Transport measurements

Patrick R. LeClair

May 26, 2010

Contents

This report is a short guide to the transport measurements and equipment used in the LeClair laboratory at the University of Alabama, located in Bevill 180. The goal is to give the reader a basic understanding of the most common measurements performed, and a somewhat detailed overview of the equipment. Some general guidelines are also given for performing good transport measurements.

1 Electrical Resistivity

In order to understand the subtleties of the probe techniques used to measure the resistivity in this laboratory, it is useful to very quickly review a simple classical model of conduction in metals and semiconductors. If you are familiar with the concepts of mobility, carrier concentration, and resistivity, you may proceed to Sec. 3.

1.1 Electric Current

First, we will consider a homogeneous conductor, whose primary conduction is by electrons, and later we will generalize our results to the case of semiconductors with both electron and hole conduction. If we take a cross section of a conductor, such as a circular wire, *an electric current is said to exist if there is a net flow of charge through this surface.* The *amount* of current is simply the *rate* at which charge is flowing, the number of charges per unit time that traverse the cross-section. Strictly speaking, we try to choose the cross-sections for defining charge flow such that the *charges flow perpendicular to that surface.* Current is a flux of charge through a wire in the same way that water flow is a flux of water through a pipe. Qualitatively, this is a reasonable way to think about electric circuits – current always has to flow somewhere, and you don't want an open connection any more than you would want an open-ended water pipe. Voltage is more like pressure – you can have a voltage even when nothing is flowing, it just means there is the *potential* for flow.

If a net amount of charge ΔQ flows perpendicularly through a particular surface of area A within a time interval Δt, we define the electric current to be simply the rate at which charge passes that surface:

$$I \equiv \frac{dQ}{dt} \tag{1}$$

In other words, current is charge flow per unit time. We should get one thing out of the way right off the bat: the definition for the current *direction* is somewhat confusing. The historical definition is that **current flow is defined as the direction that positive charges would be moving**. Of course, at this point we know that usually it is really electrons doing all the moving, but the definition of electric current has held fast.

1.1.1 Getting Current to Flow

Current in real conductors is due to the (net) motion of microscopic charge carriers. How much current flows depends on the average speed of these charge carriers, the number of charge carriers per unit volume (the density of charge carriers), and how much charge is carried by each. But how do we get charges to flow through a conductor in the first place?

In order to get a net flow of charges, we need to provide a potential difference (voltage![i] The presence of a voltage gives rise to an electric field across the conductor, which in turn causes an electric force, which accelerates the charges. The effectiveness of a potential difference to cause a current depends on the density of charge carriers, their average speed, and microscopic properties of the conductor itself.

The free charges in conductors are extremely numerous and fairly mobile. Inside a normal conductor, like copper, there is a fantastic density of charge carriers, $\sim 10^{22}$ electrons per cm^3! So many, in fact, that they continuously scatter off of each other and the fixed atoms in the conductor (about once every 10^{-14} sec or so, even in a good conductor!). Typical drift speeds in copper are $\sim 10^{-3} - 10^{-4}$ m/s for moderate electric fields, compared to the speed of random thermal electron motion of $\sim 10^5$ m/s. Any particular charge carrier has a hard time getting anywhere. Even though the charges are mobile, and able to move at fantastic speeds, the time it takes to actually get anywhere is quite a bit longer than expected. A bit like pachinko.

One result of all these collisions is that the carriers in, *e.g.*, copper, cover huge *distances* in any given time interval but have a very small *displacement* – most of their movement is wasted, and they end up close to where they started out, so their *net* velocity is very small. Even when we apply a potential difference, the net flow of charges is more sluggish than we might expect, due to all these collisions. The *net* velocity of charge flow[ii] we call the *drift velocity*, v_d. In normal conductors, like copper, this drift velocity is more or less proportional to the voltage applied, a point which we will explore in depth presently.

1.2 Drift Velocity and Current

Our conceptual physical picture of current in conductors is basically complete. A voltage induces an electric field, which gives the carriers a net velocity in one direction, which is an electric current. This drift motion along the electric field is superimposed upon the random thermal motion of the charge carriers (just like the random thermal motion in an ideal gas). From here, all we need to do is apply our knowledge of electric forces and fields and kinematics to come up with a relationship between current, field, and voltage.

---

[i] From now on, we will interchangeably use the phrases "potential difference" and "voltage." From our point of view, they are the same thing.

[ii] Distinct from and not to be confused with the random thermal motion, see below.

So first: given a drift velocity v_d, through a conductor of cross section A, what is the current? The number of charges that flow through our cross section A in the time dt is just the free charge which is physically close enough to *reach* the surface A within that time. Those charges close enough must cover the distance dx in the time dt. Since the average speed of the carriers is v_d, then we must have $dx = v_d dt$. This is illustrated schematically in Figure 1.

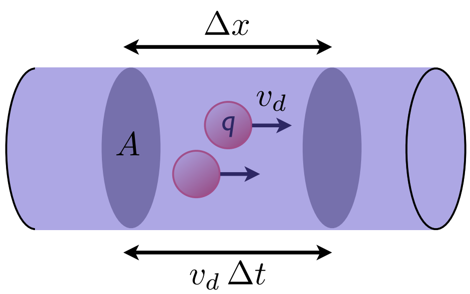


Figure 1: *A small piece of a conductor of cross-sectional area A. The charge carriers move with a speed v_d, and are displaced by $dx = v_d dt$ in a time interval dt. The number of carriers in a section of length dx is, on average, $nAv_d dt$, where n is the density of the charge carriers.*

The number of charges which cross the surface A, those close enough to reach it in a time dt, is just the number contained within the volume $A \cdot dx$, or $Av_d dt$. A bit more mathematically, we can write this:

$$\begin{aligned}
\text{number of charge carriers} \equiv N &= \text{charge density} \times \text{volume} && (2)\\
&= \text{charge density} \times \text{area} \times \text{distance covered in time} dt && (3)\\
&= nAdx = nAv_d dt && (4)
\end{aligned}$$

Here we have used n to represent the number of charges per unit volume, the carrier density. The total amount of charge is the number of charge carriers times how much charge each one carries, which we'll call q. The current then is just the total amount of charge, Nq divided by the total amount of time, dt:

$$I = \frac{dQ}{dt} = \frac{Nq}{dt} = \frac{nqAv_d dt}{dt} = nqAv_d \tag{5}$$

We can see that the drift velocity and resulting current are larger when the carriers carry more charge q, or when their mass is small. However, it would be nice to have expressions that didn't directly involve the cross-sectional area of the conductor, so we can calculate general properties independent of any particular conductor shape or size. For this reason, it is common to introduce *current density*, J, which is just the current per unit area. Rewriting Eq. 5 in terms of current density, we come up with a simpler and more general expression:

$$J \equiv \frac{I}{A} = nqv_d \tag{6}$$

Now we can calculate the current density for any given material of *arbitrary geometry*, and later specify a cross-sectional area to determine absolute currents.

1.3 Resistance and Ohm's Law

From Equation 5, we saw that the current through a conductor can be expected to scale with the drift velocity. You might expect that the effect of increasing the applied voltage across a conductor ΔV is to increase the drift velocity. This is basically true, but justifying that statement will require a few more steps.

More accurately, the presence of a potential difference between two points on the conductor means that those two points are at different potential energies. Recall that negative charges want to move from regions of lower potential to regions of higher potential. In a conductor, even when a current flows, the charges like to spread out as evenly as possible. This even and moving distribution of charge gives rise to a uniform *electric field*. If the potential difference ΔV is applied over some distance l, and the electric field is uniform, we know that the electric field along the length of the conductor must be given by:

$$E = \frac{\Delta V}{l} \tag{7}$$

The presence of the electric field causes an acceleration of the charge carriers:

$$a = \frac{F_e}{m} = \frac{q}{m}E \tag{8}$$

Thus the acceleration of the charge carriers depends only on the electric field and their charge-mass ratio, q/m, about 1.76×10^{11} C/kg for electrons. In order to figure out how much current will flow for a given potential difference, we need to find a way to take into account the dissipative effect of all the collisions the carriers are constantly undergoing. In a sense, the collection of charge carriers is a bit like an ideal gas, and our treatment here is reminiscent of an ideal gas law derivation. The analogy is a close one – the innumerable electrons in a conductor are often called an *electron gas*.

1.3.1 Drift Velocity and Collisions

If we assume the charge carriers are electrons, of mass m_e (and charge $-e$), then each has an *average* momentum $p = m_e v_d$. We expect on average that each collision an electron experiences will completely destroy all forward momentum – they are stopped cold by every single collision. This makes some sense, since most of the collisions will be with the atoms making up the conductor, which are very heavy compared to electrons, rather than with other electrons. If all forward momentum is destroyed, then the electron is left with only its random thermal motion. If there were no electric force present to accelerate the electrons, the random thermal motion of all the electrons will cancel out, and there is no net flow or current.

We can easily find the thermal velocity of the carriers just like we do for an ideal gas – the thermal energy of the electrons is $\frac{3}{2}k_B T$, where k_B is Boltzmann's constant, and we equate this to the carriers' kinetic energy:

$$\frac{3}{2}k_B T = \frac{1}{2}mv_{th}^2 \tag{9}$$

$$\implies v_{th} = \sqrt{\frac{3k_B T}{m}} \sim 10^5 \text{ m/s} \ \ \text{(at 295 K)} \tag{10}$$

Here we use v_{th} to specify the thermal velocity distinctly from the electric-field-induced *drift* velocity. As it turns out, the thermal velocity typically greatly exceeds the drift velocity (by ten million times or so!) – the acceleration of the carriers by the electric field induces only a tiny velocity compared to that given by the random thermal motion of the carriers. Again, this is what leads to carriers covering huge *distances* but having very small *displacements*. The overall motion is terribly chaotic, and even fairly large electric fields only alter the carrier velocity in conductors by parts per million at best. Still, the random thermal velocities do not contribute to the electric current,[iii] it is only the *tiny* field-induced drift velocity that gives rise to electric current.

We should also keep in mind that the collisions the carriers undergo are not continuous, but happen one after another with some average time between them τ.[iv] In that time interval, the electron loses its momentum $m_e v_d$ due to a collision, and thereafter regains it due to the action of electric field present, only to lose it again about τ seconds later. As stated above, the presence of the electric force F_e gives the electron an acceleration $a = F_e/m_e$, which allows it to regain its former drift velocity. From kinematics, we would expect a *mean* displacement $v_d \approx a\tau$.[v]

The starting and stopping motion of the carriers gives us an average rate at which the electrons are losing momentum due to the collisions and associated impulse forces. We can straightforwardly find this momentum change as:

$$\left.\left(\frac{\Delta p}{\Delta t}\right)\right|_{\text{loss}} = \frac{m_e v_d}{\tau} \tag{11}$$

Once the scattering event is over, the electron regains momentum through the action of the electric *force*

[iii] They do give rise to electrical *noise*, however.
[iv] For Cu, we can estimate $\tau \sim 2 \times 10^{-14}$ s.
[v] Depending on the method of derivation, there may be a factor of 2 in this expression, but the physics is the same.

caused by the electric field. We can easily write down the momentum gained up until the next collision:

$$\left(\frac{\Delta p}{\Delta t}\right)\bigg|_{\text{gain}} = F_e = qE = -eE \tag{12}$$

Now, the total momentum loss has to equal the total momentum gain for there to be a steady state. If this were not true, the momentum would quickly build up, and the whole wire would start to move! So we must impose conservation of momentum:

$$\left(\frac{\Delta p}{\Delta t}\right)\bigg|_{\text{loss}} = \left(\frac{\Delta p}{\Delta t}\right)\bigg|_{\text{gain}} \tag{13}$$

$$\frac{m_e v_d}{\tau} = -eE_x \tag{14}$$

$$v_d = \frac{-e\tau}{m_e}E \tag{15}$$

Really, this is just an application of Newton's laws – $\Delta p/\Delta t$ is a force, and the equations above are also essentially a force balance between the electric force and the impulse force due to the collision. Now we have an expression for the average drift velocity of electrons flowing along the wire, in terms of the average time between carrier collisions:

$$v_d = \frac{-e\tau}{m_e}E \tag{16}$$

The minus sign makes sense here, by the way. Since electrons are negatively charged, they move in the opposite direction that the electric field lines point. It is also reassuring that the drift velocity *increases* as τ increases, since more time between collisions means more time spent accelerating, and that in principle lighter carriers would have a higher velocity since they are more easily accelerated. Finally, the proportionality with the electric field is what we expect.

