An analysis of exciton transport, electron tunneling, and electromigration in nanotube and nanowire systems

by

Darin O. Bellisario

B.S., Chemical Engineering, 2009
Tufts University

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY
IN CHEMISTRY

at the

Massachusetts Institute of Technology

February 2016

Signature of Author: Signature redacted
Department of Chemistry
December 14th, 2015

Certified by: Signature redacted
Michael S. Strano
Thesis Supervisor

Accepted by: Signature redacted
Robert W. Field
Chairman, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a Committee of the Department of Chemistry that included,

Professor Troy Van Voorhis
Chair Signature redacted

Professor Keith A. Nelson Signature redacted

Professor Michael S. Strano
Thesis Supervisor Signature redacted

AN ANALYSIS OF EXCITON TRANSPORT, ELECTRON TUNNELING, AND ELECTROMIGRATION IN NANOTUBE AND NANOWIRE SYSTEMS

by

Darin O. Bellisario

Submitted to the Department of Chemistry
on January 15, 2016 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Science in
Chemistry

ABSTRACT

Herein three systems – electromigration in metal nanowires, electron tunneling in single-molecules, and carbon nanotube photovoltaics – are investigated. In the first area, electromigrative failure of metal nanowires has been shown to form single-molecule tunnel junctions, but the process has remained unpredictable, limiting the yield of devices under current methods. Electromigration in micron diameter and larger wires is well understood as the migration of vacancies in the bulk crystal, but both the quantitative predictions and qualitative features of that mechanism break down at the nanometer scale. We propose, and validate against experimental data, that as the wire diameter falls below a micron, the increased surface-area-to-volume ratio and the low barrier to surface atom translation shift the dominant mechanism of electromigration from bulk transport to surface transport. We then apply the model to design a process controller to guide gradual electromigration. We then turn to investigating the tunnel junctions themselves. Diverse physical insights have been gained from electron tunneling measurements of single molecules, but to date all observations have been static *i.e.* subject to long integration times. We performed temporally resolved measurements, revealing underlying molecule dynamics. In particular we find that molecules can stochastically switch between discrete inelastic transport states, suggesting discretized molecule reconfiguration consistent with the body of literature from Scanning Tunneling Microscopy. Finally, we investigate carbon nanotube (CNT) network solar cells. The large parametric space associated with the nanometer-scale heterogeneous material, including the mixture of nanotube length, chirality, orientation, *etc.*, has prevented proof-of-concept devices from revealing a research pathway to practical efficiencies. To address this empirical limitation, we derived a model of CNT photovoltaic steady-state operation from the light absorption and exciton transport behaviors of single and aggregate nanotubes. To do so, we treated single nanotube properties as random variables, describing the nanotube network as distributions of those properties. Applying the model to different solar cell architectures, we predict that efficiencies will be dramatically higher in high density films of vertically-aligned nanotubes. We also show that the film thickness must be at an optimum, and that as a rule of thumb the film thickness should be approximately the exciton diffusion length.

Thesis Supervisor: Michel S. Strano
Title: Carbon P. Dubbs Professor of Chemical Engineering

ACKNOWLEDGEMENTS

The author would like to express his feelings of gratitude towards the following people, each of whom was instrumental to the completion of this research. Foremost among them is his advisor Prof. Michael S. Strano, who holds at least as much claim to this work as Mr. Bellisario. In addition to his scientific contributions to the daily management of research, Prof. Strano brought an openness to risky ideas, optimism, and energy that in combination continuously demanded more from our research, challenging us to hold lofty aspirations. He taught Darin, among other things, one of the two most important lessons of his Ph.D. tenure: do not do things just for the sake of doing them, but rather for the impact they have, on the scientific community and on society. Steven Shimizu, Joel Anthony Paulson, and Zackary Ulissi each aided crucial experiments and analysis. The precocious undergraduate researchers Rebecca Han, Jasmine Harris, and Brandon Zabala used their learning experiences to also substantially advance our research. Qing Hua Wang and Kumar Varoon Agrawal provided crucial insight guiding and clarifying the work, and raised the author's spirits through the inevitable periods of experimental frustration. Mark Mondol at MIT's Research Laboratory for Electronics' (RLE) Scanning Electron Beam Lithography (SEBL) facility aided in the design of EBL fabrication processes. The author's friends, especially Geraldine L.C. Paulus, Scott Cordeiro, Rishabh M. Jain, Markita Landry, Ben Singer, John Barbieri, Shannon Morey, and Liz Mundy; his girlfriend Sarah G. Miller; and his family, Sebastian, Diane, Craig, Afarin, and Lee Bellisario; all gave him a necessary sense of value and fulfillment in the most trying periods of this work. Without the boosts to confidence and energy they provided, the author might not have persevered. Finally the author would like to credit his former advisor, Prof. E. Charles H. Sykes, who gave him the tools and confidence to set out on this endeavor in the first place.

Several shared facilities provided the necessary tools to perform much of the device fabrication discussed in this research. These include Harvard's Center for Nanoscale Studies (CNS), MIT's Institute for Soldier Nanotechnologies (ISN), and the aforementioned Research Laboratory for Electronics' (RLE) Scanning Electron Beam Lithography (SEBL) facility. The equipment at these shared facilities would have been prohibitively expensive to invest in to conduct the research performed in this thesis; without the availability of these centers, we would never have engaged in this work to begin with. As a result, the author would like to express his opinion that communal research equipment dramatically increases the efficiency of research, enabling otherwise impossible projects.

As a final note the author would like to express gratitude to the late Prof. Robert 'Bob' Silbey. Darin was among the most minor characters in the Professor's life, but *vice versa* Prof. Silbey by his example played an outsized role in Darin's. Prof. Silbey exhibited the rare and estimable quality of being unassuming, unpresumptuous. He was supportive not by directly involving himself, but by making you feel unjudged, unpressured. In Prof. Silbey Darin found a new, positive model for engaging with others. By giving everyone the benefit of the doubt, by listening to them carefully, you can learn from anyone. Everyone has something to offer. And when you listen to people earnestly, not out of politeness or mock respect, you will nearly always earn their trust and have your respect for them reciprocated.

OUTLINE & LIST OF PUBLICATIONS

This thesis is divided into six chapters reviewing three primary areas of investigation:

Nanometer-scale electromigration

I. Darin O. Bellisario, Zackary Ulissi, Michael S. Strano, 'A Quantitative and Predictive Model of Electromigration-Induced Failure of Metal Nanowires.' *J. Phys. Chem. C* 117, 12373−12378 2013
II. Darin O. Bellisario, Rebecca Han, Jasmine K. Harris, Robert Brandon Zabala, Qing Hua Wang, Kumar V. Agrawal, Youngwoo Son, Michael S. Strano. 'Monitoring Electromigration Kinetics for Continuous Control'. *In Preparation*
III. Darin O. Bellisario, Steven Shimizu, Rebecca Han, Qing Hua Wang, Michael S. Strano, 'Nanometer-scale Electromigrative Failure with a Metal-Covalent Interface of a Au Nanowire and Carbon Nanotube.' *In Preparation.*

Single-molecule electron tunneling spectroscopy

IV. Darin O. Bellisario, Rebecca Han, Jasmine K. Harris, Robert Brandon Zabala, Qing Hua Wang, Kumar V. Agrawal, Youngwoo Son, Michael S. Strano. 'Electrical Observation of Molecule Dynamics in a Single-Molecule Tunnel Junction'. *In Preparation.*

Carbon nanotube photovoltaics

V. Darin O. Bellisario, Rishabh M. Jain, Zackary Ulissi, Michael S. Strano, 'Deterministic Modelling of Carbon Nanotube Near-Infrared Solar Cells.' *Energy & Environmental Science* 7, 3769-3781 2014
VI. Darin O. Bellisario, Joel A. Paulson, Richard D. Braatz, Michael S. Strano, 'Single-Parameter Determination of Carbon Nanotube Photovoltaic Efficiency.' *In Preparation.*

In addition to the lead-author research listed above, the author has made secondary contributions to the following investigations in the course of this thesis research. This work is not surveyed in this document.

Carbon Nanotube Nanopores:

1. W Choi, ZW Ulissi, S Shimizu, DO Bellisario, MD Ellison, MS Strano. 'Diameter Dependent Ion Transport through the Interior of Isolated Single Walled Carbon Nanotubes.' *Nature Comm.* 4, 2397 2013
2. S Shimizu, M Ellison, K Aziz, QH Wang, Z Ulissi, Z Gunther, DO Bellisario, MS Strano. 'Stochastic Pore Blocking and Gating in PDMS-Glass Nanopores from Vapor–Liquid Phase Transitions.' *J. Phys. Chem. C* 2013

Carbon Nanotube Chemistry:

3. AJ Hilmer, DO Bellisario, S Shimizu, TP McNicholas, QH Wang, SA Speakman, MS Strano. 'Formation of High-Aspect-Ratio Helical Nanorods via Chiral Self-Assembly of Fullerodendrimers.' *J. Phys. Chem. Lett.* 2014

Carbon Nanotube Sensors:

4. NF Reuel, *et al.* 'Emergent Properties of Nanosensor Arrays: Applications to Monitoring IgG Affinity Distributions, Weakly-Affined Hypermannosylation, and Colony Selection for Biomanufacturing.' *ACS Nano* 2013
5. QH Wang, DO Bellisario, LW Drahushuk, RM Jain, S Kruss, MP Landry, SG Mahajan, SFE Shimizu, ZW Ulissi, MS Strano, 'Low Dimensional Carbon Materials for Applications in Mass and Energy Transport.' *Chemistry of Materials* 2014

Carbon Nanotube Separation:

6. TP McNicholas, V Cantu, AJ Hilmer, K Tvrdy, RM Jain, R Han, DO Bellisario, J Ahn, PW Barone,

B Mu, MS Strano. 'Magnetoadsorptive Particles Enabling the Centrifugation-Free, Preparative Scale Separation and Sorting of Single Walled Carbon Nanotubes' *Advanced Materials* 2013

Molybdenum Oxide Transistors:

7. MMYA Alsaif, S Balendhran, DO Bellisario, Y Son, AF Chrimes, T Daeneke, MR Field, W Zhang, H Nili, EP Nguyen, K Latham, J van Embden, S Sriram, M Bhaskaran, MS Strano, JZ Ou, K Kalantar-zadeh. 'Tunable Field Effect Transistors of High Free carrier Mobilities based on Two-dimensional Molybdenum Oxide Nanoflakes' *ACS Nano* 2015

CONTACT

Please direct inquiries to either the author at *dob@alum.mit.edu* or Prof. Michael S. Strano, the Principal Investigator on all of these studies, at *strano@mit.edu*. On request we are happy to provide any further details of experimentation or results, raw data, or computer code.

TABLE OF CONTENTS

Overview

Complete

INTRODUCTION

In the sections below, we survey some of the relevant concepts and literature necessary to understand and contextualize the investigations in the subsequent Chapters. In-depth exploration is outside the scope of this document, but where relevant references are provided at the beginning of each section for readers that would like a deeper understanding.

Metal Electromigration

For insight into metal electromigration beyond the content in this Introduction, please refer to the following publications:

1. Paul S. Ho and Thomas Kwok. *Electromigration in Metals*. Rep. Prog. Phys. 52, 301-349, 1989.
2. James R. Black. *Electromigration – A Brief Survey and Some Recent Results*. IEEE Transactions on Electron Devices 16, 4 338-347, 1969.
3. Donald G. Pierce and Phillip G. Brusius. *Electromigration: A Review*. Microelectron. Reliab. 37, 1053-1072, 1997.

The basic mechanism

When a voltage bias is applied across a metal crystal, lattice atoms feel two forces that can inspire translation. The first force is the direct electrostatic, or Coulomb, force driven by the attraction of the partially-shielded nuclei to the negative electrode and repulsion from the positive electrode,

$$\boldsymbol{F_d} = qZ_d\boldsymbol{E} \tag{1}$$

where q is the elementary charge, Z_d is the effective valence of the nuclear ion,[i] and $\boldsymbol{E}$ is the electric field vector. While this force is intuitive, it has yet to be unambiguously observed and its existence has frequently been called into question.[2,3] In particular, beginning with *Bosvieux and Friedel*[4] the extent of charge screening, and hence coulomb force cancellation, has been under investigation (see *e.g. Lodder et al.*[1]). Regardless, outside of exotic chemistries and architectures[5] the direct force is generally considered negligible.[2,3]

In bulk metals the direct force is substantially weaker than the second force imposed on lattice nuclei in an electric field, the so-called 'electron wind' force.[2,3] Electron scattering *via* interaction with crystal lattice phonons imparts momentum to the lattice nuclei. A net current flux therefore creates a net force on metal atoms in the direction of electron flow,

$$\boldsymbol{F_w} = qZ_w\boldsymbol{E} \tag{2}$$

---

[i] Z_d is the effective charge expressed by the nucleus on the length scale of the electric field, *i.e.* the valence minus screening effects.[1]

where Z_w is the effective valence associated with the wind force. Z_w is therefore some function of the current flux, and its particular form is generally the focus of theoretical investigations into electromigration such as by *Rous et al.*,[3] *Dekker et al.*,[6] and *Bly et al.*[7]

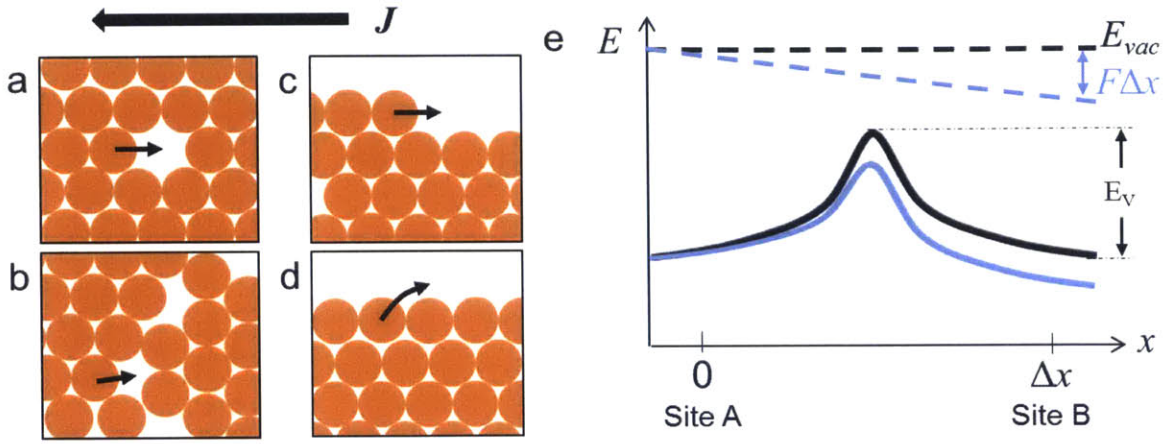


Figure 1. (a,b,c,d) Cartoon illustrations of metal atom displacement due to electromigration in bulk (a,b) and on the surface (b,c) under a current flux ***J***. Barriers to nuclear translation may be low enough to exhibit electromigration at vacancies (a), grain boundaries (b), and surface sites (c,d). (e) Example potential energy diagram for atom displacement mediated by a vacancy, as in (a). Translation requires overcoming an energy barrier E_V.

When a voltage and resulting current are applied across a metal wire, these forces can inspire the dislocation of atoms to adjoining sites when the energy barrier to nuclear displacement is low, such as when the atom is adjacent to a lattice vacancy or is on the crystal surface (Figure 1). While such individual energy-activated lattice site hopping events are stochastic, averaging over the crystal of interest[ii] and over minimum accessible timescales (for electrical measurements, that is microseconds at fastest) yields a continuous flow of metal atoms called electromigration. Increasing the current flux, such as by increasing the applied voltage bias, increases the rate of mass transport. Electromigration is most rapid in areas where the energy barrier to atom/vacancy translation is lowest, such as grain boundaries and lattice defects, or where the current flux is highest, such as at constrictions and sharp edges.

[ii] In the Introduction, we will focus on applying a bias across wires of micron and larger effective diameter, according to the state of the field up to the work presented in this thesis. In the thesis chapters, and in discussing electromigrative tunnel junction formation in later Introduction sections, we will consider wires down to tens of nanometers in effective diameter. In all such cases, electromigration remains a continuous process. In Chapter I and Chapter II, we observe and discuss measurable *stochastic* electromigration when wire effective diameters fall below approximately 10 nm, as similarly observed by *Strachan et al.[8]*

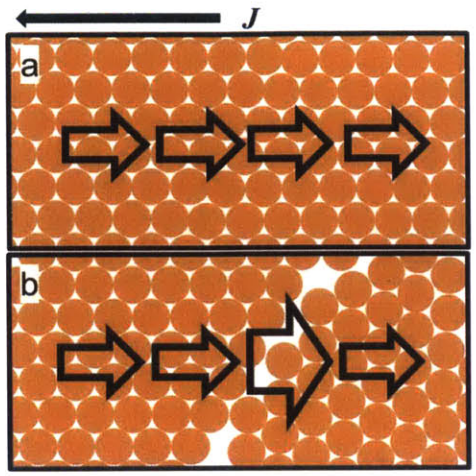


Figure 2. Cartoon illustration of the role of heterogeneities in electromigrative evolution of metal crystals. J is the current flux, of positive charge carriers by convention (opposite polarity of the electron flux vector). Orange circles represent atoms in a crystal cross-section, and the width of the hollow black arrows represent the magnitude of the atom flux. (a) a perfect crystal; the nuclear flux is uniform. As a result, there is no net loss of mass at any point. (b) a grain boundary; the nuclear flux increases at the grain boundary. As a result, more atoms will leave the grain boundary per time than are replenished. There is thus a net loss of mass from that point over time.

In a perfect crystal this process yields no macroscopic change in morphology because the net flux of atoms in any cross-section is zero; the same number of atoms enter as leave. Heterogeneities such as grain boundaries, temperature gradients, and material interfaces however create a gradient in atom flux from the gradient in vacancy concentration, mean translocation barrier, or current density. That gradient causes regions of the crystal to experience a net gain or loss of atoms with time (Figure 2). The process leads to the formation of extrusions at some points on a wire and voids at others, until eventually any finite-diameter wire will break.

The Black equation and electromigration at micron scales

At least in wires of micron diameter and larger, electromigration has been well understood since 1967 when Black first proposed that the rate of electromigration was dominated by the drift of vacancies in the bulk crystal.[2,9,10] From the vacancy transport mechanism he derived and empirically verified the celebrated Black equation of metal wire time-to-failure t_f due to electromigration under an initial current flux j,[9,11]

$$t_f = Ae^{\frac{E_V}{kT}}j^{-2} \quad (3)$$

where A is an empirical prefactor, E_V is the activation barrier to vacancy diffusion, k is the Boltzmann constant, and T is temperature. In the intervening years the bulk vacancy transport model and resulting Black equation have been verified across numerous experimental platforms, simulations, and theoretical treatments.[1,2,10,12-15] This bulk vacancy transport mechanism of electromigration has proven invaluable in the semiconductor industry, where Black's predicted time-to-failure limits the passage of current in metal wires and interconnects.

In addition to the quantitative treatment of electromigration, several qualitative phenomena are ubiquitously observed. One is void formation. As vacancy concentrations rise they aggregate to form macroscopic voids in the crystal. This effect is discussed further in Chapter I. Another is failure at

material interfaces. As explained above, the gradients in metal atom flux are what cause morphological changes and ultimately wire failure. These gradients arise from any heterogeneity in the material, including temperature gradients and grain boundaries, but among them interfaces between two materials, even of the same chemistry, dominate the rate of electromigration. Across diverse architectures and chemistries, failure occurs at the interface of materials.[16-21] The explanation for this phenomenon is that a material interface even between two identical compounds acts as a grain boundary that spans the entire width of the crystal. Grain boundaries produce the highest local material flux amongst heterogeneities, but grains are smaller than the crystal diameter and as such do not span its entire width.[21] This effect is discussed further in Chapter III.

Breakdown of the Black equation: electromigration at the nanometer scale

Despite the broad validity and technological impact of the Black equation and the bulk vacancy transport mechanism of electromigration, systematic deviations from the model began to be observed in the 1980s as miniaturization brought semiconductor metal interconnects to sub-micron dimensions. Deviations from the quadratic scaling of electromigration rate[iii] with current flux were increasingly observed as wires became narrower. Empirical corrections to the current exponent became necessary, with power coefficients in the absence of Joule heating being measured between -1 and -2.[10] These corrections were initially explained as due to material-specific properties, but eventually were attributed to diminishing wire diameter.[10] Prior to the investigations of this thesis, there was no explanation for this breakdown of the bulk vacancy transport mechanism of electromigration. In the work presented in Chapter I and Chapter II, we propose that as metal wire diameters fall to tens or hundreds of nanometers the highest-rate electromigration pathway shifts from bulk vacancy transport to surface transport. The energy barrier to metal atom dislocation on surfaces is known to in general be significantly, by orders of magnitude lower than bulk vacancy dislocation; as the surface area to volume ratio rises, the population of surface sites becomes comparable to the population of bulk vacancies and the rate of electromigration therefore becomes dominated by the faster migration of surface atoms. We validate this hypothesis both quantitatively and qualitatively using experimental data from our own group (Chapter II) and others (Chapter I). This surface transport model yields a linear, rather than Black's quadratic, scaling of mean electromigration rate with current density, explaining and thereby being supported by the empirical corrections to Black's model mentioned above.

Although it has not been explicitly recognized prior to this work, the qualitative features of vacancy-mediated electromigration also break down as wire effective diameters enter the nanometer regime. As we explore in Chapter I, void formation is less commonly observed in nanowires. Instead,

---

[iii] The mean electromigration rate is the inverse of the time to failure.

under electromigration nanowires thin at the point of most rapid migration, typically a grain boundary,[22] 'necking' gradually until the wire breaks. This change is consistent with mass transport, and hence mass loss, from the surface rather than the bulk. In Chapter III, we show that the phenomenon of ubiquitous failure at material interfaces also breaks down at the nanometer scale, with electromigration through a Au nanowire-carbon nanotube (CNT) interface yields necking and failure along the nanowire rather than at the connection between the two, despite the substantial gradients in vacancy concentration (the interface is effectively a grain boundary), atom mobility (the covalently bonded carbon atoms do not electromigrate, so Au should be ejected from the interface but not replenished), and temperature (substantial differences in both thermal conductivity and Joule heating between the materials).

Nanowire Electromigration for Tunnel Junction Formation

In the previous section we discussed the current knowledge of metal electromigration, specifically the bulk-vacancy transport mechanism, its validity for micron-diameter and larger wires, and the gradual realization that as wire dimensions shrink to the nanometer scale the model appears to break down. In Chapters I through III, we show that both the quantitative and qualitative features of micron-scale electromigration do indeed diverge, and we provide and validate a new dominant mechanism for metal electromigration in nanometer-diameter wires: surface transport. The value of those findings goes beyond the semiconductor industry however, as the electromigration of nanowires has recently been exploited as a method for a new technology: on-chip single-molecule tunnel junctions. In this section, we discuss that application.

In the first sub-section, we cite the discovery of electromigrative tunnel junctions (ETJs), and in the 5th sub-section, *features of electromigrative formation of tunnel junctions*, we explain its basic features. In the sections *progress to date in electromigrative single-molcule tunnel junctions* and *potential technologies based on on-chip tunnel junctions*, we survey applications of the technology to both illustrate its potential and to contextualize our own exploitation of the platform in Chapter IV. In the fourth sub-section *comparison to alternative techniques*, we compare electromigrative tunnel junctions to alternative techniques. In the final section, *progress to date in the formation of on-chip tunnel junctions*, we provide the context surrounding and motivating our investigations in Chapters II and III into understanding the mechanism of nanowire electromigration and exploiting it to control the formation of electromigrative tunnel junctions.

Discovery

In 1999 *Park et al.* discovered a new technological application of electromigration.[23] They drove electromigration by gradually increasing the bias applied across a nanowire approximately 200 nm in width and 10 nm in thickness (Figure 3). They found that by using such a thin wire and arresting the

voltage as soon as the wire broke at failure, as detected by the sudden rise in resistance, the resulting break was frequently small enough to measure an electron tunneling current across.

Electron tunneling, described in more depth in the section *Electron Tunneling Spectroscopies* below, is the phenomenon of electrons traversing energy barriers, such as the vacuum between two electrodes, due to the exponentially-decaying spatial projection of a wavefunction into a finite barrier; when the barrier to tunneling is low enough and the separation between two electrodes is small enough, typically a few angstroms, a measurable current[iv] can be conducted through a gap, called a 'tunnel gap' or a 'tunnel junction.' Parallel plates with nanometer separation are used to study large-area tunnel junctions,[24,25] but small area, rough electrodes such as the tip of a Scanning Tunneling Microscope or those created here by *Park et al.* (Figure 3b) usually produce a current dominated by a single point of the 2-5 closest-proximity atoms; the exponential decay of the tunneling probability, and hence the tunneling current,[v] mean that even if the second-narrowest constriction between electrodes is only half an angstrom wider than the narrowest constriction, its resulting parallel tunneling current will be negligible. This exponential decay principle is what allows STM tips to be maintained with angstrom precision above a substrate. For example making effectively 'atomically-sharp' STM tips is almost trivial; simple wire cutters can be used to cut a tip out of a metal wire by hand. Often the tunnel gap is not a single metal atom at each electrode however; even in STM 'trimers' are frequent, as imaged in Field Ion Microscopy by *e.g. Kuk et al.* in 1986 and again by *Cross et al.* in 1998.[26,27]

[iv] Depending on the device and apparatus, direct currents down to femtoamps can be measured. In the research presented in this Thesis, our apparatus reliably reaches pico-amp sensitivity, as limited by leakage of current through the insulators surrounding cabling. Periodic (AC) signals are generally more limited in amplitude sensitivity due to parasitic capacitance.

[v] The relationship between the tunneling current and tunneling probability is explored in greater depth in the section *Electron Tunneling Spectroscopies* below. Approximately, the derivative with voltage (electric field) of the tunneling current is proportional to the tunneling probability.

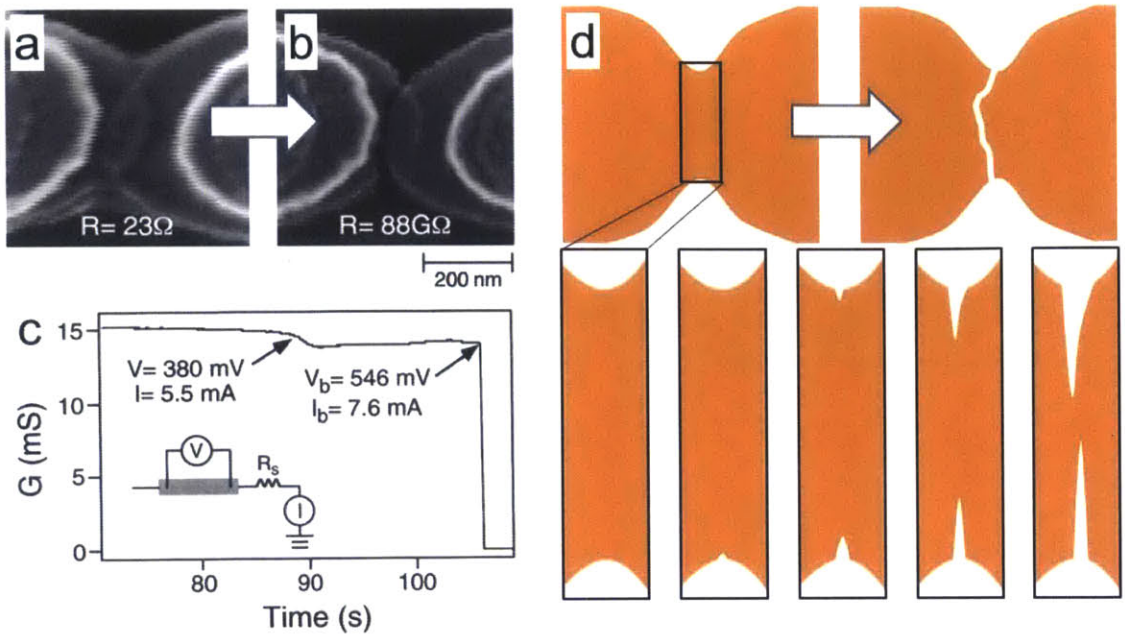


Figure 3. First discovery, by Hongkun Park *et al.*, of electromigrative formation of tunnel junctions, reproduced from reference [23] with permission. (a) Scanning Electron Micrograph (SEM) of the 'nanowire' device prior to electromigration, with approximately 10 nm thick Au in the middle and 90 nm thick Au on the two ends. R is the initial resistance of the device. (b) SEM of the device after electromigration, showing a break has formed between the electrodes. The resistance R suggests that transport between the two electrodes is in the electron tunneling regime. (c) the evolution of wire conductivity observed during the electromigration process. See text for description. (d) cartoon of the wire necking processes during electromigration.

In the section below, *features of electromigrative formation of tunnel junctions*, we describe in greater depth the electromigrative failure of the junction as revealed by research over the past decade. Simple features revealed in the *Park et al.* work however are worth recounting here. During the electromigration process, the status of the nanowire could be monitored indirectly by its resistance R (or equivalently conductivity $G = 1/R$). As electromigration proceeds, the ultimate point (cross-section) of failure along the length of the nanowire gradually loses mass; most later investigations, including our own, show that this occurs by thinning, or 'necking,' of the nanowire, *e.g. Strachan et al.* 2006,[28] 2008,[29] or *Taychatanapat et al.*[22] As the neck thins, the resistance of the wire increases due to the smaller cross-section and, as we show in Chapter I, increased specular scattering. Measuring the current during the voltage application as in Figure 3c, the progress of electromigration can be monitored and when failure occurs, as indicated by a drop in the conductivity to below the conductance quantum (see *features of electromigrative formation of tunnel junctions* below), the applied bias is quickly attenuated.[vi]

[vi] We find, and it is implicit in the publications of other groups, that the resulting tunnel junctions cannot survive arbitrarily high voltages. The sharp electrode 'tips,' a cluster of a few atoms, are not equilibrium configurations and as such relax over time and temperature, as investigated by *Prins et al.*,[30] and with

Other methods of forming a tunnel gap, specifically Scanning Tunneling Microscopy (STM), had existed previously. In the section below, *comparison to alternative techniques*, we survey them. This electromigration method however was the first that could produce an 'on-chip' junction, one which could survive without active feedback from an electro-mechanical apparatus. Below we discuss the advantages of that architectural feature.

The following year, Park *et al.* found that by coating the wire amply in a molecule of interest beforehand, in that case C_{60} (fullerene), the gap would generally form with one or two of the molecule trapped in the interstitial.[32] These single-molecule tunnel junctions (SMTJs) opened up new possibilities for both physical exploration and technological application. In the following two sections, *progress to date in electromigrative single-molecule tunnel junctions* and *potential technologies based on on-chip tunnel junctions*, we survey the research to date realizing those possibilities before returning to the issue of nanowire electromigration in the final two sections, *features of electromigrative formation of tunnel junctions* and *progress to date in the formation of on-chip tunnel junctions*.

Progress to date in electromigrative single-molecule tunnel junctions

Since their discovery by Park *et al.* in 2000,[32] single molecules trapped in on-chip tunnel junctions (SMTJs) have produced diverse physical discoveries.[31-46] From 2003 to 2013, Park *et al.*,[32,33] Kubatkin *et al.*,[34] Perrin *et al.*,[35] Liang *et al.*,[36] Pasupathy *et al.*,[37] and Calvo *et al.*[38] have, among others, shown Coulomb Blockade and the Kondo effect in fullerenes, chelated metal complexes, alkanethiols, and aromatics. Reddy *et al.* in 2007 published the first tunnel junction thermoelectric measurements.[47] Ward *et al.*, also in 2007, exploited the electrode morphology to perform Surface-Enhanced Raman Spectroscopy (SERS).[31,41] In 2008 Grose *et al.* showed spin-selective transport.[43] Song *et al.*[40] in 2009 performed inelastic tunneling spectroscopy, building on 1998 work by Stipe *et al.*[39] doing the same in Scanning Tunneling Microscopy (STM). Winkelmann *et al.* in 2009 discovered superconductivity in SMTJs.[42] In 2010, Ward *et al.* showed optical rectification of the junctions.[44]

---

applied bias. The best tunnel junctions formed in our group are stable enough to survive testing for several hours around 100 mV of applied bias, but most relax within a few seconds or minutes above 10 mV. As a result, it is important to attenuate the applied bias used to form the gap before the gap itself relaxes. Bias-inspired relaxation may be due to local heating, which can reach hundreds of Kelvin (see Chapter IV),[24,31] or further electromigration.

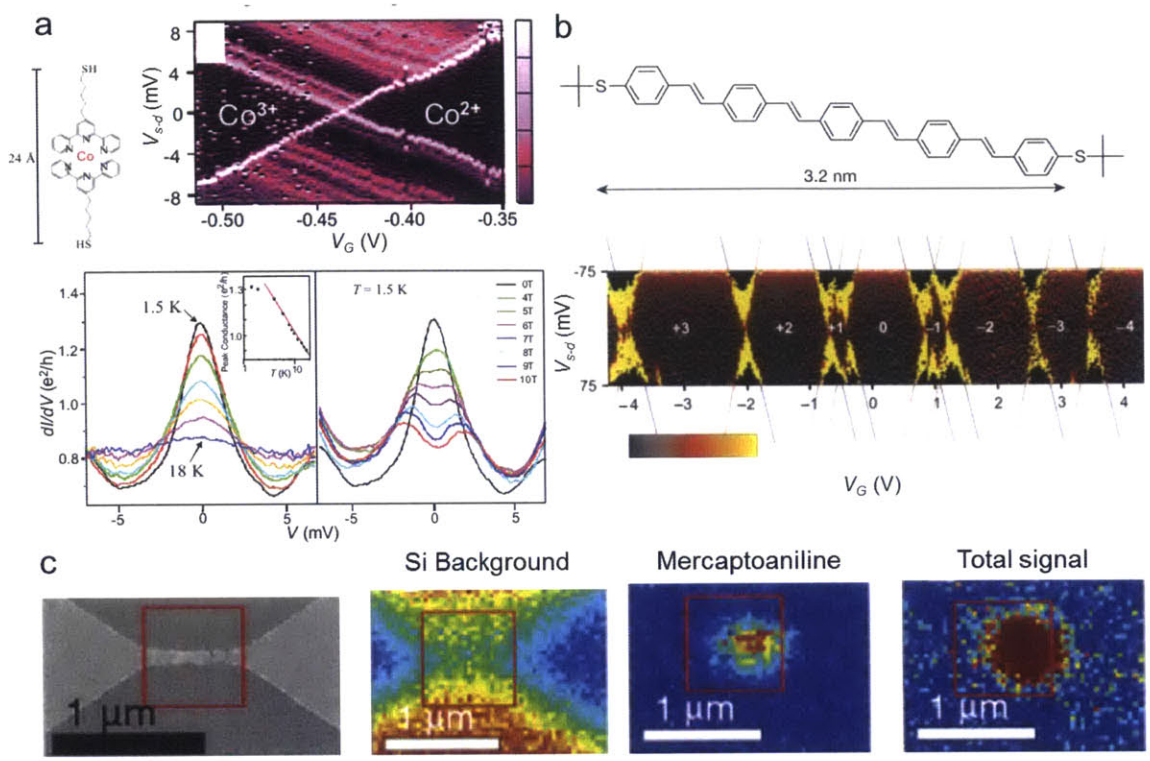


Figure 4. (a) Observation by *Park et al.* of Coulomb Blockade (top) and the Kondo effect (bottom) in a chelated Cobalt compound (cartoon inset).[33] (top) Conductivity (colormap) versus the source-drain bias and gate bias, showing characteristic increases in conductivity when molecular orbitals are accessed between discrete charge states of Co. (bottom) observation of the Kondo effect by observing the differential conductivity near zero bias at varying temperature (left) and electric field (right). (b) Observation of Coulomb Blockade in a polyarmomatic hydrocarbon by *Kubatkin et al.*, showing the ability to access many discrete oxidation states of the molecule. Again the conductivity is the color map plotted against the applied source-drain and gate biases.[34] (c) Surface Enhanced Raman Spectroscopy (SERS) in a single-molecule tunnel junction from *Ward et al.*[41] SEM image (top) shows geometry to interpret Raman mapping (bottom). The total area of Raman peaks of interest are plotted (colormap) as a function of position. The characteristic Si peak show attenuation under the Au wire, while the molecule being studied (mercaptoanline) shows a substantial raman enhancement near the tunnel junction. The total signal is also larger in proximity to the junction. All works reproduced with permission.

To date however all investigations have treated SMTJs, in theory and measurement, as static. DC and AC (harmonic) electronic measurements have been integrated over minutes or hours to maximize the signal, at the cost of any temporal resolution. That is despite extensive evidence from Scanning Tunneling Microscopy (STM) that even at liquid He temperatures the electric field in a tunnel junction can inspire molecule dynamics.[48,49] Beginning with Stipe *et al.*[48] it has been commonly observed that at low temperatures functional groups on the surface studied by STM can flip between adsorption sites, using the energy supplied by the applied bias to overcome the reorientation barrier,[48,49] or that the thiol anchor itself can hop between Au adsorption sites.[50] In addition to the ubiquity of molecular dynamics under applied bias observed in STM, recent work by Ward *et al.*[31] and Ioffe *et al.*,[24] has shown, by measuring the stokes-anti-stokes shift in SERS, that the local temperature in the junction is substantially elevated, reaching over 200 K under a 200 mV bias.

In Chapter IV, we investigate this unexplored time dimension by performing the first time-resolved electrical measurements of single-molecule tunnel junctions. We observe reversible stochastic fluctuations between discrete inelastic transport states that suggest that molecules do indeed exhibit dynamics even at low temperature.

Potential technologies based on on-chip tunnel junctions

In addition to the physical insights gained with them, on-chip tunnel junctions offer unique potential to enable three categories of technology: single-molecule electronics (SMEs),[43,51-54] optoelectronics,[44] and biomolecule assays.[46,55]

In the first category, the molecular orbitals and spin states of single molecules can in theory be exploited as ideal transistors. Ballistic transport reduces heat production and energy requirements. Electron state quantization means an identical readout on every cycle and no hot carriers causing degradation of the dielectric over time. The switching speed of single molecules could be orders of magnitude faster than current transistors.[56] Finally single molecules represent an extreme of miniaturization, reducing the size and energy requirements of electronics; the first fullerene transistor presented by Park *et al.* in 2000[32] is more than 15 times smaller than the smallest commercial transistor today and 5 times smaller than current end of the International Technology Roadmap for Semiconductors in 2021.[vii]

To this end, several groups have demonstrated transistor behavior in single molecule tunnel junctions. The coulomb blockade effect discussed in the previous section is transistor behavior; work including, among others, Park *et al.*,[32,33] Kubatkin *et al.*,[34] Perrin *et al.*,[35] Liang *et al.*,[36] Pasupathy *et al.*,[37] and Calvo *et al.*[38] have in effect made transistors from fullerenes, chelated metal complexes, alkanethiols, and aromatics. In 2012 Lortscher *et al.* used a molecule with an electric dipole to create a single molecule diode *i.e.* current rectifier (Figure 5a).[57] In 2013 Schirm *et al.* created a single atom transistor by halting junction formation, also with a mechanical break junction, just prior to failure (Figure 5a).[52] In 2008, Grose *et al.* formed a transistor with a magnetic nitrogen endofullerene, allowing magnetic-field manipulation of the spin state and hence conductivity (Figure 5c).[43]

[vii] These statements are somewhat hyperbolic in that in the transistor we must include the dielectric thickness in addition to the channel (molecule) size, and tunneling from the channel through the dielectric to the gate is expected to ultimately limit the extent of miniaturization.

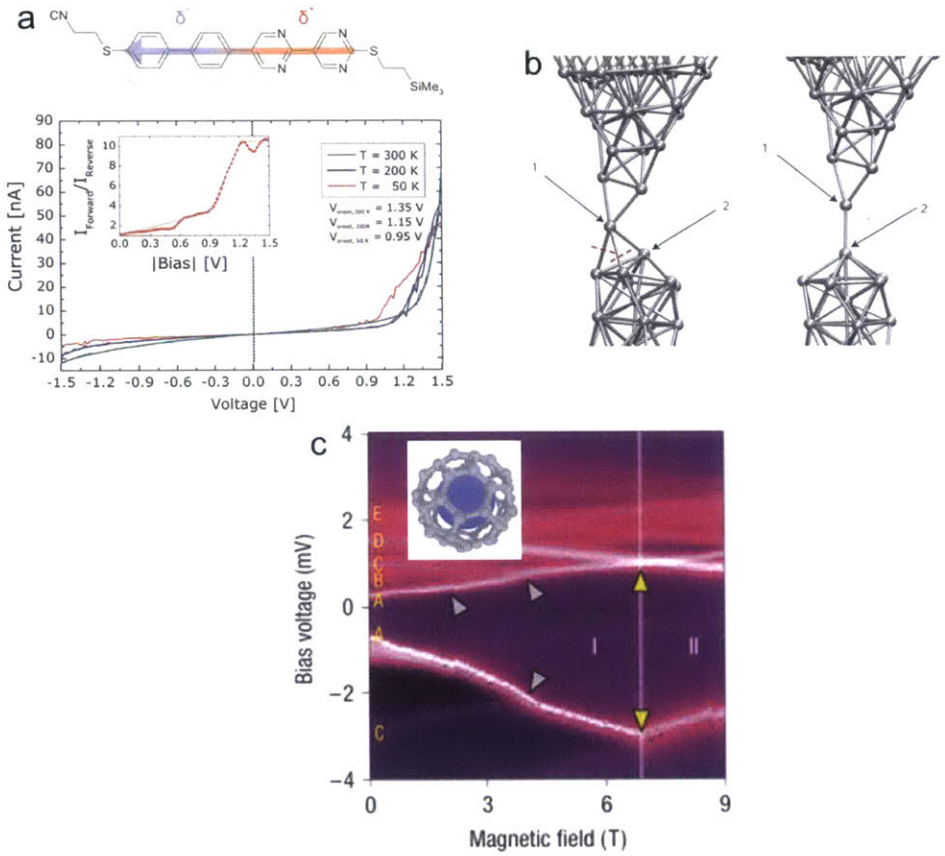


Figure 5. Previous work applying tunnel junctions towards single-molecule electronics development. (a) Single-molecule diode produced by *Lortscher et al.*[57]Top is cartoon of the molecule, bottom is $I(V)$ traces showing the polarity-asymmetric conductivity; at positive 1 V applied bias the channel conductivity increases to 10 times the conductivity at negative 1 V. (b) Hypothetical cartoon of the single-atom transistor observed by *Schirm et al.*, where current application switches the single atom bridge between metastable states.[52] (c) Spin-transistor from *Grose et al.*[43] Colormap of channel conductivity with the source-drain voltage and magnetic field. Inset cartoon of the nitrogen endofullerene molecule in the junction. All works reproduced with permission.

The second promising area of technological application of on-chip tunnel junctions is optoelectronics. As surveyed in the previous section, the electrodes with nanometer separation act as an optical antenna, spatially concentrating some frequencies of the incident field through resonant surface plasmon states.[44] This property has been exploited for surface-enhanced spectroscopy as pioneered by *Ward et al.*[41,44] Technologically, optical rectification and amplification can be exploited to couple photonics to conventional electronics.

Finally, on-chip tunnel junctions may uniquely offer spatial resolution needed for single-biomolecule assays such as peptide, sugar, and DNA sequencing.[46,55,58] Nanometer-diameter pores have generated excitement over the last decade with their ability to translocate single molecules at a time in solution phase.[55,58-60] To do so, a single pore is placed between two electrolyte reservoirs, at least one with the analyte of interest, and a current is passed through the solution between the reservoirs. The

electric field drives the charged analyte through the pore. As it passes through the pore, the ion current measured between the electrodes is modulated by its passage because the molecule is of comparable dimensions to the pore itself; in occupying a substantial cross-sectional area of the pore, the molecule kinetically limits the ion transport, called 'pore blocking.' Nanopores used to translate single-stranded DNA have been made from Silicon oxide and nitride,[61-63] α-Hemolysin[64] and MspA[65] proteins in a phospholipid bilayer, single-walled carbon nanotubes,[59] and graphene.[66,67] In theory, the sequence of a polymeric biomolecule can be identified from the changes in current; each moiety should present a different surface charge, solvent shell, and steric size and therefore create a different ion channel conductivity. While this effect has indeed been demonstrated,[58,59,64-67] the technology has reached an impasse in spatial resolution.[viii] As Henk Postma[68] showed, even in the thinness limit of a graphene nanopore, 1 carbon atom thick, the electric field drops perpendicularly to the pore over more than a nanometer, encompassing for example 4 or more DNA nucleotides. That creates ambiguity in the resulting electrical measurement. The leading proposed solution has been to combine nanopores with tunnel junctions; the exponential decay of tunneling probability with spaces would make the tunneling current sensitive to only a few angstroms of pore length, providing the needed spatial sensitivity to differentiate the polymer sequence.[55,58,60,68] Differentiating between monomers, such as different nucleotides, could be achieved most easily by functionalizing the electrodes with recognition molecules presenting different bonding patterns to different monomers. To demonstrate this latter concept in 2006 *Ohshiro et al.* functionalized an STM tip with a nucleotide self-assembled monolayer, showing by scanning over a DNA strand on the surface that complementary binding could be read out from the tunneling current.[69] More recently Zhao *et al.* placed electrodes with approximately 10 nm separation in a solution and functionalized them with a recognition molecule to yield a measurable conductivity only when analyte binding events occurred;[46] the duration and frequency of the resulting conductivity spikes were found to be characteristic of the analyte.

[viii] There is a second major barrier to nanopore sequencing: control of the translocation rate. Even with ~ 1 nm diameter pores, long biomolecule chains, viscous solutions, and the minimum achievable electric field biomolecules tend to pass through the pore too rapidly, usually by several orders of magnitude, to temporally differentiate between the signals produced by individual monomers.

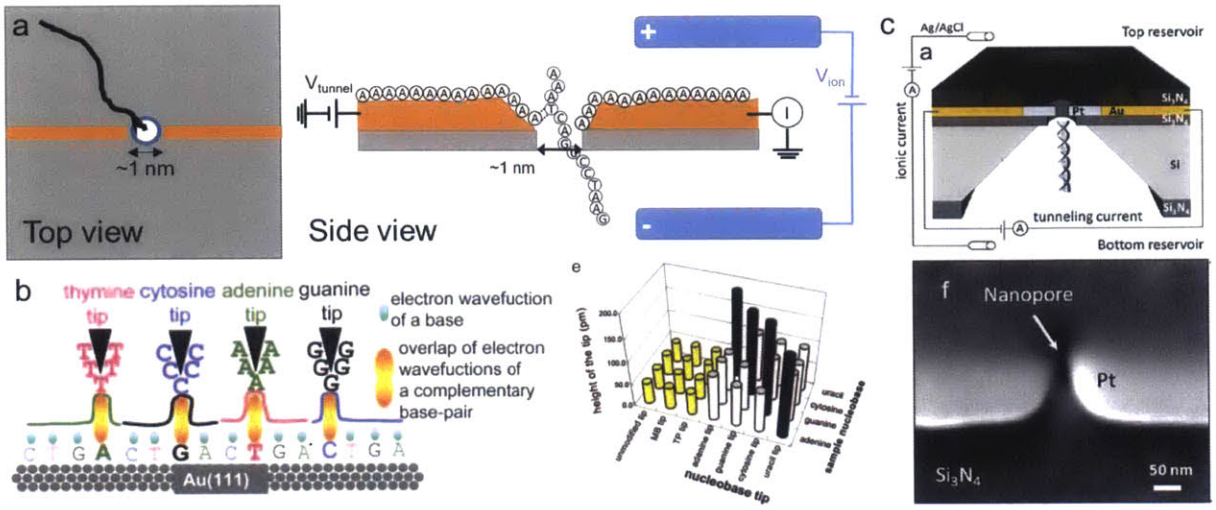


Figure 6. (a) Cartoon of the basic proposed operation of a tunnel junction interrogation of a polymeric biomolecule (in this case DNA) sequence as it translocates through a nanopore. The bias applied in the solution V_{ion} drives the translocation of the charged DNA strand through the nanopore. As it passes through, a bias V_{tunnel} is applied across a tunnel junction collaocated with the nanopore. The tunnel junction metal is coated with a monolayer of a single nucleotide, providing a tunneling current readout that is highly sensitive to the formation of a complementary base pair due the resulting increase in molecular orbital coupling between molecules upon hydrogen bonding. Multiplexing four pores functionalized with each of the four nucleotides would provide a total sequence. (b) Demonstration of the recognition principle by *Ohshiro et al.* in STM, where functionalizing the STM tip with different nucleotides yielded a tunneling current sensitive to the formation of base pairs with DNA sequences on the surface.[69] (c) An experimental attempt to realize this concept by *Ivanov et al.*[60] The shortcoming was co-locating the nanopore and electrodes to sufficient accuracy that the tunnel current was sensitive to the translating analyte.

While appealing in theory, this combination of tunnel junction with nanopore has proven challenging. As discussed in the next section, electromigrated tunnel junctions remain the only successful means of fabricating on-chip tunnel junctions to date. As discussed in *progress to date in the formation of on-chip tunnel junctions* however, they cannot be spatially positioned with nanometer resolution. As a result, fabrication of a tunnel junction co-located with a nanopore remains unaccomplished despite published attempts.[60] In part towards this end, we investigate in Chapter III the point of electromigrative failure in nanowires, specifically whether failure occurs at material interfaces as it does in micron-diameter wires, and find that on the contrary the nanowire crystal structure controls the point of failure.

Comparison to alternative techniques

Electromigrative failure of nanowires is not the sole means of forming single-molecule tunnel junctions. Techniques can be separated into two categories: on-chip and off-chip. On-chip methods, such as electromigration, create a tunnel junction that, once formed, is stand-alone *i.e.* can be disconnected from the apparatus. Off-chip methods, such as Scanning Tunneling Microscopy (STM) and Mechanical Break Junctions (MBJs) require active feedback to maintain the junction. In this section I will briefly survey the alternative techniques in these two categories. In general, off-chip methods offer greater

reliability and control in the formation of tunnel junctions, while sacrificing the possibility of scaling the devices for technological application.

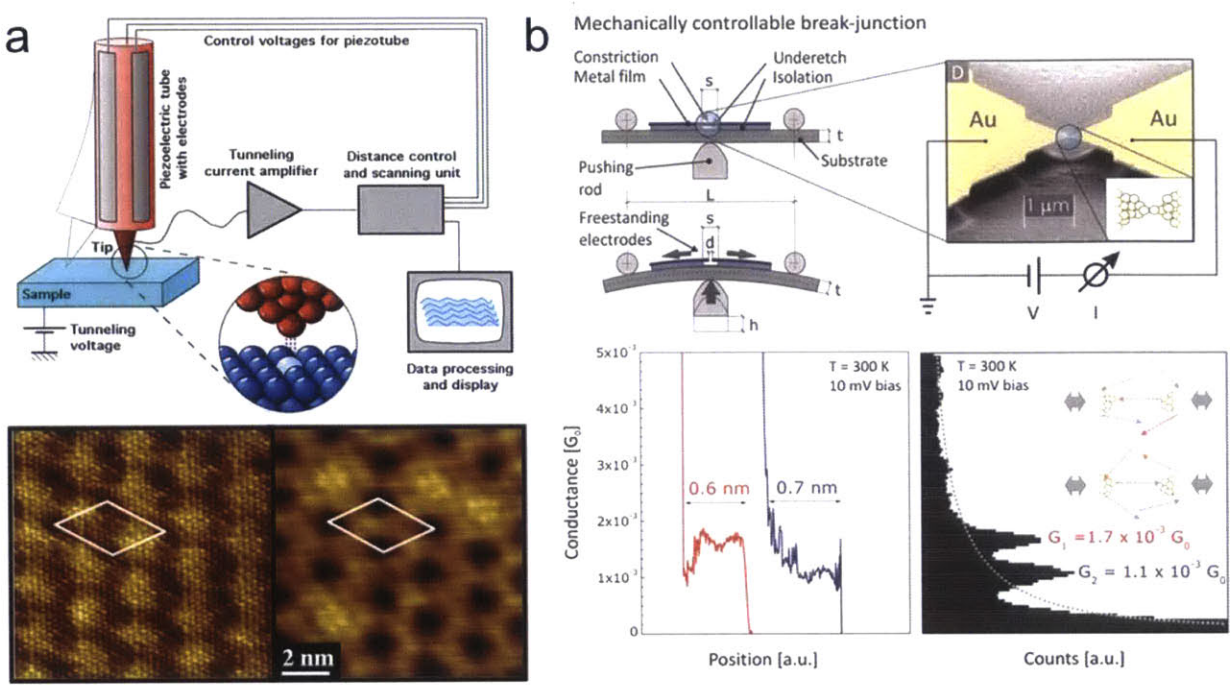


Figure 7. Examples of off-chip tunnel junction devices. (a) Scanning tunneling microscopy. Cartoon illustrating operational principle (top) and example STM images (bottom) of a Ag metal layer on top of Cu[111] (left), showing atomic resolution of the metal atoms, and the same substrate with four thioether molecules shown rotating on the surface (right). Cartoon courtesy of Michael Schmid of TU Wein. (b) Mechanical Break Junctions. Cartoon illustration operational principle (top) with false color SEM image of the resulting device, and example conductance traces measured with displacement (bottom). Reproduced from Lortscher *et al.*[57] with permission.

The most common off-chip tunnel junctions are made in Scanning Tunneling Microscopy (STM). A sharp tip is brought within a nanometer distance of a metal surface by monitoring the current passing between the two; with the exponential dependence of tunneling current on tip-surface separation, the current measurement can be used to feed back to piezoelectric *z*-axis modulators, maintaining the tip-surface gap at a fixed current value. The tip can then be scanned across the surface, producing an atomic-resolution image of the surface electronic structure at the applied bias (Figure 7a). This can be used to find features on the surface, such as isolated molecules, and park the tip over them to perform tunneling spectroscopies such as discussed below. Because of its sensitivity to impurities such as organics coating the surface, STM is usually performed at low temperature and ultra-high vacuum, but it can also be performed in solution by putting a bias on the surface relative to a counter electrode to electrochemically clean the substrate. Among electron tunneling techniques with single-molecule spatial resolution, STM affords the greatest versatility in studying diverse molecules and the greatest reliability in forming tunnel junctions and performing tunneling spectroscopy. For example, inelastic tunneling spectroscopy is regularly measured in STM,[39,70-73] but in other single-molecule tunnel junction techniques only 6

have been reported to our knowledge.[40,74-77] Nevertheless the platform has limitations in both available interrogation techniques and application to technology development. In the former case, STM presents only a two-electrode system; its architecture precludes the addition of a back-gate to inject charge into the channel. As a result gating experiments such as investigation of Coulomb Blockade cannot be performed. Technologically, in addition to being off-chip, STM devices do not scale; parallelizing the apparatus would require independent piezoelectric feedback for each tip. The former problem is addressed by another popular off-chip apparatus: mechanical break junctions (MBJs). Mechanical break junctions are constructed by suspending a nanowire above a trench with a piezoelectric bulb (or 'screw') beneath the substrate right under the wire (Figure 7b). The bulb is then gradually expanded, bending the substrate towards the wire and thereby stretching the wire, thinning it. The conductivity of the wire can be monitored as the bulb is expanded and the wire can be gradually stretched until it finally breaks to form a single tunnel junction, similar to electromigrated tunnel junctions. All the experiments that can be performed with an electromigrated tunnel junction, including gating the channel and testing in solution, can be performed with MBJs. With the piezoelectric feedback allowing slow expansion and contraction of the gap, the formation of tunnel junctions with MBJs tends to be more reliable than electromigrated tunnel junctions, but at the cost of being off-chip; the device cannot be dis-engaged from the piezo-electric screw as the substrate would then relax. The junction must be actively maintained. Additionally, like STM, MBJs cannot be parallelized.

Electromigrated tunnel junctions, as explored in Chapter II, form less reliably and are harder to maintain than STM and MBJ tunnel junctions, but afford the benefit of being scale-able. ETJs can be fabricated in parallel,[78] and can be disengaged entirely from the testing apparatus and re-contacted. No other on-chip fabrication method to date has been able to produce a tunnel junction. Any technology seeking to exploit tunnel junctions, such as those described in the previous section, must therefore use electromigratively-formed ones. In addition to the improved throughput in SMTJ investigations, this technological need motivates our research in Chapters I and II into understanding the mechanism of nanowire electromigration and developing an electromigration process control based on that mechanism to improve the reproducibility of tunnel junction formation.

Because of the contemporary short-comings in the formation of electromigrative tunnel junctions, explored in greater depth in the section *progress to date in the formation of on-chip tunnel junctions* below, alternative methods of on-chip tunnel junction formation have been explored. None have produced gaps small enough to measure a tunneling current across, but they nevertheless have generated excitement for their potential. The first is electroplating. First two Au electrodes with a large, of order 100 nm, separation are fabricated with electron beam lithography. Then one or both of the electrodes are used as working electrodes in the solution-phase reduction of another metal ion, plating the electrodes and

gradually reducing the gap size. The challenge is achieving suitable control over gap sizes; the metal deposition must be arrested before the gap is closed, but not before a sub-nanometer gap is formed. Deshmukh *et al.* in 2003 made the first attempt at this approach (Figure 8a);[79] they attempted to control the gap size by monitoring the current between the two electrodes during the deposition process. As the gap shrunk, the tunneling current rose. The process remained too rapid however to provide suitable time-sensitivity in the current measurement to gap size, and they could only produce gaps of 2 nm and larger. More recently, Serdio *et al.* developed a self-limiting electrodeposition process (Figure 8b).[80] They added a molecule spacer, C_x-TAB, of an alkane chain with a trimethyl ammonium head that would adsorb to the electrode surface during the metal deposition process. As the electrodes grow close to each other, the self-assembled monolayers on each electrode would bond to each other by van der Waals interactions of the alkane tails, stabilizing and sterically hindering further electrode growth. In this manner the gap size would be limited by the alkane chain length, and by varying the chain length they were able to vary the gap size. Despite the efficacy of the approach, they gain could not achieve sub-nanometer gap sizes. The final and most recent on-chip fabrication method has been Focused Ion Beam (FIB) milling of nanowires. This year (2015) both Li *et al.*[81] and Cui *et al.*[82] indepdently applied FIB to metal nanowires, etching nanometer separations between the two (Figure 8c). Unfortunately neither group produce a gap smaller than 4 nm.

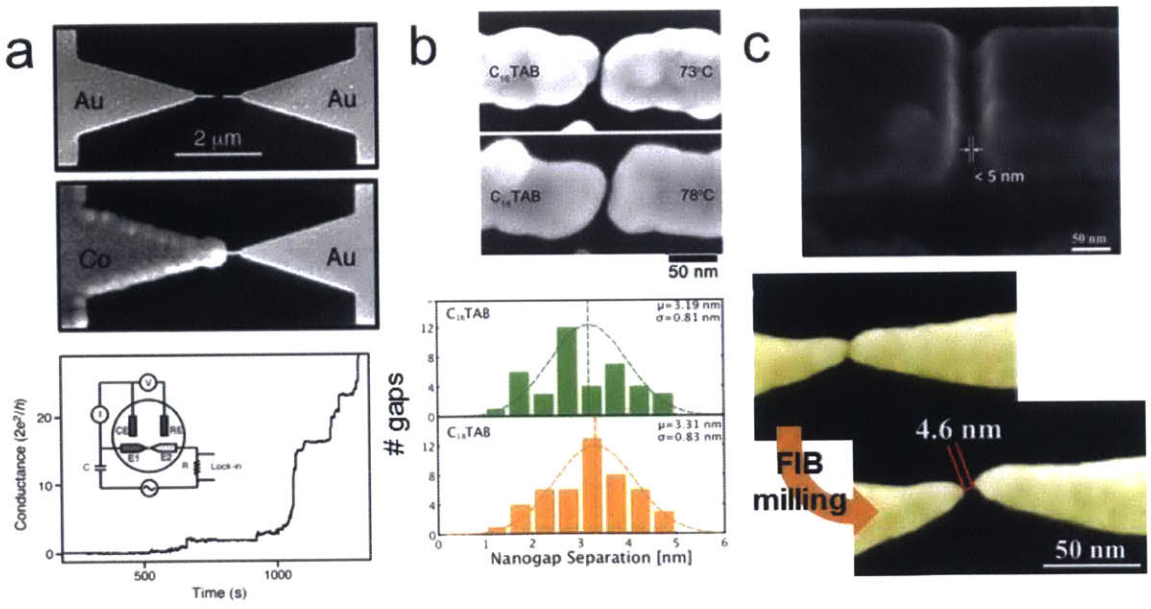


Figure 8. Alternative methods of on-chip tunnel junction formation. (a) Electrodeposition of Co from solution onto Au electrodes from *Deshmukh et al.* using feedback from the tunneling current.[79] (b) Self-limiting electrodeposition using C-TAB spacers between electrodes, changing the length to change the gap size, by *Serdio et al.*[80] (c) FIB milling to form nanometer gap electrodes. Top from *Li et al.*,[81] bottom from *Cui et al.*[82] All figures reproduced with permission.

Features of electromigrative formation of tunnel junctions

In this section we survey the basic features of the electromigration of nanowires towards the formation of tunnel junctions. In the next section, *progress to date in the formation of on-chip tunnel junctions*, we survey the progress improving the method in the literature to date. In Chapters I and II, our investigations reveal the mechanism underlying these features and exploit that understanding to improve tunnel junction formation.

A bias V is applied across the nanowire, driving metal atom electromigration. The morphological evolution of the wire during electromigration was investigated using *in situ* SEM and TEM in 2007 by Taychatanapat *et al.*,[22] 2008 by Strachan *et al.*,[29] and 2009 by Huang *et al.*[83] As described in the section *Metal Electromigration* heterogeneities along the nanowire, such as grain boundaries or material interfaces, create local gradients in the flux of metal atoms, leading to a net loss or gain of material at different points along the nanowire length, cross-sections perpendicular to the applied electric field. In nanowires, there is generally one point that has a dominant gradient, such that the loss of material elsewhere is often negligible. In micron-sized wires, loss of material yields voids as vacancies aggregate with increased concentration. In nanowires however the point of most rapid electromigration begins to thin, or 'neck.' As the neck thins, reducing the area of the charge carrier channel, the measured resistance rises, providing an electrical observation of the current state of the wire.[22,29] When the applied bias is too low, the rate of electromigration can be too slow to observe on reasonable experimental timescale, but when the applied bias is too high electromigration proceeds so rapidly that it becomes explosive. The bias is therefore started low and gradually increased, monitoring the wire for increases in resistance. We found in Chapter II that at low bias a reversible increase in resistance without any observable accompanying change in nanowire morphology is observed, which we assign to Joule heating of the wire. When electromigration proceeds slowly enough and the wire thins to a hundred atoms or so, an effective diameter around 10 nm, electromigration is no longer continuous on measurement timescales (typically tens or hundreds of milliseconds). The wire becomes stationary until groups or single atoms displace suddenly, producing steps in the measured resistance (or conductivity) such as we observe in Chapter IV, Figure 39. Eventually the wire breaks, forming a gap between the resulting electrodes. As *Park et al.* showed,[23] when a nanowire is used there is often a point between the two electrodes where the gap is small enough, approximately 1 nm or less, to produce a measurable tunneling current. As described in the *Discovery* section, the exponential decay of tunneling probability within a few angstroms leads the single ray, 1-3 atoms on each electrode, of smallest separation to dominate the tunneling current. When the conductivity drops below the conductance quantum as described in the next paragraph, the voltage is attenuated to avoid relaxing the junction (see footnote vi).

The existence of a tunnel junction after formation is generally taken to be implicit in the conductivity. The conductance quantum, $G_0 = 2q^2/h = 7.75\text{x}10^{-5}$ S,[ix] is the conductivity of a single (spin degenerate) ballistic channel *i.e.* the conductivity of a single electron state in the absence of any scattering. A tunnel junction necessarily has a conductivity less than G_0 due to the sub-unity tunneling probability. When the wire breaks, the conductivity falls below G_0, indicating that a tunnel junction has formed. That conclusion is further supported by observations of the stochastic regime; Strachan *et al.*[8] in electromigrative tunnel junctions and Schirm *et al.*[52] in mechanical beak junctions published observations similar to ours of the stochastic regime, showing that the steps in conductivity are often integers of the conductance quantum, indicating the formation of parallel single-atom bridges. In some work, including our own, further evidence supporting the presence of a tunnel junction is collected. First, in the absence of molecules, the tunneling current is expected to be Ohmic *i.e.* $I(V)$ is linear. Second, ballistic conductivity can be verified with temperature-variation experiments. The resistivity of bulk metals increases substantially with temperature as electron-phonon scattering, the typical electron scattering mechanism, increases. Ballistic conduction entails no scattering mechanism however[x] and as such is temperature invariant. By measuring $I(V)$, therefore $G = I/V$, at a range of temperatures the junction current can be confirmed to be dominated by tunneling (*e.g.* see Chapter II, Figure 29c). Finally, the nanowire can be imaged in electron microscopy after formation to confirm that the gap between electrodes is sub-nanometer. Unless performed *in situ* and with high resolution TEM, this piece of evidence is not considered compelling however for several reasons: most SEM apparatus' can only confirm that the gap is less than several nanometers;[xi] when an adhesion layer such as Ti is present the tunnel junction can form in the Ti, which usually can't clearly be imaged, rather than the Au;[30,84] removing the sample from the low-temperature chamber can cause the sharp, non-equilibrium Au features to relax;[28] and the electrostatic discharge or testing of the device and neighboring electrically-contacted devices eventually causes the junction to expand.

Progress to date in the formation of on-chip Single-Molecule Tunnel Junctions

Applying electromigrated tunnel junctions to gain physical insight or develop technologies requires not only advances in post-formation interrogation as surveyed above, but also reliable formation of the tunnel junctions themselves. Although a valuable proof of concept, the nanowire fabrication and electromigration process used by *Park et al.* in the first devices proved to be highly variable and have a

---

[ix] where q is the elementary charge, h is planck's constant, and the factor of 2 is for two degenerate spin channels.

[x] Electrons in tunnel junctions do scatter by coupling to intersititial molecules, as discussed in the next section, but these events are both approximately temperature invariant and increase the conductivity with increased coupling, rather than decrease it.

[xi] Sub-nanometer imaging of insulative substrates with SEM remains non-trivial.

low yield of successful tunnel junctions. We define 'successful' junctions as those that are small enough for a tunneling current to measured across, approximately 1 nanometer or smaller. Since the first devices, the field has made gradual improvements to the electromigration process through empirical observation. Based on the first method presented by Park *et al.*,[23] investigators initially placed a resistor in series with the nanowire device,[33,36,37,85,86] but in 2005 Esen and Fuhrer[87] found that actually minimizing the series resistance improved outcomes. We hypothesize that as the neck thins and its resistance rises, its share of the total resistance increases and more of the applied bias drops across nanowire, in effect accelerating the voltage application, hence accelerating electromigration and leading to more catastrophic failure. In 2009, *Xiang et al.*[88] found that using nanowires that are too large, in their case of order 1 μm^2 cross-sectional area, led to periodic re-formation of the nanowire after gap formation. In our own experiments (see Chapter II), we found that not only was this the case for any wire greater than 0.1 μm^2 cross-sectional area, but that such large wires exhibit explosive formation more frequently. In general we have found that the smaller the nanowire diameter, the higher the yield of successful junctions.[xii]

Perhaps the greatest contribution to the electromigrative formation of tunnel junctions to date was the development of a feedback control algorithm by Strachan *et al.* in 2005.[8] Rather than ramping the voltage and attenuating only when the wire has broken, they monitored the nanowire resistance whenever it rose above a preset percentage threshold, typically 0.1%, and then quickly attenuated the voltage to zero before gradually ramping again, repeating the cycle until the wire broke. This technique, which we call the 'Penn method,' produces a more gradual electromigration pathway and increases the yield of successful junctions substantially. It has since been adopted in varying forms by every group performing electromigration of nanowires for tunnel junction formation. Although influential, the Penn method had prior to our work not been explored in depth. The formula had not been parameterized, with parameter values that could be shared in publication, and characteristic features of the formation pathway ubiquitously observed had not been identified or explained. In Chapter II we investigate the Penn method in greater depth.

Despite the progress surveyed above, adoption and advancement of electromigration as a means of tunnel junction formation has nevertheless remained hampered by poor yield and process variability.[8,22,85,89-92] We posit that the root limitation is in the lack of knowledge of the mechanism

---

[xii] The lower boundary on nanowire diameter is not merely set by fabrication techniques, but by practicalities of testing. We have found that nanowires with less than 100 nm^2 cross-sectional area electromigrate too easily, such that even weak static fields destroy the wires. More recently this was verified by *Cui et al.*[89] who found that in such small nanowires relaxation of the morphology can occur at grain boundaries, breaking the wire at the boundary to form smooth-ended electrodes, was fast enough to occur within a few days a room temperature.

of electromigration at the nanometer scale. The above questions, for example, of series resistance, cross-sectional area, and process control could be accessed empirically but not explained theoretically. As described in the *Metal Electromigration* section above and investigated in Chapters I and III, the well-understood behaviors of electromigration in micron-diameter and larger wires break down at the nanometer scale. As a result, it is unclear how the method can be improved to a level of reliability that is both commercially viable and increases experimental throughput. For example, the Penn method is neither continuous nor has any connection to electromigration kinetics. It has provided a means of slowing down electromigration, but is not a traditional process controller in that it has no setpoint; it ostensibly 'controls' R, but to date no group has been able to 'plug-in' a resistance pathway. The pathway of electromigration, *i.e.* the evolution of wire radius and resistance, remains spasmodic and unpredictable. A Joule heating mechanism for failure has been proposed with some indicators, but has yet to be validated or exploited to alter the electromigration process.[91,93] This lingering ambiguity in both the appropriate design of nanowires and electrical control of the electromigration process continue to restrict the technology to early developmental stages (*cf.* references [82,84]).

To address this situation in Chapter I we propose a mechanism of electromigration in nanowires and validate it against experimental data from Karim *et al.*[94] in both the wire radius and time dimensions. We posit that as wire diameters fall to tens or hundreds of nanometers, the dominant mechanism of electromigration shifts from bulk vacancy transport to surface atom transport as the surface area to volume ratio falls. In Chapter I and Chapter II we show that this mechanism provides physical explanations for the observed features of the Penn method wire evolution, and we exploit it to design a better means of controlling nanowire electromigration. We find that dR/dt (specifically $dlnR/dt$) reflects the mean rate of electromigration and can be maintained at a constant setpoint to produce gradual formation and a predictable $R(t)$ pathway. The pathway is also sensitive to the setpoint's value, providing further validation for our proposed mechanism and offering the potential to design the electromigration pathway by designing the setpoint pathway rather than holding it at a single value for the duration. We find that the setpoint can be optimized to raise device yield, and that under the right conditions it is possible to achieve a 100% success rate.

Electron Tunneling Spectroscopies

For a more in-depth discussion of electron tunneling than that provided in this Introduction, please refer to the following publications:

1. Supriyo Datta. *Electrical Resistance: An Atomistic View*. Nanotechnology 15, S433, 2004.
2. Abraham Nitzan and Mark A. Ratner. *Electron Transport in Molecular Wire Junctions*. Science 300, 1384, 2003.

3. K.W. Hipps and Ursula Mazur. *Inelastic Electron Tunneling: an alternative molecular spectroscopy.* J. Phys. Chem. 97, 7803, 1993.
4. Abraham Nitzan. *Electron transmission through molecules and molecular interfaces.* Annu. Rev. Phys. Chem. 52, 681-750, 2001.
5. Robert Quan, Christopher S. Pitler, Mark A. Ratner, Matthew G. Reuter. *Quantitative Interpretations of Break Junction Conductance Histograms in Molecular Electron Transport.* ACS Nano 2015.

Electron tunneling through an energy barrier

As illustrated in Figure 9, the stationary wavefunction of any particle faced with a finite energy barrier, *i.e.* a potential greater than the particle's energy, persists into the barrier with some decay over depth. The greater the energy barrier the more rapid the decay.[xiii] When the barrier is spatially narrow enough and energetically low enough the particle can penetrate, or 'tunnel through,' the barrier, achieving some reasonable probability of existence on the opposite side. For example, for a plane wave particle, *e.g.* an electron in a flat potential, in one dimension, the probability $T(E)$ of transmission through a square potential barrier such as in Figure 9b with particle energy E is

$$T(E) = \frac{16E(V_b - E)}{V_b^2} e^{-2d\sqrt{\frac{2m(E-V_b)}{\hbar^2}}} \quad (4)$$

where m is the particle mass and $\hbar$ is the reduced Planck's constant, and the remaining variables are as defined in the figure. While a truly one-dimensional plane wave particle facing a square barrier is rarely encountered in reality, the basic features of this system are generally preserved. In particular, the probability of tunneling through the barrier exponentially decays with the width of the barrier, the height of the barrier, and the mass of the particle. In the absence of coupling to other particles (see section below), the process is elastic, with no loss of particle energy.

[xiii] In all cases that the author can conceive of.

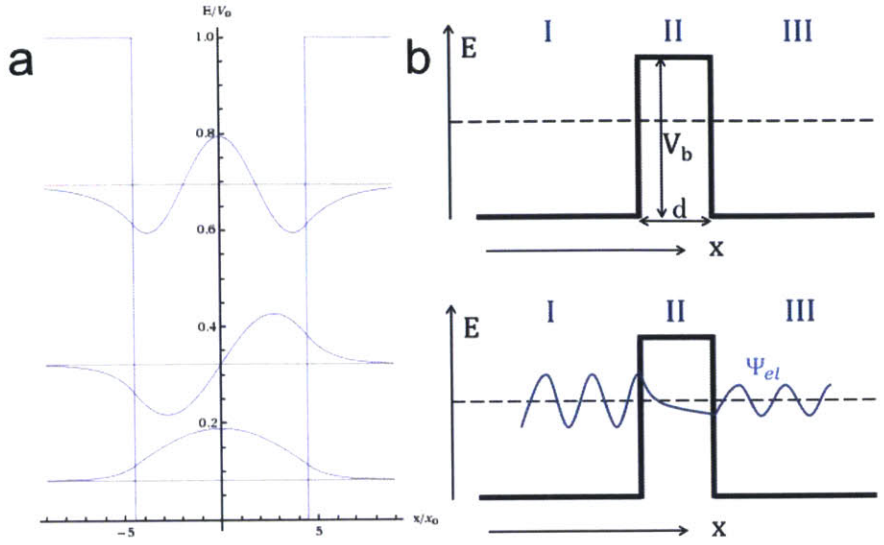


Figure 9. (a) Particle in a box wavefunctions (blue) and energies (red) for a finite potential well. Attribution Anna Fitzgerald. (b) Cartoon illustrating one dimensional tunneling of a plane wave (free-space) electron through a finite square potential. (top) definition of variables, barrier height V_b, spatial with d, cartesian coordinate x, electron energy E (dashed line). Black line is the semi-infinite (in x) flat potential field the electron is exposed to, producing a plane wave in regions I and III. (bottom) cartoon illustration of electron wavefunction Ψ_{el} in the three regions (I, II, III) of interest, 'incident' from the left. The spatial occupation of the electron persists through the barrier with exponentially-decaying probability, producing a plane wave on the opposite end of the same energy (frequency) but attenuated amplitude.

In this manner metal electrons can traverse a vacuum or insulator separating two electrodes, such as a substrate and STM tip, or the two electrodes formed from an electromigrated nanowire. It should be noted that the specific transmission probability function for a given system is especially complicated because the constituent electron stationary states are dictated by the local configuration of metal surface atoms, not merely the bulk dispersion. Nevertheless the same general principles apply, with the electron tunneling probability exponentially decaying with electrode separation, particle mass, and barrier height. The latter depends on whether the interstitial is a vacuum, solvent, or metal oxide insulator. In tunneling through a vacuum, separations greater than approximately 2 nm yield negligible tunneling probabilities in our experiments of interest.

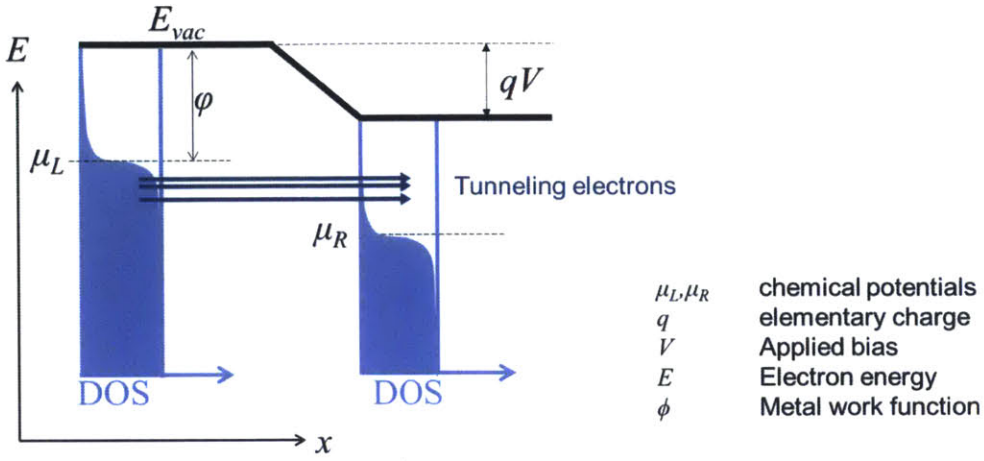


Figure 10. Illustration of electron tunneling between two electrodes under an applied electric field. y axis is energy but x axis is two-fold: electrode density of occupied electron states (DOS) in blue and the spatial coordinate along the ray of minimum electrode separation in black. μ is the chemical potential of the left and right electrodes, φ is the work function of the metal (assumed the same for both electrodes), E_{vac} is the vacuum energy, q is elementary charge, and V is the applied bias.

