Statistical Tools for Analyzing Measurements of Charge Transport

Statistical Tools for Analyzing Measurements of Charge Transport

William F. Reus,[1] Christian A. Nijhuis,[2] Jabulani R. Barber,[1] Martin M. Thuo,[1] Simon Tricard,[1] George M. Whitesides[1,3]*

[1] Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA, 02138

[2] Department of Chemistry, National University of Singapore, 3 Science Drive, Singapore 117543

[3] Kavli Institute for Bionano Science & Technology, Harvard University, School of Engineering and Applied Sciences, Pierce Hall, 29 Oxford St., Cambridge, MA 02138

*Author to whom correspondence should be addressed: gwhitesides@gmwgroup.harvard.edu

Abstract: This paper applies statistical methods to analyze the large, noisy datasets produced in measurements of tunneling current density (J) through self-assembled monolayers (SAMs) in large-area junctions. It describes and compares the accuracy and precision of procedures for summarizing data for individual SAMs, for comparing two or more SAMs, and for determining the parameters of the Simmons model (β and J_0). For data that contain significant numbers of outliers (i.e. most measurements of charge transport), commonly used statistical techniques—e.g. summarizing data with arithmetic mean and standard deviation, and fitting data using a linear, least-squares algorithm—are prone to large errors. The paper recommends statistical methods that distinguish between real data and artifacts, subject to the assumption that real data (J) are independent and log-normally distributed. Selecting a precise and accurate (conditional on these assumptions) method yields updated values of β and J_0 for charge transport across both odd and even n-alkanethiols (with 99% confidence intervals), and explains that the so-called odd-even effect (for n-alkanethiols on Ag) is largely due to a difference in J_0 between odd and even n-alkanethiols. This conclusion is provisional, in that it depends to some extent on the statistical model assumed, and these assumptions must be tested by future experiments.

Introduction

Understanding the relationship between the atomic-level structure of organic matter, and the rate of charge transport by tunneling across it, is relevant to fields from molecular biology to organic electronics. Self-assembled monolayers (SAMs) should, in principle, be excellent substrates for such studies.[i] In practice, the field has proved technically and experimentally to be very difficult (for reasons we sketch, at least in part, in following sections), and measurements of rates of tunneling across SAMs have produced an abundance of data with often uncharacterized reliability and accuracy.

Although a number of experimental factors contribute to the difficulty of the field, there is an additional problem: namely, analysis of data. Many of the experimental systems used to measure tunneling across SAMs generate noisy data (in some cases, for reasons that are intrinsic to the type of measurement, in some cases because of poor experimental design or inadequate control of experimental variables). Regardless, with the exception of work done using scanning probe techniques[ii,iii,iv,v,vi,vii,viii] and break junctions,[ix] and by Lee *et al.*,[x] the data have seldom been subjected to tests for statistical significance, and papers have sometimes been based on selected data, or on (perhaps) meaningful data winnowed from large numbers of failures, without a rigorous statistical methodology.

We have worked primarily with a junction composed of three components: i) a "template-stripped"[xi] silver or gold electrode (the "bottom" electrode)—template stripping provides a relatively flat (rms roughness = 1.2 nm, over a 25 μm^2 area of Ag) surface;[xi] ii) a SAM; and iii) a top-electrode, comprising a drop of liquid eutectic GaIn alloy, with a surface film of (predominantly) Ga_2O_3. (We abbreviate this junction as

"Ag^{TS}-SR//Ga_2O_3/EGaIn", following a nomenclature described elsewhere.[xii,xiii,xiv]) Figure 1 shows a schematic of an assembled junction, including examples of defects in the substrate, SAM, and top-electrode that affect the local spacing between electrodes. (The composition of the junction has been discussed elsewhere in detail[xiv]). The metric for characterizing charge transport through this junction is the current density (J, A/cm^2) as a function of applied voltage (V). We calculate J by dividing the measured current by the cross-sectional area of the junction, inferred (assuming a circular cross-section) from the measured diameter of the contact between the Ga_2O_3/EGaIn top-electrode and the SAM.

This paper is a part of a still-evolving effort to use statistical tools to analyze the data generated by this junction, and to identify factors that contribute to the noise in the data. This analysis is important for our own work in this area, of course. It is also—at least in spirit—important in analyzing data generated using many SAM-based systems. We acknowledge that our analysis contains a number of approximations. It is, however, extremely useful in identifying sources of error, and in providing the basis of future evolutions of these types of systems into simpler and more reliable progeny.

To develop and demonstrate our analysis, we use data collected[xiv] across a series of n-alkanethiolate SAMs, $S(CH_2)_{n-1}CH_3$, for n = 9 – 18. Figure 2 powerfully conveys the magnitude of the challenge faced by any analysis of charge transport in Ag^{TS}-SR//Ga_2O_3/EGaIn junctions (and, we strongly suspect, other systems as well). It shows two histograms (see the Supporting Information for details on plotting histograms) of J on a log-scale for n-alkanethiols at opposite ends of the series: the histogram of $S(CH_2)_{17}CH_3$ (black) is superimposed on that of $S(CH_2)_9CH_3$ (gray). Given that the

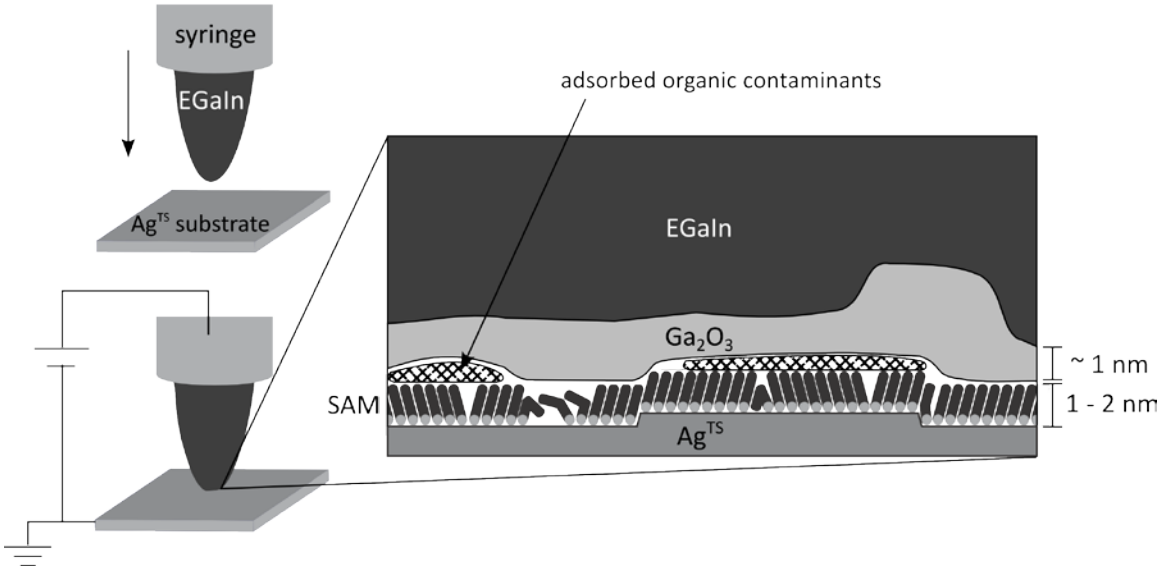


Figure 1:

The formation and structure of an Ag^{TS}-SR//Ga_2O_3/EGaIn junction. To form the junction, a conical tip of Ga_2O_3/EGaIn, suspended from a syringe, is lowered into contact with a SAM on a Ag^{TS} substrate. The substrate is grounded; an electrometer applies a voltage to the syringe and measures the current flowing through the junction. The schematic representation of the junction shows defects in the Ag^{TS} substrate and SAM, as well as adsorbed organic contaminants, and roughness at the surface of the Ga_2O_3 layer. Note that some of these defects produce "thick" areas, while others produce "thin" areas.

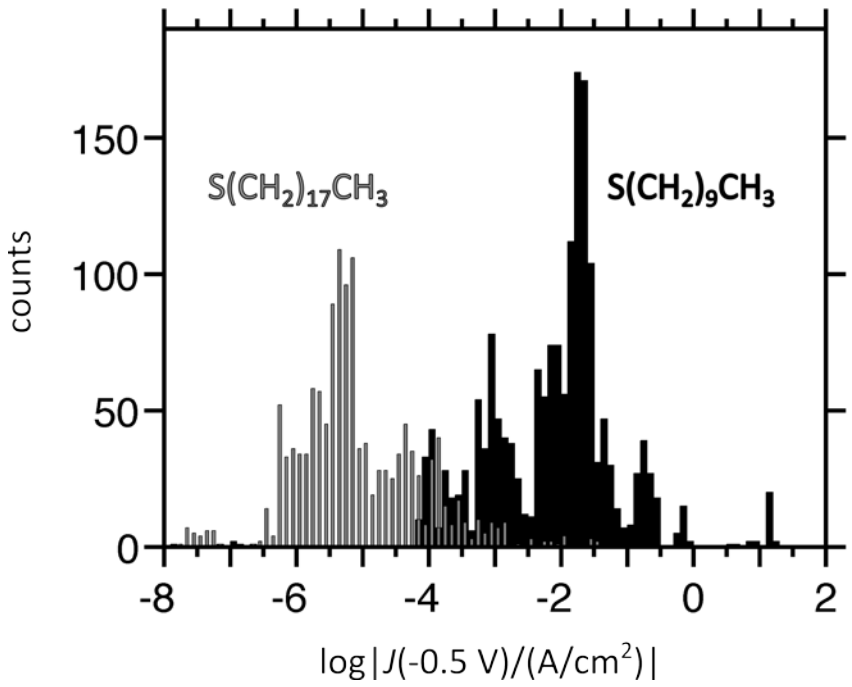


Figure 2:

Histograms of $\log|J/(A/cm^2)|$, at V = -0.5 V, for SAMs of two n-alkanethiols: $S(CH_2)_{17}CH_3$ (gray bars), and $S(CH_2)_9CH_3$ (black bars). These histograms overlap to a significant degree, despite being on opposite ends of the series of n-alkanethiols. In other words, the dispersion (spread) in the data for these two SAMs (which are representative of other n-alkanethiols), is similar to the effect of changing n from 10 to 18.

lengths of these two alkanethiols differ by almost a factor of two, the overlap between the data generated by these two SAMs is surprising—and there are seven more compounds that lie between these two. This overlap is not as severe when the Ga_2O_3/EGaIn used to contact the SAM is stabilized in a microfluidic channel,[xv] or when a single, experienced user (rather than a group of users with different levels of experience) collects the data. Even under these favorable circumstances, however, the spread in the data is still significant. One question that this paper seeks to address is how to how to draw confident conclusions about molecular effects when i) the spread of the data is comparable to the magnitude of the effect being investigated, and ii) the noise in the data make it difficult to separate real results from artifacts.

Foundational Assumptions of Statistical Analysis of Charge Transport through SAMs. Statistical analysis generally (and our analysis in particular) seeks to describe populations—i.e., groups of items that are all related by a certain characteristic.[xvi,xvii,xviii] An example of a population that we study is the set of all possible Ag^{TS}-$S(CH_2)_{12}CH_3$//Ga_2O_3/EGaIn junctions that could be prepared according to our standard procedure.[xiv] Obviously, such a population is immeasurably large, and, as is typical in statistical analysis, it is impossible to measure the entire population directly. We must, therefore, measure a representative sample of the population, and then use statistical analysis to draw conclusions, from the sample, about the general population. Hence, for example, we measure current density, J (at a particular bias, V) for a certain sample of Ag^{TS}-$S(CH_2)_{12}CH_3$//Ga_2O_3 junctions, and make generalizations about J for the population of all Ag^{TS}-$S(CH_2)_{12}CH_3$//Ga_2O_3/EGaIn junctions.

In order to draw conclusions, from a random sample, about the population from which it is derived, it is necessary to have a statistical model that identifies the meaningful parameters of the population, and describes how the observations in a sample can be used to estimate those parameters. We currently use a statistical model to describe how values of *J* arise from a population of Ag^{TS}-SR//Ga_2O_3/EGaIn junctions. Our model is a statistical extension of the approximate but widely used Simmons model[xix] (eq. **1**), which describes tunneling through an insulator, at a constant applied bias; the issues raised in the analysis would, however, apply as well to other models.

$$J = J_0 e^{-\beta d} \qquad (\mathbf{1})$$

In eq. **1**, *d* is the molecular length (either in Å, or number of carbon atoms), J_0 is a (bias-dependent) pre-exponential factor that accounts for the interfaces between the SAM and the electrodes, and β is the tunneling decay constant. The Simmons model predicts only individual values of *J* through a junction of known, and constant, thickness. It is not, therefore, a *statistical* model—one that predicts the properties of a random sample comprising measurements of many junctions.

To develop a statistical model, we began with the assumption that the junctions we fabricate fall into two categories: i) junctions in which, despite the presence of defects, the basic Ag^{TS}-SR//Ga_2O_3/EGaIn structure dominates charge transport, in keeping with the Simmons model, and ii) junctions in which experimental artifacts alter the basic structure of the junction to something radically different from Ag^{TS}-SR//Ga_2O_3/EGaIn—e.g. penetration of the SAM by the Ga_2O_3/EGaIn electrode yields a junction of the form Ag^{TS}//Ga_2O_3/EGaIn—and invalidate the Simmons model as a description of charge transport. The first type of junctions give data that are "informative" about charge

transport, while the second type give data that are difficult to interpret, within the framework of the Simmons model, and thus "non-informative". The goal of our statistical analysis, therefore, is to characterize data that are informative, and ignore data that are non-informative, by using some method to discriminate between the two.

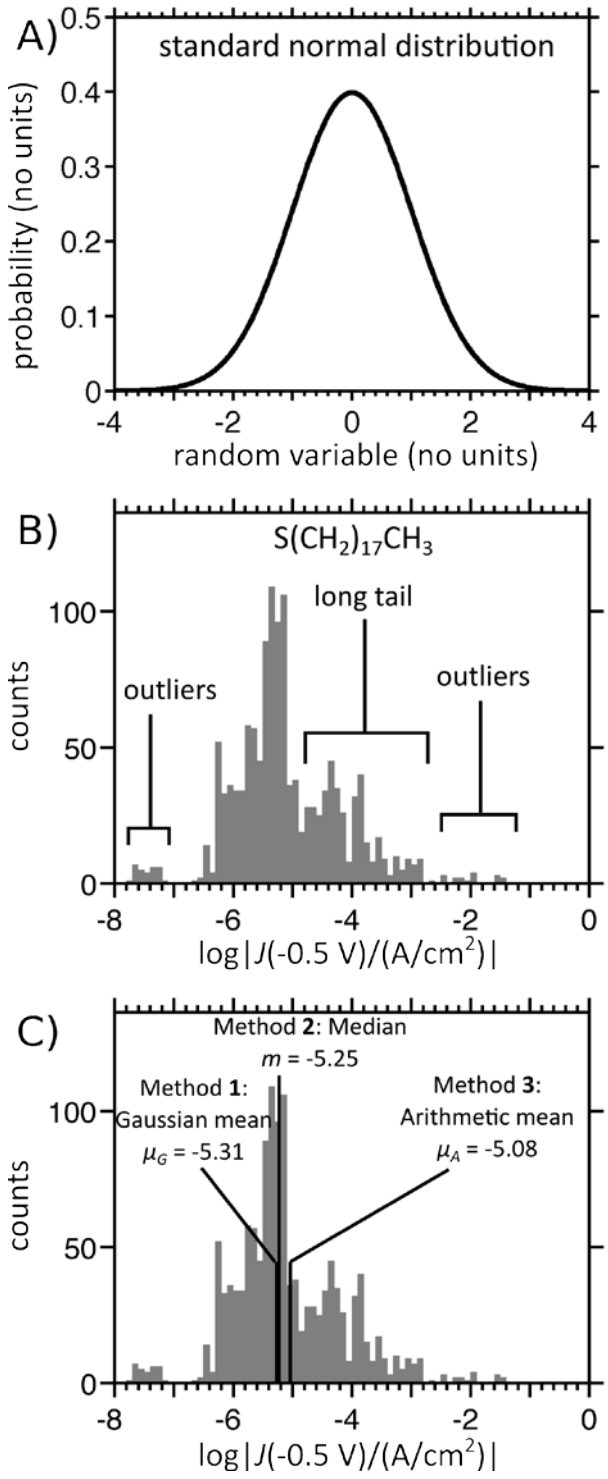
There are two major ways to draw a distinction between informative and non-informative data: i) construct a parametric statistical model[xvi,xvii,xviii] that assumes that informative data follow a certain probability distribution, while non-informative data follow a different distribution, or ii) assume that the majority of the data are informative, and choose a methodology that is insensitive to relatively small numbers of extreme data (that is, a "robust" method[xx,xxi]), since these data are likely to be non-informative. In this paper, we discuss techniques that follow each of these approaches, and argue that they are superior to other, more common techniques (which we also discuss) that do not distinguish between informative and non-informative data.