For typical metals, we can estimate drift velocities of about 5×10^{-3} m/s for a moderate electric field of 1 V/m, about *eight orders of magnitude below the thermal velocity!* Really, the effect of the electric field is quite negligible in one sense, though it has profound consequences.

1.3.2 Mean Free Path and Mobility

Instead of dealing with the mean time between collisions, we could just as easily have started with the mean *distance* that electrons travel before undergoing a collision.[vi] This quantity is known as the *mean free path*, λ_{mfp}, and it has essentially the same meaning as it does in the kinetic theory of gasses. The

[vi] Here we do mean the *distance* covered between collisions, not the *displacement*

shorter the time between collisions, the smaller the mean free path, and vice versa. The mean time and mean free path are easily related through kinematics:

$$\lambda_{\text{mfp}} = \tau(\nu_d + \nu_{th}) \approx \tau\nu_{th} \tag{17}$$

Here we are considering the total distance covered not just the net displacement, so we need to use the *total* velocity, $\nu_d + \nu_{th}$. For the last relationship, we have made use of the fact that $\nu_{th} \gg \nu_d$. What this means is that the mean distance (and mean time) between collisions does *not* really depend on the applied electric field, but really *only* comes from the random thermal motion of the carriers.

The proportionality constant between drift velocity and electric field in Eq. 16 is commonly called the carrier *mobility*, which is just what it sounds like. In this case, we write $\nu_d = \mu E$, where μ is the mobility:

$$\nu_d = \mu E \qquad \text{with} \qquad \mu = \frac{q\tau}{m} \tag{18}$$

From the units of μ ($m^2/V{\cdot}s$) and E (N/C or V/m), we can see that mobility is a quantity that tells us how far a charge is able to move per second per unit of electric field (V/m). Now we have a nice expression for *exactly* what we mean by mobility, rather than just a vague notion.

1.3.3 Current, Electric Field, and Voltage

Plugging Eq. 16 into Eq. 6, we find the relationship between current density and electric field, Ohm's law:

$$J = \frac{I}{A} = nq\nu_d = -ne\frac{-eE\tau}{m_e} = \frac{ne^2\tau}{m_e}E \equiv \frac{1}{\rho}E \tag{19}$$

In the end, it turns out that current density (or current) and electric field are simply proportional. We could almost have guessed this in the first place, but now we have a formal relationship between the two, and we even know the constant of proportionality. Typically, we define a new quantity ρ, the *electrical resistivity*, which is the constant of proportionality between current density and electric field:[vii]

$$\rho = \frac{m_e}{ne^2\tau} = \frac{1}{ne\mu} = \frac{1}{\sigma} \tag{20}$$

The conductivity σ is simply the inverse of the resistivity. Resistivity represents the effectiveness with which a given electric field or potential difference causes a current to flow, and is a (strongly) material-

[vii] We will use a slightly different rho character for resistivity, ρ, to distinguish it from the one we use for *mass density*, ρ.

dependent property – it is a measure of the resistance of a material to current flow. We see that the resistivity gets larger when the time between electron collisions gets smaller, just as we would expect, and it gets larger when we increase the density of free carriers. Similarly, resistivity an mobility are inversely proportional. We can go further in our analysis by noting that the potential difference and electric field are simply related in a conductor by $E = \Delta V/l$, which leads us to:

$$J = \frac{I}{A} = \frac{1}{\rho}\frac{\Delta V}{l} \qquad \text{or} \qquad \Delta V = \frac{\rho l}{A} I = \rho l J \tag{21}$$

In other words, we find $J \propto I \propto \Delta V$ – the current flow in a conductor is proportional to the magnitude of the applied voltage, and the amount of current one gets for a particular applied voltage depends on the conductor's resistivity and geometry. We can make this simpler by introducing a new constant of proportionality $R = \frac{\rho l}{A}$. This, along with the definition of current density ($J = I/A$), will allow us to relate I and ΔV directly. **This new constant of proportionality R between I and ΔV is known as the resistance of the conductor**, and it allows us to connect ΔV and I in the traditional form known as Ohm's[viii] law:

$$\Delta V = IR \qquad \text{or} \qquad I = \frac{\Delta V}{R} \qquad \text{or} \qquad R = \frac{\Delta V}{I} \tag{22}$$

1.4 Transport in Semiconductors

The primary difference between metals and semiconductors, from our point of view, is that semiconductors can have electrical conduction by both electrons and holes.[ix] In practical semiconducting devices, the carrier concentrations are such that the approximation of a small drift velocity compared the the thermal velocity holds, and our model of collision-dominated conduction above well-describes both electron and hole motion.

We can consider the electron and hole contributions to the conductivity separately. If an electron meets a hole in the semiconductor, the electron will simply annihilate the hole (i.e., occupy the empty state that the hole represents), and neither will contribute to conduction. For this reason, we may consider the two channels of conductivity to be essentially non-interacting (other than the presence of one influencing the mean scattering time of the other), and treat them as two parallel conductors within the same material. We can then add the electron and hole resistivities in the same fashion that we add resistors in parallel, since the geometrical factors relating resistivity to resistance are the same for both. That is, we add the

[viii] After Georg Simon Ohm (1789–1854) a German physicist who first found the relationship between current, voltage, and resistance.

[ix] A "hole" is the conceptual opposite of an electron, and describes the *lack* of an electron at a position where one could exist. It is not the same as a positron.

resistivities inversely (or simply add the conductivities). If we denote the electron density and mobility as n and μ_n, and the hole density and mobility as p and μ_p, this leads to a total resistivity ρ_{tot}

$$\frac{1}{\rho_{tot}} = \frac{1}{\rho_n} + \frac{1}{\rho_p} = ne\mu_n + pe\mu_p \tag{23}$$

$$\rho_{tot} = \frac{1}{ne\mu_n + pe\mu_p} \tag{24}$$

If one species of carrier dominates the conductivity – either by sheer numbers or by a vastly larger mobility – the expression reduces to that of a normal conductor above, which is the situation we will encounter. Which type of carrier is dominant cannot be determined by a simple resistivity measurement, being insensitive to the sign of the charge carrier. The Hall effect, discussed below, can make this determination. If we simply measure the resistivity of a semiconductor without regard to whether both charge carriers play a significant role or not, we can extract an effective mobility μ_{eff} and carrier concentration N:

$$\rho_{tot} \equiv \frac{1}{Ne\mu_{eff}} \tag{25}$$

$$N\mu_{eff} \equiv n\mu_n + p\mu_p \tag{26}$$

For a doped semiconductor, one can show that the product np is constant, $np = n_i^2$ where n_i is the *intrinsic carrier concentration*. Without doping, the carrier concentrations must be equal, and $n = p = n_i$. Except for very high carrier densities, approaching that of a metal, n and p are highly temperature-dependent, increasing as temperature increases. This is due to the fact that in a lightly-doped semiconductor the concentration of free carriers of either type is strongly determined by thermal activation. The mobility shows a strong temperature dependence as well, with mobility decreasing strongly as temperature increases. For pure silicon, $n_i \approx 1.5 \times 10^{10}\,\mathrm{cm}^{-3}$ at $300\,\mathrm{K}$, about 12 orders of magnitude below that of copper. Doped silicon can have $n_i \approx 10^{13} - 10^{18}\,\mathrm{cm}^{-3}$, above $n_i \approx 10^{18}\,\mathrm{cm}^{-3}$, one usually considers the semiconductor so highly doped that it is for all intents and purposes a metal.

2 Measuring resistive devices

2.1 Sourcing Voltage

A current can only be maintained in a closed circuit by a source of electrical energy. The simplest way to generate a current in a circuit is to use a *voltage source*, such as a battery. A voltage source essentially raises or lowers the potential energy of charges that pass through it. The amount of energy gained *per charge* that passes through a device is the potential difference that the voltage supplies, ΔV, measured in Joules per Coulomb (J/C), i.e., Volts (V). Though voltage is strictly an energy per unit charge, it is often useful to think of a voltage as a "pressure" of sorts, which tries to force charges through an electric circuit. Just like hydrostatic pressure, the presence of a voltage does not necessarily lead to a current, this only occurs when a completed circuit is present. In this way of thinking, a voltage source is a sort of generalized power supply which can be thought of as a "charge pump" that tries to force charges to move within an electric field inside the source. Many batteries, for instance, are "electron pumps" in which negatively charged electrons move opposite to the direction of the electric field. In an idealized voltage source, the output terminals provide a constant potential difference ΔV, and can pump any amount of charge through any closed circuit connected to the output terminals.

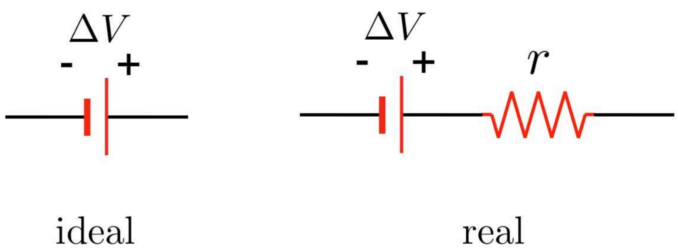


Figure 2: *A real voltage source provides a voltage ΔV, but has an internal resistance r. The actual output voltage developed at its terminals depends on r and the resistance of the external circuit connected to the battery.*

Real voltage sources, however, always have internal resistances, resulting in parasitic voltage losses within the source itself, and they have power limits which restrict the amount of current that can be sourced. In general, we can model a real voltage source as an ideal voltage source ΔV in series with an internal resistance r, as illustrated in Figure 2. The effect of the internal resistance is clear: as soon as an external load is connected to the voltage source and a current flows, the voltage at the battery terminals is always less than that of the ideal internal source. The only way to realize the ideal voltage of a source is if it drives no current – hardly useful for our purposes.

As a concrete example, consider the circuit in Figure 3, a voltage source ΔV with internal resistance r connected to an external resistor R.[x] If we neglect the internal resistance of the battery, the potential difference across the battery terminals is ΔV.

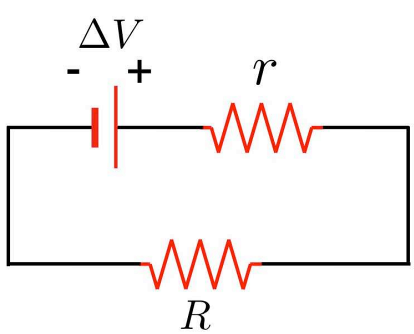


Figure 3: *A voltage source ΔV with internal resistance r connected to an external resistor (load) R.*

Once the external (load) resistance R is connected to the source, a single current I is produced in this

[x] We are assuming, for now, that wires connecting to the source have no resistance.

single-loop circuit. Conservation of energy (a.k.a. Kirchhoff's voltage law) requires the sum of potential differences around the entire circuit be zero:

$$0 = \Delta V - Ir - IR \tag{27}$$

Thus, the voltage delivered to the external load resistance R is only

$$\Delta V_{\text{load}} = \Delta V - Ir = IR = \Delta V \frac{R}{r + R} \tag{28}$$

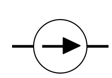
This makes it clear that the voltage across the load is the same as the ideal voltage ΔV only when the current is zero. This is why another name for the rated voltage is the **open-circuit voltage** – rated and actual voltages are only the same for a real voltage source when nothing is connected and no current flows. For completeness, given a load resistance R, internal resistance r, and an ideal open-circuit source voltage ΔV, we can also determine the current:

$$I = \frac{\Delta V}{R + r} \tag{29}$$

Clearly, the current delivered by the battery through the resistor depends on both the resistor's value *and* the internal resistance of the battery. If $R \gg r$, of course we need not worry about the internal resistance of the battery, and this is the regime we prefer to operate in. In a nutshell: voltages sources like high load resistances, compared to their internal resistance. [xi]

2.2 Sourcing Current

A current source is nothing more than a device that delivers and absorbs a constant current, sourcing and sinking a constant number of charges per unit time. An ideal current source (which exists only on paper) delivers a constant current to any closed circuit connected to its output terminals, no matter what the voltage or load resistance. Though a battery provides a simple example of a voltage source, there is no correspondingly simple realization of a current source.