When an electric field is placed across such two electrodes, such as in Figure 10, electrons in the higher-bias electrode with energy greater than the chemical potential of the lower bias electrode will spontaneously translate to unoccupied states in the opposite electrode. If excess charge is drained from each electrode by completing the circuit, a non-equilibrium current is produced, analogous to placing a bias across a resistor. As no scattering is involved, this conduction is ballistic. The cumulative tunneling probability for all electron states therefore produces a corresponding tunneling current I under an applied bias V. The tunneling current can be described in the Landauer-Büttiker formalism by integrating the tunneling probability over the dispersion curve as

$$I(V) = \frac{q}{2\pi\hbar}\int_{-\infty}^{\infty} T(E)\big(f(E-\mu_L) - f(E-\mu_R)\big)dE \tag{5}$$

where $f(E-\mu_{L/R})$ is the density of occupied states for the left and right electrodes, typically a Fermi-Dirac distribution around the chemical potentials of each electrode,[xiv]

$$f(E) = \frac{1}{1+\exp\left[\frac{E}{k_B T}\right]} \tag{6}$$

where k_B is Boltzmann's constant and T is temperature. As an example of its application, in Chapter IV we employ this formalism to simulate inelastic transport in our tunnel junctions. Expression (5) has a fairly simple conceptual construction. The difference term between the density of occupied states between

[xiv] For a bulk metal, the total density of states (occupied and unoccupied) near the Fermi level is approximately flat.

the two electrodes is the density of occupied electron states in the high-bias electrode with a corresponding degenerate unoccupied electronic state in the opposite electrode *i.e.* the density of electrons that will produce a tunneling current *via* elastic transitions, illustrated in the following figure. The current can therefore be thought of as the sum of the product of all electron states making the transition times the transition probability for each of them.

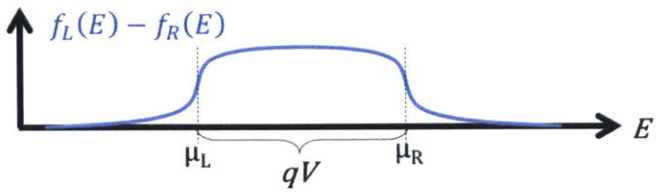


Figure 11. Illustration of the integrand of equation (5).

Measurement of the current-voltage, or $I(V)$, behavior of the tunnel junction is therefore the predominant means of interrogating tunnel gaps. While the exact tunneling probability $T(E)$ is complicated and electrode-morphology-specific, in the absence of an interstitial molecule in the gap between electrodes (discussed in the next section) and over a small applied bias (a few hundred millivolts) it is approximately constant with energy. As a result, the 'direct tunneling' between electrodes produces an approximately linear increase in current with applied bias, as the number of electrons making the transition linearly rises (*e.g.* the integral of Figure 11). Tunnel junctions in the absence of a molecule therefore usually behave as an approximately Ohmic resistance.

Molecular-orbital-mediated electron tunneling

When a molecule is in the interstitial space between the electrodes, its electronic states couple to those of tunneling electrons, providing an additional pathway for tunneling electron transport through the molecular orbitals. Upon coupling to either or both of the contacts (Figure 12b), the molecule's electron chemical potential equalizes to the fermi level of the contact *i.e.* the molecule accepts or donates electron density such that its ionization energy equilibrates with the work function of the metal. Coupling to the bath of electrode electronic states also broadens the molecular orbitals (MOs), and combined with the partial charging perturbs the molecular orbital energies. As discussed in the previous section, when a steady-state electric potential gradient is maintained the higher-bias electrode Fermi level is raised above the low-bias electrode. If a molecule is in the interstitial, its chemical potential μ_M may follow that of either electrode or fall somewhere in between, depending on its coupling to each electrode. Regardless, if the applied bias is high enough a molecular orbital will be between the Fermi levels of the two electrodes. In such a case, elastically transmitting electrons in the high-bias electrode, moving to unoccupied orbitals in the low-bias electrode, will be degenerate with part of the molecular orbital density of states (as it is broadened by the perturbation of the electrodes, see above). Elastic transmission into and out of the

molecular orbital in order to traverse the electrodes is therefore possible, called orbital-mediated tunneling (OMT).

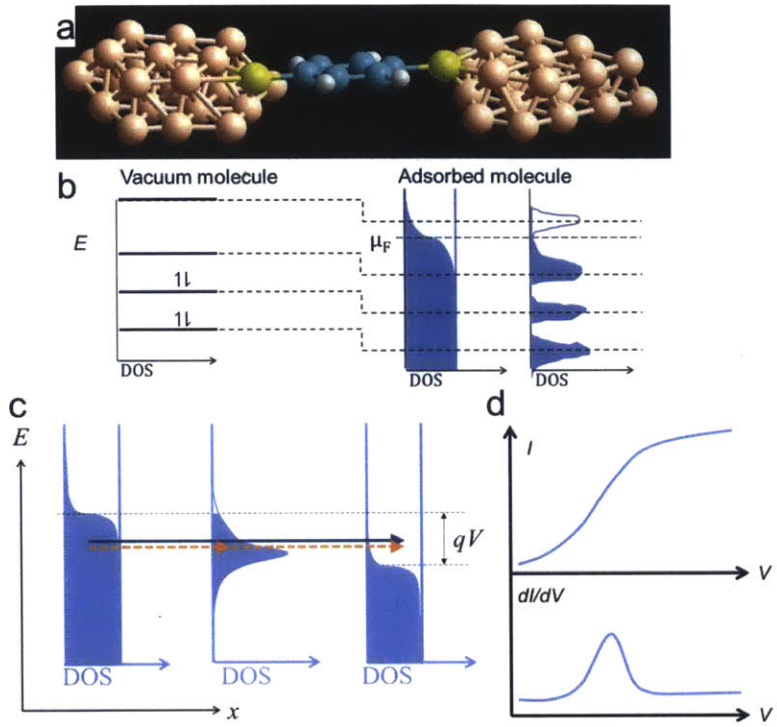


Figure 12. Electron tunneling through a molecule channel. (a) cartoon of a benzene dithiol molecule in a tunnel junction. (b) Illustration of the effect of coupling between the molecular orbitals and the electrode electronic states. The molecule and electrode density of occupied states (DOS) are plotted against electron energy. There are two major effects. First, the electron chemical potential of the molecule shifts to equalize to the Fermi level of the electrodes *via* transfer of charge until equilibrium is reached, illustrated by the dashed lines. Charge transfer also alters the electronic structure as with any ionization. Second, the molecular orbitals broaden due to their equilibrium superposition with the bath of electrode states. (c) Illustration of elastic electron conduction through a molecule. In a completed circuit, the applied bias V creates a persistant energy drop between the electrodes qV that yields elastic transport of electrons from the left electrode to unoccupied states in the right electrode, as in Figure 10. When a molecular orbital of the interstitial molecule is energetically between the electron chemical potentials of the two electrodes, it provides an additional pathway for electron transfer. (d) The current-voltage behavior resulting from molecular orbital tunneling. As qV rises to encompass the DOS of the molecular orbital, the additional conduction channel opens up, increasing the current for every additional electron that can transfer. Once qV is such that the molecular orbital is entirely encompassed between the electrode chemical potentials, additional increases in bias will not provide access to any further conduction channels *via* that molecular orbital. As a result, orbital-mediated tunneling produces a step in current, and corresponding peak in dI/dV, at the relevant bias.

This tunneling pathway is approximately[xv] additive with the direct tunneling, thereby increasing the tunneling probability and hence tunneling current. In developing transport simulations in Chapter IV the Landauer-Buttiker $I(V)$ expressions analogous to equation (5) for elastic and inelastic (next section) transport through molecules are presented; briefly, the energy dependence of OMT is the density of states

[xv] Recent theory suggests that conduction channels are not entirely independent and have multiplicative effects on current *i.e.* can be coupled. See *Quan et al.*[95]

$D(E)$ of the molecular orbital of interest and the full expression is attenuated by the magnitude of coupling of the molecular orbitals to the metal electronic states. As the voltage rises to produce tunneling electrons degenerate with the MO, the current rises commensurately; when the MO electron density is fully contained between the chemical potentials of the two electrodes, further increases in voltage provide no further gains in OMT, and the current again increases merely linearly with voltage *via* direct tunneling. As a result, molecular orbitals appear as broadened steps in $I(V)$ and corresponding peaks in $dI/dV(V)$, the differential conductance (Figure 12d). The spacing between MO peaks, symmetry between positive and negative bias, and broadening of the peaks all depend on the particular coupling of the MO states to each of the two electrodes. For example, if the separation between electrodes is large compared to the size of the molecule and the molecule adsorbed onto one of the electrodes, its chemical potential is expected to track approximately with that of the electrode it is adsorbed on, producing an asymmetric (around 0 bias) OMT spectrum ($dI/dV(V)$) corresponding to the series of MOs. If instead the molecule is equidistant between the two electrodes, its chemical potential may always be halfway between the two, yielding a symmetric OMT spectrum as positive or negative bias would encompass the same MOs. The magnitude of the $I(V)$ step is proportional to the strength of the reduced coupling to the two electrodes.

Inelastic tunneling

Tunneling electrons can couple to the vibrational modes of the molecule, yielding scattering events where the vibration is excited in exchange for a corresponding loss of electron energy, analogous to the mechanism by which hot electrons relax *via* crystal phonon excitation. This inelastic pathway offers an additional transmission mechanism between electrodes, increasing the tunneling current. Unlike orbital-mediated tunneling, any electron of arbitrarily high energy can inelastically scatter with approximately equal probability; inelastic scattering therefore linearly increases with applied bias *i.e.* the population of available tunneling electrons, increasing the slope of $I(V)$. To scatter however the tunneling electron requires an unoccupied state at the lower energy to relax to, and as such only electrons with energy $E > \mu_{LB} + \hbar\omega$ can inelastically tunnel, where μ_{LB} is the Fermi level of the lower-bias electrode (the electron 'drain') and ω is the frequency of the vibrational mode (hence $\hbar\omega$ is the quantum of vibrational energy). As a result, the slope of the $I(V)$ curve increases due to a given vibration only when $qV > \hbar\omega$. Each vibrational mode therefore appears as a step in dI/dV, the differential conductance, at $V = \hbar\omega/q$ and a corresponding peak in the d^2I/dV^2. The d^2I/dV^2 peaks over V is called the inelastic tunneling spectrum (IETS), such as we perform in Chapter IV. Molecular orbital tunneling peaks are generally significantly stronger, albeit one to two orders of magnitude broader, than IET vibrational peaks; to clarify the spectrum, it is usually divided by the differential conductance *i.e.* typically $d^2I/dV^2 \,/\, dI/dV$ is plotted.

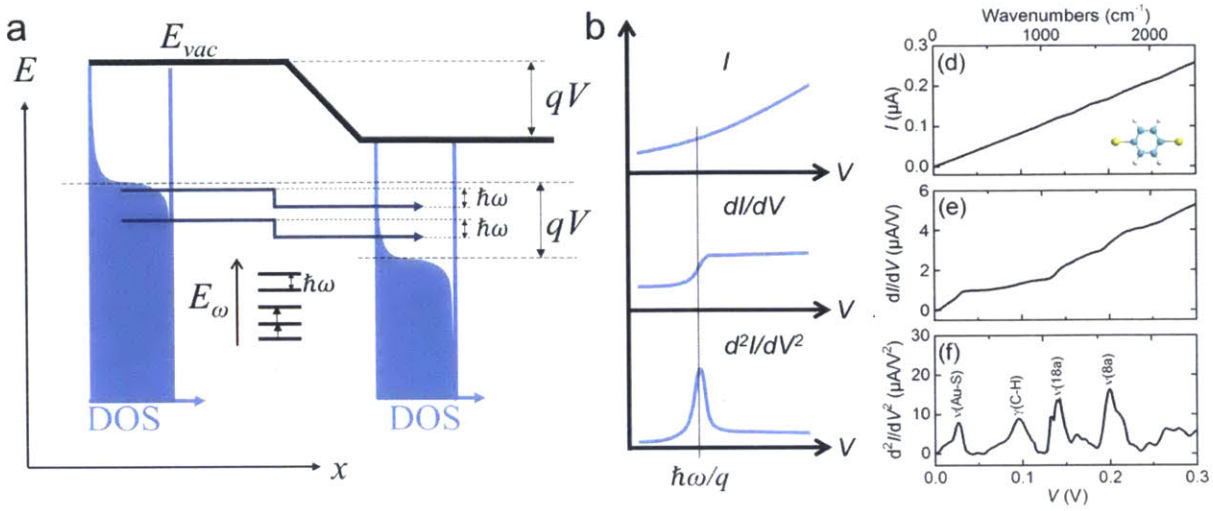


Figure 13. (a) Illustration of the inelastic scattering process. An electron traversing electrodes relaxes during tunneling to excite a vibrational mode on the molecule (center of figure). *N.B.* the figure is merely figurative; the electron and vibrational energies are on independent axes, and the *x*-axis for the vibrational modes is meaningless. (b) Cartoon illustration (left) and experimental data from Song *et al.*[40] (right) of the inelastic tunneling spectrum produced as d^2I/dV^2. Data taken for 1,8-octane dithiol in a Au electromigrated tunnel junction.

Carbon Nanotube Electronic Structure

For insight into carbon nanotube electron states beyond the content in this Introduction, please refer to the following publications:

1. Susan Reich, Christian Thomsen, and Janina Maultzsch. *Carbon nanotubes: basic concepts and physical properties*. Wiley-VCH, Weinheim ; Cambridge, 2004.
2. Tsuneya Ando. *Theory of Electronic States and Transport in Carbon Nanotubes*. Journal of the Physical Society of Japan. 74, 777-817, 2005.

Single-walled carbon nanotubes (SWNT) are hollow cylinders with walls comprised of sp^2-bonded carbon atoms in a honeycomb lattice (Figure 14a). The lattice structure is equivalent to folding a graphene sheet over on itself, as illustrated in Figure 14b. The particular lattice arrangement and resulting nanotube diameter depend on the corresponding perpendicular vectors in the graphene plane, the axial vector T (parallel to the length of the nanotube) and the radial vector C_h (wrapping the circumference of the nanotube). Together these two vectors defining the nanotube structure are called the nanotube 'chirality.' The radial vector is generally expressed as sum of the graphene unit cell basis vectors a_1 and a_2 with integer coefficients, $C_h = na_1 + ma_2$. The chirality is therefore generally expressed by the coefficient set (n,m). As the name implies, many such vectors produce chiral enantiomer structures.

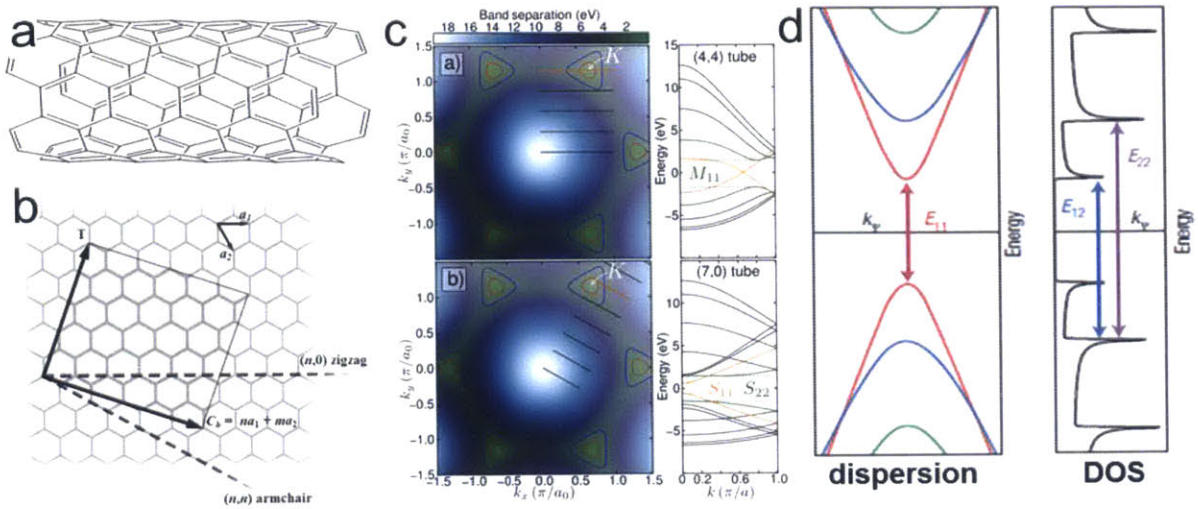


Figure 14. (a) Cartoon illustration of a SWNT, (6,6) chirality. (b) Cartoon of graphene lattice (hexagons) with arbitrary nanotube chiral vectors T and C_h, axial and radial, respectively, overlaid. Rolling the graphene sheet along C_h, head-to-tail, produces the SWNT lattice. The box formed by those vectors (shaded grey) forms the unit cell of the associated chirality nanotube when folded. C_h can be exressed as integer values of the graphene lattice basis vectors a_1 and a_2, as in the inset expression. The coefficients (n,m) therefore define the nanotube unit cell, or 'chirality.' Two particular chiral families, zig-zag and arm-chair nanotubes, are indicated. (c) Graphene dispersion surface (colormap) with dirac cone (K-point) indicated. A given nanotube chirality's dispersion curve is, under the zone-folding approximation, a series of slices out of the graphene dispersion surface, as described in the main text. Slices for two nanotube chiralities, (4,4) and (7,0), are produced here as an illustration, with the resulting one-dimensional cross-sections for each slice plotted on the right. (d) Illustration of nanotube density of states. The dispersion curve (left) produced from the zone-folding approximation (c) has parabolic bands resulting from slices through the dirac cone. The extrema of those parabolas produce corresponding Van Hove singularities in the Density of States (DOS) at those energies (right).

The electronic structure of the nanotube can to first order be derived directly from the graphene dispersion surface using the 'zone-folding' approximation. In folding the graphene sheet into the nanotube, the space of stationary electronic states becomes confined to the more limited periodicity. Graphene electronic states with wavevectors $k_{\parallel}$ along the nanotube chirality's axial vector are, under this approximation, preserved; the nuclear potential periodicity they are exposed to in the graphene lattice and nanotube lattice are, again neglecting curvature effects, the same. Almost all of the electron states with wavevectors $k_{\perp}$ along the radial vector however no longer exist because the nanotube has been made spatially finite in that direction; only states the a periodicity that is a harmonic of the circumference of the nanotube can persist at equilibrium. That discretizes the allowed $k_{\perp}$ into a small finite set. The carbon nanotube therefore essentially has a one-dimensional dispersion curve $E(k_{\parallel})$ comprised of a band for each discrete $k_{\perp}$ (Figure 14d). Each band is a slice out of the graphene dispersion surface along $k_{\parallel}$ (Figure 14c). The dispersion therefore depends exquisitely on the nanotube chirality.

As slices out of the graphene dispersion surface, these one-dimensional bands derive their character near the Fermi level from the Dirac cone in graphene. When the chirality is such that $k_{\parallel}$ passes through the center of the Dirac cone, the nanotube has metallic character corresponding to the behavior of

graphene. This occurs at chiralities where $(n - m)/3$ is an integer. Otherwise, the zone-folding slices intersect the sides of the cone, yielding parabolas in $E(k_{//})$ (Figure 14d). These nanotubes therefore have semi-conducting character with the corresponding band gap.

As usual the density of electronic states can be constructed from the dispersion curve; in one dimension it is simply proportional to the inverse derivative of the dispersion curve. The parabolas thereby generate their usual Van Hove singularities – sharp and extreme but finite spikes in the density of states at the extrema of the parabola (Figure 14d). Transitions between these singularities dominate the optical excitation behavior of nanotubes as explained in the next section.

While the zone-folding approximation is strong when treating large diameter nanotubes ($\gtrsim$ 1.5 nm), improved accuracy for smaller diameter tubes requires accounting for curvature effects. First, strain from bending the otherwise flat sp^2 lattice changes the carbon-carbon bond distance, which differs from atom to atom as there is no axial, only radial, strain. Second, the proximity of carbon atoms *via* the interior of the nanotube causes repulsion of electron density internally. The electron cloud therefore projects outward from the nanotube, particularly at diameters well below a nanometer. This effect can be treated as rehybridization of the σ and π orbitals, as the curvature makes them no longer orthogonal. Including these factors as perturbations on the zeroth-order zone-folding approximation, high accuracy has been demonstrated in theoretical prediction of nanotube electronic structure, such as *Tsuneya Ando*.[96]

Carbon Nanotube Optical Absorption

For insight into the optical properties of carbon nanotubes beyond the content in this Introduction, please refer to the following publications:

1. Susan Reich, Christian Thomsen, and Janina Maultzsch. *Carbon nanotubes: basic concepts and physical properties*. Wiley-VCH, Weinheim ; Cambridge, 2004.
2. Tsuneya Ando. *Theory of Electronic States and Transport in Carbon Nanotubes*. Journal of the Physical Society of Japan. 74, 777-817, 2005.

The reduced dimensionality of SWNT (*i.e.* rods) combined with the low atomic number of the constituent atoms (carbon) greatly reduces the screening of any localized charges as compared to metal solids. As a result, there exist excited exciton states in SWNT that are more stable than the excited crystal stationary states (*i.e.* an electron and hole at the band edges); the low screening increases the Coulomb potential gain by localizing the positive and negative charge. The difference in energy, termed the binding energy, significantly exceeds kT at room temperature. As it is due to spatial confinement and charge screening, the magnitude of this effect is almost entirely dependent on the nanotube diameter, with negligible contributions from chirality. Maultzsch *et al.* experimentally verified for example that binding energy falls with diameter.[98] Optical excitations of carbon nanotubes therefore generate excitons, rather than free carriers. As a superposition of several similar-energy crystal states, the dispersion relation of excitons approximately matches that of the stationary states. Because of the above-mentioned Van Hove singularities at the SWNT band edges, semiconducting nanotubes exhibit strong optical absorption peaks corresponding to the energy of the band edge-to-band edge transitions (Figure 15). While all band edges appear to align in $k_{\parallel}$ as plotted in Figure 14d, they have different $k_{\perp}$ and therefore inter-band optical transitions are forbidden.[xvi] Absorption peaks therefore appear at each intra-band pair of Van Hove singularities, with energies labelled E_{1u}, E_{2u}, and so on for increasingly high-energy bands. The band gap of the chirality therefore determines the peaks of its absorption spectrum (Figure 15).

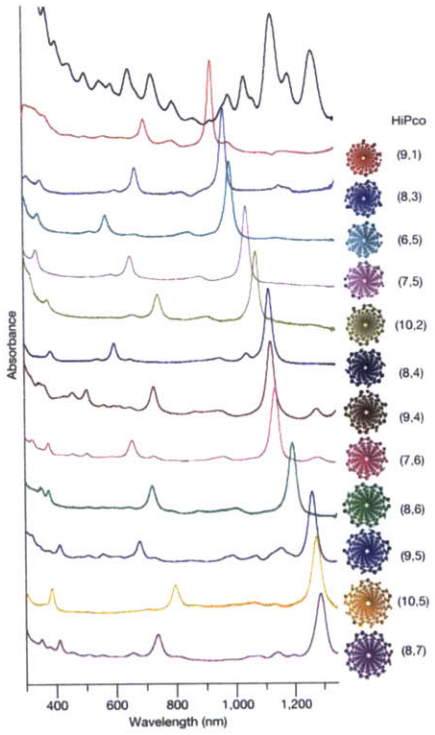


Figure 15. Optical absorption of carbon nanotubes, reproduced with permission from Tu *et al.*[97] Black line is the absorption spectrum of a solution-phase mixture of nanotubes. Subsequent lines are the absorption spectra of single-chirality solutions, showing peaks associated with the E_{1u} transitions of each isolated chirality.

As the described optical absorption corresponds to electronic transitions in $k_{\parallel}$, the transition dipoles are parallel to the axis of the nanotube. Therefore only light with a polarization component parallel to the nanotube axis is absorbed, giving carbon nanotubes anisotropic optical absorption. Independent radial excitations also exist, but do not reflect a band structure (Figure 16, below).

[xvi] To first order. Indirect band-gap transitions can be realized *via* coupling to phonon modes.

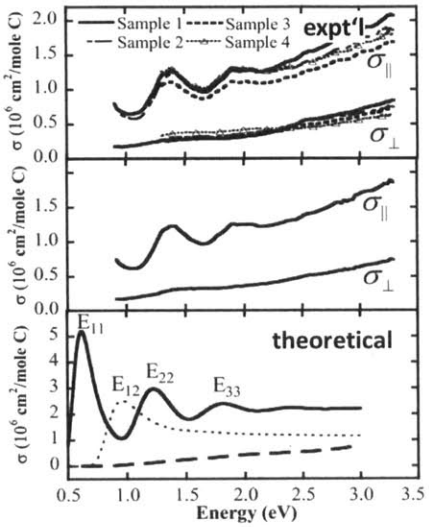


Figure 16. Polarization-dependent absorption cross-section, reproduced with permission from Islam *et al.*[99] The top panel is the absorption cross-section measured with light polarized parallel to ($\sigma_{\parallel}$) and perpendicular to ($\sigma_{\perp}$) the longitudinal axis of (6,5) nanotubes. The middle panel is the average of the four samples in the top panel. The bottom panel is a theoretical prediction of the absorption cross-section in the two orientations.

Integrated over the solar spectrum, which is of interest to our work in Chapters V and VI, the total radial absorptivity of nanotubes is roughly $1/5^{th}$ of the axial absorptivity.

While semi-conducting nanotubes can absorb light of any frequency greater than the band gap energy, coupling of excitons to nanotube crystal phonon modes is very strong such that non-radiative relaxation of hot excitons to the bandgap occurs within a few femtoseconds. Therefore in all but the most extreme experimental circumstances, excitons are almost always at the band-gap energy regardless of the excitation that produced them.

Carbon Nanotube Exciton Transport

For insight into exciton states in carbon nanotubes beyond the content in this Introduction, please refer to the following publications:

1. Susan Reich, Christian Thomsen, and Janina Maultzsch. *Carbon nanotubes : basic concepts and physical properties*. Wiley-VCH, Weinheim ; Cambridge, 2004.
2. Tsuneya Ando. *Theory of Electronic States and Transport in Carbon Nanotubes*. Journal of the Physical Society of Japan. 74, 777-817, 2005.
3. Phaedon Avouris, Marcus Freitag, and Vasili Perebeinos. *Carbon-nanotube photonics and optoelectronics*. Nature Photonics 2, 341, 2008.

Once generated, excitons on single walled carbon nanotubes can undergo a variety of processes in their lifetimes. As mentioned in the previous section, they non-radiatively relax to the band gap within a few femtoseconds. Once generated, excitons translate along the nanotube axis. They do so diffusively *via* stochastic scattering off of crystal phonon modes. Their ballistic translation length, *i.e.* their translation mean free path, remains unknown because of the challenges of deconvoluting exciton size and mean free path.[100,101] Measurements of exciton diffusion coefficients have varied many orders of magnitude, from less than 1 cm^2/s (Cognet *et al.* and Luer *et al.*) [102,103] and 7.5-10.7 cm^2/s (Crochet *et al.* and

Hertel *et al.*)[104,105] to 150-350 cm^2/s (Ruzicka *et al.* and Anderrson *et al.*)[106,107]. Comparison of the environments of those measurements, combined with comparison of other environmentally sensitive investigations,[104,108-111] strongly suggests to the author that this variation is due to influence of the local environment on exciton-phonon scattering. In particular, lower diffusivity values (and associated diffusion lengths) have been observed uniformly observed in surfactant-wrapped SWNT in solution phase, while SWNT in air and vacuum exhibit the higher diffusivity values. Changing the surrounding dielectric constant intuitively should impact the dispersion and binding energy of the exciton states and perhaps also their coupling to the crystal vibrational modes.

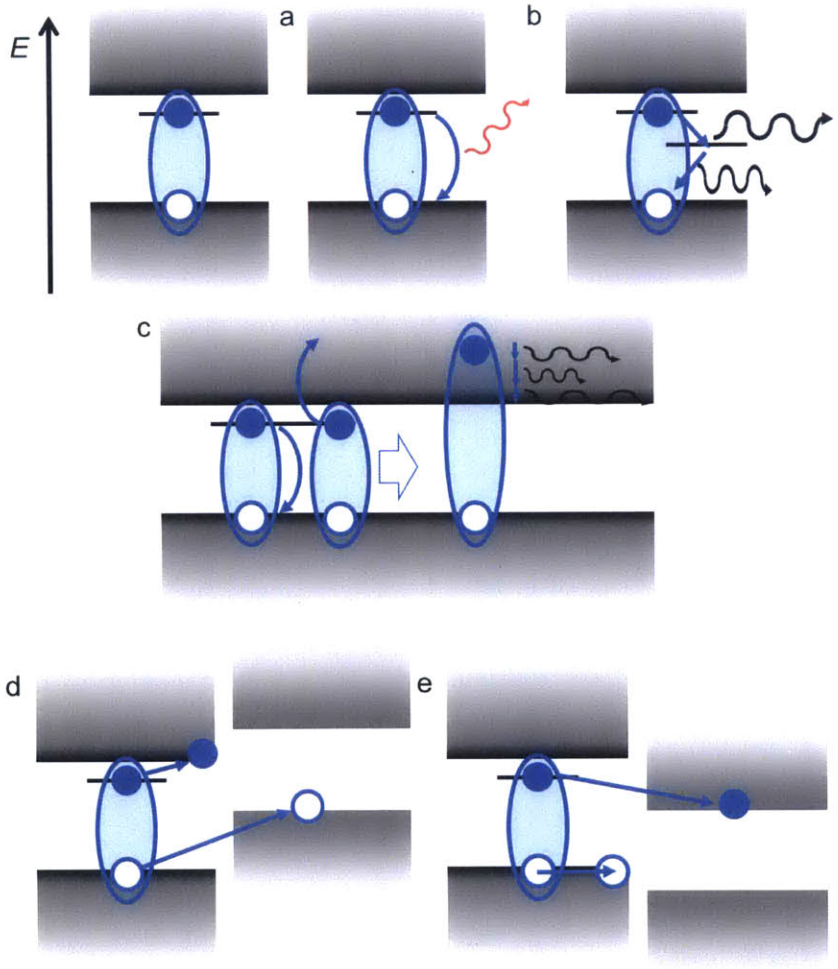


Figure 17. Cartoons of exciton decay processes in a semiconducting carbon nanotube. Greyscale bands are the valence and conduction free carrier bands (electron states). The line beneath the conduction band represents the energy of the ground state exciton, with the difference in energy between the exciton and the conduction band edge being the binding energy. The blue filled and hollow dots represent an electron and hole, respectively, with the oval encompassing the two representing their coupling into an exciton quasiparticle. (a) Radiative relaxation of the exciton to produce a photon. (b) Non-radiative decay of the exciton *via* an interband state, such as introduced by an impurity adsorbed on the nanotube, in the nanotube lattice, or at the nanotube ends, producing phonon mode excitations. (c) Auger recombination of two excitons, briefly producing a single hot exciton which then rapidly decays back to the band edge non-radiatively, producing phonon excitations. (d,e) dissociation of the exciton at a

Type-II heterojunction, where the band edges align to stabilize either the electron (e) or the hole (d) by more energy than the exciton binding energy.

While non-radiative pathways to decay are low probability (the exciton must simultaneously excite several phonon modes), excitons at the band edge do not persist indefinitely. Excitons can radiatively decay, emitting a photon with the band gap energy. Relative to most excitonic materials, a long radiative decay lifetime is usually measured, of order 0.5 ns.

Non-radiative decay pathways also exist. Local impurities, such as adsorbed Nitrogen Oxide or sp^3 carbon lattice defects, can introduce inter-band electronic states that allow relaxation *via* coupled excitation of only one or two phonon modes at a time. The nanotube ends are often terminated with carboxyl groups or other oxidized carbon states that can introduce similar inter-band electronic states. Two excitons can also scatter off each other in Auger recombination, relaxing one exciton to ground state while doubly exciting the other. Because of the aforementioned rapid non-radiative decay of excited excitons to the band edge state, the surviving exciton soon returns to its original energy and the net effect is simply annihilating one of them.

When the nanotube is brought into contact with another semi-conducting material with bands that appropriately align with those of the nanotube (*i.e.* allowing SWNT valence holes to be injected into the material but not conduction electrons, or *vice versa*), the exciton can dissociate via Type-II exciton dissociation (Figure 17d,e), breaking into its constituent electron and hole charge carriers. When the excited electron (Figure 17e) or hole (Figure 17d), depending on the electrode, produced by dissociation of the exciton can relax in the electrode to produce an energy savings the same or greater than the binding energy of the exciton, there is a significant probability of exciton dissociation. The electron or hole then resides on the electrode after relaxing, as it is energetically unable to return to the nanotube, and the opposite charge carrier remains on the nanotube. The appropriate material depends on the nanotube chirality's binding energy, but for the most chirality, (6,5), effective Type-II materials have been discovered, including Fullerene (typically in the polymerized form, Phenyl-C_{61}-butyric acid methyl ester or PCBM), conductive polymers (especially Poly(3-hexylthiophene-2,5-diyl) or P3HT), and n-doped Silicon.

When nanotubes are in close proximity, such as in bundles (see next section), excitons can transfer from one nanotube to another. The first direct evidence for this effect was shown, by Tan *et al.*, by mapping the emission intensity of a mixed-chirality nanotube solution as a function of emission frequency *versus* the excitation frequency (by sweeping the excitation), producing a two-dimensional excitation-emission plot such as in Figure 18.[112] This experiment shows emission peaks corresponding to one chirality appearing at the excitation peaks of a different chirality of lower band gap energy, indicating the excitons in the high-bandgap nanotubes were able to transfer to the lower-bandgap

nanotube and relax (prohibiting return) before they radiatively decayed. Until recently, the transfer mechanism was presumed to be a simple electron tunneling process. Recent evidence from *Crochet et al.* however suggests that exciton states may actually delocalize across bundles of even only two nanotubes;[113] they showed that bundling increased the rate of exciton non-radiative scattering at impurities proportional to the size of the bundle. By delocalizing across multiple nanotubes, the excitons were able to scatter off of impurities on any of the tube, not just one of them, increasing the number of impurities they were effectively exposed to.

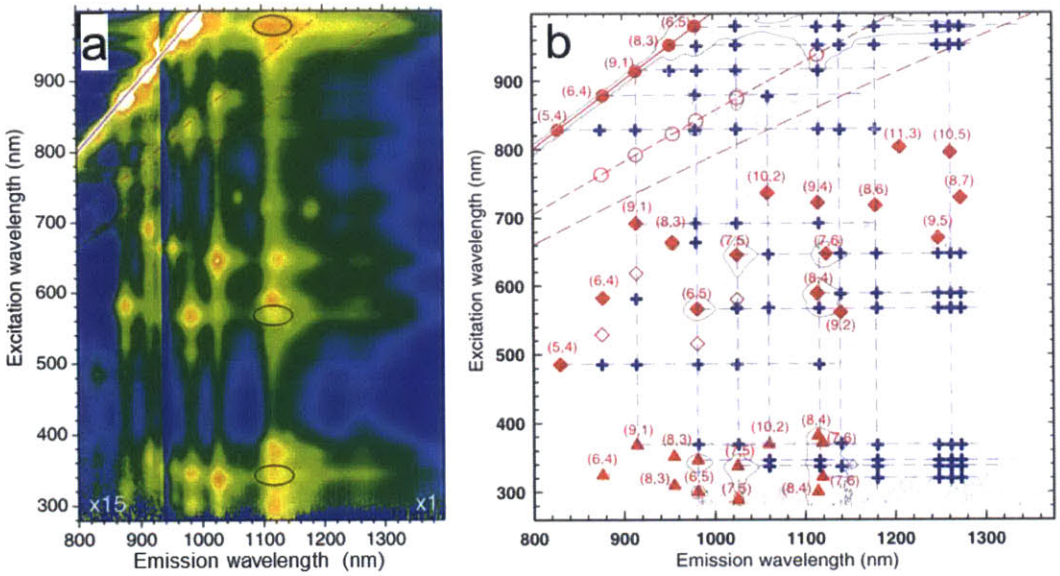


Figure 18. First direct evidence of exciton energy transfer between nanotubes, from Tan *et al.*[112] (a) two-dimensional plot of stimulated emission intensity (colormap) *versus* excitation and emission wavelengths. (b) peaks extracted from (a), showing that excitations at frequencies resonant with large bandgap chiralities are yielding emissions at frequencies resonant with lower bandgap chiralities (blue crosses), implying that excitons are transferring prior to radiative decay.

Carbon Nanotube Photovoltaics

For insight into carbon nanotube photovoltaics beyond the content in this Introduction, please refer to the following publications:

1. Brian A. Gregg. *Excitonic Solar Cells*. The Journal of Physical Chemistry B. 107, 20, 4688-4698, 2003.
2. Phaedon Avouris, Marcus Freitag, and Vasili Perebeinos. *Carbon-nanotube photonics and optoelectronics*. Nature Photonics 2, 341, 2008.
3. Michael S. Arnold, Jeffrey L. Blackburn, Jared J. Crochet, Stephen K. Doorn, Juan G. Duque, Aditya Mohite, Hagan Telg. *Recent developments in the photophysics of single-walled carbon nanotubes for their use as active and passive material elements in thin film photovoltaics*. Physical Chemistry Chemical Physics 15, 14896-14918, 2013.

Photovoltaic energy is generated from carbon nanotubes by pairing them with one or two Type-II exciton dissociating interfaces, as described above. When light is absorbed any generated excitons that arrive at a dissociating electrode have an opportunity to split into an electron and a hole, creating an electric field from the Coulomb attraction between the two and thereby providing the opportunity to generate energy. At this interface, as with any photovoltaic, an electric potential drop is generated from the charge separation. As more excitons are generated and dissociate, the potential grows. This electric field, which lowers the energy of electrons (or holes) in the electrode relative to the SWNT, eventually creates an equilibrium at the interface, where the electron stabilization from dissociation is equal to the exciton binding energy, halting dissociation. The potential at that steady state is the open circuit voltage. Completing the circuit by connecting the electrode to the SWNT film either as a short or through a load creates a steady state current, relieving the charge buildup. In this manner, energy can be harvested from the dissociation of excitons at the heterojunction and drift-diffusion of the resulting free carriers through the materials to metal contacts. Heterojunctions can be planar, where the SWNT film and dissociating material are parallel plates, or bulk, where the two materials are intermixed.

Networks of carbon nanotubes are an appealing material for photovoltaic energy generation. Through continual refinement in design and manufacture, silicon p-n junction solar cells (SCs) have achieved un-subsidized cost parity with the grid in several emerging and advanced economies.[114] The gains of refinement are slowing down however, with the remaining returns diminished to insufficiency for achieving global energy parity.[114] To increase the solar-sourced share of global energy production, the next generation of photovoltaic technologies aim to reduce cost and increase the space of suitable conditions for solar energy collection. These cost and performance requirements have made semiconducting single-walled carbon nanotubes (SWNT) attractive as photo-absorbers for near-infrared photovoltaic (nIR PV) applications.[115-121] Economically, their solution-process-ability,[xvii] earth-abundant source materials, and recently scale-able fabrication and purification raise the potential for low cost manufacture.[117,119,120,122-124] In performance properties, SWNT PVs can augment the photo-conversion efficiencies of conventional visible-PV systems by absorbing the nIR wavelengths that comprise approximately 22% of the solar spectrum but fall within a silicon PV bandgap. Low unit-operations costs particularly enable such application by making stacked independent PV layers (four-terminal SCs) feasible, *versus* the constraints of one-step fabrication (*i.e.* lattice, bandgap, and workfunction matching constraints) that have challenged tandem multi-junction (two-terminal) SCs.[125]

[xvii] Solution process-ability can greatly reduce manufacturing costs because the material can be printed under ambient conditions (rather than vacuum growth or deposition), continuously (*versus* in batches), and without the lattice and energy matching constraints of exotic semiconductor crystal alternatives for supplementing Si solar cells.

SWNT are also attractive relative to some organic and inorganic PV materials in their exceptional exciton and free carrier diffusivity and mobility, exceptional absorption cross-sections, and resistance to oxidation by water and oxygen.[106,117,119,120,126-128] Over the past five years these properties have motivated the development of the first photovoltaics with SWNT active layers. To date, single-walled nanotube (SWNT) SCs with 0.01-3.1% external quantum efficiency (EQE) have been developed employing single chiralities,[116,118] mixed semiconducting chiralities,[115,117] and mixed semiconducting and metal chiralities[119,120] in bulk heterojunction[117,121] and planar heterojunction[115,116,118] configurations using polymers[129] and fullerenes[115-118] for exciton dissociation.

This investment in proof of concept devices has been important in establishing the technology's feasibility, but has not provided insight into design. The parametric space of SWNT networks, due to their awkward nanometer-scale heterogeneity, is too large to investigate empirically. For example, basic questions such as the effect of *e.g.* nanotube density, chirality, orientation, film thickness, or impurity concentration remained unanswered. The extreme non-linearity of the dependence of film performance on such properties additionally confounds simple intuition. Even the full set of variables describing a carbon nanotube solar cell have not been identified, let alone their optimization. This gap in knowledge has obscured what research, especially into nanotube purification and film casting, is necessary to advance the technology. Only one attempt has been made to define and elucidate the effect of a design variable, specifically thickness,[115] but the observed behavior is limited to films with all other properties, such as nanotube orientation and density, which were not measured, held constant. In Chapter V, we address this shortcoming by taking a modelling approach. By deriving a model of steady-state SWNT network photovoltaic efficiency directly from single- and paired-nanotube properties and physical processes, we are able to predict the dependence of efficiency on arbitrary values of nanotube properties, including length, orientation, film thickness, and chirality. Optimizing the model isolates the critical design variables, and predicts the highest priority research progress necessary to raise SWNT solar cell efficiency to technologically viable levels. In Chapter V we solve the model numerically to optimize efficiency over the parametric space; while that produces film design rules however, it does not connect them explicitly to the intrinsic properties of the film *e.g.* exciton diffusion length. In Chapter VI, we develop an analytically solvent approximation to the model that reveals that connection.

I. SURFACE TRANSPORT MODEL OF NANOWIRE METAL ELECTROMIGRATION

Summary of Findings

As discussed in the Introduction, the mechanism of electromigration in micron-diameter and larger wires has been thoroughly understood since James Black developed the widely verified bulk vacancy drift model in 1969. As miniaturization has progressed however the semiconductor industry has had to add empirical correction factors to the model, and the recent use of electromigration of nanowires for tunnel junction formation has been unable to exploit the mechanism to improve yield. These observations suggest that the predominant mechanism of metal electromigration changes when wire diameters fall to tens or hundreds of nanometers.

In this work we propose that as a wire shrinks to nanometer dimensions the kinetics of electromigration shift from being limited by bulk vacancy transport, as in the Black model,[9] to being limited by surface transport. The barrier to surface diffusion of metal atoms is in general substantially lower, by as much as ten orders of magnitude, than bulk vacancy diffusion;[xviii] we posit that as the wire diameter gets smaller and the surface area to volume ratio increases, the increase in relative population of surface atoms to bulk vacancies shifts the kinetically dominant electromigration pathway correspondingly. From this principle we derive a model of nanowire electromigration and the resulting change in nanowire diameter with time. We validate this model against experimental measurements from *Karim et al.* of the electromigrative failure of Au nanowires varying in diameter from 80 nm to 700 nm.[94,132] We find that a bulk vacancy transport model cannot fit the observed failure times and time-dependent resistance evolutions, providing the first quantitative verification that the Black model breaks down at sub-micron diameters. In contrast, our proposed surface transport model matches the data entirely, with deviation negligible compared to experimental error. The time to failure predicted by a surface transport mechanism is shown to have an analogous form to the time-to-failure predicted by Black, but with a change in the power dependence on current density from -2 to -1; such a correction in the power dependence is already commonly applied in the empirical literature,[10] further supporting our surface transport theory and providing a physical explanation for those experimental findings. In providing access to the mechanism of nanowire electromigration, this model could enable prediction and optimization of the process for desired outcomes. We show that for these nanowires the critical failure current at constant ramp scales with initial wire radius to the 3/2 power. Applying to the case where the electromigration process is feedback controlled according to the process of *Strachan et al.*,[8] further

---

[xviii] On a Au(111) surface for example, the surface self-diffusion at room temperature has been measured of order 10^{-15} cm^2/s or higher, *versus* order 10^{-40} cm^2/s for bulk vacancy diffusion.[130,131]

explored in Chapter II, the surface transport model reproduces all of the observed characteristic features *e.g.* the 'turnaround point' and 'shark fins.' In particular the surface transport model accounts for the precipitous (accelerating) failure of nanowires; as the radius reduces, the surface area to volume ratio increases such that the rate of radius reduction also increases, leading to an intrinsic positive feedback loop. The original method of *Strachan et al.* does not prescribe any parameters for the control algorithm, but we predict and then apply the parameterization that governs the process; it is the ratio of the current or voltage attenuation to the resistance threshold. In Chapter II, we use this model to gain greater control over the process of nanowire electromigration for tunnel junction formation by designing a new controller based on its predictions.

Surface Transport Model

To describe the electromigration process we adopted a surface transport mechanism, anticipating that for small diameter wires translocation barriers are lower and available sites more numerous than vacancy drift in the bulk crystal. This hypothesis is later verified by deriving equivalent expressions for bulk diffusion and finding that, unlike the surface transport model, they are inconsistent with experimental results (Figure 21). For our geometry and driving force application we focused on cylindrical metal nanowires with time-varying current forcing, but the treatment is analogous for other geometries or under applied voltage. An electron scattering treatment where surface atoms interact with the current flux through an empirical scattering cross-section provides a convenient representation of the wind force atom dislocation rate F_S,

$$F_s = F_0(r,t) \cdot \sigma_{cs} \cdot N(r) \tag{7}$$

where $F_0(r,t)$ is the fraction of the momentum imparted by the incident current flux in the surface region as a function of time t in a wire of radius r, σ_{CS} is the atom electron scattering cross section, $N(r)$ is the number of atoms in the surface region, and F_S is the resulting number of scattering events per time. Experimentally it has been observed that well before breakage 'necking' occurs where some region of the nanowire is thinned, often with a downstream extrusion forming from the lost material.[22,28,83] We approximate that the specific shape of the necking region can be reduced to some equivalent symmetric representation, cylindrical in this case, and that the length of the hypothetical region δ is constant (Figure 19).

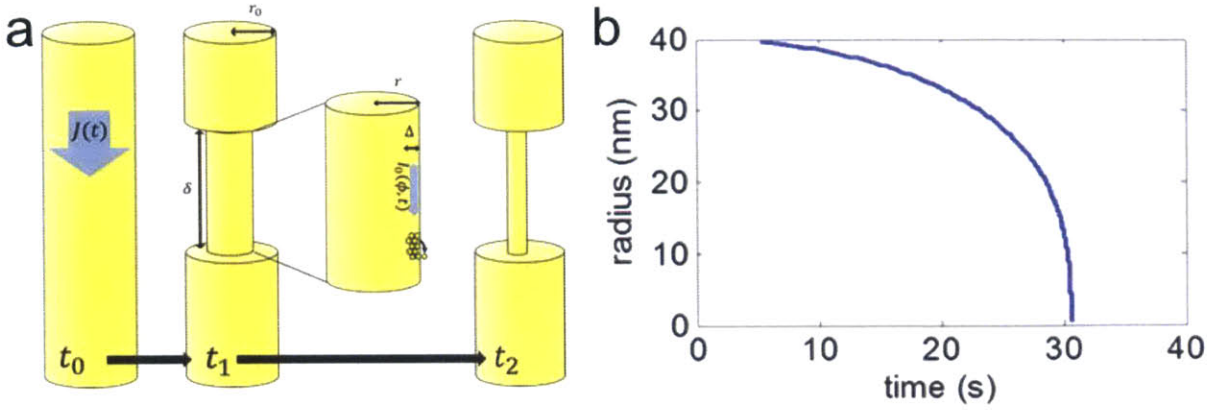


Figure 19. a) Cartoon of the model geometry. b) Predicted nanowire radius versus time trace for a NW of initial radius of 40 nm (using equation 4).

For a fixed length of the necking region δ the number of Au atoms in the surface region is $N = \pi\delta n_{Au}(2r\Delta - \Delta^2)$, where r is the radius of the neck, Δ is the depth of the surface region under consideration, one atomic layer, and n_{Au} is the number density of bulk gold. In this surface region we have a resulting electron flux of $I/q\pi r^2$, where q is the elementary charge. If we define a new empirical cross-section ϕ_{CS} as the product of the electron scattering cross section σ_{CS} and the fraction of scattering events that yield atomic loss, we have a simple expression for atomic loss in the neck region

$$\frac{dN}{dt} = -\phi_{CS}\left[\frac{F}{q\pi r^2}\right][\pi\delta n_{au}(2r\Delta - \Delta^2)] = -\delta\phi_{CS}n_{au}\frac{I}{q}\frac{(2r\Delta - \Delta^2)}{r^2} \tag{8}$$

Substituting in the number of atoms in the necking region $N = \pi r^2\delta \cdot n_{Au}$ we have

$$\frac{dr}{dt} = -\gamma I\frac{(2r\Delta - \Delta^2)}{r^3} \qquad \gamma \equiv \frac{\phi_{CS}}{4\pi e}. \tag{9}$$

For a constant current ramp at rate α A/s ($I = \alpha t$) we can integrate analytically,

$$\begin{aligned} &\frac{1}{2}\gamma\alpha\Delta \cdot t^2 + k_0 = -\Delta^3\ln(2r - \Delta) - 2\Delta^2 r - 2\Delta r^2 - \frac{8}{3}r^3 \\ &k_0 \equiv -\Delta^3\ln(2r_0 - \Delta) - 2\Delta^2 r_0 - 2\Delta {r_0}^2 - \frac{8}{3}r_0^3. \end{aligned} \tag{10}$$

Choosing representative parameters, we can qualitatively observe a precipitously accelerating reduction in radius with time (Figure 19b), as we expect ubiquitously from experimental results.[23,133]

Comparison to micron-scale transport

The relationship between our proposed surface transport mechanism of metal electromigration and the Black model bears explicit consideration. As related in the Introduction, section *The Black equation and electromigration at micron scales*, Black in 1969 proposed that the dominant pathway of

metal electromigration is *via* vacancy drift in the bulk lattice.[9] In the intervening years that bulk vacancy mediated transport model has been broadly verified in experiments, and his celebrated equation for the predicted time-to-failure of wires due to electromigration has become the standard in the semiconductor industry. As fabrication techniques have improved to allow the ready production of metal wires tens or hundreds of nanometers in diameter however, experimental results have deviated from the Black model, requiring empirical correction factors.[10] Later in this Chapter and in Chapter II we both quantify that deviation and show that there is a parallel breakdown in the qualitative characteristics of bulk-vacancy-mediated electromigration. Our surface transport mechanism seeks to explain this divergence. We are proposing that as the wire diameter shrinks to make the surface are to volume ratio non-negligible, the dominant electromigrative transport pathway shifts to drift of surface atoms as described above. If true, as our validations in this Chapter and Chapter II suggest, this hypothesis implies two electromigration regimes with different dominant kinetics. For micron and larger diameter wires, the Black proposal of drift of bulk vacancies irrefutably holds; for nanometer diameter wires, surface transport dominates.

In Chapter II, section '*Time-to-failure in Nanowires*', we derive the surface transport mechanism analogue of Black's time-to-failure equation. This analytic comparison bears reproduction here. Black's time to failure t_f of a wire due to electromigration under a fixed bias, which from a Kinetics perspective is the inverse of the mean rate of electromigration over the lifetime of a wire, is[9]

$$t_f = Ae^{\frac{E_V}{kT}}j^{-2} \tag{11}$$

where A is an empirical prefactor, E_V is the activation barrier to vacancy diffusion, k is the Boltzmann constant, T is temperature, and j is the initial current density in the wire. The empirical corrections described above have corrected (11) by empirically adjusting the power of j to between -1 and -2, justifying the change as material-specific non-ideality.

The equivalent expression in our surface transport model, which we propose holds for nanometer-diameter wires, is

$$t_f = Ae^{\frac{E_S}{kT}}j^{-1}, \tag{12}$$

where E_s is the activation barrier to vacancy diffusion. The expression is identical in form, with the differences being that A and E have different values (we show, for example, that A has a dependence on the wire initial radius) and, most notably, that current density has a linear rather than quadratic influence on the rate of electromigration.

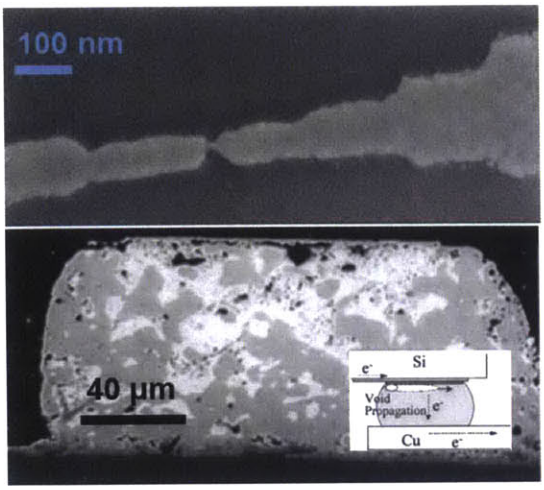


Figure 20. Illustration of necking *versus* void formation in nanowires *versus* micron-diameter wires. (top) nanowire produced according to methods in Chapter II, showing that electromigration led to the formation of a neck as similarly observed by *Taychatanapat et al.*[22]. (bottom) Figure reproduced from *Yeh et al.* with permission showing void formation during electromigration.[18]

Our regime change hypothesis, with the dominant electromigration mechanism changing from bulk vacancy transport to surface transport as the wire diameter shrinks to the nanometer scale, can therefore explain the empirical corrections to the Black equation.

The surface transport mechanism is also consistent with consistently observed qualitative differences between micro-wire and nano-wire electromigration. In particular, electromigration in micron-size wires characteristically yields cavity ('void') formation, consistent with bulk transport and aggregation of vacancies.[10,18,134,135] In contrast however, nanowires ubiquitously exhibit thinning ('necking'),[22,83] as described above and in the Introduction (section *features of electromigrative formation of tunnel junctions*), consistent with mass transport being in the surface rather than the bulk. In later work, Chapter III, we find that another hallmark of electromigration at the micron scale, failure at interfaces, breaks down at nanometer dimensions.

Model Limitations

We show below that despite its simplicity the model quantitatively and qualitatively predicts the empirical observations of electromigration in nanowires. For future evaluation and extensibility however it is useful to note a few particular limitations. First, our model is isothermal *i.e.* we have neglected Joule heating. Temperature increases of 100-300K have been proposed,[22,87,93] yielding, if true, higher mobility of Au atoms, higher resistivity of the NW, and possibly a temperature-gradient-driven net diffusive flux. Some researchers have used this mechanism to account for the failure of their nanowires.[87,93] A more general version of our surface transport model should include thermal effects. Nevertheless, in the experimental system below surface transport alone reproduces the observed results, without consideration of a change in temperature.

Second, our reduced representation of the necking region as symmetric and fixed in length is not, and is not intended to be, precisely observed. The actual neck region could be asymmetric *i.e.* hold a dependence of r on position, for example bowing in the middle providing a higher current density over a smaller region. Additionally, the necking region length could evolve over the course of the process – in

fact asymmetric behavior could be approximated as a reduction in the length of effective neck, δ, as the highest current density region shows the dominant kinetics. Our simplified reduction of these effects has proven sufficient to describe the current-voltage behavior of the formation process, but a finer examination of δ and ϕ should take them into account. As is, our model does not predict actual wire morphology. Doing so would provide a clearer physical picture, but does not impact the predictive ability we demonstrate.

Finally, the actual failure at the end of the process we expect to be stochastic (see Chapter II), as fewer than a hundred atoms form the final electrodes at failure. As a result conductivity behavior with time at the end of the process cannot be well described with our deterministic model. For a cutoff of ~100 atoms per nanometer of neck length, this corresponds to a maximum valid radius of ~1.5 nm for Au. By the same token, the model cannot account for the shape and size of gap formed. Nevertheless, our deterministic model describes the necking process leading up to the final break, providing anticipation and adjustment of the speed at which the radius reduces and the voltage applied at failure.

Experimental Data Set

To validate our proposed model we used the experimental results of *Karim et al.*[94] We chose this work because they report experimental data from a well-defined and prototypical[xix] process – increasing the current flux at a fixed rate through a cylindrical NW until breakage – and because they provide a broad data set including several important vectors for extracting model parameters including initial resistance, current and voltage with time, and failure current information all as functions of a range of well-characterized initial radii.[94,132]

They used ion beam track-etched polycarbonate membranes as templates to form cylindrical Au nanowires using electrochemical metal deposition. The resulting nanowires were left in the membrane and were 30 μm long and 40-360 nm in radius. Their radii were verified by SEM and conductometry. Single nanowires were contacted with an upper Cu electrode at room temperature, and contact resistances were estimated to be of order 10 Ω. They then applied increasing current to the NWs at a fixed rate of 0.1 mA/s until breakage occurred, measuring the required voltage over this time.[94,132]

[xix] As explored in Chapter II, a similar electromigration process is used to form tunnel junctions from nanowires.

Predicting critical failure current

Karim et al. ramped a current applied to their NWs at a fixed rate of 0.1 mA/s until breakage occurred, observing a trend in the critical failure current density *versus* initial wire radius.[94] We evaluated our model by applying it to this data, optimizing the scattering cross-section for fit.[xx] Solving equation (10) for current and substituting r_c (critical failure radius) for r,

$$I_{crit} = \left[\frac{2\alpha}{\gamma\Delta}(-\Delta^3 \ln(2r_c - \Delta) - 2\Delta^2 r_c - 2\Delta r_c^2 - 8r_c^3 - k_0)\right]^{\frac{1}{2}} \quad (13)$$

The result (Figure 21) shows strong agreement with the experimental data. We can benchmark the model by comparing to a pure power law regression (Figure 21 inset), finding them to agree to well within experimental error. In other words, the model error is negligible relative to experimental error. The log-log slope we predict, which is the order of the r dependence of the rate of radius change, is -0.5; this corresponds to a 3/2 power scaling between the critical current and initial radius (Equation (14)). The fit yields a ϕ_{CS} of 8.86×10^{-27} cm^2.

$$I_{crit} = A\,\pi r_0^{3/2} \quad (14)$$

Our extracted value for A in this system is $10^{29.17}$ A/m^2. We can further validate our surface transport premise by comparing to a bulk transport model, derived and optimized in the same fashion. Plotted in Figure 21 (inset), we can see that the resulting $-1/r$ necking rate dependence is unsuitable for describing the data. This finding provides quantitative evidence that the Black model, *i.e.* bulk vacancy dominated electromigration, does not hold for nanowires smaller than micron diameter.

[xx] They defined the critical failure point as when an arbitrarily high bias (20 V) was reached; we equivalently defined it as when the radius reached *4r*. The rapid reduction in radius at the end of the formation process introduces large tolerances in either definition, and for critical radii up to 10 nm the resulting scattering cross section is unchanged to within 3 orders of magnitude.

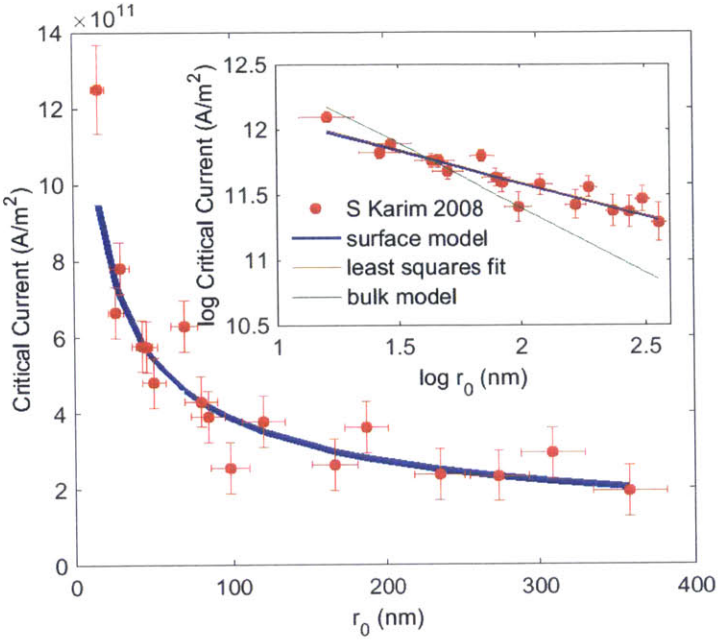


Figure 21. Observed critical failure current from *Karim et al.* compared to surface diffusion model predictions (Equation (13)). Data is linearized on a log-log plot (inset), showing that the model prediction is as strong as a power law regression. Comparison to an equivalent bulk model validates the surface diffusion treatment.

Note that while the data has a clear overall trend, there remains considerable variability in the observed failure current, as high as 10 mA, well beyond the reported measurement error bars. Uncontrolled experimental variables, *e.g.*, organic impurities at the NW surface, surface charging at the polymer interface, and variation in contact resistance, may provide this experimental scatter. This observation is relevant later when we reproduce the time-evolution traces from single experiments.

Predicting resistance evolution during necking

In order relate our $r(t)$ predictions to time-dependent observables, we require valid expressions for resistivity. It has been broadly demonstrated that a bulk Au resistivity – or even a modified but radius-independent resistivity – is insufficient to describe nanowire resistance.[132,136-141] In this size regime two effects in addition to phonon scattering strongly influence resistivity: the change in grain boundary scattering behavior from bulk due to smaller grain sizes, and the increased importance of specular carrier scattering off of surface states. Mayadas and Shatkes (MS) pioneered the commonly accepted behavior in the former case, expressed here for a cylinder:[132,136,137,139]

$$\frac{\rho_{Au}}{\rho_G} = 3\left[\frac{1}{3} - \frac{\alpha_G}{2} + {\alpha_G}^2 - {\alpha_G}^3 \ln\left(1+\frac{1}{\alpha}\right)\right] \qquad \alpha_G \equiv \frac{l_e}{D_g}\frac{R_G}{1-R_G} \tag{15}$$

where ρ_G is the resistivity due to enhanced grain boundary scattering, ρ_{Au} is the bulk Au resistivity, l_e is the mean free path of electrons (which we take to be 40 nm in Au at 298K), R_G is the reflectivity of the grain boundaries, and D_G is the mean grain diameter.

Fuchs and Sondheimer (FS), with later refinements for cylinders by Dingle, described the effect of surface scattering as dependent on the specularity of the material, p, with $p=0$ representing a completely diffuse surface scattering behavior and $p=1$ representing a totally reflective surface.[138,140,141] Two regimes are considered: when the radius is sufficiently above or below the electron mean free path. They are expressed in Equations (16) and (17) for a cylinder.

$$r \gg l_e$$

$$\frac{\rho_{Au}}{\rho_s} = 1 - \frac{3}{4k}(1-p) + \frac{3}{8k^3}(1-p)^2 \sum_{v=1}^{\infty} \frac{p^{v-1}}{v^2} \tag{16}$$

$$r \ll l_e$$

$$\frac{\rho_{Au}}{\rho_s} = \frac{1+p}{1-p}k - \frac{3k^2}{8}\left[\frac{1+4p+p^2}{(1-p)^2}\left(\ln\left(\frac{1}{k}\right) + 1.059\right) - (1-p)^2 \sum_{v=1}^{\infty} v^3 p^{v-1} \ln v - \frac{2k^3}{15}\frac{1+11p^2+p^3}{(1-p)^3}\right] \tag{17}$$

where $k \equiv r/l_e$ and v is the electron velocity.[138] In all cases we take the infinite series to convergence. For radii intermediate between the extreme limits we approximated the resistivity by calculating the value at that radius from both equations and averaging those two values weighted by the proximity to each, choosing demarcations of 60 nm and 20 nm.

Note that while FS scattering depends continuously on the NW radius, the MS grain scattering is only connected to the wire dimensions indirectly through the crystallization kinetics. The grain size and specularity have been shown to vary between experimental systems for the same material due to differences in grain, surface facets, and impurity levels resulting from fabrication methods.[132,136,137] For Au, values from 0.4 to 0.9 for R_G and 0 to 0.6 for p have been inferred from different experiments.[132,136]

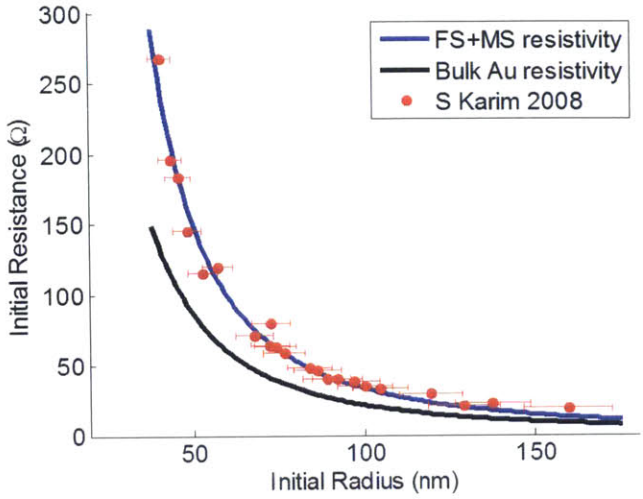


Figure 22. Nanowire resistance versus radius data from *Karim et al.* with a fit using FS and MS models of resistivity.

To extract the crystal properties of our experimental data set, we calculated the initial resistance for a NW of given radius using Matthiessen's rule and optimized p and α_G given *Karim et al.*'s measurement. As a first approximation, verified by the quality of the fit, we took the grain size to be constant across the entire diameter range. This yields an α_G of 0.22 and specularity of 0.36 with a resulting rms error in $R(r)$ of 8.7 Ω.[xxi] Figure 22 shows that with these parameter values the FS and MS models describe the data exceptionally well over the wide diameter range. We observe that, especially below 100 nm wire radius, the bulk resistivity of Au is insufficient to describe the observed resistances. At radii above 110 nm one can observe a possible slight systematic deviation, which could be attributed to a change in grain size in fabrication of these larger NWs. If desired, this effect could be accounted for by fitting a different α_G for initial radii in this range. Note again that, just as in the case of the critical failure point, undetermined variables are providing variation in the initial resistance of the data for specific data points.

[xxi] Note there are experimental errors in the measurement of radius, see Figure, that may account for much of this deviation.

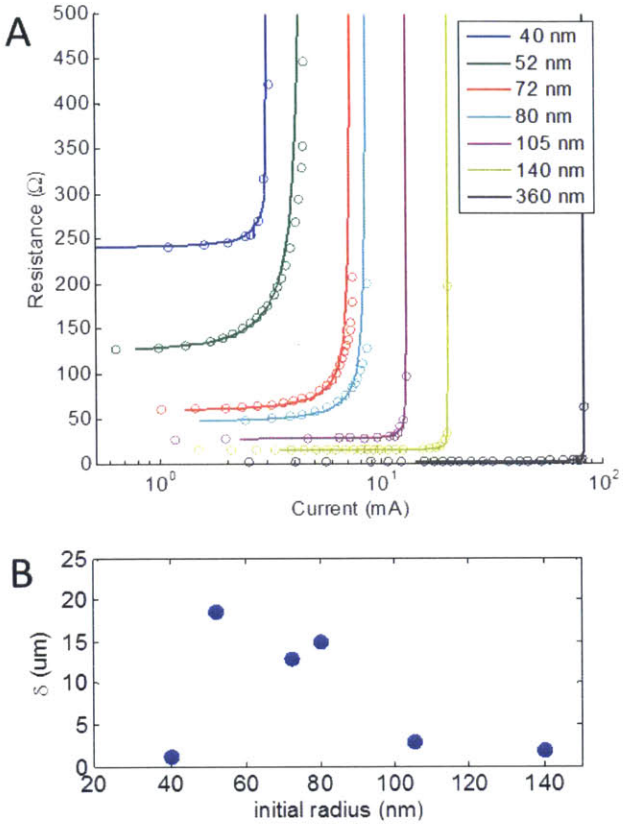


Figure 23. A) Observed (circles) and model-predicted (lines) resistance versus current traces for different initial radii. Since it is ramped at a fixed rate, current is a proxy for time in this system. B) The resulting necking region lengths optimized to fit the curves in (A).

Using these resistivity parameters, we modeled the time traces of individual experiments, and used those results to determine the effective necking region length, δ, for each radius (Figure 23). As earlier noted, the critical failure current and the initial resistance data both showed the influence of uncontrolled variables, leading to variation in the data that we could not simulate. In modeling the time evolution of single NW experiments, this provides an unpredictable offset in the initial NW resistance and the final breakage current. As we are interested in prototypical cases, we rectify this experimental variation by proportionally shifting the current data and the resistance data from each experiment such that the failure current and initial resistance collapse to their respective trend lines.

Despite the simplicity of the model assumptions on evolution of the necking region dimensions, strong prediction of time-evolution behavior is achieved. With the validity of the model's prediction established through experimental observables, it can be used it examine unmeasured quantities, the most relevant of which is the effective neck radius, shown in Figure 19 for the 40 nm radius nanowire.

Predicting nanowire response to a control scheme

One application of the model predictions is devising optimal strategies for electromigrative tunnel junction formation. As described in the Introduction, electromigrative failure (cleavage) of nanowires can

be exploited to produce a gap between the resulting electrodes small enough that an electron tunneling current can be measured.[8,23] In particular, there is ongoing debate in the literature over the source of the efficacy of and appropriate parameterization for the predominant method of controlling nanowire electromigration (towards tunnel junction formation) *via* manipulation of the applied bias.[22,133,142] In the method, developed by *Strachan et al.*, the resistance of the NW is monitored as the current or voltage is linearly increased, as in *Karim et al.*'s system, but when the resistance rises by above a threshold percentage, the bias is rapidly attenuated to arrest the formation, and the process is repeated.[133] Please see Chapter II for a full discussion of this technique, which we call the 'Penn method,' including our experimental implementation of it. The purpose of the technique is to increase the yield of tunnel junctions *i.e.* electrode gaps small enough, generally less than 1 nm, to measure a tunneling current across as opposed to forming a larger break upon failure. The reason that the Penn method improves yield however remains ambiguous. We simulated the effect of the method by applying the feedback control as an exogenous input to our nanowire surface-transport model. We used a finite-difference method with time steps of 1 ms to predict for *Karim et al.*'s nanowires the result of using this control technique with a variety of parameterizations (Figure 24).

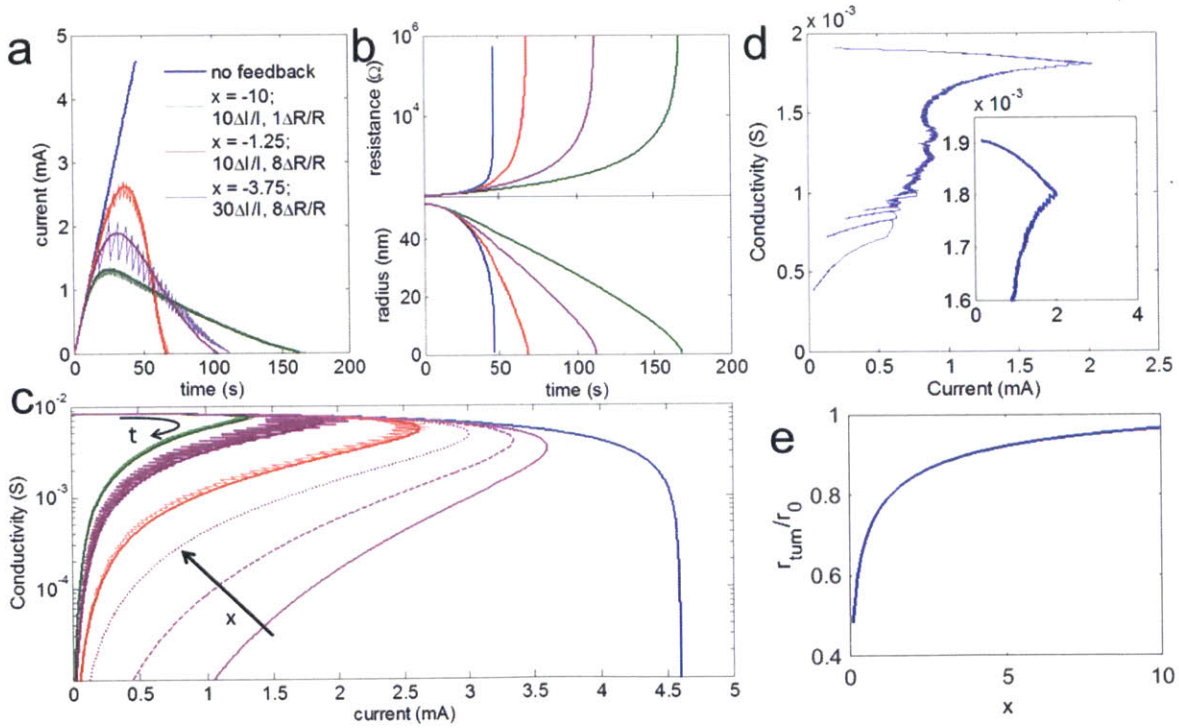


Figure 24. a) Simulated current versus time trace in a 50 nm initial radius nanowire for cases of no feedback and feedback with different parameters. Discrete feedback control evolution with a percentage reduction in current ($\Delta I/I$) in response to a percentage reduction in resistance ($\Delta R/R$) is overlayed on the equivalent continuous feedback behavior (see text). b) Simulated evolution of radius and resistance with time. c) Evolution of the conductivity with the applied current. Curves are included (in pink) for x = -0.25 (solid), -0.4 (dash), and -0.7 (dot) to show trend in deviation from uncontrolled scenario. d) experimental results of a feedback loop in our own

electron-beam lithography defined nanowires showing qualitative agreement. e) The effect of the feedback parameter on the turnaround radius.

The experiments of *Karim et al.* did not employ a feedback control mechanism during electromigration breakdown. However, our model can be used to predict the behavior under these conditions, despite crystal structure differences associated with electron-beam lithography-based fabrication in those systems such as longitudinal contact with an oxide substrate; thin, rectangular NW dimensions; and electron beam evaporation for crystal formation.[133] Under the control scheme, the time for formation is lengthened and the reduction in radius is made more gradual. This phenomenon provides a potential explanation for the increased yield of tunnel junctions under the Penn method,[xxii] versus holding at a single voltage or ramping the voltage without attenuation; instead of allowing the system positive feedback to run away and cause sudden failure of the wire, more gradual failure may yield a smaller break. In Chapter II, we indeed find that under the Penn method more gradual formation correlates with higher yield, and use our surface transport model to design more efficacious process control. On the latter point of more effective control, note from Figure 24b that while the Penn method maintains more gradual electromigration for a period of time, towards the end of formation it eventually is unable to prevent sudden formation; it cannot maintain a linear resistance pathway.

The surface transport model also predicts characteristic features of the resistance evolution under the Penn method. We observe these features experimentally not only in applying the technique to our own devices (see Chapter II; *e.g.* Figure 24d) but also in the electromigration traces published by all previous researchers.[8,88,92,143] The first feature of interest is the 'shark-fins' in $R(I(t))$ corresponding to each voltage/current ramp and attenuation cycle; rather than a linear rise in resistance, we observe an acceleration. This phenomenon is discussed in greater detail in Chapter II, but of note here is that it is reproduced by our surface transport model as a consequence of the intrinsic positive feedback; as the radius falls, the surface area to volume ratio rises, causing the radius to fall faster. The second distinguishing feature that we observe is the 'turnaround' or 'critical' voltage/current; once a critical current is reached, successively lower and lower currents are required to provide the same resistance increases, providing a maximal current at some 'turnaround radius.' A surface transport mechanism reproduces this effect (Figure 24), and explains the turnaround point as the point at which the formation pathway $R(I(t))$ under control diverges from the control-free pathway (Figure 24b). We can understand this point by considering the continuous analog of the discrete feedback process

[xxii] 'Yield' circumscribes successful versus unsuccessful tunnel junction formation, with the former being a gap between electrodes small enough (generally less than 1 nm) to measure a tunneling current across and the latter being larger. These two outcomes are discussed in more detail in Chapter II.

(18) $$\frac{dI}{dt} = \alpha + \frac{Ix}{R}\frac{dR}{dt}$$

which with Equation (8) defines the smooth curves in Figure 24. x is the negative unitless control parameter corresponding to the proportional change in I to be induced by a proportional change in R, or in the language of discrete feedback, $x \equiv \Delta I/I \cdot R/\Delta R$. More negative x values therefore correspond with more aggressive control – a longer time to failure and lower maximal current. As x increases, the turnaround occurs at an earlier time and higher radius despite a longer process, leading to more thinning occurring in the 'tail' portion of the current-time trace (Figure 24a,c). Furthermore, from x the radius reduction at the turnaround radius can be extracted, providing physical intuition on observed current traces. In defining the resistance evolution pathway, x is suggested to be the parameter defining application of the Penn method.

II. MONITORING ELECTROMIGRATION KINETICS FOR CONTINUOUS CONTROL

Summary of Findings

As described in the Introduction, Electromigration of metal nanowires is the predominant method of on-chip fabrication of a single-molecule tunnel junction,[51,144] which is considered a uniquely enabling technology for molecular electronics,[145,146] biomolecule sequencing,[147,148] nanopore interrogation,[55,60,148] and single-molecule physics.[37] Electromigration inspired by an applied bias thins the wire until it breaks, sometimes forming a gap between the resulting electrodes small enough, less than 1 nm, to measure a tunneling current across. Adoption and advancement of the method has been hampered however by poor yield and process variability;[8,22,85,89-92] the mechanism of electromigration, *i.e.* the migration pathway and resulting kinetics, at this scale was unknown, yielding a dearth of voltage-control algorithms. In the previous chapter, we proposed that as wire diameter shrinks to O(100 nm), the kinetically-dominant pathway for mass transport becomes surface transport, rather than the bulk vacancy migration that is known to dominate at micron and larger diameters.[149] We tested this model against experimental data from *Karim et al.*[94], showing that surface transport but not bulk transport could describe the initial-radius-dependent time-evolution of resistance in nanowires from 40 nm to 360 nm radius. If our model is accurate, as suggested by that data, it can be applied to improve the control of the electromigrative formation of tunnel junctions.