Introduction to Our Parametric Statistical Model for Measurements of Charge Transport. In constructing our parametric statistical model, we used the Simmons model[xix] as a starting point. We assumed that β and J_0 are constants, and that the actual values of d in *informative* Ag^{TS}-SR//Ga_2O_3/EGaIn junctions vary according to a normal distribution (Figure 3A; see Supporting Information and ref. *xvi* for a discussion of statistical distributions), as a result of non-catastrophic defects[xxii] in the Ag^{TS} substrate, the SAM, and the Ga_2O_3/EGaIn electrode.[xxiii] When the Simmons model holds, J depends exponentially on d, so a normal distribution of d would translate to a log-normal distribution of J (i.e. a normal distribution of $\log|J/(A/cm^2)|$; hereafter, $\log|J|$ for

Figure 3:

Deviations of $\log|J|$ from normality, and their effects on Methods 1 – 3. A) The standard normal distribution, with a mean of 0 and a standard deviation of 1 (these quantities are unitless). B) The first of two identical histograms of $\log|J(\text{-}0.5\ \text{V})/(\text{A/cm}^2)|$ for $S(CH_2)_{17}CH_3$. This plot shows two primary deviations of $\log|J|$ from normality: i) a long tail (i.e. a larger share of the sample to the right of the peak than in a normal distribution), and ii) outliers (data that lie far from the peak). C) Methods 1 – 3 respond differently to these deviations of $\log|J|$ from normality, as shown by their estimates for the location of the sample. Methods 1 responds the least to the long tail and outliers on the right, Method 2 responds moderately to them, and Method 3 responds strongly to them.

Figure 3 (Continued)



convenience). In other words, our model predicts that informative measurements of log|*J*| are normally distributed. Based on this assumption, our statistical model predicts that log|*J*| for any population of Ag^{TS}-SR//Ga_2O_3/EGaIn junctions will have two components: i) a normally distributed component that is informative, and ii) a component comprising non-informative values of log|*J*| that follow an unknown and unspecified distribution. Aside from having some *a priori* physical justification, these two components predicted by our statistical model are observed in experimental results (an example—log|*J*| for $S(CH_2)_{17}CH_3$—is shown in Figure 3; these data have been published previously[xiv]). In all cases, a prominent, approximately Gaussian peak[xvi] is easily identifiable, but anomalies (Figure 3B) are also present: i) long tails (portions of data that extend beyond the Gaussian peak, to the left or right, and cause the peak to be asymmetric) and ii) outliers (individual data, or clusters of data, that are separated from the main peak of the histogram by regions of "white space").[xx,xxiv] The difference between these two categories is somewhat subjective and arbitrary, and we present them only as guides to aid the reader in visualizing the pathologies of distributions of log|*J*|. None of the methods of analysis described in this paper require distinguishing between long tails and outliers; we, therefore, refer to them collectively as "deviations of log|*J*| from normality".

A key implication of our statistical model is that the normally distributed component of log|*J*| is the only component that gives meaningful information about the SAM. According to the model, deviations of log|*J*| from normality arise from processes that dramatically alter the typical structure of the junction, and may mislead a naive analysis that treats these deviations as informative. If the model is correct, the analysis of log|*J*| should, therefore, be designed to ignore any deviations of log|*J*| from normality.

We believe that this model offers a reasonably accurate description of $\log|J|$ (we offer further justification for our model in the Experimental Design section), but we recognize that our model could be wrong in an important way. Specifically, if the component of $\log|J|$ arising from the typical behavior of the junction follows something other than a normal distribution (i.e. if d is not normally distributed, or if β or J_0 vary significantly between junctions), then, by definition, even informative measurements of $\log|J|$ should deviate from normality, probably to a small extent.

If our statistical model is wrong in this way, then ignoring deviations of $\log|J|$ from normality will lead to similarly small, but possibly significant errors. On the other hand, an approach that incorrectly treats all data as informative will be prone to large errors from the influence of extreme data. Between these two approaches would lie methods that neither assume that $\log|J|$ is normal, nor respond strongly to extreme values. Each of the methods of statistical analysis discussed in this paper (Figure 4) fall into one of these three categories, according to how strongly they respond to deviations of $\log|J|$ from normality. The relative accuracy of these different methods of analysis will depend on whether our statistical model is eventually confirmed or discredited.

Also, the precision (but not the accuracy) of all of the methods of analysis described in this paper depends on the assumption that our measurements of $\log|J|$ are independent and uncorrelated to one another. We are relatively certain that this assumption is wrong (for instance, values of $\log|J|$ measured within the same junction correlate more with one another than values of $\log|J|$ measured from two different junctions), and we discuss, in the Results and Discussion section, a procedure to correct for violations of this assumption. If this correction is insufficient, then all methods of analysis will overstate

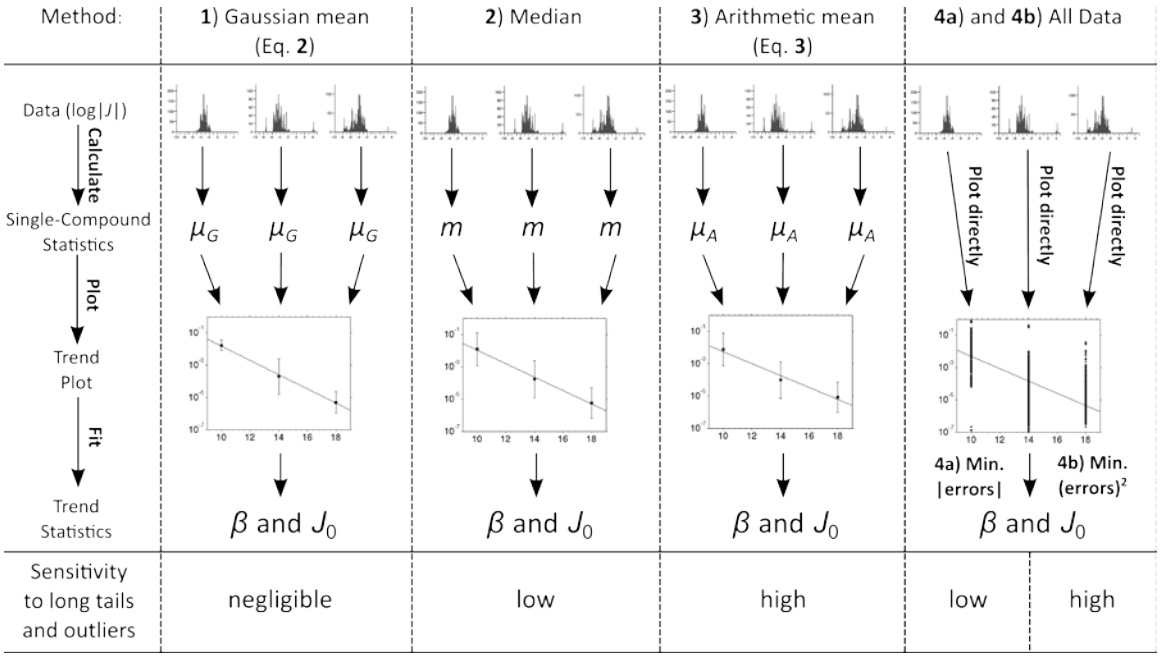


Figure 4:

Schematic of the four methods of analyzing charge transport discussed in this paper. Methods 1 – 3 use the data (samples of log|*J*|) to calculate single-compound statistics; plotting those statistics, and fitting the plots, yields trend statistics. For Method 1, μ_G is the Gaussian mean; for Method 2, *m* is the median; and for Method 3, μ_A is the arithmetic mean. Methods 4a proceed directly to plotting and fitting the raw data to determine trend statistics. The bottom row gives the sensitivity of each method to common deviations of log|*J*| from normality (long tails and outliers).

the statistical confidence of their conclusions, and underestimate the widths of confidence intervals (see Results and Discussion). Further research is needed to determine to what extent measurements of log|*J*| are correlated, and how strong of a correction must be applied to account for this correlation.

Comparison of Methods for Analyzing Charge Transport. We are interested in two categories of statistical results: i) *single-compound statistics*, which summarize measurements of charge transport through a single type of SAM (i.e. a particular compound), and enable comparisons between two SAMs, and ii) *trend statistics*, which lead to conclusions about the dependence of charge transport on some parameter (such as molecular length) that varies across a series of SAMs. In this paper, we describe five methods of analyzing charge transport. Figure 4 schematically shows that Methods 1 – 3 begin by calculating single-compound statistics, and then use those results to determine trend statistics, while Methods 4a and 4b do not produce single-compound statistics, and instead proceed directly from the data (samples of log|*J*|) to calculation of trend statistics. It is not necessary to choose only one method of analysis for both single-molecule and trend statistics; in fact, we shall show that, in some cases, it is best to use one method to estimate single-compound statistics, and another to estimate trend statistics.

Single-Compound Statistics. Methods 1 – 3 generate single-compound statistics that describe samples of log|*J*| for SAMs of a particular compound. Each method has a procedure for calculating i) the *location* (sometimes called the center, or central tendency) of the sample, ii) the *dispersion* (sometimes called the scale, or spread) of the sample, and iii) a *confidence interval* that surrounds the location and has a width related to the dispersion and the number of data in the sample.[xvi,xvii,xviii,xx] To estimate the

location and dispersion, respectively, Method 1 uses the mean (μ_G) and standard deviation (σ_G) determined by fitting a Gaussian function to the sample of log|J|, Method 2 uses the median (m) and adjusted median absolute deviation (σ_M) or interquartile range of the sample, and Method 3 uses the arithmetic mean (μ_A) and standard deviation (σ_A) of the sample. (These quantities, and the procedures used by each method for calculating confidence intervals, are detailed in the Results and Discussion section.)

Each method makes different assumptions about the distribution of log|J|. Method 1 employs an algorithm that essentially "selects" the most prominent peak in the sample of log|J|, and disregards the rest. In other words, Method 1 closely follows the statistical model described above, in that it assumes that deviations of log|J| from normality (i.e. long tails or outliers; see Figure 3B) are not informative about charge transport through the SAM, and ignores them. The appropriateness of Method 1 for statistical analysis depends, therefore, on the correctness of this assumption. Methods 2 and 3 both depart from the statistical model by taking into account, to different degrees, deviations of log|J| from normality. Method 2 (m and σ_M) responds only moderately to such deviations, while even a few extreme data can have a significant effect on Method 3 (μ_A and σ_A).

Figure 3C briefly demonstrates the different responses to deviations of log|J| from normality of the locations estimated by these three methods. The histogram of $S(CH_2)_{17}CH_3$ exhibits a long tail to the right (towards high values of log|J|). This tail strongly influences the arithmetic mean (Method 3) and "pulls" it to the right by 0.23 log-units, in comparison with the Gaussian mean (Method 1). By comparison, the median (Method 2) is only moderately affected by the tail, and differs from the Gaussian mean by 0.06 log-units. Although even the divergence between Methods 1 and 3 may not appear

significant, there are two reasons to take it seriously. i) For the two n-alkanethiols on the extreme ends of the series, $S(CH_2)_8CH_3$ and $S(CH_2)_{17}CH_3$, the locations of the distributions of $\log|J|$ only differ by about 3.0 – 3.5 log-units (depending on the method used to estimate the locations). The divergence, evident in Figure 3, between Method 1 and Method 3, therefore, represents approximately 8% of the total change in $\log|J|$ across the entire series of n-alkanethiols. When comparing two adjacent n-alkanethiols (e.g. n = 17 and 18), the differences between Methods 1 – 3 will be even more significant. ii) The data discussed in this paper—in particular, n-alkanethiols for which n is even—were collected by experienced users, and probably exhibit fewer deviations from normality than would data collected by inexperienced users. For inexperienced users, then, the effect of outliers and tails on Method 3 may be large, and using Method 1 or 2 is important to draw accurate conclusions from the data.

The goal of single-compound statistics is to estimate the location (and, secondarily, the dispersion) of the population in a way that is both *precise* and *accurate*. The precision of the estimated location is determined by the width of the confidence interval, such that a narrow confidence interval indicates a precise (although not necessarily accurate) estimate.[xvi] The accuracy of the location estimated by a given method depends on how well the assumptions of the method conform to reality. If, through its assumptions, a method correctly discriminates between informative and non-informative data, then its estimate for the location will generally be accurate, in the sense that the true location of the $\log|J|$ for the population will, with a stated confidence (e.g. 99%), lie within the confidence interval.[xvii] If the method makes incorrect assumptions about the data, then the confidence interval cannot be trusted to contain, at the stated level of

confidence, the true location of the population. Because we believe that our statistical model comprises reasonably correct assumptions about the data, we recommend the use of either Method 1 (Gaussian mean) or 2 (median), but not Method 3 (arithmetic mean), for estimating single-compound statistics.

Trend Statistics. When calculating trend statistics, such as β and J_0 (see eq. **1**), it is possible to use any of the four methods discussed in this paper. In general, the process of calculating trend statistics involves plotting log|*J*|, measured across a series of compounds, against some molecular characteristic that varies across the series (e.g., molecular length, n), and then fitting the plot to a model that specifies the relationship between the desired trend statistics and the data. The plotted values of log|*J*| are either single-compound statistics summarizing log|*J*| for each molecule (Methods 1 – 3), or the raw data themselves; that is, all measured values of log|*J*| (Methods 4a and 4b).

All methods use an algorithm to fit (values summarizing) log|*J*| vs. n to a linear model (the Simmons model predicts a linear relationship between log|*J*| and n, *via* the parameter *d*). The choice of the fitting algorithm determines the influence exerted on the outcome by data that lie far from the fitted line (these data are roughly equivalent to those that cause log|*J*| to deviate from normality—i.e. long tails and outliers in each sample). For Methods 1 – 3, the choice of the fitting algorithm has little effect on the outcome, because, in the process of estimating single-compound statistics, each Method has already made the assumptions that determine how it responds to extreme data. By summarizing log|*J*| for each compound and passing those summaries to the fitting algorithm that determines trend statistics, these Methods have compressed the wealth of information in each sample, and essentially ensured that the fitting algorithm will not

"see" any extreme data. In calculating trend statistics, therefore, Methods 1 – 3 carry forward all of the respective assumptions and biases that they exercise in the calculation of single-compound statistics.

For Methods 4a and 4b, by contrast, the choice of the fitting algorithm has a (potentially) large impact on the results, because deviations from normality are invariably present in the raw data. Method 4a uses an algorithm that minimizes the sum of the *absolute values* of the errors between the data and the fitted line (a "least-absolute-errors algorithm"), while Method 4b employs an algorithm that minimizes the sum of the *squares* of those errors (a "least-squares algorithm"). Method 4a responds to deviations of $\log|J|$ from normality in a manner analogous to that of Method 2; both methods are only moderately affected by such deviations. Method 4b, on the other hand, is analogous to Method 3, in that it responds strongly to deviations from normality.

Although Methods 4a and 4b cannot give single-compound statistics, they have the advantage of offering far greater precision than Methods 1 – 3 in estimating trend statistics. Because Method 4a has assumptions similar to Method 2, the accuracies of the two methods will also be similar (by the same token, the accuracy of Method 4b will be similar to that of Method 3). We, therefore, recommend using either Method 1 (fitting Gaussian means of $\log|J|$ with a least-squares algorithm) or Method 4a (fitting all values of $\log|J|$ with a least-absolute-errors algorithm) to estimate trend statistics. If estimates produced by these two methods agree, then it may be preferable to use Method 4a, because of its high precision.

Figure 4 schematically depicts how each of the four methods progress from histograms of $\log|J|$ for single compounds (three compounds are shown, but each method

can involve an arbitrary number) to analysis of the trend across those compounds; the figure also indicates the type of information given by each of the four methods. All four methods assume respond differently to deviations of $\log|J|$ from normality, and, as we shall demonstrate, the choice of method can affect the conclusions about charge transport drawn by the analysis.

Background

Methods for Measuring Charge Transport. Many approaches exist for measuring charge transport through self-assembled monolayers (SAMs) of thiol-terminated molecules. These approaches can be segregated into those that produce small-area (from single-molecule to ~ 100 nm^2) junctions comprising relatively few molecules (scanning probe techniques[ii,iii,iv,v,vi,vii,viii] and break junctions[ix]) and those that produce large-area (> 1 μm^2) junctions.

