMENT2*DEXP(ALPHAL+B))

1133 CONTINUE

c integrand is some of elements over range of incremented energies
IGRAND(ISTEP)=SQRT((EL*ER)/(1.0D0/ALPHAL+1.0D0/ALPHAR))*SUMERATE

c increment the energies on left and right
EL=EL+DE
ER=ER+DE

1122 CONTINUE

IGRAL=IGRAND(0)+IGRAND(100)+4.0D0*IGRAND(99)

DO 707 I=1,49
IGRAL=IGRAL+4.0D0*IGRAND(2*I-1)+2.0D0*IGRAND(2*I)
707 CONTINUE

IGRAL=IGRAL*DE/3.0D0
VINEL(N,IFREQ)=IGRAL/(X*PI*PI*(HBAR*HBAR*HBAR))*8.0D0*AREA*E*E*M/(UL*UR)

944 CONTINUE
1144 CONTINUE

C *** now the elastic calculation ***

RANGE=E*ABS(V)
IF(V.LT.0.0D0)THEN
  IF(RANGE.GT.(E*EFR))RANGE=(E*EFR)
ENDIF
IF(V.GT.0.0D0)THEN
  IF(RANGE.GT.(E*EFR))RANGE=(E*EFR)
ENDIF
DEL=E*V*1.0D-2
EL=EFL+E

DO 882 I=0,100
ROOT=DSQRT(E+EL*V+EFR-EFL))
ALPHA=DSQRT(2.0D0*T*(UL-EL))/HBAR
INTEGRAND(I)=ROOT*A*ALPHA*ALPHA*EXP(-ALPHA*2.0D0*X)

EL=EL+DEL
IF(EL.LT.0.0D0)EL=0.0D0

882 CONTINUE
IGRALE=INTEGRAND(0)+INTEGRAND(100)+4.0D0*INTEGRAND(99)
DO 307 I=1,49
IGRALE=IGRALE+4.0D0*INTEGRAND(2*I-1)+2.0D0*INTEGRAND(2*I)
307 CONTINUE
IGRALE=IGRALE*DEL/3.0D0*AREA*HBAR*2.0D0/(PI*PI*UL*UR*X*M)
ELASTIC=IGRALE
RETURN
END

*****************************************************************************
*****************************************************************************
SUBROUTINE DIELECTRICLOSS(VARRAYNORM,MU1,MU2,A,GSET,NMAX,VDIEL,BROAD)
IMPLICIT REAL*8 (A-H,O-Z)
COMMON /C1/F(100),E1(100),E2(100)
COMMON /C5/E2IM(100),E1IM(100)
REAL*8 MU1,MU2
DIMENSION VARRAYNORM(0:400,0:400),GSET(0:200),XA(4),YA(4),YB(4)
DIMENSION RATE(0:200),Q1(0:200),Q2(0:200),VDIEL(0:200,0:200),T2(0:200)
DIMENSION EIM1(0:200),EIM2(0:200),BROAD(0:200)
INTEGER B

c physical constants
EPS0=8.854187818D-12
PI=3.141592653589793D0
E=1.6021892D-19
HBAR=1.054589D-34
AA=A*1.0D-10

ccosh and sinh of mu1, mu2
SHMU1=0.5D0*(DEXP(MU1)-DEXP(-MU1))
SHMU2=0.5D0*(DEXP(MU2)-DEXP(-MU2))
C1=0.5D0*(DEXP(MU1)+DEXP(-MU1))
C2=0.5D0*(DEXP(MU2)+DEXP(-MU2))

Q1(0)=0.5D0*DLOG((1.0D0+C1)/(C1-1.0D0))
Q1(1)=0.5D0*C1*DLOG((1.0D0+C1)/(C1-1.0D0))-1.0D0
Q2(0)=0.5D0*DLOG((1.0D0+C2)/(C2-1.0D0))
Q2(1)=0.5D0*C2*DLOG((1.0D0+C2)/(C2-1.0D0))-1.0D0

cc introduce safeguard against exploding values of Q
ICUTOFF1=0
ICUTOFF2=0
DO 711 I=2,200
Q1(I)=(2.0D0*1.0D0)*C1*Q1(I-1)-(I-1.0D0)*Q1(I-2))/REAL(I)
Q2(I)=C2*(2.0D0*1.0D0)*Q2(I-1)-(1.0D0-1.0D0)*Q2(I-2)
711 CONTINUE
IF(ABS(Q2(I)).GE.ABS(Q2(I-1))).AND.(ABS(Q2(I-1)).LE.ABS(Q2(I-2))))ICUTOFF2=I+1
IF((ABS(Q2(I)).GE.ABS(Q2(I-1))).AND.(ABS(Q2(I-1)).LE.ABS(Q2(I-2))))ICUTOFF1=I+1