Prior to this work, the only means of controlling electromigration was the 'Penn method,' developed by *Strachan et al.*[8] A voltage V is applied across the nanowire and increased while the wire resistance R is monitored, rising as electromigration proceeds and the wire thins. When the resistance increases by a set percentage threshold the voltage is rapidly attenuated and the process repeated until the wire breaks, sometimes forming gaps small enough to measure a tunneling current across. Described in more depth in the section below, the 'Penn Method' is neither continuous nor tied to electromigration kinetics, yielding high variation.[8,22,85,90,91] On examination (in our work, below), the resistance in the Penn method changes little as the voltage increases until an unknown threshold voltage is passed, at which point the resistance rapidly rises. The sudden rise is then caught by the R increase threshold. The wire therefore necks in an intermittent series of short bursts of rapid migration, rather than at a stable or otherwise predictable rate of mass transport. Each cycle therefore risks precipitous breakage of the nanowire *i.e.* thinning too rapidly for the voltage attenuation to arrest. This control behavior may explain the process variability.

In this work we design and apply a continuous controller of electromigration in nanowires for the formation of tunnel junctions, based on the surface transport kinetics we developed in Chapter I. By using

a control variable, *dlnR/dt*,[xxiii] that is linear function of the rate of atom electromigration, we are able to deliberately alter the electromigration pathway $R(t)$ and outcome by adjusting the controller. We find that the yield and $R(t)$ pathway are sensitive to the *dlnR/dt* setpoint, further validating our proposed mechanism and offering the potential to design the electromigration pathway by designing a setpoint pathway. We find that the setpoint can be optimized, and that it is possible to achieve tunnel junction yields of 100%.

We begin by investigating the Penn method further, developing our own nanowire devices to do so. We observe that while R, the control variable, does not have a clear relationship with V, the input variable, with the $R(V(t))$ pathway seemingly unpredictable, *dR/dt* in contrast has a highly regular response to V. That generates the characteristic 'shark-fin' patterns introduced in Chapter I and explored in more depth here. In fact, while the Penn method explicitly monitors R we show that it is *de facto* controlling *dR/dt*; each voltage-ramp cycle produces the same rate of R change as the R threshold is approached. Performing over 100 formations with the Penn method we also found that failure, defined as the creation of a gap too wide to measure electron tunneling, only occurs when formation is precipitous *i.e.* sudden. This finding is consistent with our result from Chapter I, where we applied our surface transport mechanism to study the Penn method and found that the primary effect of the Penn method was to slow the overall rate of electromigration. These results suggest that slowing the rate of wire thinning is a means of increasing the yield of electromigrative formation of tunnel junctions. As we observed in Chapter I however, the Penn method only slows electromigration at the beginning of the process; eventually it, like simpler strategies of ramping or holding constant voltage, produces sudden wire failure that we attribute to intrinsic positive feedback in the nanowire.

From our surface transport model we then derived the requisite constraints on any electromigration process to avoid such precipitous formation. We find that in our model the rate of wire necking can be experimentally observed as *dlnR/dt*. To avoid runaway therefore, V must be adjusted to maintain *dlnR/dt* at a constant value. The maximum setpoint is dictated by the noise-limited speed of measurement, τ. Specifically, $dlnR/dt \leq 2/3\tau$. We test that principal by designing a *dlnR/dt* controller and applying it experimentally to our nanowires. To handle non-idealities of electromigration such as back-fill and stochastic changes in the scattering cross-section, we employ an exponential control term in addition to integral control. Applying the controller, we find that it successfully maintains the setpoint, providing

---

[xxiii] *dlnR/dt* is a shorter expression of $(1/R)dR/dt$, the percentage change in R with time ($dlnR = (1/R)dr$). We also frequently use *dR/dt* and *dlnR/dt* seemingly interchangeably, for two-fold reasons. First, the span of R in an electromigration process is often too low (*e.g.* a typical range is 400 to 600 Ω) to differentiate between *dR/dt* and *dlnR/dt*; second, as we show below *dR/dt* often satisfies all the control criteria that *dlnR/dt* satisfies because in inequalities $dR/dt > dlnR/dt$ always.

continuous control, gradual formation, and a predictable $R(t)$ pathway. We observe that the process changes with the setpoint, falling into three regimes. When the setpoint is too high the incidence of precipitous formation rises and the yield of successful junctions drops appropriately. When the setpoint is too low however, specifically when it is in the *dR/dt* regime where Joule heating dominates *dR/dt* as indicated by *dR* being reversible, the character of formation (the $R(t)$ pathway) changes and the yield also falls. In the intermediate optimum of 'ideal' formation, which matches the *de facto dR/dt* value we observe in the Penn method, formation proceeds smoothly and the yield approaches 100% despite unpredictable and time-evolving imperfections in crystal structure. These findings have several implications in addition to establishing the first continuous control of electromigration and raising tunnel junction yields. First, the ability of the surface transport mechanism to predict the Penn method observations and the operation of the *dR/dt* controller further validates the model and our hypothesis (see Chapter I) that as a wire radius falls to the nanometer scale the mechanism of formation shifts from the Black model (bulk vacancy drift) to surface atom transport. Second, by showing that the reversible *dR/dt* regime does not lead to electromigrative failure, we provide further evidence that the Joule heating mechanism of nanoscale electromigration is incorrect. Third, the sensitivity of the electromigration pathway and outcome to the *dlnR/dt* setpoint, in addition to supporting the prediction that *dlnR/dt* observes the rate of thinning, suggests that the electromigration pathway can be closely controlled as desired by manipulating *dlnR/dt*, particularly when greater fabrication control over nanowire crystal structure is achieved. Finally, while we do not claim that our controller will guarantee tunnel junction formation in all devices, we do demonstrate that it is possible to achieve 100% yield under the right circumstances even without further advances in Au crystal fabrication.[xxiv] In total, our results establish that electromigrative formation of tunnel junctions is a commercially viable technology.

Nanowire Device Fabrication

We performed electromigration of Au using Au nanowire devices defined by Electron Beam Lithography (EBL). We exposed Poly(Methyl Methacrylate) (PMMA, MicroChem) electron resist on a substrate of 300 nm SiO_2 on top of a p-doped Si single crystal back-gate, defining nanowire negatives between 100 and 200 nm long and 40 and 100 nm wide. We found that to electromigratively form viable tunnel junctions using the 'Penn method,' discussed below, the wire cross-sectional area had to be

---

[xxiv] The fabrication techniques, centered on Electron Beam Lithography and discussed below, can produce Au nanowires of reproducible external dimensions but the crystal structure, in particular the grain boundaries and exposed facets, cannot be controlled. Grain boundaries have been widely shown to dominate the electromigration kinetics, leading to variation both between devices and within devices, in the latter case due to evolution of the boundaries during electromigration.

suitably small, less than 0.001 μm^2, as achieved by those dimensions and between 10 and 20 nm of deposited Au. Otherwise, formation was either excessively explosive, yielding gaps too wide to measure a tunneling current across, or constant back-filling would inhibit formation entirely, as also observed by *Xiang et al.*[88] In EBL we also defined 10 μm wide contact wires. To minimize specular resistance at the connection, a smooth continuous trapezoid tapered the contact wire down to the nanowire width over a 2 μm (Figure 25b).

To achieve this, 950,000 g/mol molecular weight PMMA diluted to 4 wt% in anisole was spin-coated at 3,000 RPM for 60 seconds on to the substrate. At this speed and time the viscous polymer solution approaches its asymptote spin-coat thickness, consistently yielding a 200±30 nm thick polymer layer. The polymer was then baked at 170°C ± 15°C on a hot plate, measured by IR thermometer on a Si plate, for 5 minutes to cure. Baking drives off some of the anisole solvent to reduce diffusion of polymer that is cross-linked when exposed, which would otherwise blur the feature exposure. We found that under-baking therefore yielded worse resolution, while over-baking we found to inhibit liftoff. The PMMA thickness was optimized to coincide with our electron beam acceleration voltage of 30 keV, yielding a sufficient undercut to achieve good liftoff, discussed further below, without excessively broadening our features. Immediately prior to PMMA deposition, the 300 nm SiO_2 / p-Si substrate chip was ultra-sonicated for 3 minutes in isopropyl alcohol (IPA) followed by rinsing with IPA, to remove macroscopic residue such as silicate dust, and then microscopically cleaned with an Oxygen RF plasma etch for 3 minutes at 200 W and 1.5 sccm flow of O_2 to yield a steady state pressure of 1 torr. In addition to cleaning the substrate of organic contaminants normally adsorbed from air, we found that this plasma etch was critical for strong adhesion of the nanowires, discussed further below.

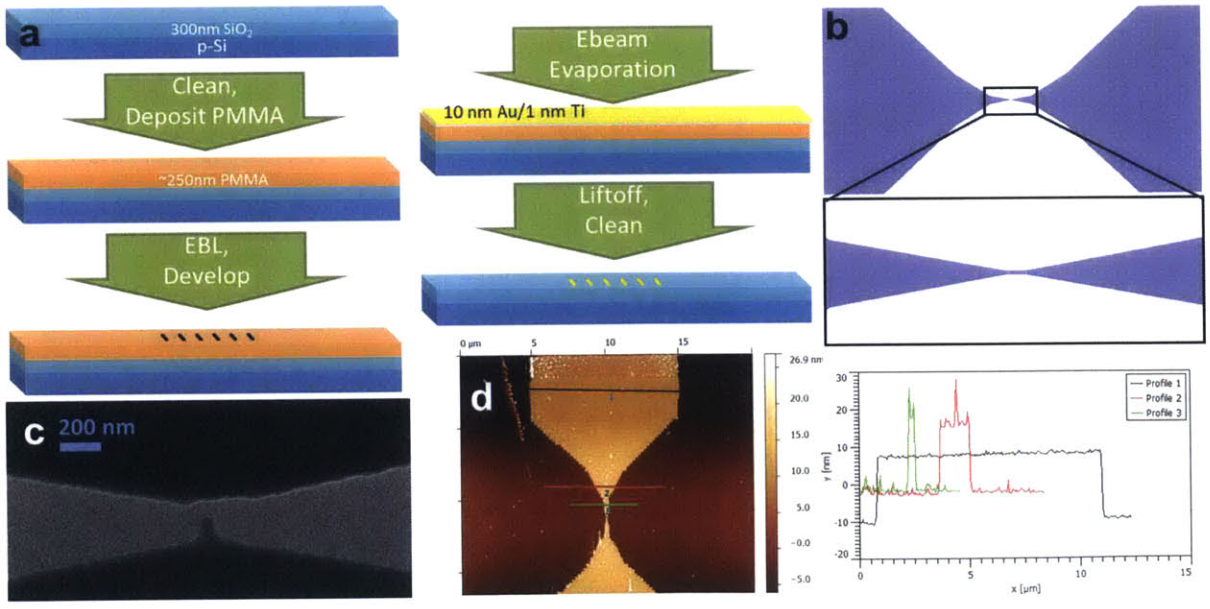


Figure 25. (a) cartoon illustrating the nanowire device fabrication process. See text for details. (b) The digital mask written onto the substrate using EBL, showing the contact wire gradually tapering to the nanowire. Dots are spaced at 500 nm (left) and 100 nm (right). (c) Scanning Electron Microscopy (SEM) image of a nanowire, showing no residual PMMA (images as clumps), clean liftoff (indicated by sharp edges), and sufficient exposure plus good adhesion to the substrate (indicated by the smooth metal surface). (d) Atomic Force Microscopy (AFM) image of the contact wire verifying the desire metal deposition thickness for this device.

The PMMA was exposed in a Raith 150 Scanning Electron Microscope (SEM) in the previously described pattern (Figure 25b). We used a 30 keV acceleration voltage to produce sufficiently sharp features while also creating an undercut for improved liftoff (see below). A 60 μm aperture was ultimately used to balance write speed with spot size. For our purposes, the nanowire widths needed to be small but precision was unimportant, *i.e.* whether the nanowire is 50 nm or 55 nm wide is not critical as long as it is less than 100 nm in width to produce a sufficiently low cross-section (see above); in the electromigration process uncontrolled heterogeneities such as grain boundaries are far more influential on outcomes than such small changes in the absolute dimensions.[22] We exploited this criterion to design a robust and high throughput EBL process. The mask we used wrote only two lines (20 nm nominal mask nanowire width, Figure 25b), but we over-exposed the wire to achieve consistent 50-70 nm final mask negatives at 320 $\mu C/cm^2$ dose (we adjusted the dose to achieve different nanowire widths). This dramatically reduced the variability in nanowire width with dose, increasing the overall device consistency. To increase throughput, we wrote 25 devices on each 12 x 14 mm^2 chip. To maintain a sharp focus across the surface over this area, we applied a plane correction to the focus, using Au nanoparticles at three points along the edges of the chip to estimate the change in focus across the surface; stigmation and aperture alignment were held constant. It was also critically important to make the chip as flat as possible. While the substrate was single-crystal silicon, contamination on the SEM sample plate and the back of the substrate would prevent it from being perfectly level. In the latter case, after PMMA spin-coating and before curing, the back of the chip was carefully cleaned with acetone-soaked cotton applicators to scrub off PMMA residue, which due to optical interference is clearly visible by eye even in thin layers. To minimize bowing of the substrate while maximizing chip size, the substrate was adhered to the SEM sample chuck using copper tape. By implementing these various measures, we nearly always achieved a surface flatness of 2-15 μm across the substrate. After exposure, the PMMA features were developed in a 1:3 Methyl Isobutyl Ketone (MIBK):IPA solution for 70 seconds at 22-24°C and checked in an optical microscope (Figure 26).

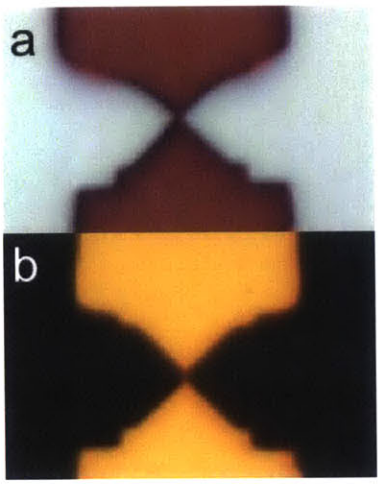


Figure 26. Optical Microscope images of the device negative in PMMA (a) and the final Ti/Au/SiO_2 device (b). For scale, contact wire is 10 μm wide.

Using electron beam evaporation, we sputtered 1 nm of Ti as an adhesion layer followed by the desired thickness of Au into the PMMA negatives. The vacuum pressure during deposition was better than 10^{-6} torr. The thickness was monitored with a quartz crystal microbalance and verified *ex post* with Atomic Force Microscopy (AFM, Figure 25d). The Ti adhesion layer was important to achieve clean liftoff, discussed below, and to achieve uniform adhesion of the Au crystal. Without the Ti, the Au layer is continuous but roughened; despite our cleaning method organic contaminants on the SiO_2 surface are inevitable and the oxidize-able Ti is able to bind to them and present the Au with a uniform crystal surface. Cleaning the surface *via* plasma etching prior to PMMA deposition is also critical to minimize the presence of organics and achieve strong adhesion. 1 nm of Ti does not form a uniform layer, but that is not necessary to present sufficient binding sites for a continuous (polycrystalline) Au crystal to form. In addition to sometimes causing poor liftoff (discussed below), poor adhesion yields a characteristic corrugation of the Au surface clearly visible in SEM imaging (Figure 27b). Poorly adhered devices can still be electromigrated, but the series resistance is higher we posit due to higher specular scattering, reducing the process reliability (discussed below). PMMA residue remaining in cases of under-exposure is also clearly and characteristically visible as chunks in SEM imaging (Figure 27c).

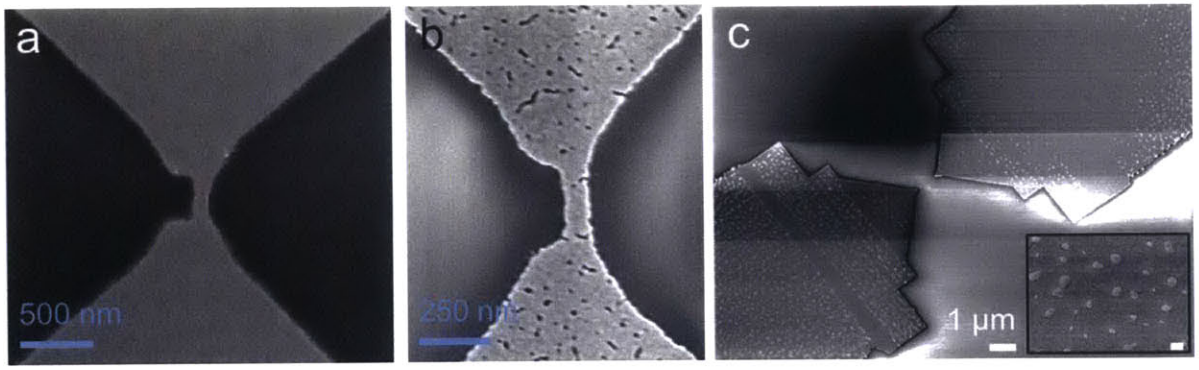


Figure 27. Au nanowire and contact wire with (a) and without (b) an adhesion layer, showing the roughness of the Au layer in the latter case. (c) Under-exposed contact wire, showing lingering PMMA residue. Inset zoom.

After metal deposition, the PMMA mask with removed in a 'lift-off' step by soaking in hot n-Methyl Pyrrolidone (NMP) solvent. Specifically, we found that holding the NMP at 95°C was ideal for achieving clean liftoff with the above EBL process. Higher temperatures can cause some Au nanowires to relax or dissolve in solution, while lower temperatures remove the PMMA too slowly such that the metal foil on top of the un-exposed PMMA is able to contact and adhere to the surface *via* van der Waals

attraction. Excessive baking of the polymer, driving off too much of the Anisole solvent, we found could inhibit clean liftoff. Poor focus in the EBL process would result, in addition to feature broadening, poor liftoff characterized by rough edges to the Au features. With process parameters appropriately optimized, liftoff was usually achieved macroscopically within a few minutes. SEM imaging confirmed the full removal of macroscopic PMMA (10+ nanometer sized globules). On a typical chip of 25 devices, 20 would remain intact after liftoff. The samples were rinsed in acetone followed by IPA after the hot NMP bath to avoid leaving solvent residue (NMP has a high vapor pressure and even HPLC-grade acetone leaves a residue).

After OM and SEM imaging to verify nanowire dimensions and integrity, the devices were rinsed in HPLC-grade Acetone (Chromasolv©, >99.95 % purity, Sigma Aldrich) followed by HPLC-grade IPA and then cleaned in an Oxygen RF plasma at 40 W for 30 seconds at 1.5 sccm flow of O_2 yielding 1 torr steady state pressure, and finally immediately placed in a Nitrogen glove box with a <0.5 ppm O_2 atmosphere before transfer to the vacuum probe station, described below, for testing. Ag paste was added to the contact pads to improve electrical contact, dried for at least 24 hours in the aforementioned nitrogen environment.

Electrical Testing

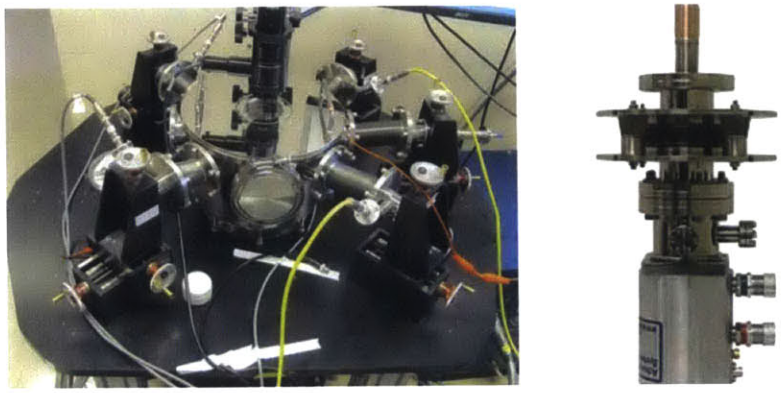


Figure 28. Photograph of probe station (left) and cryostat (right) used in this work.

To electrically interrogate the nanowires, we placed the samples in a low temperature probe station designed by ARS Cryo. Vacuum pressure was maintained below 10^{-6} mbar with a turbo pump and venting was always to dry nitrogen. The chamber was regularly leak tested and was flushed with nitrogen whenever opened to atmosphere for prolonged periods. The metal chamber was grounded and acts as a Faraday cage to minimize noise from ambient electric fields. The chamber is mounted on both vibrational isolation padding and inflated air bladders to reduce the influence of building vibrations on contact resistance. Contact was achieved with horizontal manipulators in vacuum bellows with manual XYZ

translation and tungsten probe heads. During contact the samples were exposed to light but during testing they were kept in the dark. Samples were placed on the sample stage either with the back gate grounded through the stage or isolated from ground by a glass coverslip. In all cases the sample bottom was adhered to the surface with a thin layer of Cu paste to achieve good thermal contact. Device contact resistances were checked to be always below 10 Ω. The sample stage was cooled below 9 K by cryogenically pumping on recycled He gas. The chamber was vibrationally isolated from the cryo pump with flexible bellows. The sample temperature was modulated up to 300 K with a resistive heater inside the sample stage and monitored with an embedded thermocouple.

In the work in this section, electrical testing was performed with a set of Agilent E5262A Source-Measure Units (SMUs), with which each of four probe heads independently force a voltage or current and sense a voltage or current. Sensitivity in measurement and forcing was < 1 pA and < 1 μV respectively. The SMUs were controlled in arbitrary algorithms *via* computer by programming in Matlab© and sending/receiving signals with a GPIB interface. The SMUs were interfaced with the chamber using triaxial cabling, where between the center pin (device) and outer grounded shield there is a guard shield held at the same bias as the center pin to reduce leakage to the guard. Current leakage was always measured to be less than 100 fA up to 10 V. Typical contact wire series resistance was 300 Ω, and the initial nanowire resistance varied by dimensions (width, thickness, and length) from 100 Ω to 700 Ω. To minimize destruction of the nanowires on contact from electrostatic discharge, a Transient Voltage Suppression diode with a clamping voltage of less than 1 V was placed across the source and drain during contact and removed prior to testing.

'Penn Method' Electromigrative Formation of Tunnel Junctions

To form sub-nanometer breaks in the nanowires *i.e.* electron tunneling junctions, we began by applying the 'Penn Method,' developed by *Strachan et al.*[8] See Chapter I and the Introduction for discussion of the physics of electromigration. We contact each end of the nanowire and apply a gradually increasing DC voltage bias across it. The increasing voltage increasingly drives electromigration in the wire, which occurs most rapidly over a small length (δ in our surface transport model, Chapter I) typically associated with a grain boundary as shown in *in situ* Transmission Electron Microscopy imaging of the process.[22,83] The wire thins, or 'necks,' in that regime as a result. That thinning causes the measured wire resistance R to rise. As we showed in Chapter I, even at fixed voltage the process undergoes positive feedback, with an acceleration of thinning as the wire thins. To avoid runaway, explosive formation therefore an attenuation of the voltage in response to sudden increases in resistance is necessary.

In the Penn method, when R rises above a set threshold percentage increase the voltage is rapidly attenuated to slow or halt electromigration. The cycle is then repeated, incrementally thinning the neck

and increasing the resistance, until a tunnel gap is formed as indicated by a drop in conductivity to below the conductance quantum $G_0 \approx 7.75 \times 10^{-5}$ S. A resulting trace of such a process is produced in Figure 29a. Through trial and error we have found that a percentage rise between 0.3% and 1% is appropriate. When the threshold rise is too high, the nanowire breaks precipitously as the positive feedback in the electromigration kinetics (see Chapter I) accelerates before the voltage attenuation responds. A threshold rise too low however is also detrimental; we find that under a modest applied bias there is a reversible resistance rise (it returns to the baseline value when the bias is removed) that we attribute to Joule heating. At these biases no measurable electromigration occurs; holding at a bias exhibiting this reversible rise (which eventually tapers off, which we attribute to reaching a steady-state temperature) for over 60 hours yields no discernable change in nanowire morphology in SEM imaging. When the resistance rise threshold for voltage attenuation is set too low, it attenuates the voltage in response to these mere reversible rises. The threshold must be high enough such that the observed resistance rise is reliably associated with thinning of the nanowire.

We applied the Penn method to over 100 devices, with our findings described in more depth below. In ideal cases, the resistance increase is incremental until the neck is only a few tens of atoms in diameter (conductivity on the order of $10G_0$), at which point the resistance increases are sudden stochastic jumps, which we attribute to the dislocation of single or few atoms, as in reference [133] (*e.g.* Figure 29a). Formations were performed at temperatures ranging from 9K to room temperature. Once a gap is formed and the electromigration process halted, the existence of a tunnel junction is verified in three ways. The first is the value of the conductance being below G_0; since the nanowire cannot elongate, the reduction in conductivity is assigned to thinning and eventually breaking, as verified by numerous *in situ* studies.[22,83] The second is Scanning Electron Microscopy (SEM) imaging of the gaps after testing, showing resolution-limited gap size (Figure 29b). The final is testing that the electron transport is ballistic; the resistance of tunnel junctions is limited by the tunneling probability, not phonon scattering as in bulk crystals. As a result, the conductivity of tunnel junctions is invariant with temperature, aside from a small effect from broadening of the Fermi-Dirac distribution of occupied electronic states in the electrodes, whereas bulk crystal transport is highly temperature dependent. We vary the temperature T and check that the conductivity is invariant (Figure 29c).

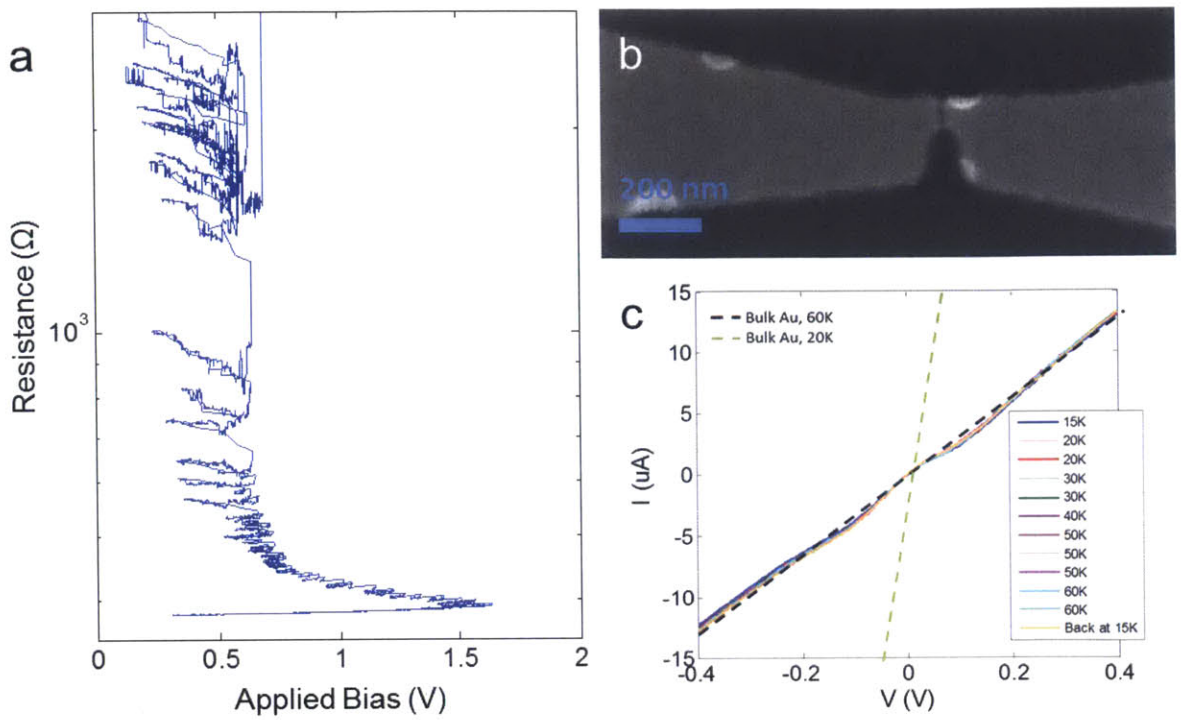


Figure 29. (a) Example resistance pathway trace from applying the Penn method to one of our devices. The voltage is gradually increased with time (not shown) and the resistance R rise monitored as electromigration proceeds and the wire neck thins. When R rises above a threshold percentage the voltage is attenuated and the cycle repeated, gradually increasing the resistance until the wire breaks. Further details in the *Results* section below. (b) Scanning Electron Microscopy (SEM) image of a nanowire after breakage, showing a resolution-limited gap formed. Within the narrow gap strip visible, there is presumably a point where the gap size is minimal. Because of the exponential dependence of the tunneling probability on gap size, the smallest-gap-point between the two electrodes will dominate the tunneling current. If the gap is less than a nanometer, the tunneling current will be measurable. (c) Test for ballistic transport. The solid curves are the *I-V* characteristics of a tunnel junction measured at different temperatures (the slope is the conductivity). There is essentially no change in conductivity with temperature, indicating ballistic conduction. For comparison, the dashed lines show the expected change in slope if the resistance was dominated by bulk Au, with resistivity almost doubling over the observed temperature range. Clearly the expected rise in resistance of bulk Au over the examined temperature range dwarfs the observed conductivity increase.

In addition to the Penn method we have developed and applied a novel means of monitoring and controlling the nanowire electromigrative failure process, *dlnR/dt* control. Those experimental methods are surveyed below however, after developing the relevant control theory, rather than elaborating on them here. Please see below.

Empirical Observation of Nanowire Electromigration using the Penn Method

We began studying electromigrative formation of tunnel junctions from nanowires by applying the standard method in the field, what we call the 'Penn method,' developed by *Strachan et al.*[8] As

described in the previous section, we contacted the nanowires and gradually increased the voltage V while monitoring the resistance R. As the neck region of the wire thins the resistance rises accordingly. When the resistance increases by a certain percentage, typically between 0.3% and 1%, the voltage is attenuated and the cycle repeated to incrementally raise the wire resistance until the wire breaks as indicated by the conductivity falling below the conductance quantum G_0 (Figure 30). Performing this process on over 100 devices, we observed only three different outcomes: (1) the resistance increases gradually until a junction is formed, and that gap is suitably small (< 1 nm) for electrical interrogation (*i.e.* a measurable electrical signal, conductivity over 10^{-8} S), which we will henceforth term 'viable' (Figure 30); (2) the wire breaks suddenly, with the resistance undergoing a sudden uncontrolled increase, but the resulting gap remains viable; (3) the resistance undergoes a sudden, uncontrolled increase, but the resulting gap is too wide to measure a tunnel current, which we consider a 'failed' device (*e.g.* Figure 31). Note that every case of failure corresponded to a sudden formation; gradual formation always produced a viable tunnel junction. Possible causes of failure on sudden formation could include thermal runaway, *i.e.* faster formation yielding a higher steady-state temperature, which at some point causes local melting; or over-voltage, since the tunnel junction once formed can only survive a few hundred millivolts of applied bias before relaxing to a larger gap, if it forms at a higher voltage it could immediately relax. The stability of a viable junction under the applied bias of testing, *i.e.* its lifetime, varied from only a few measurements to many hours. The perfect correlation of junction failure with precipitous formation implies that avoiding sudden formation may increase the regularity and viable-junction yield of the electromigration process. To realize that objective, we returned to our surface transport theory of nanowire electromigration (Chapter I) to derive conditions for which gradual formation would be guaranteed.

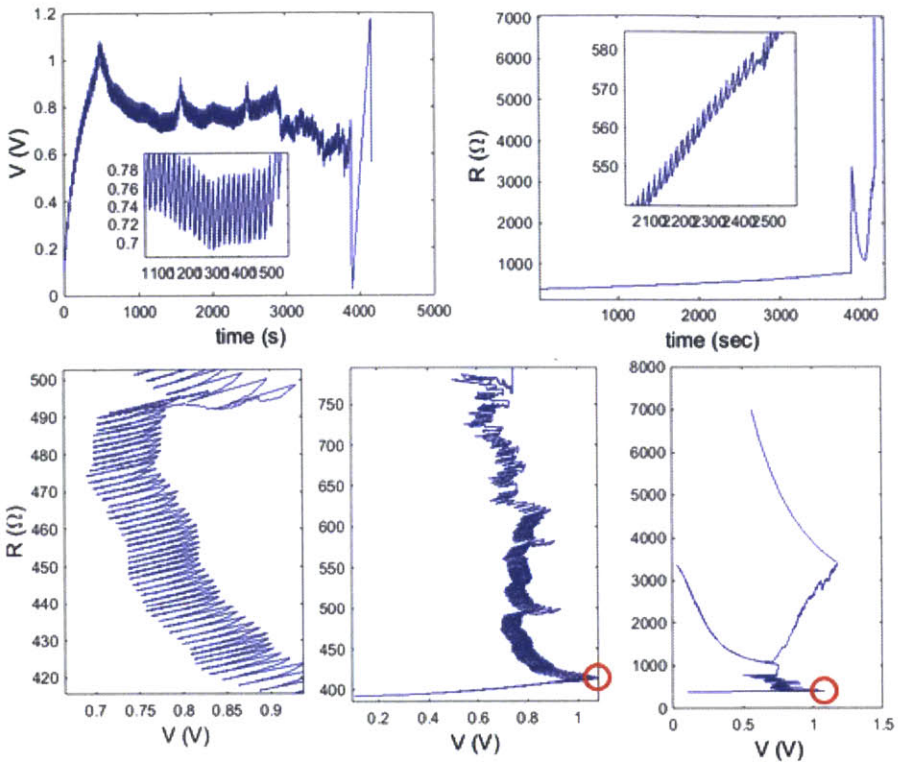


Figure 30. Application of the Penn method to our nanowires for tunnel junction formation. Example traces of the process from one experiment. The voltage *V* is gradually increased while the resistance *R* is monitored. *R* rises as the neck thins, and when it rises past a set threshold percentage the voltage is rapidly attenuated. The process is then repeated until the neck breaks, as indicated by a drop in the conductivity (rise in the resistance) below the conductance quantum. The top row is *V* and *R versus* time, with insets zoomed, and the bottom row is *R versus V* at three levels of zoom to show the process features visible at different scales. The left panel shows each cycle of the algorithm, with the characteristic 'shark-fins' discussed in the main text. The middle panel shows the overall pathway *i.e.* the envelope function bounding the cycles. The red circle indicates the 'turnaround' or 'critical' voltage. The right panel shows the end of formation, the stochastic regime where the wire is only a few atoms wide and resistance increases are in discrete steps. While some formations are gradual and successful as shown here, others happen suddenly as in Figure 31.

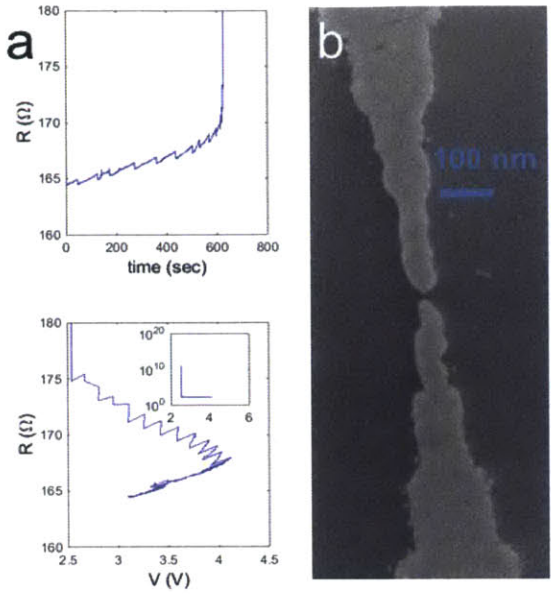


Figure 31. The positive feedback in the system causes the wire failure to accelerate faster than the voltage attenuation can detect it and respond, leading to precipitous formation (a). In these cases the resulting gap is often too large, greater than 1 nm across, to measure a tunneling current across, as for example in the SEM image of the device in (b).

Before turning to our electromigration model, we examined the Penn results further to catalogue the features of the resistance-voltage pathways it produces. Later (below), we find that they are consistent with the predictions of our surface transport model, providing further evidence of the theory's accuracy. While they have not been explicitly identified by researchers prior to this work, all of these features are visible in the application of the Penn method not only in our devices but in all previously published electromigration traces.[8,88,92,143] First, from a 'zoomed out' perspective *i.e.* examining the envelope function of $R(V(t))$ (*e.g.* Figure 30, bottom-middle panel), the most prominent feature is the 'critical' or 'turnaround' voltage, first identified in Chapter I. To increase the resistance of the nanowire irreversibly the voltage cannot be arbitrarily low; as described in the previous section (*'Penn Method' Electromigrative Formation of Tunnel Junctions*), the R rise threshold (the percent rise at which the voltage is attenuated) must be low enough to avoid sudden formation but also high enough to reflect nanowire thinning rather than the reversible R increase due to Joule heating. The voltage required to thin the neck initially increases until the 'turnaround voltage' is reached, after which point lower and lower voltages are necessary to achieve the requisite resistance rise. Our surface transport model predicted and explained this behavior without any *ex post* or empirical modification, as shown in Figure 24 and Chapter I. The turn-around voltage corresponds to the point at which the feedback algorithm pathway deviates from the no-control (voltage-hold) pathway (Figure 24c). Without control, the intrinsic positive feedback in the surface transport model causes necking to run away; as the wire thins the surface area to volume ratio increases and thus the thinning rate accelerates at fixed voltage. The voltage attenuation in the Penn method endeavors to catch that resistance rise and reduce the voltage before that occurs, leading to a more gradual formation characterized by a lower and lower voltage being required to achieve thinning, as we observe experimentally. The governing parameter defining the point of deviation between the no-control and control cases we found in Chapter I to be the ratio of the percentage increase in R that triggers the voltage response to the percentage reduction in V that responds (Figure 24).

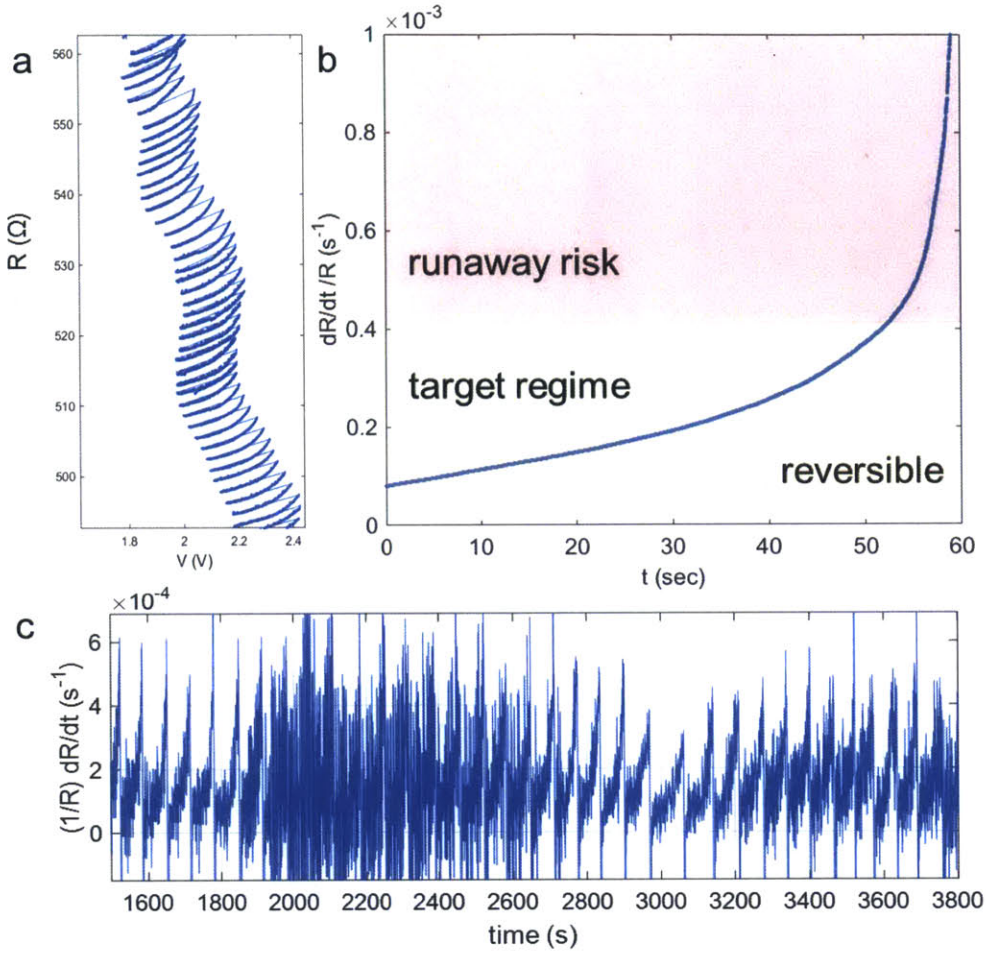


Figure 32. Penn method shark fins. (a) Zoomed section of resistance *versus* voltage trace for a nanowire formation using the Penn method. (b) Collapse of the points of trace (a) onto *(1/R) dR/dt versus* normalized time, showing the regular curvature of the positive feedback in the system; until halted by the voltage feedback response, the resistance increase accelerates. With each cycle therefore this method risks failure. We observe three regimes of different *dR/dt* behavior. At low voltage, the resistance change is reversible, which we ascribe to Joule heating. In the target regime, the resistance rise is due to necking of the nanowire. As the voltage applied becomes too high, or is not attenuated for too long allowing the natural positive feedback of the system to accelerate, the necking accelerates to where *dR/dt* risks failure. (c) The rate of resistance change *(1/R) dR/dt*, proportional to the slope of the shark-fins in (a), with time.

The second characteristic feature of the Penn method we identify is the $R(V)$ 'shark-fins' created by the voltage attenuation cycle (Figure 30, bottom-left panel; and Figure 32). The resistance does not rise linearly with applied voltage in the irreversible regime. Rather, when the bias is too low it exhibits a very small *dR/dt* that is linear with V but reversible; in Figure 32a that is visible as the beginning of each shark-fin, where one can see that the forward-moving (in time) $R(V)$ retraces the line of the backwards-moving $R(V)$, the R response to the previous cycle's voltage attenuation. When the R increase begins to deviate from this line however, it increases more and more rapidly with V; *i.e. dR/dV*, which is proportional to *dR/dt* in the linear $V(t)$ application, is not a constant but rather increasing. In fact *dR/dV*

grows exponentially, beginning to inflect towards an asymptote when it reaches a certain voltage. At that point the resistance rises suddenly, triggering the Penn algorithm to reduce the voltage immediately once the R threshold is passed. For this reason setting the R threshold too low leads to no formation (purely reversible R increases), while setting it too high risks allowing sudden formation before the voltage attenuation responds. In this manner, while the Penn method is explicitly using R as the control variable, it is *de facto* responding to dR/dt, with percent-change-in-R as a crude proxy variable; the dR/dt at the peak of the shark fin is essentially maintained within $\Delta R/R$ divided by the distribution of the shark-fin timespans. The three-regime shark-fin behavior we observe suggests that, unlike R, dR/dt is an instantaneous (*i.e.* path-independent) function of the underlying electromigration kinetics. That conclusion is supported by the consistency observed between shark-fins. While the envelope function $R(V(t))$ varies unpredictably, sometimes increasing and sometimes decreasing (Figure 32a), dR/dt, and specifically $dlnR/dt$, has a consistent trajectory (Figure 32b). From a control standpoint, dR/dt has a consistent relationship with our input variable V, whereas R does not. That suggests that dR/dt would be a more effective control variable, providing greater ability to deliberately alter the process pathway and outcome.

Surface Transport Model Implications for Electromigration Control

Empirical observation therefore suggests that dR/dt is a superior control variable for the electromigration of nanowires, and dR/dt instantaneously reflects the underlying electromigration kinetics. To understand why and derive a control mechanism, we return to our surface transport model of nanowire electromigration developed in Chapter I. As a function of voltage, we predict the rate of nanowire radius reduction to be (see Chapter I)

$$\frac{dr}{dt} \cong -\frac{2\gamma\Delta}{R_s} V(t) \frac{1}{r^2} \tag{19}$$

where r is the nanowire neck region radius, initially at r_0, t is time, γ is the scattering cross-section ϕ_{cs} divided by the constant $4\pi q$ where q is the elementary charge, Δ is the hard sphere atomic diameter of the metal, R_s is the wire series resistance, and V is the applied bias which is itself a function of time (either explicitly as in the Penn method or, in a continuous controller, implicitly *via* the control variable). The approximation sign indicates this simplified expression is valid under two very strong approximations (see Appendix L). The first is that the series resistance is substantially larger than the initial neck region resistance, and the second is that the radius is much larger than one atomic diameter ($r \gg \Delta$) for the region of the process of interest; see Appendix L for validation.

The source of the intrinsic positive feedback in the system is clear; at a fixed voltage *i.e.* in the absence of control, the shrinking radius (nanowire necking) causes the $1/r^2$ term to grow quadratically,

leading to an acceleration in dr/dt. Physically, the $1/r^2$ term arises from the surface area to volume ratio (Chapter I); as the radius shrinks, the number of surface atoms, which we showed dominate the electromigration kinetics in nanowires, relative to the number of bulk atoms increases leading to a greater portion of the momentum imparted from the current flux being allocated to the surface atoms thus speeding atom migration. That phenomenon is visualized in the simulations of Figure 33. With a fixed bias the rate at which radius declines accelerates to an asymptote, or in other words the time to nanowire failure/cleavage ($r = 0$) shrinks exponentially with radius. The higher the bias, the faster the acceleration and higher the radius at which the failure asymptote is approached. In fact for the typical nanowire conditions portrayed in Figure 33, forming at a modest 200 mV can cause essentially instantaneous failure at as high a radius as 15 nm – a wire that still contains of order 10^6 atoms in the typical neck. These asymptotes are experimentally problematic because resistance cannot be measured with arbitrary speed, due to practical limitations on the electrical noise level achievable. A given apparatus has some minimum signal measurement and response time τ corresponding to a given level of signal accuracy. The longer the current measurement integration time, the more accurate the measurement, and the higher τ.

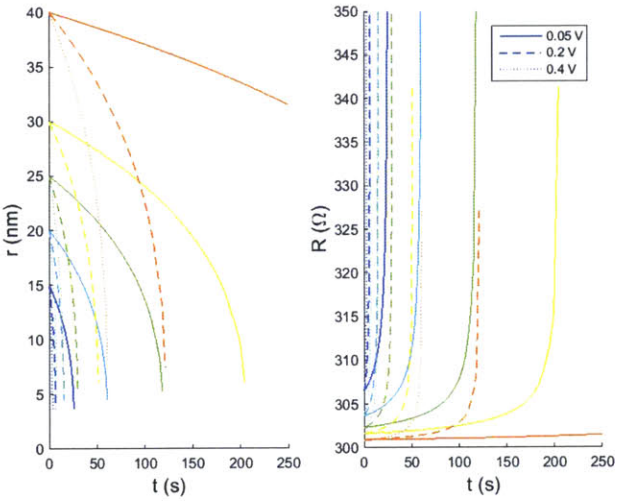


Figure 33. Illustration of positive feedback *i.e.* asymptotic failure in the surface transport model by simulating hypothetical nanowires. Solid, dashed, and double-dashed curves are for fixed voltage 0.05 V, 0.2 V, and 0.4 V respectively, with nanowires of different initial radius indicated with colors. Plotting radius (left) and resistance (right) versus time evolved. In all cases nanowires fail suddenly, with higher biases approaching the asymptotes at higher radius, providing a shorter time-to-failure which is also the controller's time to measure R and respond.

Empirically, we found in our application of the Penn method (previous section) that failure correlates with this sudden formation, making our proposed control objective avoiding it. The voltage pathway $V(t)$ must be constrained such that formation will be guaranteed, subject to τ, to halt at any $r_s>0$, *i.e.* r will not fall below zero. In other words, at any point in the evolution of the nanowire thinning, the time to $r_s > 0$, t_s, must be τ or more shorter than the time to failure, t_f where $r = 0$. r_s can in theory be arbitrary close to $r = 0$ but more practically can be thought of as the radius at which formation enters the stochastic regime, described above, where our model breaks down as resistivity becomes ballistic.[xxv] Put

[xxv] In Appendix L we estimate the effective radius at which the stochastic regime is entered to be approximately 9 nm, from both our own resistance pathway observations and from *in situ* TEM.

in simpler terms, a $V(t)$ pathway is only valid if it provides at least τ time for the controller to recognize that it has reached or surpassed the target radius r_s before $r_f = 0$ is reached. To evaluate that condition we integrate (19), the rate of radius reduction in our surface transport model, from r_0 to 0 and r_0 to r_s, yielding the constraint

$$\frac{1}{3}r_s^3 \leq -\frac{2\gamma\Delta}{R_s}\int_{t_f-\tau}^{t_f} V(t)\,dt. \tag{20}$$

To be viable, any voltage pathway must satisfy (20). This inequality places a further limit however on any single step in the electromigration process *i.e.* any single decision the discrete electrical control makes or instantaneously any value of V at a given time t. Any controller cannot measure and respond faster than τ; therefore in the final τ period of $V(t)$, integrated over above, the applied bias V is necessarily constant. As such we can perform the integration

$$\int_r^{r_s} r^2\,dr = -\frac{2\gamma\Delta}{R_s}V\int_0^{t_s} dt, \tag{21}$$

$$t_s = -\frac{R_s}{6\gamma\Delta\,V}(r - r_s)^3. \tag{22}$$

V and r fully specify dr/dt in our model (19)

$$V = -\frac{R_s}{2\gamma\Delta}r^2\frac{dr}{dt}, \tag{23}$$

yielding

$$t_s = -\frac{1}{3}\frac{(r - r_s)^3}{r^2}\left(\frac{dr}{dt}\right)^{-1}. \tag{24}$$

Assuming r_s small (< 10 nm) relative to r in the final step

$$t_s \cong -\frac{1}{3}r\left(\frac{dr}{dt}\right)^{-1}. \tag{25}$$

In fact these same expression can be derived and applied for the time-to-failure t_f with $r_f = 0$ instead of t_s

$$t_f = -\frac{1}{3}r\left(\frac{dr}{dt}\right)^{-1}. \tag{26}$$

For any given step, t_s (or t_f) must be greater than the chosen measurement-response time τ, yielding

$$-\frac{1}{r}\frac{dr}{dt} \leq \frac{1}{3\tau}. \tag{27}$$

For a nanowire to form gradually, condition (27) must always be satisfied during the electromigration process; it is the 'viable step condition.' Put in more intuitive terms (in fact the same result can be derived from this starting point), to avoid sudden formation the time to failure t_f from the present state (the current r) must always be less than the measurement-response time τ. That allows the

electromigration process to be halted prior to failure. That is visualized in Figure 34a, where the time to failure is plotted against the current nanowire radius and current rate of necking *dr/dt*. When (*r*, *dr/dt*) are such that the time to failure is above τ, formation will proceed gradually in our model. Otherwise, the wire will break suddenly, risking failure in forming a viable tunnel junction. To avoid positive feedback therefore, a controller manipulating *V* must hold $\frac{1}{r}\frac{dr}{dt}$ at or below the setpoint -1/3τ. A fixed-voltage strategy always fails the viability condition, as empirically observed by *Strachan et al.*,[8] *Karim et al.*,[94] and ourselves.[149] In an ideal nanowire such as that we have modelled, the requisite voltage pathway to achieve that could be pre-programmed by fixing

$$\frac{1}{r}\frac{dr}{dt} = K = -\frac{1}{3\tau}, \tag{28}$$

as a constant. That gives us a voltage pathway from (23),

$$V(t) = \frac{R_s}{2\gamma\Delta}K \cdot r(t)^3 = \frac{R_s}{6\gamma\Delta\tau} \cdot r(t)^3 \tag{29}$$

yielding radius evolution pathways such as in Figure 34b. As *r* falls (the nanowire necks), *V* must fall increasingly rapidly to avoid sudden failure. While this ideal *V*(*t*) is illustrative, in practice, as discussed below, non-idealities in the formation process preclude such pre-programming. The *V*(*r*) relationship does, however, provide an envelope function for the allowed values of *V*; at a given *r*, *V*(*r*) can never exceed $\frac{R_s}{6\gamma\Delta\tau} \cdot r^3$.

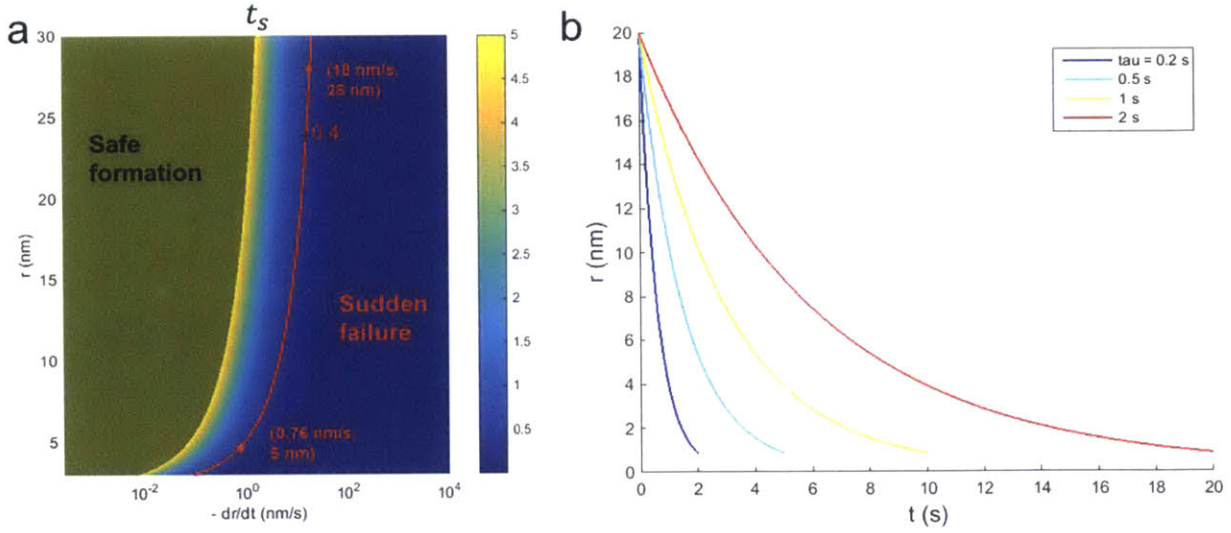


Figure 34. (a) map (color scale) of t_s, the time from the current radius *r* to reach $r_s \cong 0$ given dr/dt, in seconds. Red line indicates a given τ, 0.4 seconds, with formation times higher than τ acceptable and formation times lower risking failure. As radius decreases, a slower rate of decline is necessary to stay above τ. Points indicate *r* and *dr/dt* values along the iso- τ line. (b) *r*(*t*) traces for nanowires with the constraint (20) enforced *i.e.* a perfect controller, meeting that constraint on (*r*, *dr/dt*). As the radius shrinks, the rate of radius shrinking must slow to avoid sudden formation, indicated by the declining slopes. Enforcing a longer τ yields to more gradual formation.

Because r cannot be observed directly, for the control objective (27) to be experimentally enforceable we must change variables to the dependent observable variable resistance $R' = R + R_s$ where R' is the total resistance (measured current I divided by applied bias V), R is the resistance of the nanowire neck, and R_s is the series resistance, including the contact wire. In experiments, we estimate R_s from the initial total resistance R_0' and our known approximate R_0, then subtracting it from R' to measure R for the remainder of the experiment. R relates to r simply *via* the wire resistance as

$$R = \frac{\rho\delta}{\pi} r^{-2}. \tag{30}$$

where ρ is the material resistivity. That yields the controller objective

$$\frac{1}{R}\frac{dR}{dt} \leq \frac{2}{3\tau}. \tag{31}$$

To achieve gradual formation, any controller must maintain *dlnR/dt* at a constant value, at or below this setpoint, tied to the measurement-response rate $1/\tau$. Note that τ captures all limitations on speed, which is generally limited by the noise level (the speed-accuracy trade-off).

Our model therefore corroborates our empirical finding that *dlnR/dt* is the appropriate control variable in nanowire electromigration. We show that is because *dlnR/dt* is proportional to the uncontrolled rate of nanowire necking, including acceleration, *i.e.* the inverse of the time to failure. In Kinetics or mass transport terms, *dlnR/dt* is a measurement of the metal migration rate.

This result of surface transport changes the understanding and approach of electromigrative formation of tunnel junctions. The Penn method, which is the only method employed to date, explicitly controls R rather than *dR/dt*, but as a result the formation process is, as shown above, discretized into sporadic moments of sudden rapid thinning. The resistance response to a given voltage is unpredictable, and as such the only means of using R as a control variable has been the 'shark-fin' cycle, where the resistance is allowed to briefly run away and the voltage is rapidly brought back from the brink. Because *dlnR/dt* is the rate of electromigration, the delta-R monitored in the Penn method is, as shown in the previous section, effectively a control on *dlnR/dt*; in fact any control schema will in effect be controlling *dlnR/dt*. The choppiness of that method however risks sudden formation of the gap at each cycle. Continuous controllers, such as PID control, where a smooth trajectory, subject to noise and physical perturbations,[xxvi] is maintained have not been possible to date because any particular $R(t)$ pathway to target as a setpoint has remained elusive. In contrast, our empirical and theoretical findings suggest that if *dlnR/dt* were controlled, it could be held at a constant value without ambiguity in the formation pathway. With this control variable, constant, continuous control should be possible.

[xxvi] 'Physical perturbations' are discussed below, and are variations in the scattering cross-section which may be due to *e.g.* the reconfiguration of grain boundaries during the electromigration process.

From our inequality (31), we can also observe that because R is positive, holding dR/dt constant, rather than $dlnR/dt$, will also always satisfy the viable pathway condition so long as the setpoint is conditional on the initial nanowire resistance, which is the lowest resistance that will be present. *i.e.*

$$\frac{1}{R_0}\frac{dR}{dt} < \frac{1}{R}\frac{dR}{dt} \tag{32}$$

always holds.

Time-to-failure in Nanowires

Above, we derived the time-to-failure t_f of nanowires due to surface-transport-limited electromigration. Kinetically, $1/t_f$ is the mean rate of electromigration from the current point until failure. While we employed it to derive the conditions for controlling electromigration, it also highlights the physical differences between the surface transport mechanism of electromigration in nanowires and the bulk vacancy transport mechanism of electromigration in micron-diameter and larger wires (the Black model). To illustrate that, we derive the surface transport theory analogue to the Black equation. The Black equation for time to electromigrative failure is

$$t_f = Ae^{\frac{E_V}{kT}}j^{-2} \tag{33}$$

where A is an empirical prefactor, E_V is the activation barrier to vacancy diffusion, k is the Boltzmann constant, T is temperature, and j is the initial current density in the wire. This time to failure has been the standard expression describing electromigration in the semiconductor industry since 1969. As miniaturization has reduced the size of wires to the nanoscale however, it has been corrected by changing the -2 power to a device-specific empirical value between -1 and -2.[10][xxvii]

In a surface transport model, for a nanowire of initial radius r_0, equation (21) yields

$$t_f = \frac{R_s}{6\gamma\Delta \mathrm{V}}r_0^3. \tag{34}$$

Substituting in a fixed current rather than fixed voltage, the expression becomes

$$t_f = \frac{1}{6\gamma\Delta \mathrm{I}}r_0^3. \tag{35}$$

Substituting initial current density j for current in a cylindrical wire, and breaking γ out by its definition,

$$t_f = \frac{2}{3}\frac{qr_0}{\phi\Delta}j^{-1} \tag{36}$$

[xxvii] Exponents more negative than -2, generally no further than -3 but on rare occasion as high as -10, have been measured and assigned to Joule heating diffusion rather than current-stimulated migration.

where again φ is the electromigration scattering cross-section and q is the elementary charge. The scattering cross-section can be described as an Arrhenius process with a mean surface transport activation barrier of E_s

$$\phi = Ae^{-\frac{E_s}{kT}} \tag{37}$$

yielding a time to failure expression analogous to the Black equation,[xxviii]

$$t_f = \frac{2}{3}\frac{qr_0}{\Delta}Ae^{\frac{E_s}{kT}}j^{-1}. \tag{38}$$

or folding the constants into the empirical prefactor,

$$t_f = Ae^{\frac{E_s}{kT}}j^{-1}. \tag{39}$$

Comparing (39) with the Black equation (33) we can see that the difference between surface transport and bulk vacancy migration, *i.e.* between nanometer-diameter and micron-diameter wire regimes, is captured in the scaling of t_f (and $1/t_f$, the mean electromigration rate) with current density. Bulk vacancy kinetics yield a quadratic rate dependence on current flux, whereas surface transport yields a linear dependence. This observation both explains and is supported by the aforementioned empirical corrections to the Black model, adjusting the power between -1 and -2.[10] Previously, the change in power had been assigned to material-specific non-idealities.

Electromigration Controller Design

An electromigration controller in nanowires should therefore use *dlnR/dt* or *dR/dt* as its control variable, setting it, and hence the rate of electromigration, to some constant. To enforce that condition, a control algorithm is required to maintain the setpoint. In theory, any algorithm can be employed, *e.g.* a PID control, or even a pre-set voltage pathway based on our model predictions such as Expression (29) and Figure 34b. In practice however the controller must handle the non-idealities in the system; our model describes idealized surface atom transport, but a few secondary phenomena not included in the model are common.

First there is measurement noise *i.e.* uncertainty in the measurement of R, which is compounded in taking the derivative *dR/dt* (the derivative amplifies high frequency noise). The controller must allow for the measured *dR/dt* (*i.e.* electromigration rate) to deviate from the true *dR/dt* without increasing the voltage to a point where the sudden formation regime is entered. That uncertainty is reduced by using a longer measurement integration time, and hence increasing τ. τ cannot be made arbitrarily long however, both because formations taking longer than 24 hours are impractical and because, as described above (*e.g.*

[xxviii] Treating A as an empirical prefactor as in the Black equation, we can redefine A as $1/A$.

Figure 32b), reversible resistance rises due to Joule heating place a lower bound on the setpoint; if it is set too low *dR/dt* will be due to Joule heating, not electromigration.

Second, there are stochastic perturbations in φ_{CS}, the scattering cross-section, which we hypothesize is due to reconfiguration of the grain boundaries during electromigration as observed in *in situ* TEM (*cf.* reference [22]). When φ_{CS} increases, those sudden changes lead to a sudden jump in *dR/dt* that must be caught and attenuated faster than the system's positive feedback accelerates, as in Figure 35b. Third, the length and position of the neck are not exactly constant, with back-fill (a temporary shift in the atomic flux gradient refilling the neck) yielding occasional periods of declining *R*, as in Figure 35c.

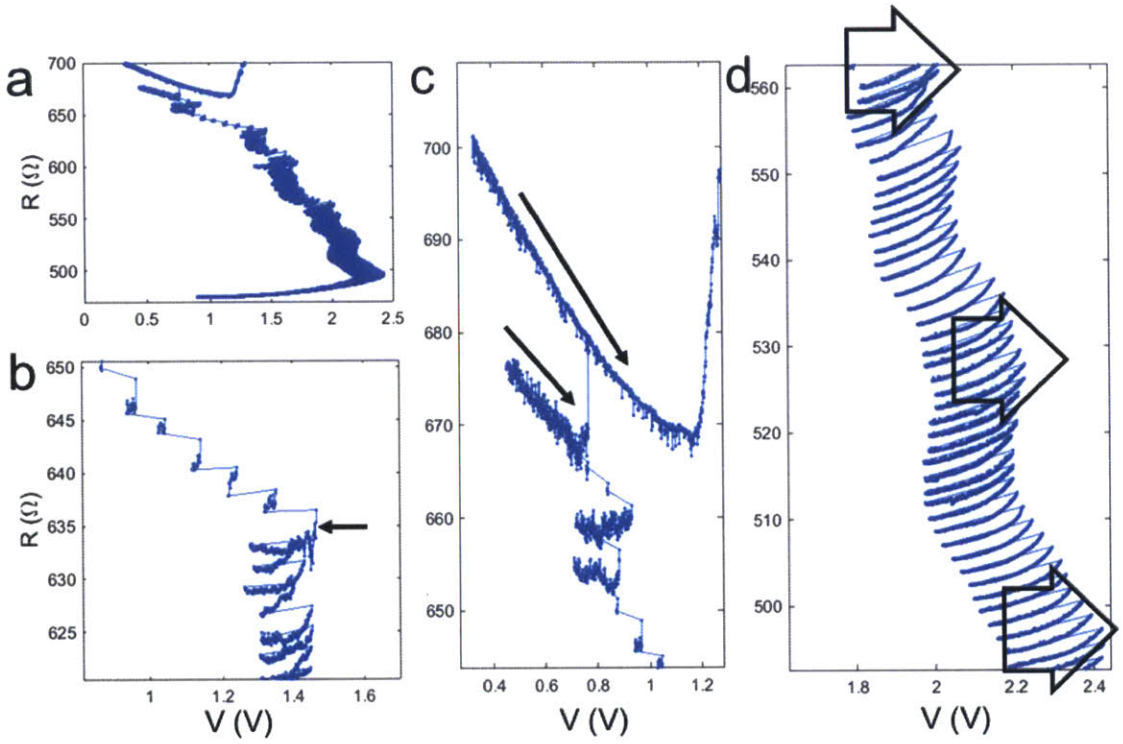


Figure 35. Examples of observed non-idealities in nanowire electromigration. (a) Overall $R(V(t))$ trace of this device. (b) Stochastic increases in φ_{CS}. At the point indicated by the black arrow, the resistance begins to rise much faster with V than previously. (c) Back-fill, with the resistance falling over time rather than increasing. (d) Fluctuations in φ_{CS}, with increasing voltages required to generate resistance rises, as indicated by arrows.

With these considerations in mind, we chose to use an exponential control term to maintain the setpoint *dlnR/dt* by manipulating the applied voltage,

$$\frac{1}{V}\frac{dV}{dt} = -K_e^{E-E_0} + \beta \tag{40}$$

where error

$$E \triangleq \frac{1}{R}\frac{dR}{dt} - \left[\frac{1}{R}\frac{dR}{dt}\right]_{SP} \tag{41}$$

is the deviation from the setpoint, constant $E_0 \triangleq \log\beta / \log K_e$ sets the origin (no response when $E = 0$), β is the desired maximum rate of voltage increase, and K_e is the sensitivity parameter. We also added integral control to manage the noise level,

$$\frac{1}{V}\frac{dV}{dt} = \beta - K_e^{E-E_0} + K_I \int_{t-T_I}^{t} (\Omega(t') - \Omega_{sp})dt' \tag{42}$$

where Ω is *dlnR/dt*, T_I is the time period backwards to integrate over, and K_I is the sensitivity parameter.

When the electromigration rate exceeds the setpoint, this nonlinear control response attenuates the voltage more rapidly the higher the rate (Figure 36a). This acceleration of the dampening voltage response with acceleration of electromigration allows the controller to compete with the wire's positive feedback to avoid runaway. Thus when measurement noise or a stochastic increase in φ_{CS} initiates run-away at the current voltage, the controller can respond to reduce the possibility of sudden formation. When the electromigration rate is too low however, the controller increases the voltage only gradually, with an asymptote rate β. That prevents back-fill, where *dR/dt* becomes negative, from instigating an aggressive, spurious rise in voltage that could quickly tip the system into run-away as a standard PID control would. The choice of K_E determines the severity of the response. By examining the speed of positive feedback in the Penn method data, we estimate that K_E between 1.2 and 1.8 are reasonable, confirming that is the case in applying the controller below.

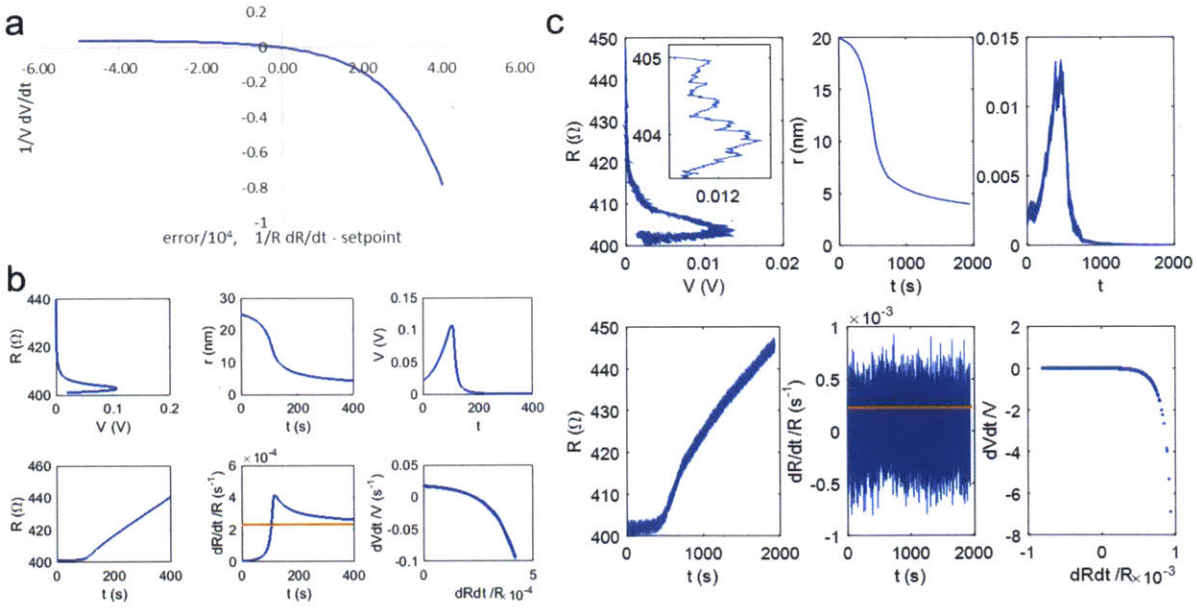


Figure 36. (a) Controller response curve, plotting the controller response *(1/V)dV/dt* to a given measured *dR/dt* deviation from the setpoint, for K = 2. Below the setpoint, this (negative) exponential raises the voltage, but does so diminishingly farther from the setpoint (*e.g.* when *dR/dt* is negative, the voltage still only gradually rises). Above the setpoint, voltage is attenuated increasingly rapidly the farther the deviation, such that an acceleration in *dR/dt* (runaway) is met by a compensating accelerating reduction in *V*. (b) Control algorithm applied to our idealized

surface transport model of a nanowire, showing the voltage and resistance pathway taken in the absence of noise. (c) Control algorithm applied to the same model but now with noise in the measured resistance, showing that while the controller can still maintain the setpoint, the resulting resistance pathway is not as smooth as the noiseless scenario, due to spurious increases and decreases in voltage. Several simulations were performed, but shown here is the highest-noise, with a Signal-to-Noise Ratio less than 1. The noise level is substantially higher than that observed experimentally, but the controller is nevertheless able to avoid sudden formation.

To simulate its effect we applied the controller to our surface transport model of electromigration (Figure 36b,c) by solving the coupled differential equations for the nanowire electromigration rate (Equation (19)) and the controller response (Equation (42)). In the absence of measurement noise (Figure 36b), the controller gradually finds and maintains the setpoint. That leads to an $r(t)$ trajectory similar, after reaching the setpoint, to the 'perfect control' case of expression (29), Figure 34b, with the radius declining at an increasingly gradual rate to maintain the time-to-failure below τ. The resulting voltage pathway has a similar $R(V(t))$ profile to that of the Penn method, with a turn-around voltage reached after which lower and lower bias is required to maintain the desired electromigration rate. The similar profile is expected because the Penn method, as discussed above, *de facto* controls *dR/dt* within bounds. Later (below), when we experimentally apply this controller, we verify that the two trajectories – the Penn method and *dR/dt* control – have similar features.

Adding white noise to the current measurement in our simulation (Figure 36c), the voltage and resistance pathways are overall similar to the noiseless simulation. The deviations from the setpoint however are more substantial. Under higher noise than experienced experimentally however, Signal-to-noise Ratios (SNR) below 1, the controller is nevertheless able to reach and maintain the setpoint, leading to gradual formation. When the SNR exceeds 0.5 there are periods of increasing and decreasing voltage observed (Figure 36c inset), due to mismatches between the measured and actual electromigration rate. For example the real *dR/dt* can rise while the noise, at this level, can cause the measured signal to fall, leading to an increase in V and hence an even faster acceleration of *dR/dt*. We observe this feature in our experimental application of the controller below. The buffering of the exponential control prevents this effect from causing sudden formation.

Application of Electromigration Control

We applied the new controller to our nanowire devices to form electromigrated tunnel junctions. For initial parameterization, *i.e.* setting K_E and the setpoint Ω_{SP}, we used data from the Penn method shark-fins to estimate reasonable values. We then had to tune K_E based on the response of the device during electromigration (Figure 37). We began with values of K_E and the setpoint that are tight (conservative) and gradually loosened them. When K_E is too high, the voltage attenuates rapidly in response to noise-driven, as opposed real, increases in resistance near the setpoint such that voltage 'cut-

backs' (sudden drops) are observed where R declines (Figure 37), indicating that V did not have to drop so substantially to avoid precipitious formation. Such cutbacks are analogous to the shark-fins observed in the Penn method, particularly when the voltage attenuation response is set high; for example in the original publication of the Penn method, the voltage was reduced to zero.[8] When K_E is optimal, voltage attenuation is associated with resistance *increases*, such as in Figure 38, indicating that the controller is competing with/managing the system positive feedback. When the voltage frequently undergoes reductions of that sort *i.e.* sharp V reductions with increasing R, that suggests that the setpoint is too high such that the system is regularly dangerously close to run-away given τ.

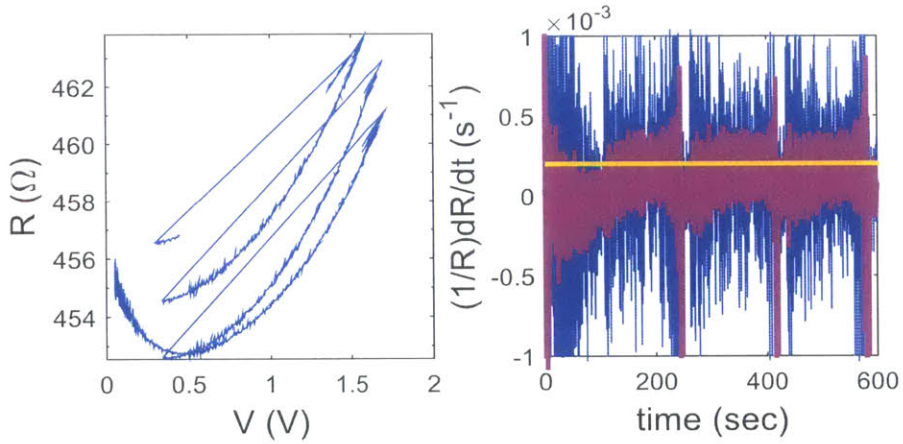


Figure 37. Example trace of tuning controller parameters with a nanowire device. In this trace, K_E is too tight (too high), leading to excessive attenuation of the voltage observable from the sharp cut-backs in voltage despite a resulting decrease in resistance. If K is tuned appropriately, sudden voltage drops will be in response to real (rather than noise-driven) increases in resistance such that R still increases even when V is dropped, indicating that the controller is competing with the system's positive feedback loop. Left plot is R-V, right plot is *dlnR/dt* with time. The purple curve is the filtered signal used for control (Gaussian lowpass filter), while the orange line is the setpoint. Despite the over-tightened K, the controller is able to generally maintain the setpoint. Even with the premature voltage reductions, the system resistance is consistently rising, suggesting that eventually, like the Johnson method, formation would occur.

Once the parameters of the controller (42) were appropriately tuned, the control algorithm worked as predicted and intended, from the very first device tested (Figure 38). To date, all electromigrative formations of tunnel junctions in the literature have been erratic. An $R(t)$ setpoint pathway has not been established; as we observed in our application of the Penn method, the response of R to V is not predictable. As a consequence, the Penn method produces discrete jumps and halts in resistance, rather than a continuous evolution. In contrast, our application of *dlnR/dt* control approaches and then maintains a stable setpoint (Figure 38b). As visible in the inset of Figure 38c, where R continues to rise when V is

attenuated,[xxix] the voltage attenuation anticipates, rather than reacts to, runaway electromigration. Also visible in the inset are the same $R(V)$ fluctuations observed in our simulations (Figure 36c), which we can therefore assign to noise-instigated deviations between the measured and actual dR/dt. This produces a stable, gradual, and smooth rise in $R(t)$, as in Figure 38a, rather than erratic jumps and plateaus. This formation, and subsequent ones with this method (below), are the first times that a chosen, pre-programmed resistance pathway has been achieved in electromigration. In each cycle of the Penn method, as discussed above, the stable dR/dt regime is passed, electromigration accelerates to cause a jump in R, which is then caught and suddenly halted. In our dR/dt control, the R-V response is continuously monitored to produce a stable dR/dt; the sudden resistance rises at the edge of the shark-fins are pre-empted. As predicted, $dlnR/dt$ control produces gradual electromigration.