Most large-area junctions employ a SAM, supported on a conductive substrate, and contacted by a top-electrode – either a layer of evaporated Au, a drop of liquid Hg supporting a SAM (Hg-SAM), or a structure of Ga_2O_3/EGaIn. While it was common, in the past, to evaporate Au directly onto the SAM,[xxv] this procedure resulted in low yields (up to 5% when executed very carefully, but usually < 1 %)[x] of non-shorting junctions, and is now known to damage the SAM.[xxvi] Most currently successful large-area junctions employ a top-electrode with an insulating or semiconducting barrier (a "protective layer") between the metal and the SAM, to protect against damage from high-energy metal atoms (during evaporation) and guard against metal filaments formed by the electromigration of metal atoms through defects in the SAM. Examples of protective layers between the

SAM and the metal of the top-electrode include conducting polymers (e.g. Au-SAM//PEDOT:PSS/Au junctions of Akkerman *et al.*[xxvii] and Au-SAM//(polymer)/Hg drop junctions of Rampi *et al.*[xxviii]), a second SAM (e.g. Ag-SAM//SAM-Hg junctions of us,[xxii,xxix] and others[xxx,xxxi,xxxii]), and a layer of metal oxides (e.g. our Ag^{TS}-SAM//Ga_2O_3/EGaIn junctions[xii,xiv]).

There are two exceptions to the rule of the protective layer in large-area junctions. i) Cahen *et al.*[xxxiii,xxxiv,xxxv] use n-Si-R//Hg and p-Si-R//Hg junctions, in which a layer of alkenes, covalently attached to a doped and hydrogen-passivated Si surface, is directly contacted by a drop of Hg. Use of a semiconducting, rather than a metallic, substrate reduces the migration of metal atoms responsible for metal filaments and shorts. ii) Lee *et al.*[x] continue to evaporate Au directly on the SAM to form Au-SAM//Au junctions. Skilled users can generate yields of 1 – 5 %, and the authors use careful statistical analysis to distinguish between real data and artifacts resulting from SAMs damaged by the direct evaporation of Au.

Ag^{TS}-SAM//Ga_2O_3/EGaIn Junctions. In our nomenclature, Ag^{TS} denotes an ultra-flat Ag substrate produced by template stripping,[xi] while Ga_2O_3/EGaIn denotes the eutectic alloy of gallium and indium (75% Ga, 25% In by weight, m.p. = 15.5 °C) with its surface layer of oxide.[xxxvi] The rheological properties of this composite material make it possible to mold it into conical shapes, but still allow it to deform under applied pressure. These properties make Ga_2O_3/EGaIn an excellent material for forming soft, microscale contacts to structures like SAMs.

The question – does Ga_2O_3/EGaIn form good *electrical* contact to SAMs? – hinges on the resistivity of the Ga_2O_3 surface film. We have measured[xxxvii] the thickness (~ 0.7 nm

is the average thickness, although some regions may be several μm thick), composition (primarily Ga_2O_3), and resistivity ($10^5 - 10^6$ Ω•cm) of the film. The film of Ga_2O_3, like all surfaces in the laboratory, supports a layer of adsorbed organic material, which is undoubtedly present in Ag^{TS}-SAM//Ga_2O_3/EGaIn junctions (and in many other junctions). The measured thickness of this layer is ~ 1 nm,[xxxvi,xxxvii] and its composition probably depends on the environment, but it is probably a discontinuous layer, rather than a continuous sheet.[xxxviii] We measured the resistivity of the Ga_2O_3 film, with its adsorbed organic layer, using two direct methods[xiii] (contacting structures of Ga_2O_3/EGaIn with Cu and ITO electrodes) and one indirect method (placing an upper bound on the resistivity of the Ga_2O_3 using the value of J_0, explained below, for n-alkanethiols). All three methods converge to a range of $10^5 - 10^6$ Ω•cm for the resistivity of the Ga_2O_3 film. This range is at least three orders of magnitude lower than the resistivity of a SAM of $S(CH_2)_9CH_3$ (~ 10^9 Ω•cm), the least resistive SAM that we have measured.[xxxix] The Ga_2O_3 film, and especially the layer of adsorbed organic matter on its surface, are certainly the least understood components of our system. Based on these measurements, however, we conclude that the Ga_2O_3 film, with its adsorbed organic layer, is sufficiently conductive that it does not affect the electrical characteristics of the junction. This conclusion is supported by our recent study[xl] that uses molecular rectification in various SAMs to show that the SAM, rather than the Ga_2O_3 layer, dominates charge transport through the junction.

Tunneling through SAMs. As stated in the introduction to our statistical model, tunneling through SAMs is widely assumed to follow eq. **1**.[xix] Typically, one of the first experiments performed with any experimental system is to measure log|*J*| through SAMs

of a series of n-alkanethiols of increasing molecular lengths (d), and to calculate the values of β and J_0. The tunneling decay constant, β, is related to the height and shape of the tunneling barrier posed by the SAM. Because the value of β theoretically depends only on the molecular orbitals of the SAM, and not on the interfaces between the SAM and the electrodes, β is expected to be largely independent of the method used to measure $\log|J|$, and is, thus, a useful standard with which to validate new techniques. By contrast, J_0 is a pre-exponential factor that accounts for factors that contribute to "contact resistance" – the resistivity and density of states of the electrodes, and any tunneling barriers at the interfaces between the SAM and the electrodes. While it is rare to find a value of J_0 in the literature, this parameter also contains important information about the electrodes and interfaces in a junction—information that is complementary to that conveyed by β. Thus, J_0 could be used to compare different techniques for measuring charge transport through SAMs.

The Simmons model contains many assumptions (it is, after all, a simplification of a model originally designed to describe tunneling through a uniform insulator with extended conduction and valence bands), the most significant of which is that tunneling is the only operative mechanism of charge transport through the SAM.[xix] Another significant assumption is that the complicated tunneling barrier posed by a particular class of SAMs can be described by a simplified "effective" barrier of a certain height, and that this height does not vary with the length of the molecules in the SAM (i.e. that β does not depend on d). Despite these assumptions, eq. **1** shows reasonable agreement with results for n-alkanethiols in the approximate range of n = 8 – 20, with values of β of

$0.8 – 0.9$ Å^{-1} $(1.0 – 1.1\ n_C^{-1})$,[xli] and it is now standard practice to report the value of β for n-alkanethiols as one (often the primary, or only) parameter of interest.

Defects in SAM-Based Junctions Necessitate Reporting of All Data. Measurements of charge transport through SAMs have often neglected the contribution of (probably) unavoidable variations in the system to the dispersion of data. Because J depends exponentially on the thickness of the SAM, J can be extremely sensitive to defects in a tunneling junction, especially those that decrease the local distance between electrodes (so-called "thin-area" defects).[xxii] For example, assuming a value of 0.8 Å^{-1} for β, a defect 5 Å thinner than the nominal thickness of the SAM would carry a current density more than 50 times that of a corresponding area on a defect-free SAM. If such "thin" defects comprised just 2% of the total area of the junction, then the same amount of current would pass through those thin-area defects as through the rest of the (defect-free) SAM. As a result of the exponential dependence of $J(V)$ on d and β, thin-area defects can easily dominate charge transport through the junction. By contrast, thick-area defects (those that increase the local separation between electrodes) can usually be ignored, because their contribution to $J(V)$ is small relative to other sources of error. Many types of defects are common in both the Ag^{TS} substrate (e.g. grain boundaries, vacancy islands, and step edges) and the SAM (e.g. domain boundaries, pinholes, disordered regions, and physisorbed contaminants).[xxii]

These defects presumably give rise to the large spread observed in distributions of $\log|J|$. The roughness of the metal substrate (whether Au or Ag) is one of the most significant factors in determining the density of defects in SAM-based junctions. In a previous paper,[xi] we showed that using template-stripped substrates results in smoother

surfaces (for Ag^{TS}, rms roughness = 1.2 nm over a 25 μm^2 area) than using surfaces as-deposited by electron-beam evaporation (for as-deposited Ag, rms roughness = 5.1 nm over a 25 μm^2 area). In Ag-SAM//SAM-Hg junctions, this decrease in roughness between as-deposited and template-stripped Ag decreased the range of measured values of J by several orders of magnitude, and increased the yield of working junctions by more than a factor of three.[xxii]

Experimental Design

***"Informative" vs. "Non-informative" Measurements of* log|*J*|.** The Simmons model[xix] (eq. **1**) of tunneling predicts individual values of J for known values of d, but does not describe actual measurements of J (or $\log|J|$). Actual measurements comprise random samples of many junctions, across which the parameters of the Simmons model (most likely d, but possibly β and J_0) certainly vary. Because the Simmons model does not specify how real data arise from random sampling of charge transport (i.e. it is not a *statistical* model), our statistical analysis must account for what the Simmons model ignores, in order to derive meaningful results from measurements of $\log|J|$.

We begin by recognizing that, in our junctions (and SAM-based junctions in general), there are two classes of defects: i) defects that preserve the basic Ag^{TS}-SR//Ga_2O_3/EGaIn structure of the junction, even while changing d, and ii) defects (perhaps better termed "artifacts") that disrupt the basic structure of the junction. Defects belonging to the first class might include, for example, domain boundaries, pinholes, disordered regions, and physisorbed contaminants on the SAM. Even though this class of defects may alter d in a junction, the Simmons model remains a valid description of charge transport through the

junction, because charge must still tunnel between the Ag^{TS} and Ga_2O_3/EGaIn electrodes, through the SAM. In the second class of defects belong artifacts, such as areas in which the Ga_2O_3/EGaIn electrode penetrates or intrudes into the SAM, regions of bare EGaIn (lacking a Ga_2O_3 film) in contact with the SAM, and metal filaments that bridge the two electrodes and bypass the SAM. These artifacts not only change *d* in the junction, but they also change (at least) J_0, and might alter the mechanism of charge transport between electrodes to some process other than tunneling. In short, these types of artifacts invalidate the Simmons model (with its assumption of a constant J_0 for Ag^{TS}-SR//Ga_2O_3/EGaIn junctions with a constant R group) as a description of charge transport through the junction.

There is nothing particularly controversial in partitioning defects into those that preserve the integrity of the Simmons model, and those that destroy it, nor in stating that the former result in measurements that are "informative" about charge transport through the SAM, while the latter result in measurements that are "non-informative". The question is how to account fully for the informative data, while minimizing the effect of non-informative data on the analysis. There are, broadly, two ways to approach this question: i) to use a parametric (or semi-parametric) statistical model[xvi,xvii] that differentiates between the distributions of informative and non-informative data in order to identify the former and discard the latter, and ii) to assume that informative data predominate over non-informative data and use an analysis that responds strongly to the bulk of the data (no matter how the data are distributed) and weakly to a small number of extreme values.[xx,xxi] As we will show, Method 1 follows the first approach, while Method 2 (and Method 4a) follows the second approach. (Methods 3 and 4b follow

another, inferior approach that does not discriminate between informative and non-informative data; we include them in this paper in order to illustrate their deficiencies, because they are, unfortunately, the methods most commonly used outside of statistical disciplines.)

Method 1 uses a statistical model to distinguish between informative and non-informative data. Method 1 depends on a statistical model that predicts that when $\log|J|$ arises from informative measurements, those data are normally distributed. The statistical model is based on the following reasoning: when the Simmons model holds, the statistical distribution of $\log|J|$ is determined by the distributions of J_0, β, and d—if one knows the distributions of these three parameters, one can predict the distribution of $\log|J|$ that results from informative measurements. Our statistical model assumes that, in junctions that generate informative data, d is normally distributed, while J_0 and β are approximately constant (i.e. if they vary, their contributions to the dispersion of $\log|J|$ are negligible, compared to the contribution of d).[xlii] Since $\log|J|$ is proportional to d, if d is normally distributed, then $\log|J|$ is also normally distributed. Method 1 follows this statistical model by employing a fitting algorithm to "seek out" the largest component of the histogram of $\log|J|$ that conforms to a normal distribution, and "ignore" the rest. If the model is correct, then Method 1 finds the informative data in the most prominent Gaussian peak in the sample, and rejects the non-informative data that deviate from this peak.

There are two lines of reasoning that support the assumptions that $\log|J|$ is normally distributed. i) That d is normally distributed is probable, because normal distributions arise frequently in nature when a variable is influenced by many uncorrelated factors.[xviii]

If the factors that determine the density and type of defects in a junction, therefore, are many and uncorrelated with one another—a plausible scenario—then *d* will be normally distributed, as will log|*J*|. ii) Previous experiments by us[xii,xiv] (and others[x]), such as the data shown in Figure 3B, are consistent with log|*J*| being approximately normally distributed. Although histograms of log|*J*| are often noisy and slightly asymmetric, there is almost always a prominent peak resembling a Gaussian function identifiable in every histogram.

Despite some theoretical and experimental support, the assumption that log|*J*| is normally distributed might still be wrong. Other distributions, such as a Cauchy distribution (sometimes called a Lorentzian distribution) or a Student's t-distribution, may also be consistent with observed histograms of log|*J*|, and we cannot entirely rule them out, but they lack the *a priori* physical justification of a normal distribution. It is, however, plausible that while *d* is normally distributed, our procedure for measuring log|*J*| does not *randomly* sample *d*, but that there is some correlation in the values of *d* for junctions measured under similar conditions. For example, the junctions formed on a common Ag^{TS} substrate (supporting a SAM) may have values of *d* that cluster around a certain value (e.g. 10 Å), whereas the values of *d* for junctions formed on a different Ag^{TS} substrate may cluster around a higher value (e.g. 12 Å), perhaps due to differences in the amount of organic contaminants present in the environment. In such a case, the first Ag^{TS} substrate would result in one normal distribution of log|*J*|, while the second substrate would result in another, overlapping normal distribution, centered at lower values of log|*J*| than the first.

Such clustering may exist at multiple levels, such as between Ag^{TS} substrates, Ga_2O_3/EGaIn electrodes, operators, or times of year during which measurements were performed. Investigating the individual contributions of each level to clustering in log|*J*| (*via* clustering in *d*), and answering the question of whether such clustering is significant enough to cause log|*J*| to deviate noticeably from normality, would entail an in-depth experimental study that is beyond the scope of this paper. We do not know whether clustering violates the assumption that log|*J*| is normally distributed, but we raise it as a concern, in order to disclose a potential failure of our statistical model, and to motivate the development of multiple methods of statistical analysis to respond flexibly to such a contingency.

Clustering would possibly affect the accuracy of Method 1, but it would also possibly affect the precision of *all* of the methods described in this paper, as expressed by the widths of the confidence intervals around parameters estimated by each method. As explained in the Results and Discussion section, the width of a confidence interval decreases as the number of data increases; that is, many measurements lead to a narrow confidence interval. For each of Methods 1 – 4, this relationship between the width of a confidence interval and the number of data depends on the assumption that the data are independent from, and uncorrelated to, one another. If there is significant clustering of measurements of log|*J*|, then this assumption will be violated, the widths of confidence intervals will be underestimated, and the precision of results will be overstated. In the Results and Discussion section, we discuss a procedure for estimating the correlation within a sample (termed "autocorrelation") and correcting for its effect on the width of confidence intervals. We do not currently have enough information to evaluate whether

or not this procedure adequately corrects for autocorrelation, and we emphasize the need for further experiments to investigate and minimize the amount of autocorrelation in our measurements.

***Methods 2 and 4a do not assume a specific distribution for* log|*J*|*, but avoid extreme values*.** Currently, we are reasonably confident that our statistical model describes real measurements of log|*J*| with enough accuracy to be useful, and we, therefore, favor Method 1 in our analysis. In case future experiments or insights cast doubt on our statistical model, we offer Method 2 as a substitute that does not depend on our model, but does an adequate job of minimizing the effect of (probably) non-informative measurements on the analysis. Method 2 uses the median and other quantiles to characterize log|*J*|. Since well-defined quantiles exist for every continuous probability distribution, Method 2 does not require that log|*J*| be normally distributed.[xvi] The median tends to follow the bulk of the data in a sample, and it is much less influenced by extreme values than the mean.[xx] If informative measurements constitute the bulk of the data, therefore, then even extreme values resulting from non-informative measurements will have a relatively small effect on Method 2. Method 4a, as we will show below, carries the same relative insensitivity[xxi] to extreme values as Method 2. Methods 3 and 4b, by contrast, are relatively sensitive to extreme values,[xx] and are likely to allow non-informative measurements to bias the conclusions of statistical analysis.

Results and Discussion

Overview of Assembly of Ag^{TS}-$S(CH_2)_9CH_3$//Ga_2O_3/EGaIn Junctions and Measurement of Charge Transport. We have previously published a large dataset

comprising log|*J*| for n-alkanethiols (n = 9 – 18), with both odd and even numbers of carbon atoms in the backbone.[xiv] We elected to use this dataset, in order to test and demonstrate the analytical methods described in this paper, for two reasons: i) because the series of even-numbered alkanethiols (n = 10, 12, 14, 16, 18) is the standard dataset used to calibrate and compare experimental techniques for measuring charge transport through SAMs, and ii) because the series of both odd and even alkanethiols (n = 9 – 18) shows a subtle effect (the "odd-even" effect) that cannot be accurately characterized without careful statistical analysis.

