Scanning Tunneling Microscopy

Wang Materials Group

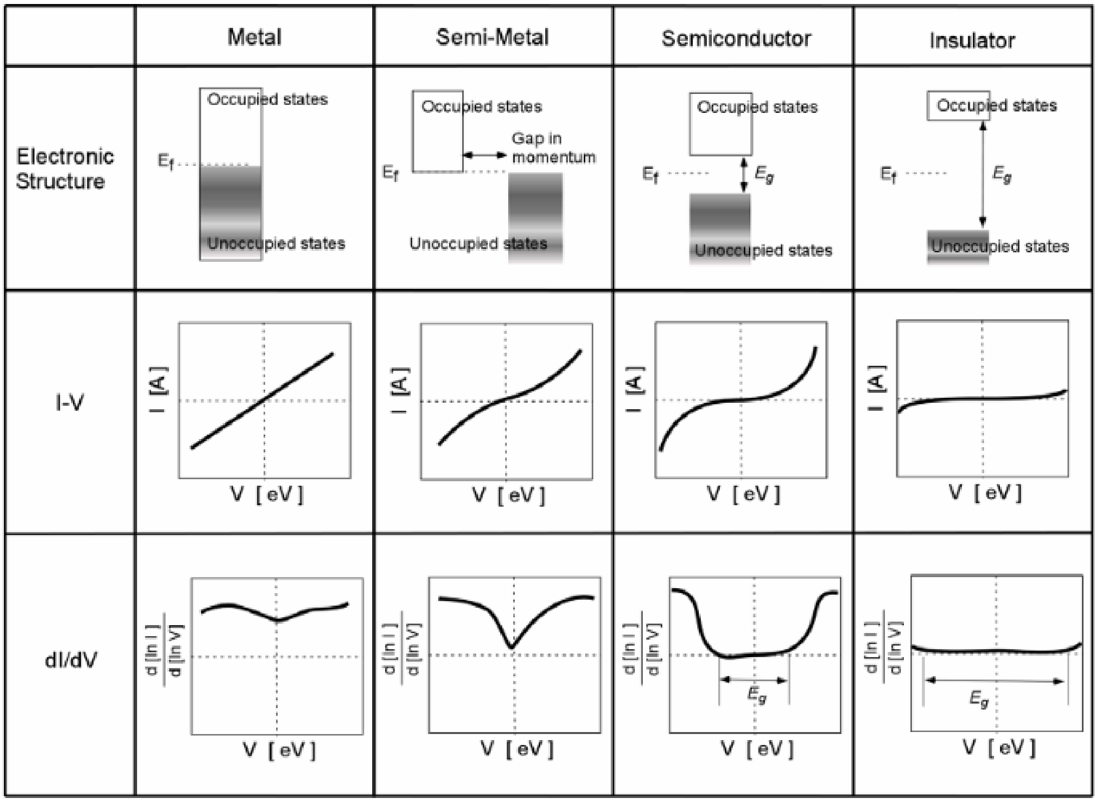
UT Austin, McKetta Department of Chemical Engineering

Exploring the theory behind scanning tunneling microscopy (STM)- some brief notes

1 General concepts

STM has two modes of operation: imaging and spectroscopy

- Imaging: constant current and constant height mode. A nice explanation and animation can be found on Wikipedia
- Spectroscopy: sweeping across V_{bias} or varying tip height from sample to obtain I-V characteristics; location specific, often based on STM image
  - a simple graphical summary from here



- For $+V_{bias}$, electrons flow from tip to sample, which probes unoccupied states of sample

- For $-V_{bias}$,electrons flow from sample to tip, which probes occupied states of sample

2 Theory

2.1 C.J Chen "Introduction to Scanning Tunneling Microscopy", Ch 2 10.1093/acprof:oso/9780199211500.001.0001

- Bardeen's transfer-Hamiltonian formalism in the spirit of Oppenheimer's treatment of field ionization

- Bardeen solved two separate subsystems; tunneling matrix element M is determined by a surface integral of the unperturbed wavefunctions of the the two subsystems at the separation surface; M is not sensitive to choice of the separation surface, fortunately

2.1.1 Bardeen theory of tunneling (1D)

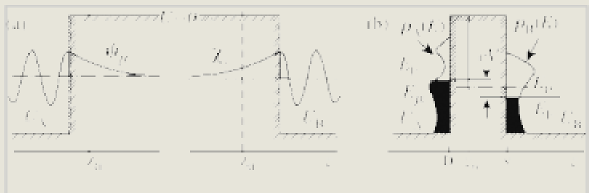


Fig. 2.3. The Bardeen tunneling theory: one-dimensional case. (a). When the two electrodes are far apart, the wavefunctions of both electrodes A and electrode B decay into the vacuum. (b). By bringing the two electrodes