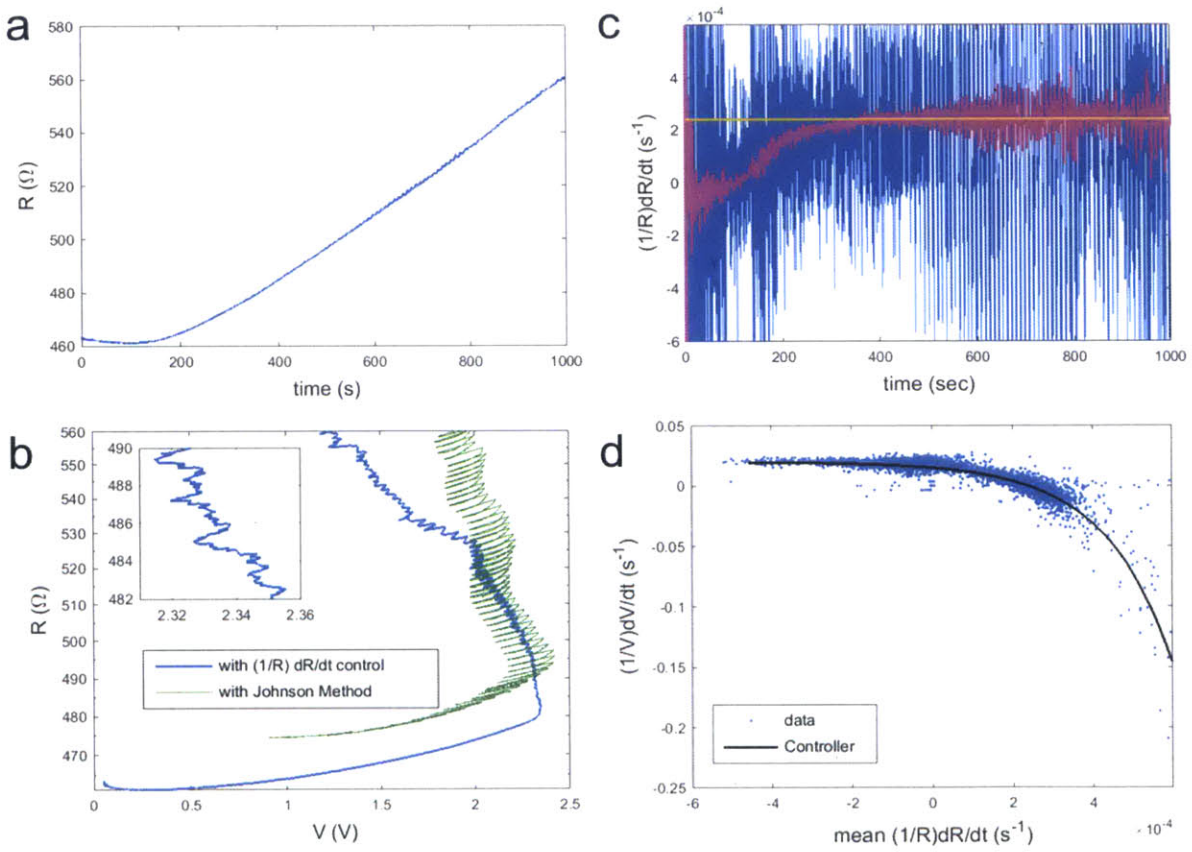


Figure 38. Experimental application of the control algorithm to a nanowire for electromigrated formation of an electron tunnel junction; the first device we tested. (a) Resistance over time, showing that the controller is able to maintain a smooth linear rise in resistance, avoiding sudden formation. (b) Resistance-voltage trace (blue), compared to data from a similar nanowire using the Penn method (green). As predicted by our surface transport model, the controller yields a qualitatively similar formation pathway as the Penn method. Once the setpoint is attained, at the 'turn-around voltage,' the resistance rises consistently. Unlike the Penn method, resistance-reducing

[xxix] As opposed to the Penn method sharkfins, visible in the same figure panel, where the voltage attenuation generally yields a *reduction* or stagnation in resistance.

cutbacks are not necessary and rapid increases in resistance are avoided outside of stochastic events. Inset zoom shows that voltage reductions are associated with resistance rises, rather than falls in the Penn method, reflecting that the *dR/dt* control is continuously managing the electromigration rate. (c) control variable *dlnR/dt* over time. Blue curve is the raw measurement data, magenta curve is the noise-filtered measurement data (applied to controller), and the orange line is the setpoint. The controller brings the system up to the setpoint, and then is able to maintain a consistent *dR/dt*. Maintaining a setpoint in *R* has eluded researchers to date, vindicating our model predictions and empirical observations suggesting *dR/dt* as a control variable. (d) Controller response, intended and measured. The blue dots are measured voltage changes corresponding to measured resistance changes. The black line is the control equation, showing that the intended relationship is generally held despite the effects of noise and controller response delay.

Our exponential controller also handles the non-idealities as anticipated. The voltage modulation maintains the target *dR/dt* setpoint despite a low signal-to-noise ratio (Figure 38b). While noise and the finite delay in controller response lead to variation in the measured *dV/dt* as a function of the control variable *dR/dt*, the desired control expression (42) is overall maintained (see Figure 38d, comparing the measurements, blue dots, to the control equation (42), the black line). Qualitatively, the behavior matches the noisy simulations, Figure 36b, with noise potentially explaining the voltage oscillations; compare the Figure 38b inset with the Figure 36c inset. When back-fill is encountered (*e.g.* the beginning of trace Figure 38a), the exponential horizontal asymptote successfully avoids an aggressive voltage response. *Vice versa*, sudden leaps in resistance are caught by the exponential vertical asymptote with a correspondingly rapid attenuation of voltage. These responses vindicate the selection of a negative exponential control term.

As predicted by our application of the control to the surface transport model (Figure 36), the same envelope features of formation as in the Penn method are observed. There is a turnaround voltage in the *R-V* curve (Figure 38b) where expeditious formation initiates, after which lower and lower voltages are required to continue formation at that rate, predicted to be caused by the thinning of the wire. Combined, our observations of the Penn method, our surface transport model predictions, and the verification of those predictions here, suggest that such a pathway is a general property of gradual formation; the voltage must decline as the wire thins.

Eventually the nanowire necks gradually to reach the 'stochastic regime,' where only a few dozen atoms remain leading to a conductivity only 5 to 20 times the conductance quantum G_0 (Figure 39). Coincidentally, in this region the controller is still able to continue formation until a final break in the wire occurs. The mechanism of control however is no longer valid; the Penn method or even a simple single-voltage hold all perform similarly in forming the final tunnel junction from this point. The process is stochastic, with a single or few atom dislocation being required to step up the resistance. As such all that is required is a reasonable applied voltage and monitoring of the resistance to ground the leads once the junction forms.

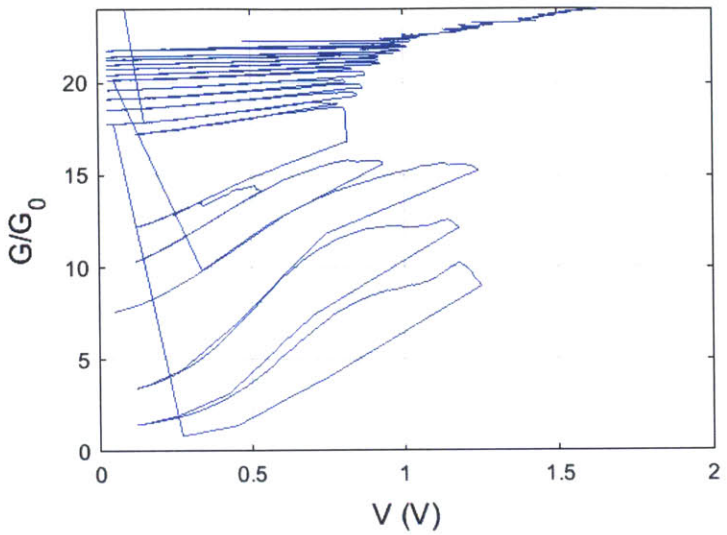


Figure 39. Controller performance in the stochastic regime. Conductivity normalized to the conductance quantum G_0 versus voltage (with time) trace. When the nanowire necks to a few conductance quantums, continued narrowing is achieved through quantized dislocations of atoms, yielding stepwise jumps in resistance, until the wire finally breaks by falling below G_0. The continuous stretches of conductance increase are due to the nonlinearity of the current-voltage curve at this small 'diameter.'

We applied the electromigration control to 24 nanowire devices, varying the target *dlnR/dt* setpoint *i.e.* the rate of electromigration to examine whether that has an impact on the formation pathway or outcome. As with the Penn method, we characterized the formation of a tunnel junction as successful or unsuccessful depending on whether a measurable tunneling current is obtained after formation, regardless of how robust the 'successful' junction is to elevated-voltage electrical testing. We found that while the *dlnR/dt* controller always generated a smooth, continuous *R(V(t))* pathway, it did not always avoid precipitous formation (Figure 40). Again we observed a correlation between precipitous formation and device failure. We found however that unlike the Penn method, the yield of successful tunnel junctions was dependent on the control parameter, the *dlnR/dt* setpoint (Figure 40a). Three regimes were observed. When *dlnR/dt* is too high, over 0.07 %*R*/s, the exponential control is constantly competing with run-away, leading to precipitous formation and failure in 60% of devices. When *dlnR/dt* is too low however, below 0.04 %*R*/s, the yield also falls. That low rate of measured resistance change convolutes any electromigrative effect with Joule heating, with the latter phenomenon identified as reversible increases in *R*. When Joule heating masks the wind-force driven electromigration, the observed *dlnR/dt* may be decoupled from the rate of electromigration, and the controller is no longer adequate. The low yield suggests that is the case, and provides further evidence that the mechanism of electromigration in nanowires is not Joule heating. When the setpoint is optimal, between 0.04 and 0.06 %*R*/s for our devices, the yield approaches 100%. That high success rate is despite the current limitations of device fabrication; with current electron beam lithography and metal deposition techniques we can produce consistent

nanowire device dimensions, but cannot control the crystal structure. In particular, grain boundaries will vary widely from device to device, such as observed in *Taychatanapat et al.*[22] and *Strachan et al.*[29] As discussed in the Introduction, these heterogeneities dominate the rate of electromigration, leading to device-to-device variation and, as the grain boundaries evolve during electromigration, variation within the device. Nevertheless when well-tuned the *dlnR/dt* control is able to manage these uncertainties and produce high yield. In demonstrating that under the right circumstances tunnel junction yield can approach 100%, our controller establishes that electromigrative tunnel junctions have the potential to be commercially viable.

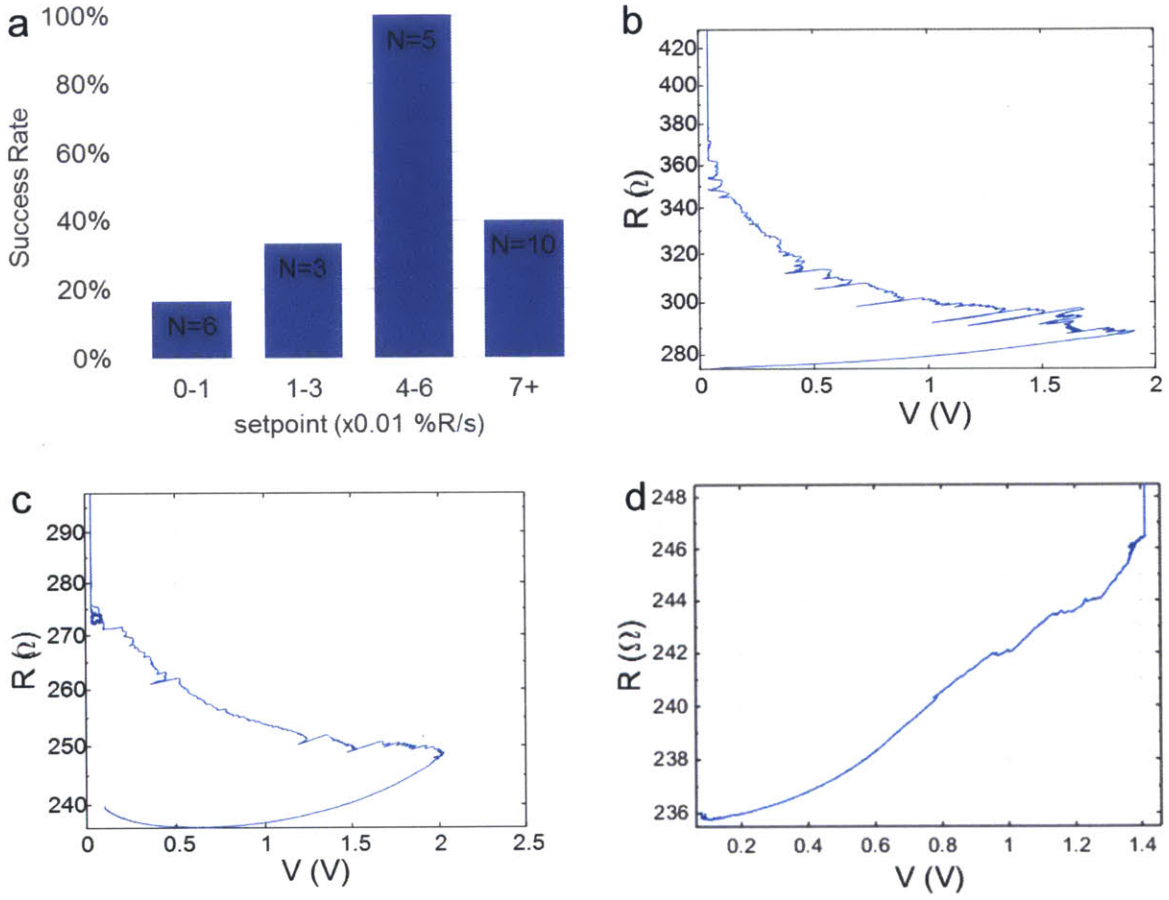


Figure 40. (a) Rate of successful formations as a function of *(1/R)dR/dt* setpoint. There is an optimum setpoint between 0.04 and 0.06 %*R*/s where *dR/dt* is irreversible (*i.e.* distinguishable from Joule heating) but not too high such that precipitous formation is avoided. At this optimum the yield of successful tunnel junctions approaches 100%. Number of devices tested at each setpoint indicated in data labels. (b,c,d) Example successful (b,c) and failed (d) formations with the *(1/R)dR/dt* control, at setpoints 0.10 (b), 0.05 (c), and 0.006 (d) %*R*/s.

The success of our *dlnR/dt* control and the dependence of electromigration pathway and outcomes on the setpoint have several implications. Continuous control of the electromigration process, where the $R(V(t))$ pathway can be pre-specified and maintained, was previously impossible. In addition to its potential to improve yields, as we've shown, continuous control could provide greater insight into the

electromigration process. Unlike the Penn method, the factors effecting electromigration success can be probed and understood. When fabrication methods advance to reduce the variance and heterogeneity in the nanowire crystal structure, both the yield and physical insight provided by *dlnR/dt* control are expected to increase. The ability of the surface transport mechanism to predict the features of *dlnR/dt*, such as gradual formation and the turnaround voltage, provides further evidence in support of the validity of that model of nanoscale electromigration. The sensitivity of electromigration yield and pathway to the *dlnR/dt* setpoint supports the proposition that *dlnR/dt* is proportional to the rate of electromigration, providing a direct measurement of kinetics. That sensitivity also affords new opportunities for designing the electromigration pathway by designing a varying *dlnR/dt* pathway, as opposed to the constant setpoint we have employed here. Finally by showing that under the right circumstances it is possible to approach total yield, we demonstrate the technological viability of the electromigrative tunnel junction technology.

III. ELECTROMIGRATION *VIA* A CARBON NANOTUBE–AU NANOWIRE INTERFACE

Summary of Findings

In the previous Chapter, we showed that the bulk vacancy transport limited kinetics (the Black Model) that describe micron-scale electromigration do not apply to sub-micron diameter wires, and rather that a surface transport limited model is consistent with experimental results.[149] Specifically we found that the time to failure of nanowires is not predicted by the Black model, but predicted without measurable error by a surface transport model. Since its discovery the Black transport mechanism has been the standard quantitative model of wire failure in the semiconductor industry, used to predict the time-to-failure and allowable current flux.[2,10] With miniaturization to the nanometer scale correction factors have been introduced, in particular changing the current scaling of electromigration rate from quadratic to linear,[10] which we showed in Chapter II are explicable as a transition from the bulk vacancy transport limited kinetics to surface transport limited kinetics. In addition to its quantitative properties however, electromigration in micron-sized wires has a variety of characteristic qualitative features which may also break down at the nanometer-scale. In Chapter I we identified one such feature – void formation. In micron-scale wires, electromigration eventually yields voids in the material from the accumulation of vacancies at the point of highest atomic flux,[10,16,18] consistent with a bulk vacancy transport mechanism. In nanowires however, vacancies do not form and the wire instead thins, or 'necks,'[29,83] consistent with the surface transport mechanism we propose.

The question remains whether other characteristics of electromigration at micron scales change at smaller scales. Of particular technological interest is the point of electromigrative failure *i.e.* the location along the wire where material is lost most rapidly and the wire ultimately breaks. In micron diameter wires it is ubiquitously observed, across nearly all materials and architectures, that electromigrative failure occurs at material interfaces whenever they are present.[16-21] The net rate of gain or loss of atoms in a given cross-section of the wire is dictated by the gradient in the electromigrative flux. In a perfect crystal, the flow of atoms into and out of any given cross-section would be equal such that there is no net loss of material or resulting change in crystal dimensions. In a real crystal however gradients in structure, chemical composition, and/or temperature lead to a net re-allocation of atomic material and the above-mentioned void, neck, or protusion formation. At material interfaces, even between two crystals of the same chemical composition,[20] the barrier to atom vacancy diffusion is substantially lower than the adjoining bulk crystal, making the flux gradient large. Under usual circumstances this electromigration gradient is substantially larger than any other along the wire such that it dominates the rate of material loss, becoming the point of failure. The interface can be thought of as a grain boundary that spans the full wire diameter rather than a single grain. Whether this phenomenon holds in nanometer-diameter wires is

of technological import. For integrated circuit design, it dictates whether interconnects must be made wider or otherwise more robust than the isolated wire lengths, increasing the circuit footprint and potentially complexity. In the formation of tunnel junctions, discussed in Chapter II, control over the location of gap formation is required to interface single-molecule junctions with an electronic network or nanopore.[54,60] In monometallic nanowires, that has yet to be achieved because failure nucleates at an unpredictably-located grain boundary, where in the absence of other material interfaces the gradient in the flux of metal atoms is maximal.[29] If material interfaces present the fastest rate of electromigration in nanowires as they do in micron-diameter wires, that can be exploited to position the tunnel junctions, which form at the point of failure, using a highly heterogenous material interface such as between a covalently-bonded carbon nanotube (which is electrically conductive) and a metallic-bonded Au nanowire. That behavior could be used to locate tunnel junctions with the requisite nanometer precision, although it would also limit the feasible designs of complex multi-component structures on the 1-10 nm scale. On the other hand, if the dominance of interfacial mass transport no longer holds then arbitrarily complex structures may be feasible from an electromigrative failure standpoint, and the precise positioning of tunnel junctions will remain possible albeit requiring fine control over crystal structure.

In this work, we investigated a covalent-metal interface in the formation of electron tunneling gaps, flowing current between a single-walled carbon nanotube and Au nanowire. We found that the rule of micron-sized wires, where failure occurs ubiquitously at interfaces even when between the same chemical composition, breaks down at this scale; interfacial heterogeneities did not limit the gap nucleation kinetics. Instead, electromigrative failure and tunnel junction formation frequently occurred within the Au nanowire or the carbon nanotube, implying that crystal heterogeneities – *e.g.* defects and grain boundaries – can have comparable or superior diffusion gradients. This system is the first reported electron tunnel junction formed by flowing current between a metal and covalent material. It is of particular interest because carbon nanotubes (CNTs) can be used as nanopores to translocate DNA,[59,150] and potentially other biomolecules such as glycans and peptides, in which case interfacing the pore with a tunnel junction is necessary to read out the sequence.[xxx] [151] Using the nanopore as an electrode to form the junction at an interface with a Au nanowire (NW) is one of few possible methods of positioning the gap appropriately. Intuition from micron-scale systems, where even single-component metal-metal interfaces limit electromigration, is that necking would occur at the carbon nanotube (CNT)-nanowire (NW) interface; with appropriate bias polarity, Au atoms would migrate *away* from the interface along the Au NW, but not *into* the interface from the CNT, leading to a large net flux of atoms

[xxx] An electron tunneling measurement is necessary for single-base-pair or single Amino Acid sensitivity because the size of those individual moieties is smaller than the Debye length of any reasonable solution, preventing single-moiety spatial resolution using nanopore conducitivity measurements.

out of the interface. We found surprisingly that the interface does not necessarily nucleate gap formation, and that it is possible for crystal defects in the CNT and Au NW to present limiting formation kinetics. Using two-terminal bias application and feedback techniques, we found that we could break either the Au NW or the CNT along their lengths, and that the interface was not inherently the dominant failure mode.

Device Fabrication

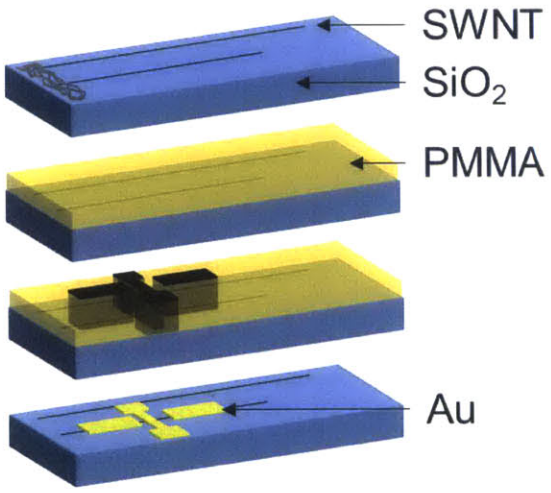


Figure 41. Cartoon illustrating carbon nanotube growth, electron beam lithography, and metal deposition fabrication process described in the text.

To be able to drive significant electromigration of Au atoms, several properties of the Au-SWNT device were required. First, the Au nanowire, as discussed in the *Experimental Methods* of Chapter II, must be less than 0.001 μm^2 in cross-sectional area. Otherwise electromigrative failure is frequently explosive or, as observed in *Xiang et al.*,[88] constant back-fill inhibits formation. Second, the carbon nanotube must directly interface with the Au crystal *i.e.* without any immobile materials interceding; that precludes the use of a Ti or Cr adhesion layer such as those we employed to stabilize our all-Au devices of Chapters I and II. Third, the lengths of the Au nanowire and carbon nanotubes must also be made small enough such that the series resistance R_S is no more than a few kiloohms. High series resistances present two impediments to electromigration. The first impediment is that it has been shown to increase the incidence of sudden failure and reduce the yield of successful tunnel junctions, due to an acceleration of the voltage drop across the nanowire as wire resistance rises.[93] The second impediment is that the higher the series resistance, the harder it is to detect the resistance increase of the nanowire due to thinning; for example if the nanowire resistance is 10 Ω and R_S is 90 Ω, a 1% rise in nanowire resistance can be measured as a ~0.1% fall in the current signal, but if R_S rises to 0.9 kΩ, a 0.01% fall in the current signal must be detected at the same noise level. In the extreme case of $R_S > 1/G_0$, the conductance quantum, it is impossible to even detect the formation of a tunnel junction without testing for ballistic conduction. To achieve R_S less than a few kiloohm, the Au nanowire must be no longer than a hundred microns or so, and for typical Single-Walled Nanotube (SWNT) resistivities of 10 MΩ/cm the nanotube must be no longer than 2 μm.

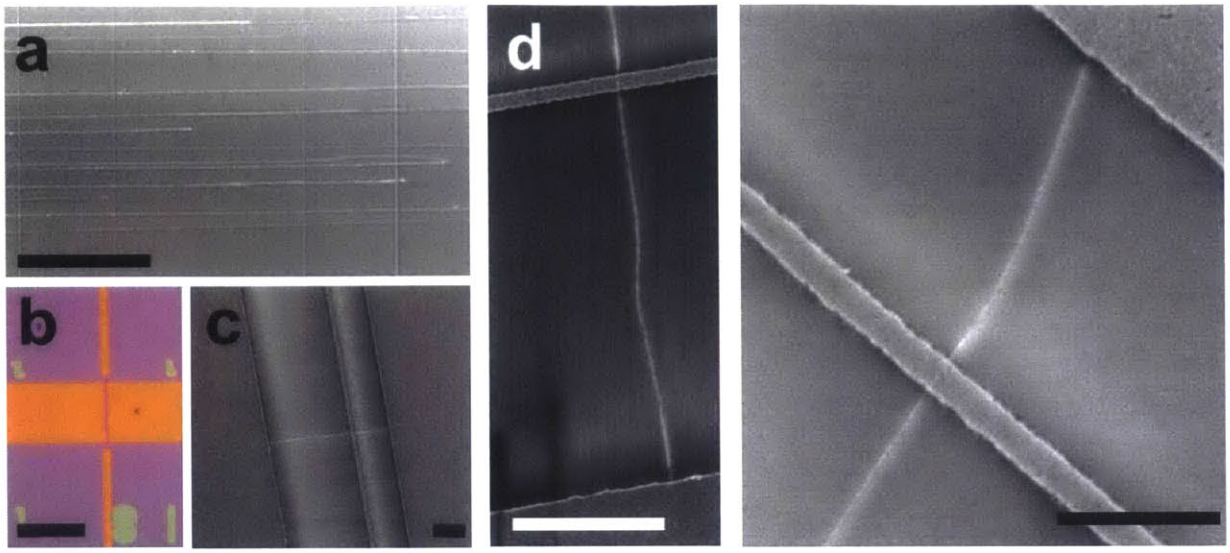


Figure 42. (a) Scanning Electron Micrograph (SEM) of carbon nanotubes grown using chemical vapor deposition on a SiO_2 substrate (scale bar 1 mm). (b) Optical Microscope (OM) image of the Au nanowire and CNT contact pads defined by EBL on top of the CNT (not visible) (scale bar 100 μm). (c,d) SEM images of the nanotube-nanowire interface (scale bars 10 μm, 1 μm, 500 nm). The nanotubes are the bright white lines, appearing larger than their actual diameters (1.1-1.8 nm, measured using Raman spectroscopy of the radial breathing mode) in SEM because they diffuse surface charge buildup on the insulative SiO_2 top layer, thereby attracting primary incident electrons.

We achieved the desired device properties using CVD growth of isolated carbon nanotubes on an SiO_2 chip, followed by Electron Beam Lithography and metal evaporation for the formation of nanowires and SWNT contact wires. This device fabrication process (Figure 41) was begun by growing centimeter-long single-walled carbon nanotubes (SWNTs) using a 'kite growth' chemical vapor deposition (CVD) process on top of an inert SiO_2 surface,[152] exactly as was done in the fabrication of SWNT nanopores by *Choi et al.*[150] and *Liu et al.*[59] To do so, SWNT (NanoC) with resident Fe nanoparticle impurities, those used to catalyze the growth of those tubes, was suspended in aqueous solution with a Sodium Dodecyl Sulfate (SDS) surfactant by sonication and centrifugation. A few micro-liters of the solution was then deposited as a thin strip on the edge of a 10 x 14 mm^2 SiO_2 chip (300 nm SiO_2 on top of a p-doped Si single crystal). The chip was then placed in a CVD chamber, where hydrogen and methane gas were flowed laminarly at a fixed rate over the sample at 900 °C. The Fe nanoparticles catalyze the formation of a nanotube ring structure, which it is hypothesized continues to elongate the nanotube at the Fe nanoparticle surface, pushing the nanoparticle away such that it is carried into the gas flow stream analogous to a kite being pushed away from the substrate by convective flow. After a period of time chosen to allow for a sufficiently long nanotube, typically 5-20 mm, the reagent flow was replaced with an inert nitrogen flow and the chamber was slowly cooled as the nanotube eases down onto the substrate, adhering strongly via van der Waals interaction. The sample was then cleaned of organics prior to the Au

nanostructure fabrication by sonicating in acetone and isopropyl alcohol followed by baking in nitrogen (to maintain SWNT chemical stability) above 300 °C to desorb impurities.

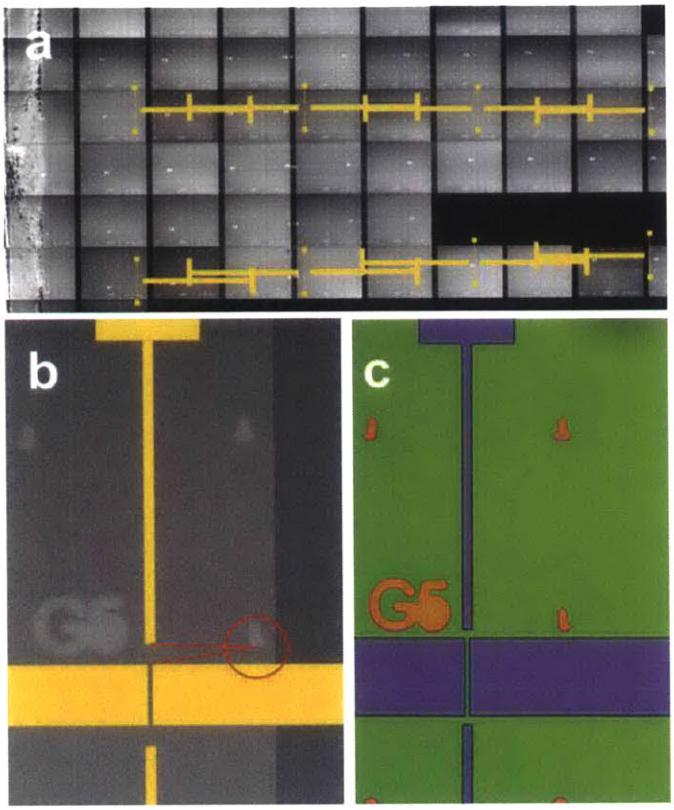


Figure 43. (a) Example alignment of EBL features to write (yellow) with the SEM images of SWNT (grayscale); image 12.1 mm long, SWNT growth catalyst visible on left (SWNT grown left to right). The SWNT are identified and their positions define relative to the chip markers by stitching together contiguous SEM frames as shown. The features to write in the EBL process are then overlaid on the SEM images and their positions measured relative to the same markers (*e.g.* Panel b). As with this chip, a single chip can have multiple viable SWNT and a single long nanotube can be used to produce multiple devices. Comparison of the planned feature positions (b) to the actual feature positions (OM of PMMA negative in (c)) show that the devices are located as anticipated. Before and after testing SEM is again performed to verify that the SWNT and the Au nanowires are indeed in the correct locations, structurally intact, and to observe outcomes on morphology.

Prior to growth, markers were defined on the SiO_2 surface to allow co-location of the Au structures and the SWNT. After growth, individually isolated nanotubes – those without another tube within 2 mm – were identified and their locations registered with respect to the markers using Scanning Electron Microscopy (SEM); despite their size, SWNT on insulative substrates appear clearly in SEM as they collect charge and project an electric field.[153] The Au nanowire, nanowire contact pads, and SWNT contact pads were then designed using CleWin© software, and their desired positions calculated relative to a reference marker using the SEM images. The nanowires were designed to be < 100 nm wide, achieved by varying the amount of electron exposure in a 60 nm wide digital mask. Prior to electron beam lithography, we performaed Raman spectroscopy on the nanotubes of interest to identify their diameter and whether they were single- or double-walled from the Radial Breathing vibrational mode (RBM); the RBM is a cylinder mode characteristic of the tube diameter, allowing us to extract diameter d_t from RBM frequency ω_t according to *Jorio et al.*[154]

$$d_t = {}^{248}/_{\omega_t} \tag{43}$$

A 950,000 g/mol molecular weight Poly(Methyl Methacrylate) (PMMA) electron beam resist diluted to 4 wt% in anisole was spin-coated at 3,000 RPM for 45 seconds on to the substrate, similar to

the process described in Chapter II. The resist was then baked at 180 °C on a hot plate, measured by IR thermometer, for 5 minutes to cure. See Chapter II for discussion of conditions and resist thickness optimization. To perform Electron Beam Lithography (EBL) we used a Raith 150 apparatus at the MIT Research Laboratory for Electronics (RLE) Scanning Electron Beam Lithography (SEBL) facility. We exposed the resist using a 30 keV acceleration voltage, but varied the aperture depending on the features. For the Au nanowire, we used a 30 μm aperture to achieve the desired feature resolution, but the write time with a 30 μm aperture was prohibitively long for the SWNT and Au nanowire contact pads. For those larger features where high spatial resolution was unnecessary, we wrote them separately with a 120 μm aperture to increase the current flux, and hence reduce the write time, by a factor of ~ 16. To align the features written with the 30 μm and 120 μm apertures, as well as align those features to the carbon nanotube, we referenced their positions to the markers on the substrate, achieving ±4 μm accuracy in position. This accuracy limited the minimum length of the Au nanowires, as to sufficiently guarantee that they would overlap with the nanotube they could be no shorter than approximately 10 μm.

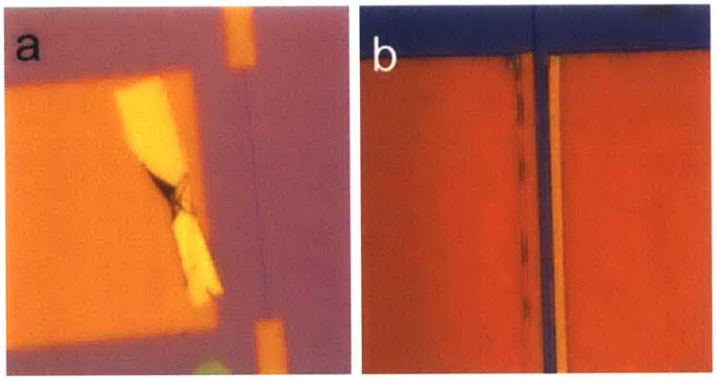


Figure 44. Optical microscope images of poor liftoff examples. Due to the lack of an adhesion layer (to have direct Au-SWNT contact), it was necessary to use low temperature acetone for liftoff, creating some device failures. (a) Au foil on the PMMA surface could be pulled back to the surface, where it gets stuck, by undissolved PMMA anchors. (b) small features, such as the 1-2 μm gap between the SWNT contact pad and the Au nanowire, would sometimes not lift off. Approximately 50% of devices were lost to liftoff, and many surviving devices have only the minimum of two contacts.

After exposure, the resist was developed in 1:3 Methyl Isobutyl Ketone (MIBK):IPA as described in Chapter II. Au was deposited into the resulting nanowire and contact pad negatives using electron beam evaporation of Au. We deposited thicknesses ranging from 15 to 20 nm. Evaporation was performed below 10^{-6} torr vacuum pressure and the thickness was monitored using an *in situ* Quartz Crystal Microbalance (QCM). The Au thickness was confirmed with Atomic Force Microscopy (AFM). Without an adhesion layer, we could not use hot n-Methyl Pyrrolidone, as we did in Chapter II, to liftoff the PMMA mask; doing so dissolved the Au nanowires. Instead, we used a gentler acetone bath held at

25 °C. This occasionally led to poor liftoff, with the Au foil falling back down onto the surface and rendering the device useless (Figure 44), but in cases where careful agitation was able to prevent this from occurring the liftoff was clean. To remove residual PMMA in this liftoff procedure it was necessary to soak the sample in warm acetone for at least 60 minutes prior to drying. Devices were examined for in SEM after fabrication to verify the liftoff, absence of PMMA residue, and correct positioning of the Au features *i.e.* their co-location with the SWNT. The samples were stored under nitrogen atmosphere until testing. As intended, the resulting Au nanowires were 100 ± 10 nm wide, 120 μm long, and with less than a 2 μm separation from nanotube contact pads (*i.e.* the SWNT lengths channel lengths were less than 2 μm). The resulting nanotube and nanowire resistances, discussed further below, were at parity with one another. Because of the length of the nanotubes, each nanotube chip produced as many as 4 Au-SWNT interface devices, although in practice after liftoff only 1 or 2 typically survived. In total, 20 viable devices were produced and tested.

Electrical Testing

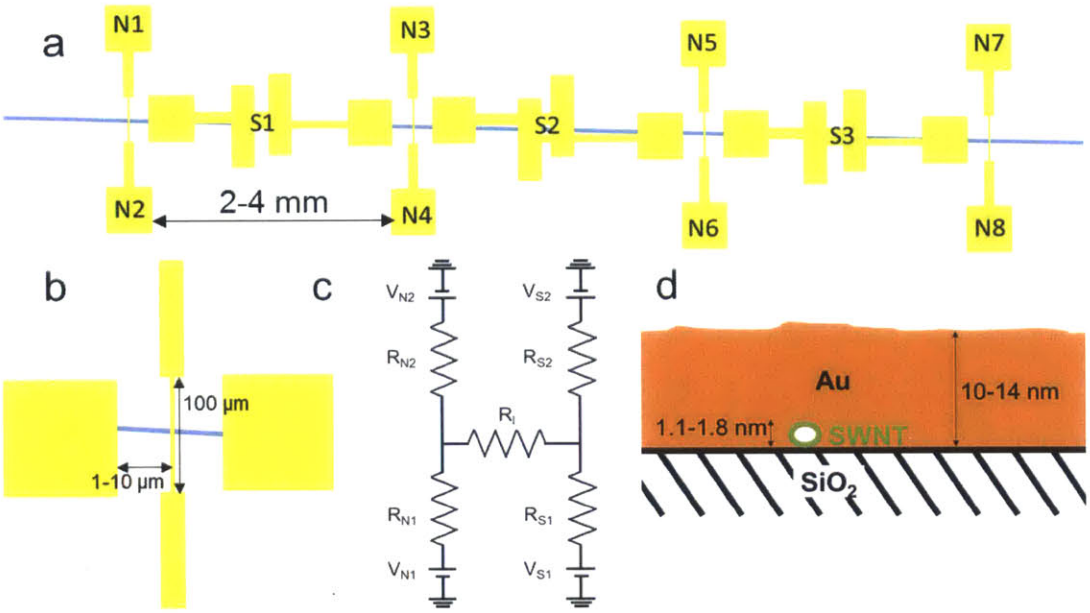
Please see Chapter II for discussion of the electrical testing apparatus; the same probe station, source-measure units, *etc.* were used here as in our examination of the monometallic Au nanowire electromigration in that chapter.

For this work, each device had up to four contacts – two SWNT contact pads and two Au nanowire contact pads. In practice however, damage during liftoff (Figure 44) constrained several devices to the minimum of two contacts – one on the nanotube and one on the nanowire – necessary to conduct the experiment. We labelled the four points of contact N1, N2, S1, and S2, corresponding to nanowire (N) and SWNT (S) contact pads (Figure 45a). For each device, we measured the resistance between each set of contacts at 100 mV bias to extract the resistance values in the device's simplified circuit diagram (Figure 45c). We found that there was a significant material interface resistance R_I between the nanotube and the nanowire, measured as

$$\begin{aligned} R_{S1-N2} + R_{S2-N1} &= (R_{S1} + R_I + R_{N2}) + (R_{S2} + R_I + R_{N1}) \\ &= 2R_I + (R_{S1} + R_{S2}) + (R_{N1} + R_{N2}) = 2R_I + R_{S1-S2} + R_{N1-N2} \\ \rightarrow R_I &= \frac{1}{2}(R_{S1-S2} + R_{N1-N2} - R_{S1-N2} - R_{S2-N1}) \end{aligned} \tag{44}$$

where $R_{X\text{-}Y}$ is the resistance measured between the contacts X and Y. Nanowire and nanotube resistances *e.g.* R_{N2}, R_{S1} were similarly calculated. Both by bisecting the nanotube with the nanowire and by varying the separation between the nanotube contact pads, the scaling of nanotube resistance with nanotube length was checked, albeit between different chiralities in the latter case, and found to be linear as expected. Across all devices, the measured nanowire conductivities ranged from $4\text{x}10^4$ S/m to $10\text{x}10^4$ S/m, SWNT

conductivities from 0.5×10^7 S/m to 3×10^7 S/m, consistent with previous literature measurements ranging from 10^5-10^8 S/m, and Au-SWNT interfacial resistances ranged from 2 to 3 kΩ.



e

	N1	N2	S1	N3	N4	S2	N5	N6	S3	N7	N8
N1	7Ω	4.7kΩ	300kΩ								
N2		4.8Ω	300kΩ								
S1			5.6Ω	∞	130kΩ 200kΩ	400kΩ 780kΩ	1.4MΩ	∞	1.8MΩ		
N3					∞						
N4						300kΩ					
S2									1MΩ		
N5								∞	560kΩ		
N6									∞		
S3										405kΩ	416kΩ
N7										4.3Ω	3.8kΩ
N8											

Figure 45. (a,b) cartoon of device numbering system for a single nanotube and device geometry. Devices with nanotube lengths (nanowire-contact pad separations) ranging from 1 to 10 μm were tested. (c) simplified electrical circuit diagram of the device. The nanotube and nanowire are each bisected by the other, allowing us to identify the resistance of each half. We also measure a significant interfacial contact resistance R_I between the two however, ranging from 1 to 5 kΩ. We measure R_I by subtracting the total nanowire resistance (the resistance measured between *N1* and *N2*) from the resistance measured of the individual segments *via* the nanotube (*e.g.* the resistance between *S1* and *N1* and *S1* and *N2*) minus the associated nanotube resistances. In principle, each of the four contact pads can be biased independently. (d) cartoon of the material interface profile. We studied carbon nanotubes ranging from 1.1 to 1.8 nm in diameter, and deposited Au nanowires ranging from 10 to 14 nm in diameter. (e) Example resistance measurements for four devices on one nanotube. Each matrix cell is the resistance measurement at 100 mV between the two associated contacts. Color coding corresponds to each of the four devices, as well as brown for contact resistance measurements (the diagonal) and red for the SWNT channel resistance measurements.

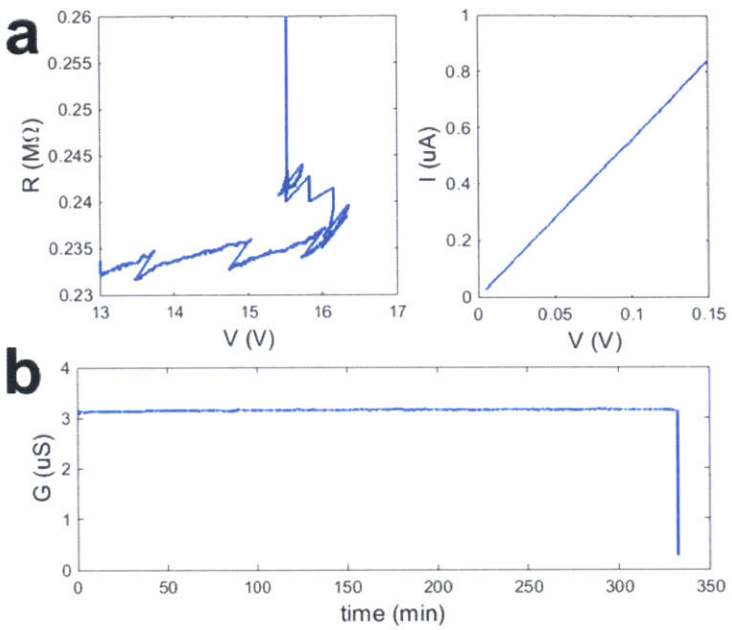


Figure 46. Electrical traces from nanowire or CNT failure. (a) Failure along the Au NW using a feedback loop. The voltage is gradually increased with time as the resistance rises through a combination of nanowire thinning and SWNT Joule Heating (left). When the resistance begins to rise rapidly, the algorithm attenuates the voltage to slow failure, controlling the gap formation until the final break. I-V trace of gap after break (right) (b) Failure along the SWNT using a 20V hold.

For each electromigration experiment, discussed below, one SWNT electrode and one nanowire electrode were contacted. For most experiments positive current was flowed into the nanotube *i.e.* electrons and Au would be migrating away from the SWNT and 'into' the nanowire, but we also treated several devices with the opposite polarity. The bias orientation determines the direction of electromigration (see Introduction), and this polarization was anticipated to maximize the probability of tunnel junction formation at the interface as the Au should be vacating the interface but has no supply for it to be replenished from.

We employed two electrical formation techniques (Figure 46): a resistance-monitoring feedback loop and a bias hold. The first, originally developed by *Strachan et al.*, involves increasing the voltage steadily and monitoring the system resistance, including the resistance of the nanotube and the nanowire.[149,155] When the resistance rises suddenly due to wire necking, the applied bias is attenuated to prevent precipitous wire failure (Figure 46a). Please see the previous Chapter for further discussion of this technique, which we call the 'Penn Method.' Our previous work modelling wire failure data[149] (Chapter I) corroborates real-time Transmission Electron Miscroscopy (TEM) results[29] suggesting that the resistance increase however is only appreciable after significant necking occurs. This process was

complicated by Joule heating of the SWNT, leading to an equilibrium temperature increase that created a reversible resistance increase convoluted with the irreversible increase of necking. The impact of Joule heating on resistance is much stronger in the nanotube than in the nanowire, and as a result unlike with electromigration in Au nanowires (Chapter II) it is difficult to distinguish from irreversible formation. The second technique we employed was simpler: we ramped the applied bias to a high value (5-20 V) but below the threshold for nanowire or CNT failure and held at that bias for a prolonged period (up to 10 days). Often within that period the nanowire or CNT would fail suddenly due to, we hypothesize, gradual electromigration in the former case and stochastic defect-driven failure in the latter.

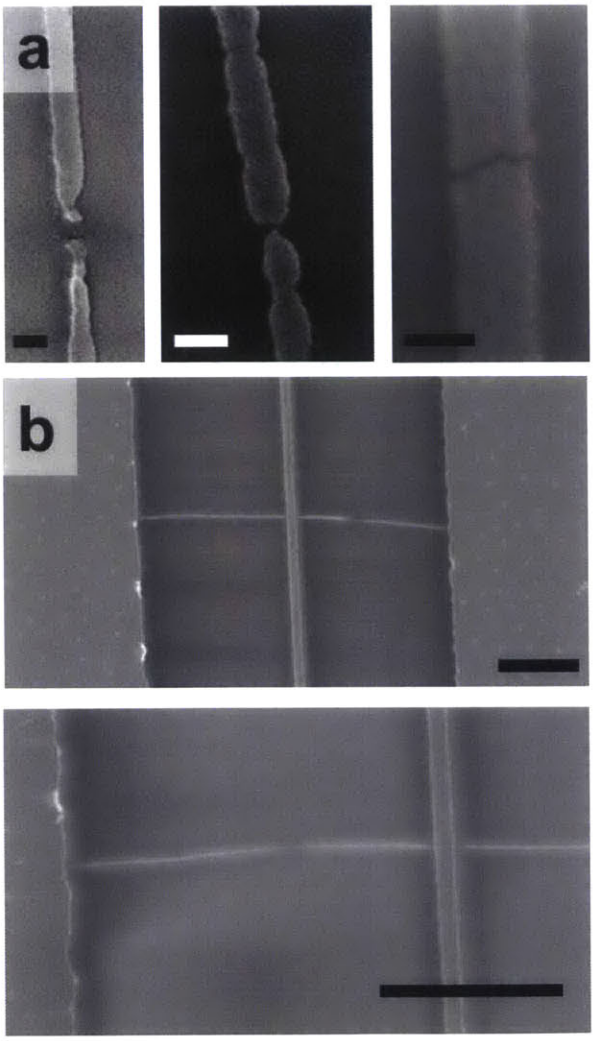


Figure 47. SEM images of (a) Au Nanowires (scale bars 100 nm) and (b) SWNT (scale bars 1 um) broken in the application of our electromigration algorithms.

In both methods we were able to drive the failure of either the CNT or the Au nanowire (Figure 47). Often in the case of Au and once in the case of SWNT we could measure electron tunneling in the

gap that formed, indicating a gap size below 2 nm, with conductivities between 10^{-4} and 10^{-1} conductance quantums. In no cases did the gaps form at the interface of the nanowire and the nanotube, and bias polarity had no noticeable effect on outcomes. That result is surprising given experimentation at the micron-scale; previous work has established that material interfaces tend to create kinetically-limiting gradients in the metal flux, even when the interface is between two metals of the same composition and crystal structure.[11,17] Our system has chemical heterogeneity at the interface, a covalently-bonded carbon material meeting a metallically-bonded Au material. The atomic spacing, crystal structures, barriers to atomic displacement, resistivity, and thermal conductivities all dramatically change at the interface. Given that the comparably minor heterogeneity of grain mismatch is enough to dominate micron-scale electromigration kinetics, the *ex ante* intuition in our system is that the interface should limit formation. By readily forming gaps at non-interface locations, our results show that nanoscale kinetics are not analogous to micron scale kinetics. At this smaller scale, defects in the materials *e.g.* grain boundaries in the Au or covalent defects in the CNT have demonstrably more rapid kinetics because their influence is no longer averaged across a larger domain. This empirical observation is consistent with the theoretical predictions of *Knowlton et al.* anticipating that reduced wire width reduces polygranular clusters, leading to more single-grain boundaries, more rapid electromigrative failure, and a greater variance in the failure time.[156] Our result shows that that effect is so strong as to make the electromigration gradient at defect sites comparable or even superior to interfacial sites.

Although this discovery makes gap formation at nanoscale interfaces more challenging, it generally makes electromigration at that scale a more adaptable tool. The implication of our result that interfaces do not strictly limit formation kinetics is that crystal structure and defects can be deliberately altered to control gap formation sites even when interfacing multiple nanostructures. Furthermore while we show that a chemical interface does not necessarily limit gap formation kinetics, that does not mean it is impossible to construct a system where it does. A larger covalent domain, such as a graphene nanoribbon, or a multi-terminal feedback control system, such as controlling the bias of 2 CNT contacts and 2 nanowire contacts simultaneously, could be used to shift formation to the interface. This tunability is a boon to the emerging fields relying on electron tunneling, including molecular electronics, nanopore sequencing, and at-scale tunneling spectroscopy. To realize that aspiration however precise control over device chemistry, geometry, and electrical circuitry is necessary, beyond present techniques. For example, the ability to grow, precisely place, and de-chelate single-crystal metal nanowires of the same aspect ratios possible with EBL would be an asset to this field. Our result is additionally beneficial to the further miniaturization of conventional electronics, where electromigrative failure of metal nanowires is problematic and carbon nanotubes and other covalently-bound conductive wires are being explored as

alternatives;[157] we have shown that it is possible to form such interfaces that are stable under the application of tens of volts.

IV. OBSERVATION OF MOLECULE DYNAMICS USING INELASTIC ELECTRON TUNNELING

Summary of Findings

As discussed in the Introduction, Single molecules trapped in on-chip tunnel junctions (SMTJs) have produced diverse high-profile physical discoveries[31-46] and hold promise for single-molecule electronics (SMEs),[43,51,52] optoelectronics,[44] and biomolecule assays.[46,55] To these ends they have been studied to show Coulomb Blockade,[32-35] the Kondo effect,[33,36-38] inelastic tunneling spectroscopy (IETS),[39,40] surface-enhanced Raman spectroscopy (SERS),[31,41] superconductivity,[42] spin-selective transport,[43] optical rectification,[44] thermoelectricity,[45] and biomolecule fingerprinting.[46] To date however all investigations have treated SMTJs, in theory and measurement, as static. That is despite extensive evidence from Scanning Tunneling Microscopy (STM) that even at liquid He temperatures an electric field can inspire molecule dynamics.[48,49]

In this work, we made conductance and differential conductance measurements with millisecond time resolution, allowing interrogation of the time evolution of on-chip single-molecule tunnel junctions (SMTJs) for the first time. Studying Cysteine and Benzene Dithiol we observed stochastic, reversible transitions between discrete inelastic transport states. State lifetimes, measured at substrate temperatures ranging from 8 to 50 K and applied potentials from 10 to 150 mV, ranged from tens of milliseconds to as high as a few seconds. Lifetimes are found to depend inversely on the applied bias, indicating an activated process with activation energies ranging from 35 to 66 meV. Comparison of our observations with theory predictions[158,159] and results from STM[48,49] and mechanical break junctions[160] among others suggest that these transitions reflect Au-thiol anchor shifts between adsorption sites and associated molecule reorientation. The bias-dependent lifetimes are consistent with and hence support theory predictions of Au-S lattice site hopping energy barriers.[161-163] In addition to our experimental findings, we simulate stochastic switching between discrete inelastic transport states such as we observe, confirming that all of the features we observe align with those expected. These reconfigurations, as posited after the first tunnel junction thermal measurements showing high local heating with applied bias,[24,31] may influence the operation and hence design of SMEs and tunnel junction assays, while our dynamic measurement approach affords new opportunities for physical exploration such as those we demonstrate.

Device Fabrication

The device fabrication was identical to that of Chapter II; please see the *Experimental Methods* section of that Chapter.

The same low temperature vacuum probe station as in Chapter II was employed. The same Source-Measure units as in Chapter II were used for Voltage forcing and Current sensing. Please see the *Experimental Methods* section of that Chapter. The noise-limited time resolution of the SMU current measurements was between 130 and 350 ms.

In addition to the SMUs, we applied a technique new to this field, delta measurement. Using a Keithley 6221 current source and 2182A voltmeter we forced current and measured voltage. While sweeping the current, a small differential was overlaid and the resulting voltage differential response was measured, providing a high bandwidth measurement of *dI/dV*, such as illustrated in Figure 48.

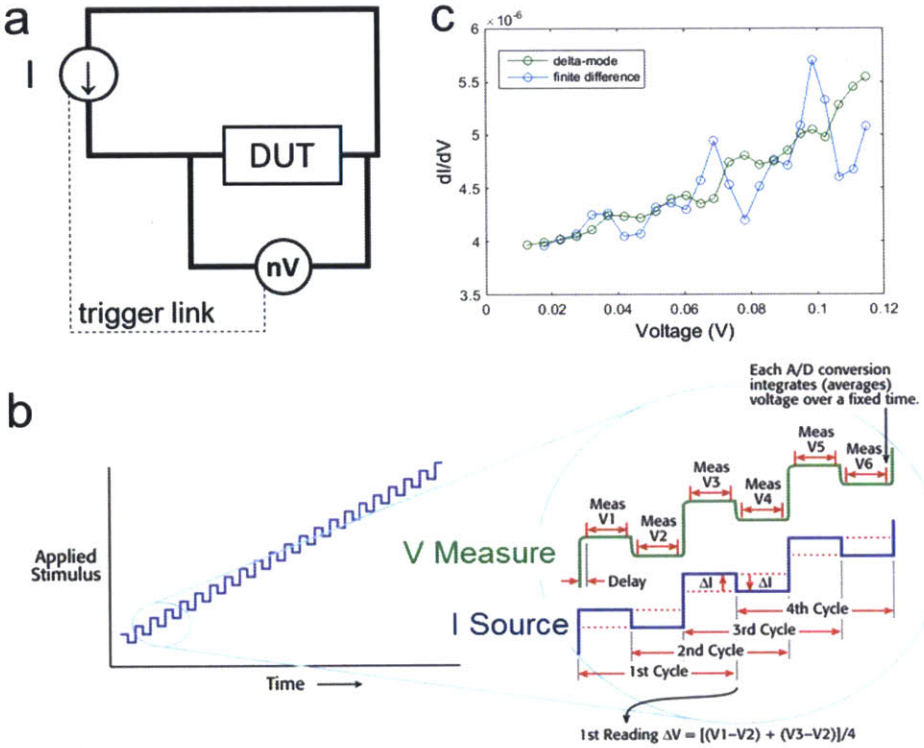


Figure 48. Simpified circuit diagram (a) and signal source-measurement illustration (b, attribution Adam Daire, Keysight Technologies) of the delta mode operation. A current signal $I(t)$ is source to the Device Under Testing (DUT) and the voltage requirement across the DUT is measured. Overall the current is swept upwards to measure $V(I)$, but at each current point I a small step up and down around that current is overlaid to measure the change in voltage in response to that change in current, providing a local measurement of *dI/dV*. (c) comparison of the delta mode measurement of differential conductance to taking the finite difference derivative of $I(V)$ of the same junction simultaneously, showing a significant improvement in noise level.

This method enabled a differential conductance measurement time resolution as low as 35 ms, a speed several orders of magnitude faster than any reported in the tunneling spectroscopy literature to date, limited by the device capacitance.

Tunnel Junction Formation

To form electromigrated tunnel junctions, we used the methods described in Chapter II. Please see that Chapter.

Molecule Deposition

We deposited the target molecule, Cysteine or Benzene Dithiol, onto the nanowires from solution as in references [31-38,40-45]. After device fabrication we soaked the nanowire chip in a 10 mM HPLC-grade anhydrous ethanol solution of the chosen molecule under a nitrogen atmosphere. The thiol and amine moieties are well known to bind strongly to the Au, forming self-assembled monolayers.[164] During the electromigration process, described in Chapters I and II, molecules are well known to migrate into the tunnel junction, as in references [31-38,40-45]. Samples were rinsed in ethanol before drying, but only lightly to allow for some crystallization of the molecule on the surface, ensuring an ample supply of molecule in addition to the chemisorbed material.

In addition to the established methodology of molecule deposition,[31-38,40-45] the presence of a molecule in the tunnel junction was verified in a few ways. First in the absence of molecule deposition we always saw ohmic resistance *i.e.* a linear current-voltage curve, whereas with a molecule deposited we nearly always saw elastic scattering (a peak in the conductivity). Second the junction dynamics we observe in this chapter are highly suggestive of a molecule, as discussed in more depth below. For example, the conductivity variations are only consistent with inelastic transport, which requires a molecule's discrete vibrational modes; the addition or removal of atoms, as shown by *Schirm et al.*,[52] yields orders of magnitude larger changes in conductivity than those observed. Third, we have been able to collect Inelastic Tunneling (IET) Spectra in some gaps, shown below, which is an unambiguous signal corresponding to the molecule.

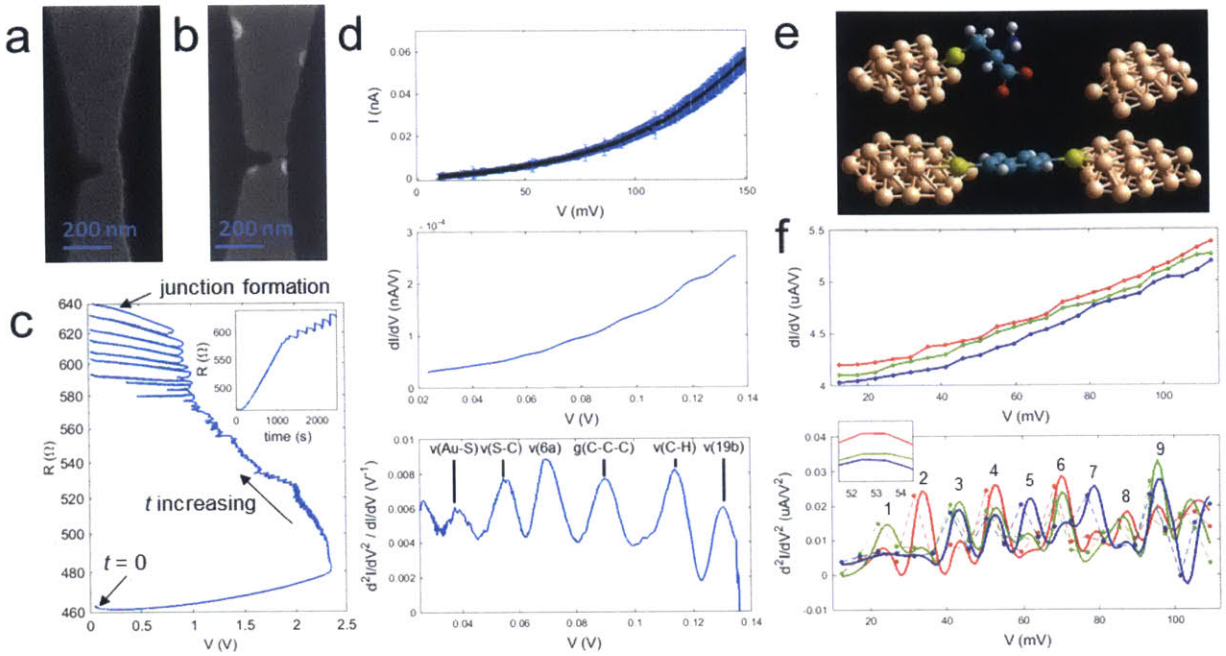


Figure 49. Scanning Electron Micrographs of nanowire devices before (a) and after (b) electromigrative failure, with the break in the nanowire visible. (c) Example trace of electromigrative tunnel junction formation, following the resistance R evolution over time t as the applied bias V is modulated to maintain a constant rate dR/dt. (d) Inelastic tunneling spectrum of one of the Benzene Dithiol (BDT) junctions. Top panel is current I measured as a function of voltage V for 300 consecutive sweeps of the voltage with 0.2 mV resolution. Mean in black and standard deviation at each voltage in blue. Middle panel is the derivative dI/dV, *a.k.a.* the differential conductance. Bottom panel is the Inelastic Tunneling Spectrum taken as the derivative of dI/dV after low-pass filtering. By convention the spectrum is divided by dI/dV to remove broad elastic contributions. The resulting peaks align as indicated with those assigned in theory and empirical measurements of BDT on Au.[40,74,158] (e) Cartoon representations of the molecules in this study, cysteine (top) and BDT (bottom), adsorbed in hypothetical configurations on the electrodes. (f) Inelastic tunneling spectrum of one of the cysteine junctions. Top panel is three differential conductance dI/dV traces as a function of voltage for three junction states (see Figure 51). Bottom panel is the finite-difference derivative, *i.e.* points in the inelastic tunneling spectrum, of each trace (dotted lines) along with a continuous derivative of the spline interpolation (solid curves). To show uniqueness of the interpolation, spline smoothing parameter varied over 10 orders of magnitude (all are plotted, see inset zoom), showing no significant variation in the resulting curve. Steps in dI/dV and peaks in d^2I/dV^2 align between the three states, varying only in amplitude. All peaks assignable to cysteine from reference [165], as follows: 1- CH_2OH torsion; 2- CH_2CHCO_2 bend; 3- SH out of plane bend; 4- CH_2CHSH bending; 5- CH_2CHN bending; 6- NH_3 torsion; 7- $CHCO_2$ stretching; 8- CS stretching; 9- CH_2 rocking.

We produced three tunnel junctions, two with BDT and one with cysteine, stable enough to test under up to 110 mV applied bias at 8.5 K for several hours. In two junctions we were able to measure inelastic tunneling spectra (Figure 49d,f) showing the expected vibrational peaks of each molecule. To interrogate molecule dynamics, we performed two kinds of electron tunneling measurement: voltage holds and delta-measurement current sweeps. In the first case, a constant voltage bias V was applied and the resulting current I, and hence conductivity G, measured over time with a sample period of 137 ms. We performed this measurement 12 times on two junctions, one BDT and one cysteine, at voltages ranging

from 20 mV to 100 mV. In both cases we observed Random Telegraph Noise in $G(t)$ (Figure 50); the conductivity stochastically, reversibly switches between different discrete values. Discrete conductivity values can be identified as peaks in the histogram of the conductivity distributions (Figure 50 right column). No drift, *i.e.* continuous monotonic evolution, of the mean conductivity over time was observed.

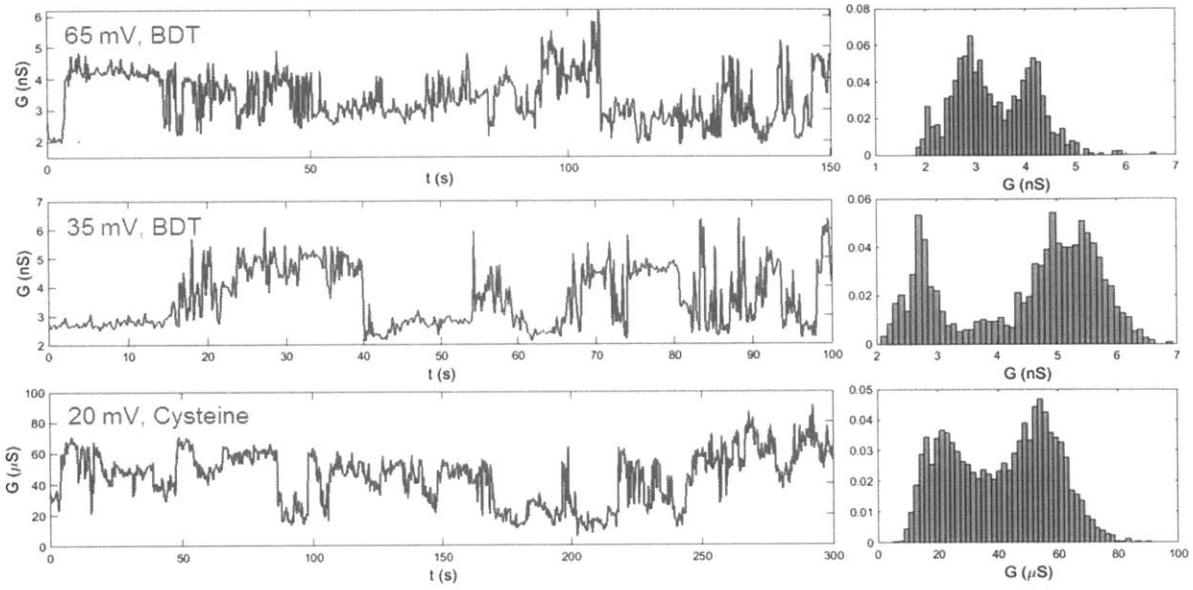


Figure 50. (left column) Conductivity G measured at fixed voltage over time in two junctions, one with BDT and one with cysteine. Excerpts of three traces reproduced here. Telegraph noise is visible. Measurement sample period of 137 ms at 8.5 K. (right column) Conductivity measurement frequency distributions (histograms) for each trace. Y axis (counts) in arbitrary units. Multiple peaks corresponding to discrete conductivity measurements are visible.

The second type of measurement we perform is delta measurements. As described in Methods, the applied bias V is adjusted to maintain a Current setpoint I around which a small current difference, or 'delta,' is superimposed to measure the resulting change in required Voltage, *i.e.* dV/dI, providing a measurement of the differential conductance dI/dV at a sampling rate of 35 ms. The Current setpoint is swept and the mean Voltage is simultaneously measured to produce $dI/dV(V)$.[166,167] Performing three consecutive delta measurement sweeps on a cysteine device (Figure 51a), we observe overlapping points (black arrows) whose variance is negligible compared with the data span, indicating measurement discretization overlaid on lower-magnitude thermal noise. Series' of consecutive overlaps (*e.g.* red and blue traces from 81 to 91 mV) indicate the ability of discrete conductivity states to persist for longer than the 35 ms sample period. That persistence also associates the discretized states with underlying $dI/dV(V)$ functions. Each trace overlaps multiple times with each of the other two, indicating reversible transitions between the discrete states. All of these observations are as expected from the telegraph signal in the Voltage-hold measurements.

To investigate these states, we performed 30 consecutive delta-measurement sweeps (Figure 51b). We again observe overlapping *dI/dV(V)* points with variance negligible compared to total span and persistence of overlaps for consecutive measurements. As apparent from the colored order of traces in Figure 51b, *dI/dV* shifts are not monotonic in time *i.e.* there is no net temporal evolution, again indicating reversible transitions.

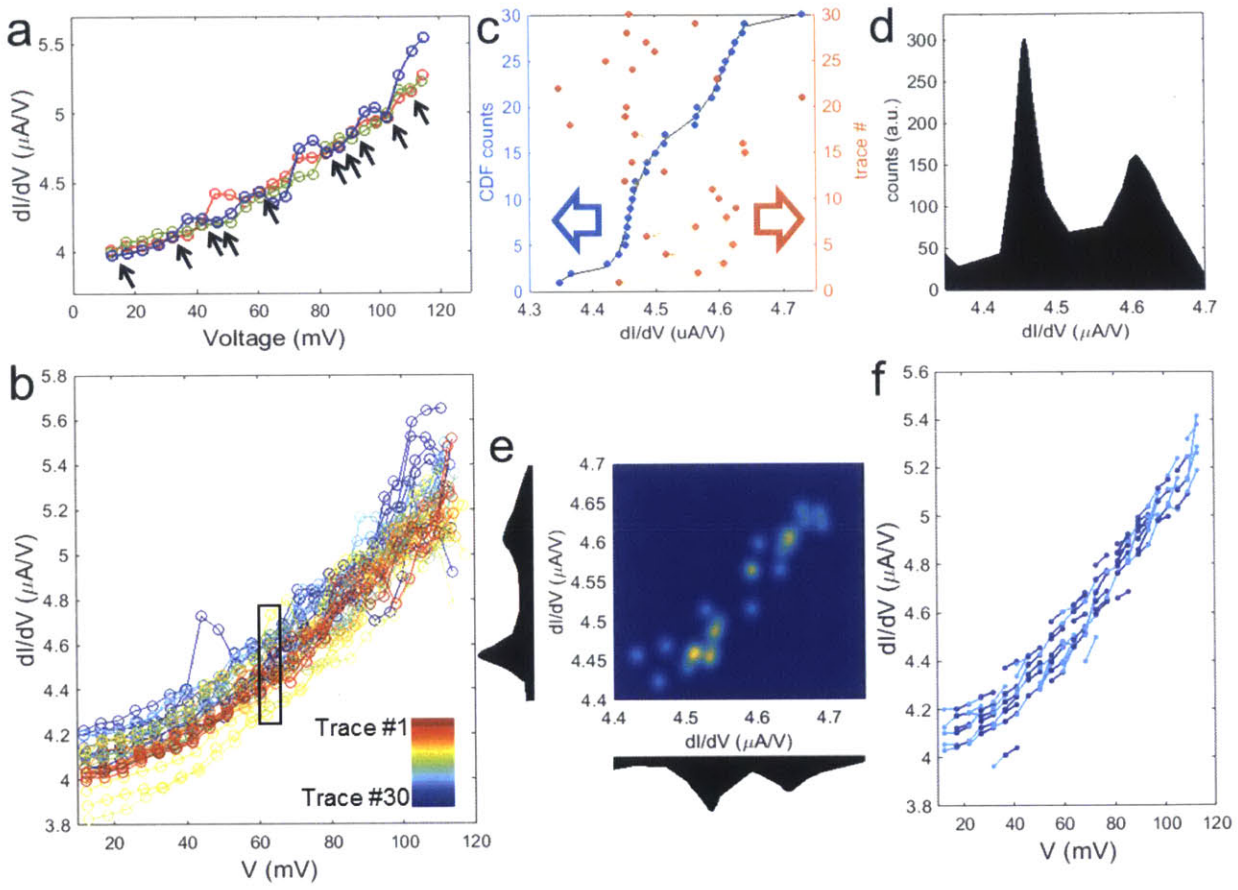


Figure 51. (a) Differential conductance *dI/dV* measured as a function of applied voltage in a cysteine tunnel junction for three representative consecutive traces. Sample period 35 ms. Arrows highlight measurements that overlap between traces, indicating discretization. Overlaps are observed between each pair of traces. Series' of consecutive overlapping measurements are observed. (b) Thirty consecutive traces as in (a), in the same device. Traces colored from red to blue in the order they were taken (inset). There is no monotonic evolution of conductivity. Black box indicates points studied in (c). (c) *dI/dV* measurements from (a) at one voltage, 64 mV (0.29 μA). (blue) measurements plotted in order of increasing *dI/dV i.e.* the cumulative distribution (CDF) of measurements, revealing clusters (discretization of *dI/dV*). For easier visualization, fitting the CDF to a spline (black line) allows continuous differentiation to yield the probability distribution function (PDF) (d), with peaks corresponding to the clusters in the CDF. (orange) measurements plotted in the order they were taken, from the first trace to the last. Arrows guide the eye along the direction of time. Again there is no monotonic evolution or irreversible shift of conductivity over time. Multi-modal *dI/dV* distributions are observed at all voltages. (e) Two-dimensional *dI/dV* measurement distributions between consecutive voltages, 64 mV and 68 mV. The measurement from each trace at 64 mV (Y axis) is plotted against the measurement from the same trace at the next voltage, 68 mV. To visualize overlap each point is plotted in the Z axis (linear colormap) as a Gaussian with standard deviation set to measurement precision. Clusters of measurement pairs produce peaks, with more common pairs producing higher

peaks. Several such clusters, *i.e.* correlated discrete measurements, are visible. Comparison (black overlay) with corresponding one-dimensional distributions from (d) shows higher resolution of measurement discretization here. (f) All peaks from the two-dimensional distributions of all consecutive voltage pairs *ij* plotted as their equivalent line segments $((V_i, V_j), (dI/dV_i, dI/dV_j))$. Colors alternated between even and odd *i* for visual clarity. Line segments frequently connect. Three nearly continuous sets of connecting segments are observed; they are isolated in Figure 49f.

To identify individual states, we extracted the distribution of differential conductance measurements at each of the 24 voltages (Figure 51c,d). To avoid the substantial binning error of a histogram analysis, we used the common technique of instead examining the cumulative distribution function (CDF), which is simply the *dI/dV* measurements sorted in order of magnitude (Figure 51c blue points).[168] All 24 CDFs show consistent behavior. Clusters of measurements in the CDFs are discrete conductivity states. For easier visualization we also converted the CDFs to population distribution functions (PDFs) by fitting them with splines (Figure 51c) and Normal distributions and taking the derivative (Figure 51d). Peaks in the distributions are discrete conductivity states at each voltage. Comparison of the CDFs with the data in the order it was measured (Figure 51c orange) show that the junction switches back and forth between states (conductivity values). The transitions are therefore reversible, as was observed in the Voltage-hold telegraph signal, rather than a single irreversible change.

To further test the observation in Figure 51a and Figure 51b that the discrete, persistent states are associated with underlying $dI/dV(V)$ functions, we plotted each measurement at each voltage against the corresponding measurement from the same trace at the next voltage 35 ms later (Figure 51e). Consecutive measurements that correlate would appear as peaks. As expected, the sequential measurements cluster into *dI/dV* pairs that correspond with one another *i.e.* are from the same state, verifying that discrete states exist and correspond to a $dI/dV(V)$ function. Compared to in the one-dimensional distributions, the discrete states are even more distinct (see side-by-side comparison in Figure 51e). Note that transitions between stable, *i.e.* common, states will also appear as peaks. Low-stability states and transitions to or from them would appear as outliers and weak peaks. Plotting all of the consecutive-measurement correlated peaks together as line segments in $dI/dV(V)$ (Figure 51f), we would expect that many peaks will correspond with adjacent peaks *i.e.* that line segment ends will connect. That is indeed the case (Figure 51f), again showing that discrete states persist for multiple sample periods and correspond with $dI/dV(V)$ functions. These end-to-end connections thereby reconstruct the $dI/dV(V)$ functions associated with each state. While five or six states might seem apparent in Figure 51f, only three appear frequently enough to fully reconstruct their $dI/dV(V)$ traces as in Figure 49f. Isolating individual states leaves only thermal and measurement noise, which in our apparatus at 9 K are small as observed in Figure 51a. As a result, clear steps in $dI/dV(V)$ emerge, characteristic of inelastic transport due to coupling of tunneling electrons to molecule vibrational states. Between the three states, the magnitude of each step changes substantially but their positions do not. As we discuss in more detail below, this behavior, in addition to

the other features we have observed, is characteristic of changes in inelastic transport between states. Taking the derivative (Figure 49f) to provide points in the inelastic tunneling spectrum reiterates this observation; peaks align in energy with variations in magnitude.

The stochastic, reversible switching between discrete conductivity states observed in all three data sets, with states mapping to $dI/dV(V)$ functions, indicate changes in the molecule's inelastic conduction channels. Because of the exponential dependence of tunneling current on gap width and barrier height, conductivity changes upon introduction or removal of an atom or molecule are known to be an order of magnitude greater than those we observe.[35,40,52,158,160,169,170] Rather, the magnitude of the conductivity difference between discrete states is only consistent with inelastic tunneling, as per references [40,52,158,160,169,170]. Furthermore as bias increases, the states monotonically increase in dI/dV (Figure 49f, Figure 51), with states differing in the amount of increase. That behavior is characteristic of inelastic transport,[40,52,158,160,169,170] where coupling to vibrational modes manifests as step increases[171] in dI/dV when the bias applied reaches the associated oscillator energies. In contrast elastic transmission appears as a peak in dI/dV, not a step.[169] Finally, the peaks we observe in d^2I/dV^2 (Figure 49d,f) have linewidths of order 1 meV and energy spacings of order 10 meV, characteristic of inelastic peaks, while elastic peaks, which correspond to molecular orbitals, have energy spacings of order 0.1-1 eV and typically peak widths of order 100 meV.[169] We illustrate the contributions of elastic and inelastic transport in simulations below.

Comparison of the observed features of these discrete inelastic states with previous literature suggests that they reflect changes in molecule conformation *i.e.* the orientation and adsorption of the molecule. In first-principles calculations Lin *et al.*,[158] Nakamura,[159] and several others have predicted that inelastic transport is extraordinarily sensitive to the orientation of the molecule and its bonding to the electrode, even when elastic (molecular-orbital-mediated) transport is unaffected. Steigerwald *et al.* then confirmed that principle[160] in mechanical break junctions by showing that bi-aromatic tunneling conductivity changes with twist angle. Kiguchi *et al.* similarly showed[172,173] that aromatic conduction changes discretely with the molecule-electrode anchor site. Molecule reorientation would therefore explain the discrete changes in inelastic tunneling probability that we observe. In the Scanning Tunneling Microscopy (STM) literature, beginning with Stipe *et al.*[48] it has been commonly observed that at low temperatures functional groups can flip between adsorption sites with a frequency activated by the applied bias,[48,49] or that the thiol anchor itself can hop between Au adsorption sites.[50]

The reversibility of the state transitions further supports a reconfiguration hypothesis. As observed in STM by Stipe *et al.*,[48] Baber *et al.*,[49] and others, when given sufficient energy to overcome transition-state energy barriers a net-static molecule hops between local-energy-minima

configurations. The energy can be supplied either thermally[49] or with an applied bias.[48] In analysis of the data below, we observe such voltage activation (Figure 52, Figure 53), where state transitions become more likely at higher applied bias. The energy supplied by the applied bias may be converted to translational energy either directly by the hot electrons[48,49] or thermally; as observed by Ward *et al.*[31] and Ioffe *et al.*,[24] a 200 mV applied bias can heat the junction molecule by more than 200 K. Furthermore under this explanation states of different relative stability should yield, within shot noise, different relative occupations exactly as we observe in the distributions of Figure 50 and Figure 51d. At least 4 states are unambiguously resolved in the 1D distributions (Figure 50, Figure 51d), and at least 5 are clear in the 2D distributions (Figure 51e,f). To resolve lower-stability states both faster and more measurements would be required due to the shorter lifetimes.

Note that if the integration time were significantly longer, the central limit theorem would cause the signal to wash out to the mean (a Normal distribution) and discrete states would be masked. The molecule would thereby appear static. The only indication of multi-state occupation would be, at intermediate integration times (between 1 and 2 orders of magnitude slower than the lifetime), a noise level greater than thermal noise. Previous presumably static observations of single-molecule tunnel junctions, such as in references [31-38,40-45], may therefore have had underlying dynamics, or 'high-noise' samples may have been prematurely excluded from consideration.