In our previous publication,[xiv] we formed SAMs of $S(CH_2)_9CH_3$ (decanethiol) on template-stripped silver (Ag^{TS}) substrates, and made electrical contact to these SAMs using cone-shaped microelectrodes of the liquid eutectic of gallium and indium (75 % Ga, 25 % In by weight, with a surface of predominantly Ga_2O_3). We denote the resulting structure a "Ag^{TS}-$S(CH_2)_9CH_3$//Ga_2O_3/EGaIn junction"; detailed procedures for forming SAMs on Ag^{TS}, fabricating cone-shaped electrodes of Ga_2O_3/EGaIn, and assembling these junctions have been given elsewhere.[xiii,xiv] After forming a junction, we grounded the Ag^{TS} substrate and applied a voltage (V) to the Ga_2O_3/EGaIn electrode while measuring the current flowing between the two electrodes. We applied the voltage in steps of 50 mV, with a delay of 0.2 s between steps, starting at 0 V, increasing to +0.5 V, decreasing to -0.5, and returning to 0V; a cycle, beginning and ending at 0 V, constituted one *J*(*V*) trace. We calculated the current density (*J*) by dividing the current through the junction by its estimated contact area, which we determined by measuring the diameter of the junction, and assuming a circular contact between the Ga_2O_3 electrode and the SAM.

Excluding Shorts Prior to Analysis. Some analytical tools are especially sensitive to outliers in distributions of log|*J*|, so it is best to begin by excluding any data that are unambiguously known to be artifacts, as long as there is a simple procedure for doing so. For one type of artifact – short circuits, or simply "shorts" – there is such a procedure. We define shorts as values of current that reach the compliance limit of our electrometer (± 0.105 A); given the range of contact areas for our junctions (~ 10^2 – 10^4 μm^2), shorts translate to values of |*J*| in the range of 10^3 – 10^5 A/cm^2 (log|*J*| = 3 – 5). Shorts clearly do not give information about the SAM and can bias the distribution towards high values of log|*J*|, so when we perform operations on the raw distribution of log|*J*|, we discard all values of log|*J*| > 2.5 (i.e. |*J*| > 3.2 × 10^2 A/cm^2). We chose this threshold because it is higher than J_0 for our junctions (see below), but also lower than all shorts, which lie in the range of log|*J*| = 3 – 5 (see Figure 5).

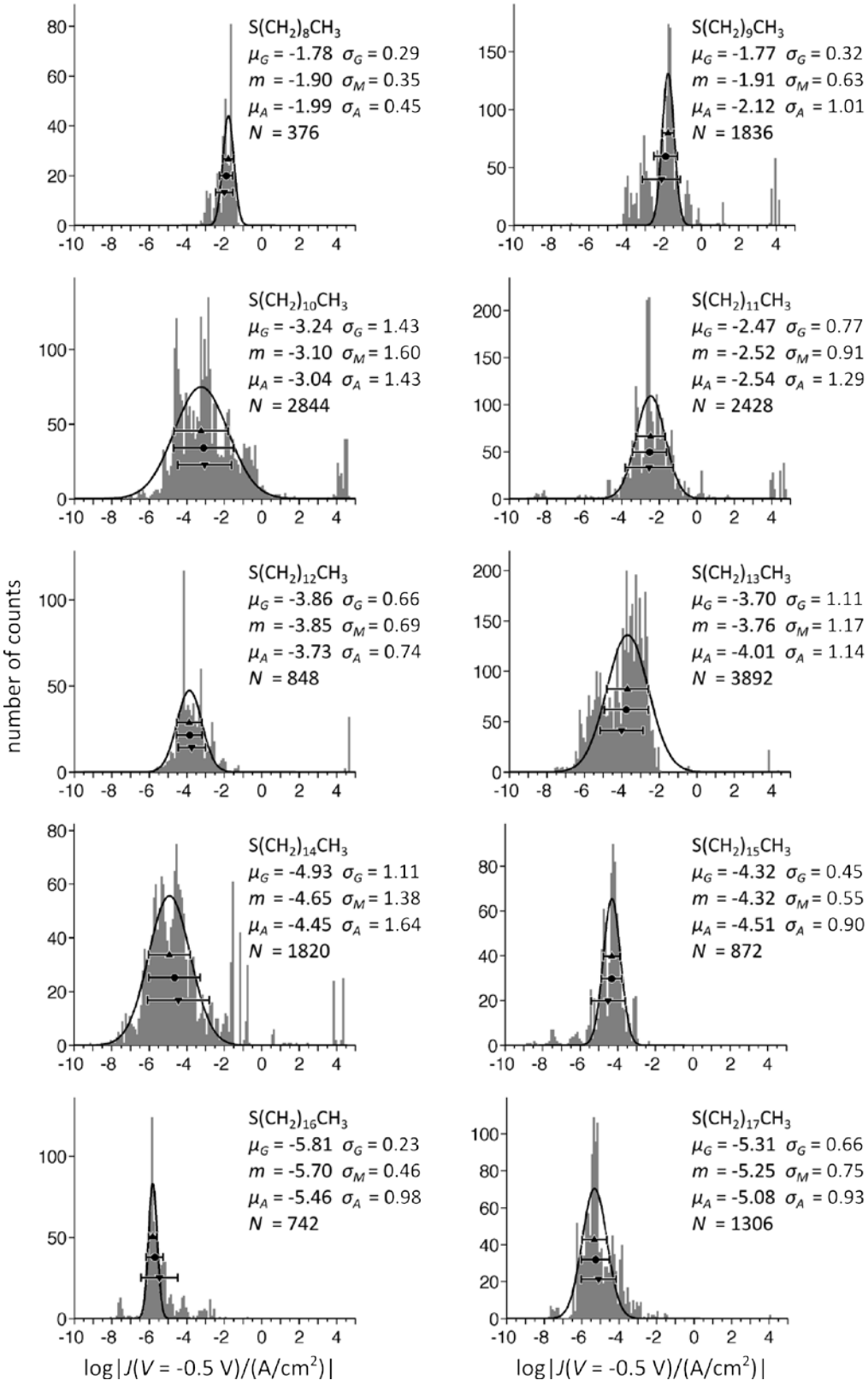
Another type of artifact that occurs in measurements of charge transport is the open circuit, but there is no reliable way to exclude open circuits, as there is for shorts. An open circuit occurs when the Ga_2O_3/EGaIn electrode fails to make contact with the SAM (the image of the junction used to judge contact can sometimes be ambiguous); charge cannot tunnel through the SAM, and the flow of current is limited to accumulating charge on the substrate and top-electrode. In such cases, the measured current is low (~ ± 10^{-12} A), and the values of |*J*| that result from these currents range from 10^{-8} – 10^{-6} A/cm^2 (log|*J*| = -8 – -6). For relatively long alkanethiols (n = 14 – 18), a significant portion of the Gaussian peak in the distribution of log|*J*| extends into this range. Unlike with shorts, therefore, there is no clear threshold that distinguishes between open circuits and real data.

***Calculating Single-Compound Statistics: the Location and Dispersion of* log|*J*|.** In the introduction, we defined the location and dispersion of a distribution.[xvi] Here, we

Figure 5:

Comparison of Methods 1 – 3 for estimating the location and dispersion of $\log|J|$, over the series of n-alkanethiols: $S(CH_2)_{n-1}CH_3$ (n = 9 – 18). The data shown in this figure have been reported previously, but have not been analyzed in this way.[xiv] Gray bars constitute histograms of $\log|J/(A/cm^2)|$ at V = -0.5 V. Black curves show the Gaussians fitted to each histogram using the fitting algorithm in Method 1. Data points (with error bars) summarize the location (and dispersion) of $\log|J|$ estimated by each method. Method 1: upward-facing triangles (and error bars) indicate the Gaussian mean, μ_G (and the Gaussian standard deviation, σ_G). Method 2: circles (and error bars) indicate the median, m (and the adjusted median absolute deviation, σ_M). Method 3: downward-facing triangles (and error bars) indicate the arithmetic mean, μ_A (and the arithmetic standard deviation, σ_A). The vertical positions of the points were chosen only for clarity, and do not convey information about the methods. Insets give the values of location and dispersion estimated by each method, as well as the size of the sample (including shorts).

Figure 5 (Continued)



define three methods for estimating the location and dispersion of distributions of log|*J*|, and discuss the results of these methods, when applied to data that we have previously reported for n-alkanethiols.[xiv]

Method 1: the Gaussian Mean and Standard Deviation. The first method involves constructing a histogram of the sample (see Figure 5, for example), and fitting a Gaussian function (eq. **2**) to the histogram (*f*(*x*) is the frequency of a particular observed value of the independent variable, *x*, and *a*, μ_G, and σ_G are fitting parameters).

$$f(x) = \frac{a}{\sigma_G \sqrt{2\pi}} e^{\frac{-(x-\mu_G)^2}{2\sigma_G{}^2}} \qquad (\mathbf{2})$$

The fitting parameters μ_G and σ_G are, theoretically, the mean and standard deviation, respectively, of the normally distributed component of log|*J*|. (we refer to these values as the "Gaussian mean", and the "Gaussian standard deviation"). To fit histograms to Gaussian functions, we used an algorithm (MATLAB 7.10.0.499, see Supporting Information for a detailed description) that minimizes the sum of the squares of the differences between the Gaussian function and the histogram – a "least-squares" algorithm.

The accuracy of the Gaussian mean and standard deviation depend heavily on the correctness of the statistical model described in the Experimental Design section—i.e. whether all informative measurements of log|*J*| are randomly sampled from a normal distribution. Figure 5 shows histograms of all n-alkanethiols for n = 9 – 18, with black curves showing the Gaussian functions fitted to the histogram under Method 1. As expected, we found that Method 1 was insensitive to those deviations of log|*J*| from normality that could be classified as long tails and outliers (see Figure 3B and the

Introduction for explanations of these terms). For example, μ_G and σ_G did not change when an exclusion rule was used to eliminate shorts (which are an extreme class of outlier). The fitting algorithm finds the global minimum of the squared error between the data and the Gaussian function, but shorts (and other outliers) only create (or affect) local minima in the squared error; in the vast majority of cases, therefore, shorts have a negligible effect on the location of the global minimum, and do not need to be excluded prior to using Method 1.

While tails and outliers were, as predicted by our statistical model, the predominant deviations of $\log|J|$ from normality in most histograms, there were two histograms that included qualitatively different types of anomalies: those of $S(CH_2)_9CH_3$ and $S(CH_2)_{13}CH_3$. For $S(CH_2)_9CH_3$, the normal component of the histogram of $\log|J|$ appeared to contain a "gap" in the data at approximately $\log|J| = -2.5$. The algorithm that fit the Gaussian function to the histogram disregarded the data to the left (towards low values of $\log|J|$) of this gap as non-informative, but it is unclear whether this "choice" was correct, since the disregarded region contained many data. The histogram of $S(CH_2)_9CH_3$ may represent a failure of our statistical model, but it is difficult to be certain.

The histogram of $S(CH_2)_{13}CH_3$ seemed to contain not one, but two, major Gaussian peaks in close proximity. The second apparent peak was more prominent than a simple tail, and the fitting algorithm used by Method 1 could not entirely ignore it. In this case, Method 1 seemed to consider both peaks as informative data. Again, it is unclear whether this "choice" was correct, but from these two cases, it is evident that a possible

weakness of Method 1 is its ambivalence in how it responds to deviations of log|*J*| from normality that cannot be classified as either long tails or outliers.

For any fitted function, it is possible to calculate R^2, the coefficient of determination (most fitting algorithms will give R^2 as one of the outputs).[xvi] While this parameter is not very useful in evaluating the "goodness" of a particular fit to a sample of data, it does convey some useful information. The value of R^2 can be interpreted as the fraction of the data that are explained by the fitted function, as opposed to the remainder of the data, which constitute random errors not explained by the function.[xvi]

If our statistical model is correct in stating that all deviations of log|*J*| from normality are non-informative, then, in our case, R^2 approximately represents the fraction of data in the sample that are informative, and $(1 - R^2)$ gives the fraction of data that are non-informative. The values of R^2 for the Gaussian fits to the n-alkanethiols ranged from a low of 0.64 (for n = 10), to a high of 0.82 (n = 16). Values of R^2 for all Gaussian fits are given in the Supporting Information. According to our statistical model, therefore, approximately 64 – 82% of the measurements of log|*J*| shown in Figure 5 are informative, while the remaining 18 – 36% (a significant fraction of the each sample of log|*J*|) are non-informative. If this interpretation is correct, then it leads to two interesting, but tentative, conclusions: i) a significant fraction of our junctions fail in ways that disrupt the basic Ag^{TS}-SR//Ga_2O_3/EGaIn structure of the junction, yet do not cause shorts, and ii) despite these failures, careful statistical analysis can reconstruct the actual characteristics of charge transport through the junctions.

Method 2: Median, Box and Whisker Plots, and Estimates of Scale. The second method for estimating the location of a sample of log|*J*| uses the median (*m*). The median

is defined[xvi,xvii,xviii] as the value for which 50% of the sample is greater than or equal to that value, and 50% of the sample is less than or equal to that value (i.e. the median is the 50th percentile of the sample).[xliii]

Method 2 includes two ways of expressing the dispersion of the sample: the interquartile range, and the (adjusted) median absolute deviation (σ_M). The interquartile range is the difference between the lower and upper quartiles, which are the 25th and 75th percentiles, respectively, of the sample.[xvi] The interquartile range is useful for visualizing the sample (see discussion of box-and-whisker plots below), but it does not attempt to express a standard deviation for the sample, so it cannot be compared directly to the Gaussian standard deviation (nor the arithmetic standard deviation; see next section). For a true normal distribution, any estimate of the standard deviation will tend to be smaller than the interquartile range. For comparison to the standard deviation, we use the adjusted median absolute deviation (eq. **3**).[xx]

$$\sigma_M = 1.4826 \cdot \text{median}\left(|x - m|\right) \qquad (\mathbf{3})$$

The quantity, median($|x - m|$), is called the median absolute deviation, and the factor of 1.4826 adjusts this quantity to correct for underestimation of the sample standard deviation.

A common and useful method for visually conveying a large amount of information about a sample (including the median and interquartile range) in a compressed form is to use a box-and-whisker plot (Figure 6). This plot compares, side by side, the medians, interquartile ranges, and relative symmetry of samples of $\log|J|$ for all ten n-alkanethiols described in this paper.

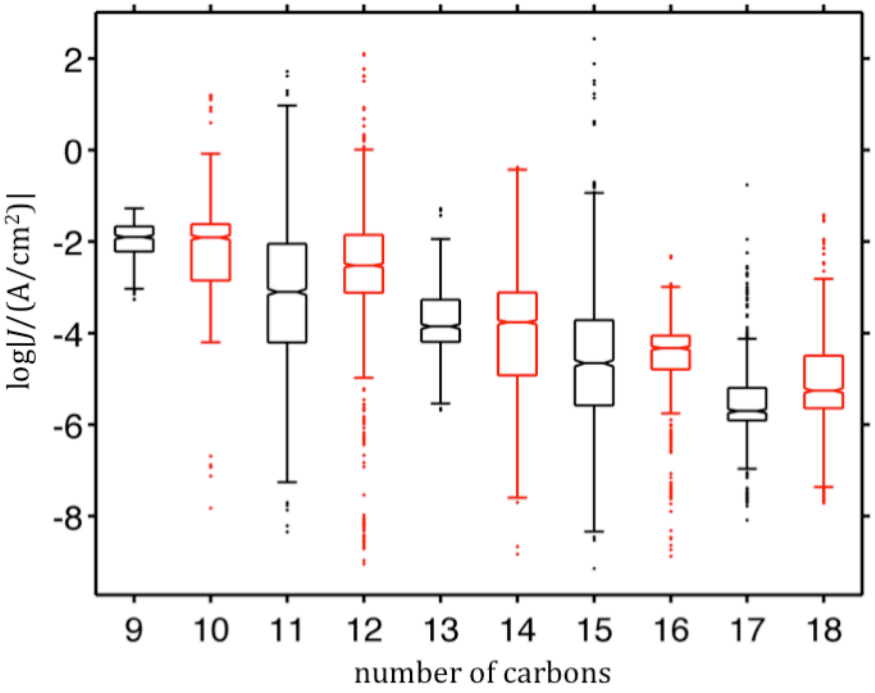


Figure 6:

Box and whisker plot of $\log|J/(A/cm^2)|$ vs. n for n-alkanethiols. The horizontal line within the box denotes the median of the distribution; the top and bottom of the box denote the upper and lower quartiles, respectively; the error bars (or "whiskers") extend to the datum furthest from the box, up to a distance of 1.5 times the interquartile range (the height of the box), in either direction. Points lying beyond this distance are defined as outliers, and appear as individual points. Shorts (values of $\log|J| > 2.5$) were excluded prior to calculating these statistics, to avoid unnecessarily skewing the distributions. Notches surrounding the median indicate the 95% confidence interval for the median.