- electrode A: $i\hbar\frac{\partial\Psi}{\partial t} = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + U_A\right]\Psi$, where $\Psi = \psi_\mu e^{-iE_\mu t/\hbar}$ and $\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + U_A\right]\psi_\mu = E_\mu\psi_\mu$; electrode B: $i\hbar\frac{\partial\Psi}{\partial t} = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + U_A\right]\Psi$, where $\Psi = \chi_\nu e^{-iE_\nu t/\hbar}$ and $\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + U_B\right]\chi_\nu = E_\nu\chi_\nu$
- combine system: $i\hbar\frac{\partial\Psi}{\partial t} = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + U_A + U_B\right]\Psi$
- state ψ_μ evolves and has probability transferring to electrode B; thus we assume

$$\Psi = \psi_\mu e^{-iE_\mu t/\hbar} + \sum_{\nu=1}^{\infty} c_v(t)\chi_\nu e^{-iE_\nu t/\hbar}$$

with $c_\nu(0) = 0$

- asuume approximately orthogonality; neither ψ_μ nor χ_ν are eigenfunctions of the combined Hamiltonian; so trial wavefunction $\Psi = \psi_\mu e^{-iE_\mu t/\hbar} + \sum_{\nu=1}^{\infty} c_v(t)\chi_\nu e^{-iE_\nu t/\hbar}$ is normalized up to second-order quantity proportional $|c_\nu|^2$
- plug trial wavefunction into combined system Hamiltonian

$$i\hbar\frac{\partial}{\partial t}\left(\psi_\mu e^{-iE_\mu t/\hbar} + \sum_{\nu=1}^{\infty} c_v(t)\chi_\nu e^{-iE_\nu t/\hbar}\right) = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + U_A + U_B\right]\times$$

$$\left(\psi_\mu e^{-iE_\mu t/\hbar} + \textstyle\sum_{\nu=1}^{\infty} c_v(t)\chi_\nu e^{-iE_\nu t/\hbar}\right)$$

$$i\hbar\left(-\tfrac{iE_\mu}{\hbar}\psi_\mu e^{-iE_\mu t/\hbar} + \textstyle\sum_{\nu=1}^{\infty}\frac{\partial c_v}{\partial t}\chi_\nu e^{-iE_\nu t/\hbar} - \sum_{\nu=1}^{\infty}\frac{iE_\nu}{\hbar}\chi_\nu e^{-iE_\nu t/\hbar}\right)$$

terms grouped by $i\hbar\frac{\partial\Psi}{\partial t} = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + U_A\right]\Psi$ will cancel each other out on each side leaving

$$i\hbar\sum_{\nu=1}^{\infty}\frac{\partial c_v}{\partial t}\chi_\nu e^{-iE_\nu t/\hbar} = U_B\psi_\mu e^{-iE_\mu t/\hbar} + U_A\sum_{\nu=1}^{\infty}c_v(t)\chi_\nu e^{-iE_\nu t/\hbar}$$

- apparently the term with U_A is a second-order infinitesimal quantity (why?), leaving

$$i\hbar\frac{\partial c_v}{\partial t} = \int_{z>z_0}\psi_\mu U_B\chi_\nu^* e^{-i(E_\mu - E_\nu)t/\hbar}$$

- define tunneling matrix element $M_{\mu\nu} = \int_{z>z_0} d^3r\psi_\mu U_B\chi_\nu^*$, evaluted only on right side where $U_B$is non-zero
- rewrite matrix elements to be

$$M_{\mu\nu} = \int_{z>z_0} d^3r\psi_\mu U_B\chi_\nu^* = M_{\mu\nu} = \int_{z>z_0} d^3r\psi_\mu(E_\nu + \frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2})\chi_\nu^*$$

- considering only elastic tunneling where $E_\mu = E_\nu$

$$M_{\mu\nu} = \int_{z>z_0} d^3r\left(\chi_\nu^* E_\mu\psi_\mu + \psi_\mu\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2}\chi_\nu^*\right)$$

- using $\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + U_A\right]\psi_\mu = E_\mu\psi_\mu$ and noting $U_A = 0$ on the side of electrode B

$$M_{\mu\nu} = -\frac{\hbar^2}{2m}\int_{z>z_0} d^3r\left(\chi_\nu^*\frac{\partial^2}{\partial z^2}\psi_\mu + \psi_\mu\frac{\partial^2}{\partial z^2}\chi_\nu^*\right)$$