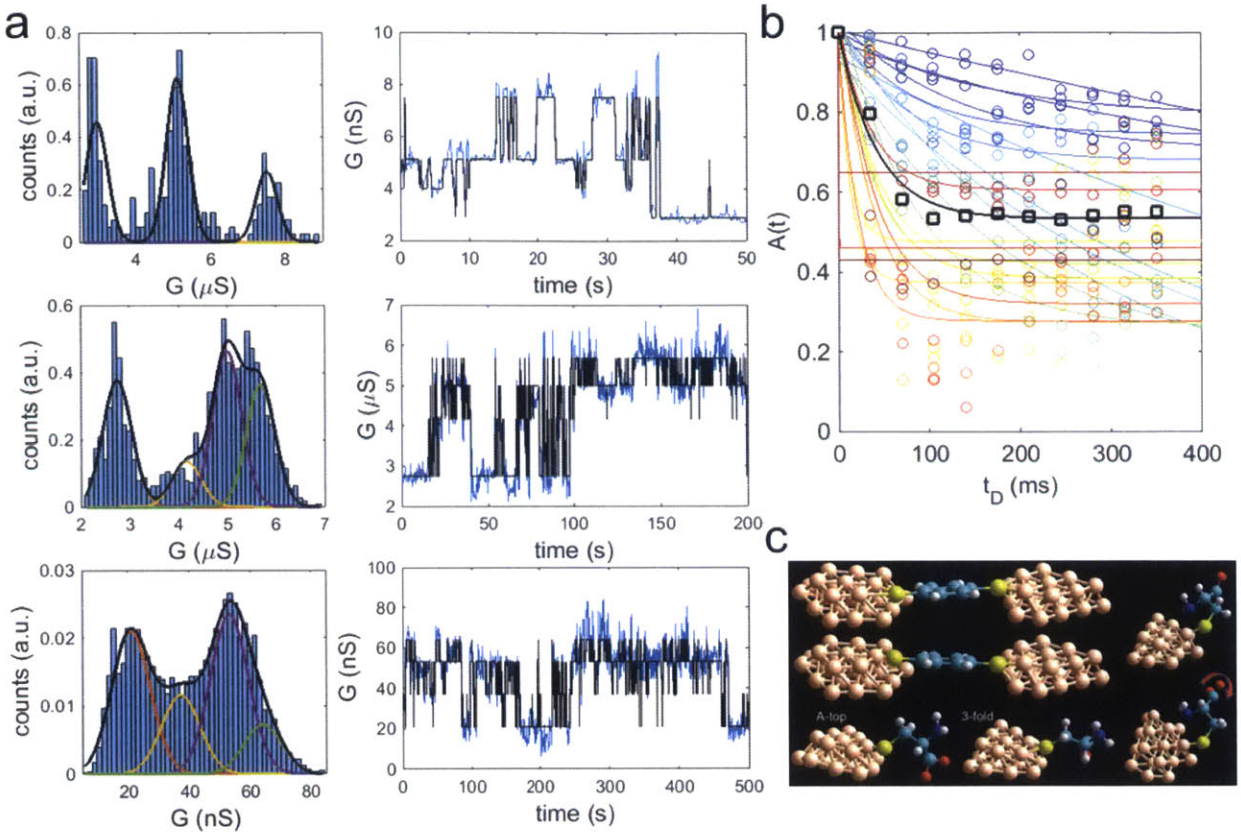


Figure 52. (a) Markov chain analysis of telegraph noise in constant-voltage conductivity measurements. (left) Deconvolution of conductance measurement distributions into constituent discrete states for three traces. (right) Using that deconvolution, each measurement in the corresponding constant-voltage conductivity trace is assigned to its most probable state (black line is the mean conductivity of the associated state, blue line is the raw data *cf.* Figure 50), yielding the underlying Markov chain. Three traces analyzed here: from top to bottom, 25 mV and 35 mV in a BDT device and 20 mV in a cysteine device. (b) Time-delay autocorrelation $A(t)$ of the dI/dV measurements from Figure 51. The time-delay autocorrelation is the correlation between measurements made in the same $dI/dV(V)$ trace t_D time apart. $A(t)$ calculated from different starting voltages, colored from blue, 12 mV, to red 110 mV. Circles are data, lines are fit exponential decay functions. An increase in the rate of decay is observed with increasing bias. (c) Cartoons illustrating molecule reconfigurations consistent with the observed signals, including changes in both molecule orientation and thiol anchor adsorption site, such as between 'atomic-top' (A-top) and 'three-fold hollow' (3-fold) sites.

To explore the molecule reconfiguration hypothesis further, we measured the state lifetimes. To do so from the Voltage-hold data, for each trace we identified the mean conductivity of each abundant state by deconvoluting the distributions as in Figure 52a. We then assigned each measurement at each time to the most probable state (black lines in Figure 52a). Reversible switching between molecule configurations can be described as a Markov Chain, with a probability at each time step (measurement) of either remaining in the current state or overcoming an energy barrier to move to one of the other local-energy-minima in an Arrhenius transition. The process is described by a transition matrix T where T_{ij} is the probability in a single time step of transitioning from state i to state j. We optimized transition

matrices to maximize the log-likelihood of the state sequence associated with each trace. The mean lifetime of each state τ_i is a function of the transition matrix diagonals,

(45) $$\tau_i = \frac{1}{1 - T_{ii}} \Delta t$$

where $1/\Delta t$ is the sampling rate.

To measure the state lifetimes from the delta-measurement sweeps, we calculated the two-dimensional time-delay autocorrelation A_{ij} between each voltage,

(46) $$A_{ij} \equiv \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{\left(\langle Y_i^2 \rangle - \langle Y_i \rangle^2\right)\left(\langle Y_j^2 \rangle - \langle Y_j \rangle^2\right)}}$$

(47) $$\langle Y_i \rangle = \frac{1}{R} \sum_{r=1}^{R} Y_i^r, \qquad Y_i^r = \frac{dI}{dV}(I_i, r)$$

where R is the total number of runs (traces), I is current and V is voltage. A_{ij} is the Pearson correlation between dI/dV measurements at voltage i (made at time t_i) and voltage j (made at time t_j), *i.e.* it is the probability that the measurement at t_i predicts the measurement in the same trace some time $t_j - t_i$ later. Longer state persistence increases the correlation between measurements made farther apart. Perfectly independent measurements yield $\langle Y_i Y_j \rangle = \langle Y_i \rangle \langle Y_j \rangle$ and $A_{ij} \to 0$, a perfect correlation (*i.e.* state lifetimes $\to \infty$) yields $A_{ij} \to 1$, and a perfect anti-correlation yields $A_{ij} \to -1$.

In Figure 52b we plot A_{ij} as a function of time delay $t_D = t_j - t_i$ for each initial voltage (colored from blue, 12 mV, to red, 110 mV). Confirming the high consecutive-measurement state persistence observed in Figure 51e, consecutive-measurement correlations are often over 0.8. We further observe long range correlation, gradually decaying with time after neighboring measurements. That confirms the qualitative observation of single-state persistence for multiple measurements in Figure 51a. The decay of long-range correlations is slow at low bias, but as bias increases the correlations decay more rapidly, indicating that as expected the inter-state transitions are activated. Exponential decay functions (solid lines) are fit to each curve, and the resulting mean state lifetimes τ_D are the reciprocals of the decay constants.

The state lifetimes as a function of applied bias are plotted for our three data sets in Figure 53a. Consistent with our hypothesis of molecule reconfiguration and the associated literature, state lifetimes drop precipitously after a threshold voltage, suggesting an associated reconfiguration transition energy such as shown by Sainoo *et al.*[73] The resulting activation energies for the three systems are between 35 and 50 meV, 40 and 50 meV, and 43 and 66 meV. The hot electron energy available under even our modest applied biases far exceeds the available thermal energy kT even up to 100 K; an electron with 35 meV excess energy in the drain electrode has the thermal energy equivalent of being at 400 K. As shown

by Ioffe *et al.*[24] and Ward *et al.*[31] that power input to the junction even manifests as a local temperature rise perhaps *via* excitation of molecule vibrational modes and electrode surface phonon modes; at 200 mV junction temperatures measured by stokes-anti-stokes Raman peak shifts exceeded 250 K. As a result ambient heating from 8 to 50 K is not expected to effect the junction and indeed we see no effect.

The activation energies measured are consistent with reversible hopping between thiol anchor adsorption sites. Discrete changes in inelastic tunneling can be due to any combination of molecule reorientation and re-adsorption.[48,49,158,159,172,173] BDT however has only one axis of rotation and therefore only rare opportunities for discretized reorientation without changes in adsorption (**Error! Reference source not found.**f). Lin *et al.*[158] and Nakamura *et al.*[159] independently developed *ab initio* models of BDT tunnel junction transport, both showing that changes in thiol adsorption site (*e.g.* between A-top and 3-fold-hollow) yield dramatic changes in inelastic coupling. No experimental measurement of the energy barrier to changes in thiol adsorption site have been measured to date, but numerous theoretical predictions exist on the Au(111) surface for various thiol compounds.[161-163] Predicted barriers typically range from 43 to 97 meV, consistent with the activation energies we have measured. While the crystal facets of our electrodes are unknown, these values provide a reasonable estimate of magnitude.

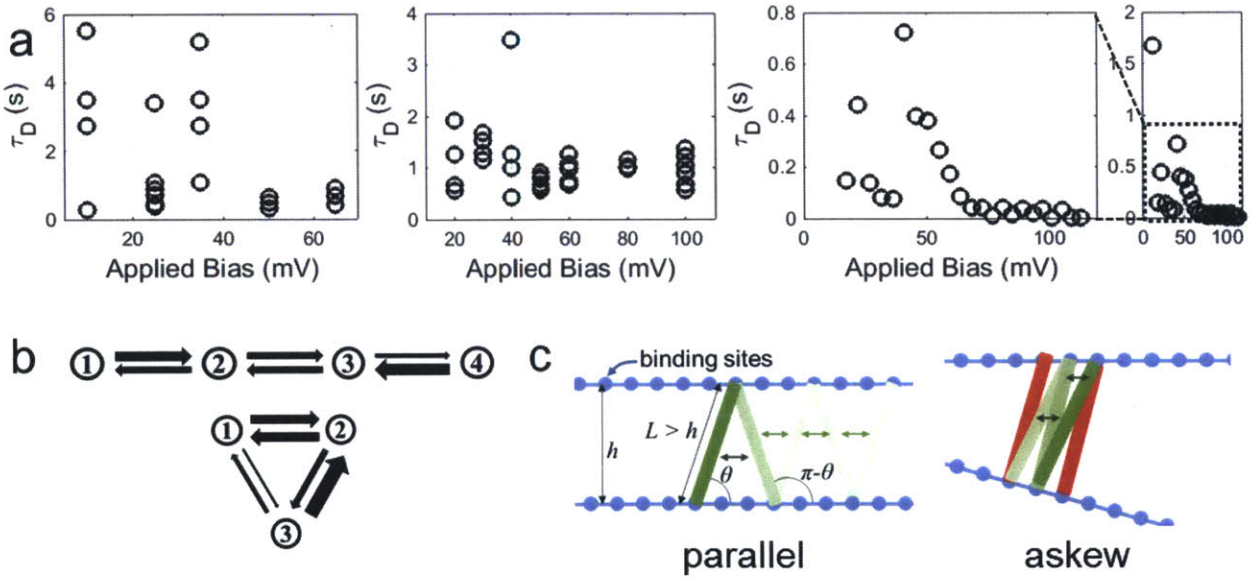


Figure 53. (a) Lifetimes τ_D of conductivity states as a function of applied bias for three tunnel junction devices, from left to right: BDT constant-voltage traces, cysteine constant-voltage traces, and cysteine delta measurements (mean state lifetimes; zoom out to the far right). In each case a sudden drop in lifetimes is observed past a voltage threshold. (b) Graphs of relative transition rates between states in one BDT (top) and one Cysteine (bottom) junction, with arrow widths proportional to the kinetic rate constant. Inter-state transition rates measured as the off-diagonal terms from the Markov process transition matrices T for the constant-voltage data sets, averaged over all voltages. (c) Cartoons illustrating geometric constraints on a molecule in the junction in two dimensions. Blue lines (locally) represent the

electrodes, with molecule adsorption sites indicated by circles. The molecule is represented as a rigid rod of length L which must conform to the fixed electrode geometry, constraining the number of possible orientations. In the askew electrode case several possible positions and orientations are presented, with the number constrained by the limits of compression or expansion of the molecule (red configurations).

Although thermal noise limits the accuracy of our Markov chains, it was possible to discern some information on relative interstate transition kinetics in the voltage-hold data in addition to the lifetime measurements. In each of the two data sets, the Markov Transition Matrix elements for the most stable states are averaged across voltages. The resulting relative interstate transition probabilities (*i.e.* the non-zero off-diagonal elements) construct corresponding interstate transition graphs (Figure 53b). While many different junction structures are possible however, these graphs constrain the space of possibilities. Linear reversible graphs emerge for both junctions; although the Cysteine junction produces a triangular graph, one set of transitions are rare and can be neglected. This linearity in the graph is consistent with a single molecule dominating the tunneling current, rather than two or three, as each additional molecule would produce parallel branches. A single molecule of length L between two parallel electrodes spaced h apart such that $h < L$ has exactly two possible configurations to bridge the gap, forming an angle at one arbitrary end θ or $\pi - \theta$ such that

$$L\cos\theta = h \tag{48}$$

This constraint produces a simple two state reversible graph if the molecule is able to break one anchoring bond and reform it in the opposing configuration. If the center of mass of the molecule translates parallel along the electrodes, the same two states are occupied and have tunneling currents that are indistinguishable. However, if the electrodes are locally not parallel, a likely scenario after junction formation, additional non-degenerate states are available as shown in Figure 53c. A molecule of length L can span the electrodes by adjusting its orientation and center of mass position through constrained values, producing a truncated, linear graph. The range of positions available is confined by the limited ability of the molecule to expand or compress (*e.g.* red states in Figure 53c). We expect the corresponding transition rates at the ends of the linear graph to be less favorable such that extension to an even farther state is sterically forbidden. This picture is consistent with the observed graphs for both BDT and cysteine. In the full three dimensional picture, the same linear graph describes a path that a single molecule traverses.

In this work we have made the first time-resolved tunneling current measurements of a molecule in an electromigrated tunnel junction. Studying Cysteine and Benzene Dithiol, we observed stochastic, reversible switching between discrete inelastic transport states. The switching is activated by the applied bias, with interstate transition energy barriers ranging from 35 to 66 meV. The observed signal and activation energy are consistent with theory predictions and observations from other platforms of the molecule switching between thiol adsorption sites and reorienting. Markov chain modelling of

conductivity-over-time observations produces measurements of state lifetimes and interstate transition rates. The non-zero elements of the latter produce a graph describing the Markov chain, constraining the possible molecule configurations. These findings show that, consistent with observations in STM, molecule dynamics can be observed in tunnel junctions even at low temperatures.

Simulation

To better understand discretized inelastic transport states, we simulated such a system. The tunneling current is proportional to the electron transmission probability $T(V)$, which excluding interference[95] is approximately the addition of each available bias-dependent transmission channel:[169,174] elastic $T_E(V)$, direct $T_D(V)$, and inelastic $T_\omega(V)$

(49) $$T(V) = T_E(V) + T_D(V) + T_\omega(V).$$

Direct tunneling between electrodes agnostic of the molecule is a linear baseline

(50) $$I(V) = \frac{dI}{dV}\Big|_0 V$$

For elastic and inelastic tunneling, we use the Landauer-Buttker non-equilibrium transport formalism,[174]

(51) $$I(V) = \frac{2e}{\hbar}\int_{-\infty}^{\infty} T(E)\cdot\big(f(E-\mu_L-eV) - f(E-\mu_R)\big)dE$$

where E is electron energy, $T(E)$ is the transmission probability for the pathway of interest, $f(E)$ is the density of states, which we take as a fermi-dirac distribution,

(52) $$f(E) \equiv \frac{1}{1+\exp\left[\frac{E}{kT}\right]},$$

and $\mu_{L/R}$ is the chemical potential of the left and right electrodes, respectively. Because only relative voltage dependence is relevant, for simulation purposes we take $\mu_L = \mu_R = 0$. We arbitrarily choose the right electrode as the high-potential electrode.

Elastic transport is mediated by any molecular orbitals (chemical potential μ_M) with significant density of states within the nonequilibrium potential drop eV ($\mu_L \lesssim \mu_M \lesssim \mu_R + eV$). The transmission function is the density of states of the molecular orbital $D_M(E)$, which is broadened due to coupling with the electrodes, and the coupling strength,

(53) $$T(E|\mu_M,\gamma_1,\gamma_2) = \frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2}\cdot D_M(E|\mu_M,\gamma_M)$$

where γ_i is the coupling energy between the MO and the electronic states of electrode i, and γ_M is the sum $\gamma_1 + \gamma_2$. D_M is suitably approximated by a Lorentzian

(54) $$D_M(E|\mu_M,\gamma_M) = \frac{\gamma_M/2\pi}{(E-\mu_M)^2 + \left(\gamma_M/2\right)^2}.$$

Inelastic transport mediated by a molecule vibration mode with frequency ω shifts the relevant right electrode density of states to $f(E-\mu_R-\hbar\omega)$ to provide energy to excite the mode. Each mode ω has a corresponding coupling energy A_ω and a linewidth full-width-half-max Γ. The tunneling probability is a complicated function with several proposed theoretical expressions,[170,172,175] but empirically its derivative is closely approximated by any bell distribution. We use a Gaussian distribution,

(55) $$\frac{d}{dE}T_\omega(E|A_\omega,\Gamma,\omega) = A_\omega \cdot \frac{1}{\Gamma\sqrt{2\pi}}\exp\left[-\frac{(E-\hbar\omega)^2}{2\Gamma^2}\right].$$

Integrating, we have the transmission function's functional form

(56) $$T_\omega(E|A_\omega,\Gamma,\omega) = \frac{A_\omega}{2}\left[1+\operatorname{erf}\left(\frac{E-\hbar\omega}{\sqrt{2\Gamma^2}}\right)\right].$$

To calculate differential conductance we take the derivative of the relevant *I(V)* terms,

(57) $$\frac{dI}{dV} = \frac{dI}{dV}\Big|_0 + \frac{2e}{\hbar}\int_{-\infty}^{\infty} T(E|\mu_M,\gamma_1,\gamma_2)\frac{d}{dV}[f(eV)]\,dE + \frac{2e}{\hbar}\sum_{\omega'\in\{\omega\}}\left[\int_{-\infty}^{\infty} T(E|A_{\omega'},\Gamma,\omega')\frac{d}{dV}[f(eV)]\,dE\right]$$

where $\{\omega\}$ is the set of all molecule vibrational modes.

Three molecule configurations with different $\{A_\omega\}$ sets and a single fixed elastic mode are chosen (Figure 54). The magnitude range of $\{A_\omega\}$ is chosen from [169]. To produce *dI/dV* slopes of the same order as we observe, an elastic transmission mode dominating total tunneling probability is necessary, demonstrating the phenomenon that elastic tunneling changes are an order of magnitude greater than the observed variations and substantially broader than the observed steps.

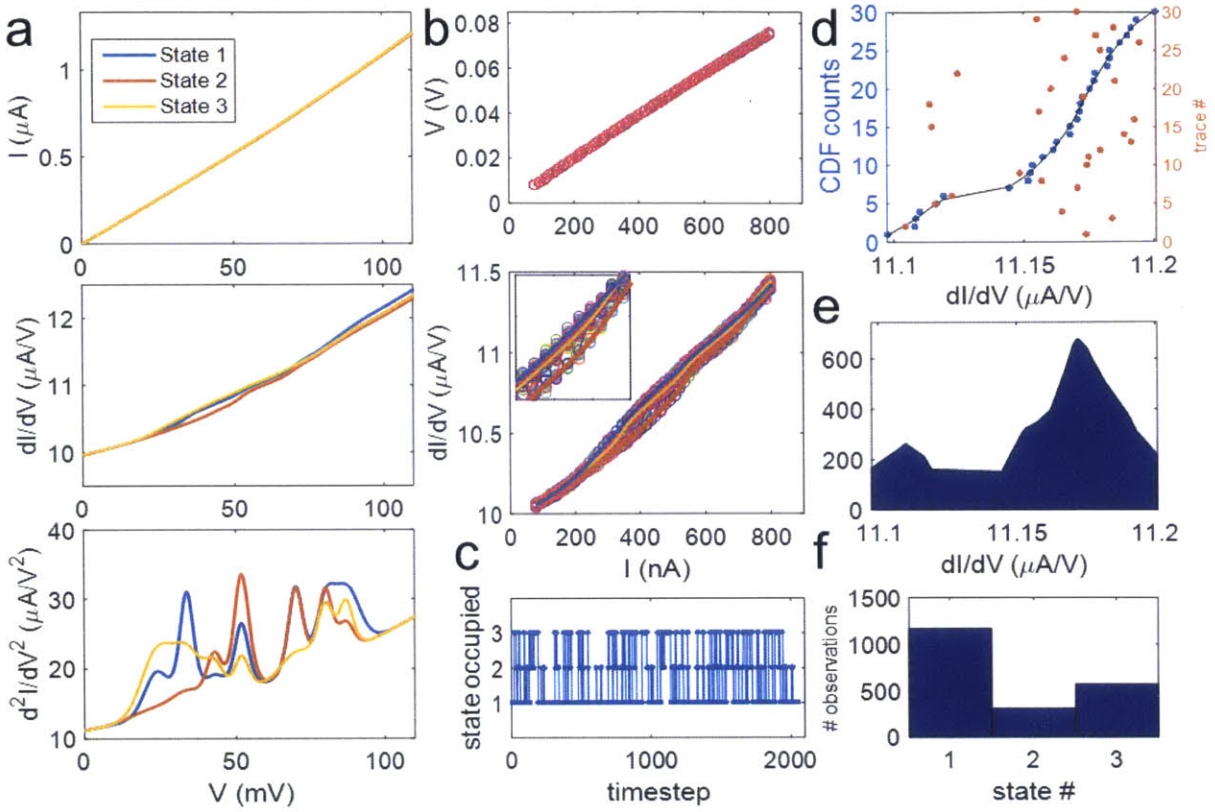


Figure 54. Simulation of stochastic switching between inelastic transport states. (a) Three different inelastic transport states of the same molecule with different couplings to each vibrational mode (Cysteine vibrational modes) and a shared molecular-orbital-mediated tunneling probability. Single-state *I-V* curves (top), *dI/dV* curves (middle), and inelastic tunneling spectra (bottom). Note that to illustrate the elastic effect we have not divided the spectrum by *dI/dV* as is convention (*cf.* Figure 53c). (b-f) Monte carlo simulation of switching between the inelastic transport states of (a). (c) Simulated state occupations over time. (b) resulting *I(V)* and *dI/dV* measurements (open circles) including thermal noise and measurement precision. 'True' single-state traces from (a) overlaid (solid lines). Inset zoom in from 600 to 750 nA. (d) Cumulative distribution function at point 42, 682 nA, performing the same analysis as Figure 50a. (e) corresponding probability distribution function approximated from the spline fit derivative. (f) total number of occupations observed of each state, a random outcome of the different relative stabilities of each state.

To simulate our observation of transitions between these discrete inelastic transport states, we created a Markov process transition matrix *T* similar to those measured. We then ran Kinetic Monte Carlo simulations subject to that probability matrix, such as in Figure 54c. The simulations match all of the features observed in the experimental results (Figure 54b,d,e). The differential conductance shows the 'overlap' behavior observed, and produces distributions qualitatively indistinguishable from those we have measured (Figure 54e). No contradictory phenomena are observed. This provides a valuable sanity check to the consistency between our physical interpretation and the observed results.

V. DETERMINISTIC MODELLING OF CARBON NANOTUBE NEAR-INFRARED SOLAR CELLS

Summary of Findings

As discussed in the Introduction, carbon nanotube (CNT) photovoltaics are a promising material for cost-effectively harvesting energy from the near-infrared region of the solar spectrum, complementing visible solar cells. That appeal spurred abundant development of proof-of-concept devices,[115-121,129] but progress quickly stagnated, with efficiencies trapped below 3%. The research pathway to achieving commercially viable efficiencies has been empirically impenetrable because of the large parametric space associated with nanometer-scale material heterogeneity, specifically CNT networks, and highly non-linear relationships between those parameters and efficiency.

In the work related in this Chapter, we addressed this SWNT PV design knowledge gap by developing a deterministic model of carbon nanotube photovoltaic steady state operation. The model is a set of integro-differential equations derived from single- and paired-SWNT optical absorption, exciton transport, and free carrier transport. It was derived by adopting an approximation of macroscopic, $O(100\ nm)$, homogeneity. That allowed us to treat relevant local, *i.e.* single-nanotube, properties as random variables, describing the network of nanotubes as a distribution of those properties. We focused on planar heterojunction devices. Our treatment is general for arbitrary networks in terms of chemical and structural properties – *e.g.* any distribution of nanotube chiralities, lengths, orientations, defect types and levels, bundle fraction and size, density, dielectric environment, or electrode combination. It is rigorously limited however to densities above the percolation threshold in the charge-collecting axis, although it is quantitatively and qualitatively consistent with recent observations at lower densities in the literature suggesting possible broader relevance.

We then applied the model to elucidate the optimal SWNT PV design as a function of nanotube orientation, packing density, defect density, and nanotube length. An optimal film thickness, T^*, emerges that decreases with density, impurity concentration, and in-plane nanotube orientation to balance light absorption with exciton diffusion length and exciton-electrode separation. We find that a given set of parameters are characterized by a minimum optimal thickness, T^*_{CP}, at close-packed density ρ_{CP}. Devices below close-packed density have a normalized optimal thicknesses $T^*/(T^* - T^*_{CP})$ that scales with density.

Nanotube alignment angle (*e.g.* vertical, horizontal) balances much faster intra- *versus* inter-nanotube exciton diffusion with higher axial versus radial absorptivity. At sufficiently high densities, above 3% of close-packed, vertically aligned films are shown to be unambiguously favored to maximize exciton transport, but as density is lowered to below 10% of close-packed a second optimum emerges for

some thicknesses at an intermediate angle, $\pi/3$ to $\pi/4$ from the vertical axis, to compensate for the decreasing light absorption. Of all film properties, density and alignment are most important in determining efficiency, with 10- to 100-fold performance improvements over current devices predicted through vertical alignment and increasing density to 10% of close-packed or higher.

Model Derivation

A summary derivation of the model is presented here. See Appendix A for a detailed derivation, as well notation conventions *e.g.* for vector quantities and random variable distributions.

We consider a network of single-walled nanotubes (SWNT) sandwiched between two electrode plates (Figure 1). We define a Cartesian z axis in the propagation direction of the incident solar photon flux $J_0(\omega)$ at $z = 0$. The film is semi-infinite in x and y and has thickness T. The back electrode has reflectance $0 \leq R \leq 1$ which can in general be frequency dependent. A variety of conduction and valence band charge collecting electrode (type II exciton dissociation interface) configurations are possible, including those where an electrode also acts as a photoabsorbing layer (*e.g.* fullerenes[115,116,121]). We focus on the relatively general case of two dissociating electrodes, but the model for the SWNT network applies to any other case by corresponding adjustment of the boundary conditions.

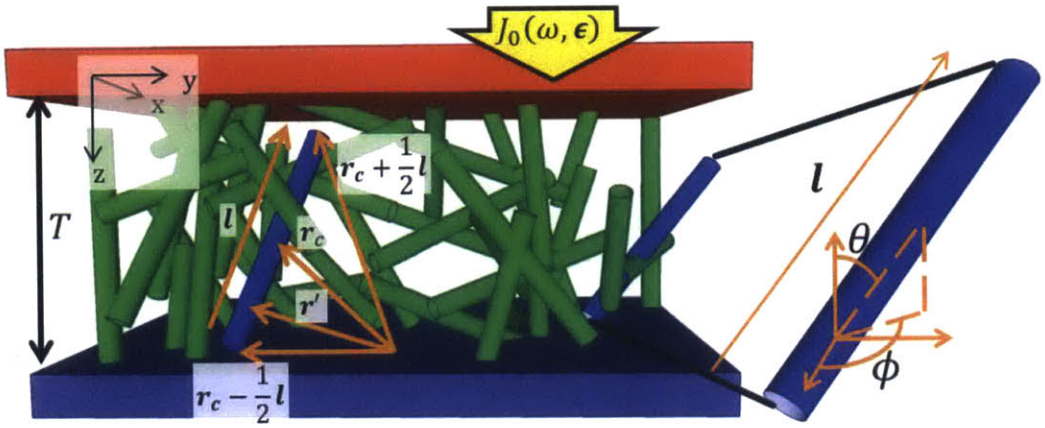


Figure 55. Cartoon of model geometry, depicting the position $\boldsymbol{r}_c$, orientation $\hat{\boldsymbol{l}} \equiv (\theta, \phi)$, and length l of nanotubes. The thickness of the film is T, and we take the z axis to be the propagation direction of incident irradiance $J_0(\omega, \epsilon)$. A point on a given tube is $\boldsymbol{r}'$.

We define an individual nanotube in our network, which we approximate as a rigid rod, by a set of relevant properties: position, length, orientation, and chirality. In the global coordinate system the center of each nanotube is located at $\boldsymbol{r}_c$ relative to an arbitrary origin in the xy plane at $z = 0$. The length l and orientation $\hat{\boldsymbol{l}} \equiv (\theta, \phi)$ of the tube form a spherical-coordinate vector $\boldsymbol{l} = (l, \theta, \phi)$. The ends of a single SWNT are therefore located, relative to the origin, at the ends of vectors $\boldsymbol{r}_c \pm \frac{1}{2}\boldsymbol{l}$. A point $\boldsymbol{r}'$ along

the central axis of the SWNT is then confined to the line segment between the ends. In describing intra-SWNT mechanics we will also consider the one-dimensional space along a nanotube axis, denoting that coordinate $r_l = \alpha_l l,\ \alpha_l \in [0,1]$, which bijectively maps to $\boldsymbol{r}'$. A nanotube also has a chirality $\boldsymbol{c} \equiv (n_1, n_2)$ where n_1 and n_2 are chiral indices.[176] For convenience we index all chiralities present in the network by integer i (chirality $\boldsymbol{c}_i$) in order of increasing exciton (optical) band gap $E_{i+1} > E_i$. Each chirality has an associated center-center diameter $d_{\boldsymbol{c},c-c}$, and an outer diameter $d_{\boldsymbol{c}}$.

The properties $\{\boldsymbol{r}_C, \boldsymbol{l}, \boldsymbol{c}\}$ completely define a given nanotube in our network. Our approach begins with treating these parameters as random variables, defining a given film by their distributions:

$$
\begin{aligned}
& p(\boldsymbol{r}_C|\boldsymbol{r}) = p(r_{cx}, r_{cy}, r_{cz}|\boldsymbol{r}) \\
& p(\boldsymbol{l}) = p(l) \cdot p(\theta, \phi) \\
& p(\boldsymbol{c}_i) = v_i \in [0,1], \qquad \sum_i v_i = 1
\end{aligned}
\tag{58}
$$

where v_i is the number fraction of chirality i in the film. In theory any of the distributions can be dependent on the others depending on the nanotube and film fabrication processes; we focus on cases where they are uncorrelated, but the derivation is general for correlated distributions. Our derivation is also general for cases where all of these properties vary systematically with depth, *i.e.* $p(\boldsymbol{r}_C, \boldsymbol{l}, \boldsymbol{c}|z)$, but in most practical applications $p(\boldsymbol{l})$ and $p(\boldsymbol{c}_i)$ are spatially invariant. It is also possible to introduce additional variables, such as the local dielectric constant, by including their distributions. That allows the inclusion of those variables in any exciton generation, diffusion, or quenching processes they are involved in. To introduce additional variables the model equations must be re-derived as below, but the approach and steps remain the same.

Solving for the steady state current in the illuminated film breaks into three distinct problems: light interaction, exciton transport, and free carrier transport. Beginning with the treatment of light, at any given point $\boldsymbol{r} = (x, y, z)$ there is a total photon flux $J_v(\boldsymbol{r})$ that is distributed over (as $J_v(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r})$, *i.e.* irradiance) frequency ω, linear polarization component $\boldsymbol{\epsilon} = (\theta_\epsilon, \phi_\epsilon)$, and propagation direction $\boldsymbol{k} = (\theta_k, \phi_k)$.

To a first approximation we treat absorption as the only light-matter interaction under consideration in the film, neglecting the weaker mechanisms of elastic scattering and photoluminescence (PL); see detailed derivation (Appendix A) and Appendix D for discussion and justification. To treat attenuation of the field we derive the polarization- and frequency-dependent absorption cross-section of the film from the single-SWNT absorption behavior. The dipole approximation of light-matter interaction yields the absorption cross-section for a single nanotube of orientation $\hat{\boldsymbol{l}} = (\theta, \phi)$,

$$
\sigma(\boldsymbol{c}_i, \boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega) = |\boldsymbol{\epsilon} \cdot \hat{\boldsymbol{l}}|^2 \sigma_{\parallel}(\boldsymbol{c}_i, \omega) + \left(1 - |\boldsymbol{\epsilon} \cdot \hat{\boldsymbol{l}}|\right)^2 \sigma_{\perp}(\boldsymbol{c}_i, \omega),
\tag{59}
$$

where $\sigma_{\parallel}$ and $\sigma_{\perp}$ are the empirically- or theoretically-determined absorption cross-sections for light polarized parallel and perpendicular, respectively, to the SWNT longitudinal axis, reflecting the two anisotropic sets of optically active electronic transitions.[99] Integrated over all frequencies, the parallel absorption cross-section is approximately five times larger than the perpendicular due to depolarization.[99] For later convenience we normalize the cross-section to per-unit-length, σ_l, rather than the more common reported value per-carbon. Over our length-scales of interest for SWNT material, < 1 μm, we neglect phase shifts such as refraction that introduce circular polarization components to the light field.

From the perspective of the gradient in the light field, the diameter of the SWNT is negligible and the film is essentially homogenous (see detailed derivation, Appendix A, below). For irradiance $J_v(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r})$ the photon absorption rate per volume due to SWNT of chirality i oriented in $\hat{\boldsymbol{l}}$ are

$$N_i(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}, \hat{\boldsymbol{l}}|\boldsymbol{r}) = v_i \cdot \rho_{\langle l \rangle} \cdot \sigma_l(\boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega, \boldsymbol{c}_i) \cdot J_v(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}) \tag{60}$$

where $\rho_{\langle l \rangle} = \rho \cdot \langle l \rangle$ is the length-normalized SWNT film density (ρ is the number density). To get the total film absorption rate we must sum absorption terms (60) for nanotubes at each possible orientation $\hat{\boldsymbol{l}}$ and chirality i, weighted by the relative populations of each. That is equivalent however to integrating over the distribution $p(\hat{\boldsymbol{l}})$ and summing over chiralities, forming the crux of our methodology,

$$N(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}) = \sum_i \int_0^{2\pi} p(\phi) \int_0^{\pi} p(\theta)\, N_i(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}, \hat{\boldsymbol{l}}|\boldsymbol{r})\, d\theta\, d\phi. \tag{61}$$

In this manner, we can describe the light absorption of the film by integrating the single-SWNT behavior over the distribution of independent single-SWNT properties. We will take an analogous approach to exciton and free carrier transport below.

A steady-state balance on the light intensity yields

$$\nabla \cdot J_v(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}) = -N(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}). \tag{62}$$

Two useful reductions then emerge. First, neglecting Rayleigh scattering and photoluminescence the $\boldsymbol{k}$ dependence drops out. Second, in the x and y dimensions where our film is infinite we take periodic solutions as trivial, making our boundary conditions in those dimensions uniformity, *i.e.*

$$\frac{dJ_v}{dy} = 0, \qquad \frac{dJ_v}{dx} = 0. \tag{63}$$

Our balance then simplifies to

$$\frac{dJ_v(\omega, \boldsymbol{\epsilon}|z)}{dz} = -\sum_i N_i(\omega, \boldsymbol{\epsilon}|z). \tag{64}$$

The linear differential equation (64) can be solved for a transparent back electrode subject to the boundary condition of perpendicularly incident light unpolarized in the xy plane,

(65) $$J_v(\omega, \boldsymbol{\epsilon}|z=0) = J_0(\omega, \boldsymbol{\epsilon}) = J_0(\omega) \cdot \delta\left(\theta_\epsilon - \frac{\pi}{2}\right) \cdot \frac{1}{2\pi}, \qquad \phi_\epsilon \in [0, 2\pi].$$

We treat reflection off the back electrode without the propagation vector $\boldsymbol{k}$ by breaking the light field into approximately non-interfering 'forward' and 'reverse' fluxes $J_F(\omega, \boldsymbol{\epsilon}|z)$ and $J_R(\omega, \boldsymbol{\epsilon}|z)$ (See detailed derivation, Appendix A, below), where the boundary condition of the latter is the reflection of the former at $z = T$.

We then turn to the exciton transport problem. In SWNT, optical electronic excitations result in exciton generation – bound carrier states – rather than free charges, due to one dimensional confinement resulting in a weak dielectric environment.[98,100,177,178] To collect charges in a solar cell, excitons must be dissociated either at a Type II interface or by an electric field imparting coulomb force greater than the binding energy. Examining the relevant physical quantities (see detailed derivation, Appendix A below), we validate the approximation that nonradiative decay of hot excitons to the band gap is essentially instantaneous over our length scales of interest. This simplification allows us to treat the exciton generation rate at band gap energy E_i as the total photon absorption rate at all energies $\hbar\omega > E_i$ for chirality i, $N_i(z)$.

A single empirical diffusion coefficient describing exciton transport in a film would be limited to only a particular geometry and chemistry. Instead, we start with the intra-tube 1D exciton reaction-diffusion behavior, then deriving the contribution of a network of such systems to three-dimensional exciton transport, and finally coupling nanotubes via exciton hopping (EH), sometimes termed exciton energy transfer (EET).

Along a single nanotube axis there is an exciton concentration profile $n(r_l)$. We construct a steady-state exciton population balance along r_l including the generation rate $N(r_l) = N(\boldsymbol{r}')$, longitudinal diffusion, and relaxation due to localized impurities, exciton-exciton annihilation (EEA), radiative decay, and end-quenching. This yields the steady-state differential equation (see detailed derivation, Appendix A, below)

(66) $$\frac{dn}{dt} = N(r_l) + D_l \frac{d^2n}{dr_l^2} - k_\Gamma n - k_{EEA} n^2 - \sum_{im} k_{im} n_{im} n - k_{end}\big(\delta(r_l - 0) + \delta(r_l - l)\big)n = 0,$$

where D_l is the longitudinal diffusion coefficient, k_Γ, k_{EEA}, k_{im}, and k_{end} are, respectively, the radiative, Auger recombination, impurity (of type im, *e.g.*, oxidative agents, catalyst nanoparticles), and end-quenching rate constants, and n_{im} is the number of impurity contacts per length of SWNT.

To determine the network behavior, we construct a three dimensional population balance subject to diffusion in one dimensional channels. Starting with a single chirality, the diffusion flux (not yet including SWNT-SWNT interactions) due to nanotubes of orientation $\hat{\boldsymbol{l}}$ is

(67) $$D_l(\sin\theta\cos\phi)^2\frac{\partial c}{\partial x} + D_l(\sin\theta\sin\phi)^2\frac{\partial c}{\partial y} + D_l(\cos\theta)^2\frac{\partial c}{\partial z}$$

where $c(\boldsymbol{r})$ is the 3D exciton density. This expression is arrived at by considering 1D channels penetrating the faces of a differential volume (see detailed derivation, Appendix A, below).

Note that while Equation (67) has the form of what one would empirically intuit, there is the crucial difference that D_l is exactly the single-SWNT exciton diffusion coefficient, not a net film coefficient. That represents the primary advantage of deriving our network model from single- and paired-SWNT physics rather than using aggregate empirical parameters; in the latter case, we would not be able to predict the scaling of parameters with the distributions of nanotube properties.

Next we extend film relaxation kinetics to 3D. We multiply the reaction rates per length of SWNT by $\rho_{\langle l\rangle}$ to convert to rate per volume network under an assumption of macroscopic homogeneity:

(68)
$$\begin{aligned}
&-k_\Gamma c \\
&-\rho_{\langle l\rangle}\cdot k_{im}n_{im}n = -k_{im}n_{im}c \\
&-\rho_{\langle l\rangle}\cdot k_{EEA}n^2 = -k_{EEA}n\cdot c = -\frac{\rho_{\langle l\rangle}}{\rho_{\langle l\rangle}}\cdot k_{EEA}n\cdot c = -\frac{k_{EEA}}{\rho_{\langle l\rangle}}c^2 \\
&-\frac{k_{end}}{\rho_{\langle l\rangle}}c_{end}(\boldsymbol{r})\cdot c,
\end{aligned}$$

where $c_{end}(\boldsymbol{r}) = p(\boldsymbol{r}_{end} = \boldsymbol{r}_C \pm \boldsymbol{l}|\boldsymbol{r})$ is the distribution of nanotube ends in the film. Note that all of $\{k_{im}, k_{EEA}, k_{end}\}$ are single-nanotube rate constants, which can be evaluated from single-SWNT experimental measurement, theory, or simulation.

Finally we consider exciton transport between nanotubes. The mechanisms of exciton transfer/hopping (EH) between nanotubes of the same and different chirality, as well as the mechanisms of Type II exciton dissociation at donor-acceptor interfaces, remain under investigation and debate.[113,115,117,179] Explicitly, in the model we treat EH as a tunneling process in bundles and at interconnects, neglecting the real and significant phenomena of exciton delocalization[113,180] and Schottky barriers/band bending.[181] What is important for the future use and evolution of this framework however is that the relevant independent parameters are available, allowing a different functional form to be introduced without contradicting the remainder of the model. This consideration highlights the extensibility of our method beyond our explicit form. At present, neither exciton delocalization nor SWNT band bending at interconnects are sufficiently well understood for us to posit meaningful expressions describing them.

We divide the film into two microscopic environments: bundles and interconnects. The bundling coefficient, b_c, we define as the length fraction of SWNT in bundles. Along with the mean number of

SWNT in a bundle, M_b, we determine the diffusion coefficent in a bundle orthogonal to the longitudinal axis (see detailed derivation, Appendix A, below),

$$D_{EH,b} = \chi \frac{1}{4} k_{EH,b} (\Delta_B + d_c)^2, \tag{69}$$

where $\chi(M_b)$ is the mean number of nearest neighbors, Δ_B is the inter-SWNT separation in a bundle, and $k_{EH,b}$ is the elastic exciton transition rate. Beginning with a monochiral film, this yields a film exciton flux due to bundling,

$$\begin{aligned} & b_c(1 - \sin\theta\cos\phi)^2 D_{EH,b} \frac{\partial c}{\partial x} + b_c(1 - \sin\theta\sin\phi)^2 D_{EH,b} \frac{\partial c}{\partial y} \\ & + b_c(1 - \cos\theta)^2 D_{EH,b} \frac{\partial c}{\partial z}. \end{aligned} \tag{70}$$

At interconnects, we consider the number of intersections per length of a nanotube in a control volume around a nanotube,

$$n_I = \left(1 - b_c\left(1 - \frac{1}{M_B}\right)\right) \rho \frac{15}{4} \pi d_c^2. \tag{71}$$

That yields a film exciton flux, still for a monochiral film,

$$\begin{aligned} & Let\ \gamma_I \equiv \left(1 - b_c\left(1 - \frac{1}{M_B}\right)\right) d_c \cdot n_I = \rho\left(1 - b_c\left(1 - \frac{1}{M_B}\right)\right)^2 \frac{15}{4} \pi d_c^3 \\ & (1 - \sin\theta\cos\phi)^2 \cdot \gamma_I \cdot D_{EH,I} \frac{\partial c}{\partial x} + (1 - \sin\theta\sin\phi)^2 \cdot \gamma_I \cdot D_{EH,I} \frac{\partial c}{\partial y} \\ & + (1 - \cos\theta)^2 \cdot \gamma_I \cdot D_{EH,I} \frac{\partial c}{\partial z}, \end{aligned} \tag{72}$$

where the orthogonal diffusivity at interconnects is

$$D_{EH,I} = \frac{1}{2} k_{EH,I} \lambda^2 (\Delta_I + d_c)^2, \tag{73}$$

where Δ_I is the inter-SWNT separation at contacts, $k_{EH,I}$ is the elastic transition rate, and λ is the co-alignment:

$$\lambda \equiv \left|\hat{\boldsymbol{l}}' \cdot \hat{\boldsymbol{l}}''\right| = \lambda(\theta', \theta'', \phi', \phi''), \tag{74}$$

for an arbitrary pair of nanotubes, with a distribution $p(\lambda)$ that is a function of $p(\hat{\boldsymbol{l}})$ (see detailed derivation, Appendix A, below).

Including all three diffusion mechanisms, we then have the volume balance for exciton transport in a monochiral network. We simplify that expression by again recognizing that periodic solutions in the xy plane are trivial for our purposes, reducing our steady state problem to a differential equation in z,

$$Let \quad D_{tot} \equiv (\cos\theta)^2 D_l + (1 - \cos\theta)^2 \left(\gamma_I D_{EH,I} + b_c D_{EH,b}\right) \tag{75}$$

$$\frac{dc}{dt} = N(z) + D_{tot}\frac{d^2c}{dz^2} - k_\Gamma c - c\sum_{im} k_{im}n_{im} - \frac{k_{EEA}}{\rho_{\langle l\rangle}}c^2 - \frac{k_{end}}{\rho_{\langle l\rangle}}c_{end}(z)\cdot c.$$

Analogous to the light absorption problem, we have constructed a differential equation describing exciton transport due to nanotubes with orientation $\hat{\boldsymbol{l}}$, but to get the total balance we must sum contributions from all possible orientations. Again, that weighted sum is equivalent to integrating (75) over the distribution $p(\hat{\boldsymbol{l}})$. This treatment is only valid under our approximation of a homogenous film. That is our method: treat single-SWNT chemical and geometric parameters as random variables, derive network behavior from single-SWNT physics as a function of those parameters, and finally integrate the solution over the distribution of those parameters. We illustrate this process when we apply the model to specific film structures below.

Note that equation (75) has the intuitive reaction-diffusion form one might guess empirically, but the net diffusivity D_{tot} and several of the kinetic rate constants depend on the film properties $(p(\hat{\boldsymbol{l}}), \rho_{\langle l\rangle}, p(l))$, preventing broader applicability of any single experimental diffusivity measurement. Additionally, by considering the distributions of single-SWNT parameters and the microscopic mechanisms of exciton transport and decay we have determined the dependence of network diffusion and reaction constants on film properties and single-exciton physical constants.

SWNT light absorption and exciton transport mechanics, particularly in two-nanotube systems, is an ongoing area of active research; our expressions describing those mechanics are unlikely to be final. It is under that consideration that the extensibility of our framework, built on treating single-SWNT properties as random variables, becomes particularly valuable. Specific terms, such as exciton hopping diffusivities, can be modularly altered as their physics are better understood. New parameters with associated distributions can be introduced, such as dielectric constant and local density (radial distribution function), and variable independence assumptions can be relaxed, increasing complexity and nuance without grossly deteriorating numerical solvency.

Multiple chiralities, including metallic SWNT (mSWNT), which quench excitons through rapid nonradiative decay, are then introduced. We construct volume balances for each chirality *i* analogous to the monochiral balance (75), but with additional quenching terms due to EH to mSWNT in bundles and at interconnects, and pairwise coupling terms to one another due to EH to lower-bandgap chiralities. This yields a system of ordinary differential equations,

$$\frac{dc_i}{dt} = v_i N(z) + D_{tot}\frac{d^2 c_i}{dz^2} - k_\Gamma c_i - c_i \sum_{im} k_{im} n_{im} - \frac{k_{EEA}}{\rho_{\langle l \rangle}} c_i^2$$
$$- \frac{k_{end}}{\rho_{\langle l \rangle}} c_{end}(z) \cdot c_i - k_{EH}\langle \lambda^2 \rangle \cdot v_m \gamma_I c_i$$
$$- b_c D_{EH,b} v_m \frac{2}{\sqrt{3}} d_c^{-2} \cdot c_i$$
$$- c_i \sum_{j<i} \left(k_{EH}\langle \lambda^2 \rangle \cdot v_j \gamma_I + b_c \cdot D_{EH,b} v_j \frac{2}{\sqrt{3}} d_c^{-2} \right)$$
$$+ \sum_{j>i} c_j \cdot \left(k_{EH}\langle \lambda^2 \rangle \cdot v_j \gamma_I + b_c \cdot D_{EH,b} v_j \frac{2}{\sqrt{3}} d_c^{-2} \right), \quad (76)$$

where v_m is the fraction of mSWNT in the film, and again we have indexed the chiralities as monotonically increasing in optical band gap ($E_{i+1} > E_i$).

The system of ODEs (76) can be integrated over the relevant distributions and solved at steady state subject to boundary conditions at the electrodes ($z = 0, T$). A variety of electrode configurations can be considered; here we focus on two electrodes that can drive Type II exciton dissociation – one hole accepting and one electron accepting. For each electrode we therefore have Robin boundary conditions from the exciton splitting rate,

$$\left.\frac{dc_i}{dz}\right|_{z=0} = \frac{k_{d0,i}}{D_{tot}} c_i(z=0)$$
$$\left.\frac{dc_i}{dz}\right|_{z=T} = -\frac{k_{dT,i}}{D_{tot}} c_i(z=T), \quad (77)$$

where $k_{d0,i}$ and $k_{dT,i}$ are the dissociation rate constants that in general could depend on the chirality i.

We complete the derivation by considering the drift-diffusion transport of charge carriers resulting from exciton dissociation. Free carrier transport in the nanotube network is in many ways more challenging than exciton diffusion. In addition to the analogous geometric effects on film charge mobility and diffusivity, charge transfer at the interface of nanotubes, particularly of different chiralities, can block or trap charges.[181] Asymmetric electron and hole generation (at the two electrodes) can lead to an electric field development that feeds back on the exciton dissociation problem, inhibiting dissociation and/or driving spontaneous in-film dissociation when the field strength exceeds the binding energy. Finally, carrier densities can significantly impact mobility.[182] In this work we do not seek to accurately describe free carrier transport in the SWNT film. Rather, we proceed under the substantial simplification that even under a modest electric field performance will be limited by exciton diffusion, due to exceptioanlly high carrier mobilities.[182] This assumption reduces coupling between the free carrier and exciton problems to the exciton dissociation rate. We also continue to neglect any charge-transfer effects

at inter-SWNT contacts, as those effects have been neither theoretically nor experimentally quantified. These rough assumptions allow, as we show, important results in the optimization of film performance, but inflate quantitative predictions of external quantum efficiency (EQE).

We estimate the geometric impact on valence and conduction band carrier mobilities in the film from the exciton behavior. We attenuate single-SWNT electron and hole mobilities, μ_e^* and μ_h^*, by the network attenuation of single-SWNT exciton diffusivities. We then balance the non-equilibrium populations of conduction band carriers, $f_e(\boldsymbol{r})$, and valence band carriers, $f_h(\boldsymbol{r})$, which are coupled to one another and the electric field $\mathcal{E}_z(z)$. This yields population balances

$$\frac{df_e}{dt} = 0 = D_e\frac{d^2f_e}{dx} + \mu_e\left(\mathcal{E}_z\frac{df_e}{dx} + f_e\frac{d\mathcal{E}_z}{dx}\right) - k_{e-h}f_ef_h. \tag{78}$$

$$\frac{df_h}{dt} = 0 = D_h\frac{d^2f_h}{dx} - \mu_h\left(\mathcal{E}_z\frac{df_h}{dx} + f_h\frac{d\mathcal{E}_z}{dx}\right) - k_{e-h}f_ef_h, \tag{79}$$

where $D_{e/h}$ and $\mu_{e/h}$ are the conduction/valence band carrier network diffusivity and mobility, respectively, and k_{e-h} is the electron-hole recombination rate constant. Equations (78) and (79) are subject to generation boundary conditions from exciton dissociation at one electrode, and a Robin (reaction) boundary condition due to extraction at the opposite electrode (see detailed derivation below). The electric field evolves in the z axis from the gradient in excess charge populations,

$$\frac{d\mathcal{E}_z}{dz} = \frac{q}{\varepsilon}\left(f_h(z) - f_e(z)\right), \tag{80}$$

where ε is the mean permittivity of the film. The intrinsic bias across the film from mismatched electrode work function forms the boundary condition to (80). The three differential equations can be solved simultaneously at steady state.

Model Limitations

Several explicit and implicit assumptions have been made in the derivation of the model. To encourage both application of the model and its further development, we would like to discuss some of the most significant limitations. Most of these are failings in our application of the underlying principle rather than the overarching method of treating single-SWNT parameters as random variables, deriving the photon, exciton, and free carrier population balances from single- and paired-SWNT physics, and integrating the resulting deterministic equations over the parameter distributions that define the film.

In both the exciton and free carrier transport problems, we neglect equilibrium charge transfer at the interface of nanotubes, particularly those of different chirality; in reality, Schottky barriers/band bending likely occur at these contacts,[181] which can alter inter-SWNT transport. We expect this phenomenon to play a substantial role in free carrier transport in particular, which is one reason we

adopted a simplified model of steady state charge distribution. Presently, experimental or theoretical examinations are necessary to elucidate the energetic and spatial character of the effect, such as achieved theoretically by *Topinka et al.* in 2009[181] and experimentally by *Jakukba et al.* in 2013.[183] A more accurate model should account for these interfacial interactions and local electric fields in the inter-SWNT coupling terms. Along the same lines, we have neglected the possibility of coherent delocalization of excitons in bundles, which recent experiments suggest to be the case,[113] contrary to previous assumptions of excitons in nanotube bundles; this effect could substantially raise orthogonal exciton diffusion lengths in highly aligned filmed.

We have also neglected Rayleigh scattering and photoluminescence processes in the light-matter interaction. These assumptions simplified our application of the model tremendously, at the potential cost of accuracy. See the detailed derivation, Appendix A, and Appendix D, below for an evaluation of the strength of these assumptions and the utility of the simplifications to the model.

Finally, the validity of our integration over single-SWNT properties rests on an approximation of homogeneity. This limitation has significant implications for certain cases, particularly in-plane ('horizontally') aligned, low-density films. The conditions and implications of this error, including estimation of the impact, are discussed in depth in Appendix B *Breakdown of Macroscopic Heterogeneity*. The homogeneity approximation also implies that we can only treat cases with SWNT densities above the percolation threshold.

Estimating Physical Constants

To both demonstrate how the framework can be practically employed and to quantitatively examine key SWNT PV properties we have applied the model to monochiral (6,5) films with both isotropic and aligned orientation distributions (Figure 2).

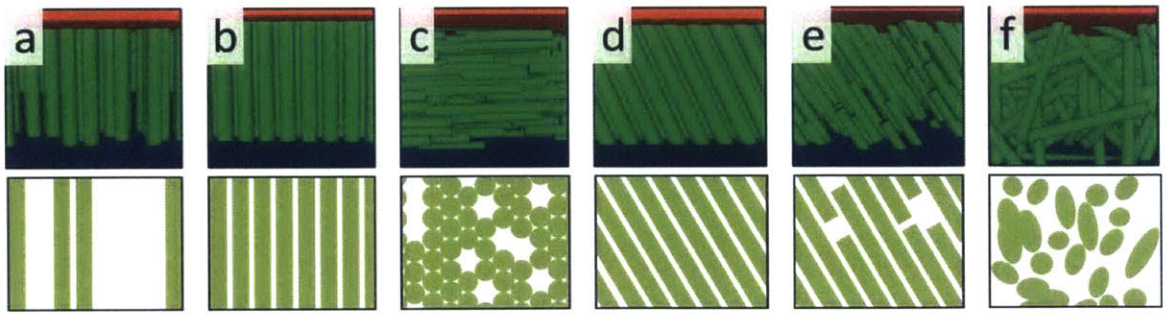


Figure 56. Cartoon representations of film geometries under consideration, side view (top) and cross-section (bottom). Vertically aligned films with densities ranging from sparse (a) to close-packed (b), horizontally aligned films at densities above the percolation threshold (c), aligned films at arbitrary angles with single-layer (d) and multi-layer (e) depositions (see text), and isotropic films (f).

To do so, we first had to determine the relevant physical constants from experimental measurement or through estimation.

Incident light flux We take the intensity and frequency distribution of incident light to be the AM1.5 solar spectrum. Even in numerical evaluation of the model, a continuous form of the spectrum is desirable for interpolation of experimental data and fine discretization of the spectrum. The spectrum can be fit to better than 99% rms error by subtracting a series of Gaussian profiles from a geometrically attenuated blackbody spectrum. The blackbody spectrum attenuated by the atmosphere at an angle of 0.26° to the sun (angle of earth to sun) is[184]

$$b_s(E) = \frac{2F_s}{h^3c^2}\left(\frac{E^2}{e^{E/kT_s}-1}\right) \tag{81}$$

in units photons per area-time per energy photon, where $F_S = 2.16 \times 10^{-5}\,\pi$ is a geometric factor, h is Planck's constant, c is the speed of light in a vacuum, T_s is 5960K (temperature of the surface of the sun), k_B is boltzmann's constant, and E is the energy of the light. To get the flux distributed in wavelength we convert

$$b_s(\lambda = hc/E) = \frac{1}{\lambda}\cdot E \cdot b_s(E). \tag{82}$$

From the spectrum we subtract a series of Gaussian functions to fit the AM1.5 spectrum, e.g. for 60 Guassians we have

$$J_0(\lambda) \cong \frac{E}{\lambda}\cdot\frac{2F_s}{h^3c^2}\left(\frac{E^2}{e^{\frac{E}{kT_s}}-1}\right) - \sum_{i=1}^{60} A_i e^{-\frac{(\lambda-\lambda_i)^2}{2\cdot\sigma_{\lambda i}^2}}. \tag{83}$$

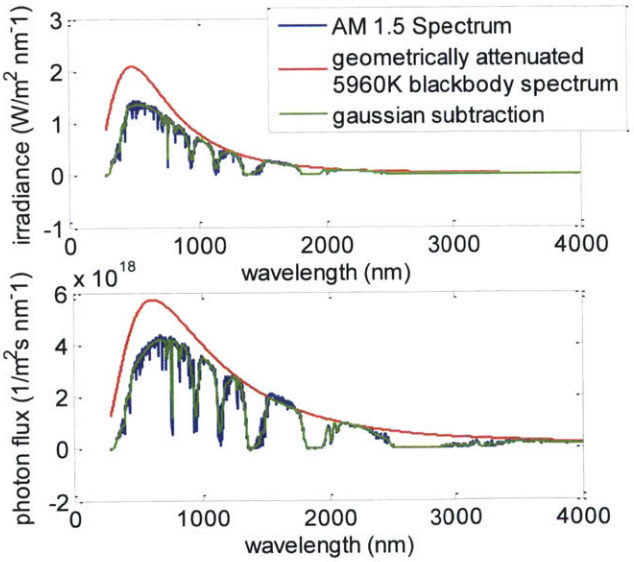


Figure 57. photon flux and irradiance distributions in wavelength for AM1.5 light, showing the experimental data (blue), 5960K blackbody spectrum attenuated by the earth-sun ray (red), and the same with 40 Gaussian curves subtracted to fit the data.

Table 1. Results of multi-Guassian function fit to the AM1.5 spectrum (Figure 57), where for each curve the mean μ, variance σ^2, and maximum A of the curve were adjusted.

μ (nm)	100	205.3	298.7	374.02	513.62	603.05	706.77	941.36	945.56	1052.6
σ (nm)	20	27.144	23.678	58.316	51.216	34.191	126.1	31.948	11.867	31.996
A (W/m^2)	0.43013	6.6622	0.47979	1.2682	0.52368	0.15331	0.42537	0.29952	0.27733	0.08644
μ (nm)	1131.7	1189.8	1314.3	1344.9	1436.9	1523	1627.7	1822	1915	2011.1
σ (nm)	20.299	23.864	24.212	43.435	56.254	23.855	72.786	33.832	46.216	9.2091
A (W/m^2)	0.46103	0.09104	-0.18011	0.3014	0.23276	-0.07346	-0.01938	0.1312	0.12825	0.07103
μ (nm)	2061	2194.6	2299	2394.2	2505.2	2600.2	2700.3	2800.1	2898.3	2999.3
σ (nm)	12.398	20.316	34.468	43.054	44.269	49.389	49.978	49.997	49.196	46.776
A (W/m^2)	0.03767	0.00369	0.00610	0.02119	0.04202	0.03648	0.03102	0.02726	0.02394	0.01759
μ (nm)	3100.1	3215	3305.8	3399	3500.9	3594.2	3696.5	3798.4	3899.2	3998.4
σ (nm)	46.272	31.193	42.886	39.739	34.614	32.971	30.404	27.578	30.045	23.088
A (W/m^2)	0.02027	0.01778	0.01521	0.00778	0.00415	0.00486	0.0039	0.00324	0.00409	0.00280

(6,5) Absorption Cross-Section On resonance with the S_{22} transition the parallel polarized absorption cross-section of a single (6,5) SWNT has been measured as 3.2×10^{-17} cm^2 per carbon atom.[185] With $8.8271 \times 10^{10}\ m^{-1}$ carbon atoms per length,[185] the absorption cross-section at S_{22} per length of nanotube is

$$\sigma_{l,meas}(S_{22}) = 2.92986 \times 10^{-10} \frac{m^2}{m}. \tag{84}$$

The wavelength dependence of $\sigma_{l,meas}(\omega)$ in general will depend on the environmental conditions, e.g., surfactant type, polymer, solid film or solution phase.[116] We estimate the dependency for a surfactant-free solid-state SWNT film by normalizing the absorbance spectrum of an isotropic film, from reference [116]. We scale that dependency by the S_{22} peak, at 582 nm in our data (Figure 59). Analogously to the incident solar flux, for a continuous expression we fit the data with a series of Lorentzian curves (Figure 59, Table 2)

$$\sigma_{l,meas}(\lambda) \cong \sum_{i=1}^{20} \frac{A_i}{(\lambda - \lambda_i)^2 + \gamma_{\lambda i}^2}. \tag{85}$$

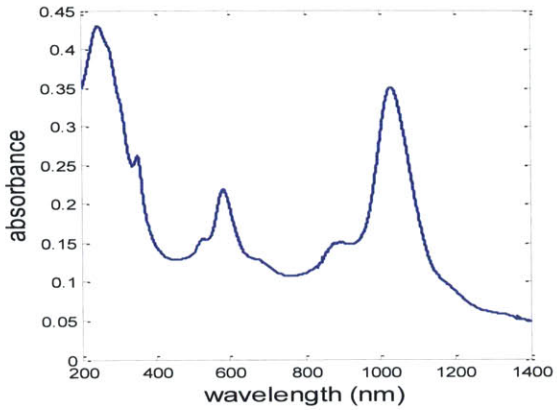


Figure 58. absorbance of isotropic film of (6,5) SWNT, from [116].

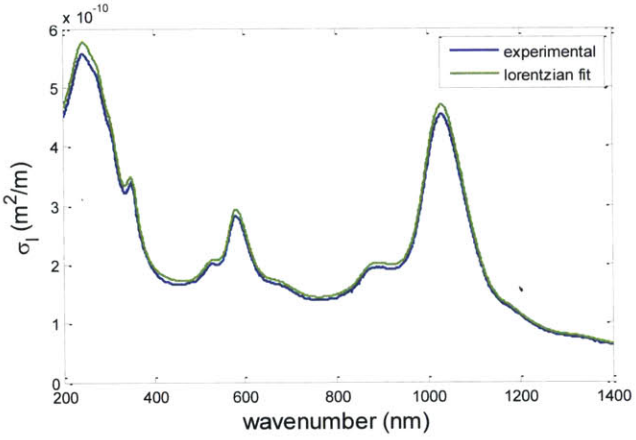


Figure 59. absorption cross-section of (6,5) SWNT as a function of wavelength, estimated from the absorbance spectrum of an isotropic surfactant-free film. For a continuous expression, data has been fit with a series of Lorentzian curves.

Table 2. Fit parameter values for Lorentzian series fit of the absorption cross-section spectrum.

Mean, λ_i	$\gamma_{\lambda i}$	Amplitude, A_i ($\times 10^9$)
232.83	103.61	5000.00
242.77	19.452	22.27
308.35	12.665	7.31
278.12	27.524	69.72
352.19	13.986	18.31
489.88	89.528	408.84
524.62	21.783	17.31
579.12	22.764	42.15
594.89	35.838	127.10

676.72	54.687	130.03
788.72	200	3085.30
867.95	24.518	8.67
894.02	58.505	228.62
1005.6	32.675	86.77
1029.9	43.087	417.75
1064.2	60.168	631.40
1178.9	34.132	20.28
1216.2	57.532	62.27
1338.6	117.2	590.47
1397.7	22.647	3.18

While the frequency dependence of $\sigma_{\parallel}(\omega)$ and $\sigma_{\perp}(\omega)$ will be different due to different optical transition modes, no polarization-dependent absorption data is available for (6,5) SWNT. We therefore crudely approximate as the same frequency dependence, and introduce the depolarization effect by an expected 1:5 net ratio of perpendicular:parallel absorbance,[186]

$$\sigma_{\parallel}(\omega) = \frac{5}{6}\sigma_{l,meas}(\omega) \tag{86}$$

$$\sigma_{\perp}(\omega) = \frac{1}{6}\sigma_{l,meas}(\omega). \tag{87}$$

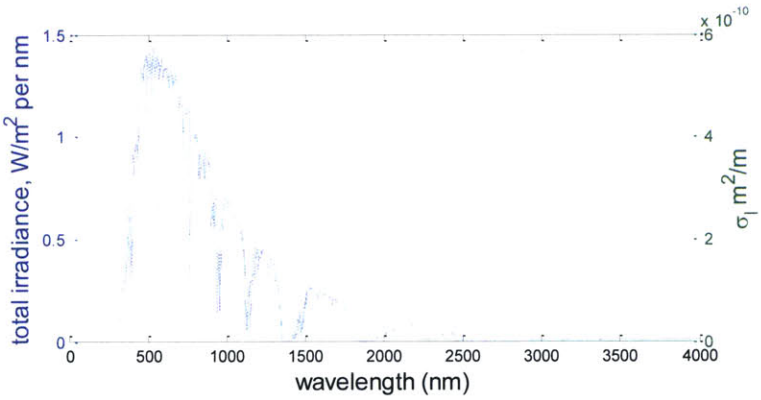


Figure 60. (6,5) isotropic absorption cross-section overlaid over solar spectrum.

Density The diameter of (6,5) SWNT is $d_{c,c-c} = 0.76\ nm$,[187] yielding an approximate $d_c = 1.095\ nm$ (see Appendix A). We vary film density to investigate its influence, but benchmark values to the maximum density of a SWNT film – the close-packed density (see section below), $\rho_{\langle l \rangle}^{CP} = 9.63 \times 10^{17}\ m^{-2}$. Current fabrication methods exhibit varying densities, but at present many are far short of

close-packed; vertical forests are <0.1% $\rho_{\langle l \rangle}^{CP}$,[188] and isotropic films are typically lower,[116] although horizontally aligned films have been produced up to 35% of CP density by some methods.[189,190]

The unit-cell area of hexagonally-packed SWNT, including interstitial space, is that of a rhombus with interior angle $\pi/3$ (Figure 61),

$$2 \cdot d_{6,5} \cdot d_{6,5} \cdot \sin\frac{\pi}{3}\cos\frac{\pi}{3} = \frac{\sqrt{3}}{2} d_{6,5}^2. \quad (88)$$

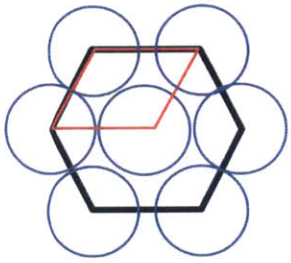


Figure 61. Cartoon of cross-section of SWNT bundle demonstrating the single-SWNT unit cell area.

This makes the length density in a bundle

$$\rho_{\langle l \rangle}^{CP} = \frac{2}{\sqrt{3} d_{6,5}^2} = 9.63 \times 10^{17}\ m^{-2}. \quad (89)$$

Kinetic and Diffusion Constants Measured values for longitudinal diffusion coefficients in SWNT span three orders of magnitude, ranging from 0.1-0.4 cm^2/s[102,103] and 7.5-10.7 cm^2/s[104,105] to 150-350 cm^2/s[106,107]. Comparison of the environments of those measurements, along with comparison of other environmentally sensitive investigations,[104,108-111] strongly suggests that this variation is due to influence of the local environment on exciton-phonon scattering. Under this hypothesis, lower diffusivity values (and associated diffusion lengths) have been observed in surfactant-wrapped SWNT in solution phase, while SWNT in air exhibit the higher diffusivity values. For a solid-state film therefore we estimate $D_l \cong 100\ cm^2/s$.

For radiative decay we take a time constant of $\tau_\Gamma = 0.45\ ns$,[104,111] yielding $k_\Gamma = 2.2 \times 10^9\ s^{-1}$. For impurity quenching, a variety of potential impurity types could be present in a film depending on the SWNT growth, separation, purification, and deposition methods; common possibilities would include lingering metal catalyst particles, sonication- or oxidation-induced sp^3 functionalization of nanotube sidewalls (lattice defect formation), and adsorbed oxygen.[191] We define some generic impurity that introduces inter-bandgap electronic states with a time constant consistent with nonradiative relaxation, $O(200\ fs)$. From that we estimate the rate constant as

(90) $$k_{im} = \sqrt{\frac{D_l}{\pi\tau_{im}}} = \sqrt{\frac{100\times10^{-4}}{\pi\cdot 200\times10^{-15}}} = 1.26\times10^{5}\frac{m}{s}$$

and take a representative population of two generic impurities per micron of SWNT length, $n_{im} = 2\times 10^{6}\ m^{-1}$ (which we vary to examine the impact of below).

The exciton-exciton annihilation rate constant we arrive at from [192] (see Appendix A), with $E_{BE,(6,5)} = 0.285\ eV$ and $E_{11,(6,5)} = 1.245\ eV$,

(91) $$k_{EEA}(c_i) = \left(6\times10^{5}\frac{m}{s}\right)\left(\frac{E_{BE,i}}{E_i}\right)^{3}\left(\frac{1.3\ eV}{0.3\ eV}\right)^{3} \cong 6\times10^{5}\frac{m}{s}.$$

End quenching we treat analogously to impurity quenching as the decay pathway is similar, giving $\tau_{end}\sim\tau_{im} \rightarrow k_{end} \cong 1.26\times10^{5}\ m/s$. Finally we estimate the exciton hopping rate constant from [115,193] as $k_{EH} \cong 0.1\ ps^{-1} = 10^{11}\ s^{-1}$.

For free carrier physical constants, we estimate a single-SWNT free carrier mobility from [128,182,194] in the $10^{4}-10^{5}\ cm^{2}V^{-1}s^{-1}$ regime, taking $\mu_e^* = \mu_h^* = 5\times10^{5}\ cm^{2}V^{-1}s^{-1}$; under the Einstein relation (see Appendix A) the resulting diffusivities at 300K are $D_e^* = D_h^* = 1.3\ m^{2}/s$. The electron-hole scattering length of $L_{eh} \cong 0.5\ um$ [128] then allows us to estimate the recombination rate constant,

(92) $$\tau_{eh} = \frac{L_{eh}^{2}}{D_e^*} = 1.93\times10^{-13}\ s$$

(93) $$k_{eh} \cong d_{6,5}^{2}\sqrt{\frac{D_e^* + D_h^*}{\pi\tau_{eh}}} = 2.5\times10^{-12}\ m^{3}/s.$$

Finally we estimate the relative film permittivity as that of SWNT, ~5.

Modelling Aligned Films

It has been intuited in the literature that an aligned carbon nanotube film at some intermediate angle (between horizontal and vertical) will present maximal photovoltaic performance by balancing light absorption with exciton transport;[120] to evaluate that hypothesis and establish the optimal angle and thickness of an aligned film, we apply the model to that case. While current fabrication methods can achieve neither arbitrary alignment angle nor near-close-packed densities, our predictions motivate progress in both. The network geometry is defined by the distributions of single SWNT properties and a few constants; in using the model to describe a particular experimental device, those distributions must be measured or estimated. We also examined films with isotropic orientation distributions, yielding very similar trade-offs between light absorption and exciton transport (following section).

For our hypothetical film, we take the length distribution as log-normal,[195]

(94) $$p(l) = \frac{1}{l \cdot \sigma_{\langle l \rangle}\sqrt{2\pi}} e^{-\frac{(\ln l - \mu_l)^2}{2\sigma_{\langle l \rangle}^2}}.$$

We look at the effect of varying it later, but as a baseline take mean length $\langle l \rangle = 400\ nm$ and second moment 20^2 nm^2. The fixed angle of alignment relative to the z axis we define θ', and we focus on the case of perfect alignment,

(95) $$p(\theta) = \delta(\theta - \theta'), \qquad \theta' \in \left[0, \frac{\pi}{2}\right].$$

In the xy plane all angles are equivalent in our semi-infinite film; we choose orientation along the x axis,

(96) $$p(\phi) = \delta(\phi - 0).$$

The film thickness, T, we will vary, but it can be limited by deposition morphology: if the aligned film is grown or stacked in a single 'forest' or 'layer' (Figure 56d), T cannot exceed the height in z of that single layer, $\langle l \rangle \cos\theta'$; additionally the distribution of $c_{end}(z)$ will be heterogenous (log-normal). If instead the film is 'sliced' out of a network mixed in all dimensions, like a typical solution-processed horizontally aligned film,[196] (Figure 56e) any thickness is hypothetically permissible, and $c_{end} = 2\rho$ is homogenous. These two film types and our handling of them are discussed in Appendix H.

Beginning our calculation with the irradiance, the absorption cross-section polarization dependence is described as equation (59) for our single chirality. We are in a position now to evaluate the dot product, $\boldsymbol{\epsilon} \cdot \hat{\boldsymbol{l}}$, by integrating the photon balance over $p(\hat{\boldsymbol{l}})$ (equation (61)). That yields the absorption cross-section (see Appendix I)

(97) $$\sigma_l(\boldsymbol{\epsilon}, \omega) = \sigma_{l,meas}(\omega)\left(\frac{5}{6}(\sin\theta_\epsilon \cos\phi_\epsilon \sin\theta' + \cos\theta_\epsilon \cos\theta')^2 + \frac{1}{6}(1 - |\sin\theta_\epsilon \cos\phi_\epsilon \sin\theta' - \cos\theta_\epsilon \cos\theta'|)^2\right).$$

Solving the photon balance (64) with no z dependence to $\rho_{\langle l \rangle}$ or σ_l, subject to the incident unpolarized solar flux boundary condition (65), and integrating over θ_ϵ, our forward flux becomes (see Appendix I),

(98) $$J_F(\omega, \phi_\epsilon | z) = \frac{1}{2\pi} J_0(\omega) \cdot \exp\left[-\rho_{\langle l \rangle}\sigma_{l,meas}(\omega)\left(\frac{5}{6}(\cos\phi_\epsilon \sin\theta')^2 + \frac{1}{6}(1 - |\cos\phi_\epsilon \sin\theta'|)^2\right) z\right].$$

Except in the vertically-aligned case ($\theta' = 0$), there is an xy-plane linear polarization dependence to the field absorption. With a boundary condition of reflectance R off the back electrode we solve the reverse flux differential equation, again substituting in $\sigma_l(\boldsymbol{\epsilon}, \omega)$ and integrating over θ_ϵ to yield

$$J_R(\omega, \phi_\epsilon|z) = \frac{1}{2\pi} \cdot R \cdot J_0(\omega) \cdot \exp\left[-\rho_{\langle l\rangle}\sigma_{l,meas}(\omega)\left(\frac{5}{6}(\cos\phi_\epsilon \sin\theta')^2 + \frac{1}{6}(1 - |\cos\phi_\epsilon \sin\theta'|)^2\right) \cdot (2T - z)\right]. \quad (99)$$

The total light field gradient would then be

$$J_v(\omega, \phi_\epsilon|z) = J_R(\omega, \phi_\epsilon|z) + J_F(\omega, \phi_\epsilon|z), \quad (100)$$

for example plotted in Figure 62 for horizontally and vertically aligned films.

The exciton generation rate is the light absorption rate integrated over frequency and polarity,

$$N(z) = \int_0^{2\pi}\int_0^{\infty} \rho_{\langle l\rangle} J_v(\omega, \phi_\epsilon|z)\sigma_l(\phi_\epsilon, \omega)\, d\omega d\phi_\epsilon. \quad (101)$$

We then resolve the monochiral exciton balance (75) with the physical constants estimated in the previous section. With an aligned film, our macroscopic homogeneity approximation limits the density space we can probe; for films with an in-plane (horizontal) orientation component, our treatment of exciton hopping is valid only at densities above the percolation threshold ($\rho/\rho_{CP} > 0.65$ for a horizontally-aligned film $\theta' = \pi/2$). Otherwise, we would be erroneously assuming in our orthogonal diffusion coefficient that percolation pathways uniformly exist. Above the threshold, we safely have a uniform orthogonal diffusivity ($\gamma_I \to 0$). Integrating the balance (75) over $p(\hat{\boldsymbol{l}})$ ((95) and (96)), the net diffusion coefficient reduces to

$$D_{tot} = D_l(\cos\theta')^2 + (1 - \cos\theta')^2 D_{EH,b}. \quad (102)$$

With $\Delta_B \ll d_c$, Equation (69) yields $D_{EH,b} = 1.8 \times 10^{-7}\ m^2/s$; note this diffusivity is 5 orders of magnitude lower than D_l ($\cong 10^{-2}\ m^2/s$). We then numerically solve (75) at steady state, subject to electrode boundary conditions (77).

We can first compare the two extreme cases, vertically ($\theta' = 0$) and horizontally ($\theta' = \pi/2$) aligned, at the same density, thickness, *etc.* ($T = 40$ nm, $\rho_{\langle l\rangle} = 0.8\rho_{\langle l\rangle}^{CP}$, $R = 1$, $\langle l\rangle = 800\ nm$, no added impurities; Figure 62). The vertical forest presents the incident light with only the perpendicular absorption cross-section (axial transition dipole), while the horizontal film can be excited along both the longitudinal and the axial transition dipoles, leading to anisotropic absorption (Figure 62b). The higher parallel absorption cross-section makes a substantial difference in total absorptivity even at this high density and with the aligned-film diminishing returns of more rapidly depleting parallel-polarized irradiance (Figure 62b). For these particular vertical and horizontal films, the total absorbances are 1.51 and 2.13 respectively. In absolute intensity absorption of visible light is higher than nIR (Figure 62a) due to greater visible irradiance in the incident solar flux, but normalized for intensity nIR absorption is

greater (Figure 62d). As light travels through the film, the anisotropy of the horizontal network induces a linear polarization (Figure 62b).