Method 2 does not attempt to discriminate between different components of log|*J*| (as does Method 1, which uses our statistical model), but rather, follows the (probably informative) bulk of the data and resists extreme values that are probably non-informative. The influence of any single datum on the median does not depend on its value, but on its ordinal position with respect to other data in the sample. In fact, one could select any outlier (or even several of them) and move it arbitrarily far from the center of the distribution, without changing the median at all.[xvi] This action would, however, cause the arithmetic mean (see below) to grow arbitrarily large, following the value of the outlier. Long tails do affect the median, but not to the extent that they affect the arithmetic mean. For these reasons, Method 2 is less sensitive to outliers (and also long tails) than Method 3.

Because the median responds relatively weakly (compared to the arithmetic mean) to tails and outliers, but does not ignore them (as does the Gaussian mean), we observed (in Figure 5) that the estimates of Method 2 typically lay between those of Method 1 and Method 3. Although Method 2 is insensitive to the values of outliers, it is still affected by their presence. We still, therefore, chose to exclude shorts (using the procedure described in the previous section) before calculating m, σ_M, and the interquartile range, because we know *a priori* that shorts are non-informative. Again, while we defined shorts as values of log|*J*| > 2.5, the specific rule for excluding shorts will vary depending on what constitutes a short in a particular experimental system.

Method 3: Arithmetic Mean and Standard Deviation. The third method for estimating the location and dispersion of log|*J*| involves calculating the arithmetic mean (the first moment, eq. **4a**) and the arithmetic standard deviation (the square root of the second

moment about the arithmetic mean, eq. **4b**) of the sample.[xviii]

$$\mu_A = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \textbf{(4a)}$$

$$\sigma_A = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu_A\right)^2} \qquad \textbf{(4b)}$$

Here, x is the sampled variable (log|J|, in this case), and x_i is the ith observation (i.e. measurement) of x. In general use, the term "mean" most commonly refers to the first moment. With Method 1, even more so than with Method 2, it is essential to apply an exclusion rule to eliminate shorts, which bias the arithmetic mean much more strongly than the median.

In general, Method 3 responded strongly to long tails and outliers in histograms of log|J|. For the histograms shown in Figure 5, the arithmetic mean typically fell on the side of the peak with the longest tail, or the most outliers. Also, since most histograms had tails and/or outliers, the arithmetic standard deviation was usually greater than the Gaussian standard deviation, a fact that also affected the widths of the respective confidence intervals given by the two methods.

Confidence Intervals on Estimates of Location. The widths of the distributions of log|J| in Figure 5 (expressed by their dispersions), give the misleading impression that the estimates of the location for these distributions are imprecise. In fact, because of the large numbers of data in each distribution, the Gaussian mean, median, and arithmetic mean can all potentially be estimated with great precision, despite the dispersion in log|J|. An important way to express the precision and certainty of an estimated value is with a confidence interval.[xvi,xvii,xviii] If the assumptions underlying the method of estimation are correct (an important qualifier), then a confidence interval gives, with a specified

confidence level (usually 95%, 99%, or 99.9%), the range within which the true value being estimated lies. A 99.9% confidence level, for example, means that, if 1000 samples were collected from a population with a known location, then for 999 of those 1000 samples, the confidence interval surrounding the location estimated from the sample would contain the true location of the population. Figure 7 compares the 99.9% confidence intervals on the median, first moment, and Gaussian mean, for both odd and even n-alkanethiols, plotted against n.

Confidence intervals are closely related to statistical tests, to the extent that every confidence interval on an estimated value specifies the "acceptance region" of a statistical test—i.e. a test that checks for a statistically significant difference between the estimate and some other value will conclude that there is a statistically significant difference if, and only if, the value lies outside the confidence interval. Since every type of confidence interval corresponds to a different statistical test, there are many possible types of confidence interval that could be used.

Confidence Intervals for Methods 1 and 3. A useful confidence interval for both the Gaussian mean and Arithmetic mean corresponds to the so-called Z-test. Although the Z-test technically performs less well than another test—the t-test—when the population standard deviation is unknown (as with our measurements), when the number of data is large, the results of the two tests asymptotically converge.[xvi] There is some disagreement over what constitutes a "large" number of data, but for $N > 50$ the two tests are practically indistinguishable. Since we, therefore, have large numbers of data, we choose to define confidence intervals based on the Z-test, because they are computationally simpler than those based on the t-test. Both the Z-test and the t-test make three assumptions: i) that the

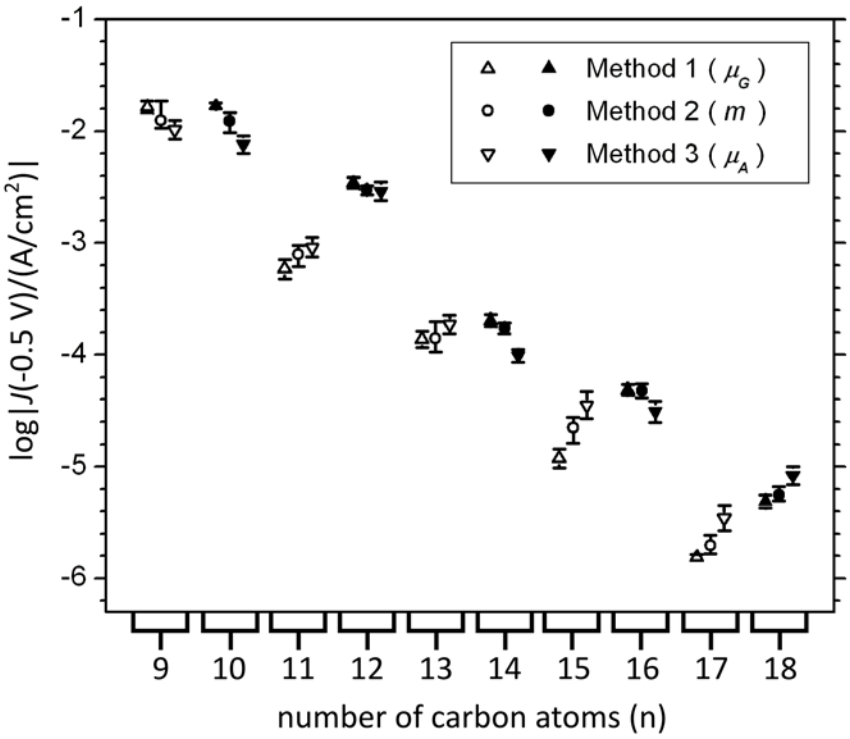


Figure 7:

Comparison of the locations, and the precisions of those locations, estimated by Methods 1 – 3 for n-alkanethiols (n = 9 – 18). All error bars indicate the 99.9% confidence interval; choosing the 99.9% confidence level for individual confidence intervals allows the set of all pairwise comparisons, across the series of n-alkanethiols, to have a collective confidence level of 99% (see text). The error bars do *not* signify the standard deviation (or other estimates of dispersion). Upward-facing triangles indicate Method 1 (the Gaussian mean), circles indicate Method 2 (median), and downward-facing triangles indicate Method 3 (Arithmetic mean). Open symbols denote odd n-alkanethiols, while closed symbols denote even n-alkanethiols.

parameter being estimated (the Gaussian mean or the arithmetic mean) is normally distributed,[xliv] ii) that this normal distribution has mean equal to the population mean, and iii) that this normal distribution has standard deviation equal to $s/N^{1/2}$ (where s is the population standard deviation).

The first assumption is rendered probable, even for non-normally distributed data, by the Central Limit Theorem.[xvi] The second assumption is only as reliable as the method on which it is based. For instance, it is probably closer to being true for Method 1 than for Method 3. The third assumption depends heavily on the independence of measurements of $\log|J|$. If measurements of $\log|J|$ are correlated, or "clustered" (as they probably are), then this assumption has been violated, and N must be corrected, as we discuss below. The extent to which our data violate this third assumption, and the magnitude of the correction needed to account for this violation, are two of the most crucial questions facing our analysis. The answers to these questions could significantly affect the confidence intervals we estimate and, thus, the conclusions we are able to draw from the data. For now, we give our best procedures, based on our current knowledge, with the cautionary note that further research is needed to address the independence of measurements of $\log|J|$.

For the Gaussian mean and arithmetic mean, the formula for confidence intervals based on the Z-test is given by eq. **5**, in which σ represents either σ_G or σ_A, as appropriate.

$$CI = z_{\alpha/2}\frac{\sigma}{\sqrt{N_{eff}}} \qquad (\mathbf{5})$$

The parameter $z_{\alpha/2}$ corresponds to the confidence level of $(1 - \alpha)$, and is the inverse of the cumulative distribution function for the standard normal distribution, evaluated at $\alpha/2$

($z_{\alpha/2}$ is, essentially, the number of standard deviations away from the mean one must go, in order to reach a value of $\alpha/2$ in the normal probability density function, eq. **2**). Some common values of $z_{\alpha/2}$ are: $z_{0.025} = 1.96$ for the 95% confidence level ($\alpha = 0.05$), $z_{0.005} = 2.576$ for the 99% confidence level ($\alpha = 0.01$), and $z_{0.0005} = 3.291$ for the 99.9% confidence level ($\alpha = 0.001$). The value N_{eff} is the effective sample size (eq. **6**).

$$N_{eff} = N\frac{1-\rho}{1+\rho} \tag{6}$$

N is the number of values of log|J| (the sample size) for the given SAM, and ρ is the average, normalized autocorrelation (eq. **7**) of all pairs of values of log|J|.[xlv]

$$\rho = \frac{2\sum_{i}\sum_{k>i}\left(\log|J_i| - \langle\log|J|\rangle\right)\left(\log|J_k| - \langle\log|J|\rangle\right)}{\sigma^2 N(N-1)} \tag{7}$$

If the individual values in a distribution are all independent and uncorrelated, then N_{eff} is equal to N. Because log|J| is measured within a hierarchy (of samples, tips, junctions, and traces), individual values are correlated, to some degree (as a result of the "clustering", discussed in the Experimental Design). For instance, two values of log|J| measured on different traces in the same junction tend to be more similar than two values of log|J| measured using different junctions, formed with different tips. Although measurement of many traces on the same junction, many junctions using the same tip, many tips on the same sample, and multiple samples per compound is necessary to guard against anomalies, this practice leads to artificially high values of N, because of the decreasing amount of new information that each subsequent measurement offers, in comparison with other measurements at the same level of the hierarchy. To account for this tendency, ρ is defined so that if values of log|J| measured close to each other are

similar, N_{eff} decreases and the confidence interval expands to correct for this "oversampling".[xlv]

We caution that using the corrected sample size, N_{eff}, does not necessarily and automatically validate a confidence interval. Further experiments are needed to determine whether this procedure offers a strong enough correction for the oversampling in our measurements, or whether a stronger correction is needed. A good way to avoid the need for such a correction is to refrain from collecting data that are known, in advance, to be probably correlated to one another. For instance, since two values of $\log|J|$ measured using the same junction will probably be correlated, it is advisable to collect only one, or a few, values of $\log|J|$ for each junction.

Note that, because of the large numbers of data (even after correcting for oversampling), the confidence interval of, for instance, the Gaussian mean is far smaller than the interval defined by $\mu_G \pm z_{\alpha/2}\sigma_G$. The latter is called a prediction interval, and denotes the range within which $(1 - \alpha)$% of the actual data lie, in a normal distribution.[xvi] The confidence interval, by contrast, expresses the range of probable values of a particular estimate (e.g. of the location of the data), not of the data themselves. In practical terms, the difference between a prediction interval (on a sample) and a confidence interval (on a statistic) means that, while individual measurements of $\log|J|$ may be widely scattered, the statistic (e.g. the Gaussian mean) describing the distribution can be estimated with great precision.

Confidence Intervals for Method 2. A confidence interval for the median is defined by quantiles.[xlvi] The q quantile (where q is a number between zero and one) of a distribution divides the distribution such that the fraction q of the data are less than or equal to the

quantile, and the fraction $(1 - q)$ of the data are greater than or equal to the quantile. For instance, the median is the quantile with $q = 0.5$ (i.e. the 50th percentile) and the lower and upper quartiles have $q = 0.25$ and 0.75, respectively (see Supporting Information for more details). The 99.9% confidence interval for the median is defined by two quantiles, q_- and q_+, that are given by eq. **8a** and **8b**, respectively, where N_{eff} and $z_{\alpha/2}$ are defined as above ($z_{\alpha/2} = 3.291$ for the 99.9% confidence interval).

$$q_- = 0.5\left(1 - \frac{z_{\alpha/2}}{\sqrt{N_{eff}}}\right) \quad \textbf{(8a)}$$

$$q_+ = 0.5\left(1 + \frac{z_{\alpha/2}}{\sqrt{N_{eff}}}\right) \quad \textbf{(8b)}$$

If $Q(q)$ is the value of the q quantile (e.g. $Q(0.5) = m$, the median), then the 99.9% confidence interval on the median is $(Q(q_-), Q(q_+))$.[xlvi]

Confidence Intervals, Precision, and Accuracy. As stated above, a confidence interval corresponds to the region in which the corresponding statistical test would fail to reject the null hypothesis (e.g. that there is no statistically significant difference between the estimated parameter and a fixed value) at the stated confidence level.[xvi] Confidence intervals can be used to check for a statistically significant difference between the locations of two samples of $\log|J|$. If, for instance, the Gaussian means (μ_{G1} and μ_{G2}) of $\log|J|$ for two compounds satisfy inequality **9**, then one can conclude, at the specified confidence level, that the two values are different (i.e. the null hypothesis can be rejected).

$$\left|\mu_{G1} - \mu_{G2}\right| > \sqrt{CI\left(\mu_{G1}\right)^2 + CI\left(\mu_{G2}\right)^2} \quad \textbf{(9)}$$

Note that this inequality can be satisfied (and there can be a statistically significant difference between two values) even if the two confidence intervals overlap somewhat.

If the confidence intervals do not overlap, then there is automatically a statistically significant difference between the two values. For instance, comparing the 99.9% confidence intervals around μ_G for $S(CH_2)_{10}CH_3$ and $S(CH_2)_{11}CH_3$ shows that they do not overlap. We conclude that the two values of μ_G are different (i.e. we reject the null hypothesis, that they are the same), at the 99.9% confidence level. The confidence intervals around μ_G for $S(CH_2)_8CH_3$ and $S(CH_2)_9CH_3$, however, overlap *and* fail to satisfy inequality **9**, so we cannot conclude that the two values are different (i.e. we cannot reject the null hypothesis).

The confidence intervals in Figure 7 show, at a glance, the precision of the locations estimated by the Gaussian mean, median, and arithmetic mean. The confidence intervals are generally narrow, indicating that μ_G, m, and μ_A are precise. Figure 7 also shows, however, the extent to which Methods 1 – 3 can differ from one another. In many cases, the 99.9% confidence intervals for these three statistics do not overlap. Obviously, although μ_G, m, and μ_A are all precise, they cannot all be accurate estimators of the locations of the populations of Ag^{TS}-SR//Ga_2O_3/EGaIn junctions.

As we argued in the Experimental Design section, the accuracy of each method depends on how it distinguishes between informative and non-informative data. We have confidence in our statistical model—that informative measurements of $\log|J|$ are approximately normally distributed—and we believe, therefore, that Method 1 is accurate. We also trust the accuracy of Method 2, because it does not respond strongly to

extreme values, which are likely to be non-informative. We do not trust the accuracy of Method 3, because it is strongly influenced by data that are likely to be non-informative.

The Odd-Even Effect Revisited: Confidence Intervals in Multiple Comparisons. The difference between the three approaches is not superficial; they lead to statistically different conclusions about, for example, the difference (or similarity) between odd and even alkanethiols. In our previous paper,[xiv] we gave two statistical justifications for our claim that there is an odd-even effect—i.e. that odd and even alkanethiols could not both be described by eq. **1**, using the same parameters. One justification depended on statistical tests to compare the Gaussian means for adjacent n-alkanethiols (a procedure equivalent to comparing confidence intervals, as noted above). In that paper, we used Student's t-tests to compare μ_G, at the 95% confidence level, for every pair of adjacent alkanethiols (e.g. $S(CH_2)_8CH_3$ and $S(CH_2)_9CH_3$). Because every odd alkanethiol except $S(CH_2)_8CH_3$ had a lower μ_G than *both* of the adjacent even alkanethiols, and because each comparison was performed using a t-test at the 95% confidence level, we concluded, with 95% confidence, that there exists an odd-even effect. We now know that, although our conclusion happened to be correct, the manner in which we arrived at that conclusion was flawed.