Circuit diagram symbol for a current source: 

We can *approximate* a current source, however, with a single battery and resistor. In the circuit of Fig. 4, a battery with internal resistance connected to a load resistor, the current through the load is given by Eq. 29. If we make the load resistor very small (or equivalently, make the internal resistance of the battery very large), $r \gg R_{\text{load}}$, then the current through the load resistor is $I \approx \Delta V / r$. This does provide a roughly constant current, but the power loss in the internal resistor will be severe, and it is generally impractical

[xi] Good laboratory voltage sources can have internal resistances well below $1\,\Omega$.

to construct a current source in this way (that is not to say that it is not very commonly done anyway, however).

How more realistic constant current sources work internally is a bit beyond the scope of our discussion. However, that does not prevent us from seeing how they behave when connected to a circuit. In the same way that a real voltage source can be considered an ideal voltage source in series with a resistor, a real current source can be considered an ideal current source in *parallel* with a resistor, as shown in Fig. 4.

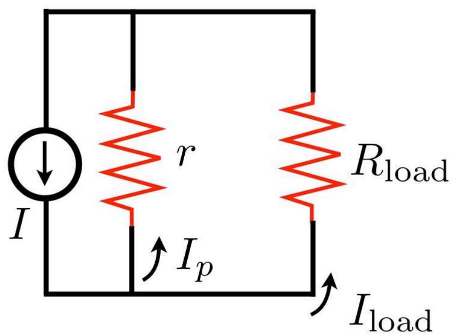


Figure 4: *A real current source can be considered as an ideal current source in parallel with an internal resistance* r. *The internal resistance "steals" some of the current, depending on the value of the load resistor.*

If the internal resistance is very large, almost all of the current goes through the load, and the current source is nearly ideal. If the load resistance becomes comparable to the internal resistance, however, a significant portion of the current takes the "parasitic" path (I_p in the figure) through the internal resistance, and the source is no longer close to ideal. The current through the load resistance is easily calculated using charge and energy conservation (a.k.a., Kirchhoff's current and voltage laws):

$$I_{load} = I\frac{r}{r+R} \tag{30}$$

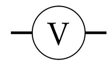
The current through the load is independent of the load resistance R and nearly equal to the source current I when $r \gg R_{load}$. In other words, current sources want low load resistances, in contrast to voltage sources. This brings up one answer to a common question: is it better to source current or voltage? If the load you are trying to source has a large resistance, as might be the case for a tunneling device, sourcing voltage is generally better. If the load is small, as is usually the case for all-metal GMR devices, sourcing current is generally better.[xii]

2.3 Measuring Voltage

A voltmeter is just what it sounds like – a device that measures voltage, or potential difference, between two points. A typical voltmeter has two input terminals, and one connects wires from these input

[xii]For sources, internal resistance is often called "output resistance." Good laboratory current sources can have internal resistances above $10^{14}\,\Omega$, while good laboratory voltage sources can have internal resistances well below $1\,\Omega$, so with good equipment either I or ΔV can usually be sourced for most common measurements without issues. Noise is often what actually determines which is used, but even so, the rule of thumb is still useful.

terminals to the points within a circuit between which one wants to know the potential difference. If we wish to measure the potential difference across a particular component in a circuit, we connect the voltmeter in *parallel* with that component.

Circuit diagram symbol for a voltmeter: 

Of course, the idea is to measure the potential difference while disturbing the circuit as little as possible. For this reason, voltmeters have very high internal resistances, such that their current draw is negligible, as shown in Fig. 5. An ideal voltmeter probes the potential difference between its inputs, but since no current flows through it, it does not affect the circuit. Thus, an ideal voltmeter should be connected in parallel with the device to be measured.

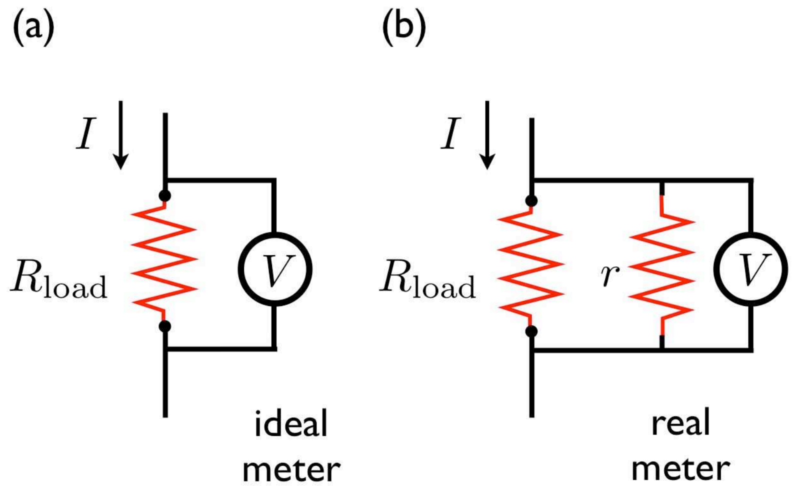


Figure 5: (a) *An ideal voltmeter has an infinite internal resistance, and no current flows through it. Hence, it measures the true voltage drop across the resistor,* $\Delta V = IR$. **(b)** *A real voltmeter has a finite internal resistance* r*, and forms a voltage divider with the load resistor. Some current flows through the voltmeter itself if* R_{load} *is comparable to* r*, and the measured voltage is less than the true voltage on the resistor.*

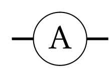
Real voltmeters have a finite internal resistance, and their current draw is not zero What the voltmeter really measures then is not just the load, but the equivalent resistance of the load in parallel with its own internal resistance r. Put another way, the voltmeter forms a *current divider* with the load, and "shunts" part of the current through the load. The voltmeter shunting part of the current obviously leads to inaccurate results, and the measured voltage drop across the resistor is no longer IR_{load} like we expect. If we assume there is a current I in the wire leading to the resistor, we can readily calculate the voltage measured by the voltmeter:

$$\Delta V_{\text{measured}} = IR_{\text{eq}} = \frac{rR_{\text{load}}}{r + R_{\text{load}}} I = \frac{IR_{\text{load}}}{1 + \frac{R_{\text{load}}}{r}} = \frac{\Delta V_{\text{expected}}}{1 + \frac{R_{\text{load}}}{r}} \tag{31}$$

The ratio between the measured voltage and the expected value is $1/(1 + R_{\text{load}}/r)$, which tells us two things. First, the measured value is always *smaller* than the true value, since $1/(1 + R_{\text{load}}) \leqslant 1$. Second, so long as the load resistor is small compared to the internal resistance of the meter, $R_{load} \ll r$, the measured and expected values will be very close. Given the enormous internal resistance of most modern voltmeters, this is usually the case, but one must still exercise caution. Using a meter with insufficient internal resistance is known as "measuring the meter," and is something you will encounter in your laboratory experiments.[xiii]

2.4 Measuring Current

An ammeter is the device that measures current, and it behaves rather differently than a voltmeter. Measuring the flow of charge has similarities with measuring the flow of fluids. A flow meter measures fluid flow by allowing the fluid of interest to pass through it. Similarly, an ammeter measures charge flow by allowing current to pass through it. Ammeters therefore connect in *series* with the device to be measured, but one should be aware that real ammeters have an internal resistance which is thus introduced in series with the load. Ammeters typically have tiny internal resistances compared to the devices of interest, and current flows readily through them. If an ammeter is connected incorrectly in *parallel* with the load, it will create current divider (parallel resistor network) with the load resistor. The small internal resistance of the ammeter can shunt most of the current from the load resistor, and an improper measurement results. Since the ammeter resistance is small, it can be connected in *series* with the load, and the voltage drop across the ammeter is usually negligible – it measures the current without disturbing the circuit.

Circuit diagram symbol for an ammeter: 

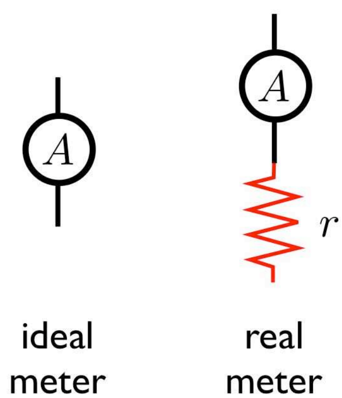


Figure 6: *A real ammeter measures current passing through it, but introduces a series resistance* r*, creating an additional voltage burden on the source.*

A simple ammeter can be constructed using a precise resistor and a good voltmeter, as shown in Fig. 7. A precise resistor placed in series with the device to be measured (in place of the ammeter in Fig. 6, for instance), and a voltmeter measures the voltage drop across this precise resistor.

[xiii] Good laboratory voltmeters can have internal resistances on the order $10^{10}\ \Omega$ or more. For meters, internal resistance is often called "input resistance."

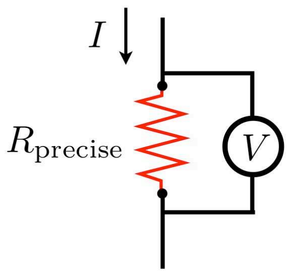


Figure 7: *A simple ammeter can be constructed from a precise resistor and a good voltmeter. Since the value of the resistance is known, the measured voltage drop across it yields the current.*

Since the value of the resistor is known precisely, the measured voltage drop across it yields the current *via* Ohm's law:

$$I = \frac{\Delta V_{\text{measured}}}{R_{\text{precise}}} \tag{32}$$

In this way currents can be measured reasonably accurately, but this is far from an ideal ammeter. First, this technique of current measurement brings in all the non-idealities associated with real voltmeters as discussed above. Second, placing a resistor within the circuit of interest introduces an additional voltage drop, which can affect other components. Care must be exercised when using this technique. The precise resistor can be chosen carefully as not to introduce a sufficiently large voltage drop to alter the circuit too much, the voltages on other components in the circuit must be independently measured to take this effect into account, or the circuit must be designed from scratch to account for this additional voltage drop. More accurate instruments, such as current preamplifiers, allow far more precise measurements with very low internal resistances. While their internal complexity is beyond the scope of the current discussion, they are in principle used just as as the ammeters discussed above.[xiv]

Finally, we leave you with a few more rules of thumb: voltmeters have a high internal resistance and connect in *parallel* with the device to be measured; ammeters have a low internal resistance and connect in *series* with the device to be measured.

3 Four point probe techniques

Armed with a knowledge of quasi-realistic devices and instruments, we are ready to discuss the most basic and essential task: measuring the resistivity of a specimen, such as a homogenous wire, a thin or thick film, or a "bulk" sample. Resistivity is of primary interest since it is a geometry-independent quantity, characteristic of a given material and its processing. However, resistivity must be determined from a geometry-dependent *resistance* measurement. Thus, in order to determine resistivity we must account

[xiv] Good laboratory current preamplifiers can have internal resistances far less than $1\,\Omega$, depending on the level of current being measured.

for the specimen and measurement geometry.

We will assume in the following discussion that the necessary precautions have been taken with regard to instrumentation, and that all sources and meters can be considered essentially ideal.[xv] Given a homogeneous conducting sample of one (wire), two (film), or three (thick film or bulk) dimensions, a primary quantity of interest is the resistivity ρ (or the conductivity $\sigma = 1/\rho$).

3.1 Four-point Probe Measurements

Resistivity measurements are commonly performed in a linear four-point geometry using current sourcing, as shown in Fig. 8. In this arrangement, the sample of interest is contacted by four collinear probes of negligible size compared to any sample dimension. The outer two probes introduce a current I, while the inner two probes measure a potential difference ΔV. It is clear that this arrangement will not be subject to artifacts due to finite wire and contact resistances as discussed in the Appendices. If the voltmeter measuring potential difference ΔV is nearly ideal (i.e., the sample resistance between points V^+ and V^- is small compared to its internal resistance), the current drawn by the voltmeter is essentially zero. Thus, the potential difference measured is characteristic of the specimen only. From the measured potential difference and the known source current, the resistivity of the sample may be determined from by geometric considerations. We need only determine for various sample geometries how the current flows outward from a source probe and the resulting potential difference at the inner probes, superposition and symmetry will do the rest.