- using the identity

$$\chi_\nu^*\frac{\partial^2}{\partial z^2}\psi_\mu + \psi_\mu\frac{\partial^2}{\partial z^2}\chi_\nu^* = \frac{\partial}{\partial z}\left(\chi_\nu^*\frac{\partial}{\partial z}\psi_\mu + \psi_\mu\frac{\partial}{\partial z}\chi_\nu^*\right)$$

  integrate over z to get Bardeen's tunneling matrix for 1D, which is a surface integral of wavefunctions and their normal derivatives of two free electrodes evaluated at separation surface z_0

$$M_{\mu\nu} = -\frac{\hbar^2}{2m}\int_{z=z_0}\left(\chi_\nu^*\frac{\partial}{\partial z}\psi_\mu + \psi_\mu\frac{\partial}{\partial z}\chi_\nu^*\right)dxdy$$

  no information about potential barrier information appears, only depends on wavefunctions; symmetric wrt both electrodes, which is the basis of reciprocity principle in STM and AFM

- tunneling spectroscopy- interpretation of the tunneling current with bias voltage

- tunneling current with bias voltage V, at finite T; factor of two for spin degeneracy

$$I = \frac{2\pi e^2}{\hbar}|M_{\mu\nu}|^2 \rho_B(E_F)\rho_A(E_F)V$$

- can also write the matrix element using solution to square potential with barrier width s;$\kappa_\mu = \kappa_\nu \approx \frac{\sqrt{2m\phi}}{\hbar}$; ϕ is like an average work function of the two sides

$$\psi_\mu(z) = \psi_\mu(0)e^{-\kappa_\mu z}$$

$$\chi_\nu(z) = \chi_\nu(s)e^{-\kappa_\nu(z-s)}$$

which makes the matrix tunneling element to be

$$\begin{aligned} M_{\mu\nu} &= \frac{\hbar^2}{2m}\int_{z=z_0} 2\kappa_z\psi_\mu(0)\chi_\nu(s)e^{-\kappa_\mu z_0}e^{-\kappa_\mu(z_0-s)}dxdy \\ &= \left[\frac{\hbar^2}{2m}\int_{z=z_0} 2\kappa_z\psi_\mu(0)\chi_\nu(s)dxdy\right]e^{-\kappa_\mu s} \end{aligned}$$

- term in brackets is a constant; matrix element is independent of separation surface $z = z_0$;

- asymmetric tunneling spectrum

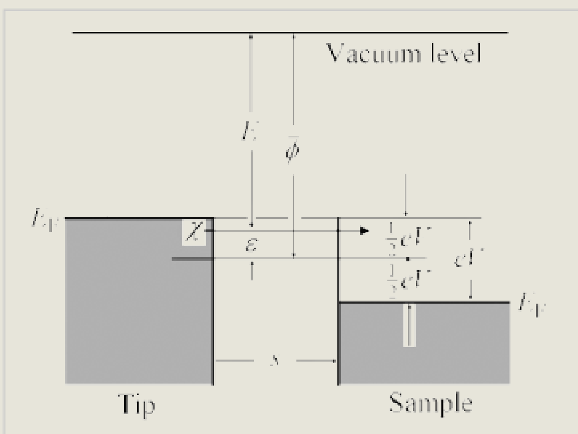


Fig. 2.5. Energy dependence of tunneling matrix element and the asymmetry of tunneling spectrum. A

- tunneling current in asymmetric tunneling spectrum is

$$I = \frac{4\pi e}{\hbar}\int_{-\frac{1}{2}eV}^{\frac{1}{2}eV} \rho_s(E_F + \frac{1}{2}eV + \epsilon)\rho_t(E_F - \frac{1}{2}eV + \epsilon)|M(\epsilon)|^2 d\epsilon$$

$$\kappa = \frac{\sqrt{2mE}}{\hbar} = \frac{\sqrt{2m(\bar{\phi}-\epsilon)}}{\hbar} \approx \frac{\sqrt{2m\bar{\phi}}}{\hbar}\left(1 - \frac{\epsilon}{2\bar{\phi}}\right) \equiv \kappa_0\left(1 - \frac{\epsilon}{2\bar{\phi}}\right)$$

$$M(\epsilon) = M(0)\exp\left(\frac{\kappa_0 \epsilon s}{2\bar{\phi}}\right)$$

- in the limiting case for large s, the main contribution of integral for tunneling current comes from small energy interval $\epsilon \approx eV/2$
- low bias-voltage limit
- $$I = \frac{2\pi e^2}{\hbar} V \sum_{\mu\nu} |M_{\mu\nu}|^2 \delta(E_\nu - E_F)\delta(E_\mu - E_F) = \frac{2\pi e^2}{\hbar} V |M_{\mu\nu}|^2 \rho_s(E_F)\rho_t(E_F)$$
- tunneling conductance