Our argument in that paper suffered from two deficiencies: i) we did not use eqs. **6** and **7** to correct the value of N for oversampling, and ii) we did not account for a pitfall that occurs when a single inference is supported by the repeated use of a statistical test. A statistical test with confidence level $(1 - \alpha)$ has a probability α of falsely rejecting the null hypothesis—i.e. concluding that a statistically significant difference exists, when in fact, it does not (a so-called "type I error").[xlvii] When c separate tests with confidence

level $(1 - \alpha)$ are performed, the probability of a type I error increases to $1 - (1 - \alpha)^c$; thus, the confidence level of the *entire set* of c tests decreases to $(1 - \alpha)^c$. Because we performed $c = 9$ separate t-tests (one for each adjacent pair in the range n = 9 – 18) at the 95% confidence level, the true confidence level of our conclusion was only 63%.

In order to achieve a true confidence level of $(1 - \alpha)$ for c tests, it is necessary to increase the confidence level for each individual test to $(1 - \alpha_{new}) = (1 - \alpha)^{1/c}$; this procedure is called the Šidák correction.[xlvii] It is for this reason that we have chosen to plot, in Figure 7, the 99.9% confidence intervals: for $c = 9$, choosing $(1 - \alpha_{new}) = 0.999$ for each test leads to an overall confidence level, for all comparisons, of 99.1%. With 99.9% confidence intervals on each estimated location, we can then draw conclusions about the entire dataset at the 99% confidence level. A procedure designed to safeguard against type I errors from performing several separate tests in a row is called a "multiple comparison" test.[xvi]

In Figure 7, the odd-even effect is quite apparent when comparing the 99.9% confidence intervals around values of μ_G Gaussian means: there is a statistically significant zig-zag alternation in μ_G with increasing n, as opposed to the monotonic decrease expected from eq. **1** if there were no difference between odd and even n-alkanethiols. In four out of five cases, μ_G for a given odd n-alkanethiol is less than μ_G for *both* adjacent (even) alkanethiols (e.g. n = 11 has a lower μ_G than both n = 10 and n = 12). Using the Gaussian means, n = 11, 13, 15, and 17 all meet this criterion (but n = 9 does not). Comparing the medians, rather than the Gaussian means, of n-alkanethiols shows a statistically significant alternation in three out of five cases (n = 11, 15, and 17); still a majority. The odd-even effect becomes less apparent, however, when using the

arithmetic mean: only n = 11 and 17 meet the above criterion, while n = 9, 13, and 15 fail. If one had no guiding principle for choosing between these three methods of analysis, one might not conclude that there is an odd-even effect, from these results, at the 99% confidence level. Based on our confidence in the accuracy of μ_G and *m*, and our lack of confidence in the accuracy of μ_A (as explained in the Experimental Design section), however, we choose to trust Methods 1 and 2 over Method 3. We can, thus, affirm our previous conclusion that there is an odd-even effect.

Our intent in discussing the dramatic differences between descriptive statistics is not to cast doubt on the existence of the odd-even effect (we now have even stronger evidence for it than we did in our previous paper; see below), but to highlight the fact that the choice of method for analyzing log|*J*| can have a large effect on statistical inferences based on confidence intervals or statistical tests. When performing any statistical analysis, it is, therefore, important to i) identify the method(s) used to estimate the parameters being compared, ii) state the assumptions underlying the method(s), and iii) offer convincing justifications for those assumptions (or alternative methods, in case those assumptions are later shown to be incorrect).

Calculating Trend Statistics: β and J_0. Because eq. **1** predicts a linear dependence of log|*J*| on *d* (or n), determining β and J_0 for a series of compounds involves i) plotting values representing log|*J*| (at a given applied bias) against n, followed by ii) fitting this plot with a line (see Figure 8 for examples). The slope of this line is -log(*e*)β, and the y-intercept is log|J_0|. Accordingly, there are two areas in which Methods 1 – 4 differ from one another: i) the data used to represent log|*J*| in the plot of log|*J*| vs. n, and ii) the algorithm used to fit a line to this plot.

Methods 1 – 3 and Trend Statistics. Methods 1 – 3 use their respective estimators for location (μ_G, m, and μ_A), in order to represent $\log|J|$ for each compound in plots of $\log|J|$ vs. n. All of Methods 1 – 3 then use the same algorithm to fit their respective plots. The linear, "least-squares" algorithm constructs the line that minimizes the sum of the squares of all errors (differences between values of $\log|J|$ and the fitted line).[xxi] This algorithm is the standard algorithm used in most procedures for performing linear fits.

For example, using Method 1, Figure 8A shows a plot of μ_G (at a bias of V = -0.5 V) vs. n for the ten n-alkanethiols, and indicates the linear fits (solid lines) for both odd and even n-alkanethiols. The dotted lines in Figure 8A (as well as 7B and 7C) represent the so-called 99% confidence bands of the fitted function; these bands contain, with 99% confidence, the region within which lies the true linear fit to the data (these confidence bands are subject to the same assumption of independence, and the same caveat about correlation of data, as the confidence intervals defined above for single-compound statistics).[xviii]

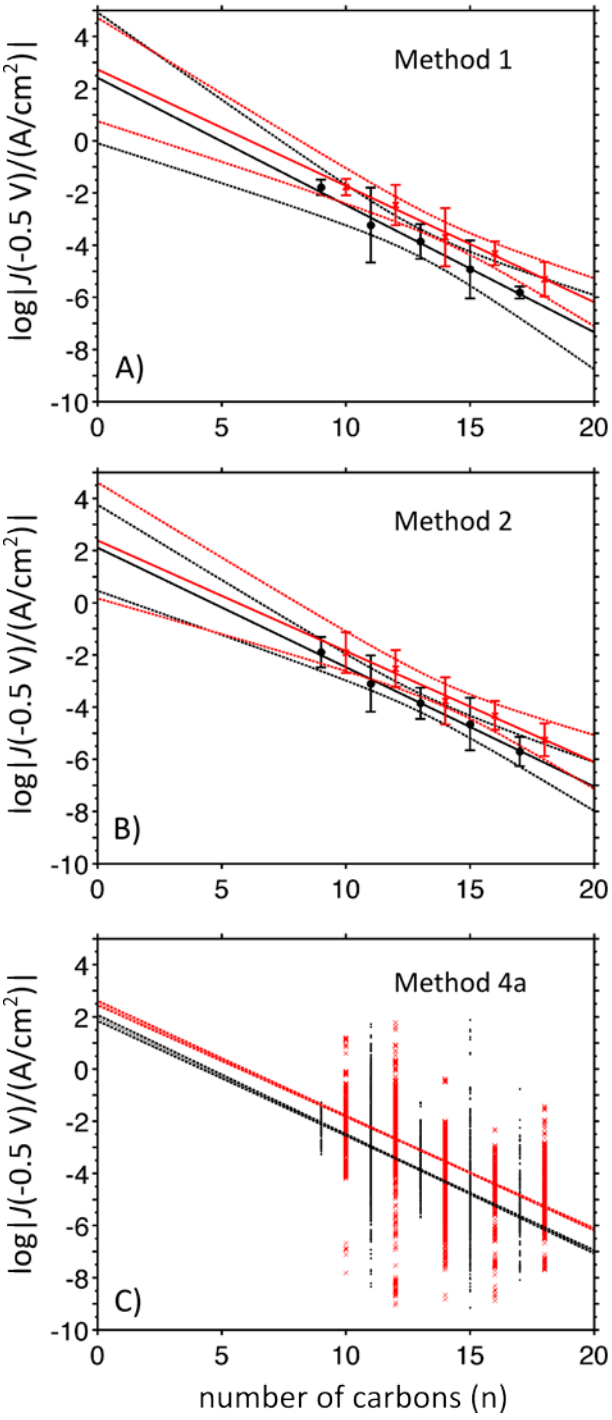
Because each method uses estimates of the location of $\log|J|$ for each compound in order to calculate trend statistics, each method carries forward, into the trend statistics it calculates, the assumptions and relative accuracy it had when estimating the location of $\log|J|$. Just as with single-compound statistics, therefore, Methods 1 and 2 are more accurate than Method 3 in their estimates of β and J_0.

Because Methods 1 – 3 compress all of the dispersion in the sample of $\log|J|$ into a single value for each compound, the choice of how to separate informative data from non-informative data has already been made. As such, the sensitivity of the linear fitting algorithm to deviations of $\log|J|$ from normality is irrelevant in Methods 1 – 3, because

Figure 8:

Comparison of Methods 1, 2, and 4a for determining β and J_0 from measurements of $\log|J|$ for n-alkanethiols (n = 9 – 18). A) Method 1: linear fits (solid lines), with least sum of squares due to error, of the Gaussian means (μ_G) of odd-numbered (black circles) and even-numbered (red "x"s) n-alkanethiols, respectively. Each line is fit to five data points (the fit, therefore, has three degrees of freedom). Dotted lines show the 99% confidence bands of the each fit – these bands denote (with 99% confidence) the region that contains the true fit (i.e. the fit that gives the true values of β and J_0). Error bars representing the Gaussian standard deviation (σ_G) are shown for reference only and do not affect the fit. B) Method 2: linear fits (solid lines), with least sum of squares due to error, of the medians (m) of odd-numbered (black circles) and even-numbered (red "x"s) n-alkanethiols, respectively. Again, each line is fit to five data, and dotted lines indicate the 99% confidence bands. Error bars representing the adjusted median absolute deviation (σ_M) do not affect the fit. C) Linear fits (solid lines), with least sum of absolute errors, of *all* values of $\log|J|$, for odd (black circles) and even (red "x"s) n-alkanethiols. Dotted lines indicate the 99% prediction bounds. No weights were applied to the data; however, shorts ($\log|J| > 2.5$) were excluded to avoid biasing the fits towards high values of $\log|J|$. The plotted values of $\log|J|$ are the same as those in the histograms in Figure 4, but the large numbers of data cause many points to be superimposed on one another. Thus, this representation visibly distorts (i.e. flattens) the data and disguises the concentration of data at the center of each distribution.

Figure 8 (Continued)



those deviations have already been taken into account. The step of calculating β and J_0 using Methods 1 – 3 is, therefore, relatively straightforward, but the use of only one datum per compound as the input to the fitting algorithm leads to estimates that are relatively imprecise, compared to those of Methods 4a and 4b.

Method 4a and Trend Statistics. Method 4a (along with Method 4b) does not use any summary, such as the location, to represent log|*J*| in plots of log|*J*| vs. n. Instead, after excluding shorts from each sample, it plots all values of log|*J*| (at a given applied bias) for each compound (often resulting in thousands of data on a single plot; see Figure 8C). Both informative and non-informative data are included on this plot; thus, the algorithm used to fit a line to the plot must, in some way, distinguish between them. Method 4a uses an algorithm that minimizes the sum of the *absolute values* (rather than the *squares*) of all errors,[xxi] whereas Method 4b employs the traditional, least-squares approach. In this sense, Method 4a is somewhat analogous to Method 2, because the median of a sample also minimizes the sum of the absolute differences between the median and the data.[xx] As such, Method 4a is less sensitive to extreme values than Method 4b (which is roughly analogous to Method 3), and is, therefore, more accurate. Because the precision of a fitted line increases with approximately the square root of the number of data used to estimate these parameters,[xxi] Method 4a is much more precise than Methods 1 – 3, in estimating trend statistics.

Method 4b and Trend Statistics. Like Method 4a, Method 4b plots all values of log|*J*| (at a given applied bias) vs. n, after excluding shorts. Method 4b differs from Method 4b, however, in that it uses a least-squares algorithm (that minimizes the sum of the squares of all errors) to fit a line to this plot. In other words, the influence of an extreme value on

the fit is proportional to the square of its distance from the fit line.[xxi] In this way, Method 4b resembles Method 3, because the arithmetic mean of a sample minimizes the squares of the differences between itself and the data in the sample (i.e. the variance). Method 4b, therefore, responds strongly to extreme values (which are likely to be non-informative), and is not an accurate method for calculating β and J_0. We include a discussion of this method simply because it is a commonly used technique.

Precision and Accuracy of Methods 1 – 4, with Respect to Trend Statistics. One of the most significant factors affecting the accuracy of any fit, regardless of the method used, is the number of compounds for which the data have been collected (i.e. the number of distinct values of n). In our first publication on the use of Ga_2O_3/EGaIn electrodes,[xii] we reported a value of β that was erroneous, partly because we performed a linear fit using data from only three compounds. Having a small range in the independent variable (n) gives inordinate influence to extreme values of log|*J*| on the slope and y-intercept of the fitted line, whereas having a large range in n helps to "fix" the fitted line at both ends and, therefore, to reduce the error in the position of the line. In our subsequent publication,[xiv] we were able to correct our earlier error by using five compounds, instead of three, to perform the fitting. (Due to experimental limitations, we were only able to measure compounds for which n = 9 – 18; in other words, five odd n-alkanethiols and five even n-alkanethiols). Even more important than choosing the correct method of analysis or collecting many values of log|*J*|, therefore, is measuring the full range of accessible compounds.

Table 1 and Figure 9 compare the values of β and log|J_0|, for V = -0.5 V, determined using each method, and give the 99% confidence intervals around these values.

Confidence intervals for β and $\log|J_0|$ are defined in exactly the same way as confidence intervals for μ_G, m, and μ_A (and are subject to the same assumptions and caveats):[xvi] if, for example, β were determined 100 times from 100 different random samples of the same data, then 99 times out of 100, the 99% confidence interval around the *estimated* values of β will contain the *true* value of β.[xvii] Unlike with single-compound statistics, however, calculating confidence intervals on trend statistics involves mathematical techniques that are outside the scope of this paper to explain. We used statistical software (the curve-fitting tool in MATLAB 7.10.0.499 R2010a) to calculate the confidence intervals in Table 1.

Recall that, for single-compound statistics, the confidence intervals around, for example, μ_G and μ_A for the same compound often did not overlap. It was, therefore, clear that Methods 1 – 3 were truly different from one another in their estimates of the location of $\log|J|$. With trend statistics, however, the confidence intervals of the values estimated by Methods 1 – 3 all overlap. Paradoxically, even though we have already shown that Methods 1 – 3 importantly differ in their approach to the data, and in their estimates of single-compound statistics, when it comes to estimating trend statistics, the differences between these methods blur into statistical insignificance. This vexing result arises because the differences between Methods 1 – 3 are overshadowed by the lack of precision associated with fitting a line to only five data (i.e. with three degrees of freedom). Clearly, when Methods 1 – 3 pre-process the data (*via* single-compound statistics), much useful information is being lost.

Methods 4a and 4b have much greater precision than Methods 1 – 3, because they fit many data (hundreds or thousands, with as many degrees of freedom). For both β and

Table 1:

Estimates of β and J_0 using Method 4a are precise and agree with those of Method 2

Method	Dataset	d.o.f.[a]	R^2 [b]	$\log\lvert J_0/(A/cm^2)\rvert$ [c]	$\beta\,(n_C^{-1})$ [c]	$\beta\,(Å^{-1})$ [c]
1 (Gaussian means)	Odd	3	0.9870	2.4 ± 2.7	1.12 ± 0.47	0.89 ± 0.37
	Even	3	0.9916	2.7 ± 2.1	1.03 ± 0.34	0.81 ± 0.27
2 (Medians)	Odd	3	0.9936	2.1 ± 1.8	1.05 ± 0.31	0.84 ± 0.25
	Even	3	0.9883	2.4 ± 2.4	0.98 ± 0.38	0.77 ± 0.30
3 (Arithmetic means)	Odd	3	0.9939	1.7 ± 1.6	0.96 ± 0.28	0.76 ± 0.22
	Even	3	0.9596	1.9 ± 4.2	0.91 ± 0.67	0.72 ± 0.53
4a (all data, least absolute errors)	Odd	6383	0.8575	1.96 ± 0.12	1.033 ± 0.021	0.819 ± 0.017
	Even	10054	0.8539	2.53 ± 0.09	1.000 ± 0.015	0.792 ± 0.012
4b (all data, least square errors)	Odd	6383	0.3234	1.34 ± 0.26	0.903 ± 0.046	0.716 ± 0.036
	Even	10054	0.4277	1.99 ± 0.18	0.938 ± 0.030	0.744 ± 0.024

[a] Degrees of freedom, with respect to error (not total degrees of freedom), for the fit. For Methods 4a and 4b, the d.o.f. is less than the sum of N for odd (even) alkanethiols, because shorts have been excluded from the data, prior to fitting.