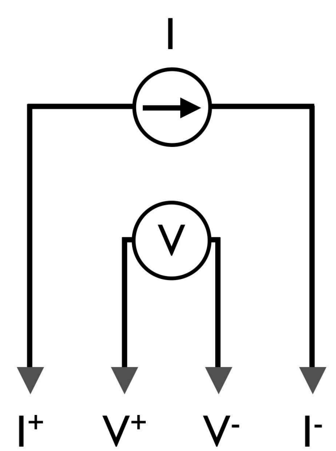


Figure 8: *Linear four-point-probe measurement. Current is introduced with the outer probes, and potential difference is measured between the inner probes.*

It is simplest to start by considering a current I introduced into an infinite bulk specimen of constant resistivity ρ through a single probe somewhere in the interior. Charge conservation dictates that the total current through a sphere of radius r centered on the probe must be constant, and thus the current density

[xv] Finite wire and contact resistances remain, however, necessitating a four-terminal measurement; see Appendix 3.

a distance r from the probe must be

$$J(r) = \frac{I}{4\pi r^2} \tag{33}$$

If the conductor obeys Ohm's law, we must have

$$J(r) = \frac{1}{\rho}E = -\frac{1}{\rho}\nabla V \tag{34}$$

The potential (relative to a distant point) due to this current a distance r from the source point then follows readily since the problem is radially symmetric and the current density is constant at a given radius r:

$$J(r) = \frac{I}{A(r)} = \frac{I}{4\pi r^2} = -\frac{1}{\rho}\frac{\partial V}{\partial r} \tag{35}$$

$$V(r) = \frac{I\rho}{4\pi r} \tag{36}$$

3.2 Bulk samples

Next, we consider a "half-infinite" bulk specimen with one free surface, Fig. 9, where the current is injected through a single probe I^+. Assuming current still spreads out uniformly, it is now confined entirely in one hemisphere, doubling the current density at an arbitrary point r and therefore doubling the potential:

$$J(r) = \frac{I}{2\pi r^2} \tag{37}$$

$$V(r) = \frac{I\rho}{2\pi r} \tag{38}$$

In the four-point probe measurement, the potential difference of interest is measured with probes V^+ and V^-, at distances a and $a+b$ from the source probe, respectively. The potentials at the probes due to the current injected at I^+ are now easily found:

$$V^+ = \frac{I\rho}{2\pi a} \quad \text{due to } I^+ \tag{39}$$

$$V^- = \frac{I\rho}{2\pi(a+b)} \quad \text{due to } I^+ \tag{40}$$

Of course, this is not the whole story: we have a second current probe I^-, our current "sink," which we

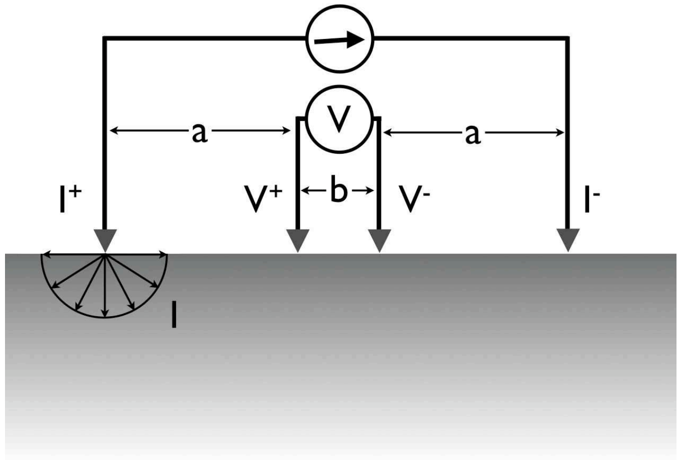


Figure 9: *Linear four-point-probe measurement on a bulk sample. The outer and inner probes are separated by a distance* a*, while the inner probes are separated by* b*. Current is introduced by the outer probes, and potential difference measured on the inner probes.*

must take into account. By symmetry, the result is the same as above if $+$ and $-$ are interchanged and the sign of the current reversed. The total potential difference between the voltage probes V^+ and V^- due to a current through both probes I^+ and I^- follows by superposition, or simply doubling the potential difference due to the single current probe I^+:

$$\Delta V = V^+ - V^- = \frac{I\rho}{\pi}\left(\frac{1}{a} - \frac{1}{a+b}\right) = \frac{I\rho b}{\pi a\,(a+b)} \tag{41}$$

This can be readily inverted to determine the resistivity ρ in terms of the measured potential difference and known current and probe spacing:

$$\rho = \frac{\pi a\,(a+b)\,\Delta V}{bI} \tag{42}$$

For the special (but common) case of equally spaced probes ($a = b$), we have

$$\rho = \frac{2\pi a \Delta V}{I} \tag{43}$$

3.3 Thin Films

More relevant for spintronics is the case of a thin film specimen, Fig. 10. Specifically, an infinite two-dimensional sheet whose thickness is small compared to the probe spacing, such that we may approximate the current density as uniform in the direction perpendicular to the film plane. In this case, the current no longer spreads evenly in a hemisphere, but is confined within the film's thickness d. In the plane of the film, the current spreads evenly leading to circular equipotential surfaces. The crucial difference is that at a lateral distance r from a current probe the total current I must pass through an area dictated by the circumference a circle of radius r and the film thickness, and thus the current density is

$$J(r) = \frac{I}{2\pi r d} \tag{44}$$

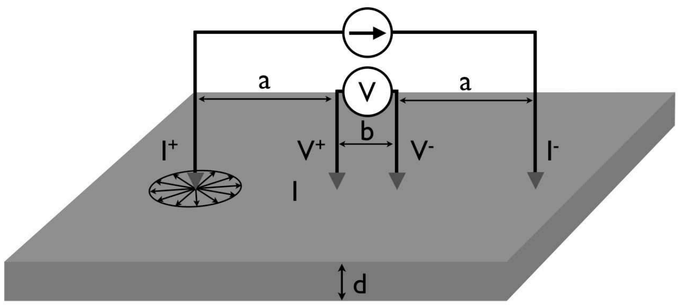


Figure 10: *Linear four-point-probe measurement on a thin film sample of thickness* d*. The outer and inner probes are separated by a distance* a*, while the inner probes are separated by* b*,* $\{a, b\} \gg d$*. Current is introduced by the outer probes, and potential difference measured on the inner probes.*

The potential at a distance r is then

$$V(r) = \frac{\rho}{2\pi d} \ln r \tag{45}$$

Following similar reasoning as above, the potential difference between the voltage probes is

$$\Delta V = \frac{\rho}{\pi d} \ln\left(\frac{a+b}{a}\right) \tag{46}$$

Which gives the resistivity of the film as

$$\rho = \frac{\pi \Delta V d}{I} \frac{1}{\ln\left(\frac{a+b}{a}\right)} \tag{47}$$

With equally spaced probes ($b = a$),

$$\rho = \frac{\pi \Delta V d}{I \ln 2} \approx 4.53 \frac{\Delta V}{I} \tag{48}$$

We note in passing that for thin film specimens, it is common to quote a *sheet resistivity*, $\rho_s = \rho/d$, the resistivity per unit thickness, as well as a *sheet resistance* R_s, the resistance per unit thickness. For equally-spaced probes the, sheet resistivity is $\rho_s = \pi \Delta V d / I \ln 2$.

3.4 Wires

Finally, in the one-dimensional case we consider a wire of sufficiently small cross-sectional area A that the current may be considered uniform along the radial direction of the wire. That is, we consider a wire whose radial dimensions are smaller than any of the contact spacings or dimensions, Fig. 11. In this case, the current spreads uniformly throughout the wire's cross section, and the current density is constant throughout, $J = I/A$. At a distance z from a current probe, the potential is

$$V(z) = \rho J z = \frac{\rho I z}{A} \tag{49}$$

and thus the potential difference between the voltage probes is

$$\Delta V = \frac{\rho I b}{A} \tag{50}$$

In the one-dimensional case, the key result is that the potential difference is independent of the spacing of the current probes a. From this the resistivity is also independent of a, and is determined by

$$\rho = \frac{\Delta V A}{I b} \tag{51}$$

As expected, in one dimension, we simply recover the usual expression for a uniform current density.

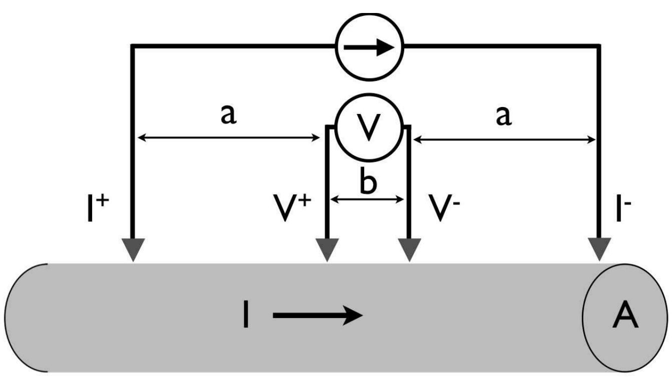


Figure 11: *Linear four-point-probe measurement on a narrow wire sample of cross-sectional area A. The outer and inner probes are separated by a distance a, while the inner probes are separated by b, $\{a, b\} \gg \sqrt{A}$. Current is introduced by the outer probes, and potential difference measured on the inner probes.*

3.5 The van der Pauw Technique

Though the resistivity determinations above are perfectly sensible, they are limited by their rather strict measurement geometry. In 1958, L.J. van der Pauw proposed a technique for measuring the resistivity of thin samples of arbitrary shapes. Due to its convenience, it is commonly used technique to measure the sheet resistance of a material. It can also be used to measure the Hall effect, which means sheet resistance, carrier type (electron or hole), carrier density, and carrier mobility can all be determined from a single set of resistance measurements (with the addition of a magnetic field for the Hall effect).

Though the full derivation of this technique is somewhat beyond the scope of this lab manual, van der Pauw's original papers on the subject are quite readable, and easily found online.[xvi] As originally devised by van der Pauw, one uses an arbitrarily-shaped (but simply connected, i.e., without holes), thin plate sample of thickness d containing four very small contacts placed on the periphery of the plate. A schematic of a quasi-rectangular configuration is shown below in Fig. 12.

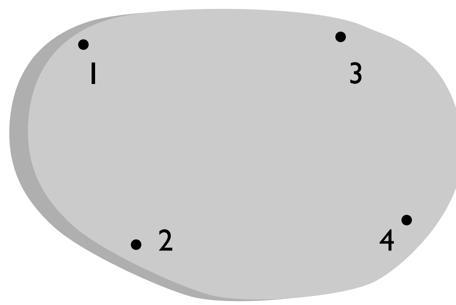


Figure 12: *Contact configuration for a van der Pauw measurement.*

The objective of the measurement is to determine the sheet resistance $R_s = \rho/d$. Using conformal mapping techniques related to the analysis above, van der Pauw demonstrated that two characteristic resistances R_a and R_b are sufficient to determine the sheet resistance through the following equation:

$$e^{-\pi R_a/R_s} + e^{-\pi R_b/R_s} = 1 \tag{52}$$

which can be numerically solved for R_s. If the thickness of the specimen d is known, ρ may be calculated. The two characteristic resistances are obtained from the configuration above in the following manner:

$$R_a = \frac{V_{43}}{I_{12}} = \frac{\text{voltage between contacts 4 and 3}}{\text{current applied through contact 1 and out of contact 2}} \tag{53}$$

$$R_b = \frac{V_{14}}{I_{23}} = \frac{\text{voltage between contacts 1 and 4}}{\text{current applied through contact 2 and out of contact 3}} \tag{54}$$

Specifically, one first sources a current I_{12} from contact 1 to contact 2, and measures the voltage developed V_{43} between contacts 4 and 3. The ratio V_{43}/I_{12} is the characteristic resistance R_a. Subsequently, sourcing a current I_{23} from contact 2 to contact 3 and measuring the voltage V_{14} between contacts 1 and

[xvi] See Philips Technical Review, vol. 20, pp. 220-224, (1958) and Philips Research Reports, vol. 13, pp. 1-9, (1958). These articles are available on the course web site in the templates directory.