$$\left(\frac{dI}{dU}\right)_{U=V} \approx \rho_s(E_F + eV)\rho_t(E_F)$$

2.1.2 Bardeen theory of tunneling (3D)

- matrix elements in 3D

$$M_{\mu\nu} = -\frac{\hbar^2}{2m}\int_{\Omega_t} d^3r \left(\chi_\nu^* \nabla^2 \psi_\mu + \psi_\mu \nabla^2 \chi_\nu^*\right) = \frac{\hbar^2}{2m}\int_\Sigma d\mathbf{S}\left(\chi_\nu^* \nabla \psi_\mu + \psi_\mu \nabla \chi_\nu^*\right) = M_{\mu\nu}^*$$

- using Green's theorem for separation surface Σ

$$\int_{\Omega_t} d^3r \left(\chi_\nu^* \nabla^2 \psi_\mu + \psi_\mu \nabla^2 \chi_\nu^*\right) = -\int_\Sigma d\mathbf{S}\left(\chi_\nu^* \nabla \psi_\mu + \psi_\mu \nabla \chi_\nu^*\right)$$

- wavefunction correction with Green's method; tip distorts sample potential and vice versa

$$V_s = U_s - U_{s0}, V_t = U_t - U_{t0}$$

  - to first order, perturbed wavefunction is $\psi(\mathbf{r}) = \psi_0(\mathbf{r}) + \int G(\mathbf{r},\mathbf{r}')V(\mathbf{r}')\psi_0(\mathbf{r}')d\mathbf{r}$
  - Green's function defined by $\left(-\frac{\hbar^2}{2m}\nabla^2 + U_{s0} - E\right)G(\mathbf{r},\mathbf{r}') = -\delta(\mathbf{r},\mathbf{r}')$

- original paper of Bardeen formulated using occupation number

2.2 Papers

- G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel Phys. Rev. Lett. 49, 57 (1982) 10.1103/PhysRevLett.49.57
- Gerd Binnig and Heinrich Rohrer Rev. Mod. Phys. 59, 615 (1987) 10.1103/RevModPhys.59.615
- Tutorial and derivation of Bardeen's approach with modern interpretations: Gottlieb, A. D. & Wesoloski, L. Nanotechnology 17, R57–R65 (2006).

2.3 Ways to write the tunneling current

- Some useful group websites with summaries (for STM measurements at low temperatures and low biases)
  - Hoffman Group, Havard
  - Zelijkovic Group, Boston College
- Tersoff1985 10.1103/PhysRevB.31.805; based on Bardeen's approach

$$I = \frac{2\pi e}{\hbar} \sum_{\mu,\nu} f(E_\mu)([1 - f(E_\nu + eV)]|M_{\mu\nu}|^2 \delta(E_\mu - E_\nu)$$

- approximating the tunneling of electrons across vacuum a the transmission of electrons across a square barrier, we can write an expression for the tunneling current using WKB approximation; based on here
  - we know the solution of the wavefunction in the barrier is $\psi(x) = \psi(0)e^{-\kappa x}, \kappa = \frac{\sqrt{2m(\phi - E)}}{\hbar}$
  - the probability of finding electron past barrier of width d is $|\psi(d)|^2 = |\psi(0)|^2 e^{-2\kappa d}$
  - the definite of the local DOS is $\rho(z, E) = \frac{1}{\varepsilon} \sum_{E_n = E - \varepsilon} |\psi_n(z)|^2$ for $\varepsilon \to 0$
  - in the limit that the work function $\phi \approx 1/2(\phi_s + \phi_t) \gg eV_{bias}$, $\kappa \approx \frac{\sqrt{2m\phi}}{\hbar}$

2.4 Tersoff1985: Theory of the scanning tunneling microscope 10.1103/PhysRevB.31.805

- theory for tunneling between real surface and model probe tip- here the tip is modeled most simply as s-wave
- tunneling current proportional to local DOS at surface at the tip

- for tip of radius R and vacuum gap distance d, laterial resolution $\sim [2\text{Å}(R+d)]^{1/2}$; applied to 2x1 and 3x1 reconstructions of Au(110) and GaAs (110)
- in STM, height is adjusted to maintain same tunneling resistance between surface and tip −> contour map of surface
- analogy with planar tunneling, current decays with $\hbar(8m\phi)^{-1/2}$for ϕ work function
- surface treated "exactly" when tip is modeled as locally sphereical potential well (at the time, did not understand local geometry of the tip)
- STM relatively insensitive for position of surface layer relative to underlying layers (at least for Au)
- first-order PT (ish)

$$I = \frac{2\pi e}{\hbar} \sum_{\mu,\nu} f(E_\mu)([1 - f(E_\nu + eV)]|M_{\mu\nu}|^2 \delta(E_\mu - E_\nu)$$