[b] Coefficient of determination for the fit

[c] Values are given with 99% confidence intervals

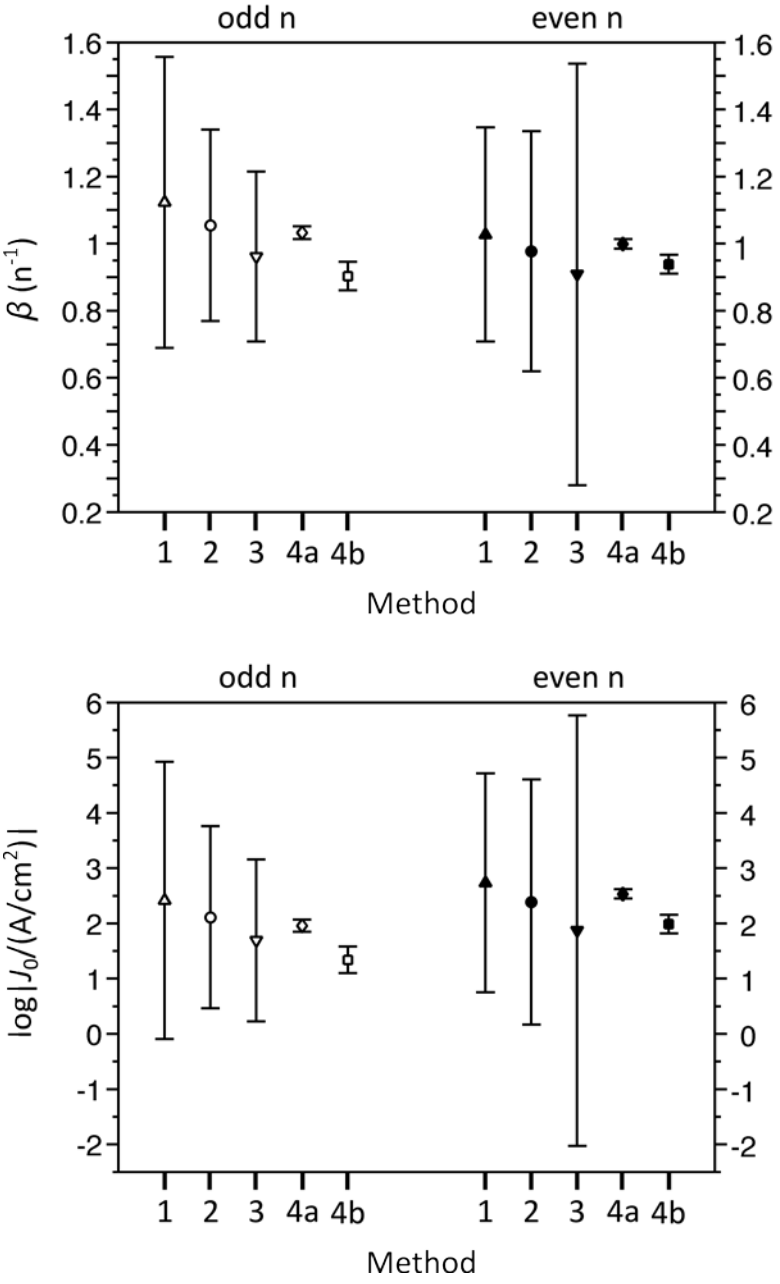


Figure 9:

Values of β (n^{-1}; top) and $\log|J_0|$ (bottom), at V = -0.5 V, determined by all methods for odd (open symbols) and even (closed symbols) n-alkanethiols. The error bars indicate the 99% confidence intervals.

$\log|J_0|$, the confidence intervals around the values estimated by Methods 4a and 4b do not overlap (Figure 9), and are roughly one order of magnitude smaller than the confidence intervals for Methods 1 – 3.

We desire a method that is both accurate and precise for estimating trend statistics. If our statistical model is valid, then the relative accuracy of the methods can be expressed in the following series: 1 > (2, 4a) >> (3, 4b); if our model is incorrect, then the series is: (2, 4a) > 1 >> (3, 4b). Regardless of the validity of our model, the relative precision of the methods follows the series: (4a, 4b) > (1, 2, 3). We can rule out Methods 3 and 4b on the grounds of inaccuracy, and Method 2 on the grounds that it is less precise, but no more accurate, than Method 4a. We are then faced with a choice between Method 1 and Method 4a. If our statistical model is correct, then Method 1 is more accurate than Method 4a. These two methods, however, agree reasonably well—both β and $\log|J_0|$ are lower for Method 4a than for Method 1, but the large confidence intervals for Method 1 completely engulf the values of Method 4a (Figure 9)—so if one method is reasonably accurate, then by extension, both methods are reasonably accurate. Since, for these particular data, the accuracy of Method 4a has been checked by comparison to Method 1, we have grounds for using Method 4a in this specific case.

The coefficient of determination (R^2) does not measure precision or accuracy. Table 1 gives the coefficient of determination, R^2, for each linear fit. The value of R^2 is not, in general, a guide to the precision or accuracy of a fit.[xvi,xviii] Rather, it represents the fraction of the variation in the data that is explained by the model used to fit the data (eq. **1**, in this case).

For Methods 1 – 3, the data being explained are not the primary data (log|*J*|), but the locations (μ_G, *m*, or μ_A) estimated for each compound. The fact that, for example, Method 1 yields large values of R^2 means that changing the length of the alkanethiol used to form the SAM explains the vast majority of the variation in the *Gaussian mean* of log|*J*| across the series of n-alkanethiols.

The relatively low values of R^2 for Methods 4a and 4b reflect the fact that the data being explained are no longer pre-processed single-compound statistics, but rather all measurements of log|*J*|. According to the values of R^2 for Methods 4a and 4b, therefore, changing the length of the alkanethiol used to form the SAM explains only some of the variation observed in all measurements of log|*J*| across the series of alkanethiols. This fact is unsurprising, because we are already aware that defects (which are largely independent of n) in Ag^{TS}-SR//Ga_2O_3/EGaIn junctions explain a significant portion of the dispersion of individual samples of log|*J*|. Furthermore, many of the data in each sample of log|*J*| are non-informative, and we should hope that the Simmons model would *not* explain these data.

While the coefficient of determination can convey useful information, it is not necessarily an indicator of either the accuracy or the precision of the method used. Its value should not, therefore, determine the choice of method.

Method 4a identifies J_0 as the major source of the odd-even effect. Methods 1 – 3 are too imprecise (for estimating trend statistics) to locate the origin of the odd-even effect in either parameter of the Simmons model (eq. **1**). As with single-compound statistics, statistical tests can be performed on trend statistics by comparing the confidence intervals of two values.[xvi] Table 1 shows that, for each of Methods 1 – 3, the confidence intervals

around the values of β for odd and even alkanethiols overlap. The same is true of $\log|J_0|$. Using Methods 1 – 3, therefore, one cannot reject the null hypothesis that the values of β (or $\log|J_0|$) for odd and even alkanethiols are the same.

Method 4a is, however, precise enough to offer an instructive comparison between odd and even n-alkanethiols, with respect to J_0 and β. In order to perform a comparison, we could simply observe that the 99% confidence intervals on the two values of, for example, $\log|J_0|$ do not overlap. To achieve a quantitative comparison, however, we use the confidence intervals on both values of $\log|J_0|$ to calculate the probability (p) of the null hypothesis that $\log|J_{0,odd}| = \log|J_{0,even}|$.[xvi] We explain the procedure for calculating p, given two values with confidence intervals, in the Supporting Information. Using this procedure, we find that $p = 1.34 \times 10^{-4}$ for the null hypothesis, so we can reject it, and conclude, with over 99% confidence, that $\log|J_{0,odd}| < \log|J_{0,even}|$. Note that, if future experiments show the need for a stronger correction for correlation among values of $\log|J|$ than that provided by equations **6** and **7**, then the value of p would increase, and the confidence level of this conclusion would be weakened (perhaps significantly). For now, we tentatively conclude that there is a statistically significant difference between $\log|J_{0,odd}|$ and $\log|J_{0,even}|$.

On the other hand, with respect to β, the null hypothesis that $\beta_{odd} = \beta_{even}$ has $p = 0.20$, so it cannot be rejected at the 99% confidence level. (See Table 2 for p values for all methods). According to Method 4a, therefore, the difference in J_0 is significant (i.e. certain), but there is no significant difference in β.

The fact that the difference in J_0 is significant does not automatically mean that this difference is sufficient to explain the magnitude odd-even effect. To investigate this

Table 2:

Results (*p*-values) from t-tests show that a difference in J_0 between odd and even alkanethiols gives rise to the odd-even effect (when using Method 4a)

Method	$p(\beta_{odd} = \beta_{even})$[a]	$p(J_{0,odd} = J_{0,even})$[b]
1 (Gaussian means)	0.87	0.93
2 (Medians)	0.88	0.93
3 (Arithmetic means)	0.94	0.97
4a (all data, least absolute errors)	0.20	1.34×10^{-4}
4b (all data, least square errors)	0.52	0.041

[a] Probability of the null hypothesis: that the values of β for odd and even n-alkanethiols are the same.

[b] Probability of the null hypothesis: that the values of J_0 for odd and even n-alkanethiols are the same.

question, it is necessary to compare the magnitude of the difference in J_0 ($\log|J_{0,even}| - \log|J_{0,odd}| = 0.57$) with the magnitude of the difference between $\log|J|$ for odd and even alkanethiols. The difference in $\log|J|$ can be estimated by comparing the values of the linear fits to each dataset at the midpoint of the series (n = 13.5). The linear fit to the even n-alkanethiols interpolates a value at n = 13.5 of $\log|J_{even}| = -3.33$, while the linear fit to the odd n-alkanethiols interpolates a value of $\log|J_{odd}| = -4.10$ at the same point. The magnitude of the odd-even effect is 0.77 (i.e. $\log|J_{odd}| - \log|J_{even}| = 0.77$, or $J_{even}/J_{odd} = 5.89$), at the point n = 13.5. Method 4a estimates that the difference in $\log|J_0|$, therefore, accounts for about 74% of the odd-even effect at the midpoint of the series. At the beginning of the series (n = 9), the difference in $\log|J_0|$ explains ~ 81% of the odd-even effect, while this difference explains only ~ 68% of the odd-even effect at the end of the series (n =18), because the linear fits diverge as n increases. We conclude, therefore, that J_0 accounts for about 70 – 80% of the odd-even effect, among the n-alkanethiols we have measured. The remaining 20 – 30% of the odd-even effect is, so far, unexplained; it may be due to β (i.e. in the case that there is a difference in β, but our analysis is not powerful enough to detect it), or to other factors that the Simmons model fails to take into account.

Conclusions

Difficult problems can be made tractable with careful statistical analysis.

Analyzing charge transport through SAMs is a difficult problem for two principal reasons. i) Measurements of $\log|J|$ are noisy and contain artifacts that are difficult to separate from real data (Figure 3). ii) The dispersion in each sample of $\log|J|$ can be a

significant fraction of the range of log|*J*| across a series of SAMs (Figure 2); as a result, the spread in the data threatens to drown out the effect being studied. Improvements in experimental techniques will probably mitigate, but not eliminate, these problems. Artifacts are inevitable when contacting an area of several hundred μm^2 on a disordered monolayer that is ~ 1 nm thick. Furthermore, even if the average thickness of the SAM varies by only 1 C–C in either direction, *J* will vary over more than an order of magnitude.

The primary purpose of this paper is to show that, in the face of such difficulties, careful statistical analysis can still extract useful information and generate confident conclusions, or at least bound uncertainty and lack of confidence. What is required, in order to draw useful conclusions from the data, is: i) *accuracy*, which requires an idea (better yet, a statistical model) of how to distinguish data that convey information about charge transport through the SAM from data that do not, and ii) *precision*, which requires large numbers of data, in order to reduce the size of confidence intervals and give power to statistical tests.

Proper design of data collection is important to accuracy. In this paper, we have discussed two ways in which the collection of data, *prior to* any analysis, can influence the accuracy of the conclusions. i) Because measurements of log|*J*| are not completely independent of one another (e.g. two values of log|*J*| measured using the same Ag^{TS}-SR//Ga_2O_3/EGaIn junction will be more correlated than two values of log|*J*| measured using different junctions), there is a tendency to overestimate the sample size, and to underestimate the widths of confidence intervals. We have discussed a possible correction for this tendency using equations **6** and **7**, but the best way to escape this

pitfall is to avoid collecting correlated data in the first place. In our previous paper,[xiv] we collected approximately 40 values of J (i.e. ~ 20 $J(V)$ traces, each giving two values of J at each voltage) for each junction, but as a result of the current analysis, we now realize that this number was far too large. The first $J(V)$ trace is often noisy, so it may be desirable to collect more than one trace, but we recommend collecting about three, and no more than five, $J(V)$ traces (i.e. six to ten values of J) for each junction. We are currently testing and refining specific recommendations related to the protocol for collecting data; we will publish these results in a separate paper.

ii) Using few compounds (values of n) for determining trend statistics can lead to estimates that appear precise, but are probably inaccurate. Increasing the range of values of n over which data are collected can dramatically improve the accuracy of the linear fit to a plot of $\log|J|$ vs. n. It is, therefore, desirable to use as wide a range of n as the experimental system will allow, in order to ensure accurate results. Given the choice between measuring a greater number of compounds or a greater number of data for each compound, always choose the former; 10^3 measurements across five compounds will give more accurate results than 10^4 measurements across three compounds. Even though the latter approach will lead to greater precision than the former, precision without accuracy is useless.

Methods 1, 2, and 4a are all acceptably accurate, while Methods 3 and 4b are not. The accuracy of each method of statistical analysis depicted in Figure 4 depends on the correctness of the assumptions on which it rests, with respect to how to interpret real data. Method 1 assumes that informative measurements of $\log|J|$, for which the Simmons model (eq. **1**) is a valid description, constitute a normal distribution, and that any

deviations of log|*J*| from normality are not informative. Methods 2 and 4a assume that informative measurements of log|*J*| (regardless of how they are distributed) represent the bulk of the data, and that non-informative measurements comprise extreme values. These two assumptions are similar, but the first assumption is stronger than the second, in that it rises to the level of a true statistical model by positing a (normal) shape for the distribution of log|*J*|. We have offered justifications for this assumption (see Experimental Design), and we are reasonably confident that it is correct, but we recognize that it could still fail. In light of this possibility, we have included Methods 2 and 4a in order to add flexibility to our analysis. If future research overturns the assumption that log|*J*| is normally distributed, then the conclusions of this paper will still be valid, because they are supported by Methods 2 and 4a, which do not assume normality.

While we favor Method 1 (fitting Gaussian functions to histograms of log|*J*|) over Method 2 (using the median and interquartile range), they are probably both accurate enough to use in reporting single-compound statistics (i.e. the location and dispersion of samples of log|*J*|, as well as confidence intervals that allow comparisons between two compounds). For trend statistics (e.g. β and J_0) that involve fitting plots of log|*J*| vs. n, Method 4a (plotting all data, and fitting to minimize the sum of the absolute values of errors) is just as accurate as Method 2, and about an order of magnitude more precise. We, therefore, recommend using Method 4a for calculating trend statistics, as long as the results do not contradict those of Methods 1 and 2. We do not recommend using Method 3 (the arithmetic mean and standard deviation) or Method 4b (fitting lines to log|*J*| vs. n

using an ordinary, least-squares algorithm), because these methods respond too strongly to extreme values of $\log|J|$, which probably do not represent informative data.

Alongside these general recommendations, we must emphasize that the choice of method for statistical analysis should be made on a case-by-case basis, after careful consideration of factors influencing the data. For example, in a recent analysis of charge transport through molecules with different degrees of conjugation, Chiechi *et al.*[xlviii] encountered a situation where Method 1 was clearly superior to Method 2. They measured values of J for these molecules approached the range ($J \approx 10^2$ A/cm^2) where the resistance[xiii,xiv] of the Ga_2O_3/EGaIn electrode begins to limit charge transport through the junction. This artifact, which invalidated the Simmons model as a description of J above this point, caused the high end of the histogram of $\log|J|$ to be truncated, and the distribution to deviate strongly from normality. Method 1 ignored the deviation from normality and essentially extrapolated the missing tail of the histogram, in order to reconstruct an accurate picture of charge transport based on the Simmons model. In other cases, however, there may be reasons to avoid Method 1 and use Method 2 instead. One advantage of Method 2 is that it can summarize many data in a clear and visually accessible format: the box-and-whisker plot (Figure 6).

Precision depends on large numbers of data. The precision of any method of statistical analysis is most clearly seen in the confidence intervals it produces. The width of a confidence interval is proportional to the dispersion in the data, and inversely proportional to the square root of the sample size (assuming independence of the measurements). In other words, when the spread in measurements of $\log|J|$ is large, many data are required to achieve precise estimates of parameters. In this paper, we were able

to detect effects with magnitudes smaller than the spread in samples of $\log|J|$ because we had approximately 10^3 data for each compound. It was these large samples, and the precise confidence intervals they afforded, that allowed us to demonstrate the odd-even effect and assign its origin primarily to J_0, as opposed to β. Analyses that seek larger effects, or operate on data with smaller dispersions, than what we have done here, may require much fewer than 10^3 measurements for each compound. As a rule of thumb, however, we suggest collecting at least ~ 200 data per compound, in a manner that minimizes correlation between measurements.