4 yields the characteristic resistance R_b. In the rectangular configuration, this amounts to measuring a single resistance and then rotating the contacts by 90°, as shown in Fig. 13

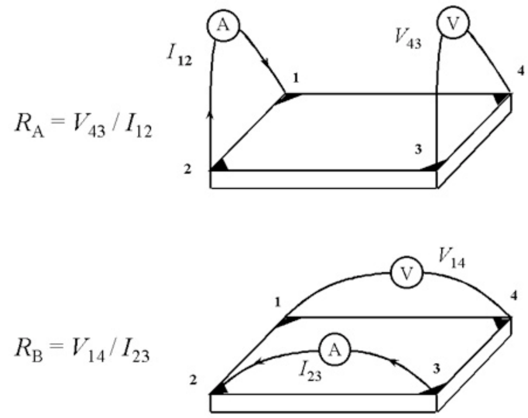


Figure 13: *Contact configuration for a van der Pauw measurement showing the determination of the two characteristic resistances* R_a *and* R_b. *From* `http://www.eeel.nist.gov/812/hall.html`.

In order to use the van der Pauw method, the sample thickness must be much less than the width and length of the sample. In order to reduce errors in the calculations, it is preferable that the sample is symmetrical. There must also be no isolated holes within the sample. Further, the contacts must be on the boundary of the sample (or as close to it as possible). Strictly, the contacts must be infinitely small. Practically, they must be as small as possible; any errors given by their non-zero size will be of the order D/L, where D is the average diameter of the contact and L is the distance between the contacts.

In addition to this, any leads from the contacts should be constructed from the same batch of wire to minimize thermoelectric effects. For the same reason, all four contacts should be of the same material.

4 Solving the van der Pauw equations numerically

NIST has published a relatively simple method for accurately solving the van der Pauw equation for sheet resistance $R_s = \rho/d$, given the two appropriate van der Pauw resistance measurements. Below, we briefly reproduce their algorithm.[xvii]

[xvii] The box below is reproduced from `http://www.eeel.nist.gov/812/samp.htm`

The sheet resistance R_s can be obtained from the two measured characteristic resistances R_A and R_B by numerically solving the van der Pauw equation:

$$e^{-\pi R_A/R_s} + e^{-\pi R_B/R_s} = 1 \tag{55}$$

using the following iterative routine:

- Set the error limit $\delta = 0.0005$, corresponding to 0.05%.

- Calculate the initial value of z_i, or

$$z_o = \frac{2\ln 2}{\pi(R_A + R_B)}$$

- Calculate the i^{th} iteration of

$$y_i = \frac{1}{\exp(\pi z_{i-1} R_A)} + \frac{1}{\exp(\pi z_{i-1} R_B)}$$

- Calculate the i^{th} iteration of z_i, where

$$z_i = z_{i-1} - \frac{(1 - y_i)/\pi}{R_A/\exp(\pi z_{i-1} R_A) + R_B/\exp(\pi z_{i-1} R_A)}$$

- When $(z_i - z_{i-1})/z_i$ is less than δ, stop and calculate the sheet resistance $R_s = 1/z_i$

- The resistivity ρ is given by $\rho = R_s d$, where d is the thickness of the conducting layer.

What follows is a very basic C program that can be used to solve for the sheet resistance, given the two appropriate resistances from a van der Pauw measurement. This file should be available for download on the course web site.

```
// Follows the NIST algorigthm from http://www.eeel.nist.gov/812/samp.htm

#include <stdlib.h>
#include <stdio.h>
#include <math.h>

#define PI 3.14159

float y_i(float z_prev, float Ra, float Rb);
float z_i(float z_prev, float y, float Ra, float Rb);
float z_o(float Ra, float Rb);
float van_der_pauw(float Ra, float Rb);

int main (int argc, const char * argv[]) {

        float Ra, Rb, Rs;
```

```c
        if (argc <=2) {
                fprintf(stdout,"Usage: 'van_der_pauw Ra Rb' \n Returns: Rs = rho / d\nAll units: Ohms\n");
                return(-1);
        }

        Ra = atof(argv[1]);
        Rb = atof(argv[2]);
        fprintf(stdout,"Ra=%g \t Rb=%g\n\n",Ra,Rb);
        Rs = van_der_pauw(Ra, Rb);
        fprintf(stdout,"Sheet Resistance (rho / d) Rs=%g\n",Rs);

        return(0);

}

float van_der_pauw(float Ra, float Rb)
{
        float Rs, err, y, z, z_prev;
        int count=0;
        float TOL = 1E-8;    /* how accurately you want to solve!*/

        z_prev = z_0(Ra,Rb);
        z=z_prev;

        do {
                y = y_i(z_prev,Ra,Rb);
                z = z_prev - z_i(z_prev,y,Ra,Rb);
                err = fabs((z-z_prev)/z);
                z_prev=z;
                count++;
        } while (err >= TOL);

        Rs = 1.0 / z;

    return (Rs);
}

float z_0(float Ra, float Rb)   /* starting value for iteration*/
{
        float z_0=0;
        z_0 = 2.0*log(2.0) / (PI*(Ra + Rb));
        return (z_0);
}

float z_i (float z_prev, float y, float Ra, float Rb) /* z_i ... calculate each iteration*/
{
        float z_i=0;
        z_i = ((1.0-y)/PI) / ( Ra/exp(PI*z_prev*Ra) + Rb/exp(PI*z_prev*Rb));
        return (z_i);
}

float y_i(float z_prev, float Ra, float Rb)
{
        float y_i=0;
```

```
        y_i = 1.0 / exp(PI*z_prev*Ra) + 1.0 / exp(PI*z_prev*Rb);
        return (y_i);
}
```

5 dI/dV and d^2I/dV^2 measurements

Many times our aim is to measure relatively small ($\lesssim 1\%$) changes in transport as a function of bias, for example, the slight increase in tunneling current when a vibrational mode in the tunnel barrier is excited. Measuring an I(V) curve one might hope to see a small change in slope at the threshold energy, but usually this is not the case. Looking at the slope of the I(V) curve is a much more accurate way to probe small changes in transport. In principle, one can numerically differentiate the I(V) curve. However, this is typically undesirable for several reasons, the most important being the rather low accuracy and noise due to discretization. With some simple mathematics, though, it can be shown that dI/dV(V) can be directly measured, with the addition of an ac modulation voltage and a lock-in amplifier.

5.1 Mathematical preliminaries

A typical I(V) measurement slowly ramps the applied dc voltage and measures the dc current through the device of interest. For a derivative measurement, in addition to the dc voltage, we additionally add a small, constant amplitude ac modulation voltage, i.e.,

$$V_{tot} = V_{dc} + \delta V_{ac} \cos \omega t$$

The current response we wish to measure is then the function

$$I(V_{tot}) = I(V_{dc} + \delta V_{ac} \cos \omega t) \tag{56}$$

where $I(V_{dc})$ would be the current-voltage characteristic in the absence of the ac modulation voltage. If we insist that the ac modulation amplitude be small compared to the dc voltages of interest, $\delta V_{ac} \ll V_{dc}$, we may perform a Taylor expansion about V_{dc}:

$$\begin{aligned}
I(V_{dc} + \delta V_{ac} \cos \omega t) &= I(V_{dc}) + \left.\frac{dI}{dV}\right|_{V_{dc}} \delta V_{ac} \cos \omega t + \frac{1}{2} \frac{d^2I}{dV^2} (\delta V_{ac})^2 \cos^2 \omega t &(57)\\
&= I(V_{dc}) + \left.\frac{dI}{dV}\right|_{V_{dc}} \delta V_{ac} \cos \omega t + \frac{1}{4} \frac{d^2I}{dV^2} (\delta V_{ac})^2 (1 + \cos 2\omega t) &(58)\\
&\approx I(V_{dc}) + \left.\frac{dI}{dV}\right|_{V_{dc}} \delta V_{ac} \cos \omega t + \frac{1}{4} \frac{d^2I}{dV^2} (\delta V_{ac})^2 \cos 2\omega t &(59)
\end{aligned}$$

for the last line, we used the power-reduction formula for $\cos$ and dropped terms of third order and higher in modulation voltage. What we see now is that the response of the system to the applied dc voltage and ac modulation has three characteristic features:

1. a dc response to the applied dc voltage $I(V_{dc})$
2. an ac response at the frequency of the ac modulation $\frac{dI}{dV}\big|_{V_{dc}} \delta V_{ac} \cos \omega t$
3. responses at the harmonics of the ac modulation

Measuring the dc current response as a function of the applied dc voltage thus reproduces the $I(V)$ characteristic, while measuring the ac current response at the frequency of the modulation voltage gives a signal proportional to dI/dV. This allows us to measure $dI/dV(V)$. For non-ohmic devices, measuring the ac current response at the second harmonic (2ω) gives us a signal proportional to d^2I/dV^2.

5.2 Ohmic and non-ohmic devices

It is instructive to see what would result when performing this measurement on an ideal Ohmic resistor. For a simple resistive device, $I = V/R$ independent of frequency. In this case,

$$I(V_{dc} + \delta V_{ac} \cos \omega t) = \frac{V_{dc}}{R} + \frac{\delta V}{R} \cos \omega t \tag{60}$$

$$\text{(shorthand)} \qquad I(V + dV \cos \omega t) = I_{dc} + dI_{ac} \cos \omega t \tag{61}$$

For convenience, we usually refer to the amplitude of the current response at the fundamental frequency as simply dI, and the dc component of the current as simply I. Similarly, we will often call the dc component of the applied voltage V and the amplitude of the ac modulation dV.

The ratio of the dc voltage to the dc component of the current for a resistive device gives us the resistance, as does the ratio of the ac modulation voltage to the ac component of the current at the fundamental frequency. More specifically, if we apply the ac modulation at frequency ω and measure the current at the same frequency ω, the amplitude of the current signal will be $\delta V/R$ as expected. For a typical measurement, one sweeps the dc voltage V and maintains a constant ac modulation dV. The measured dc response is I, and the measured response at the fundamental frequency is dI. Dividing the measured ac response by the ac modulation then gives dI/dV, and dividing the measured dc response by the applied dc voltage gives us $R = V/I$. For a simple resistive device, R independent of frequency and dc voltage, and measuring $dI/dV(V)$ should just give a constant horizontal line at $dI/dV = 1/R$.

For a non-ohmic device, we have two additional concerns. First, dI/dV will now be a function of the applied dc bias, since $I(V)$ is no longer a simple linear relationship. Thus, the $dI/dV(V)$ characteristic will no longer be constant. Second, signals will now appear at harmonics of the modulation frequency since it is unlikely that the higher derivatives (e.g., d^2I/dV^2) are zero everywhere. Measuring the current response at 2ω while driving the system with a modulation voltage at ω thus gives a signal proportional to d^2I/dV^2, and one can measure the n^{th} derivative by measuring the current response at the n^{th} harmonic of the modulation frequency. This is of course in principle; the amplitude of the resulting signal is reduced by a factor $(\delta V)^n/2n!$, making measurements of even d^2I/dV^2 more often than not quite challenging.

5.3 Circuitry

In order to accomplish the measurement described above, we need at minimum the following functions:

- dc voltage source
- dc current meter
- ac voltage source
- ac current meter

The first two are standard items, and present no problems. The third and fourth are most easily realized with a lock-in amplifier and current to voltage converter. A modern lock-in amplifier essentially provides a very stable ac voltage source and a phase-sensitive voltmeter synchronized to the same frequency. In the simplest case, the voltage to current converter can simply be a precision resistor in series with the device of interest – measuring the ac voltage across a known resistance in series allows one to determine the current. A better solution is to use a current preamplifier, which acts as a current to voltage converter without the burden of a large series resistance (often undesirable due to the extra noise introduced as well as the burden placed on the voltage sources).

In addition to these basic functions, a few ancillary items also become necessary in most practical cases.

First, one cannot simply connect the ac and dc voltage sources together. This will lead to cross-talk and increased noise, at least, and more often than not the ac source will simply not tolerate a dc voltage being applied across its output terminals (or vice versa). One solution is to couple the ac voltage into the dc-biased circuit with a transformer, but this can place serious restrictions on the available frequency range, and lead to large and undesirable inductive pickup unless careful precautions are taken. A better solution is to use an op-amp-based summing amplifier, with pre-amplifying stages to buffer the two sources from one another. This is the solution we have chosen, the details of which are below. Second, if one wishes to avoid measuring the lead and electrode resistances of the device, a four terminal measurement must be performed. This means that the device voltage must be measured with two additional wires, since the presence of any unwanted series resistances in the primary circuit will make the actual device voltage smaller than the total output voltage. Finally, amplifiers are typically necessary for both ac and dc current and voltage measurements, as the signals involved can be extremely small.