711 CONTINUE
IF(ICUTOFF1.GT.0) THEN
DO 345 I=ICUTOFF1+1,200
Q1(I)=Q1(ICUTOFF1)
345 CONTINUE
ENDIF

IF(ICUTOFF2.GT.0) THEN
DO 346 I=ICUTOFF2+1,200
Q2(I)=Q2(ICUTOFF2)
346 CONTINUE
ENDIF
CONTINUE
ENDIF

DO 922 IFREQ=11,120
FREQ=0.05D0*IFREQ

c interpolate values of imaginary parts of epsilon for tip and sample

DO 633 I=1,87
IF ((F(I).LE.FREQ).AND.(F(I+1).GE.FREQ))B=I
633 CONTINUE

IF (B.EQ.0)PAUSE'B=0'

DO 522 M=1,4
XA(M)=F(B-2+M)
YA(M)=E1IM(B-2+M)
YB(M)=E2IM(B-2+M)
522 CONTINUE

L=4
CALL POLINT(XA,YA,L,FREQ,Y,DY)
EIM1(IFREQ)=Y
CALL POLINT(XA,YB,L,FREQ,Y,DY)
EIM2(IFREQ)=Y

922 CONTINUE

DO 677 MODE=0,NMAX
G=GSET(MODE)
c calculate double sums of legendre polynomials, and triple sums with Q

SUM1=0.0D0
SUM2=0.0D0
DO 733 N=0,NMAX
AN=VARRAYNORM(2*N,MODE)
BN=VARRAYNORM(2*N+1,MODE)
ANP=AN+BN*EXP((2.0D0*N+1.0D0)*MU1)
BNP=BN+AN*EXP(-MU2*(2.0D0*N+1.0D0))
PPP0=2.0D0/(2.0D0*N+1.0D0)

PPP1=PPP0
DO 744 M=0,N
AM=VARRAYNORM(2*M,MODE)
BM=VARRAYNORM(2*M+1,MODE)
AMP=AM+BM*DEXP((2.0D0*M+1.0D0)*MU1)
BMP=BM+AM*DEXP(-MU2*(2.0D0*M+1.0D0))

DFACTOR=2.0D0
IF (M.EQ.N)DFACTOR=1.0D0

PPP=PPP1
LL=0
DO 755 LL=0,2*M,2
L=N-M+LL
SUM1=SUM1-SHMU1*DFACTOR*ANP*AMP*Q1(L)*(2.0D0*L+1.0D0)*DEXP(-MU1*(N+M+1.0D0))*PPP*0.5D0
SUM2=SUM2+SHMU2*DFACTOR*BNP*BMP*Q2(L)*(2.0D0*L+1.0D0)*DEXP(MU2*(N+M+1.0D0))*PPP*0.5D0
PPP=PPP*(L+1.0D0)*(L+2.0D0)*(2.0D0*N+LL+3.0D0)*(2.0D0*M-LL)*(2.0D0*M-LL-1.0D0)
PPP=PPP*(L+1.0D0)*(L+2.0D0)*(2.0D0*N-M+LL+1.0D0)*(2.0D0*N-M+1.0D0)+LL
755 CONTINUE

PPP1=PPP1*(2.0D0*M+1.0D0)/(M+1.0D0)*(N-M)/(2.0D0*(N-M)-1.0D0)
CONTINUE

SUM1 = SUM1 + ANP*ANP*DEXP(-MU1*(2.0D0*N+1.0D0))
SUM2 = SUM2 + BNP*BNP*DEXP(MU2*(2.0D0*N+1.0D0))

DO 688 I=1,87
   IF ((F(I).LE.G).AND.(F(I+1).GE.G))B=I
688 CONTINUE

IF (B.EQ.0)PAUSE'B=0'
DO 599 M=1,4
   XA(M)=F(B-2+M)
   YA(M)=E1IM(B-2+M)
   YB(M)=E2IM(B-2+M)
599 CONTINUE

L=4
CALL POLINT(XA,YA,L,G,Y,DY)
EIM1G=Y
CALL POLINT(XA,YB,L,G,Y,DY)
EIM2G=Y

b broadening of the mode
BROAD(MODE)=PI*AA*EPS0/HBAR*(SUM1*EIM1G+SUM2*EIM2G)

DO 933 IFREQ=11,120

b dielectric loss rate
VDIEL(MODE,IFREQ)=PI*AA*EPS0/HBAR*(SUM1*EIM1(IFREQ) +SUM2*EIM2(IFREQ))
933 CONTINUE

CONTINUE

RETURN
END