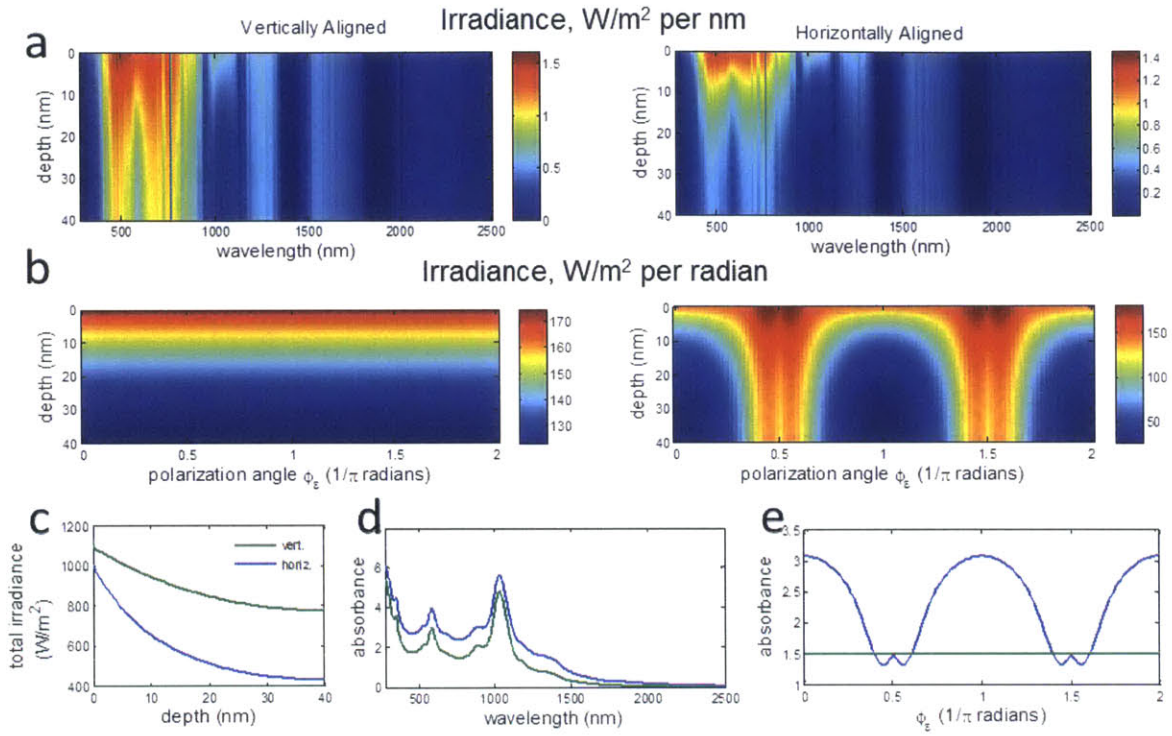


Figure 62 Comparison of vertically (green, $\theta' = 0$) and horizontally (blue, $\theta' = \pi/2$) aligned films at $T = 40\ nm$, $\rho_{\langle l \rangle} = 0.8\rho^{CP}_{\langle l \rangle}$, $R = 1$. a) irradiance per wavelength as light travels through the film. Note the more rapid attenuation in the horizontally aligned case. b) irradiance per linear polarization angle in the xy-plane as light travels through the film, showing the isotropic absorption of the vertically aligned film and induced linear polarization in the horizontally-aligned film. c) total irradiance with depth in the film, integrated over all wavelengths and polarization angles. d) total film absorbance as a function of wavelength. e) total film absorbance as a function of polarization angle. For exciton concentration gradients and the relative impact of each quenching mechanism, see supplemental Figures in Appendices.

Despite the higher absorbance, the substantially lower diffusivity in the charge-collecting axis of the horizontally aligned film ($D_{tot} = O(10^{-2}\ m^2/s)$ versus $O(10^{-7}\ m^2/s)$) greatly limits the efficiency – for these films EQEs are 64.6% and 2.7% respectively for vertically and horizontally aligned. In the horizontally aligned case a diffusion length <10 nm is observed (Figure 63), consistent with experimental findings.[115] With current experimental techniques however, horizontally aligned films can be made with substantially higher densities than vertically aligned films, which presently are grown as forests.[116,188-190] To evaluate the impact of this limitation, we explored the density-thickness space at the two orientations.

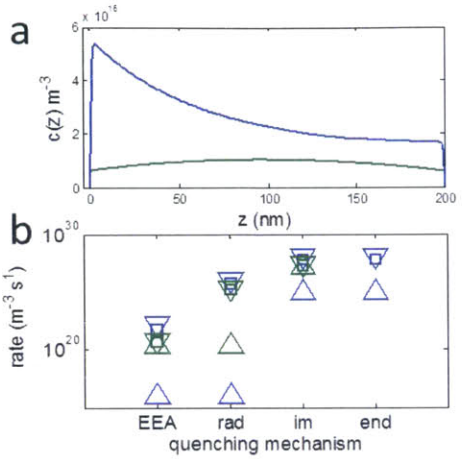


Figure 63. Comparison of exciton transport behavior in vertically (green) and horizontally (blue) aligned films at $T = 40\,nm$, $\rho_{\langle l \rangle} = 0.8\rho_{\langle l \rangle}^{CP}$, $R = 1$. a) exciton concentration at depth z in the film. Note the short (<10 nm) diffusion length of excitons in the horizontally-aligned film, leading to poor collection even at high concentrations. b) mean (squares) and minimum/maximum (triangles) quenching rates for each mechanism of interest in the film.

We began by considering an idealized case where the nanotubes are infinitely long and there are no impurities of any kind (Figure 64).[xxxi] Several features of interest emerged. First, we observed that performance is not monotonically increasing in the amount of light absorbed, and that rather at a given density there is an optimal thickness that shifts lower with higher density (that trend is clearer when we introduce finite length, see Figure 65c). This result reflects a previously posited trade-off with increasing thickness between greater total exciton generation (more light absorption) but greater mean distance a generated exciton has to travel to reach an electrode, what *Wu et al.* termed absorption and diffusion length scales.[197] The same tradeoff is reflected in the much lower optimal thickness of horizontal compared to vertical films; as noted, the diffusivity in the charge-collecting axis is several orders of magnitude lower in the horizontal film, leading to a diffusion length of order 10 nm. As a result, above a few tens of nanometers there are diminishing returns to greater absorption due to low exciton collection, consistent with recent experimental realization of higher performance in order-10-nm essentially-horizontal films versus previous order-100-nm films.[118] In contrast, vertically aligned films, where exciton dissociation is driven in the axis parallel to longitudinal diffusion, shows optimal thicknesses two orders of magnitude larger than horizontal films. The enormous longitudinal SWNT diffusion coefficient further manifests as substantially higher performance of vertically aligned films versus horizontally aligned films above 1%CP density. The higher absorptivity of the horizontally aligned film, due to depolarization, is not sufficient to compensate for the difference in diffusivity except at very low thicknesses and densities. In both cases, as density rises the optimal thickness shifts lower: higher density increases the light absorbed per thickness, shifting the optimum more towards favoring exciton transport (Figure 65).

[xxxi] While we do extract EQE, given the estimation of many parameters and the approximate consideration of charge transport we are not fixated on the absolute values of EQE; while they may be representative they are not fundamental, and are primarily valuable for comparing the differential impact of geometric and chemical parameters.

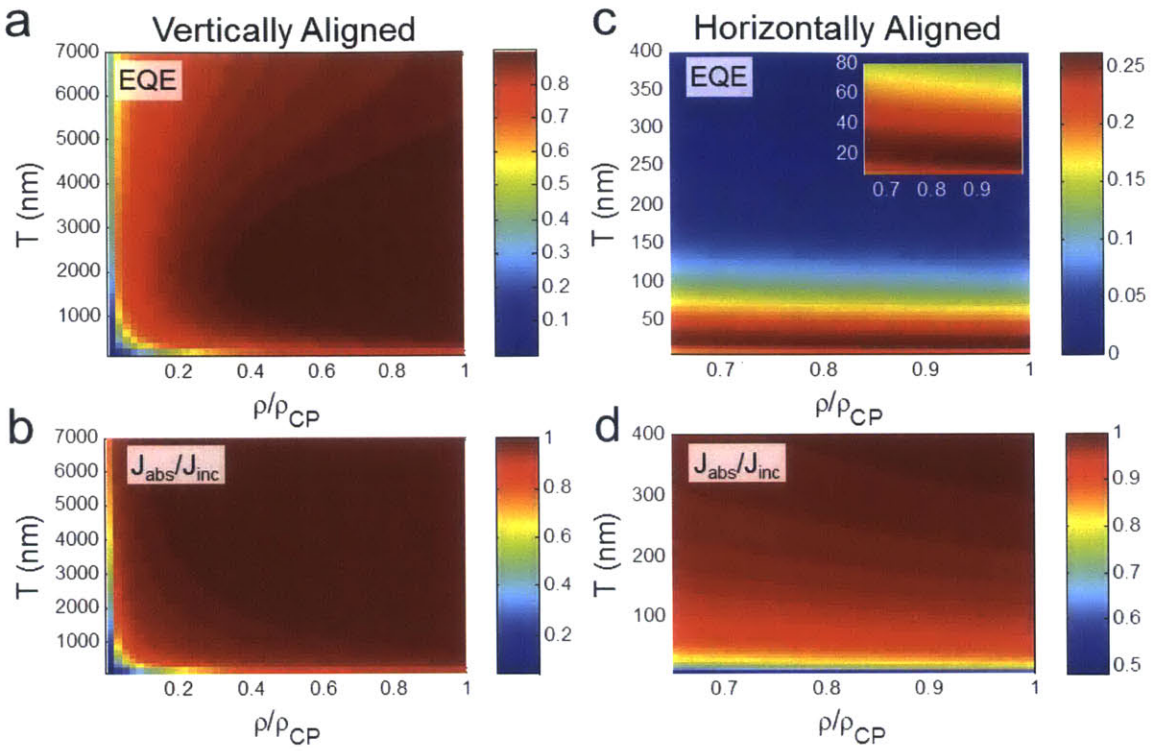


Figure 64. External quantum efficiency (EQE) (a,c) and total fraction of incident light absorbed (b,d) versus nanotube density (ρ/ρ_{CP}) and device thickness (T) for horizontally and vertically aligned films with infinitely long tubes and no impurities. For horizontal films, we look only at the density space above the percolation threshold (see text). R = 1.

We then examined the effect of introducing finite nanotube length to the network, permitting end-quenching of excitons (Figure 65). We observed that not only does EQE across the parametric space drop, as we would expect, but that the lower exciton diffusion length has the effect of shifting the optimal thickness from the absorption-diffusion tradeoff lower, compressing the density-thickness trend in the T axis. At a fixed density, shorter tubes reduce the optimal thickness (Figure 65c). This trend is analogous to the inverse relationship of density and optimal thickness: as exciton transport weakens (in this case due to quenching) relative to light absorption, the balance shifts to increase exciton diffusion by reducing the thickness.

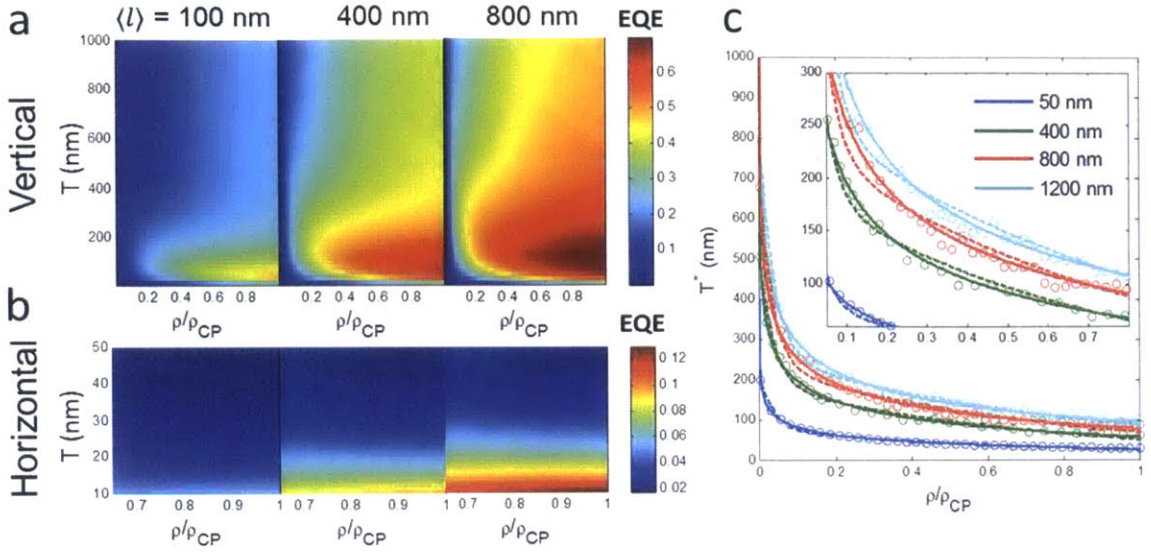


Figure 65. EQE versus thickness T and density for vertically- (a) and horizontally- (b) aligned films at different mean nanotube lengths (100, 400, and 800 nm). See Addendum for single-curve visualization of these calculations showing shift in optimal thickness. (c) Trend of optimal thickness T^* with density at different fixed mean lengths ($\langle l \rangle$ = 50, 400, 800, and 1200 nm) for vertically-aligned films. Circles are numerical result, solid lines are power law fits and dashed lines are biexponential fit (see text). Numerical results exhibit noise because of low sensitivity ($O(10\,nm)$) of EQE to thickness close to the optimum. Inset: zoom showing the 'cusp' where $dT^*/d\rho$ rapidly lowers and the biexponential trend is least accurate.

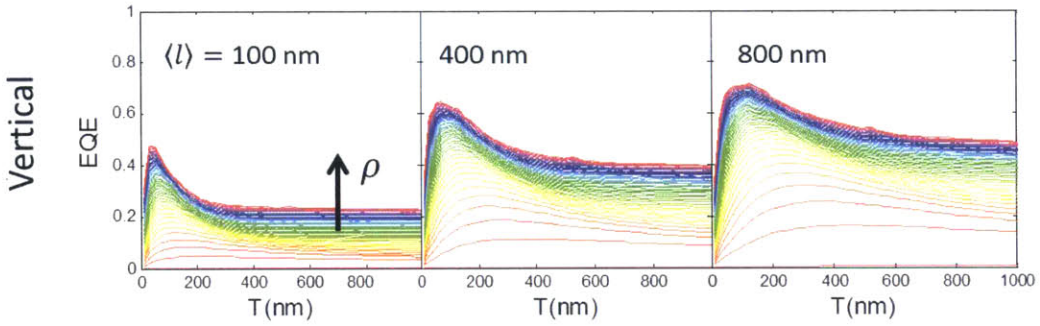


Figure 66. Performance versus thickness curves with increasing density for vertically aligned films. These plots are the same data as the 2D maps presented in Figure 65, but visualized as a series of curves to show the shift in T^*. Black arrow indicates the direction of increasing density (red to blue to purple). The shift of the optimal thickness downward is more visible, particularly the asymptote to $T^* \to \infty$ as $\rho \to 0$.

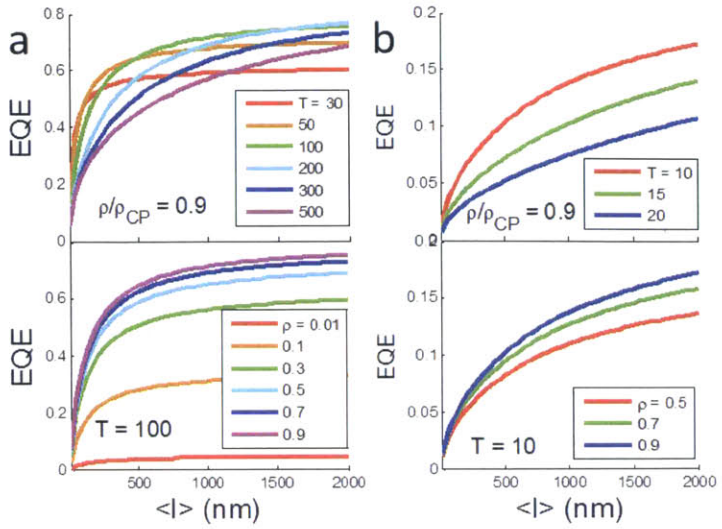


Figure 67. Trends of performance with mean length for vertically (a) and horizontally (b) aligned films. Top charts compare trends at different thicknesses and 0.9 CP density. Bottom charts compare trends at different densities and 100 nm thickness. In all cases we find that performance monotonically increases with nanotube length, as we would intuit given that end-quenching is a purely undesirable phenomenon, but here we see diminishing returns with increasing length – again as we would intuit – leading to an asymptote as $\langle l \rangle \rightarrow \infty$, where performance is limited by the balance of light absorption and the thickness of the film.

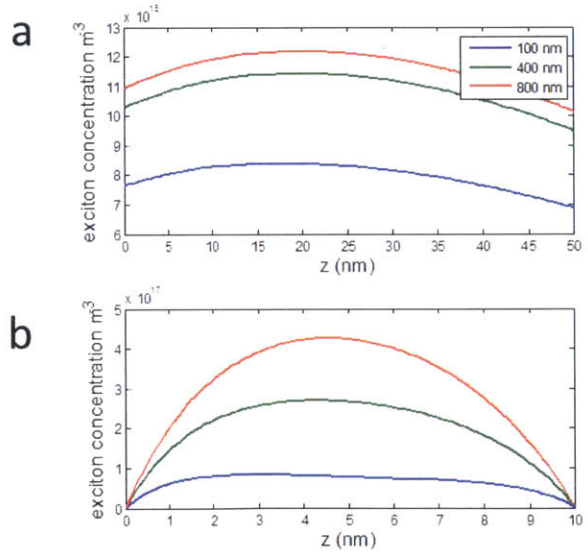


Figure 68. Exciton concentration gradients in vertically (a) and horizontally (b) aligned films with different mean lengths and 30% CP density.

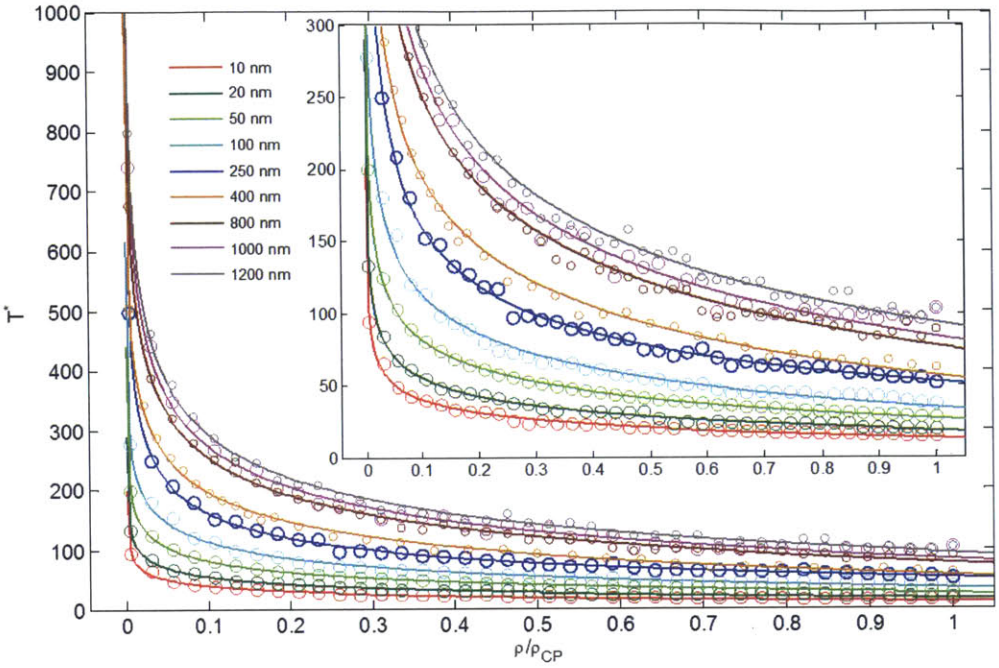


Figure 69. Trend of optimal thickness T^* with density at different mean lengths $\langle l \rangle$, showing more lengths. Open circles are numerical optimization results, and curves are fits of equation (106). Numerical results exhibit noise because of low sensitivity ($O(10\ nm)$) of EQE to thickness close to the optimum.

The trends in T^* with density are captured by a power law relationship (Figure 65c, Figure 69, Table 3):

$$T^* = a\left(\frac{\rho}{\rho_{CP}}\right)^b + c \tag{103}$$

To illustrate the quality of the power law trend, a biexponential trend is compared (Figure 65c inset, Table 3), showing systematic error at the 'cusp' between low-density and high-density regimes where the details of the light absorption-exciton transport balance are most influential.

Table 3. Power and biexponential fits of trends in optimal thickness with density, at different mean lengths. Note that error in the power law trend is entirely within optimization limits, which are $\pm O(10)\ nm$.

	Power Fit $\left(T^* = a\left(\frac{\rho}{\rho_{CP}}\right)^b + c\right)$					Biexponential Fit	
Nanotube Length $\langle l \rangle$	a (nm)	b $\left(\frac{\rho}{\rho_{CP}}\right)^{-1}$	c (nm)	R^2	rms error (nm)	R^2	rms error (nm)
10 nm	56	-0.172	-42.6	0.992	1.5		
20 nm	68	-0.187	-50	0.996	1.5		
50 nm	101	-0.189	-75	0.999	1.0	0.992	3.1
100 nm	148	-0.186	-113	0.992	4.2		
400 nm	230	-0.205	-174	0.998	3.6	0.990	8.6
800 nm	254	-0.229	-176	0.997	6.5	0.987	13.2
1000 nm	248	-0.246	-166	0.995	9.0		
1200 nm	245	-0.255	-155	0.997	6.8	0.989	13.7

While it fits the results well, expression (103) is not entirely satisfying because the coefficients a, b, and c all vary with $\langle l \rangle$, which is the inverse of the quenching site concentration. Furthermore, while the power coefficient b follows a linear trend, the proportionality and intercept coefficients, a and c, have no discernable relationship with $\langle l \rangle$. Insight is gained however by considering the limit of close-packed density, where $\rho/\rho_{CP} \rightarrow 1$, and the optimal thickness is minimal ($T^* \rightarrow T^*_{CP}$):

$$T^*_{CP} = a + c. \tag{104}$$

T^*_{CP} itself depends on quenching site concentration, with lower $\langle l \rangle$ yielding lower optimal thickness due to reduced diffusion length. Motivated by this observation, we examined the numerical data at $\rho/\rho_{CP} = 1$ and found that T^*_{CP} exhibits a strict power law trend with $\langle l \rangle$ (Figure 70 inset):

$$T^*_{CP} = 52\, nm^{0.6} \cdot \langle l \rangle^{0.4}. \tag{105}$$

Armed with this relevant length value that is a function of impurity density, we hypothesized a new relationship that we anticipated would collapse our $\langle l \rangle$ curves by making $T^* \rightarrow T^*/T^*_{CP}$ dimensionless:

$$\begin{aligned} &\frac{T^*}{T_{CP}} = a\left(\frac{\rho}{\rho_{CP}}\right)^b + c \\ &a + c \equiv T_{CP} \rightarrow c = T_{CP} - a \\ &\rightarrow \frac{T^* - T_{CP}}{T_{CP}} = a\left[\left(\frac{\rho}{\rho_{CP}}\right)^b - 1\right]. \end{aligned} \tag{106}$$

Plotting our numerical results as $(T^* - T^*_{CP})/T^*_{CP}$ versus ρ/ρ_{CP} (Figure 70), we find that indeed as expected all of the curves at different $\langle l \rangle$ collapse to a single curve, and that furthermore equation (106) captures the trend (blue line in Figure 70). The values we get for the coefficients, which again are invariant with quenching site concentration, are $a = 2.7 \pm 0.1$ and $b = -0.241 \pm 0.005$.

Equation (106) is an analytic equation describing the optimal device thickness as a function of density and quenching site concentration (equation (105)). This tool enables device-makers to calculate their optimal film thickness, which is one of the more feasible design parameters to control, given their film density and nanotube length, which can be measured. Because the model is deterministic the unitless coefficients, a and b, have some physical interpretation, which we expect relates to the balance of exciton diffusion length and light absorptivity, but have been unable to identify the underlying combination of physical constants that comprise them.

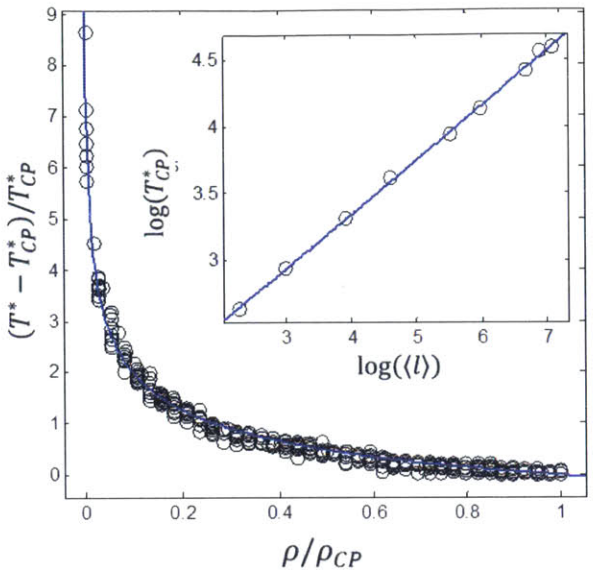


Figure 70. Trend of optimal thickness numerical results with density, plotted relative to optimal thickness at close-packed density (see text). The data (black circles) includes results from devices with $\langle l \rangle = 10, 20, 50, 100, 250, 400, 800, 1000,$ and $1200\,nm$, showing that on this plot they all collapse to a single trend. The blue curve is a fit of equation (106), showing perfect agreement (well within numerical result variation). The inset shows the trend of the optimal thickness at close-packed density (T^*_{CP}) with mean nanotube length (inverse impurity concentration) on a log-log scale, showing that they exhibt a strict power law relationship.

Next we introduced impurity scattering of excitons, providing varying concentrations of a generic local impurity that rapidly quenches excitons, choosing a representative $\langle l \rangle = 400\,nm$ (Figure 71). The same trend as end quenching emerged, with higher impurity concentrations lowering the diffusion length and both decreasing the mean performance and compressing the trend it in the T dimension.

With these results we were able to consider the differences among to-date experimental SWNT SC observations. More closely examining the density trend, we see that at low densities (< 10 %CP, where all devices have been) the increase in performance with density is rapid (Figure 64, Figure 72);[xxxii] higher density increases absorption without decreasing the diffusion length (neglecting coherent delocalization of excitons between tubes, see 'Model Limitations' section above), leading to significant performance gains at low density. Comparison to currently published SWNT photovoltaics[115-118] suggests 1) that experimentally observed EQE variations may be due to relatively low densities, and 2) that large gains can be realized in performance with higher density films. The caveat is a need to be reasonably close ($0.5T^* < T < 3T^*$) to the high-performance thickness regime. For example in primarily in-plane ('horizontal') films where the z diffusion length is short, density gains are negligible if the thickness is already limiting generated exciton collection significantly. With reasonable length and impurity terms, this is consistent with T^* below 15 nm for horizontally aligned films, explaining the improvements in performance by reducing thickness in reference [118]. In general, the existence of a thickness optimum has been experimentally observed in both planar[115] and bulk[121] heterojunctions, albeit at densities below our cutoff of validity.

[xxxii] To observe this effect in horizontal films, we must go below the percolation threshold, which is not shown here. To an order of magnitude our predictions in that regime match up with experimental devices, but because our approximation does not hold in that regime, we do not claim to reproduce that device performance.

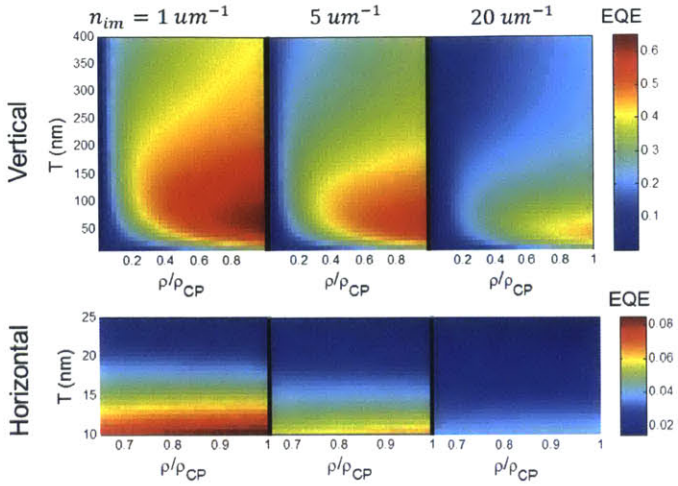


Figure 71. Efficiency versus thickness and density for vertically and horizontally aligned films with $\langle l \rangle = 400\ nm$ at different local quenching impurity concentrations (1, 5, and 20 per micron).

Understanding that thickness and density play critical roles, we turned to examining the alignment angle choice. At a given density we look at the thickness-angle space (Figure 72). Below 10 %CP density, two local thickness-angle optima emerge, one at the extreme of vertical alignment, and the other at an intermediary angle. As thickness increases, the optimal alignment angle shifts gradually towards more vertical, compensating for the increased mean exciton-electrode separation with a higher z-diffusion coefficient (larger longitudinal intra-SWNT diffusion component, lower orthogonal inter-SWNT); also as thickness rises, the ratio of vertical performance to intermediate performance shifts towards vertical. As density rises, the intermediate peak more rapidly converges to the vertical with increasing T, and the ratio becomes more favorable across all T to vertical alignment. As a result, at low density (<3 %CP) the intermediate peak is the global optimum of performance, while at high density the vertical is superior. In absolute efficiency, we again observe the strong impact of density (note scale bars in Figure 72), particularly at lower densities. Again we can consider experimental results, and note that as existing SWNT SC devices have been made with primarily nanotubes in the xy plane at low concentrations (<1 % CP density), the observed EQEs from 0.1% to 1% are to an order of magnitude consistent with our predictions.[115-118] Our results suggest that developing vertically aligned films above 10% of the close-packed density would yield substantial EQE improvements. Films with high in-plane components, such as most to-date devices, are the worst configurations of nanotubes.

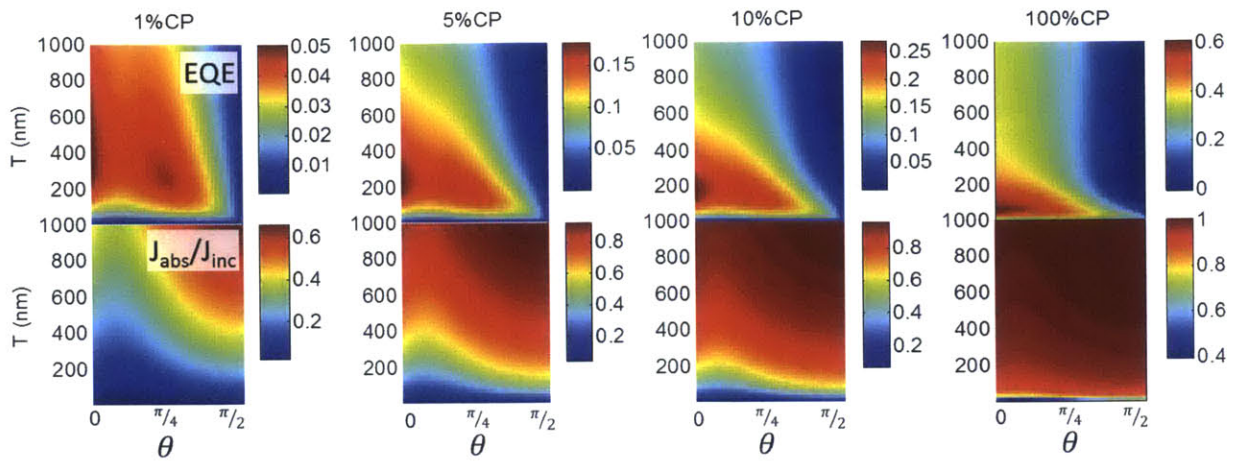


Figure 72. Performance and fraction of light absorbed versus device thickness T and aligned film nanotube orientation θ at 1%, 5%, 10%, and 100% close-packed density. $\theta = 0$ is a vertical film, $\theta = \pi/2$ is a horizontal film. $\langle l \rangle = 400\ nm$, $n_{im} = 2e6$ m^{-1}, $R = 1$.

The optimal orientation results can be interpreted again by understanding the balance of light absorption and exciton diffusion constraints. Vertical alignment provides more rapid diffusion via the longitudinal SWNT axis, but absorbance is lower due to the depolarization effect. Horizontal alignment comes at the cost of much slower exciton-transfer-dependent diffusion, but provides more rapid light absorption, which to some extent mitigates the diffusion constraints by generating more excitons closer to the top electrode. Longitudinal diffusivity, $O(10^{-2}\ m^2/s)$, is so much greater than orthogonal diffusivity, $O(10^{-7} m^2/s)$, that in all but nearly-horizontal films it dominates the net diffusion coefficient. Because of the rapid exciton transport longitudinally between electrodes, absorptivity increases from somewhat non-vertical alignment at lower densities yield efficiency gains. Those gains are highest at low T where absorption is weakest and diminish (shifting the intermediate peak more vertical) with increasing thickness. At sufficiently high densities the absorption gains are unnecessary, leading to a global vertical-growth optimum. In general, the aligned film limits light absorption off of the parallel axis (Figure 62e), reducing the efficacy of alignment.

Modelling Isotropic Films

The second case we consider is an isotropic film, where SWNT orientation in any direction is equally likely,

(107) $$p(\theta) = \frac{1}{\pi}, \qquad \theta \in [0, \pi]$$

(108) $$p(\phi) = \frac{1}{2\pi}, \qquad \phi \in [0, 2\pi].$$

This system differs from current 'isotropic' experimental systems in that those systems in practice restrict θ, making them more of a 'sandwich' architecture and more similar to off-horizontal aligned films. Even when caste with surfactant we expect some level of bundling to be typical in an isotropic network, and take a baseline for comparison of $b_c = 0.1$, $M_b = 4$.

Our treatment of light absorption is analogous to the aligned case application, but the integral over $p(\hat{\boldsymbol{l}})$ becomes less trivial,

$$\sigma_l(\boldsymbol{\epsilon}, \omega) = \frac{1}{\pi}\frac{1}{2\pi}\frac{1}{6}\sigma_{l,meas}(\omega)\int_0^{2\pi}\int_0^{\pi} 5{\varepsilon_\parallel}^2 + \left(1 - |\varepsilon_\parallel|\right)^2 d\theta d\phi. \tag{109}$$

(164) is solvent, and yields (see Appendix J)

$$\sigma_l(\boldsymbol{\epsilon}, \omega) = \left(\frac{5}{12} - \frac{4}{3\pi^2}\right)\sigma_{l,meas}(\omega). \tag{110}$$

Note that we no longer have any ϕ_ϵ dependence to absorption, as we would intuit from an isotropic distribution. Our light field becomes

$$J_F(\omega|z) = J_0(\omega)\cdot\exp\left[-\left(\frac{5}{12} - \frac{4}{3\pi^2}\right)\rho_{\langle l\rangle}\sigma_{l,meas}(\omega)\cdot z\right]. \tag{111}$$

$$J_R(\omega|z) = R\cdot J_0(\omega)\cdot\exp\left[-\left(\frac{5}{12} - \frac{4}{3\pi^2}\right)\rho_{\langle l\rangle}\sigma_{l,meas}(\omega)\cdot (2T - z)\right]. \tag{112}$$

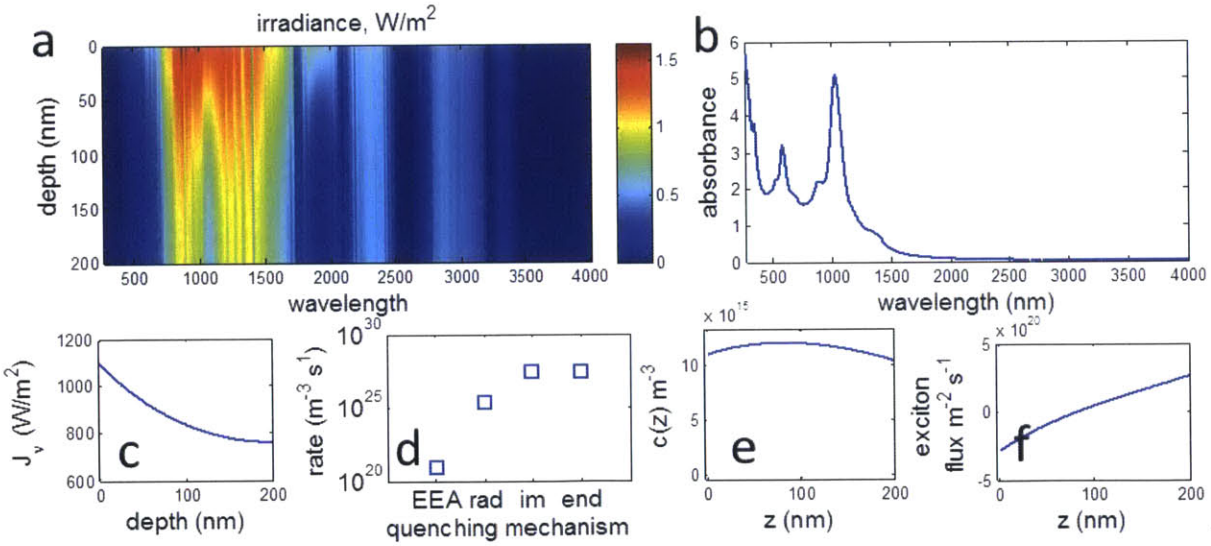


Figure 73. Light field attenuation and steady-state exciton transport for an isotropic monochiral (6,5) film with $T = 200\ nm$, $R = 1$, $\rho_{\langle l\rangle} = 0.1\rho_{\langle l\rangle}^{CP}$, $b_c = 0.1$, and $M_b = 4$. a) irradiance gradient as a function of wavelength. Note stronger attenuation on resonance. b) total film absorbance as a function of wavelength. c) total irradiance as a function of depth in the field. d) mean quenching rates by each mechanism. Note log scale. e) exciton concentration gradient. f) exciton flux through the depth of the film.

Moving on to the exciton concentration gradient, our expressions are the same as in the aligned film case, with the end concentration being $c_{end} = 2\rho$. Integrating the monochiral exciton balance (240) over orientation distribution $p(\theta)$ we get the diffusion coefficient,

$$D_{tot} = \frac{1}{\pi}\int_0^{\pi} D_l(\cos\theta)^2 + (1 - |\cos\theta|)^2\left(\gamma_I D_{EH,I} + b_c D_{EH,b}\right) d\theta \tag{113}$$

which resolves to

$$D_{tot} = \frac{1}{2\pi}\left(\pi D_l + (3\pi - 8)\left(\gamma_I D_{EH,I} + b_c D_{EH,b}\right)\right). \tag{114}$$

D_l, $D_{EH,I}$, $D_{EH,b}$, and γ_I follow as their definitions, with the mean squared coalignment becoming

$$\langle\lambda^2\rangle = \frac{1}{4}\frac{1}{\pi^4}\cdot\frac{9\pi^4}{64} = \frac{9}{256}. \tag{115}$$

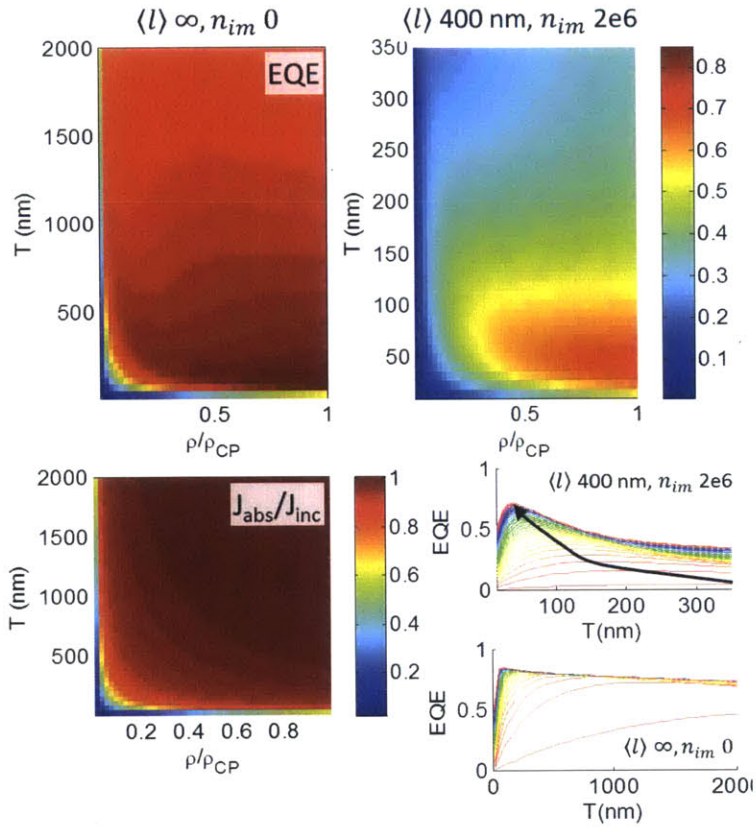


Figure 74. Density-thickness optimization of isotropic films, with = 1, $b_c = 0.1$, and $M_b = 4$. a) performance for 'idealized' film. b) fraction of light absorbed. c) performance for film with $\langle l\rangle = 400\ nm$, 2 impurities per micron. d) performance versus T plots at each density sampled in (a) and (c), the arrow indicates path of T^* in (c).

Beyond examining particular cases (Figure 73), we can proceed to explore the parameter space, particularly the thickness-density relation (Figure 74). Again we see a strong optimum T^* emerge as a function of density. As density rises, the enhanced absorbance makes lower thicknesses with shorter generated exciton-electrode separations preferable. Again we see a large improvement in performance with density at low density; higher density improves absorbance, increases inter-SWNT hopping, and reduces (at fixed concentration) exciton-exciton annihilation and end-quenching. In general we note similar performance trends compared to aligned films with modest vertical components.

VI. SINGLE-PARAMETER DETERMINATION OF CARBON NANOTUBE PHOTOVOLTAIC EFFICIENCY

Summary of Findings

In the previous Chapter we derived a deterministic model of CNT planar heterojunction photovoltaic steady-state operation directly from single- and paired-nanotube photophysics. We did so with a novel approach, treating local, *i.e.* tube-specific, properties as random variables to describe the bulk film as distributions of those properties. We optimized that model to identify design principles for increasing device efficiency by two orders of magnitude over current generation devices, such as maximizing density and the vertical alignment component. In particular, a given film morphology has an optimal thickness regime outside of which efficiency falls off. The resulting system of non-linear non-homogenous integro-differential equations however obliged numerical evaluation, obscuring how intrinsic film properties such as exciton diffusion length generated our predicted trends. While the trends guide device design and provide some physical intuition, we could not connect them explicitly to underlying physical constants.

In this Chapter we present an analytic solution to the model that reveals those fundamental connections, explicitly balancing exciton generation, reaction, and diffusion. All relevant intrinsic physical quantities are reduced to a single group, α, that dictates maximum possible efficiency; a master curve of optimal film thickness, T^*, as a function of α is produced; and we find that to within 99.5% of the maximum possible EQE the optimal thickness is approximately the exciton diffusion length.

We find that the characteristic spatial length scale describing SWNT PV efficiency is the exciton diffusion length, $L_D = \sqrt{D_z/k_1 n_1}$, where D_z is the exciton diffusivity in the charge-collecting axis and $k_1 n_1$ is the first-order quenching rate constant; *i.e.* the balance of exciton reaction and diffusion. Non-dimensionalizing device thickness as $\eta \equiv T/L_D$, the efficiency-maximizing thickness η^* depends on a single grouping of parameters, $\alpha = \phi \cdot L_D$, where ϕ is the SWNT number density normalized to close-packed. That generates a master $\eta^*(\alpha)$ curve (Figure 77D) that any device can be placed on, empowering device-makers to know their optimal thickness by measuring their material's α. We further find that close to the optimum, external quantum efficiency (EQE) is only weakly variant with η, yielding the rule of thumb $0.6 \leq \eta^* \leq 1.4$, or equivalently $T^* \cong L_D$, the thickness equals the diffusion length, which provides an EQE within 0.5% of the maximum. Finally, our solution is enabled by a new method of approximating solar absorption that is applicable to any film with bell-shaped, spatially-isotropic absorption peaks (*e.g.* quantum dots, polymers, dyes). The light absorption (carrier generation) gradient is typically the most difficult component of photovoltaic performance to solve due to the convolution of incident intensity and

absorption spectrum over frequency and polarization. Our treatment as a whole applies to any exciton-transport-limited film with those absorption properties, with an analytic *vs.* numerical error of less than 1% of the resulting EQE.

Exciton Transport Model

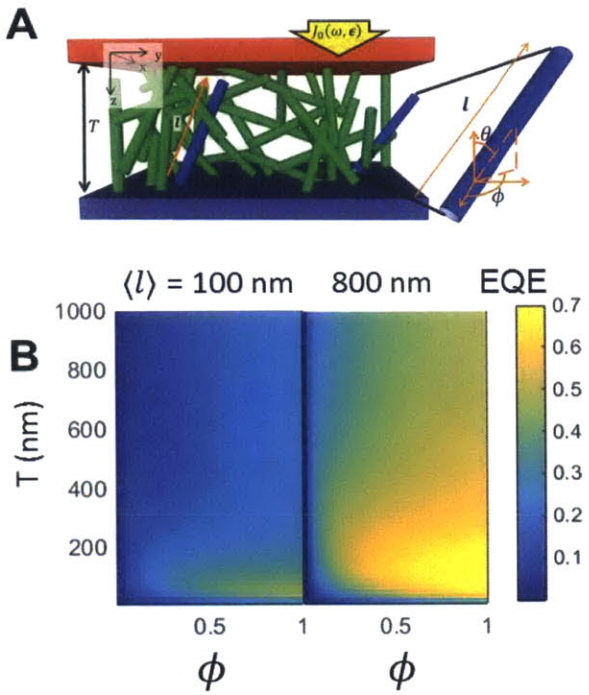


Figure 75. (A) Cartoon of a SWNT solar cell of thickness T, with diameters exaggerated. (B) Numerical solution to the steady-state behavior of the vertically-aligned film, with the External Quantum Efficiency (EQE) plotted (colormap) versus the film thickness T and the density ϕ as a fraction of the close-packed density, at two different mean nanotube lengths $\langle l \rangle$. Longer tubes exhibit less end-quenching, increasing EQE and allowing the film to be thicker to capture more light while still collecting excitons at the electrodes.

In Chapter V, we considered a monochiral network of single-walled nanotubes (SWNT) sandwiched between two Type-II exciton dissociation electrode semi-infinite plates separated by a distance T (Figure 75).[198] The Cartesian z axis is normal to the plates, with the incident solar photon flux $J_0(\omega)$ normally incident at $z = 0$. A given nanotube in the film has orientation $\hat{\boldsymbol{l}} \equiv (\theta, \phi)$, where θ is the angle with the z axis. When the network density is high enough (above the percolation threshold), the film can then be meaningfully described by the distribution $p(\theta, \phi)$, with $p(\theta, \phi) = \delta(\theta - \theta')\delta(\phi)$ for example representing an aligned film at angle θ'.[198]

Exciton diffusion occurs via three channels. Longitudinal transport is along the nanotube length, with diffusion coefficient D_l. Exciton energy transfer occurs between neighboring tubes either in bundles or at misaligned interconnects, in both cases being orthogonal to the longitudinal axis of the originating nanotube. Those effective diffusion coefficients are $D_{EET,b}$, dependent on the mean bundle size, and $D_{EET,I}$, respectively. D_l is approximately 5 orders of magnitude larger than the latter two. The net exciton diffusivity in the z axis is then achieved by integrating these effects over the film orientation distribution[198]

$$D_z \equiv \iint p(\theta,\phi)\left[(\cos\theta)^2 D_l + (1-\cos\theta)^2\left(\gamma_I D_{EET,I} + b_c D_{EET,b}\right)\right]d\theta d\phi = \tag{116}$$

where b_c is the bundle fraction in the film (typically near unity) and γ_I is a sparsity coefficient capturing the density of interconnects. The exciton population balance can then be constructed by accounting for fluorescent emission, non-radiative decay, Auger recombination, and tube-end quenching,

$$\frac{dc}{dt} = N(z) + D_z\frac{d^2c}{dz^2} - k_\Gamma c - k_1' c - \frac{k_{EEA}}{\rho_{\langle l\rangle}}c^2 - \frac{k_{end}}{\rho_{\langle l\rangle}}\frac{2}{\langle l\rangle}\cdot c. \tag{117}$$

where $c(z)$ is the exciton concentration, $N(z)$ is the exciton generation rate from light absorption, k_{EEA} is the auger recombination rate constant, k_Γ is the radiative decay rate constant, k_1' is the impurity quenching rate constant proportional to the concentration of impurities, k_{end} is the end-quenching rate constant, and $\langle l\rangle$ is the mean nanotube length. This expression holds only for films of densities above the percolation threshold and with a uniform distribution of nanotube end-sites. For a more detailed derivation from single-SWNT behavior, including broader case analysis, see reference [198].

At steady state the exciton transport expression (75) is therefore described by a nonlinear, nonhomogenous ordinary differential equation (ODE) of simplified form,[198]

$$D_z\frac{d^2c}{dz^2} - k_{1'}c - k_{2'}c^2 = -N(z) \tag{118}$$

Note that all parameters are extracted from individual/aggregate SWNT photophysics and the distribution of nanotube orientations in the film. The ODE is subject to Robin boundary conditions describing Type II exciton dissociation at each interface, $z = 0,T$ for a film of thickness T,

$$\pm D_z\left.\frac{dc}{dz}\right|_{z=0,T} = k_d c|_{z=0,T} \tag{119}$$

where k_d, the dissociation rate constant, we take to infinity in the rapid-dissociation limit, checking that the flux converges.

Analytic Solution to Photoadsorption of the Solar Spectrum in SWNT films

The first obstacle to analytic evaluation of (118) is *N(z)*, the exciton generation rate, which is the photon absorption rate. A valid form of *N(z)* that enables facile solution to the exciton transport ODE is a common inhibitor to analytic or even deterministic solutions to exciton-transport-limited nanomaterial photovoltaic models. Herein we report an analytic expression with material-specific constants that captures that light absorption behavior.

N(z) is the convolution over frequency and linear polarization space of the absorption cross-section per nanotube $\sigma_l(\boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega)$, the number density $\rho_{\langle l\rangle}$, and the solar flux at depth z $J_\nu(\omega, \boldsymbol{\epsilon}|z)$,

(120) $$N(\omega, \boldsymbol{\epsilon}|z) = \rho_{\langle l \rangle} \cdot \sigma_l(\boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega) \cdot J_v(\omega, \boldsymbol{\epsilon}|z).$$

$N(\omega, \boldsymbol{\epsilon}|z)$ carries a frequency ω and polarization $\boldsymbol{\epsilon}$ dependence, the latter of which demands a SWNT orientation $\hat{\boldsymbol{l}}$ dependence. *N(z)* is then a quadruple integral over the nanotube orientation distribution $p(\theta, \phi)$, linear in-plane polarization ϵ (we assume normally-incident light), and light frequency ω,

(121) $$N(z) = \rho_{\langle l \rangle} \int_0^{\infty} \int_0^{2\pi} J_v(\omega, \epsilon|z) \cdot \int_0^{2\pi} \int_0^{\pi} \sigma_l(\epsilon, [\theta, \phi], \omega) p(\theta, \phi) d\theta \, d\phi \, d\epsilon \, d\omega$$

where $\rho_{\langle l \rangle}$ is the nanotube density (length of nanotube per volume of film), J_v is the photon flux, and σ_l is the single-SWNT absorption cross-section per length of nanotube.

To proceed we focus on devices with isotropic in-plane light absorption and no reflection of light off the back electrode. This parametric space is the most relevant to technological interests; in the former case it includes isotropic and vertically-aligned films, which we previously showed present dominant efficiency relative to films with anisotropic in-plane alignment; in the latter case a transparent back electrode takes advantage of SWNTs near-infrared (nIR) absorption by allowing the solar cell to complement existing visible PVs or coat building materials. Those approximations make the polarization and nanotube orientation integrals trivial. We then solve the differential equation describing the light field attenuation as it passes through the film,

(122) $$\frac{dJ_v(\omega|z)}{dz} = -N(\omega|z) \Rightarrow J_v(\omega|z) = J_0(\omega) \cdot \exp\left[-\rho_{\langle l \rangle} \cdot \sigma_l(\omega) \cdot z\right]$$

where $J_0(\omega)$ is the incident solar flux. That produces the final integral over frequency space

(123) $$N(z) = \rho_{\langle l \rangle} \int_0^{\infty} J_0(\omega) \sigma_l(\omega) \exp\left[-\rho_{\langle l \rangle} \cdot \sigma_l(\omega) \cdot z\right] d\omega.$$

To solve that integral, we recognize several simplifications. First we can closely approximate the nanotube absorption spectrum as a series of Gaussian functions for each absorption peak, with magnitude S_a,

(124) $$\sigma_l(\omega) = \sum_a S_a \cdot \exp\left[-\frac{(\omega - \mu_a)^2}{2\sigma_a^2}\right]$$

where μ_a is the energy of the transition and σ_a is the peak width ($FWHM = 2\sqrt{2 \ln 2}\, \sigma_a$). Next, we find that because the SWNT absorption peaks lie on only the red side of the solar spectrum peak, the spectrum can be sufficiently approximated as a Gaussian as well,

$$J_0(\omega) \cong J_0 \cdot \exp\left[-\frac{\left(\omega-\mu_J\right)^2}{2\sigma_J^2}\right]. \tag{125}$$

Those simplifications make $N(z)$ an insolvent convolution of a Gaussian and an exponential decay of a Gaussian (see Appendix K). We can observe however that while the Exponential-Gaussian term is in general not well captured by the first-order Taylor expansion, its product with the Gaussian term *is* because either the function extrema are close *or* the Gaussian-Exponential reduces to unity (See Appendix K). That observation simplifies the integral, allowing us to solve it as

$$N(z) = \rho_{\langle l\rangle} J_0 \sum_{a=1}^{2} S_a \gamma_a \sqrt{2\pi} \cdot e^{-\rho_{\langle l\rangle} S_a z} \tag{126}$$

where γ_a can be understood as the area of overlap of the solar flux and absorption peak a,

$$\gamma_a \equiv \sigma_{Ja} e^{-\frac{(\mu_J-\mu_a)^2}{2(\sigma_J^2+\sigma_a^2)}}, \qquad \sigma_{Ja} \equiv \sqrt{\frac{\sigma_J^2\sigma_a^2}{\sigma_J^2+\sigma_a^2}} \tag{127}$$

The sum must include two peaks for each SWNT chirality, corresponding to the E_{1u} and E_{2u} transitions. For a solid film of (6,5) SWNT,[116] the empirical constants are $S_1 = 1.64 \times 10^{-10}\ m^2/m$, $\gamma_1 = 293\ nm$, $S_2 = 0.13 \times 10^{-10}\ m^2/m$, and $\gamma_2 = 74\ nm$.

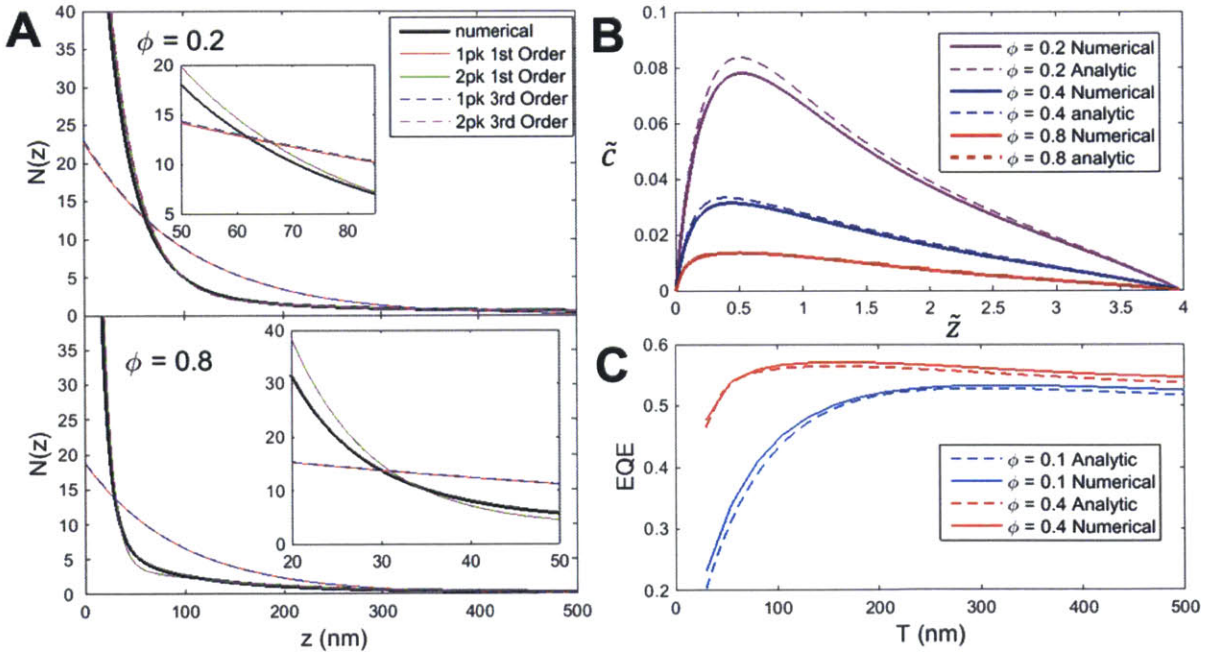


Figure 76. Comparison of analytic and numerical solutions. (A) Solution to the photon absorption rate, comparing the numerical result (black) to the analytic solutions with one or two absorption peaks and first or third order Gaussian-exponential Taylor expansion. Solutions for 20% and 80% close-packed density are shown, with insets zoomed for clarity. While it is necessary to include absorption from both SWNT electronic transitions, we can see that a third-order Taylor expansion of the frequency integrand is unnecessary. (B) Nondimensional exciton

concentration profiles comparing the numerical and analytic solutions at different densities for a film of thickness $\tilde{z} = 4$. (C) Film EQE as a function of film thickness at different densities.

This biexponential analytic expression matches the numerical solution (Figure 76A). The curve can be conceptually understood as light absorption on two length-scales, one for each portion of the frequency spectrum absorbed. Because there are two absorption peaks with different absorption cross-sections, with the near-IR absorption being stronger than visible absorption, light at nIR frequencies is, relative to initial intensity, attenuated rapidly while visible light is absorbed more gradually, providing a long tail to our generation rate. That effect is emphasized by the greater incident solar intensity at the visible *vs.* nIR frequencies. We can also see validation that our first-order Taylor expansion of the Gaussian-exponential is sufficient, with essentially no benefit to a third-order expansion.

Solution to the Exciton Transport Model

With an analytic expression for *N(z)* we are able to non-dimensionalize the ODE (118). The problem symmetry yields the characteristic exciton concentration and Cartesian length,

$$c_0 \equiv \frac{J_0 \cdot \rho_{\langle l \rangle} S_1 \cdot \gamma_1}{k_1 n_1} \tag{128}$$

$$z_0 \equiv \sqrt{\frac{D_z}{k_1 n_1}} = L_D \tag{129}$$

where the near-IR peak (S_1, γ_1) is the dominant absorption mode and $k_1 n_1$ is the product of the single-nanotube first-order quenching rate constant and the concentration of impurities including nanotube ends. Multiple impurities with different rate constants can be treated with a weighted sum. The characteristic variables c_0 and z_0 reflect the physical trade-offs of the system. The nondimensional concentration is a ratio of the exciton generation rate to the dominant exciton quenching rate, which compete to increase and decrease, respectively, the exciton concentration. The nondimensional Cartesian coordinate normalizes the film thickness as $\eta \equiv T/L_D$, which balances the film thickness against the exciton diffusion length. A thicker film or lower diffusion length reduces the number of excitons that reach the electrodes.

Substituting nondimensional concentration $\tilde{c} \equiv c/c_0$ and Cartesian dimension $\tilde{z} \equiv z/z_0$ into the exciton balance, we are left with

$$\frac{d^2\tilde{c}}{d\tilde{z}^2} - \left(\frac{k_\Gamma}{k_1 n_1} + 1\right)\tilde{c} - \frac{k_2\left(\rho_{\langle l\rangle} J_0 S_1 \gamma_{11}\right)}{(k_1 n_1)^2}\tilde{c}^2$$

$$= -\sqrt{2\pi}\sum_{a=1}^{2}\frac{S_a\gamma_a}{S_1\gamma_1}\cdot e^{-\rho_{\langle l\rangle}S_a\sqrt{\frac{D_z}{k_1 n_1}}\tilde{z}} \tag{130}$$

$$= -\sqrt{2\pi}\left(e^{-\rho_{\langle l\rangle}S_1 L_D\tilde{z}} + \frac{S_2\gamma_2}{S_1\gamma_1}\cdot e^{-\rho_{\langle l\rangle}S_2 L_D\tilde{z}}\right).$$

We then compare the magnitude of terms, finding that those much less than unity are negligible to yield the far simplified exciton balance

$$\frac{d^2\tilde{c}}{d\tilde{z}^2} - \tilde{c} = -\sqrt{2\pi}\left(e^{-\tau_1\alpha\tilde{z}} + \frac{S_2\gamma_2}{S_1\gamma_1}\cdot e^{-\tau_2\alpha\tilde{z}}\right) \tag{131}$$

$$\left.\frac{d\tilde{c}}{d\tilde{z}}\right|_{\tilde{z}=0} = \kappa\tilde{c}(0), \qquad \left.\frac{d\tilde{c}}{d\tilde{z}}\right|_{\tilde{z}=\eta} = -\kappa\tilde{c}(\eta) \tag{132}$$

$$\alpha \equiv \phi\cdot L_D, \qquad \phi \equiv \frac{\rho_{\langle l\rangle}}{\rho_{\langle l\rangle}^{CP}}, \qquad \eta \equiv \frac{T}{L_D}, \qquad \kappa \equiv \frac{k_d}{\sqrt{D_z k_1 n_1}}, \qquad \tau_a \equiv S_a\rho_{\langle l\rangle}^{CP} \tag{133}$$

Solving with the method of undetermined coefficients and taking the rapid-dissociation limit, we find the concentration profile is a balance of four competing exponentials,

$$\tilde{c}(\tilde{z}) = \beta_1 e^{\tilde{z}} + \beta_2 e^{-\tilde{z}} - \frac{1}{(\tau_1\alpha)^2 - 1}e^{-\tau_1\alpha\tilde{z}} - \sqrt{2\pi}\frac{S_2\gamma_2}{S_1\gamma_1}\frac{1}{(\tau_2\alpha)^2 - 1}e^{-\tau_2\alpha\tilde{z}} \tag{134}$$

$$\beta_1 = -\frac{1}{e^{\eta} - e^{-\eta}}\left(\frac{e^{-\eta} - e^{-\tau_1\alpha\eta}}{(\tau_1\alpha)^2 - 1} + \sqrt{2\pi}\frac{S_2\gamma_2}{S_1\gamma_1}\frac{e^{-\eta} - e^{-\tau_2\alpha\eta}}{(\tau_2\alpha)^2 - 1}\right) \tag{135}$$

$$\beta_2 = +\frac{1}{e^{\eta} - e^{-\eta}}\left(\frac{e^{\eta} - e^{-\tau_1\alpha\eta}}{(\tau_1\alpha)^2 - 1} + \sqrt{2\pi}\frac{S_2\gamma_2}{S_1\gamma_1}\frac{e^{\eta} - e^{-\tau_2\alpha\eta}}{(\tau_2\alpha)^2 - 1}\right). \tag{136}$$

The result we again check against the numerical solution (Figure 76B), finding strong agreement to validate our approximations. The bi-exponential light field decay due to a weak and strong absorption peak manifests in a pronounced tail in exciton concentration towards the back electrode; nIR light (the stronger peak, τ_1) is rapidly collected, while visible light (the weaker peak, τ_2) is more weakly absorbed to generate excitons deeper in the film.

In the exciton-transport-limited approximation, we can estimate the EQE from the boundary solutions

$$EQE(\eta) \propto D_z\left(\left.\frac{dc}{dz}\right|_{z=0} - \left.\frac{dc}{dz}\right|_{z=T}\right) = J_0\gamma_1\tau_1\alpha\left(\left.\frac{d\tilde{c}}{d\tilde{z}}\right|_{\tilde{z}=0} - \left.\frac{d\tilde{c}}{d\tilde{z}}\right|_{\tilde{z}=\eta}\right). \tag{137}$$

We can optimize (137), setting its derivative equal to zero to find the maximum as (Figure 76C),

$$\frac{dEQE}{d\eta} = 0 = F_1\Big((1-p_1)e^{-\eta^*(\tau_1\alpha)} + (2p_1)e^{-\eta^*(\tau_1\alpha-1)} + (2m_1p_1-2)e^{-\eta^*(\tau_1\alpha-2)} + (-2m_1)e^{-\eta^*(\tau_1\alpha-3)} + (1+m_1)e^{-\eta^*(\tau_1\alpha-4)}\Big) + F_2\Big((1-p_2)e^{-\eta^*(\tau_2\alpha)} + (2p_2)e^{-\eta^*(\tau_2\alpha-1)} + (2m_2p_2-2)e^{-\eta^*(\tau_2\alpha-2)} + (-2m_2)e^{-\eta^*(\tau_2\alpha-3)} + (1+m_2)e^{-\eta^*(\tau_2\alpha-4)}\Big) + (2m_1p_1F_1 + 2m_2p_2F_2)\left(2e^{2\eta^*} - e^{\eta^*} - e^{3\eta^*}\right) \tag{138}$$

$$m_a \equiv \frac{1}{\tau_a\alpha - 1}, \qquad p_a \equiv \frac{1}{\tau_a\alpha + 1}, \qquad F_a \equiv \sqrt{2\pi}\frac{S_a\gamma_a}{S_1\gamma_1}. \tag{139}$$

The most surprising result is that there is only one film parameter in this expression, α, the product of film density and exciton diffusion length. The optimal thickness therefore depends on only this single variable, and all possible films collapse to a single η-α space, as in Figure 77A.

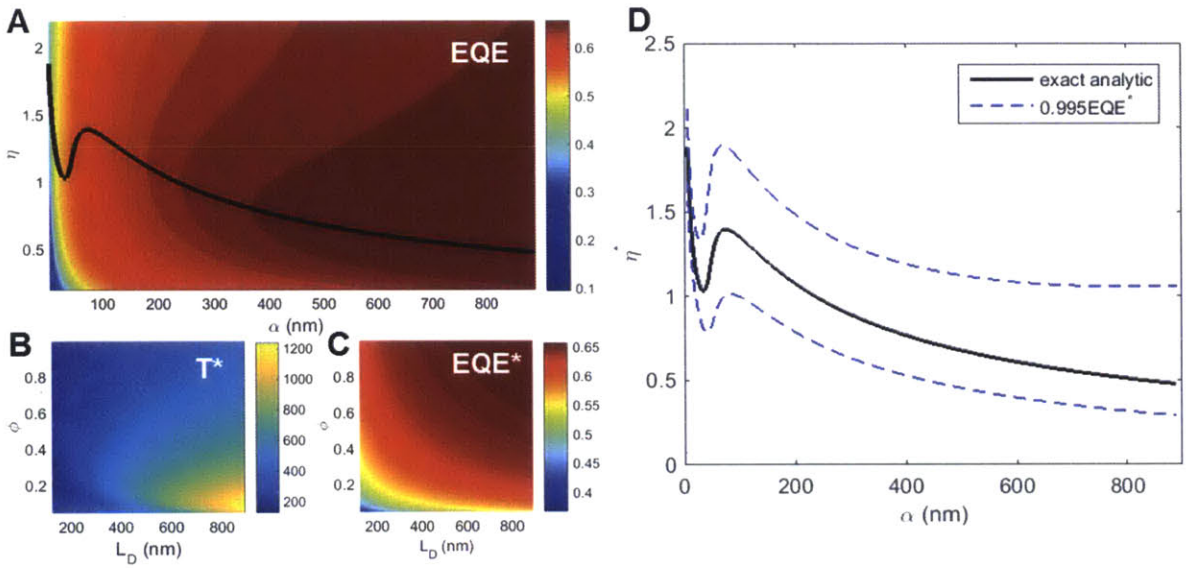


Figure 77. (A) Photovoltaic external quantum efficiency (EQE) as a function of film thickness η and parameter grouping α. Note that a nonlinear color bar was used for visual clarity. At a given α there is an optimal thickness η^* that maximizes the EQE, indicated by the black line overlay. (B) the EQE-maximizing thickness $T^* = \eta^* L_D$ versus density ϕ and exciton diffusion length L_D. (C) Maximum achievable EQE versus density ϕ and exciton diffusion length L_D, showing the monotonic increase in both. In both (B) and (C), α is the product of the axes. (D) Master curve (black) of optimal device thickness, non-dimensionalized with the exciton diffusion length $\eta^* \equiv T^*/L_D$, versus α, the product of diffusion length and normalized density and the sole system parameter. Measuring a SWNT film's α dictates the optimal device thickness for that material. Blue dashed curves indicate the bounds of higher and lower thickness that yield an EQE above 99.5% of the maximum. The broad EQE tolerance suggests that in lieu of measuring α, choosing η between 0.6 and 1.4 *i.e.* $T^* \cong L_D$ will provide essentially optimal performance.

Discussion

Examining the efficiency variation with thickness and α (Figure 77A), we can observe that EQE monotonically increases with α; higher α materials, which have higher densities and exciton diffusion lengths, are unambiguously better (Figure 77C). In particular there is a sharp drop in efficiency for α below 10 nm, indicating that a mean exciton diffusion length of over 100 nm for a 10% close-packed film or 10 nm for a 100% close-packed film is a crucial material property that must be achieved to advance SWNT Solar Cell technology. Additionally, at a given α there is an optimal η^*, indicated by the black line in Figure 77A. The optimum represents a balance between a thicker film collecting more light but increasing the distance that excitons must travel to reach the electrodes. At higher densities (ϕ) more light is collected per unit thickness, increasing the maximum EQE and reducing the optimal thickness required to achieve it. At longer diffusion lengths (L_D) excitons are collected more efficiently, allowing the film to be thicker to achieve the same efficiency or to increase the EQE at the same thickness. These cooperative effects are captured in the product α, with the only difference being that L_D's thickness-increasing attribute is balanced by η's normalization to L_D, which is clear in Figure 77B; as L_D increases and the density decreases, the optimal thickness rises. From this examination we can finally address the practically-achievable morphology trade-off between increasing density and increasing diffusion length, such as encountered when choosing between isotropic and vertically-aligned films using current techniques: because the product of ϕ and L_D is the driver of EQE, they should match. That is demonstrated in the symmetry of Figure 77C. Roughly, EQE is maximal when ϕ is about $1/1000^{\text{th}}$ of L_D in nm.

In hindsight the reduction of the system to α is intuitive: the rate of light absorption (exciton generation) is dictated by the film density, while the rate of exciton collection is dictated by the diffusion length (and film thickness). Higher density and higher L_D increase the rate of charge collection cooperatively. Having reduced the system to this single parameter, we can construct the master curve Figure 77D. To employ this curve, an experimentalist only needs to measure her SWNT film's α, ideally having increased it as much as possible (Figure 77C). Above very low α (10 nm), The EQE has a plateau-like cross-section (Figure 76C) that makes it drop off suddenly far from η^*, but fairly invariant close to η^*. As a result, there is a high degree of EQE-tolerance in the region around η^*, exemplified by the blue bounds in Figure 77D within which the EQE is ≥99.5% of the maximum possible EQE. That tolerance means that even if experimenters do not measure their α, they can comfortably approximate $0.6 \leq \eta^* \leq 1.4$ or, equivalently, $T^* \cong L_D$. In other words, near-maximal performance can be achieved by setting the device thickness equal to the exciton diffusion length in the charge-collecting axis.

Under the simplifications we have made, that result is applicable to any excitonic photovoltaic material with bell-shaped absorption modes. The rule of thumb breaks down at two extremes: very high and very low α. The latter case is physically realistic; it is sparse films, or materials with very low optical absorptivity. In those cases, the optimal thickness becomes many times the diffusion length due to the poor light collection (see the low α end of Figure 77D). The other limit of very high α can be imagined as extremely high light absorption, where the exciton diffusivity becomes irrelevant and the film should be as thin as possible. *i.e.* $\eta^* \ll 1, T^* \ll L_D$. This case is not physically realistic with existent materials however; SWNT already have very high optical absorptivity compared to most photovoltaic materials, and density is constrained to a maximum of close-packed density. As a result, our maximum α considered in Figure 77D is unlikely to be exceeded by any other material.

APPENDICES

Appendix A. Detailed Carbon Nanotube Solar Cell Model Derivation

Outline of approach and homogeneity approximation

To aid application of our methodology to other systems, we can summarize our approach in three steps:

1. Treat single-nanotube properties that are relevant to the network performance as random variables. The priority properties are position, length, orientation, and chirality. The network is defined by the distribution of those parameters.
2. Derive the network behavior from single-nanotube physics as a function of those parameters by performing photon, exciton, and free carrier volume balances. That produces sets of differential equations dependent on the single-SWNT properties.
3. Integrate those differential equations over the distributions of those properties (*e.g.* integrate over the orientation distribution). That integration is equivalent to summing up the contributions to the balances from each independent population of nanotubes with each possible value of the random variables.

This treatment is only valid under an approximation of macroscopic homogeneity; for the distributions to be valid, any slice of the film that is large enough to observe the film's steady state behavior must contain the same distribution of properties regardless of where you take the slice. In other words, the film must be homogenous on the length scale that the film's photovoltaic efficiency is reproducible. Experimentally, the invalid case has been observed with conductivity measurements.[199] This approximation not only constrains our application of the model to films with densities above the percolation threshold in the charge-collecting axis, but also excludes cases where the film is highly porous or otherwise exhibits clusters of density, length, chirality, or orientation. It is possible to relax this approximation and treat such cases – see the Appendix *Breakdown of Macroscopic Homogeneity* – but in this work we focus on taking the approximation. The approximation is ultimately (see *Results* section above) quite useful, in that it allows us to treat a space of cases that are of top priority technologically – higher density films where device performance is ideal. We point out that the depletion zone approximation in p-n junction Si solar cells has proven extraordinarily useful, despite also being an approximation. We furthermore find quantitative agreement between our predicted efficiencies and those reported in the literature at densities below what we can rigorously treat, indicating that our approximation may be generally quite good. To truly evaluate our model, experimental devices must have the key structural properties – nanotube density, chirality, length distribution, impurity content, and film thickness – measured. By pointing out the quantities of relevance in this work, and highlighting the positive impact they can, when controlled, have on device efficiency, we hope to encourage experimental groups to produce such measurements.

Extensibility

This model has been deliberately derived from single- and aggregate-nanotube photophysics in order to provide it with a sufficient extensibility for practical utility. For example, we derive an expression for the film diffusion coefficient dependent on the film's distribution of nanotube orientations, chiralities, impurity concentration, density, *etc.* An empirically-measured diffusion coefficient for a particular film would not be extensible outside of films with matching distributions of those network-defining variables. In contrast by expressing the diffusion coefficient as a function of the measurable intrinsic qualities of the network, our model can be applied to disparate films without adjustment. If our model is ultimately proven valid, empirical measurements of one film can use our comparative statics to predict the properties of other films, translating *via* the intrinsic properties.

Notation

Bold ($\boldsymbol{v}$) denotes a vector quantity, which may be represented in Cartesian

$$\boldsymbol{v} = (v_x, v_y, v_z) = v_x\hat{\boldsymbol{x}} + v_y\hat{\boldsymbol{y}} + v_z\hat{\boldsymbol{z}} \tag{140}$$

or spherical

$$\boldsymbol{v} = (r, \theta, \phi), v_x = r\sin\theta\cos\phi\,, v_y = r\sin\theta\sin\phi\,, v_z = r \tag{141}$$

coordinates. We will switch between spherical and Cartesian coordinates for convenience without comment for brevity – which is used should be clear from the context. The space on which a vector is defined varies by context, for example $\boldsymbol{c} = (n_1, n_2)$ is the chirality of a nanotube with chiral indices n_1 and n_2. Hats ($\hat{\boldsymbol{x}}$) denote unite vectors. Subscripts x, y, z will be used to denote the scalar components of a vector in the corresponding Cartesian axis. Unbolded variables (v) corresponding to vectors are the magnitudes of those vectors.

Probability density functions of a random variable are denoted $p(\cdot)$. Mean values are bracketed as $\langle\cdot\rangle$.