Figure 14 shows a basic block diagram of the system we have constructed, which contains all of the elements above. At the far left, the dc and ac voltage sources feed into a home-built amplifier (see Fig. 15), which contains pre-amplification and buffering stages followed by a unity-gain summing amplifier. The output of this amplifier is the dc voltage plus ac modulation, which is applied across the device under test in series with a current preamplifier. The current preamplifier outputs a voltage signal proportional to the current passing through it, both the ac and dc current signals. This output is fed directly into a dc voltmeter to measure the dc component of the current, and through an additional amplifier to a lock-in amplifier to measure the ac component of the current. With two separate wires, the device's ac and dc voltages are measured with a second lock-in amplifier and dc voltmeter, respectively, after passing through an additional signal amplifier.

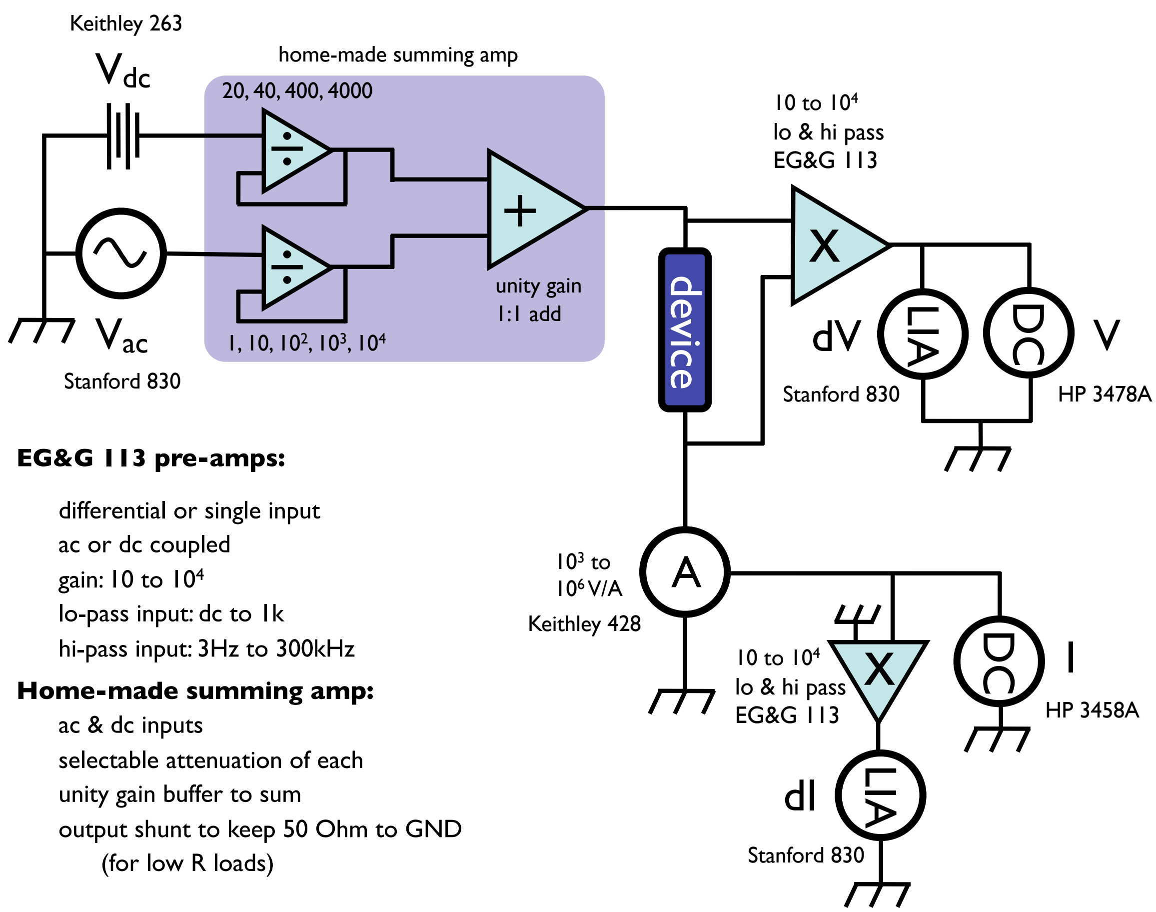


Figure 14: *Block diagram of the electronics used for measuring* dI/dV *and* d^2I/dV^2.

5.3.1 Sourcing

The dc voltage is supplied by a Keithley 263 Calibrator/Source (K263), which can supply ±20 V. The sample voltages we desire are usually in the range of 1 V or less, and often on the scale of 1 mV. In order to avoid discretization noise, we put out a relatively large voltage and attenuate it. The attenuation factor is selectable with a four-position switch: 20, 40, 400, 4000. For instance, we might source 4 V from the K263, but attenuate it by a factor 4000 to supply 1 mV to the sample and its wiring.

To illustrate the advantage in attenuation, let us assume that the 20 V range is covered by a 12 bit analog to digital converter. That means that the 20 V range is split in to 4096 discrete voltage values, so the voltage can only take on discrete values in increments of ~4.9 mV. If we desire a modulation voltage of 3 mV, and can only put out a minimum voltage of 4.9 mV, we have a relative error of over 60%. Attenuating by a factor of 400 gives us a step size of 12 μV, and a relative error of only 0.4%.

The ac voltage is supplied by the output of the Stanford 830 (SR830) lock-in amplifier, which has a range of $0-5$ V. The ac voltage is used as a modulation for measuring dI/dV, for example, and therefore must be small compared to the dc voltage. In extreme cases, it might be only 20 μV peak-to-peak. As with the dc voltage, we apply a fairly large signal from the SR830, and attenuate it by a large factor, selectable in 5 ranges from 1 to 10k. This ac modulation represents the dV part of the dI/dV measurement. In order to have a constant resolution measurement, its magnitude must remain constant. However, since we typically measure non-linear elements such as tunnel junctions, it is not sufficient to simply set this voltage once for a given measurement. Software feedback (a simple PID loop) is employed to constantly maintain the ac modulation on the sample of interest to within 10% or better of the desired value (more details below).

A dI/dV or d^2I/d^2V measurement involves sourcing the dc sample voltage with a small ac modulation voltage superimposed on it, which means we must add the two source voltages. The attenuation for both ac and dc voltages is performed by a home-made amplifier, pictured in Fig. 15, which also serves to buffer the two signals from each other. They are then added in a 1 : 1 ratio with a home-made summing amplifier.

5.3.2 Measuring Voltage

Measuring the sample voltage is accomplished with two independent leads to the sample (for a true four-point measurement), which are fed into a battery-powered EG&G 113 wideband amplifier (gain $10-10$k, selectable lo- and hi-pass filters). The dc portion of the signal is measured by a HP 3478A digital voltmeter to measure the dc voltage on the sample, while the ac voltage on the sample, the dV part of dI/dV, is measured by the same SR830 lock-in amplifier that supplies the ac modulation. Since the sample resistance is not always large compared to the wiring (not to mention the fact that the amplifier gain might not be precisely constant) it is necessary to separately measure the actual ac voltage on the

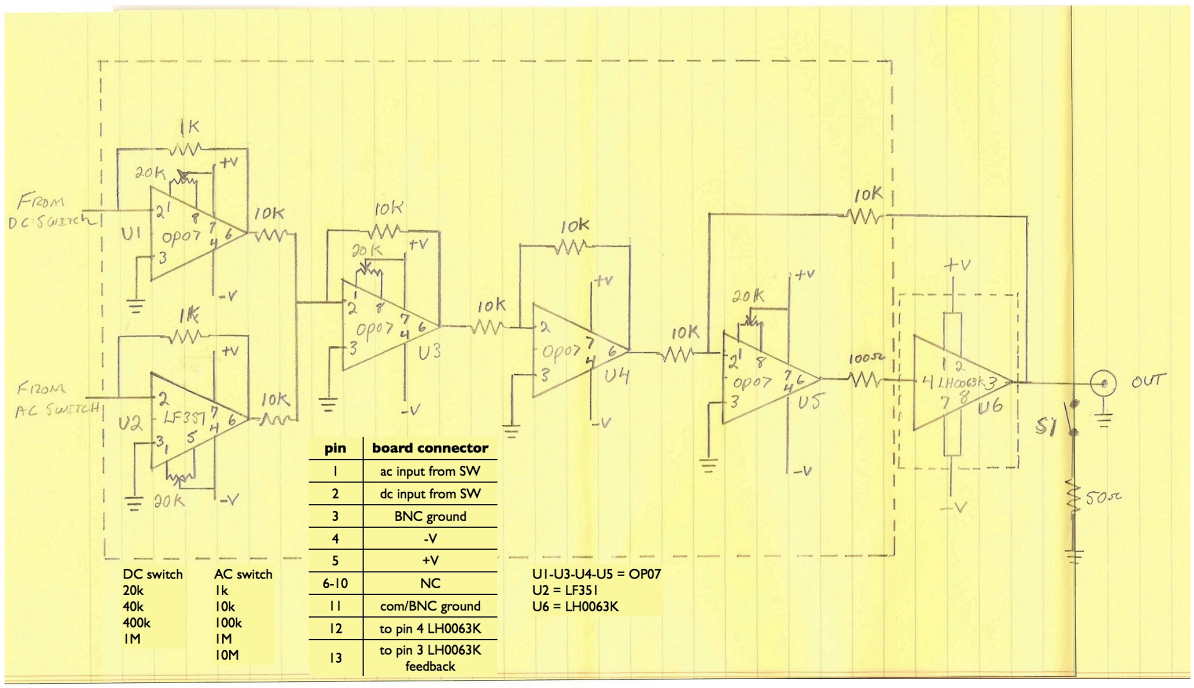


Figure 15: *Schematic of the home-built amplifier for attenuating, buffering, and summing the ac and dc voltages. Courtesy D. Whitcomb.*

sample. The ac voltage measured on the sample is compared with the desired ac modulation given by the user, and a simple software PID loop adjusts the lock-in amplifier output to maintain the actual ac voltage on the sample to with 10% of the desired value.

5.3.3 Measuring Current

Measuring the sample current is accomplished by a Keithley 428 (K428) current pre-amplifier in series with the sample of interest. The current pre-amplifier appears (essentially) as a zero-ohm load to the sources, and outputs a voltage signal proportional to the current through it. It has a selectable gain of $10\mathrm{k} - 100\mathrm{G}\,\mathrm{V/A}$, so measuring pA currents is possible on a good day. This signal is fed directly into a HP 3458A digital voltmeter to measure the dc current. The ac current signal (the $\mathrm{d}I$ part of $\mathrm{d}I/\mathrm{d}V$) is fed through a battery-powered EG&G 113 wideband amplifier (gain 10–10k, selectable lo- and hi-pass filters) and then to a Stanford Research 830 lock-in amplifier to measure the $\mathrm{d}I$ signal. This lock-in amplifier is synchronized with the ac source from the first lock-in at f (for $\mathrm{d}I/\mathrm{d}V$) or $2f$ (for $\mathrm{d}^2I/\mathrm{d}V^2$).

The EG&G amplifiers are necessary for two main reasons. First, the ac voltage and current signals can be quite small, and amplification can be necessary. Second, the input impedance of the lock-in amplifiers is only $10\,\mathrm{M}\Omega$a, so these high-input-impedance amplifiers act as buffers when measuring high resistance samples. The current pre-amplifier is preferable over a simple resistive shunt to reduce noise: having a small resistance in series for measuring current leads to small and noisy signals, while having a large resistance in series leads increased Johnson-Nyquist noise.