V is applied voltage, $M_{\mu\nu}$is matrix element between $\psi_\mu$of the probe and $\psi_\nu$of the surface (which in general are nonorthogonal states of different Hamiltonians), $E_\mu$is energy of state $\psi_\mu$in the absence of tunneling

- since experiments are taken at small voltage and low temperature (i.e., ignore reverse tunneling) (in the context of metal surfaces- not sure for semiconductor)
- Bardeen showed that $M_{\mu\nu}$ is essentially the current operator

$$M_{\mu\nu} = \frac{\hbar^2}{2m} \int d\vec{S} \cdot (\psi_\mu^* \vec{\nabla} \psi_\nu - \psi_\nu \vec{\nabla} \psi_\mu^*)$$

where integral is over any surface lying entirely within the vacuum (barrier) region separating the probe and surface

- Surface wave function $\psi_\nu$is expanded as

$$\psi_\nu = \frac{1}{\Omega_s^{1/2}} \sum_G a_G \exp[\kappa^2 + |\vec{\kappa}_G|^2)^{1/2} \exp(i\vec{\kappa_G} \cdot \vec{x})$$

where $\kappa = \hbar^{-1}(2m\phi)^{1/2}$is the minimum inverse decay length for wave functions in vacuum; $\vec{\kappa}_G = \vec{k}_{||} + \vec{G}$ denotes surface Bloch and reciprocal-lattice vectors

- Probe wave function $\psi_\mu$is modeled as locally spherical potential well (assuming work function of tip is same as surface)

$$\psi_\mu = \frac{1}{\Omega_t^{1/2}} c_t \kappa R \frac{\exp(\kappa R)}{(\kappa|\vec{r}-\vec{r}_0|)\exp(\kappa|\vec{r}-\vec{r}_0|)}$$

- This (+ further assumptions and simplifications) leads to

$$M_{\mu\nu} = \frac{\hbar^2}{2m}\frac{4\pi}{\kappa\Omega_t^{1/2}}\kappa R \exp(\kappa R)\psi_\nu(\vec{r}_0)$$

where $\vec{r}_0$is the position of the center of curvature of the tip

- This leaves the final form for the tunneling current as

$$I = 32\pi^3 e^2 V\phi^2 D_t(E_F)\frac{R^2 e^{2\kappa R}}{\hbar\kappa^4}\sum_\nu |\psi_\nu(\vec{r}_o)|^2 \delta(E_\nu - E_F) \equiv 32\pi^3 e^2 V\phi^2 D_t(E_F)\frac{R^2 e^{2\kappa R}}{\hbar\kappa^4}\rho(\vec{r}_o, E)$$

where $D_t$is the DOS of the probe tip and $\rho(\vec{r}_o, E)$ is the surface LDOS

- spherical tip approx enters only in normalization of ψ_μ; this model less accurate for large R, where higher l becomes more important
- approximate methods for STM- STM provides limited information for smooth, low-Miller-index surfaces
- small voltages- tunneling from states near E_F- for semiconductors- n- and p-doping can give different STM images
- low doping or high voltages: voltage polarity may determing whether tunneling involves VBs or CBs (e.g,, Si (111) need large 2.5 V)
- GaAs(110) 1x1 reconstruction surface- got a qualitative difference of VBs and CBs

Numerical things to consider

- if the slab is too thiin- inaccurate work function; thin sparse sampling of bulk continuum leads to numerical noise in energy-projected quantities
- not good accuracy at distances far from surface
- limitations of plane-wave part in describing exponential decay inside deep troughs
- how many states from band edge to include in LDOS; within 1eV used for GaAS (110)

In summary, things that could lead to deviations from experiment

- tip geometry not properly taken into account (e.g., not have spherical potential well); s-wave treatment of tip without details of the geometry

- assumptions of tip radius and vacuum gap distance
- if can use similar assumptions of metal surface for which beginning is referenced to for a semiconducting surface
- if ϕ assumed same for tip and sample
- limits of validity of model
  - tip and surface do not have too large difference in work function- affects the effective $\kappa(\vec{r})$
  - implicit assumption that potential goes to zero in region between surface and tip- actually the electron is never more than 3-6 Ang from surface so magnitude of potential is never less than ~ 1 eV