Network Geometry

We consider a network of single-walled nanotubes (SWNT) sandwiched between two electrode plates (Figure 78a). We define a cartesian z axis as perpendicular to the incident solar photon flux J_0 at $z = 0$ (the 'top' of the film), with $z \in \mathbb{R}^+$ being the depth of the film up to thickness T and the film being infinite in cartesian dimensions x and y. In practice the solar flux may be incident at an angle, which would be further altered by the top electrode index of refraction. These considerations merely change the boundary conditions of the light field problem below. The back electrode has reflectance[xxxiii] $0 \leq R \leq 1$ which can in general be frequency dependent. A variety of conduction and valence band charge collecting

[xxxiii] Specular reflectance; we neglect diffuse reflection as a simplification in our treatment.

electrode (type II exciton dissociation interface) configurations are possible (Figure 78c), including those where an electrode also acts as a photoabsorbing layer (e.g., fullerenes, references [115,116]). We focus on the relatively general case of two dissociating electrodes (first cartoon in Figure 78c), but the model for the SWNT network applies to any other case by corresponding adjustment of the boundary conditions (next sections). Particular material selection and band alignment issues associated with exciton dissociation are well outside the scope of this work, and instead we take the electrode properties as given.[117]

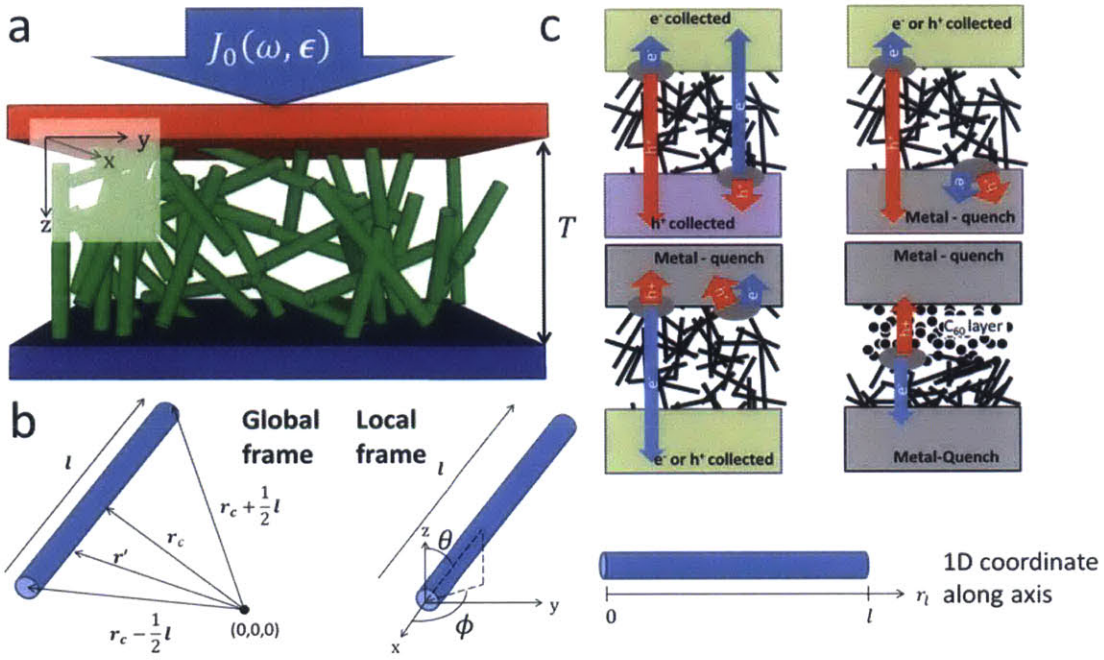


Figure 78. Cartoon of model geometry. See text for description.

Throughout this work we adopt several spatial coordinate references (Figure 78b). An individual nanotube, which we approximate as a rigid rod, has a length and orientation described by the vector in spherical coordinates as $\boldsymbol{l} = (l, \theta, \phi)$, or equivalently in Cartesian coordinates

$$\begin{aligned} \boldsymbol{l} &= (l_x, l_y, l_z) \\ l_x &= l \sin\theta \cos\phi \\ l_y &= l \sin\theta \sin\phi \\ l_z &= l \cos\theta. \end{aligned} \tag{142}$$

In relation to our global coordinate system the center of each nanotube is located at $\boldsymbol{r}_c$ relative to an arbitrary origin in the xy plane at $z = 0$. The ends of a single SWNT are therefore located, relative to the origin, at the ends of vectors $\boldsymbol{r}_c \pm \frac{1}{2}\boldsymbol{l}$. The set of points along the central axis of the SWNT is then constrained as

(143) $$\left\{\boldsymbol{r}'\middle|\boldsymbol{r}_c+\left(\alpha_l-\frac{1}{2}\right)\boldsymbol{l},\alpha_l\in[0,1]\right\}.$$

In describing intra-SWNT mechanics we will also consider the one-dimensional space along a nanotube axis, the coordinate along which we will denote $r_l \in [0, l]$. Each r_l corresponds with an $\boldsymbol{r}'$ via $r_l = \alpha_l l$ and Equation (143); there is a bijective map between $\{\boldsymbol{r}'\}$ and $\{r_l\}$. A nanotube also has a chirality $\boldsymbol{c} \equiv (n_1, n_2)$ where n_1 and n_2 are chiral indices.[176] For convenience we index all chiralities present in the network by integer i (chirality $\boldsymbol{c}_i$) in order of increasing exciton (optical) band gap $E_{i+1} > E_i$. The center-center diameter of chirality $\boldsymbol{c}$ is approximately

(144) $$d_{\boldsymbol{c},c-c} \cong \frac{a_0}{\pi}\sqrt{n_1^2+n_2^2+n_1n_2}$$

where $a_0 = 2.461 \times 10^{-10}\ m$ is the graphene lattice constant.[176] The effective outer diameter $d_{\boldsymbol{c}}$ is estimated as the diameter within which the bulk of the electron shell resides, ~0.335 nm greater than $d_{\boldsymbol{c},c-c}$ for tubes with $d_{\boldsymbol{c},c-c} > 0.7\ nm$, based on the thickness of graphene.[200]

The properties $\{\boldsymbol{r}_C, \boldsymbol{l}, \boldsymbol{c}\}$ completely define a given nanotube in our network. Our approach begins with treating these parameters as random variables, defining a given film by their distributions:

(145)
$$\begin{aligned} &p(\boldsymbol{r}_C|\boldsymbol{r}) = p(r_{cx}, r_{cy}, r_{cz}|\boldsymbol{r}) \\ &p(\boldsymbol{l}) = p(l)\cdot p(\theta,\phi) \\ &p(\boldsymbol{c}_i) = v_i \in [0,1], \qquad \sum_i v_i = 1 \end{aligned}$$

where v_i is the number fraction of chirality i in the film. In theory any of the distributions can be dependent on the others depending on the nanotube and film fabrication processes; we focus on cases where they are uncorrelated, but the derivation is general for correlated distributions. We will generate forms of these distributions for different relevant cases – e.g. isotropic, vertically aligned, horizontally aligned, monochiral, mSWNT impurities, short SWNT, long SWNT, etc. when we apply the model to relevant cases (see main paper). Our derivation is also general for cases where all of these properties vary systematically with depth, *i.e.* $p(\boldsymbol{r}_C, \boldsymbol{l}, \boldsymbol{c}|z)$, but in most practical applications $p(\boldsymbol{l})$ and $p(\boldsymbol{c}_i)$ are spatially invariant. Note we are also implicitly assuming that SWNT locations are independent, even though in some film casting processes they can be correlated; this phenomenon represents an important shortcoming of our model, and is discussed further along in the derivation and in the section *Breakdown of Macroscopic Homogeneity*.

Light Field

At any given point $\boldsymbol{r} = (x, y, z) \in \big((-\infty, \infty), (-\infty, \infty), [0, T]\big)$ there is a total photon flux $J_v(\boldsymbol{r})$ that is distributed over (as $J_v(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r})$, *i.e.* irradiance) frequency ω, linear polarization component $\boldsymbol{\epsilon} = (\theta_\epsilon, \phi_\epsilon)$, and propagation direction $\boldsymbol{k} = (\theta_k, \phi_k)$,[xxxiv] i.e.

$$J_v(\boldsymbol{r}) = \int_0^\infty \int_0^\pi \int_0^\pi \int_0^{2\pi} J_v(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}) d\phi_\epsilon \, d\theta_\epsilon \, d\theta_k \, d\omega. \tag{146}$$

We do not consider magnetic field effects in this model, although it could be introduced analogously to the electric field if desired subject to $\boldsymbol{b} = \boldsymbol{\epsilon} \times \boldsymbol{k}$. Over our length-scales of interest (< 1 μm) for SWNT material, we neglect electric-field phase shifts such as circular dichroism that introduce and/or manipulate circular polarization components to the light field.

Distributions in polarization and propagation direction of electromagnetic radiation are not independent, they are orthogonal,

$$\boldsymbol{\epsilon} \cdot \boldsymbol{k} = 0, \tag{147}$$

reducing our degrees of freedom by one, leaving only 3 independent dimensions. In practice all four coordinates can be *effectively* taken as independent if proper 'accounting' is done in all the mechanics that follow – i.e. manipulations of the distributions of $\boldsymbol{k}$ are accompanied by appropriate transformation of the distributions of $\boldsymbol{\epsilon}$ and vice versa. Alternatively, that implicit dependency can be made explicit. For example, I take $(\theta_k, \theta_\epsilon, \phi_\epsilon)$ as the independent set, and trigonometrically extract ϕ_k as a function of them:

$$\begin{aligned} \boldsymbol{\epsilon} \cdot \boldsymbol{k} &= 0 \\ &= \epsilon_x k_x + \epsilon_y k_y + \epsilon_z k_z \\ &= \sin\theta_\epsilon \cos\phi_\epsilon \sin\theta_k \cos\phi_k + \sin\theta_\epsilon \sin\phi_\epsilon \sin\theta_k \sin\phi_k \\ &\qquad + \cos\theta_\epsilon \cos\theta_k = 0, \end{aligned} \tag{148}$$

yielding

$$\phi_k(\theta_k, \theta_\epsilon, \phi_\epsilon) = \pi + \phi_\epsilon \pm \operatorname{acos}(\cot\theta_\epsilon \cot\theta_k). \tag{149}$$

As an example of how this basis practically manifests itself, consider a function $F(\boldsymbol{k}) = F(\theta_k, \phi_k)$. We can evaluate its first moment at $\boldsymbol{r}$

[xxxiv] Usual convention is to bundle the direction of propagation and the frequency of the electric field, i.e. $k=2\pi/\omega$, to be consistent with the plane wave representation of light. Since we will not exploit the plane wave character of light here however it becomes, for reasons clear later, more convenient to separate the frequency of light from the orientation of it, thus k is an orientation vector only, it does not connect with the spatial variation of the electric field.

$$\langle F(\theta_k,\phi_k)\rangle_{\boldsymbol{r}} = \int_{\boldsymbol{k}} F(\boldsymbol{k})\, J_\nu(\boldsymbol{k}|\boldsymbol{r})\, d\boldsymbol{k}$$

$$= \int_0^{\pi}\int_0^{2\pi} F(\theta_k,\phi_k) J_\nu(\theta_k,\phi_k|\boldsymbol{r}) d\phi_k\, d\theta_k \tag{150}$$

$$= \int_0^{\pi}\int_o^{\pi}\int_o^{2\pi} F\big(\theta_k,\phi_k(\theta_k,\theta_\epsilon,\phi_\epsilon)\big) J_\nu(\theta_k,\theta_\epsilon,\phi_\epsilon|\boldsymbol{r})\, d\phi_\epsilon\, d\theta_\epsilon d\theta_k$$

where

$$J_\nu(\theta_k,\theta_\epsilon,\phi_\epsilon|\boldsymbol{r}) = \int_0^{\infty} J_\nu(\omega,\boldsymbol{\epsilon},\boldsymbol{k}|\boldsymbol{r}) d\omega. \tag{151}$$

The SWNT network can interact with the field through three predominant mechanisms:

1. Absorption, generating excitons of energy $\hbar\omega$.
2. Rayleigh scattering, shifting the distribution of propagation vectors $J_\nu(\boldsymbol{k}|\boldsymbol{r})$.
3. Photoluminescence (PL), radiative decay of excitons to contribute to the photon flux at the energy of the band gap $\hbar\omega_{1u}$, position of the relaxation, and polarization of the transition dipole.

All three phenomena could be included in the derivation that follows, but to first order we neglect the latter two; omitting reflection allows us to remove $\boldsymbol{k}$ from our calculations, and omitting PL greatly simplifies coupling to the exciton transport equations. The former case we will show below, and the latter in the next section. See Appendix D below for consideration of the consequences of this choice.

To treat attenuation of the field we must derive the linear-polarization- and frequency-dependent absorption cross-section of the film from the single-SWNT absorption behavior. The absorption cross-section of SWNT of chirality $\boldsymbol{c}$ can be broken down as

$$\sigma(\boldsymbol{c},\boldsymbol{\epsilon},\omega) = \frac{w(\boldsymbol{c},\boldsymbol{\epsilon},\omega)}{J_\nu(\boldsymbol{c},\boldsymbol{\epsilon},\omega)} \tag{152}$$

where w is the photon absorption rate at that polarity and frequency, and σ and w are normalized per atom C or mole C (with no length dependence, see references [127,201]). Theoretical evaluation of absorbance is possible,[202] but given sources of variation, such as dependence on the electric and dielectric environment,[100,101,116,126] empirical measurement, $\sigma_{meas}(\boldsymbol{c},\boldsymbol{\epsilon},\omega)$,[108,126,185,203] is most immediately appealing. Of relevance theoretically is the polarization dependence. A time-dependent perturbation theory treatment of the light-matter interaction yields, to first order under the dipole approximation,[204]

$$\sigma(\boldsymbol{\epsilon},\omega) \propto \left|\boldsymbol{\epsilon}\cdot\boldsymbol{\mu}_{if}\right|^2 \rho(E_0+\hbar\omega) \tag{153}$$

where $\boldsymbol{\mu}_{if}$ is the transition dipole between initial and final electronic states, $\langle i|\hat{\boldsymbol{r}}|f\rangle$, E_0 is the energy of ground state $|i\rangle$, and $\rho(E)$ is the density of states at energy E. Electronic transitions in SWNT exist with dipoles parallel and perpendicular to the longitudinal axis (the latter could be called axial as well). The

latter set contribute only ~1/5th of the total absorbance (integrated across all ω), largely due to the depolarization effect.[176,205]

These transitions provide us with two orthogonal sets of polarization dependence:

$$\sigma(\boldsymbol{c}, \boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega) \propto \left|\boldsymbol{\epsilon} \cdot \boldsymbol{\mu}_{\parallel}(\boldsymbol{c}, \omega)\right|^2 + \left|\boldsymbol{\epsilon} \cdot \boldsymbol{\mu}_{\perp}(\boldsymbol{c}, \omega)\right|^2, \tag{154}$$

where $\boldsymbol{\mu}_{\parallel}$ and $\boldsymbol{\mu}_{\perp}$ are sums of all the dipole moments for transitions of energy $\hbar\omega$ oriented parallel and perpendicular, respectively, to the longitudinal SWNT axis. We therefore effectively have two experimentally- or theoretically-determined absorption cross-section functiona $\sigma_{\parallel}$ and $\sigma_{\perp}$, giving us a total absorption cross-section for a single SWNT of orientation $\hat{\boldsymbol{l}} = (\theta, \phi)$ as

$$\sigma(\boldsymbol{c}, \boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega) = \left|\boldsymbol{\epsilon} \cdot \hat{\boldsymbol{l}}\right|^2 \sigma_{\parallel}(\boldsymbol{c}, \omega) + \left(1 - \left|\boldsymbol{\epsilon} \cdot \hat{\boldsymbol{l}}\right|\right)^2 \sigma_{\perp}(\boldsymbol{c}, \omega), \tag{155}$$

where the second coefficient has a simple form because both polarization and orientation are unit vectors, and any component of linear polarization not parallel to the SWNT axis is necessarily in the plane perpendicular to it. We will define our coefficients for convenience as

$$\varepsilon_{\parallel}(\theta_{\epsilon}, \phi_{\epsilon}, \theta, \phi) \equiv \boldsymbol{\epsilon} \cdot \hat{\boldsymbol{l}} \tag{156}$$

$$\varepsilon_{\perp}(\theta_{\epsilon}, \phi_{\epsilon}, \theta, \phi) \equiv 1 - \varepsilon_{\parallel}(\theta_{\epsilon}, \phi_{\epsilon}, \theta, \phi). \tag{157}$$

In the derivation that follows we will find that normalizing σ per length, σ_l, rather than carbon units is more convenient. For σ per mole of C, as is more usually reported, we can convert readily. The SWNT circumference is

$$2\pi\left(\frac{1}{2} d_{\boldsymbol{c},c-c}\right) = \pi \cdot d_{\boldsymbol{c},c-c}, \tag{158}$$

which gives us

$$\frac{1}{N_A} \rho_{graph} \cdot \pi d_{\boldsymbol{c},c-c} \tag{159}$$

moles of carbon per length of SWNT of chirality $\boldsymbol{c}$, where N_A is avogadro's constant and ρ_{graph} is the density of graphene. This relation gives us

$$\sigma_l(\boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega, \boldsymbol{c}) = \frac{1}{N_A} \rho_{graph} \cdot \pi d_{\boldsymbol{c},c-c} \cdot \sigma(\boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega, \boldsymbol{c}). \tag{160}$$

This conversion is approximate as it neglects bond stress relaxation from curvature, and so measured constants can be used when available; for (6,5) SWNT this estimates 9.1558×10^{10} carbon atoms per length versus 8.8271×10^{10} measured in reference [185].

We can now consider the network. In the remainder of the derivation, two spatial magnifications will be traversed, and the transition between the two will define much of the approach. At the 'macroscopic' scale, $\geq O(100\ nm)$, where the diameter of the nanotubes is negligible, we approximate that the film is essentially homogenous. At the 'microscopic' scale, on the order of $d_{\boldsymbol{c}}$, heterogeneity in the local environment is highly relevant. There are realistic situations where the homogeneity

presumption in a strict sense breaks down; the circumstances and impact of that error, as well the means of relaxing the assumption (accounting for macroscopic heterogeneity), are laid out in the Appendix *Breakdown of Macroscopic Heterogeneity* below.

From the perspective of the gradient in the light field, the diameter of the SWNT is negligible and the film is essentially homogenous. The length-density of SWNT we will define as the length of SWNT per volume of the film,

$$\rho_{\langle l \rangle} = \int_0^\infty \rho \cdot l \cdot p(l) dl = \rho \cdot \langle l \rangle, \tag{161}$$

where ρ is the number density of SWNT and $\langle l \rangle$ is the average length of SWNT. Without any depth-variation in density, the absorption cross-section per volume of film due to SWNT of chirality i and orientation $\hat{\boldsymbol{l}}$ is

$$v_i \cdot \rho_{\langle l \rangle} \cdot \sigma_l(\boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega, \boldsymbol{c}_i). \tag{162}$$

where $v_i \rho_{\langle l \rangle}$ can be understood as the density of chirality i in the film. Note that we are omitting explicit dependence on the dielectric environment of each SWNT; that is justifiable under homogeneity if $\sigma_\parallel$ and $\sigma_\perp$ are measured in or corrected for the appropriate state, but alternatively σ_l can be conditioned on local dielectric constant and integrated over the distribution of it.

A reader familiar with SWNT literature will be more accustomed to seeing quantities liked density expressed in atom or mole Carbon units, rather than by length. That treatment is entirely equivalent to using the length of SWNT, mediated by the carbon atoms per length SWNT, so long as chirality is appropriately accounted for. The per-length quantities however are more useful, as we will see, in translating between one-dimensional and three-dimensional diffusion, where we care about the length of the SWNT present independent of the number of carbon atoms.

For irradiance $J_\nu(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r})$ the photon absorption rate per volume due to SWNT of chirality i oriented in $\hat{\boldsymbol{l}}$ are

$$N_i(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}, \hat{\boldsymbol{l}}|\boldsymbol{r}) = v_i \cdot \rho_{\langle l \rangle} \cdot \sigma_l(\boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega, \boldsymbol{c}_i) \cdot J_\nu(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}). \tag{163}$$

To get the total film absorption rate we must sum absorption terms (163) for nanotubes at each possible orientation $\hat{\boldsymbol{l}}$ and chirality i, weighted by the relative populations of each. That is equivalent however to integrating over the distribution $p(\hat{\boldsymbol{l}})$ (and summing over chiralities), forming the crux of our methodology,

$$
\begin{aligned}
N(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}) &= \sum_i \int_0^{2\pi} p(\phi) \int_0^{\pi} p(\theta)\, N_i\left(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}, \hat{\boldsymbol{l}}|\boldsymbol{r}\right) d\theta\, d\phi \\
&= \sum_i v_i \int_0^{2\pi} p(\phi) \int_0^{\pi} p(\theta)\, \rho_{\langle l \rangle} \sigma_l\left(\boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega, \boldsymbol{c}_i\right) J_v(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r})\, d\theta\, d\phi \\
&= \rho_{\langle l \rangle} J_v(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}) \sum_i v_i \int_0^{2\pi} p(\phi) \int_0^{\pi} p(\theta)\, \sigma_l\left(\boldsymbol{\epsilon}, \hat{\boldsymbol{l}}, \omega, \boldsymbol{c}_i\right) d\theta\, d\phi \\
&= \rho_{\langle l \rangle} J_v(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}) \sum_i v_i \sigma_l(\boldsymbol{\epsilon}, \omega, \boldsymbol{c}_i).
\end{aligned} \tag{164}
$$

In this manner, we can describe the light interaction behavior of the film by integrating the single-SWNT behavior over the distribution of independent single-SWNT properties. We will take the same approach to exciton and free carrier transport below, where will also consider its utility in more depth. In the last two equalities of (164) we are showing that $\sigma_l(\boldsymbol{\epsilon}, \omega, \boldsymbol{c}_i)$ can be safely evaluated from the film properties before any consideration of absorption, which we exploit in the Results section of the main paper.

Consider a balance on the hypothetical number of photons in a differential volume of the solar cell, $N_P(\boldsymbol{r})$. With absorption as the only light-matter interaction we treat, the change in photons with time is due to absorption events $N(\boldsymbol{r})$ and any gradient in the flux,

$$
\frac{dN_p(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r})}{dt} = -\nabla \cdot J_v(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}) - \sum_i N_i(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}). \tag{165}
$$

At steady state we conveniently lose our $N_p(\boldsymbol{r})$ dependence,

$$
\nabla \cdot J_v(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}) = -\sum_i N_i(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|\boldsymbol{r}). \tag{166}
$$

We are now in a position to make two useful reductions. First, in the x and y axes where our film is infinite we will treat periodic solutions as trivial, making our boundary conditions in those dimensions

$$
\begin{aligned}
\frac{\partial J_v}{\partial y} &= 0 \\
\frac{\partial J_v}{\partial x} &= 0.
\end{aligned} \tag{167}
$$

This simplifies our balance to

$$
\frac{dJ(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|z)}{dz} = -\sum_i N_i(\omega, \boldsymbol{\epsilon}, \boldsymbol{k}|z). \tag{168}
$$

Second, neglecting Rayleigh scattering and photoluminescence the $\boldsymbol{k}$ dependence drops out. We can see this by considering the source of incident light, which below we will treat as a boundary condition. Being perpendicularly incident into the film, we have

$$J_v(\theta_k|z=0) = J_v(z=0)\cdot\delta(\theta_k-\pi). \tag{169}$$

Without fluorescence or Rayleigh scattering, our balance (Equation (62)) contains no transformations of $\boldsymbol{k}$, and so the initial distribution $J_v(\boldsymbol{k}|z) = J_v(z)\cdot\delta(\theta_k-\pi)$ is maintained. θ_k then trivially integrates out of the balance; for the left hand side we have

$$\begin{aligned}\int_0^{2\pi} \nabla J_v(\omega,\boldsymbol{\epsilon}|\boldsymbol{r})\delta(\theta_k-\pi)\,d\theta_k &= \nabla J_v(\omega,\boldsymbol{\epsilon}|\boldsymbol{r})\int_0^{2\pi}\delta(\theta_k-\pi)\,d\theta_k \\ &= \nabla J_v(\omega,\boldsymbol{\epsilon}|\boldsymbol{r}),\end{aligned} \tag{170}$$

and for the right hand side

$$\begin{aligned}&-\int_0^{2\pi}\sum_i N_i(\omega,\boldsymbol{\epsilon},\boldsymbol{k}|\boldsymbol{r})\,d\theta_k \\ &\quad= -\sum_i v_i\rho_{\langle l\rangle}\sigma_l(\boldsymbol{\epsilon},\hat{\boldsymbol{l}},\omega,\boldsymbol{c}_i)\int_0^{2\pi} J_v(\omega,\boldsymbol{\epsilon},\boldsymbol{k}|\boldsymbol{r})d\theta_k \\ &\quad= -\sum_i v_i\rho_{\langle l\rangle}\sigma_l(\boldsymbol{\epsilon},\hat{\boldsymbol{l}},\omega,\boldsymbol{c}_i)J_v(\omega,\boldsymbol{\epsilon}|\boldsymbol{r})\int_0^{2\pi}\delta(\theta_k \\ &\quad-\pi)\,d\theta_k = -\sum_i v_i\rho_{\langle l\rangle}\sigma_l(\boldsymbol{\epsilon},\hat{\boldsymbol{l}},\omega,\boldsymbol{c}_i)J_v(\omega,\boldsymbol{\epsilon}|\boldsymbol{r}) \\ &\quad= -\sum_i N_i(\omega,\boldsymbol{\epsilon}|\boldsymbol{r}).\end{aligned} \tag{171}$$

This simplifies our balance to

$$\nabla\cdot J_v(\omega,\boldsymbol{\epsilon}|\boldsymbol{r}) = -\sum_i N_i(\omega,\boldsymbol{\epsilon}|\boldsymbol{r}). \tag{172}$$

Combining the two simplifications, our balance reduces to

$$\frac{dJ_v(\omega,\boldsymbol{\epsilon}|z)}{dz} = -\sum_i N_i(\omega,\boldsymbol{\epsilon}|z). \tag{173}$$

With a transparent back electrode (at $z = T$), the boundary condition can be defined as an incident unpolarized AM1.5 solar flux,

$$J_v(\omega,\boldsymbol{\epsilon}|z=0) = J_0(\omega,\boldsymbol{\epsilon}). \tag{174}$$

Note that 'unpolarized' is still constrained to be in the incident plane, giving us distributions

(175) $$J_0(\omega, \theta_\epsilon, \phi_\epsilon) = J_0(\omega) \cdot \delta\left(\theta_\epsilon - \frac{\pi}{2}\right) \cdot \frac{1}{2\pi}, \qquad \phi_\epsilon \in [0, 2\pi]$$

where $J_0(\omega)$ is the AM1.5 solar spectrum. We cannot treat reflection off the back electrode as an alternative boundary condition ($J_\nu'(T) = 0$), as omitting $\boldsymbol{k}$ prevents us from describing the incident flux with a generation term in the balance (64), such as

(176) $$+J_0(\omega, \epsilon, \boldsymbol{k}) \cdot \delta(z - 0)$$

(177) $$J_0(\boldsymbol{k}) = J_0(\phi_k) = \delta(\theta_k - \pi).$$

Instead, we recognize that absorption events are independent, allowing us to treat the total light field as the sum of two other fields – one representing the 'forward' flux, $J_F(z)$, and one representing the 'reverse' flux, $J_R(z)$. Interference of incident and reflected light is constrained to a small band of frequencies around harmonics of the film thickness and we therefore neglect it. The fluxes have identical ODEs with opposite generation sign,

(178) $$\frac{dJ_F(\omega, \boldsymbol{\epsilon}|z)}{dz} = -\sum_i N_i(\omega, \boldsymbol{\epsilon}|z)$$

(179) $$\frac{dJ_R(\omega, \boldsymbol{\epsilon}|z)}{dz} = \sum_i N_i(\omega, \boldsymbol{\epsilon}|z),$$

but with different boundary conditions. In the forward direction we have the incident solar flux BC, and in the reverse we reflect the forward flux at $z = T$,

(180) $$J_R(\omega, \boldsymbol{\epsilon}|z = -T) = R \cdot J_F(\omega, \boldsymbol{\epsilon}|z = T),$$

where $R \in [0,1]$ is the reflectivity of the back electrode. Solving the forward and reverse ODEs in sequence, we get the total light field

(181) $$J_\nu(\omega, \boldsymbol{\epsilon}|z) = J_F(\omega, \boldsymbol{\epsilon}|z) + J_R(\omega, \boldsymbol{\epsilon}|z).$$

Integrating over $(\omega, \boldsymbol{\epsilon})$ yields the flux gradient $J_\nu(z)$ and the photon absorption rate $N(z)$.

We should also note that by neglecting $\boldsymbol{k}$ we also complicate treatment of cases where light is not incident along z, i.e. light incident at an angle. In that case we would need to treat not only the longer path length of light both on incidence and reflection, but also we would need to distinguish between the direction that the light gradient is formed along and the direction that excitons diffuse in in the next problem. The situation can be rectified by performing the entire light derivation above not along the dimension z, but along some propagation direction κ with a z component of $\cos\theta_\kappa$. The entire derivation would be the same, except that the gradient in x and y must still be zero, so a 3D balance must be solved to enforce that and extract the gradient in z.

Exciton Transport

In SWNT, optical electronic excitations result in exciton generation – bound carrier states – rather than free charges, due to one dimensional confinement.[98,100,177,178] To collect charges in a solar cell, excitons must be dissociated either at a type II interface or by an electric field imparting coulomb force greater than the binding energy. Nonradiative decay of hot excitons to the band gap occurs on a timescale of $O(10\ fs)$;[206,207] with a longitudinal diffusion coefficient D_l of $O(100\ cm^2/s)$ (see Chapter V), the *hot* exciton diffusion length is $< 10\ nm$ and the exponential decay of excited-state energy brings it close to ground state within angstroms. Given that, additionally, the inter-SWNT transport occurs on a timescale of $O(10\ ps)$,[193] we make the key assumption that for $T \gtrsim 10\ nm$, excitons nonradiatively relax to the band gap E_i of the relevant SWNT chirality instantly relative to any other processes we consider. This approximation allows us to neglect hot exciton behavior entirely, and treat the exciton generation rate at band gap energy E_i as the photon absorption rate for chirality i, $N_i(z)$. We further assume that excitons do not interact with free charges, allowing us to couple the two systems only through exciton dissociation. While local dielectric environment – which impacts the exciton diffusion coefficient[106,110,111] – can be inserted in our model as another random variable without changing the derivation that follows (see *Including Dielectric Environment* section below), we have chosen for simplicity to omit it; for a fairly homogenous dielectric environment, it could be rolled into D_l without much loss of accuracy.

As we will show, a single empirical diffusion coefficient describing exciton transport in a film is limited to *only one* particular geometry and chemistry. Thus, an empirical approach alone would be misguided. Instead, we start with the single-SWNT 1D exciton reaction-diffusion behavior. We then derive the contribution of a network of such systems to three-dimensional exciton transport, and couple them via exciton hopping (EH). The goal is to extract the rate at which excitons arrive and dissociate at the electrodes at $z = \{0, T\}$.The transport is thereby treated in a highly general manner, accommodating for arbitrary network density, co-alignment, net orientation, impurity types and concentrations, length distribution, bundle fraction and size, chirality mixture, and any other properties dependent on the random variables that we used to define the film. There are several ways to approach the coupling problem; the one adopted here was chosen for tractability and generality, but elegant alternatives are mentioned in the section *Appendix G. Alternative methods of coupling the infinite number* of line segments.

Beginning with a single nanotube of length l, there is a one dimensional coordinate

$$r_l \in [0, l] \text{ or } \alpha_l = \frac{r_l}{l} \in [0,1], \tag{182}$$

with an exciton concentration $n(r_l)$ excitons per length. Excitons are generated at a constant rate $N(r_l)$ that corresponds with an associated $N(\boldsymbol{r}')$ in 3D space from the light absorption problem. In general the

exciton concentration is also a function of time, $n(r_l, t)$, but we will be concerned only with steady state operation of the system. We take the excitons to be point particles, i.e. neglecting the finite electron-hole correlation length. Intra-SWNT exciton transport is diffusive via elastic exciton-phonon scattering.[102,106,191,208] This allows us to construct a 1D differential volume balance

$$\frac{dn}{dt} = N(r_l) + D_l \frac{d^2 n}{dr_l^2}, \tag{183}$$

where D_l is the theoretically- or empirically-evaluated longitudinal exciton diffusion coefficient in the dielectric environment of interest (and at the temperature of interest – we focus isothermal operation, see *Relaxing Isothermal Approximation* for how to introduce temperature dependency).[106]

We can then introduce relevant relaxation mechanisms. Radiative (photoluminescent, PL) decay is first order with rate constant $k_\Gamma = 1/\tau_\Gamma$, where τ_Γ is the PL time constant. Exciton quenching from impurities can either be localized via inter-band states allowing nonradiative relaxation (e.g. covalent sp^3-bonded[208,209] or ionically bonded[191,210] moieties) or delocalized due to doping.[208] Terms should be included for each species of interest (e.g., oxidative agents, catalyst nanoparticles), with delocalized quenching exhibiting first-order kinetics and localized quenching for an impurity of type im with uniformly distributed concentration distribution n_{im} providing quenching rate

$$-k_{im} n_{im} n(r_l) \tag{184}$$

where k_{im} (in length^2 time^{-1}) is the associated 'bimolecular' rate constant. Note that n_{im} can be understood as the mean number of impurity contacts per length of SWNT; for example if each catalyst nanoparticle on average contacts 3 SWNT, then n_{im} would be three times the film's particle density (number density) divided by the SWNT length density $\rho_{\langle l \rangle}$. Exciton-exciton annhiliation (EEA) via auger recombination is also possible at high fluences,[108,192,211,212] yielding the bimolecular reaction term

$$-k_{EEA} \cdot n^2 \tag{185}$$

where k_{EEA} is the rate constant. Following [192] the EEA rate constant can be calculated as

$$k_{EEA} = 128 \frac{\omega_{vc}}{k_{e0}} \left(\frac{\mu}{m_0}\right) \left(\frac{E_{BE}}{E_{11}}\right)^3 \tag{186}$$

where E_{11} is the direct band gap energy, $\mu = m_e m_h/(m_e + m_h)$ is the exciton reduced mass where those masses are the effective masses, $m_e \sim m_h \sim 0.1 m_0$, E_{BE} is the exciton binding energy, m_0 is the free electron mass, ω_{vc} is the interband transition strength, and

$$k_{e0} \equiv \sqrt{\frac{\mu(E_g - 2E_{BE})}{\hbar}}. \tag{187}$$

k_{EEA} is approximately $6 \times 10^5\ m/s$ for chiralities in the regime of $E_{BE,i} \cong 0.3\ eV$.[192] Neglecting changes in the exciton coupling potential, we can scale this for different chiralities as

(188) $$k_{EEA}(\boldsymbol{c}_i) = \left(6 \times 10^5 \frac{m}{s}\right)\left(\frac{E_{BE,i}}{E_i}\right)^3 \left(\frac{1.3\ eV}{0.3\ eV}\right)^3.$$

For examinations of single SWNT PL intensities, end quenching is typically taken as a boundary condition.[104,213] Anticipating expansion to three dimensions however, instead we treat end quenching with another reaction term,

(189) $$-k_{end} \cdot n \cdot \left(\delta(r_l - 0) + \delta(r_l - l)\right)$$

where k_{end} is the end quenching rate constant (length^2 time^{-1}). Our 1D volume balance at steady state then becomes

(190) $$\frac{dn}{dt} = N(r_l) + D_l \frac{d^2 n}{dr_l^2} - k_\Gamma n - k_{EEA} n^2 - \sum_{im} k_{im} n_{im} n - k_{end} n\left(\delta(r_l - 0) + \delta(r_l - l)\right) = 0.$$

To determine the network behavior, we construct a three dimensional volume balance subject to diffusion in one dimensional channels. For clarity we will start with a single chirality in the film and omit index i, bringing in multichiral transport afterwards. We start, as we did in the light absorption case, by focusing on a nanotube population oriented in some direction $\hat{\boldsymbol{l}}$, which will yield a set of differential equations describing exciton transport; to include terms in those exciton balances for each possible orientation, we will find that we simply need to integrate the ODEs over the orientation distribution $p(\hat{\boldsymbol{l}})$.

For a hypothetical network where all nanotubes lie along the x axis, we have a one dimensional Fick's law exciton flux in a given SWNT, from above, of

(191) $$J_{1D} = -D_l \frac{dn}{dx}.$$

The resulting three-dimensional flux through the x-face (area $dy \cdot dz$) of a differential volume is the product of the one dimensional flux and the number of channels per area. The number of channels per area however is equivalent to the length of SWNT per volume, $\rho_{\langle l \rangle}$, providing a three dimensional flux of

(192) $$J_{3D,x} = \rho_{\langle l \rangle} J_{1D} = -\rho_{\langle l \rangle} D_l \frac{dn}{dx} = -D_l \frac{dc}{dx}$$

where $c(\boldsymbol{r}, t)$ is the exciton concentration per volume and we have exploited homogeneity in the frame of macroscopic diffusion for the relation

(193) $$\rho_{\langle l \rangle} n = c.$$

Note that while Equation (192) has the form of what one would empirically intuit, there is the crucial difference that D_l is exactly the single SWNT exciton diffusion coefficient, not a net film diffusivity; if we had only isolated (uncoupled) SWNT all lying in the x axis then the network diffusivity would be exactly D_l, as one would expect. For a network with orientation distribution $p(\hat{\boldsymbol{l}})$, the diffusion

coefficient for each dimension can be broken down by deconvoluting the contributions to the longitudinal diffusivity,

$$D_l = \frac{1}{2} \cdot v_{ex} \cdot l_{mfp} \tag{194}$$

where v_{ex} is the mean exciton velocity, l_{mfp} is the mean free path between elastic exciton-phonon interactions, and the ½ geometric factor reflects the single translation dimension. This relation allows us to infer orientation factors for the longitudinal diffusion coefficient in each dimension,

$$D_{l,x} = \frac{1}{2}(\sin\theta\cos\phi)v_{ex} \cdot (\sin\theta\cos\phi)l_{mfp} = (\sin\theta\cos\phi)^2 D_l \tag{195}$$

$$D_{l,y} = \frac{1}{2}(\sin\theta\sin\phi)v_{ex} \cdot (\sin\theta\sin\phi)l_{mfp} = (\sin\theta\sin\phi)^2 D_l \tag{196}$$

$$D_{l,z} = \frac{1}{2}(\cos\theta)v_{ex} \cdot (\cos\theta)l_{mfp} = (\cos\theta)^2 D_l. \tag{197}$$

Note that (θ, ϕ) are random variables, not Cartesian coordinates, i.e. $\nabla_{\boldsymbol{r}} \cdot f(\theta, \phi) = 0, \forall f$. This yields a diffusion flux in three dimensions of

$$D_l(\sin\theta\cos\phi)^2 \frac{dc}{dx} + D_l(\sin\theta\sin\phi)^2 \frac{dc}{dy} + D_l(\cos\theta)^2 \frac{dc}{dz} \tag{198}$$

Note that diffusion in Equation (198) is only due to longitudinal intra-SWNT transport, we have not yet accounted for exciton hopping (EH) between nanotubes.

Next we extract film quenching kinetics. Radiative decay, being monoexcitonic, remains simply linear in concentration and independent of geometry,

$$-k_\Gamma c. \tag{199}$$

Impurity/defect scattering is similar; multiplying by $\rho_{\langle l \rangle}$ to convert the reaction rate per length of SWNT to rate per volume network

$$-\rho_{\langle l \rangle} \cdot k_{im} n_{im} n = -k_{im} n_{im} c. \tag{200}$$

Treating EEA in the same manner we are left with a lingering n factor remedied by an identity:

$$-\rho_{\langle l \rangle} \cdot k_{EEA} n^2 = -k_{EEA} n \cdot c = -\frac{\rho_{\langle l \rangle}}{\rho_{\langle l \rangle}} \cdot k_{EEA} n \cdot c = -\frac{k_{EEA}}{\rho_{\langle l \rangle}} c^2. \tag{201}$$

Note that we therefore predict an inverse scaling of EEA rate with carbon density, an observation that can be experimentally evaluated. Again note that k_{im} and k_{EEA} are the single SWNT 1D rate constants, *not* empirical film constants, rewarding our microscopic derivation.

Note that in general the quenching rates can be made chirality dependent if desired, just sum and attenuate by chiral fraction, as we did with the absorption cross-section $\sigma(\boldsymbol{c}_i)$ in the light absorption problem, *e.g.*,

(202) $$-k_\Gamma c \to -\sum_i v_i k_{\Gamma,i} c.$$

Finally we can consider quenching due to the ends of nanotubes. In three dimensions in our homogenous picture an exciton at position $\boldsymbol{r}$ could be on any SWNT of length l, and, constrained by the distribution $p(\boldsymbol{r}_C|\boldsymbol{r})$, could be at any relative distance to the end. As a result, we can treat the SWNT ends as impurities with some concentration, $c_{end}(\boldsymbol{r})$, that is in general non-uniform. For example in a vertically aligned film with one layer of SWNT, there will be more SWNT ends near the two electrodes than in the center of the film (Figure 79b). This can be thought of as the ability to, constrained by $p(\boldsymbol{r}_C|\boldsymbol{r})$, slide SWNT around any point $\boldsymbol{r}$, bringing the end of the SWNT closer or farther away. This observation gives us a rate of end quenching

(203) $$-\frac{k_{end}}{\rho_{\langle l \rangle}} c_{end}(\boldsymbol{r}) \cdot c,$$

where again k_{end} is the one dimensional rate constant. The distribution of end locations $\boldsymbol{r}_{end}$, $c_{end}(\boldsymbol{r}) = p(\boldsymbol{r}_{end}|\boldsymbol{r})$ can either be constructed independently from the film geometry or calculated from the dependency on $p(\boldsymbol{r}_C|\boldsymbol{r})$ and $p(\boldsymbol{l})$ via their relation,

(204)
$$\begin{aligned}
&\boldsymbol{r}_{end} = \boldsymbol{r}_c \pm \frac{1}{2}\boldsymbol{l} \\
&\Rightarrow p(\boldsymbol{r}_{end}|\boldsymbol{r}_c, \boldsymbol{l}) = \delta\left(\boldsymbol{r}_{end} - \boldsymbol{r}_c + \frac{1}{2}\boldsymbol{l}\right) + \delta\left(\boldsymbol{r}_{end} - \boldsymbol{r}_c - \frac{1}{2}\boldsymbol{l}\right) \\
&c_{end}(\boldsymbol{r}) = p(\boldsymbol{r}_{end}|\boldsymbol{r}) = \int_{\mathbb{R}^3} \int_{\substack{(0,\infty),\\(0,\pi),\\(0,2\pi)}} p(\boldsymbol{r}_{end}|\boldsymbol{r}_c, \boldsymbol{l})\ p(\boldsymbol{r}_c|\boldsymbol{r}) p(\boldsymbol{l}) d\boldsymbol{l} d\boldsymbol{r}_c.
\end{aligned}$$

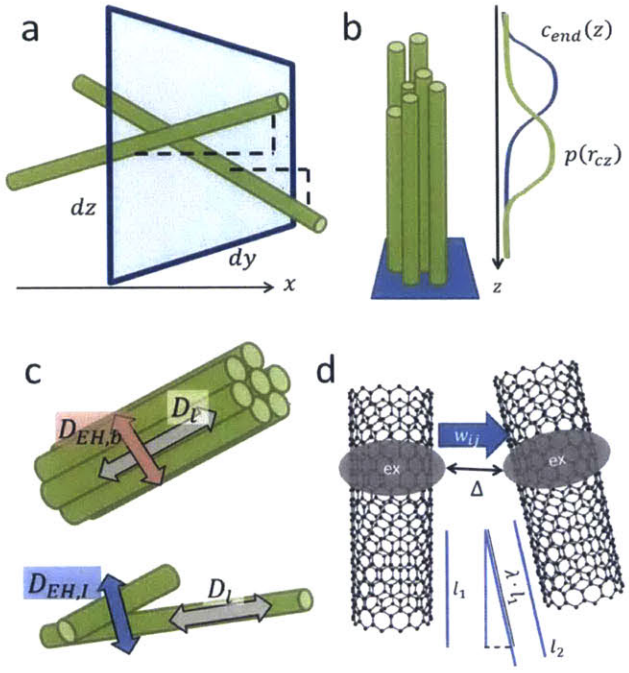


Figure 79. (a) cartoon of nanotubes penetrating the x-face of a differential volume balance. (b) an example cartoon of a vertical SWNT forest demonstrating construction of the SWNT end distribution. (c) illustration of diffusivity perpendicular to longitudinal axis at SWNT intersections and bundles. (d) illustration of tunneling process parameters.

Finally, we consider exciton transport between nanotubes. The mechanisms of exciton transfer/hopping (EH) between nanotubes of the same and different chirality, as well as the mechanisms of type II exciton dissociation at donor-acceptor interfaces, remain under investigation and debate.[113,115,117,179] Explicitly, in the model we treat EH as a tunneling process in bundles and at interconnects, neglecting admittedly vital phenomena such as exciton delocalization[113,180] and Schottky barriers/band bending.[181] What is important for the future use and evolution of this framework however is that the relevant independent parameters are available, allowing a different functional form to be introduced without contradicting the remainder of the model. This consideration highlights the extensibility of our method beyond our explicit form.

We assume that on our timescale of interest ($O(1\ ps)$) after a transition to a neighboring SWNT an exciton does not maintain momentum in the direction of the transition, yielding random walk character. The result of EH events is therefore a diffusive flux of excitons at inter-SWNT contacts perpendicular to the longitudinal axes (the normal between tubes), a crucial phenomenon for aligned films.[115] For a pair of points $(\boldsymbol{r}_i', \boldsymbol{r}_j')$ on two SWNT of orientations $\hat{\boldsymbol{l}}_i, \hat{\boldsymbol{l}}_j$ with separation (Figure 79d)

$$\Delta \equiv \left|\boldsymbol{r}_i' - \boldsymbol{r}_j'\right| - d_c \tag{205}$$

Fermi's golden rule provides a quadratic dependence of the single exciton elastic transition rate w_{EH} on the interaction potential between excitons at the two points, $V_{ij}(\Delta, \hat{\boldsymbol{l}}_i, \hat{\boldsymbol{l}}_j)$:

$$w_{EH} \propto \left|V_{ij}(\Delta, \hat{\boldsymbol{l}}_i, \hat{\boldsymbol{l}}_j)\right|^2. \tag{206}$$

For a tunneling process we anticipate exponential decay of rate with separation [113]

$$\left|V_{ij}(\Delta, \hat{\boldsymbol{l}}_i, \hat{\boldsymbol{l}}_j)\right|^2 \propto e^{-2\beta\Delta} \tag{207}$$

where β (length^{-1}) is the spatial decay constant for the equivalent monoexponential decay of the exciton wave function radially away from the SWNT, e.g.,[xxxv] $\left|\psi_{ex}\right|^2 \propto e^{-2\beta r_\perp}$. The relative alignment of the two SWNT should linearly attenuate the interaction potential by momentum conservation,

$$w_{EH} \propto \left|\hat{\boldsymbol{l}}_i \cdot \hat{\boldsymbol{l}}_j\right|^2. \tag{208}$$

Given the exponential decay of the transition rate and the strong van der Waals attraction between SWNT, we approximate that EH occurs only at intimate SWNT contacts with some fixed $\Delta = \underline{\Delta}$. This approximation allows us to define a fixed proportionality constant k_{EH} such that

$$\lambda \equiv \left|\hat{\boldsymbol{l}}_i \cdot \hat{\boldsymbol{l}}_j\right| \tag{209}$$

$$w_{EH} = k_{EH}\lambda^2. \tag{210}$$

k_{EH} can be determined experimentally or from theoretical evaluation.

From the transition rate at contacts we can derive the resulting diffusion coefficients. In the 1D axis along the transition direction (perpendicular to both SWNT longitudinal axes $\hat{\boldsymbol{l}}_i, \hat{\boldsymbol{l}}_j$, i.e. $\hat{\boldsymbol{l}}_i \times \hat{\boldsymbol{l}}_j$) the diffusivity can be expressed as a product of the velocity and mean free path as in Equation (194). Since each transition covers the effective distance $\underline{\Delta} + d_c$ and are independent events, the effective velocity is $w_{EH}(\underline{\Delta} + d_c)$ and mean free path is $\underline{\Delta} + d_c$, giving us

$$D_{EH} = \frac{1}{2} \cdot w_{EH}(\underline{\Delta} + d_c) \cdot (\underline{\Delta} + d_c) = \frac{1}{2} k_{EH}\lambda^2(\underline{\Delta} + d_c)^2. \tag{211}$$

The network diffusion resulting from this transport at contacts is intimately dependent on the microscopic heterogeneity of the film; both the co-alignment and contact density (correlated position) distributions, rather than mean values, determine the resulting film EH diffusivity. For example, a film comprised entirely of bundles – which have large contact areas and near-perfect alignment – will exhibit enhanced EH diffusion relative to an isotropic film. As another example, a network with larger voids (at the same

[xxxv] For a square tunneling barrier between SWNT this is rigorous, but in practice even with no interstitial contaminants the image potential in each SWNT gives curvature to the tunneling barrier. The wavefunction decay therefore won't be exactly monoexponential, but will always be rapid.

total density) will exhibit more efficient EH transport due to higher correlation of positions (and therefore a higher density of contacts).

We divide the film into two distinct microscopic environments: bundles and interconnects. The bundling coefficient, b_c, we define as the length fraction of SWNT in bundles. Along with the mean number of SWNT in a bundle, M_b, we can determine the diffusivity due to bundling (again, neglecting exciton delocalization[113]). In a close-packed bundle, a single SWNT can be surrounded by anywhere from 1 to 6 neighbors, and each neighbor of the same chirality provides the pairwise diffusion pathway described in (211). For χ nearest neighbors around each SWNT, and $\lambda = 1$ in bundles, the diffusivity in the plane perpendicular to the bundle would be

$$D_{EH,b} = \chi \frac{1}{4} k_{EH,b} (\Delta_B + d_c)^2. \tag{212}$$

If we define χ_i as the probability of a member of the bundle having i neighbors, then we can determine χ as

$$\chi = \sum_{i=1}^{6} i \cdot \chi_i. \tag{213}$$

We can calculate the mean value of χ from only the average bundle size. As $M_B \to \infty, \chi \to 6$, giving us a functional form (with constants n and Q)

$$\chi(M_B) = \frac{6 \cdot M_B^n - Q}{M_B^n}. \tag{214}$$

At a minimum bundle size of 2 we must have $\chi = 1$, giving us

$$\chi(2) = 1 \to Q = 5 \cdot 2^n. \tag{215}$$

From the series of maximal $\chi(M_B)$ we can evaluate n as ~0.51,

$$\chi(M_B) = \frac{6 \cdot M_B^{0.51} - 5 \cdot 2^{0.51}}{M_B^{0.51}}. \tag{216}$$

With diffusivity $D_{EH,b}$ (Equation (69)) in the plane perpendicular to a bundle, we have the resulting exciton flux in Cartesian coordinates

$$\begin{aligned} &b_c(1 - \sin\theta\cos\phi)^2 D_{EH,b} \frac{dc}{dx} + b_c(1 - \sin\theta\sin\phi)^2 D_{EH,b} \frac{dc}{dy} \\ &+ b_c(1 - \cos\theta)^2 D_{EH,b} \frac{dc}{dz} \end{aligned} \tag{217}$$

Outside of bundles, exciton transport between nanotubes is still expected at interconnects between SWNT. The resulting flux will in general be lower than bundled SWNT EH due to 1) imperfect alignment ($\lambda < 1$) and 2) lower overlap lengths (in bundles the entire length of the SWNT overlap,

giving $b_c\rho_{\langle l \rangle}$ length of SWNT in bundles per volume, whereas at interconnects only a small region of length $\sim d_c$ is contacted between SWNT.

An ideal treatment of the number of SWNT contacts per volume would include the correlation of nanotube locations, for example by deriving it from the nanotube radial distribution function (See Appendix *Breakdown of Macroscopic Homogeneity* below). As discussed in that section however, a good first approximation is to assume an independent dispersion (i.e. uniform distribution). We can determine the density of interconnects by considering a control volume surrounding a nanotube of length l (Figure 80).

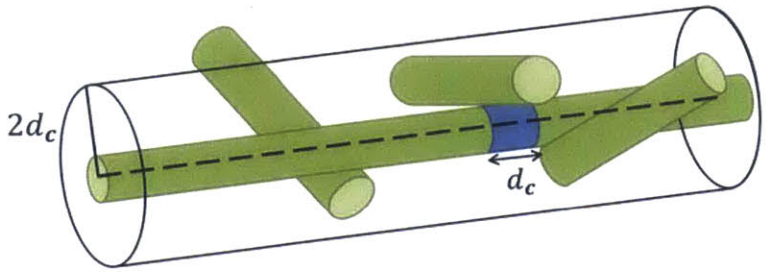


Figure 80. Cartoon of a control volume around an unbundled nanotube.

The number of other SWNT with central axes in the volume up to d_c away is the number of interconnects in the control volume (whether the locations are correlated or not, see Appendix). In other words, the number of contacts is the number of other nanotubes within one diameter of the control nanotube. To roughly approximate the bending ability of the nanotubes (deviation from rigid rods) the contact distance can be extended to twice the mean displacement r_{bend}

(218) $$d_c + 2r_{bend},$$

for example an additional $\sim d_c$. The volume excluding the central nanotube is then

(219) $$l\left(\pi(2d_c)^2 - \pi\left(\frac{d_c}{2}\right)^2\right) = l\frac{15}{4}\pi d_c^2.$$

In that volume there are, based on the number of SWNT per volume ρ,

(220) $$\left(1 - b_c\left(1 - \frac{1}{M_B}\right)\right)\rho \cdot l\frac{15}{4}\pi d_c^2$$

other SWNT present, accounting for SWNT pulled into bundles.[xxxvi] This yields the number of intersections per length of the control nanotube,

[xxxvi] i.e. each bundle itself can behave as a nanotube, providing the $\left(1 - \frac{1}{M_B}\right)$ coefficient.

$$n_I = \frac{1}{l}\cdot\left(1-b_c\left(1-\frac{1}{M_B}\right)\right)\rho\cdot l\frac{15}{4}\pi d_c^2$$
$$= \left(1-b_c\left(1-\frac{1}{M_B}\right)\right)\rho\frac{15}{4}\pi d_c^2. \quad (221)$$

Multiplying by the length of SWNT not in bundles per volume we have

$$\left(1-b_c\left(1-\frac{1}{M_B}\right)\right)\rho_{\langle l\rangle}\cdot\left(1-b_c\left(1-\frac{1}{M_B}\right)\right)\rho\frac{15}{4}\pi d_c^2$$
$$= \rho_{\langle l\rangle}\rho\left(1-b_c\left(1-\frac{1}{M_B}\right)\right)^2\frac{15}{4}\pi d_c^2 \quad (222)$$

intersections per volume of solar cell. If each intersection to provides $\sim d_c$ of contact length along the SWNT, the length of intersections per volume is

$$\rho_{\langle l\rangle}\rho\left(1-b_c\left(1-\frac{1}{M_B}\right)\right)^2\frac{15}{4}\pi d_c^3. \quad (223)$$

For each intersection, we have the diffusivity expression (211),

$$D_{EH,I} = \frac{1}{2}k_{EH}\lambda^2\left(\underline{\Delta}+d_c\right)^2, \quad (224)$$

providing a flux in the dimension along the intersection $\hat{\boldsymbol{l}}_i\times\hat{\boldsymbol{l}}_j$

$$J_{EH,I,\hat{l}_i\times\hat{l}_j} = D_{EH,I}\frac{dn}{dx}. \quad (225)$$

Multiplying by the number of intersections in a cross-sectional area we arrive at the film fluxes arising from interconnects,

$$J_{EH,I,x} = (1-\sin\theta\cos\phi)^2\cdot\rho\left(1-b_c\left(1-\frac{1}{M_B}\right)\right)^2\frac{15}{4}\pi d_c^3\cdot D_{EH,I}\frac{dc}{dx} \quad (226)$$

$$J_{EH,I,y} = (1-\sin\theta\sin\phi)^2\cdot\rho\left(1-b_c\left(1-\frac{1}{M_B}\right)\right)^2\frac{15}{4}\pi d_c^3\cdot D_{EH,I}\frac{dc}{dy} \quad (227)$$

$$J_{EH,I,z} = (1-\cos\theta)^2\cdot\rho\left(1-b_c\left(1-\frac{1}{M_B}\right)\right)^2\frac{15}{4}\pi d_c^3\cdot D_{EH,I}\frac{dc}{dz} \quad (228)$$

where the leading trigonometric terms are again projections onto the plane perpendicular to the control nanotube longitudinal axis, and we employed relation (193). To account for significant reductions in inter-SWNT contacts arising from nanotube coatings (e.g., polymer, surfactant), an interstitial modifier A_I can be included to attenuate the number of interconnects with the proportional factor $1-A_I$,

$$J_{EH,I,x} = (1-A_I)\cdot(1-\sin\theta\cos\phi)^2\cdot\rho\left(1-b_c\left(1-\frac{1}{M_B}\right)\right)^2\frac{15}{4}\pi d_c^3\cdot D_{EH,I}\frac{dc}{dx} \quad (229)$$

$$J_{EH,I,y} = (1 - A_I) \cdot (1 - \sin\theta \sin\phi)^2 \cdot \rho \left(1 - b_c \left(1 - \frac{1}{M_B}\right)\right)^2 \frac{15}{4} \pi d_c^3 \cdot D_{EH,I} \frac{dc}{dy}$$

$$J_{EH,I,z} = (1 - A_I) \cdot (1 - \cos\theta)^2 \cdot \rho \left(1 - b_c \left(1 - \frac{1}{M_B}\right)\right)^2 \frac{15}{4} \pi d_c^3 \cdot D_{EH,I} \frac{dc}{dz}.$$

This adjustment is not necessary in bundles, as the definition of the bundle coefficient explicitly rejects lengths of SWNT that are not in intimate contact.

The final consideration in exciton diffusion at contacts is the co-alignment (λ) dependence. Since no other term in our balance will depend on the relative orientation,

(230) $$\lambda \equiv \left|\hat{\boldsymbol{l}}' \cdot \hat{\boldsymbol{l}}''\right| = \lambda(\theta', \theta'', \phi', \phi''),$$

integrating over $p(\lambda)$ yields

(231) $$D_{EH,I} = \frac{1}{2} k_{EH} \left(\underline{\Delta} + d_c\right)^2 \cdot \langle\lambda^2\rangle.$$

$\langle\lambda^2\rangle$ is readily calculable given the distribution of $p(\hat{\boldsymbol{l}})$. For example two methods would be:

1) Integrate λ^2 over the distribution of angles directly

(232) $$\langle\lambda^2\rangle = \int\int\int\int \lambda^2 p(\theta') p(\theta'') p(\phi') p(\phi'') d\theta' d\theta'' d\phi' d\phi''$$

2) Evaluate the product distribution

(233) $$p(\lambda|\theta', \theta'', \phi', \phi'') = \delta\left(\lambda - \left|\hat{\boldsymbol{l}}' \cdot \hat{\boldsymbol{l}}''\right|\right)$$

$$\left|\hat{\boldsymbol{l}}' \cdot \hat{\boldsymbol{l}}''\right| = l_x' \cdot l_x'' + l_y' \cdot l_y'' + l_z' \cdot l_z''$$

(234) $$= |\sin\theta' \cos\phi' \sin\theta'' \cos\phi'' + \sin\theta' \sin\phi' \sin\theta'' \sin\phi'' + \cos\theta' \cos\theta''|$$

$$p(\lambda)$$

(235) $$= \int_0^{2\pi} \int_0^{2\pi} \int_0^{\pi} \int_0^{\pi} p(\lambda|\theta', \theta'', \phi', \phi'') p(\theta') p(\theta'') p(\phi') p(\phi'') d\theta' d\theta'' d\phi' d\phi''.$$

The expected value is then

(236) $$\langle\lambda^2\rangle = \int_0^1 \lambda^2 p(\lambda) d\lambda.$$

Including all three diffusion mechanisms, the volume balance for exciton transport in a monochiral network is

(237) $$Let\ \gamma_I \equiv \rho \left(1 - b_c \left(1 - \frac{1}{M_B}\right)\right)^2 \frac{15}{4} \pi d_c^3 = \left(1 - b_c \left(1 - \frac{1}{M_B}\right)\right) d_c \cdot n_I$$

$$\frac{dc}{dt} = N(\boldsymbol{r}) + \Big(D_l(\sin\theta\cos\phi)^2 + (1-\sin\theta\cos\phi)^2\left((1-A_I)\gamma_I D_{EH,I} + b_c D_{EH,b}\right)\Big)\frac{d^2c}{dx^2} + \Big(D_l(\sin\theta\sin\phi)^2 + (1-\sin\theta\sin\phi)^2\left((1-A_I)\gamma_I D_{EH,I} + b_c D_{EH,b}\right)\Big)\frac{d^2c}{dy^2} + \Big(D_l(\cos\theta)^2 + (1-\cos\theta)^2\left((1-A_I)\gamma_I D_{EH,I} + b_c D_{EH,b}\right)\Big)\frac{d^2c}{dz^2} - k_\Gamma c - c\sum_{im} k_{im}n_{im} - \frac{k_{EEA}}{\rho_{\langle l\rangle}}c^2 - \frac{k_{end}}{\rho_{\langle l\rangle}}c_{end}(\boldsymbol{r})\cdot c. \tag{238}$$

We can recognize γ_I as representing the 'sparsity' influence on orthogonal diffusivity, with $\gamma_I \to 0$ for a fully-bundled film (independent of density). Analogously to the light field balance, we can recognize that periodic results in x and y are trivial, simplifying the balance to the z coordinate,

$$\frac{dc}{dx} = \frac{dc}{dy} = 0 \tag{239}$$

$$\frac{dc}{dt} = N(z) + \Big((\cos\theta)^2 D_l + (1-\cos\theta)^2\left((1-A_I)\gamma_I D_{EH,I} + b_c D_{EH,b}\right)\Big)\frac{d^2c}{dz^2} - k_\Gamma c - c\sum_{im} k_{im}n_{im} - \frac{k_{EEA}}{\rho_{\langle l\rangle}}c^2 - \frac{k_{end}}{\rho_{\langle l\rangle}}c_{end}(z)\cdot c. \tag{240}$$

The balance is subject to steady state conditions and electrode boundary conditions, treated below. For simplicity we can now define a total network diffusion coefficient,

$$D_{tot} \equiv (\cos\theta)^2 D_l + (1-\cos\theta)^2\left((1-A_I)\gamma_I D_{EH,I} + b_c D_{EH,b}\right). \tag{241}$$

Analogous to the light absorption problem, we have constructed a differential equation describing exciton transport due to nanotubes with orientation $\hat{\boldsymbol{l}}$, but to get the total balance we must sum contributions from all possible orientations. Again, that weighted sum is equivalent to integrating (241) over the distribution $p(\hat{\boldsymbol{l}})$. This treatment is only valid under our approximation of a homogenous film. That is our method: treat single-SWNT chemical and geometric parameters as random variables, derive network behavior from single-SWNT physics as a function of those parameters, and finally integrate the solution over the distribution of those parameters. In the *Results* section we illustrate this process when we apply it to specific cases.

This is an important result; equation (240) has the intuitive reaction-diffusion form one might guess for an empirical fit of a given experimental solar cell, but

A) the net diffusivity D_{tot} and several of the kinetic rate constants depend exquisitely on the film properties ($p(\hat{l}), \rho_{\langle l \rangle}, p(l)$), preventing broader applicability of any single diffusivity measurement, and
B) by considering the distributions of single-SWNT parameters and the microscopic mechanisms of exciton transport and decay, we have been able to derive the dependence of network diffusion and reaction constants on film properties and fundamental single-exciton physical constants.

Furthermore, this framework built on treating single-SWNT parameters as random variables and defining the network through their distributions is flexible and adaptive; particular terms – such as exciton hopping diffusivities – can be changed as SWNT exciton physics is better understood. New parameters with associated distributions can be introduced, such as dielectric constant and local density (radial distribution function), and variable independence assumptions can be relaxed, increasing complexity and nuance without grossly deteriorating numerical solvency.

Multiple chiralities can now be introduced. Separate balances can be constructed for excitons residing on each chirality, $c_i(z)$, coupled by inter-SWNT interactions (exciton hopping, limited by band gap alignment rules). In applying the model we make the simplifying assumptions of homogenous chirality distributions (v_i independent of z), and chirality-independent length and orientation distributions; all of those assumptions can be relaxed without invalidating the remaining derivation. The total concentration of excitons is thus

$$c_i = v_i c \tag{242}$$

$$c(\boldsymbol{r}) = \sum_{i \in \{c_i\}} c_i(\boldsymbol{r}) = c(\boldsymbol{r}) \sum_{i \in \{c_i\}} v_i = c(\boldsymbol{r}). \tag{243}$$

Beginning with quenching due to metallic nanotubes (mSWNT), we denote v_m as the mSWNT fraction. Since nonradiative quenching on mSWNT is rapid compared to residence times, we do not need to construct a concentration balance for them, and the impact of EH to mSWNT is an exciton loss term for all other chiralities. Along the length of an unbundled SWNT every contact with mSWNT quenches excitons at rate

$$k_{EH}\langle \lambda^2 \rangle \cdot n. \tag{244}$$

The length of intersection per length SWNT is the intersections per length times the length of each intersection,

$$n_{l,m} \cdot d_c, \tag{245}$$

where $n_{l,m}$ is the number of intersections with a metal SWNT per length SWNT, which is related to our earlier quantity n_l number of intersections total per length SWNT (Equation (71)) via

(246) $$n_{I,m} = v_m n_I = v_m \left(1 - b_c\left(1 - \frac{1}{M_B}\right)\right)\rho\frac{15}{4}\pi d_c^2.$$

Multiplying the rate per intersection times the length of intersection per length SWNT gives the total quenching rate for a single-nanotube 1D system

(247) $$-k_{EH}\langle\lambda^2\rangle \cdot v_m n_I \cdot d_c \cdot n.$$

We extend to the film quenching rate by multiplying by the length of unbundled mSWNT per cell volume,

(248) $$\begin{aligned} &-\left(1 - b_c\left(1 - \frac{1}{M_B}\right)\right)\rho_{\langle l\rangle} \cdot k_{EH}\langle\lambda^2\rangle \cdot v_m n_I \cdot d_c \cdot n \\ &= -\left(1 - b_c\left(1 - \frac{1}{M_B}\right)\right) \cdot k_{EH}\langle\lambda^2\rangle \cdot v_m n_I \cdot d_c \cdot c \\ &= -k_{EH}\langle\lambda^2\rangle \cdot v_m \gamma_I c \end{aligned}$$

where in the final equality we substitute in definition (237).

Note that d_c varies by chirality, making the relevant control volume chirality-dependent; properly, the coupling term for EH to/from chirality j at interconnects in the balance of excitons on chirality i should have in Equation (219) $d_c \to (d_{c,i} + d_{c,j})$ in the first term $d_{c,i}$ in the second term, i.e.

(249) $$l\left(\pi\left(d_{c,i} + d_{c,j} + 2r_{bend}\right)^2 - \pi\left(\frac{d_{c,i}}{2}\right)^2\right).$$

For practical purposes however when diameters are similar the effect is small compared to r_{bend}.

To treat mSWNT quenching in bundles, consider the number of mSWNT per cross-sectional area of a bundle,

(250) $$v_m \cdot \frac{2}{\sqrt{3}} d_c^{-2}.$$

By definition of the diffusion coefficient, excitons in a cross-section cover $D_{EH,b}$ area per time, yielding an exciton quenching rate (mean rate of arrival to a mSWNT) of

(251) $$-D_{EH,b} \cdot v_m \cdot \frac{2}{\sqrt{3}} d_c^{-2} \cdot n$$

per length of SWNT. Multiplying by the length of SWNT in bundles per volume yields the film quenching rate due to mSWNT in bundles

(252) $$-b_c \rho_{\langle l\rangle} \cdot D_{EH,b} v_m \frac{2}{\sqrt{3}} d_c^{-2} \cdot n = -b_c \cdot D_{EH,b} v_m \frac{2}{\sqrt{3}} d_c^{-2} \cdot c.$$

We can now construct exciton concentration balances for each chirality, $c_i(z)$. In addition to all of the monochiral balance (240) and mSWNT quenching terms ((248) and (252)), the balances for each chirality are coupled to one another through EH interactions in bundles and interconnects. Recall that we

have chosen indices such that band gap ($E_i \equiv E_{1u,i}$) and index are monotonically increasing ($E_{i+1} > E_i$). Because transitions are elastic, exciton transitions from higher bandgaps to lower bandgaps occur at rates analogous to mSWNT quenching; transitions 'up' bandgaps may be thermally activated for small band gaps, in which case they are attenuated by an Aarhenius factor,

$$j > i, \qquad k_{EH} \to k_{EH} e^{-\frac{E_j - E_i}{k_B T}}. \tag{253}$$

In the balance for chirality i, we have exciton loss due to unbundled coupling,

$$-c_i \sum_{j<i} \left(1 - b_c \left(1 - \frac{1}{M_B}\right)\right) k_{EH} \langle \lambda^2 \rangle \cdot v_j n_l \cdot d_c = -c_i \sum_{j<i} k_{EH} \langle \lambda^2 \rangle \cdot v_j \gamma_I , \tag{254}$$

exciton gain due to unbundled coupling,

$$+\sum_{j>i} \left(1 - b_c \left(1 - \frac{1}{M_B}\right)\right) k_{EH} \langle \lambda^2 \rangle \cdot v_j n_l \cdot d_c \cdot c_j = +\sum_{j>i} k_{EH} \langle \lambda^2 \rangle \cdot v_j \gamma_I \cdot c_j, \tag{255}$$

exciton loss due to bundled coupling,

$$-c_i \sum_{j<i} b_c \cdot D_{EH,b} v_j \frac{2}{\sqrt{3}} d_c^{-2}, \tag{256}$$

and exciton gain due to bundled coupling,

$$+\sum_{j<i} b_c \cdot D_{EH,b} v_j \frac{2}{\sqrt{3}} d_c^{-2}\ c_j. \tag{257}$$

Note again that we are neglecting band bending and Schottky barriers at the interfaces of SWNT, which may play a significant role in transition rates;[117,181] as understanding of such phenomena is reached, the coupling expressions (254)-(257) can be changed appropriately.

The final expression for multichiral network exciton transport is then the system of ordinary differential equations

$$\frac{dc_i}{dt} = v_i N(z) + D_{tot}\frac{d^2c_i}{dz^2} - k_\Gamma c_i - c_i\sum_{im} k_{im}n_{im} - \frac{k_{EEA}}{\rho_{\langle l\rangle}}c_i^2$$
$$- \frac{k_{end}}{\rho_{\langle l\rangle}}c_{end}(z)\cdot c_i - k_{EH}\langle\lambda^2\rangle\cdot v_m\gamma_I c_i$$
$$- b_c D_{EH,b} v_m \frac{2}{\sqrt{3}} d_c^{-2}\cdot c_i$$
$$- c_i\sum_{j<i}\left(k_{EH}\langle\lambda^2\rangle\cdot v_j\gamma_I + b_c\cdot D_{EH,b}v_j\frac{2}{\sqrt{3}}d_c^{-2}\right)$$
$$+ \sum_{j>i} c_j\cdot\left(k_{EH}\langle\lambda^2\rangle\cdot v_j\gamma_I + b_c\cdot D_{EH,b}v_j\frac{2}{\sqrt{3}}d_c^{-2}\right)$$

(258)

for each chirality i.

The system of ODEs (258) can be solved at steady state subject to boundary conditions at the electrodes ($z = 0, T$). A variety of electrode configurations can be considered (Figure 78c); in proceeding we focus on two electrodes that can drive type II exciton dissociation – one hole accepting and one electron accepting. For each electrode we therefore have Robin boundary conditions from the exciton splitting rate,

$$\left.\frac{dc_i}{dz}\right|_{z=0} = \frac{k_{d0,i}}{D_{tot}}c_i(z=0) \quad (259)$$

$$\left.\frac{dc_i}{dz}\right|_{z=T} = -\frac{k_{dT,i}}{D_{tot}}c_i(z=T). \quad (260)$$

where $k_{d0,i}$ and $k_{dT,i}$ are the dissociation rate constants that in general could depend on the chirality i. In applying the model in this work we approximate perfect collection efficiency, i.e. $k_{d0}, k_{dT} \to \infty$, $c_i(z = 0, T) \to 0$, and check convergence to a finite value of $k_{d0,dT}c(z = 0, T)$.

Charge Transport

Free carrier transport in the nanotube network is in many ways more challenging than exciton diffusion. In addition to the analogous geometric effects on film charge mobility and diffusivity, charge transfer at the interface of nanotubes, particularly of different chiralities, can block or trap charges.[181] Asymmetric electron and hole generation (at the two electrodes) and slow mobilities can lead to an electric field development that feeds back on the exciton dissociation problem, inhibiting dissociation and/or driving spontaneous in-film dissociation when the field strength exceeds the binding energy. Finally carrier densities can significantly impact mobility.[182]

In this work we do not attempt to accurately describe free carrier transport in the SWNT film. Rather, we proceed under the large simplification that performance will be limited by exciton diffusion, due to high longitudinal mobilities in any intrinsic electric field.[182] This assumption reduces coupling between

the free carrier and exciton problems to the exciton dissociation rate. We also continue to neglect any charge-transfer effects at inter-SWNT contacts, as those effects have been neither theoretically nor experimentally defined. These rough assumptions allow, as we show, important results in the optimization of film performance, but prohibit accurate quantitative prediction of external quantum efficiency (EQE). In future work, we will develop a more complete consideration of free carrier transport.

A variety of electrode materials and configurations are possible (Figure 78c), the choice of which impacts charge carrier transport profoundly, just as in any excitonic photovoltaic system. In this work we focus on an illustrative case of one electron-collecting and one hole-collecting electrode, arbitrarily choosing $z = 0$ to be electron-collecting (Figure 81).

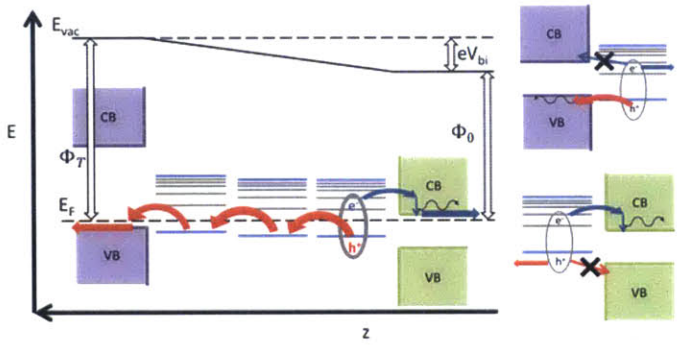


Figure 81. cartoon of the free carrier generation and translation process at the electron-collecting electrode, where band alignment is chosen such that holes are blocked and electron relaxation in the electrode conduction band compensates for the binding energy of the nanotube. Intrinsic work function differences between the two electrodes inspire a potential drop of $\Phi_T - \Phi_0$.

We estimate the geometric impact on valence and conduction band carrier mobilities in the film from the exciton behavior. We attenuate single-SWNT electron and hole mobilities, μ_e^* and μ_h^*, by the network attenuation of single-SWNT exciton diffusivities,

$$D_e = D_e^* \frac{D_l}{D_{tot}} = \mu_e^* \frac{k_B T}{q} \frac{D_{tot}}{D_l} \tag{261}$$

$$D_h = D_h^* \frac{D_l}{D_{tot}} = \mu_h^* \frac{k_B T}{q} \frac{D_{tot}}{D_l} \tag{262}$$

where we have applied the Einstein relation at low bias, q is the elementary charge, k_B is Boltzmann's constant, and (here only) T is the isothermal film temperature. Reversing the Einstein relation we extract the film mobilities,

$$\mu_e = \frac{q}{k_B T} D_e \tag{263}$$

$$\mu_h = \frac{q}{k_B T} D_h. \tag{264}$$

We now balance the non-equilibrium populations of conduction band carriers, $f_e(\boldsymbol{r})$, and valence band carriers, $f_h(\boldsymbol{r})$, which are coupled to one another and the electric field $\mathcal{E}_z(z)$. We again identify that in any practical case

$$\frac{df_e}{dx} = \frac{df_e}{dy} = \frac{df_h}{dx} = \frac{df_h}{dy} = 0 \tag{265}$$

and our problem is confined to the z axis. For each carrier there are drift and diffusion fluxes,

$$J_e^{diff} = -D_e \frac{df_e}{dz} \tag{266}$$

$$J_e^{drift} = -\mu_e \mathcal{E}_z \cdot f_e \tag{267}$$

$$J_h^{diff} = -D_h \frac{df_h}{dz} \tag{268}$$

$$J_h^{drift} = \mu_h \mathcal{E}_z \cdot f_h. \tag{269}$$

The fluxes contribute to a population balance on a differential volume as

$$\frac{df_e}{dt} = -\frac{dJ_e^{dif}}{dz} - \frac{dJ_e^{drift}}{dz} = D_e \frac{d^2 f_e}{dz} + \mu_e \left(\mathcal{E}_z \frac{df_e}{dz} + f_e \frac{d\mathcal{E}_z}{dz} \right) \tag{270}$$

$$\frac{df_h}{dt} = -\frac{dJ_h^{dif}}{dz} - \frac{dJ_h^{drift}}{dz} = D_h \frac{d^2 f_h}{dz} - \mu_h \left(\mathcal{E}_z \frac{df_h}{dz} + f_h \frac{d\mathcal{E}_z}{dz} \right). \tag{271}$$

Interacting excess charges also threaten recombination, coupling the differential equations,

$$-k_{e-h} f_e f_h. \tag{272}$$

Equations (78) and (79) have boundary conditions defined at the electrodes. At the hole collecting electrode $z = T$ electrons are generated in the film with no possibility of collection, coupling to the exciton dissociation rate,

$$\left(-J_e^{diff} - J_e^{drift}\right)\Big|_{z=T} = \sum_i k_{dT,i} \cdot c_i(z = T). \tag{273}$$

At $z = 0$ the equivalent process occurs for holes,

$$\left(-J_h^{diff} - J_h^{drift}\right)\Big|_{z=0} = -\sum_i k_{d0,i} \cdot c_i(z = 0). \tag{274}$$

At the opposite boundary, carrier collection can be represented by a Robin boundary condition,

$$\left(-J_e^{diff} - J_e^{drift}\right)\Big|_{z=0} = k_{abs} \cdot f_e(z = 0) \tag{275}$$

$$\left(-J_h^{diff} - J_h^{drift}\right)\Big|_{z=T} = -k_{abs} \cdot f_h(z = T). \tag{276}$$

For perfect collection efficiency we have $k_{abs} \to \infty$, which we can take as arbitrarily high and check that $k_{abs} f_e(\boldsymbol{z} = 0, T)$ converges.

The electric field evolves in the z axis from the gradient in excess charge populations,

(277) $$\frac{d\mathcal{E}_z}{dz} = \frac{q}{\varepsilon}\left(f_h(z) - f_e(z)\right),$$

where ε is the permittivity of the film. The intrinsic bias across the film from mismatched electrode work functions,

(278) $$\mathcal{E}_z(z = 0) = T \cdot V_{bi}$$

(279) $$V_{bi} \equiv \Phi_0 - \Phi_T,$$

forms the boundary condition to (80), where Φ_0 and Φ_T are the work functions of materails at $z = 0$ and T respectively.