5.3.4 Miscellanea

5.4 Software

5.4.1 Basic principles

5.5 Further practical considerations

6 Tunnel Junctions & other non-linear elements

6.1 Simmons & Brinkman models

6.2 dI/dV measurements

6.3 d^2I/dV^2 measurements

6.4 Parallel RC model

A tunnel junction is a metal-insulator-metal structure, which means that it also acts as a parallel-plate capacitor. One may view the structure as a "leaky" capacitor, modeled as a parallel resistor and capacitor. For most of our work involving tunnel junctions, the resistive current dominates the device response at the measurement frequency, and *most of the time* we can consider the device to be purely resistive. However, there is no sharp boundary between a purely resistive junction and a purely capacitive one: essentially all devices we measure will have both resistance and capacitance, we only try to ensure that most of the time the former is dominant and the latter negligible. The conduction through a tunnel junction at finite frequency can be resistive via tunneling or via the capacitance, and there are simple mechanisms for determining which contribution might be dominant. Of course, for simple dc measurements, the capacitive component causes us no concern beyond setting a minimum $\tau = RC$ time scale for measurement perturbations.

6.4.1 Magnitude of the capacitive current

For most work involving tunnel junctions, the capacitive current is considered to be parasitic, and should be avoided. Hence, in doing an ac measurement, care must be taken that most of the current across the barrier is due to tunneling process, and not by capacitance. If we model the tunnel junction as a resistor R and capacitor C in *parallel*, we can get a feeling for how large this parasitic effect can be under typical conditions.

Of course, at "sufficiently" high frequency, most current will be via the capacitor, since its impedance is inversely proportional to the driving frequency, $Z_c = 1/i\omega C$. At "sufficiently" low frequency, the current is mostly due to tunneling because the capacitor has essentially an infinite impedance. If we take as our criterion in choosing the excitation frequency that the resistive current should be at least 10 times the capacitive current:

$$\left|\frac{I_R}{I_C}\right| \geqslant 10 \tag{62}$$

Conservation of current dictates that

$$I_s \cos(\omega t) = I_C + I_R = -\omega C V_0 \sin(\omega t + \alpha) + \frac{V_0}{R} \cos(\omega t + \alpha) \tag{63}$$

where $I_s \cos(\omega t)$ is the excitation current, $V = V_0 \cos(\omega t + \alpha)$ is the voltage across the junction, and α is the phase lag in the voltage response. Combining (62) and (63), we have

$$\frac{1}{\omega RC} \geqslant 10 \qquad \Longrightarrow \qquad \omega = 2\pi f \leqslant \frac{1}{10RC} = 0.1\tau^{-1} \qquad \Longrightarrow \qquad f \lesssim 0.016\tau^{-1} \tag{64}$$

In other words, the modulation frequency should be about 2% of the inverse RC time constant of the junction for the resistive contribution to dominate. We can estimate the capacitance of a typical (large) junction of area $300 \times 300\,\mu\text{m}^2$ and barrier thickness of 2.0 nm, assuming the relative dielectric constant $\epsilon_r \sim 10$:

$$C = \frac{A\epsilon_0\epsilon_r}{d} = \frac{(300 \times 300 \times 10^{-12})(8.85 \times 10^{-12})(10)}{2 \times 10^{-9}}\,\text{F} = 1\,\text{nF} \tag{65}$$

If we assume an excitation frequency of 1 kHz, then we can estimate the maximal resistance for which the tunnel current is still an order of magnitude greater than the capacitive current:

$$R \leqslant \frac{1}{\omega C_J} = \frac{1}{2\pi f C_J} = \frac{1}{(2\pi)(1000)(1 \times 10^{-9})} \approx 40\,\text{k}\Omega \tag{66}$$

Within this crude model, then, for resistances higher than $\approx 40\,\text{k}\Omega$, a 1 kHz excitation frequency will be too high to be considered purely resistive. Experimentally, whether the current through the barrier is mostly resistive or capacitive can be easily checked by observing the phase of the current signal. If the current is mostly resistive, the current signal should be more or less in-phase (say, $\pm 10°$) with the modulation voltage. If the phase shift is significantly larger, this typically means the capacitance is too high, and a lower excitation frequency should be used to ensure that the resistive contribution is dominant.

6.4.2 Parallel RC model in more detail

Let us assume for the moment that we are unlucky, and the resistive current does *not* dominate, or at least that we are not certain. If the electrode and lead resistances are negligible, we can model the tunnel junction itself as a parallel RC circuit driven by a time-varying signal V_{in}. The currents through the resistor I_r and the capacitor I_c and the total current I_t are readily determined:

$$I_c = i\omega C V_{in} \tag{67}$$

$$I_r = V_{in}/R \tag{68}$$

$$I_t = \sqrt{I_r^2 + I_c^2} = V_{in}\sqrt{\frac{1}{R^2} + \omega^2 C^2} \tag{69}$$

The total impedance Z_t of this circuit, the ratio of V_{in} to the total current I_t, is

$$Z_t = \frac{V_{in}}{I_t} = \frac{R}{1+\omega^2 C^2 R^2}\left(1 - i\omega RC\right) \tag{70}$$

$$|Z_t| = \sqrt{\frac{R^2}{1+\omega^2 R^2 C^2}} \tag{71}$$

The phase angle for the current is given by

$$\tan\alpha = \omega RC \tag{72}$$

This already illustrates one important point: if the capacitance is negligible, the phase angle of dI approaches zero, whereas if the capacitance becomes large, the phase angle approaches $90°$. This is the easiest way to determine which contribution dominates: if the phase is near zero, the device is predominantly resistive; if the phase is near $90°$, the device is predominantly capacitive. Anything in between, and a more complete analysis must be performed.

Since this is a parallel circuit, the voltages on the resistor and capacitor are simply equal to V_{in}. In our measurement, we will measure V_{in}, the total current I_t, and its phase α, and we wish to extract R and C.

From Eq. 70 or 69 we can already calculate dI/dV for the circuit as a whole:

$$\frac{dI}{dV} = \sqrt{\frac{1}{R^2} + \omega^2 C^2} \tag{73}$$

This will be the result of a measurement of dI/dV which consists of dividing the *magnitude* of the ac current by the *magnitude* of the ac modulation voltage, with the phase of the dI signal being α relative to the modulation voltage. If the capacitance is negligible, we recover the familiar result $dI/dV = 1/R$ for an ohmic resistor, while if the capacitance is dominant, we find $dI/dV = i\omega C = 1/Z_c$.

Upon rearranging, we can extract the resistive and capacitive contributions from the measured dI/dV and phase angle α:

$$\frac{1}{R} = \frac{dI/dV}{\sqrt{1+\tan^2\alpha}} \tag{74}$$

$$C = \frac{dI/dV}{\omega\sqrt{1+1/\tan^2\alpha}} \tag{75}$$

This brings up the last important point: by measuring the *magnitude* of dI/dV (dividing the *magnitude* of the ac current by the *magnitude* of the ac modulation voltage) and the phase angle α, one can extract both the resistive and capacitive contributions to conduction. For example, one can simultaneously measure variation of conductance with dc bias and the variation of capacitance with voltage. As another example, what we typically consider to be a conductance measurement is also a form of dc-bias-dependent impedance spectroscopy, if the measurement is performed at various frequencies.

Finally, the table below gives a few numerical examples. As a rule of thumb, when the phase α is greater than $\sim 5-10^\circ$, the capacitive contribution is no longer negligible. At our standard operating frequency of $1\,\mathrm{kHz}$, junctions below $R \sim 10\,\mathrm{k}\Omega$ typically present no problems, while above $R \sim 100\,\mathrm{k}\Omega$ problems are frequently encountered (consistent with our estimated capacitance of $\sim 1\,\mathrm{nF}$ for a large junction). Reducing the frequency to $100\,\mathrm{Hz}$ has marginal gain, and reducing the frequency further both increases noise and unacceptably increases measurement time. For very high resistance junctions ($\gtrsim 100\,\mathrm{k}\Omega$) the capacitive contribution is almost always a factor. Not to mention the fact that the resistive signal is very small, presenting its own problems.

Table 1: *Impedance and phase angle for parallel* RC *circuits.*

	$R = 1\,k\Omega$, $f = 1\,kHz$						
C	1pF	10pF	100pF	1nF	10nF	100nF	1µF
$\|Z\| = 1/dI/dV$ (Ω)	1,000.00	1,000.00	1,000.00	999.98	998.03	846.85	157.25
α (°)	0.00	0.00	0.04	0.36	3.59	32.13	80.95
	$R = 10\,k\Omega$, $f = 1\,kHz$						
C	1pF	10pF	100pF	1nF	10nF	100nF	1µF
$\|Z\| = 1/dI/dV$ (Ω)	10,000.00	10,000.00	9,999.80	9,980.34	8,468.55	1,572.54	159.22
α (°)	0.00	0.04	0.36	3.59	32.13	80.95	89.09
	$R = 100\,k\Omega$, $f = 1\,kHz$						
C	1pF	10pF	100pF	1nF	10nF	100nF	1µF
$\|Z\| = 1/dI/dV$ (Ω)	99,999.98	99,998.03	99,803.39	84,685.45	15,725.45	1,592.15	159.24
α (°)	0.04	0.36	3.59	32.13	80.95	89.09	89.91
	$R = 1\,k\Omega$, $f = 10\,kHz$						
C	1pF	10pF	100pF	1nF	10nF	100nF	1µF
$\|Z\| = 1/dI/dV$ (Ω)	1,000.00	1,000.00	999.98	998.03	846.85	157.25	15.92
α (°) 0.00	0.04	0.36	3.59	32.13	80.95	89.09	
	$R = 1\,k\Omega$, $f = 100\,Hz$						
C	1pF	10pF	100pF	1nF	10nF	100nF	1µF
$\|Z\| = 1/dI/dV$ (Ω)	1,000.00	1,000.00	1,000.00	1,000.00	999.98	998.03	846.85
α (°)	0.00	0.00	0.00	0.04	0.36	3.59	32.13
	$R = 100\,k\Omega$, $f = 100\,Hz$						
C	1pF	10pF	100pF	1nF	10nF	100nF	1µF
$\|Z\| = 1/dI/dV$ (Ω)	100,000.00	99,999.98	99,998.03	99,803.39	84,685.45	15,725.45	1,592.15
α (°)	0.00	0.04	0.36	3.59	32.13	80.95	89.09

6.4.3 Impedance spectroscopy

6.5 Other concerns

7 Noise in resistive devices

In electronics and communication systems, noise is a random fluctuation or variation of an electromagnetic analog signal such as a voltage or a current. Electronic noise is a characteristic of all electronic circuits. Depending on the circuit, the noise generated by electronic devices can vary greatly and arise through several different mechanisms. Contributions such as thermal noise and shot noise are inherent to all devices, while other types depend mostly on manufacturing quality and defects. Though noise usually has a negative connotation, it does have its uses – for instance, noise power is used in low-temperature thermometry, and the study of noise can be a powerful technique for elucidating the microscopic mechanisms of conduction.

A noise signal is typically considered as a linear addition to a useful information signal, typified in Fig. 16 where a noise signal is superimposed on constant voltage signal. Noise is a random process, characterized by stochastic properties such as its variance, distribution, and spectral density. The spectral distribution of noise can vary with frequency, so its power density is measured in watts per hertz (W/Hz). Since the power in a resistive element is proportional to the square of the voltage across it, noise voltage (density) can be described by taking the square root of the noise power density, resulting in volts per root hertz ($V/\sqrt{Hz}$). Integrated circuit devices, such as operational amplifiers commonly quote equivalent input noise level in these terms (at room temperature).

Noise levels are usually viewed in opposition to signal levels and so are often seen as part of a signal-to-noise ratio (SNR). Typical signal quality measures involving noise are signal-to-noise ratio (SNR or S/N), signal-to-quantization noise ratio (SQNR) in analog-to-digital conversion, and compression, peak signal-to-noise ratio (PSNR) in image and video coding, carrier to noise ratio (CNR) before the detector in carrier-modulated systems, and noise figure in cascaded amplifiers.