One advantage of using Ag^{TS}-SR//Ga_2O_3/EGaIn junctions is that they are able to generate data quickly and conveniently, to enable precise statistical analysis. For example, the minimum requirement of ~200 data can be fulfilled for one compound in less than one day. The two primary factors that enable this rapid collection of data with Ga_2O_3/EGaIn are i) the relatively high yield (~ 80%) of non-shorting junctions, and ii) the ability to measure in air, under ambient conditions. For these reasons, Ga_2O_3/EGaIn electrodes are an improvement over electrodes based on hanging drops of Hg, and have an advantage over direct evaporation of metal electrodes onto SAMs. Conductive polymers can be used to form contacts with SAMs in high yields (> 90%)[xxvii] and with many junctions in parallel, for rapid measurements.

The updated values of β and J_0 in this paper are precise. Using Method 4a, we have improved our estimates of β and J_0 over those in our previous paper comparing odd and even alkanethiols.[xiv] At an applied bias of V = -0.5V, these updated values (Table 1) are $\beta_{odd} = 1.033 \pm 0.021$ n_C^{-1} (0.819 ± 0.017 Å^{-1}), $\beta_{even} = 1.000 \pm 0.015$ n_C^{-1} (0.792 ± 0.012 Å^{-1}), $\log|J_{0,odd}/(A/cm^2)| = 1.96 \pm 0.12$, and

$\log|J_{0,even}/(A/cm^2)| = 2.53 \pm 0.09$. The uncertainties in these values denote the 99% confidence intervals, whose validity is contingent on how well we have corrected for correlation between values of $\log|J|$ within each sample. It is rare to find uncertainties reported for values of β and J_0 in the literature, but we are confident that our values are among the most precise to date. Elsewhere,[xli] we have conducted a meta-analysis of values of β and J_0 reported in the literature, and identified a consensus for β in the range of $1.0 – 1.1\ n_C^{-1}$ ($0.8 – 0.9$ Å^{-1}; the range is approximate) from among many different systems for measuring charge transport through large-area, SAM-based junctions. (Reports of β and J_0 in the literature do not differentiate between odd and even alkanethiols). Our values agree with this consensus (although they lie towards the low end of the range), and we are, therefore, confident that our values are not only precise, but also accurate. For J_0, we could not identify a consensus across all experimental systems, but there was a loose agreement among several techniques in the range of $J_0 = 10 – 10^3$ A/cm^2 ($\log|J_0| = 1 – 3$). Our values of J_0 lie within this range.

J_0 explains the majority of the odd-even effect. The precision of method 4a allows us to conclude, with a level of confidence that is high ($p = 1.34 \times 10^{-4}$) but tempered by the potential effects of correlation between measurements, that there is a difference in J_0 between odd and even n-alkanethiols. With respect to β, we cannot claim a difference between odd and even n-alkanethiols ($p = 0.20$). We have shown that the difference in $\log|J_0|$ between odd and even n-alkanethiols ($\log|J_{0,even}| - \log|J_{0,odd}| = 0.57$) was large enough to explain approximately 70 – 80% of the magnitude of the odd-even effect. As β is the only other parameter in eq. **1** besides J_0, a difference in β between odd and even n-alkanethiols is currently the strongest candidate to explain the remaining 20 – 30% of the

odd-even effect, but we emphasize caution, since we cannot conclude that the difference in β is even significant, let alone large. The unexplained portion of the odd-even effect may simply be the result of uncertainties in β and J_0 (determining J_0 does, after all, require a long extrapolation), or deficiencies in the Simmons model.

The value of J_0 represents the contributions to charge transport of the electrodes and the interfaces between the electrodes and the SAM. Since the electrodes and (presumably) the Ag^{TS}-S interface are identical for both odd and even alkanethiols, the observation of a significant and large difference in J_0 implies that odd and even alkanethiols form different van der Waals interfaces with the Ga_2O_3/EGaIn electrode.[xlix] Indeed, because of the small tilt angle (the angle of the *trans*-extended alkyl chain with respect to the surface normal) of SAMs of n-alkanethiols on Ag (~ 12°),[i] the terminal $-CH_3$ group at the surface of a (*trans*-extended) SAM should have a different orientation, depending on whether the number of carbon atoms in the alkyl chain is odd or even. For odd n-alkanethiols, the terminal C–C bond is expected to be roughly perpendicular to the surface, whereas for even n-alkanethiols, the terminal C–C bond should be approximately parallel to the plane of the surface.[xlix] This subtle difference is apparently large enough to affect the wavefunction of charges tunneling through the Ag^{TS}-SR//Ga_2O_3/EGaIn junction—and careful statistical analysis is powerful enough to distinguish this effect from the myriad other variables (defects) that affect charge transport. The fact that such a small change in the surface of the SAM has a noticeable effect on charge transport is a testament to the (by now, well-known) importance of interfaces in molecular electronics.

Future experiments may settle the question of whether there is a significant difference in β between odd and even n-alkanethiols, and, if so, how large it is. A significant

difference in β would indicate that the tunneling barrier posed by a SAM of odd n-alkanethiols differs (in shape or height) from that posed by a SAM of even n-alkanethiols. The most we can conclude, at the moment, is that any difference between odd and even n-alkanethiols with respect to the tunneling barrier (β) has less of an influence on charge transport than the difference with respect to the van der Waals interface. We currently have no explanation for why a terminal C–C bond parallel to the surface would be more favorable to charge transport than a terminal C–C bond perpendicular to the surface, but we identify this problem for theoretical study.

Acknowledgements

This research was supported by the National Science Foundation under Award # CHE-05180055 (measurements of charge transport and support for W.F.R.) and the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering under Award # DE-FG02-OOER45852 (support for M.M.T. and J.R.B.). C.A.N. acknowledges the Netherlands Organization for Scientific Research (NWO) for the Rubicon grant supporting this research and the Singapore National Research Foundation under NRF Award No. NRF-RF2010-03.

References

[i] Love, J. C.; Estroff, L. A.; Kriebel, J. K.; Nuzzo, R. G.; Whitesides, G. M. *Chem. Rev.* **2005**, *105*, 1103.

[ii] Cui, X. D.; Primak, A.; Zarate, X.; Tomfohr, J.; Sankey, O. F.; Moore, A. L.; Moore, T. A.; Gust, D.; Harris, G.; Lindsay, S. M. *Science* **2001**, *294*, 571.

[iii] Luo, L.; Frisbie, C. D. *J. Am. Chem. Soc.* **2010**, *132*, 8854.

[iv] Kim, B.; Beebe, J. M.; Jun, Y.; Zhu, X. Y.; Frisbie, C. D. *J. Am. Chem. Soc.* **2006**, *128*, 4970.

[v] Wang, G.; Kim, T.-W.; Jang, Y. H.; Lee, T. *J. Phys. Chem. C* **2008**, *112*, 13010.

[vi] Bumm, L. A.; Arnold, J. J.; Cygan, M. T.; Dunbar, T. D.; Burgin, T. P.; Jones, L., II; Allara, D. L.; Tour, J. M.; Weiss, P. S. *Science* **1996**, *271*, 1705.

[vii] Venkataraman, B.; Breen, J. J.; Flynn, G. W. *J. Phys. Chem.* **1995**, *99*, 6608.

[viii] Venkataraman, L.; Klare, J. E.; Nuckolls, C.; Hybertsen, M. S.; Steigerwald, M. L. *Nature* **2006**, *442*, 7105.

[ix] Lee, T.; Wang, W.; Klemic, J. F.; Zhang, J. J.; Su, J.; Reed, M. A. *J. Phys. Chem. B* **2004**, *108*, 8742.

[x] Kim, T-W.; Wang, G.; Lee, H.; Lee, T. *Nanotechnology* **2007**, *18*, 315204.

[xi] Weiss, E. A.; Kaufman, G. K.; Kriebel, J. K.; Li, Z.; Schalek, R.; Whitesides, G. M. *Langmuir* **2007**, *23*, 9686.

[xii] Chiechi, R. C.; Weiss, E. A.; Dickey, M. D.; Whitesides, G. M. *Angew. Chem. Int. Ed.* **2008**, *47*, 142.

[xiii] Nijhuis, C. A.; Reus, W. F.; Whitesides, G. M. *J. Am. Chem. Soc.* **2009**, *131*, 17814.

[xiv] Thuo, M. M.; Reus, W. F.; Nijhuis, C. A.; Barber, J. R.; Kim, C.; Schulz, M. D.; Whitesides, G. M. *J. Am. Chem. Soc.* **2011**, *133*, 2962.

[xv] Nijhuis, C. A.; Reus, W. F.; Barber, J. R.; Dickey, M. D.; Whitesides, G. M. *Nano Lett.* **2010**, *10*, 3611.

[xvi] Casella, G.; Berger, R. L. *Statistical Inference*; Duxbury Press: Belmont, California, 1990.

[xvii] Bevington, P. R.; Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences*, 3rd ed.; McGraw-Hill: New York, 2003.

[xviii] *Statistical Methods for Physical Science;* Stanford, J. L.; Vardeman, S. B., Eds.; Methods of Experimental Physics Series 28; Academic Press: San Diego, CA, 1994.

[xix] Simmons, J. G. *J. Appl. Phys.* **1963**, *34*, 2581.

[xx] *Robust Statistics;* Huber, P. J.; Wiley Series in Probability and Mathematical Statistics; Wiley and Sons: New York, 1981.

[xxi] *Robust Regression and Outlier Detection;* Rousseeuw, P. J.; Leroy, A. M.; Wiley Series in Probability and Mathematical Statistics; Wiley and Sons: New York, 1987.

[xxii] Weiss, E. A.; Chiechi, R. C.; Kaufman, G. K.; Kriebel, J. K.; Li, Z.; Duati, M.; Rampi, M. A.; Whitesides, G. M. *J. Am. Chem. Soc.* **2007**, *129*, 4336.

[xxiii] For the purposes of building the model, this assumption is equivalent to assuming that the product, βd, is normally distributed. Assuming that βd is normally distributed may be physically correct, but d is likely to vary to a much greater extent than β, since β depends (to a zeroeth-order approximation) on the (constant) electronic structure of the molecules in the SAM, while d depends on the (variable) density and type of defects in the components of the junction. We, therefore, proceed under the assumption that β is a constant, and d is normally distributed.

[xxiv] The difference between a long tail and an outlier is subjective and somewhat arbitrary. The main difference between the two terms is that long tails are joined to the main peak of the histogram (as a tail is joined to a body) by a region of continuous data, while outliers are separated from the main peak of the histogram by "white space", or regions without any data. Also, outliers tend to be fewer in number than the data in a long tail. In any case, the distinction has very little bearing on our analysis. Because both long tails and outliers have a similar effect on the statistical procedures we use, we can treat them collectively as "deviations of $\log|J|$ from normality".

[xxv] Chen, J.; Reed, M. A.; Rawlett, A. M.; Jones, L., II; Tour, J. M. *Science* **1999**, *286*, 1550.

[xxvi] (a) Fisher, G. L.; Walker, A. V.; Hooper, A. E.; Tighe, T. B.; Bahnck, K. B.; Skriba, H. T.; Reinard, M. D.; Haynie, B. C.; Opila, R. L.; Winograd, N.; Allara, D. L. *J. Am. Chem. Soc.* **2002**, *124*, 5528. (b) Walker, A. V.; Tighe, T. B.; Cabarcos, O. M.; Reinard, M. D.; Haynie, B. C.; Uppili, S.; Winograd, N.; Allara, D. L. *J. Am. Chem. Soc.* **2004**, *126*, 3954.

[xxvii] (a) Akkerman, H. B.; de Boer, B. *J. Phys.: Condens. Matter* **2008**, *20*, 013001. (b) Akkerman, H. B.; Blom, P. W. M.; de Leeuw, D. M.; de Boer, B. *Nature* **2006**, *440*, 69.

[xxviii] Milani, F.; Grave, C.; Ferri, V.; Samori, P.; Rampi, M. A. *ChemPhysChem* **2007**, *8*, 515.

[xxix] (a) Haag, R.; Rampi, M. A.; Holmlin, R. E.; Whitesides, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 7895. (b) Holmlin, R. E.; Haag, R.; Chabinyc, M. L.; Ismagilov, R. F.; Cohen, A. E.; Terfort, A.; Rampi, M. A.; Whitesides, G. M. *J. Am. Chem. Soc.* **2001**, *123*, 5075.

[xxx] Slowinski, K.; Fong, H. K. Y.; Majda, M. *J. Am. Chem. Soc.* **1999**, *121*, 7257.

[xxxi] (a) York, R. L.; Nguyen, P. T.; Slowinski, K. *J. Am. Chem. Soc.* **2003**, *125*, 5948. (b) York, R. L.; Slowinski, K. *J. Electroanal. Chem.* **2003**, *550*, 327.

[xxxii] Rampi, M. A.; Whitesides, G. M. *Chem. Phys.* **2002**, *281*, 373.

[xxxiii] Salomon, A.; Boecking, T.; Chan, C. K.; Amy, F.; Girshevitz, O.; Cahen, D.; Kahn, A. *Phys. Rev. Lett.* **2005**, *95*, 266807.

[xxxiv] Salomon, A.; Boecking, T.; Seitz, O.; Markus, T.; Amy, F.; Chan, C.; Zhao, W.; Cahen, D.; Kahn, A. *Adv. Mater.* **2007**, *19*, 445.

[xxxv] Salomon, A.; Böcking, T.; Gooding, J. J.; Cahen, D. *Nano Lett.* **2006**, *6*, 2873.

[xxxvi] Dickey, M. D.; Chiechi, R. C.; Larsen, R. J.; Weiss, E. A.; Weitz, D. A.; Whitesides, G. M. *Adv. Funct. Mater.* **2008**, *18*, 1097.

[xxxvii] Cademartiri, L.; Chiechi, R. C.; Thuo, M. M.; Nijhuis, C. A.; Barber, J. R.; Sodhi, R. N. S.; Brodersen, P.; Kim, C.; Reus, W. F.; Whitesides, G. M. unpublished experiments.

[xxxviii] Barr, T. L.; Seal, S. *J. Vac. Sci. Technol. A* **1995**, *13*, 1239.

[xxxix] The concept of resistivity is not well-defined for components with non-linear $J(V)$ characteristics (such as SAMs). For comparing a SAM to the Ga_2O_3 film, an analog of the resistivity of the SAM at a particular bias (V) can, however, be estimated, given J through the SAM and the thickness of the SAM (d), using the equation $\rho = V/(Jd)$.

[xl] Reus, W. F.; Thuo, M. M.; Shapiro, N. S.; Nijhuis, C. A.; Whitesides, G. M. unpublished experiments.

[xli] Nijhuis, C. A.; Reus, W. F.; Barber, J. R.; Whitesides, G. M. manuscript submitted to *J. Am. Chem. Soc.*

[xlii] Our statistical model would also be valid under the alternative assumption that the product βd is normally distributed and J_0 is approximately constant, for then $\log|J|$ would still be normally distributed. Even if J_0 is not constant, but log-normally distributed (perhaps due to variations in the width of the tunneling barrier defined by the van der Waals interface), $\log|J|$ would still be the sum of two normal distributions, which is itself a normal distribution. There are, thus, many scenarios in which $\log|J|$ is normally distributed.

[xliii] If a sample has an odd number of elements, then the median is the value that lies in the middle, when the elements of the sample are arranged in increasing order. If a sample has an even number of elements, then the median is the average of the two middle values in the ordered list of elements.

[xliv] Assuming that, for example, the Gaussian mean of $\log|J|$ is normally distributed is not the same as assuming that $\log|J|$ itself is normally distributed; rather, this assumption implies that if many samples of $\log|J|$, each comprising N measurements, were taken from the same population, then the Gaussian means of all samples, taken together, would collectively conform to a normal distribution.

[xlv] (a) Dale, M. R. T.; Fortin, M-J. *J. Agric. Biol. Envir. S* **2009**, *14*, 188. (b) Cressie, N. A.; *Statistics for Spatial Data*, New York: Wiley, 1991.

[xlvi] Conover, W.J. *Practical Nonparametric Statistics;* John Wiley and Sons: New York, 1980.

[xlvii] Galwey, N. W. *Genet. Epidemiol.* **2009**, *33*, 559.

[xlviii] Fracasso, D.; Valkenier, H.; Hummelen, J. C.; Solomon, G. C.; Chiechi, R. C. *J. Am. Chem. Soc.* **2011**, *133*, 9556.

[xlix] Tao, F.; Bernasek, S. L. *Chem. Rev.* **2007**, *107*, 1408.