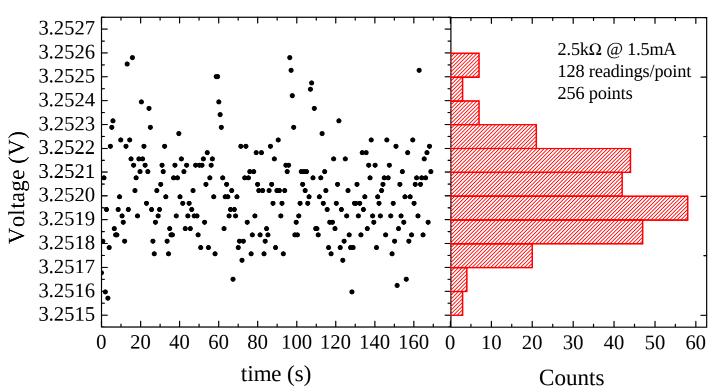


Figure 16: (left): *Time-domain signal showing random noise.* **(right):** *Distribution of the noise about the average voltage, approximately Gaussian.*

The fluctuation of, e.g., a voltage signal about its mean value is governed by a random process, namely the collisions of the electrons making up the current. In this regard, the analysis of electrical noise will closely parallel the analysis of counting statistics in radioactive decay, and both electrical signal and counting noise can be reduced through signal averaging techniques. Crucial differences arise, however, due to the differing physical origins of the fluctuations in the two systems. In the case of electrical noise, one may also effectively reduce fluctuations by controlling the measurement temperature and bandwidth and the overall resistance of the device under test.

7.1 Thermal Noise

Thermal noise, or Johnson-Nyquist noise, is the electrical noise generated by random thermal agitation of the charge carriers in an electrical conductor at equilibrium. The direction of the movement of the charge carriers is changed by collisions with, for example, the host crystal, impurities and other electrons, resulting in a random movement at zero voltage (Brownian motion), and a certain degree of randomness in their movement at finite voltages. This random movement gives rise to fluctuations in the net current in the conducting material. These fluctuations time-average to zero, but the time-average of the *square* of the fluctuations is not. Since thermal noise results from a random process, its amplitude very nearly follows a Gaussian probability density function, as shown in Fig. 16. [xviii]

Since the movement of electrons is only correlated within the time between two collisions, there is essentially no frequency-dependence of the noise for frequencies lower than the reciprocal collision time $1/\tau$ of the electrons. The collision time determines the conductivity of a material. In Cu, for example, $\tau \approx 2.5 \times 10^{14}$ s, so below $1/\tau \approx 40$ THz thermal noise is frequency-independent, or "white."[xix]

Thermal noise can be modelled as a resistor representing the device under test in series with a random voltage source, or in parallel with a random current source. This is because by using Thévenin and Norton equivalents, any linear electrical network consisting of combinations of voltage sources, current sources, and resistors can be replaced with a single resistor and a series voltage or parallel current source. Thus, an ideal, noise-free resistive device can be replaced by an ideal, noise-free resistor in series with a random voltage source. The root mean square noise voltage produced in a resistor R at temperature T is given by

$$\langle V^2 \rangle = V_{rms}^2 = \langle V \rangle^2 + \sigma^2 = 4k_B TR\Delta f \tag{76}$$

[xviii] Thermal noise is, formally, distinct from "shot noise," a type of electronic noise that occurs when the finite number of electrons is small enough to give rise to detectable statistical fluctuations. While shot noise will dominate only at very low currents, thermal noise is always present.

[xix] The frequency spectrum is the Fourier transform of the signal in the time domain. If the signal is random in the time domain, it can be thought of as containing all possible frequencies superimposed in equal amounts.

where k_B is Boltzmann's constant and Δf the bandwidth of the measurement. If your measurement uses amplifiers with a lower cutoff frequency f_l and an upper cutoff frequency f_h, the effective bandwidth of the measurement can be expressed as

$$\Delta f = \frac{\pi}{2}\frac{f_h^2}{f_h + f_l} \approx \frac{\pi}{2} f_h \qquad (f_h \gg f_l) \tag{77}$$

As a concrete example, let's suppose that $R = 10\,\mathrm{k\Omega}$, $T = 295\,\mathrm{K}$, and our bandwidth is set by filters of 1 Hz and 100 kHz ($\Delta f \approx 157$ kHz). Then we find $V_{\text{rms}} \approx 16\,\mu\mathrm{V}$. This represents the limit for the smallest voltage we can resolve across this resistor in this bandwidth. Note that if we put two resistors in series, the mean square voltage is given by

$$V_{\text{rms, tot}}^2 = 4k_B T\,(R_1 + R_2) = V_{\text{rms, 1}}^2 + V_{\text{rms, 2}}^2 \tag{78}$$

The noise powers add, **not** the noise voltages. For an arbitrary resistive circuit, we can find the equivalent noise by using a Thevenin (Norton) equivalent circuit or by transforming all noise sources to the output by the appropriate power gain (e.g. voltage squared or current squared).

We can already draw two important conclusions: the noise in a given resistive device can be effectively reduced by reducing the temperature of the device, or by reducing the measurement bandwidth. While the former is not always feasible, a number of clever solutions exist for reducing measurement bandwidth. A simple technique we will use here is *signal averaging*, which reduces the bandwidth by repeatedly sampling the quantity of interest. More sophisticated techniques, such as lock-in detection (modulation) will be explored in future labs.

A third factor influencing the noise voltage is the resistance of the circuitry under study. Even in the simplest possible system, a short-circuited resistor, the noise voltage produced increases as $\sqrt{R}$, and a noise current $\sqrt{\langle I^2 \rangle} = \sqrt{\langle V^2 \rangle}/R = \sqrt{4k_B T \Delta f / R}$ must be produced. Decreasing the resistance of the circuit under study, if feasible, will reduce the noise in the measured voltage at the expense of the noise in the measured current. Thermal noise is intrinsic to all resistors and is not a sign of poor design or manufacture, although resistors may also have excess noise.

In addition to thermal noise in the device under study, one must also consider the noise present in the components making up the sources, amplifiers, and meters used. For example, consider a simple circuit consisting of a single resistor connected to a current source, and we wish to measure the voltage produced on the resistor. Independent of the current level, the thermal noise voltage measured on the resistor will remain the same. However, the current source itself is not perfect, and the current supplied will have its own variance $\langle I^2 \rangle$. These current fluctuations lead to additional voltage fluctuations measured on the resistor. Since the current source fluctuations are independent of the thermal fluctuations in the resistor, we must add the variances in quadrature (i.e., add the noise powers):

$$\langle V^2 \rangle_{\mathrm{tot}} = \langle V^2 \rangle_{\mathrm{source}} + \langle V^2 \rangle_{\mathrm{thermal}} \tag{79}$$

7.2 Random Telegraph Noise

Random telegraph noise, also called burst, popcorn, impulse, or bi-stable noise, is another type of electronic noise that is frequently encountered, particularly in semiconductors. This type of noise consists of sudden step-like transitions between two or more discrete voltage or current levels, as high as several hundred microvolts, at random and unpredictable times. An example is shown in Fig. 17. Each shift in offset voltage or current often lasts from several milliseconds to seconds, and sounds like popcorn popping if hooked up to an audio speaker. No single source of popcorn noise is theorized to explain all occurrences, however the most commonly invoked cause is the random trapping and release of charge carriers at thin film interfaces or at defect sites in bulk semiconductor crystal. For the present experiment, we wish to avoid telegraph noise as much as possible.

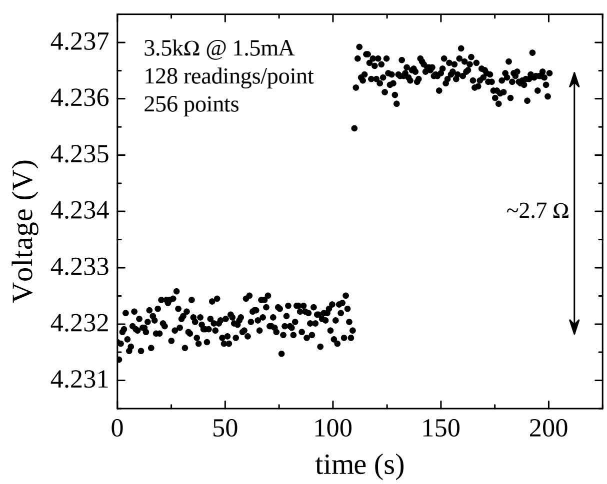


Figure 17: *A voltage signal exhibiting telegraph noise in addition to random noise. The jump in resistance is* $\sim 0.08\%$.

8 Signal averaging

Once the sources of noise in a measurement are understood, the usual question is how they may be reduced or eliminated. Thermal noise cannot be eliminated, but can be reduced by reducing device resistance, temperature, or bandwidth. Once these parameters have been controlled, the remaining noise can be further reduced through signal averaging techniques. In its simplest form, signal averaging just means the repeated measurement of the quantity of interest.

As a more concrete example, our goal is to measure the voltage across a resistor. Rather than making only a single measurement, we may improve our accuracy by taking N measurements and reporting the mean result $\overline{V}$. As we make more and more measurements, the uncertainty in our mean voltage compared to the true voltage is given by the standard deviation of the mean

$$s_{\overline{V}} = \frac{s}{\sqrt{N}} \tag{80}$$

where s is the standard deviation of our collection of measurements. The relative uncertainty of the mean decreases as $1/\sqrt{N}$, and this is at the heart of signal averaging: repeated measurements lead to increased accuracy, at the price of increased measurement time. The latter point cannot be overestimated. While the uncertainty in the mean voltage is reduced as $\mathcal{O}(1/\sqrt{N})$, the measurement time increases as $\mathcal{O}(N)$, and thus there exists for any measurement, due to practical considerations, a point at which further reduction of noise through signal averaging is no longer feasible. For example, to reduce the signal to noise ratio by a factor of two, a factor of four increase in the number of measurements, and thus measurement time, is required.

Figure 18 shows an example of the effect of averaging repeated measurements. In this measurement, a $500\,\Omega$ resistor was measured with a $1.5\,\text{mA}$ constant current, and the measurement was repeated up to 1024 times. The ratio of the average voltage to the standard deviation – the signal to noise ratio – increases steadily as $\sqrt{N}$.

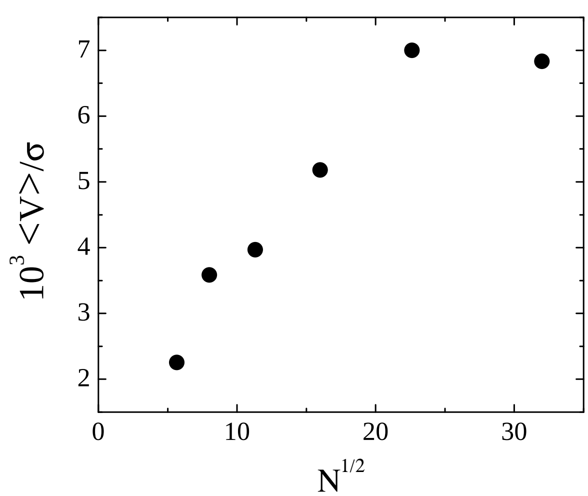


Figure 18: *Effect of point averaging on a voltage signal. The relative uncertainty $\sigma/\overline{V}$ decreases as $\sqrt{N}$, while the signal-to-noise ratio $\overline{V}/\sigma$ increases as $\sqrt{N}$.*

9 Summary of Noise and Averaging

The root-mean square (RMS) voltage is the sum of the squared average voltage and the squared standard deviation for a series of N measurements V_i is

$$\langle V^2 \rangle = \sqrt{\frac{\sum_{i=1}^{N} V_i^2}{N}} = V_{rms}^2 = \langle V \rangle^2 + \sigma^2 = 4k_B TR\Delta f \tag{81}$$

For thermal noise, the RMS voltage is governed by the resistance R and temperature T of the device under study and the measurement bandwidth Δf. With no dc component present,

$$\langle V^2 \rangle = V_{rms}^2 = 4k_B TR\Delta f \tag{82}$$

If we drive a resistive device with a current source which also has inherent fluctuations, the total noise voltages from the resistive load and current source add in quadrature. That is, we add the noise powers:

$$\langle V^2 \rangle_{tot} = \langle V^2 \rangle_{source} + \langle V^2 \rangle_{thermal} \tag{83}$$

The signal to noise ratio (SNR) when a dc component is present is defined as the ratio of the mean voltage to its standard deviation. For a large collection of measurements of size N, the SNR increases as $\sqrt{N}$:

$$SNR = \frac{\overline{V}}{\sigma} \propto \sqrt{N} \tag{84}$$