Coupled to (80), the balances (78) and (79) can be solved numerically at steady state, subject to boundary conditions (273)-(276) and (278). From the resulting carrier fluxes at $z = 0$ and $z = T$ we have the short circuit current and EQE.

Appendix B. Breakdown of Macroscopic Homogeneity

In our evaluation of the absorption cross-section in the light absorption problem and exciton hopping at nanotube contacts in the exciton transport problem we adopted an assumption of homogeneity at a macroscopic scale (on the order of the exciton longitudinal diffusion length). In certain reasonable film morphologies, however, there are significant variations in film density at the micron scale, even for unbundled SWNT. For example, in vacuum filtration or spin-coat deposition (Figure 82a), large voids are present. Essentially, the distribution $p(\boldsymbol{r}_c)$ is not uniform even at the macroscopic scale. The result of deviations from uniformity is that a mean film density, $\rho_{\langle l \rangle}$, is lower than the local density around a given nanotube, $\rho_{\langle l \rangle}^{loc}$. Viewed another way, the position of a given nanotube is not independent of the position of other nanotubes, but rather is correlated – a nanotube is more likely to be in some proximity to another nanotube than would be the case if their locations were independent (Figure 82b). Ideally, we would consider a radial distribution function in the plane perpendicular to the SWNT longitudinal axis, describing the probability density of another tube being at a given separation.

Above the percolation threshold density (65% close-packed density for an aligned film), the morphology of the film does not matter as conduction paths are relatively constant. This has been experimentally verified by *Maillaud et al.[199]* Below the percolation threshold, morphology plays a substantial role in film conductivity.[199]

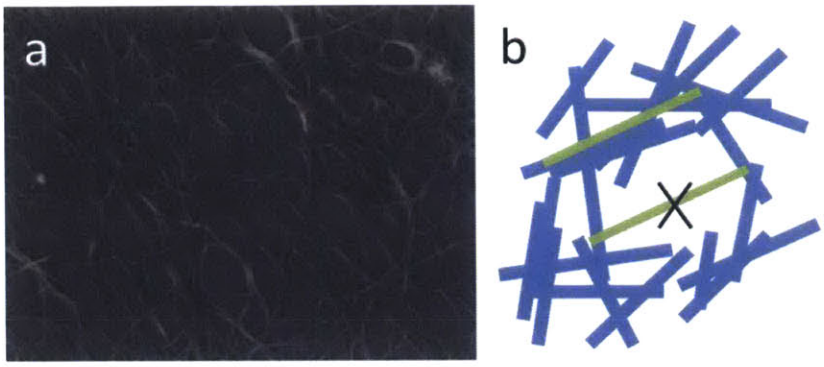


Figure 82. a) SEM image of spin-coated film of SWNT deposited from a sodium-dodecyl-sulfate solution. b) cartoon illustrating the impact of heterogeneity; a given/control nanotube (green) is not equally likely to be anywhere, but rather more likely to be close to other nanotubes.

In the light absorption problem, the effect of the error is to overestimate absorption when using a uniform $\rho_{\langle l \rangle}$; the correlation between SWNT reduces the light available at higher z in regions where SWNT is present, and no absorption occurs in voids. In the exciton transport problem, the effect of the error is to underestimate interconnect diffusivity, as the number of SWNT contacts is in reality higher. One solution to that problem is to use a SWNT-SWNT radial distribution function, $\rho(r)$, instead of a mean value of the number density, ρ, where r is the distance away from a SWNT central axis and $\rho(r)$ is the probability density that another SWNT is present at that separation. The number of contacts would then be integrated over the desired distance, rather than a fixed value,

$$\int_{d_c/2}^{d_c} \frac{1}{\frac{4}{3}\pi(2d_c)^3}\rho(r)\cdot\left(1-b_c\left(1-\frac{1}{M_B}\right)\right)\cdot l\frac{15}{4}\pi d_c^2\,dr \tag{280}$$

Fortunately, even qualitative consideration suggests that situations where this effect becomes substantial are also situations of least interest for investigation. First, high density films, within an order of magnitude of $\rho_{\langle l \rangle}^{CP}$, where this effect is minimal or nonexistent, are the films of interest for performance maximization; as we show, and has been experimentally indicated,[115] and is easy to intuit, higher density films yield higher EQE. Second, as we move towards aligned films where bundling is endemic, the exciton hopping impact of heterogeneity is mitigated.

We can also quantitatively consider the magnitude and conditions of this error. In terms of scaling, the number of interconnects increases linearly with density, and $\rho_{\langle l \rangle} < \rho_{\langle l \rangle}^{loc} < \rho_{\langle l \rangle}^{CP}$. Unless the void fraction is greater than 99%, in which case the film performance will be weak anyway, the number interconnects will be less than two orders of magnitude smaller than reality. While that seems like a large change, comparison of the diffusivities estimated puts that into perspective

$$D_{EH,I} \to O(10^{-13}\ m^2/s) \tag{281}$$

$$D_{EH,b} \rightarrow O(10^{-8}\, m^2/s)$$
$$D_l \rightarrow O(10^{-4} m^2/s).$$

Unless the density itself is high – in which case the void fraction is reduced – the diffusivity at interconnects plays a small role compared to any other available exciton diffusion pathway, even if bundling fractions are below 1%. In other words, the higher the void fraction, the less important SWNT interconnects become to exciton diffusion.

To relax our uniformity assumption, either simple approximations or rigorous measurement could be employed. In the former case, SEM and AFM images could be used to estimate the excess void volume in the film, and the density could be corrected where appropriate. For example, for a 30% void volume,

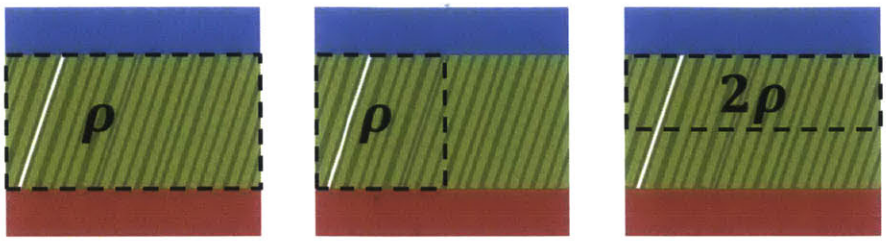
$$\rho_{\langle l \rangle}^{loc} \cong v^{loc} \rho_{\langle l \rangle}$$
$$v^{loc} = \frac{1}{1 - 0.3}. \quad (282)$$

More rigorously, X-Ray Diffraction (XRD) could be performed on the film to get the carbon-carbon radial distribution function. To extract the SWNT-SWNT radial distribution function from this, the intra-SWNT XRD spectrum must be either simulated or measured, and then subtracted from the film spectrum. This process is challenging and complicated.

Appendix C. Anisotropy of Number Density

ρ, the mean number of SWNT per volume, we employ as a defining parameter for the film. In experimental application of the present work, it is only necessary in determining $\rho_{\langle l \rangle}$, and other empirical quantities could be measured and used. In adjustments to this derivation, however, care should be taken if using ρ in any microscopic and some macroscopic scenarios for non-isotropic films. The quantity is calculated based on the total number of SWNT in the total volume. As the volume is reduced it may not scale appropriately. In a truly isotropic film, ρ can accurately describe the number of SWNT in a given volume all the way down to the point where microscopic heterogeneity arises ($O(10\ nm)$ scale). In an aligned film however, the situation depends on how the volume is shaped. Imagine a sandwiched forest of vertical SWNT. ρ is the number of SWNT divided by the total volume. If we bisect the forest vertically and consider half the volume, reducing the number of SWNT by half and the volume by half, our number density holds (Figure below). If we instead bisect it horizontally however, we cut the volume in half, but the total number of SWNT in each half hasn't changed, and our number density is grossly inaccurate. In treating SWNT interconnects we are saved because by definition the SWNT are not aligned in this case –

they are essentially isotropic – and the bundled SWNT scenario relies not on ρ but rather $\rho_{\langle l \rangle}$, which *always* scales appropriately with volume. Note that $\rho_{\langle l \rangle}$ always scales appropriately with volume.



Appendix D. Consideration of Rayleigh Scattering and Photoluminescence

In the derivation above we opted to treat absorption as the only light-matter interaction mechanism of concern, neglecting Rayleigh scattering of light within the film and reemission from radiative decay of excitons (photoluminescence, PL). Making those assumptions has clear benefits for tractability and computation. Omitting reflection allowed us to remove $\boldsymbol{k}$ from our calculations, and omitting PL greatly simplified coupling to the exciton transport equations. The error introduced by neglecting these phenomena bears consideration however.

Fluorescence requires that an exciton radiatively decay before being quenched or diffusing to the electrodes. To achieve the latter in substantial quantities case the film must be thick, with T on the order of the radiative diffusion length, (see Results section for parameter estimates)

$$L_\Gamma = \sqrt{D_l \tau_\Gamma} \cong \sqrt{10^{-2} m^2 s^{-1} \cdot 0.45 \times 10^{-9}\, s} = 2\ um. \tag{283}$$

As we see in the Results section, thicknesses on the order of a micron are rarely interesting. Furthermore, in the scenario when thickness is high, readsorption of the emitted light would be likely, making the net effect a dislocation of the exciton; a valid phenomenon to be sure, but a secondary one compared to primary absorption. To achieve a case where radiative decay could lead to a net loss of excitons, i.e. emitted light escaping, the film would have to be thin, in which case radiative decay would be slower than separation at the electrodes. Arguments for neglecting Rayleigh scattering are similar, the losses and gains of light are, by virtue of the scattering cross-section, a higher-order consideration.

To increase the nuance and complexity of the model, fluorescence and internal reflection could still be included. In Rayleigh scattering, the photon balance (Equation (62)) should include a term transforming $\boldsymbol{k}$ appropriately. To treat the fluorescent case we would include the term

$$+k_\Gamma c(\boldsymbol{r}) \cdot \delta\left(\hbar\omega - E_{1u,i}\right) \cdot \delta\left(\boldsymbol{\epsilon} - \hat{\boldsymbol{l}}\right)$$

where $c(\boldsymbol{r})$ is excitons/volume solar cell and $k_\Gamma = 1/\tau_\Gamma$ is the radiative decay rate of excitons, and $E_{1u,i}$ is the band gap energy of chirality i. This term would couple the light absorption problem to the exciton

concentration problem; given the small relative impact of fluorescence, a self-consistent method may be most efficient:

1. Solve the light absorption and exciton diffusion problems without including PL contributions to the light field.
2. Take the resulting $c(z)$, use it as an input ($\rightarrow c^{(0)}(z)$) into solving the light field problem with the fixed term
   $+k_\Gamma \cdot c^{(0)}(z) \cdot \delta(\hbar\omega - E_{1u,i})\delta(\epsilon - \hat{l})$,
   then solve the exciton concentration problem again.
3. Take that output, call it $c^{(1)}(z)$, and again plug it into the problem in the same manner.
4. Repeat the process until convergence ($c^{(n)}(z) \cong c^{(n-1)}(z)$.

Appendix E. Relaxing Isothermal Approximation

To relax the isothermal approximation and introduce temperature dependency to desired properties, a parallel set of ODEs solving the heat transfer problem must be coupled to our light absorption problem by the heat generated through nonradiaitve decay at the point of exciton generation, e.g.,

$$N(\boldsymbol{r}') \rightarrow N(\boldsymbol{r}', \omega) \rightarrow Q(\boldsymbol{r}') \cong \int_0^\infty N(\boldsymbol{r}', \omega) \cdot (\hbar\omega - E_0)\, d\omega \tag{284}$$

where Q is the rate of heat generation.

Appendix F. Including dielectric environment

The local dielectric environment has been shown to influence the size and diffusion coefficient of excitons.[100,101] That behavior is intuitive, as it is the poor dielectric screening in two of three dimensions in a nanorod like SWNT that gives rise to the strong exciton binding energies.[98,100,177,178] The impact on efficiency would be the same as adding impurities and end quenching (see Results) – it would perturb the balance between light absorption and diffusion, shifting the optimal thickness proportionally to the diffusion coefficient. Changing the dielectric would also change the absorption cross-section.[214] The present model can be made to explicitly include the local dielectric constant κ. The distribution, $p(\kappa)$, would be an additional defining property of the network, just like $p(l)$ or $\{v_i\}$. $p(\kappa)$ would include for example surfactant coverage, interstitial polymer coverage, and bundling (b_c, M_b). Relevant physical constants should then be made dependent on κ, for example $D_l \rightarrow D_l(\kappa)$.

Appendix G. Alternative methods of coupling the infinite number of line segments

The method we adopted for coupling the single-SWNT systems has no cost associated with it outside the breakdown of local (correlated) density considerations. Nevertheless it is worth mentioning two alternative approaches to treating the infinite system of line segments that might provide additional opportunities if explored.

One approach is to treat the time-evolution of a single exciton's probability distribution function, first along a single nanotube and then with multiple SWNT and exciton hopping, yielding an expression for $p(\boldsymbol{r}, t, c(\boldsymbol{r})|\boldsymbol{r}_0)$ where $\boldsymbol{r}_0$ is the position where the exciton is created, with excitons generated there at rate $N(\boldsymbol{r}_0)$. The dependence on concentration comes from the exciton-exciton annihilation (EEA), coupling the problem to other excitons at other positions. Whether coupling to other excitons or not (neglecting EEA), one can use that time-evolving PDF to calculate a diffusion length for the exciton depending on the $\boldsymbol{r}_0$ it originates at, $L_D(\boldsymbol{r})$. Using $L_D(\boldsymbol{r})$ and the generation rate, $N(\boldsymbol{r})$, one can calculate the rate at which excitons will reach the electrodes and the exciton concentration profile. I do not adopt this approach because we are interested in the system at steady state, making the laborious evaluation of the time evolution superfluous.

A second approach is to treat each SWNT as a 1D reaction-diffusion system, as we began with, but rather than using them in an exciton balance, instead leave them as an infinite set of differential equations, each corresponding to a different nanotube line segment, and couple them in 3 dimensions via exciton transfer. Each SWNT in the infinite series is indexed by the location of its center at $\boldsymbol{r}_c$, and each has pairwise coupling terms to each other SWNT, dependent on their $\boldsymbol{r}_c$ and $\boldsymbol{l}$. Because nanotubes are not infinitely long, there is a finite number of neighbors to each nanotube (e.g. in a bundle there can be only 6 neighbors per length of nanotube), i.e. the infinite sum of pairwise coupling terms drops off as the SWNT centers and orientations differ from one another. As a result, the infinite series of 1D ODEs can be taken over the entire crystal, somewhat like the treatment of crystals in solid-state physics (e.g. the construction of Bloch states from symmetry breaking). This approach can be elegant, but doesn't offer any practical gains over the more tractable one taken in this derivation.

Appendix H. Single-layer and multi-layer aligned films

As mentioned in the *Results* section above, two deposition paradigms are possible for fabrication of an aligned film. The aligned film can be grown or stacked in a single 'forest' or 'layer', which we call a single-layer (SL) film (Figure 83a). A SL film limits the possible thickness of the film, as

$$T < \langle l \rangle \cos\theta'. \tag{285}$$

It is also possible to contact a film without end-alignment, such as in a horizontally aligned film. This is the multi-layer (ML) case (Figure 83b).

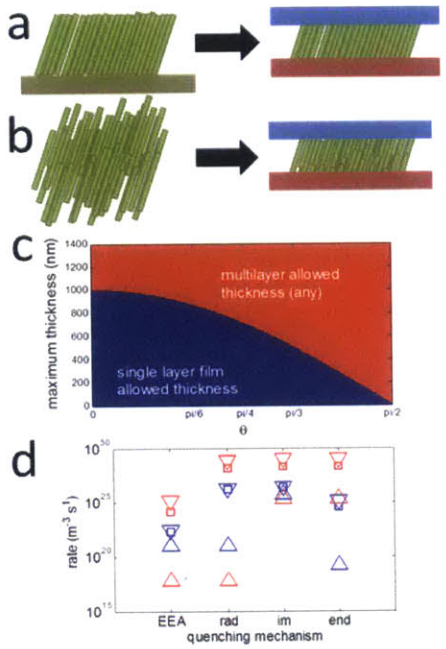


Figure 83. SL and ML film comparison. a) cartoon of a SL film, with both ends immersed in electrodes. b) cartoon of a ML film, with electrodes adhered onto a larger cake. c) allowed thickess of SL and ML films. d) comparison of quenching rates for horizontally aligned (red, $T = 200\ nm$, 80% close-packed density) and vertically aligned (blue, $T = 1$ um, 0.3% close-packed density). End quenching is not zero in the vertically aligned (SL) case because the thickness is on the order of $\langle l \rangle$. Squares are spatially-averaged means, triangles are minima and maxima values.

A SL film will always outperform a ML film all else being equal, due to lower end quenching. In practice, SL film electrodes are likely to encompass the ends of the SWNT in the SL film, leading to

$$c_{end}^{SL} = \rho^{2/3} p(l) \cos\theta' \cong 0 \tag{286}$$

where the approximation is unless the film thickness is within a standard deviation of the average SWNT length $\langle l \rangle$ (we find, however, that in optimization such thicknesses are rarely desirable). In contrast, in a ML film the end distribution is homogenous,

$$c_{end} = 2\rho. \tag{287}$$

In the cases we examine, rather than choose between the two morphologies we often allow for both – i.e. if the film thickness is low enough for a SL film to be permissible, we use a SL film, while if it

falls above that we use a ML film. In practice however the optimal film thicknesses never, except in the case of *very* low density or very low angle (e.g., horizontally aligned film), approach the limit (285).

Appendix I. Aligned Film Light Absorption

A more gradual application of the light absorption problem to aligned films is presented here, building on the *Results* section above.

Beginning our calculation with the irradiance, the absorption cross-section polarization dependence is described as equation (59) for our single chirality. We are in a position now to consider the dot product,

$$\begin{aligned}\boldsymbol{\epsilon}\cdot\hat{\boldsymbol{l}} &= \epsilon_x\hat{l}_x + \epsilon_y\hat{l}_y + \epsilon_z\hat{l}_z \\ &= \sin\theta_\epsilon\cos\phi_\epsilon\sin\theta\cos\phi + \sin\theta_\epsilon\sin\phi_\epsilon\sin\theta\sin\phi \\ &\quad + \cos\theta_\epsilon\cos\theta.\end{aligned} \tag{288}$$

Integrating the light field balance over $p(\hat{\boldsymbol{l}})$, we have from Equation (164)

$$\sigma_l(\boldsymbol{\epsilon},\omega) = \int_0^{2\pi}\int_0^{\pi}\sigma_l(\boldsymbol{\epsilon},\hat{\boldsymbol{l}},\omega)\cdot p(\theta)p(\phi)d\theta d\phi. \tag{289}$$

Subject to our distributions,

$$p(\theta) = \delta(\theta-\theta'), \qquad \theta'\in\left[0,\frac{\pi}{2}\right], \tag{290}$$

$$p(\phi) = \delta(\phi-0), \tag{291}$$

equation (289) becomes a sifting integral yielding $=\theta'$ and $\phi=0$. Our absorption cross-section becomes,

$$\epsilon_x\hat{l}_x = \sin\theta_\epsilon\cos\phi_\epsilon\sin\theta' \tag{292}$$

$$\epsilon_y\hat{l}_y = 0 \tag{293}$$

$$\epsilon_z\hat{l}_z = \cos\theta_\epsilon\cos\theta' \tag{294}$$

$$\begin{aligned}\sigma_l(\boldsymbol{\epsilon},\omega) = \sigma_{l,meas}(\omega)\Big(&\frac{5}{6}(\sin\theta_\epsilon\cos\phi_\epsilon\sin\theta' + \cos\theta_\epsilon\cos\theta')^2 \\ &+\frac{1}{6}(1-|\sin\theta_\epsilon\cos\phi_\epsilon\sin\theta' - \cos\theta_\epsilon\cos\theta'|)^2\Big).\end{aligned} \tag{295}$$

Solving the photon balance (64) with no z dependence to $\rho_{\langle l\rangle}$ or σ_l, subject to the incident unpolarized solar flux boundary condition (65), our forward flux becomes

$$J_F(\omega, \boldsymbol{\epsilon}|z) = J_0(\omega) \cdot \delta\left(\theta_\epsilon - \frac{\pi}{2}\right) \cdot \frac{1}{2\pi}$$

$$\cdot \exp\left[-\rho_{\langle l \rangle}\sigma_{l,meas}(\omega)\left(\frac{5}{6}(\sin\theta_\epsilon \cos\phi_\epsilon \sin\theta' + \cos\theta_\epsilon \cos\theta')^2\right.\right. \tag{296}$$

$$\left.\left. + \frac{1}{6}(1 - |\sin\theta_\epsilon \cos\phi_\epsilon \sin\theta' - \cos\theta_\epsilon \cos\theta'|)^2\right) \cdot z\right].$$

Integrating over θ_ϵ is trivial with the sifting property of $\delta(\cdot)$, yielding

$$J_F(\omega, \phi_\epsilon|z) = \frac{1}{2\pi} J_0(\omega)$$

$$\cdot \exp\left[-\rho_{\langle l \rangle}\sigma_{l,meas}(\omega)\left(\frac{5}{6}(\cos\phi_\epsilon \sin\theta')^2\right.\right. \tag{297}$$

$$\left.\left. + \frac{1}{6}(1 - |\cos\phi_\epsilon \sin\theta'|)^2\right) z\right].$$

Note that except in the vertically-aligned case ($\theta' = 0$), there is a polarization dependence to the field absorption. We can integrate over all frequencies to get the flux gradient at each polarization, or integrate over all polarization angles to get the flux gradient at each frequency.

With a boundary condition of reflectance R off the back electrode we solve the reverse flux differential equation, again substituting in $\sigma_l(\boldsymbol{\epsilon}, \omega)$ and integrating over θ_ϵ to yield

$$J_R(\omega, \phi_\epsilon|z) = \frac{1}{2\pi} \cdot R \cdot J_0(\omega)$$

$$\cdot \exp\left[-\rho_{\langle l \rangle}\sigma_{l,meas}(\omega)\left(\frac{5}{6}(\cos\phi_\epsilon \sin\theta')^2 + \frac{1}{6}(1 - |\cos\phi_\epsilon \sin\theta'|)^2\right)\right. \tag{298}$$

$$\left. \cdot (2T - z)\right].$$

The total light field gradient would then be

$$J_v(\omega, \phi_\epsilon|z) = J_R(\omega, \phi_\epsilon|z) + J_F(\omega, \phi_\epsilon|z), \tag{299}$$

for example plotted in Figure 62 for horizontally and vertically aligned films.

Appendix J. Isotropic absorption cross-section

For brevity, resolution of the integral

$$\sigma_l(\boldsymbol{\epsilon}, \omega) = \frac{1}{\pi}\frac{1}{2\pi}\frac{1}{6}\sigma_{l,meas}(\omega) \int_0^{2\pi}\int_0^{\pi} 5\varepsilon_\parallel{}^2 + \left(1 - |\varepsilon_\parallel|\right)^2 d\theta d\phi, \tag{300}$$

$$\varepsilon_\parallel = \boldsymbol{\epsilon} \cdot \hat{\boldsymbol{l}} = \sin\theta_\epsilon \cos\phi_\epsilon \sin\theta \cos\phi + \sin\theta_\epsilon \sin\phi_\epsilon \sin\theta \sin\phi$$

$$+ \cos\theta_\epsilon \cos\theta, \tag{301}$$

is omitted from the section above; we pursue it here. First, we recognize that when we integrate over θ_ϵ the incident light $\delta(\cdot)$ function will sift to $\theta_\epsilon = \frac{\pi}{2}$, simplifying

$$\varepsilon_\parallel \to \cos\phi_\epsilon \sin\theta \cos\phi + \sin\phi_\epsilon \sin\theta \sin\phi. \tag{302}$$

Next, we note the identity

$$(1 - |x|)^2 = |x|^2 + 1 - 2|x| = x^2 + 1 - 2|x|. \tag{303}$$

Applying this our integral simplifies to

$$6\int_0^{2\pi}\int_0^{\pi} {\varepsilon_\parallel}^2 d\theta d\phi + \int_0^{2\pi}\int_0^{\pi} d\theta d\phi - 2\int_0^{2\pi}\int_0^{\pi} |\varepsilon_\parallel|\ d\theta d\phi \tag{304}$$

which further simplifies with

$$\int_0^{2\pi}\int_0^{\pi} {\varepsilon_\parallel}^2 d\theta d\phi = \frac{1}{2}\pi^2 \tag{305}$$

giving us

$$\sigma_l(\boldsymbol{\epsilon}, \omega) = \frac{1}{\pi}\frac{1}{2\pi}\frac{1}{6}\sigma_{l,meas}(\omega)\left(3\pi^2 + 2\pi^2 - 2\int_0^{2\pi}\int_0^{\pi} |\varepsilon_\parallel|\ d\theta d\phi\right). \tag{306}$$

We then approach the absolute value integral. $\cos\theta$ is always positive over the domain, allowing us to easily evaluate the inner integral,

$$\begin{aligned}\int_0^{2\pi}\int_0^{\pi} |\varepsilon_\parallel|\ d\theta d\phi &= 2\int_0^{2\pi} |\sin(\phi)\sin(\phi_\epsilon) + \cos(\phi)\cos(\phi_\epsilon)| d\phi \\ &= 2\int_0^{2\pi} |\cos(\phi - \phi_\epsilon)| d\phi.\end{aligned} \tag{307}$$

Fortuitously, the remaining integral splits over four domains with definite sign regardless of the value of ϕ_ϵ,

$$2\int_0^{2\pi} |\cos(\phi - \phi_\epsilon)| d\phi = 8. \tag{308}$$

Our expression is then resolved,

$$\sigma_l(\boldsymbol{\epsilon}, \omega) = \left(\frac{5}{12} - \frac{4}{3\pi^2}\right)\sigma_{l,meas}(\omega). \tag{309}$$

Appendix K. Simplifying the Gaussian-Exponential Photon Population Decay Integrand

Equation (125) simplifies the *N(z)* integrand, especially as the product of Gaussians is a Gaussian,

$$J_0(\omega)\cdot\sigma_l(\omega) = J_0\sum_a S_a\cdot\exp\left[-\frac{(\omega-\mu_a)^2}{2\sigma_a^2}\right]\exp\left[-\frac{(\omega-\mu_J)^2}{2\sigma_J^2}\right]$$

(310)

$$= J_0\sum_a S_a\cdot\exp\left[-\frac{(\mu_J-\mu_a)^2}{2(\sigma_J^2+\sigma_a^2)}\right]\cdot\exp\left[-\frac{(\omega-\mu_{Ja})^2}{2\sigma_{Ja}^2}\right]$$

(311) $$\mu_{Ja} \equiv \frac{\mu_J\sigma_a^2+\mu_a\sigma_J^2}{\sigma_J^2+\sigma_a^2}$$

(312) $$\sigma_{Ja} \equiv \sqrt{\frac{\sigma_J^2\sigma_a^2}{\sigma_J^2+\sigma_a^2}}$$

Note the first exponential is constant in ω; it simply represents the normalized overlap area over the two Gaussians. This allows us to split our integral into a sum of integrals convoluting the Gaussian and our exponential decay, with a term for each $\sigma_l(\omega)$ peak

(313) $$N(z) = \rho_{\langle l\rangle}J_0\sum_a S_a\exp\left[-\frac{(\mu_J-\mu_a)^2}{2(\sigma_J^2+\sigma_a^2)}\right]\int_0^\infty\exp\left[-\frac{(\omega-\mu_{Ja})^2}{2\sigma_{Ja}^2}\right]\exp\left[-\rho_{\langle l\rangle}z\sigma_l(\omega)\right]d\omega$$

Unfortunately taking the integral of an exponential of a single Gaussian on its own is not solvent, let alone the exponential of a series of Gaussians or the convolution of a Gaussian with that exponential of a series of Gaussians, which I will call a 'Exponential-Gaussian' for brevity from here,

(314) $$\exp\left[-\rho_{\langle l\rangle}z\sigma_l(\omega)\right] = \prod_j\exp\left[-\rho_{\langle l\rangle}z\cdot S_j\cdot\exp\left[-\frac{(\omega-\mu_j)^2}{2\sigma_j^2}\right]\right]$$

$$N(z) = \rho_{\langle l\rangle}J_0\sum_a S_a\exp\left[-\frac{(\mu_J-\mu_a)^2}{2(\sigma_J^2+\sigma_a^2)}\right]\int_0^\infty\exp\left[-\frac{(\omega-\mu_{Ja})^2}{2\sigma_{Ja}^2}\right]\prod_j\exp\left[-\rho_{\langle l\rangle}z\cdot S_j\right.$$

(315)

$$\left.\cdot\exp\left[-\frac{(\omega-\mu_j)^2}{2\sigma_j^2}\right]\right]d\omega$$

As is therefore, *N(z)* an insolvent convolution of a Gaussian and an exponential decay of a Gaussian (313). Furthermore, an Exponential-Gaussian on its own is not well approximated by a Taylor Expansion (Figure 84).

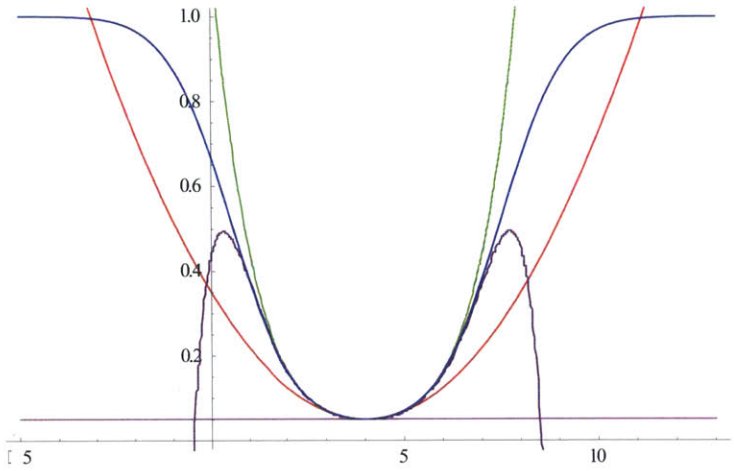


Figure 84. Taylor Expansion of a Gaussian-Exponential. Blue is exact, magenta is first-order, red is second-order, green is sixth-order, and purple is eigth-order.

We make two relevant observations of the function however; far from the symmetry center of the Exponential-Gaussian the function is unity, while at the symmetry center the first-order solution captures its value.

On its own, the Exponential-Gaussian integral is insolvent. We observe however that while the Exponential-Gaussian term is in general not well captured by the first-order Taylor expansion, its product with a Gaussian term *is* because either the function extrema are close *or* the Gaussian-Exponential term reduces to unity.

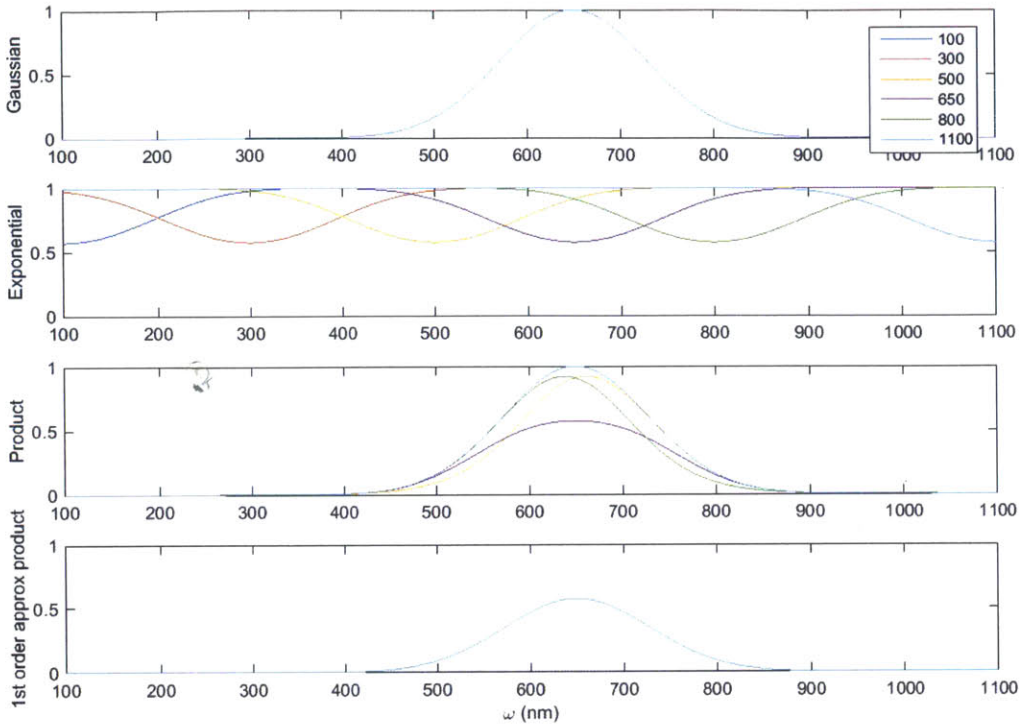


Figure 85. Comparison of Exponential-Gaussian to the product of a Gaussian with the Exponential-Gaussian of different mean values (legend).

These observations suggest a number of simplifications that make the frequency integral solvent and are revealed to be sufficient approximations in Figure 76. First, when the Exponential-Gaussian and Gaussian peaks are far from one another, the Exponential-Gaussian reduces to unity, which allows us to collapse the product series in (315) to $j = a$,

$$N(z) = \rho_{\langle l \rangle} J_0 \sum_a S_a \exp\left[-\frac{(\mu_J - \mu_a)^2}{2(\sigma_J^2 + \sigma_a^2)}\right] \int_0^\infty \exp\left[-\frac{(\omega - \mu_{Ja})^2}{2\sigma_{Ja}^2}\right] \exp\left[-\rho_{\langle l \rangle} z \cdot S_a \cdot \exp\left[-\frac{(\omega - \mu_a)^2}{2\sigma_a^2}\right]\right] d\omega. \tag{316}$$

Then, we can recognize that when μ_{Ja} and μ_a are close, the Exponential-Gaussian can be approximated by low-order Taylor expansions (Figure 84). In our system, where $\{\mu_J, \sigma_J\}$ correspond to the solar flux and $\{\mu_a, \sigma_a\}$ correspond to the absorption peaks, that is ubiquitously the case because $\sigma_J \gg \sigma_a$ *i.e.* the variance of the solar spectrum, which spans visible and nIR frequencies, is much larger than that of the SWNT absorption modes which are resonant excitation peaks; that shifts the weighted sum (311) to shift towards μ_a. We therefore approximate the Exponential-Gaussian as its first- and third-order

expansions, ultimately finding by comparison to the numerical solution that the first-order expansion turns out to be sufficient (Figure 76). The integral solution under the first-order expansion is,

(317) $$N(z) = \rho_{\langle l \rangle} J_0 \sum_{a=1}^{2} S_a \gamma_a \sqrt{2\pi} \cdot e^{-\rho_{\langle l \rangle} S_a z}$$

(318) $$\gamma_a \equiv \sigma_{Ja} e^{\frac{(\mu_J - \mu_a)^2}{2(\sigma_J^2 + \sigma_a^2)}}$$

where γ_a can be conceptually understood as the area of overlap of the solar flux and absorption peak a. The sum must include two peaks for each SWNT chirality, corresponding to the E_{1u} and E_{2u} transitions.

Appendix L. Simplification of Surface Transport Model Expression

To reach equation (19) to approximations were required. We began with the exact model from Chapter I, expressed in terms of a forced voltage,

(319) $$\frac{dr}{dt} = -\gamma \frac{V}{R'} \frac{2r\Delta - \Delta^2}{r^3}.$$

In (319) R' is the total resistance, including both the neck region resistance R which evolves over time and the series resistance R_S of the remainder of the wire which is constant. The neck resistance is simply

(320) $$R = \frac{\rho\delta}{\pi} r^{-2}.$$

Substituting in we have

(321) $$\frac{dr}{dt} = -\gamma V \frac{1}{r} \frac{2r\Delta - \Delta^2}{R_S r^2 + {\rho\delta}/{\pi}}.$$

In the denominator we can compare the magnitude of terms. Experimentally we measure a typical R_S from the contact wires of 200 Ω, before reaching the stochastic regime a reasonable minimum r is 10 nm, the resistivity of bulk Au is $2.3 \times 10^{-8}\ \Omega \cdot m$, and a typical δ is 10 nm. To an order of magnitude, this yields

$$R_S r^2 \cong 2 \times 10^{-13}\ \Omega \cdot m^2 \gg {\rho\delta}/{\pi} \cong 7.3 \times 10^{-17}\ \Omega \cdot m^2.$$

The several orders of magnitude difference easily makes the second term in the denominator negligible,

(322) $$\frac{dr}{dt} = -\frac{\gamma}{R_S} V \frac{2r\Delta - \Delta^2}{r^3}.$$

Put in other terms, this approximation is the same as recognizing that the series resistance is much greater than the nanowire resistance, *i.e.* $R_S \gg R$, which we empirically observe to be true.

The second approximation that we make is that the radius is much larger than one atomic diameter ($r \gg \Delta$) up until the stochastic regime is reached. We observe that the stochastic regime is typically entered when R rises from approximately 50 Ω to 250 Ω, which for an effective initial nanowire

diameter of 20 nm (our devices) corresponds to an effective radius of 8.9 nm. This number is also consistent with *in situ* observation of nanowire electromigration in TEM.[22] For Au, Δ is approximately 0.288 nm, much less than this radius. We can therefore approximate that for the entire trace region of interest $r \gg \Delta$ and hence

$$\frac{dr}{dt} = -\frac{2\gamma\Delta}{R_s} V r^{-2}. \tag{323}$$

REFERENCES

[1] A. Lodder. *Electromigration theory unified.* EPL (Europhysics Letters) **72**, 774 (2005).

[2] D. G. Pierce and P. G. Brusius. *Electromigration: A review.* Microelectronics Reliability **37**, 1053 (1997).

[3] P. J. Rous, T. L. Einstein, and E. D. Williams. *Theory of surface electromigration on metals: application to self-electromigration on Cu(111).* Surface Science **315**, L995 (1994).

[4] C. Bosvieux and J. Friedel. *Sur l'electrolyse des alliages metalliques.* Journal of Physics and Chemistry of Solids **23**, 123 (1962).

[5] J. Ventura, J. B. Sousa, Y. Liu, Z. Zhang, and P. P. Freitas. *Electromigration in thin tunnel junctions with ferromagnetic/nonmagnetic electrodes: Nanoconstrictions, local heating, and direct and wind forces.* Phys Rev B **72**, 094432 (2005).

[6] J. P. Dekker and A. Lodder. *Calculated electromigration wind force in face-centered-cubic and body-centered-cubic metals.* J Appl Phys **84**, 1958 (1998).

[7] D. N. Bly and P. J. Rous. *Theoretical study of the electromigration wind force for adatom migration at metal surfaces.* Phys Rev B **53**, 13909 (1996).

[8] D. R. Strachan, D. E. Smith, D. E. Johnston, T.-H. Park, M. J. Therien, D. A. Bonnell, and A. T. Johnson. *Controlled fabrication of nanogaps in ambient environment for molecular electronics.* Applied Physics Letters **86**, 043109 (2005).

[9] J. R. Black. *Electromigration: A brief survey and some recent results.* Electron Devices, IEEE Transactions on **16**, 338 (1969).

[10] P. S. Ho and T. Kwok. *Electromigration in Metals.* Rep Prog Phys **52**, 301 (1989).

[11] P. S. Ho and T. Kwok. *Electromigration in metals.* Reports on Progress in Physics **52**, 301 (1989).

[12] M. Shatzkes and J. R. Lloyd. *A model for conductor failure considering diffusion concurrently with electromigration resulting in a current exponent of 2.* J Appl Phys **59**, 3890 (1986).

[13] J. R. Lloyd and J. Kitchin. *The finite electromigration boundary value problem.* Journal of Materials Research **9**, 563 (1994).

[14] V. M. Dwyer, F. S. Wang, and P. Donaldson. *Electromigration failure in a finite conductor with a single blocking boundary.* J Appl Phys **76**, 7305 (1994).

[15] J. T. Trattles, A. G. O' Neill, and B. C. Mecrow. *Computer simulation of electromigration in thin - film metal conductors.* J Appl Phys **75**, 7799 (1994).

[16] R. Rosenberg, D. C. Edelstein, a. C.-K. Hu, and K. P. Rodbell. *COPPER METALLIZATION FOR HIGH PERFORMANCE SILICON TECHNOLOGY.* Annual Review of Materials Science **30**, 229 (2000).

[17] C. K. Hu, M. B. Small, and P. S. Ho. *Electromigration in Al(Cu) two - level structures: Effect of Cu and kinetics of damage formation.* J Appl Phys **74**, 969 (1993).

[18] E. C. C. Yeh, W. J. Choi, K. N. Tu, P. Elenius, and H. Balkan. *Current-crowding-induced electromigration failure in flip chip solder joints.* Applied Physics Letters **80**, 580 (2002).

[19] M. W. Lane, E. G. Liniger, and J. R. Lloyd. *Relationship between interfacial adhesion and electromigration in Cu metallization.* J Appl Phys **93**, 1417 (2003).

[20] C. S. Hau-Riege and C. V. Thompson. *Electromigration in Cu interconnects with very different grain structures.* Applied Physics Letters **78**, 3451 (2001).

[21] C. V. Thompson and J. R. Lloyd. *Electromigration and IC Interconnects.* MRS Bulletin **18**, 19 (1993).

[22] T. Taychatanapat, K. I. Bolotin, F. Kuemmeth, and D. C. Ralph. *Imaging electromigration during the formation of break junctions.* Nano Letters **7**, 652 (2007).

[23] H. Park, A. K. L. Lim, A. P. Alivisatos, J. Park, and P. L. McEuen. *Fabrication of metallic electrodes with nanometer separation by electromigration.* Applied Physics Letters **75**, 301 (1999).

[24] Z. Ioffe, T. Shamai, A. Ophir, G. Noy, I. Yutsis, K. Kfir, O. Cheshnovsky, and Y. Selzer. *Detection of heating in current-carrying molecular junctions by Raman scattering*. Nat Nano **3**, 727 (2008).
[25] J. Lambe and S. L. McCarthy. *Light Emission from Inelastic Electron Tunneling*. Physical Review Letters **37**, 923 (1976).
[26] Y. Kuk and P. J. Silverman. *Role of tip structure in scanning tunneling microscopy*. Applied Physics Letters **48**, 1597 (1986).
[27] G. Cross, A. Schirmeisen, A. Stalder, P. Grütter, M. Tschudy, and U. Dürig. *Adhesion Interaction between Atomically Defined Tip and Sample*. Physical Review Letters **80**, 4685 (1998).
[28] D. R. Strachan, D. E. Smith, M. D. Fischbein, D. E. Johnston, B. S. Guiton, M. Drndic, D. A. Bonnell, and A. T. Johnson. *Clean electromigrated nanogaps imaged by transmission electron microscopy*. Nano Letters **6**, 441 (2006).
[29] D. R. Strachan, D. E. Johnston, B. S. Guiton, S. S. Datta, P. K. Davies, D. A. Bonnell, and A. T. C. Johnson. *Real-Time TEM Imaging of the Formation of Crystalline Nanoscale Gaps*. Physical Review Letters **100**, 056805 (2008).
[30] F. Prins, T. Hayashi, B. J. A. de Vos van Steenwijk, B. Gao, E. A. Osorio, K. Muraki, and H. S. J. van der Zant. *Room-temperature stability of Pt nanogaps formed by self-breaking*. Applied Physics Letters **94**, 123108 (2009).
[31] D. R. Ward, D. A. Corley, J. M. Tour, and D. Natelson. *Vibrational and electronic heating in nanoscale junctions*. Nat Nano **6**, 33 (2011).
[32] H. Park, J. Park, A. K. L. Lim, E. H. Anderson, A. P. Alivisatos, and P. L. McEuen. *Nanomechanical oscillations in a single-C60 transistor*. Nature **407**, 57 (2000).
[33] J. Park *et al. Coulomb blockade and the Kondo effect in single-atom transistors*. Nature **417**, 722 (2002).
[34] S. Kubatkin, A. Danilov, M. Hjort, J. Cornil, J.-L. Bredas, N. Stuhr-Hansen, P. Hedegard, and T. Bjornholm. *Single-electron transistor of a single organic molecule with access to several redox states*. Nature **425**, 698 (2003).
[35] M. L. Perrin *et al. Large tunable image-charge effects in single-molecule junctions*. Nat Nano **8**, 282 (2013).
[36] W. Liang, M. P. Shores, M. Bockrath, J. R. Long, and H. Park. *Kondo resonance in a single-molecule transistor*. Nature **417**, 725 (2002).
[37] A. N. Pasupathy, R. C. Bialczak, J. Martinek, J. E. Grose, L. A. K. Donev, P. L. McEuen, and D. C. Ralph. *The Kondo Effect in the Presence of Ferromagnetism*. Science **306**, 86 (2004).
[38] M. R. Calvo, J. Fernandez-Rossier, J. J. Palacios, D. Jacob, D. Natelson, and C. Untiedt. *The Kondo effect in ferromagnetic atomic contacts*. Nature **458**, 1150 (2009).
[39] B. C. Stipe, M. A. Rezaei, and W. Ho. *Single-Molecule Vibrational Spectroscopy and Microscopy*. Science **280**, 1732 (1998).
[40] H. Song, Y. Kim, Y. H. Jang, H. Jeong, M. A. Reed, and T. Lee. *Observation of molecular orbital gating*. Nature **462**, 1039 (2009).
[41] D. R. Ward, N. K. Grady, C. S. Levin, N. J. Halas, Y. Wu, P. Nordlander, and D. Natelson. *Electromigrated Nanoscale Gaps for Surface-Enhanced Raman Spectroscopy*. Nano Letters **7**, 1396 (2007).
[42] C. B. Winkelmann, N. Roch, W. Wernsdorfer, V. Bouchiat, and F. Balestro. *Superconductivity in a single-C60 transistor*. Nat Phys **5**, 876 (2009).
[43] J. E. Grose, E. S. Tam, C. Timm, M. Scheloske, B. Ulgut, J. J. Parks, H. D. Abruna, W. Harneit, and D. C. Ralph. *Tunnelling spectra of individual magnetic endofullerene molecules*. Nat Mater **7**, 884 (2008).
[44] D. R. Ward, F. Huser, F. Pauly, J. C. Cuevas, and D. Natelson. *Optical rectification and field enhancement in a plasmonic nanogap*. Nat Nano **5**, 732 (2010).
[45] Y. Kim, W. Jeong, K. Kim, W. Lee, and P. Reddy. *Electrostatic control of thermoelectricity in molecular junctions*. Nat Nano **9**, 881 (2014).

[46] Y. Zhao *et al. Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling*. Nat Nano **9**, 466 (2014).
[47] P. Reddy, S.-Y. Jang, R. A. Segalman, and A. Majumdar. *Thermoelectricity in Molecular Junctions*. Science **315**, 1568 (2007).
[48] B. C. Stipe, M. A. Rezaei, and W. Ho. *Inducing and Viewing the Rotational Motion of a Single Molecule*. Science **279**, 1907 (1998).
[49] A. E. Baber, H. L. Tierney, and E. C. H. Sykes. *A Quantitative Single-Molecule Study of Thioether Molecular Rotors*. ACS Nano **2**, 2385 (2008).
[50] G. Leonhard. *Functionalized molecules studied by STM: motion, switching and reactivity*. Journal of Physics: Condensed Matter **20**, 053001 (2008).
[51] S. V. Aradhya and L. Venkataraman. *Single-molecule junctions beyond electronic transport*. Nat Nano **8**, 399 (2013).
[52] C. Schirm, M. Matt, F. Pauly, J. C. Cuevas, P. Nielaba, and E. Scheer. *A current-driven single-atom memory*. Nature Nanotechnology **8**, 645 (2013).
[53] M. L. Perrin, E. Burzuri, and H. S. J. van der Zant. *Single-molecule transistors*. Chemical Society Reviews **44**, 902 (2015).
[54] L. Sun, Y. A. Diaz-Fernandez, T. A. Gschneidtner, F. Westerlund, S. Lara-Avila, and K. Moth-Poulsen. *Single-molecule electronics: from chemical design to functional devices*. Chemical Society Reviews **43**, 7378 (2014).
[55] D. Branton *et al. The potential and challenges of nanopore sequencing*. Nat Biotech **26**, 1146 (2008).
[56] M. Rahmani, M. T. Ahmadi, H. K. F. Abadi, M. Saeidmanesh, E. Akbari, and R. Ismail. *Analytical modeling of trilayer graphene nanoribbon Schottky-barrier FET for high-speed switching applications*. Nanoscale Research Letters **8**, 55 (2013).
[57] E. Lörtscher, B. Gotsmann, Y. Lee, L. Yu, C. Rettner, and H. Riel. *Transport Properties of a Single-Molecule Diode*. ACS Nano **6**, 4931 (2012).
[58] B. M. Venkatesan and R. Bashir. *Nanopore sensors for nucleic acid analysis*. Nat Nano **6**, 615 (2011).
[59] H. Liu *et al. Translocation of Single-Stranded DNA Through Single-Walled Carbon Nanotubes*. Science **327**, 64 (2010).
[60] A. P. Ivanov, E. Instuli, C. M. McGilvery, G. Baldwin, D. W. McComb, T. Albrecht, and J. B. Edel. *DNA Tunneling Detector Embedded in a Nanopore*. Nano Letters **11**, 279 (2010).
[61] S. W. Kowalczyk, D. B. Wells, A. Aksimentiev, and C. Dekker. *Slowing down DNA Translocation through a Nanopore in Lithium Chloride*. Nano Letters **12**, 1038 (2012).
[62] M. Wanunu, T. Dadosh, V. Ray, J. Jin, L. McReynolds, and M. Drndic. *Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors*. Nat Nano **5**, 807 (2010).
[63] A. J. Storm, C. Storm, J. Chen, H. Zandbergen, J.-F. Joanny, and C. Dekker. *Fast DNA Translocation through a Solid-State Nanopore*. Nano Letters **5**, 1193 (2005).
[64] J. Clarke, H.-C. Wu, L. Jayasinghe, A. Patel, S. Reid, and H. Bayley. *Continuous base identification for single-molecule nanopore DNA sequencing*. Nat Nano **4**, 265 (2009).
[65] E. A. Manrao, I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, and J. H. Gundlach. *Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase*. Nat Biotech **30**, 349 (2012).
[66] G. F. Schneider, S. W. Kowalczyk, V. E. Calado, G. Pandraud, H. W. Zandbergen, L. M. K. Vandersypen, and C. Dekker. *DNA Translocation through Graphene Nanopores*. Nano Letters **10**, 3163 (2010).
[67] C. A. Merchant *et al. DNA Translocation through Graphene Nanopores*. Nano Letters **10**, 2915 (2010).
[68] H. W. C. Postma. *Rapid Sequencing of Individual DNA Molecules in Graphene Nanogaps*. Nano Letters **10**, 420 (2010).

[69] T. Ohshiro and Y. Umezawa. *Complementary base-pair-facilitated electron tunneling for electrically pinpointing complementary nucleobases*. Proceedings of the National Academy of Sciences of the United States of America **103**, 10 (2006).
[70] B. C. Stipe, M. A. Rezaei, and W. Ho. *Localization of Inelastic Tunneling and the Determination of Atomic-Scale Structure with Chemical Specificity*. Physical Review Letters **82**, 1724 (1999).
[71] J. R. Hahn, H. J. Lee, and W. Ho. *Electronic Resonance and Symmetry in Single-Molecule Inelastic Electron Tunneling*. Physical Review Letters **85**, 1914 (2000).
[72] N. Lorente, M. Persson, L. J. Lauhon, and W. Ho. *Symmetry Selection Rules for Vibrationally Inelastic Tunneling*. Physical Review Letters **86**, 2593 (2001).
[73] Y. Sainoo, Y. Kim, T. Okawa, T. Komeda, H. Shigekawa, and M. Kawai. *Excitation of Molecular Vibrational Modes with Inelastic Scanning Tunneling Microscopy Processes: Examination through Action Spectra of \textit{cis} -2-Butene on Pd(110)*. Physical Review Letters **95**, 246102 (2005).
[74] Y. Kim, T. Pietsch, A. Erbe, W. Belzig, and E. Scheer. *Benzenedithiol: A Broad-Range Single-Channel Molecular Conductor*. Nano Letters **11**, 3734 (2011).
[75] H. Song, Y. Kim, J. Ku, Y. H. Jang, H. Jeong, and T. Lee. *Vibrational spectra of metal-molecule-metal junctions in electromigrated nanogap electrodes by inelastic electron tunneling*. Applied Physics Letters **94** (2009).
[76] M. Rahimi and A. Troisi. *Probing local electric field and conformational switching in single-molecule break junctions*. Phys Rev B **79**, 113413 (2009).
[77] A. Xiang, H. Li, S. Chen, S.-X. Liu, S. Decurtins, M. Bai, S. Hou, and J. Liao. *Electronic transport in benzodifuran single-molecule transistors*. Nanoscale **7**, 7665 (2015).
[78] D. E. Johnston, D. R. Strachan, and A. T. C. Johnson. *Parallel Fabrication of Nanogap Electrodes*. Nano Letters **7**, 2774 (2007).
[79] M. M. Deshmukh, A. L. Prieto, Q. Gu, and H. Park. *Fabrication of Asymmetric Electrode Pairs with Nanometer Separation Made of Two Distinct Metals*. Nano Letters **3**, 1383 (2003).
[80] V. M. Serdio V, T. Muraki, S. Takeshita, D. E. Hurtado S, S. Kano, T. Teranishi, and Y. Majima. *Gap separation-controlled nanogap electrodes by molecular ruler electroless gold plating*. RSC Advances **5**, 22160 (2015).
[81] H. Li, I. H. Wani, A. Hayat, S. H. M. Jafri, and K. Leifer. *Fabrication of reproducible sub-5 nm nanogaps by a focused ion beam and observation of Fowler-Nordheim tunneling*. Applied Physics Letters **107**, 103108 (2015).
[82] A. Cui, Z. Liu, H. Dong, Y. Wang, Y. Zhen, W. Li, J. Li, C. Gu, and W. Hu. *Single Grain Boundary Break Junction for Suspended Nanogap Electrodes with Gapwidth Down to 1–2 nm by Focused Ion Beam Milling*. Advanced Materials **27**, 3002 (2015).
[83] Q. J. Huang, C. M. Lilley, and R. Divan. *An in situ investigation of electromigration in Cu nanowires*. Nanotechnology **20** (2009).
[84] M. Frimmer, G. Puebla-Hellmann, A. Wallraff, and L. Novotny. *The role of titanium in electromigrated tunnel junctions*. Applied Physics Letters **105**, 221118 (2014).
[85] K. I. Bolotin, F. Kuemmeth, A. N. Pasupathy, and D. C. Ralph. *Metal-nanoparticle single-electron transistors fabricated using electromigration*. Applied Physics Letters **84**, 3154 (2004).
[86] M. F. Lambert, M. F. Goffman, J. P. Bourgoin, and P. Hesto. *Fabrication and characterization of sub-3 nm gaps for single-cluster and single-molecule experiments*. Nanotechnology **14**, 772 (2003).
[87] G. Esen and M. S. Fuhrer. *Temperature control of electromigration to form gold nanogap junctions*. Applied Physics Letters **87** (2005).
[88] C. X. Xiang, J. Y. Kim, and R. M. Penner. *Reconnectable Sub-5 nm Nanogaps in Ultralong Gold Nanowires*. Nano Letters **9**, 2133 (2009).
[89] A. Cui, H. Dong, and W. Hu. *Nanogap Electrodes towards Solid State Single-Molecule Transistors*. Small, n/a (2015).
[90] M. Ito, K. Morihara, T. Toyonaka, K. Takikawa, and J.-i. Shirakashi. *High-throughput nanogap formation by field-emission-induced electromigration*. Journal of Vacuum Science & Technology B **33**, 051801 (2015).

[91] T. Li, W. P. Hu, and D. B. Zhu. *Nanogap Electrodes*. Advanced Materials **22**, 286 (2010).
[92] J. M. Campbell and R. G. Knobel. *Feedback-controlled electromigration for the fabrication of point contacts*. Applied Physics Letters **102**, 023105 (2013).
[93] M. L. Trouwborst, S. J. van der Molen, and B. J. van Wees. *The role of Joule heating in the formation of nanogaps by electromigration*. J Appl Phys **99** (2006).
[94] S. Karim, K. Maaz, G. Ali, and W. Ensinger. *Diameter dependent failure current density of gold nanowires*. J Phys D Appl Phys **42** (2009).
[95] R. Quan, C. S. Pitler, M. A. Ratner, and M. G. Reuter. *Quantitative Interpretations of Break Junction Conductance Histograms in Molecular Electron Transport*. ACS Nano **9**, 7704 (2015).
[96] T. Ando. *Theory of Electronic States and Transport in Carbon Nanotubes*. Journal of the Physical Society of Japan **74**, 777 (2005).
[97] X. Tu, S. Manohar, A. Jagota, and M. Zheng. *DNA sequence motifs for structure-specific recognition and separation of carbon nanotubes*. Nature **460**, 250 (2009).
[98] J. Maultzsch *et al. Exciton binding energies in carbon nanotubes from two-photon photoluminescence*. Physical Review B **72** (2005).
[99] M. F. Islam, D. E. Milkie, C. L. Kane, A. G. Yodh, and J. M. Kikkawa. *Direct measurement of the polarized optical absorption cross section of single-wall carbon nanotubes*. Physical Review Letters **93** (2004).
[100] P. Avouris, M. Freitag, and V. Perebeinos. *Carbon-nanotube photonics and optoelectronics*. Nature Photonics **2**, 341 (2008).
[101] V. Perebeinos and P. Avouris. *Exciton ionization, Franz-Keldysh, and stark effects in carbon nanotubes*. Nano Letters **7**, 609 (2007).
[102] L. Cognet, D. A. Tsyboulski, J. D. R. Rocha, C. D. Doyle, J. M. Tour, and R. B. Weisman. *Stepwise quenching of exciton fluorescence in carbon nanotubes by single-molecule reactions*. Science **316**, 1465 (2007).
[103] L. Luer, S. Hoseinkhani, D. Polli, J. Crochet, T. Hertel, and G. Lanzani. *Size and mobility of excitons in (6, 5) carbon nanotubes*. Nat Phys **5**, 54 (2009).
[104] J. J. Crochet, J. G. Duque, J. H. Werner, B. Lounis, L. Cognet, and S. K. Doorn. *Disorder Limited Exciton Transport in Colloidal Single-Wall Carbon Nanotubes*. Nano Letters **12**, 5091 (2012).
[105] T. Hertel, S. Himmelein, T. Ackermann, D. Stich, and J. Crochet. *Diffusion Limited Photoluminescence Quantum Yields in 1-D Semiconductors: Single-Wall Carbon Nanotubes*. Acs Nano **4**, 7161 (2010).
[106] B. A. Ruzicka, R. Wang, J. Lohrman, S. Q. Ren, and H. Zhao. *Exciton diffusion in semiconducting single-walled carbon nanotubes studied by transient absorption microscopy*. Physical Review B **86** (2012).
[107] M. D. Anderson, Y.-f. Xiao, and J. M. Fraser. *First-passage theory of exciton population loss in single-walled carbon nanotubes reveals micron-scale intrinsic diffusion lengths*. Physical Review B **88**, 045420 (2013).
[108] Y. F. Xiao, T. Q. Nhan, M. W. B. Wilson, and J. M. Fraser. *Saturation of the Photoluminescence at Few-Exciton Levels in a Single-Walled Carbon Nanotube under Ultrafast Excitation*. Physical Review Letters **104** (2010).
[109] H. Hirori, K. Matsuda, Y. Miyauchi, S. Maruyama, and Y. Kanemitsu. *Exciton Localization of Single-Walled Carbon Nanotubes Revealed by Femtosecond Excitation Correlation Spectroscopy*. Physical Review Letters **97**, 257401 (2006).
[110] A. J. Siitonen, D. A. Tsyboulski, S. M. Bachilo, and R. B. Weisman. *Surfactant-Dependent Exciton Mobility in Single-Walled Carbon Nanotubes Studied by Single-Molecule Reactions*. Nano Letters **10**, 1595 (2010).
[111] T. Gokus, L. Cognet, J. G. Duque, M. Pasquali, A. Hartschuh, and B. Lounis. *Mono- and Biexponential Luminescence Decays of Individual Single-Walled Carbon Nanotubes*. Journal of Physical Chemistry C **114**, 14025 (2010).

[112] P. H. Tan, A. G. Rozhin, T. Hasan, P. Hu, V. Scardaci, W. I. Milne, and A. C. Ferrari. *Photoluminescence Spectroscopy of Carbon Nanotube Bundles: Evidence for Exciton Energy Transfer.* Physical Review Letters **99**, 137402 (2007).
[113] J. J. Crochet, J. D. Sau, J. G. Duque, S. K. Doorn, and M. L. Cohen. *Electrodynamic and Excitonic Intertube Interactions in Semiconducting Carbon Nanotube Aggregates.* Acs Nano **5**, 2611 (2011).
[114] D. E. Arvizu, (National Renewable Energy Laboratory, 2012).
[115] D. J. Bindl, M. J. Shea, and M. S. Arnold. *Enhancing extraction of photogenerated excitons from semiconducting carbon nanotube films as photocurrent.* Chemical Physics **413**, 29 (2013).
[116] R. M. Jain, R. Howden, K. Tvrdy, S. Shimizu, A. J. Hilmer, T. P. McNicholas, K. K. Gleason, and M. S. Strano. *Polymer-Free Near-Infrared Photovoltaics with Single Chirality (6,5) Semiconducting Carbon Nanotube Active Layers.* Advanced Materials **24**, 4436 (2012).
[117] M. Bernardi, J. Lohrman, P. V. Kumar, A. Kirkeminde, N. Ferralis, J. C. Grossman, and S. Q. Ren. *Nanocarbon-Based Photovoltaics.* Acs Nano **6**, 8896 (2012).
[118] M. J. Shea and M. S. Arnold. *1% solar cells derived from ultrathin carbon nanotube photoabsorbing films.* Applied Physics Letters **102**, 243101 (2013).
[119] L. Wang, H. Liu, R. M. Konik, J. A. Misewich, and S. S. Wong. *Carbon nanotube-based heterostructures for solar energy applications.* Chemical Society Reviews **42**, 8134 (2013).
[120] M. S. Arnold, J. Blackburn, J. Crochet, S. Doorn, J. Duque, A. Mohite, and H. Telg. *Recent Developments in the Photophysics of Single-Wall Carbon Nanotubes for Active and Passive Material Elements in Thin Film Photovoltaics.* Physical Chemistry Chemical Physics (2013).
[121] Y. Ye, D. J. Bindl, R. M. Jacobberger, M.-Y. Wu, S. S. Roy, and M. S. Arnold. *Semiconducting Carbon Nanotube Aerogel Bulk Heterojunction Solar Cells.* Small, n/a (2014).
[122] H. Liu, D. Nishide, T. Tanaka, and H. Kataura. *Large-scale single-chirality separation of single-wall carbon nanotubes by simple gel chromatography.* Nat Commun **2**, 309 (2011).
[123] K. Tvrdy, R. M. Jain, R. Han, A. J. Hilmer, T. P. McNicholas, and M. S. Strano. *A Kinetic Model for the Deterministic Prediction of Gel-Based Single-Chirality Single-Walled Carbon Nanotube Separation.* Acs Nano **7**, 1779 (2013).
[124] K. Moshammer, F. Hennrich, and M. Kappes. *Selective suspension in aqueous sodium dodecyl sulfate according to electronic structure type allows simple separation of metallic from semiconducting single-walled carbon nanotubes.* Nano Research **2**, 599 (2009).
[125] A. S. Brown and M. A. Green. *Limiting efficiency for current-constrained two-terminal tandem cell stacks.* Progress in Photovoltaics: Research and Applications **10**, 299 (2002).
[126] F. Schoppler, C. Mann, T. C. Hain, F. M. Neubauer, G. Privitera, F. Bonaccorso, D. P. Chu, A. C. Ferrari, and T. Hertel. *Molar Extinction Coefficient of Single-Wall Carbon Nanotubes.* Journal of Physical Chemistry C **115**, 14682 (2011).
[127] A. V. Naumov, D. A. Tsyboulski, S. M. Bachilo, and R. B. Weisman. *Length-dependent optical properties of single-walled carbon nanotube samples.* Chemical Physics.
[128] N. M. Gabor, Z. H. Zhong, K. Bosnick, and P. L. McEuen. *Ultrafast Photocurrent Measurement of the Escape Time of Electrons and Holes from Carbon Nanotube p-i-n Photodiodes.* Physical Review Letters **108** (2012).
[129] S. Q. Ren, M. Bernardi, R. R. Lunt, V. Bulovic, J. C. Grossman, and S. Gradecak. *Toward Efficient Carbon Nanotube/P3HT Solar Cells: Active Layer Morphology, Electrical, and Optical Properties.* Nano Letters **11**, 5316 (2011).
[130] R. C. Jaklevic and L. Elie. *Scanning-Tunneling-Microscope Observation of Surface Diffusion on an Atomic Scale: Au on Au(111).* Physical Review Letters **60**, 120 (1988).
[131] C. L. Liu, J. M. Cohen, J. B. Adams, and A. F. Voter. *EAM study of surface self-diffusion of single adatoms of fcc metals Ni, Cu, Al, Ag, Au, Pd, and Pt.* Surface Science **253**, 334 (1991).
[132] S. Karim, W. Ensinger, T. W. Cornelius, and R. Neumann. *Investigation of size effects in the electrical resistivity of single electrochemically fabricated gold nanowires.* Physica E **40**, 3173 (2008).

[133] D. R. Strachan, D. E. Smith, D. E. Johnston, T. H. Park, M. J. Therien, D. A. Bonnell, and A. T. Johnson. *Controlled fabrication of nanogaps in ambient environment for molecular electronics*. Applied Physics Letters **86** (2005).
[134] M. R. Gungor and D. Maroudas. *Theoretical analysis of electromigration-induced failure of metallic thin films due to transgranular void propagation*. J Appl Phys **85**, 2233 (1999).
[135] L. Zhang, S. Ou, J. Huang, K. N. Tu, S. Gee, and L. Nguyen. *Effect of current crowding on void propagation at the interface between intermetallic compound and solder in flip chip solder joints*. Applied Physics Letters **88**, 012106 (2006).
[136] C. Durkan and M. E. Welland. *Size effects in the electrical resistivity of polycrystalline nanowires*. Phys Rev B **61**, 14215 (2000).
[137] W. Steinhogl, G. Schindler, G. Steinlesberger, M. Traving, and M. Engelhardt. *Comprehensive study of the resistivity of copper wires with lateral dimensions of 100 nm and smaller*. J Appl Phys **97** (2005).
[138] R. B. Dingle. *The Electrical Conductivity of Thin Wires*. Proc R Soc Lon Ser-A **201**, 545 (1950).
[139] A. F. Mayadas and M. Shatzkes. *Electrical-Resistivity Model for Polycrystalline Films - Case of Arbitrary Reflection at External Surfaces*. Phys Rev B **1**, 1382 (1970).
[140] K. Fuchs. *The conductivity of thin metallic films according to the electron theory of metals*. P Camb Philos Soc **34**, 100 (1938).
[141] J. K. Mackenzie and E. H. Sondheimer. *The Theory of the Change in the Conductivity of Metals Produced by Cold Work*. Phys Rev **77**, 264 (1950).
[142] Z. M. Wu, M. Steinacher, R. Huber, M. Calame, S. J. van der Molen, and C. Schonenberger. *Feedback controlled electromigration in four-terminal nanojunctions*. Applied Physics Letters **91** (2007).
[143] R. Hoffmann, D. Weissenberger, J. Hawecker, and D. Stöffler. *Conductance of gold nanojunctions thinned by electromigration*. Applied Physics Letters **93**, 043118 (2008).
[144] T. Li, W. Hu, and D. Zhu. *Nanogap Electrodes*. Advanced Materials **22**, 286 (2010).
[145] C. Joachim, J. K. Gimzewski, and A. Aviram. *Electronics using hybrid-molecular and mono-molecular devices*. Nature **408**, 541 (2000).
[146] H. B. Akkerman, P. W. M. Blom, D. M. de Leeuw, and B. de Boer. *Towards molecular electronics with large-area molecular junctions*. Nature **441**, 69 (2006).
[147] P. Pang *et al. Fixed-Gap Tunnel Junction for Reading DNA Nucleotides*. ACS Nano **8**, 11994 (2014).
[148] J. He, L. Lin, P. Zhang, and S. Lindsay. *Identification of DNA Basepairing via Tunnel-Current Decay*. Nano Letters **7**, 3854 (2007).
[149] D. O. Bellisario, Z. Ulissi, and M. S. Strano. *A Quantitative and Predictive Model of Electromigration-Induced Breakdown of Metal Nanowires*. The Journal of Physical Chemistry C **117**, 12373 (2013).
[150] C. Y. Lee, W. Choi, J.-H. Han, and M. S. Strano. *Coherence Resonance in a Single-Walled Carbon Nanotube Ion Channel*. Science **329**, 1320 (2010).
[151] S. Huang *et al. Identifying single bases in a DNA oligomer with electron tunnelling*. Nat Nano **5**, 868 (2010).
[152] L. X. Zheng *et al. Ultralong single-wall carbon nanotubes*. Nat Mater **3**, 673 (2004).
[153] Y. Homma, S. Suzuki, Y. Kobayashi, M. Nagase, and D. Takagi. *Mechanism of bright selective imaging of single-walled carbon nanotubes on insulators by scanning electron microscopy*. Applied Physics Letters **84**, 1750 (2004).
[154] A. Jorio, R. Saito, J. H. Hafner, C. M. Lieber, M. Hunter, T. McClure, G. Dresselhaus, and M. S. Dresselhaus. *Structural (n, m) Determination of Isolated Single-Wall Carbon Nanotubes by Resonant Raman Scattering*. Physical Review Letters **86**, 1118 (2001).
[155] D. R. Strachan, D. E. Smith, D. E. Johnston, T.-H. Park, M. J. Therien, D. A. Bonnell, and A. T. Johnson. *Controlled fabrication of nanogaps in ambient environment for molecular electronics*. Applied Physics Letters **86** (2005).

[156] B. D. Knowlton, J. J. Clement, and C. V. Thompson. *Simulation of the effects of grain structure and grain growth on electromigration and the reliability of interconnects*. J Appl Phys **81**, 6073 (1997).
[157] P. Avouris, J. Appenzeller, R. Martel, and S. J. Wind. *Carbon nanotube electronics*. Proceedings of the IEEE **91**, 1772 (2003).
[158] L.-L. Lin, C.-K. Wang, and Y. Luo. *Inelastic Electron Tunneling Spectroscopy of Gold−Benzenedithiol−Gold Junctions: Accurate Determination of Molecular Conformation*. ACS Nano **5**, 2257 (2011).
[159] H. Nakamura. *First-Principles Study for Detection of Inelastic Electron Transport in Molecular Junctions by Internal Substitution*. The Journal of Physical Chemistry C **114**, 12280 (2010).
[160] L. Venkataraman, J. E. Klare, C. Nuckolls, M. S. Hybertsen, and M. L. Steigerwald. *Dependence of single-molecule junction conductance on molecular conformation*. Nature **442**, 904 (2006).
[161] Y. Yourdshahyan and A. M. Rappe. *Structure and energetics of alkanethiol adsorption on the Au(111) surface*. The Journal of Chemical Physics **117**, 825 (2002).
[162] K. M. Beardmore, J. D. Kress, N. Grønbech-Jensen, and A. R. Bishop. *Determination of the headgroup-gold(111) potential surface for alkanethiol self-assembled monolayers by ab initio calculation*. Chemical Physics Letters **286**, 40 (1998).
[163] F. P. Cometto, P. Paredes-Olivera, V. A. Macagno, and E. M. Patrito. *Density Functional Theory Study of the Adsorption of Alkanethiols on Cu(111), Ag(111), and Au(111) in the Low and High Coverage Regimes*. The Journal of Physical Chemistry B **109**, 21737 (2005).
[164] J. C. Love, L. A. Estroff, J. K. Kriebel, R. G. Nuzzo, and G. M. Whitesides. *Self-Assembled Monolayers of Thiolates on Metals as a Form of Nanotechnology*. Chemical Reviews **105**, 1103 (2005).
[165] S. F. Parker. *Assignment of the vibrational spectrum of l-cysteine*. Chemical Physics **424**, 75 (2013).
[166] A. Daire, (Keithley Instruments, Inc., United States, 2005).
[167] P. K. Muduli and R. C. Budhani. *Tunneling magnetoresistance in spin valves exchange biased with metallic antiferromagnet La0.45Sr0.55MnO3*. J Appl Phys **106**, 103924 (2009).
[168] Fitting a Univariate Distribution using Cumulative Probabilities, http://www.mathworks.com/help/stats/examples/fitting-a-univariate-distribution-using-cumulative-probabilities.html.
[169] K. W. Hipps and U. Mazur. *Inelastic electron tunneling: an alternative molecular spectroscopy*. The Journal of Physical Chemistry **97**, 7803 (1993).
[170] M. Galperin, M. A. Ratner, and A. Nitzan. *Inelastic electron tunneling spectroscopy in molecular junctions: Peaks and dips*. The Journal of Chemical Physics **121**, 11965 (2004).
[171] The step is broadened by temperature via Fermi-Dirac distributions at the electrodes, by the measurement resolution, and by the intrinsic linewidth.
[172] M. Kiguchi, H. Nakamura, Y. Takahashi, T. Takahashi, and T. Ohto. *Effect of Anchoring Group Position on Formation and Conductance of a Single Disubstituted Benzene Molecule Bridging Au Electrodes: Change of Conductive Molecular Orbital and Electron Pathway*. The Journal of Physical Chemistry C **114**, 22254 (2010).
[173] M. Kiguchi, T. Ohto, S. Fujii, K. Sugiyasu, S. Nakajima, M. Takeuchi, and H. Nakamura. *Single Molecular Resistive Switch Obtained via Sliding Multiple Anchoring Points and Varying Effective Wire Length*. Journal of the American Chemical Society **136**, 7327 (2014).
[174] A. Nitzan and M. A. Ratner. *Electron Transport in Molecular Wire Junctions*. Science **300**, 1384 (2003).
[175] H. Nakamura, K. Yamashita, A. R. Rocha, and S. Sanvito. *Efficient \textit{ab initio} method for inelastic transport in nanoscale devices: Analysis of inelastic electron tunneling spectroscopy*. Phys Rev B **78**, 235420 (2008).
[176] S. Reich, C. Thomsen, and J. Maultzsch, *Carbon nanotubes : basic concepts and physical properties* (Wiley-VCH, Weinheim ; Cambridge, 2004).
[177] F. Wang, G. Dukovic, L. E. Brus, and T. F. Heinz. *The optical resonances in carbon nanotubes arise from excitons*. Science **308**, 838 (2005).

[178] G. Dukovic, F. Wang, D. H. Song, M. Y. Sfeir, T. F. Heinz, and L. E. Brus. *Structural dependence of excitonic optical transitions and band-gap energies in carbon nanotubes*. Nano Letters **5**, 2314 (2005).
[179] L. Luer, J. Crochet, T. Hertel, G. Cerullo, and G. Lanzani. *Ultrafast Excitation Energy Transfer in Small Semiconducting Carbon Nanotube Aggregates*. Acs Nano **4**, 4265 (2010).
[180] P. Delaney, H. J. Choi, J. Ihm, S. G. Louie, and M. L. Cohen. *Broken symmetry and pseudogaps in ropes of carbon nanotubes*. Nature **391**, 466 (1998).
[181] M. A. Topinka, M. W. Rowell, D. Goldhaber-Gordon, M. D. McGehee, D. S. Hecht, and G. Gruner. *Charge Transport in Interpenetrating Networks of Semiconducting and Metallic Carbon Nanotubes*. Nano Letters **9**, 1866 (2009).
[182] Z. Yang, A. Liao, and E. Pop. *Multiband Mobility in Semiconducting Carbon Nanotubes*. Electron Device Letters, IEEE **30**, 1078 (2009).
[183] F. Jakubka, C. Backes, F. Gannott, U. Mundloch, F. Hauke, A. Hirsch, and J. Zaumseil. *Mapping Charge Transport by Electroluminescence in Chirality-Selected Carbon Nanotube Networks*. ACS Nano **7**, 7428 (2013).
[184] J. Nelson, *The physics of solar cells* (Imperial College Press, London, 2003).
[185] L. Oudjedi, A. N. G. Parra-Vasquez, A. G. Godin, L. Cognet, and B. Lounis. *Metrological Investigation of the (6,5) Carbon Nanotube Absorption Cross Section*. The Journal of Physical Chemistry Letters **4**, 1460 (2013).
[186] L. X. Benedict, S. G. Louie, and M. L. Cohen. *Static polarizabilities of single-wall carbon nanotubes*. Physical Review B **52**, 8541 (1995).
[187] M. S. Arnold, A. A. Green, J. F. Hulvat, S. I. Stupp, and M. C. Hersam. *Sorting carbon nanotubes by electronic structure using density differentiation*. Nat Nano **1**, 60 (2006).
[188] G. D. Nessim, A. J. Hart, J. S. Kim, D. Acquaviva, J. Oh, C. D. Morgan, M. Seita, J. S. Leib, and C. V. Thompson. *Tuning of Vertically-Aligned Carbon Nanotube Diameter and Areal Density through Catalyst Pre-Treatment*. Nano Letters **8**, 3587 (2008).
[189] J. E. Fischer, W. Zhou, J. Vavro, M. C. Llaguno, C. Guthy, R. Haggenmueller, M. J. Casavant, D. E. Walters, and R. E. Smalley. *Magnetically aligned single wall carbon nanotube films: Preferred orientation and anisotropic transport properties*. Journal of Applied Physics **93**, 2157 (2003).
[190] M. Lu, M.-W. Jang, G. Haugstad, S. A. Campbell, and T. Cui. *Well-aligned and suspended single-walled carbon nanotube film: Directed self-assembly, patterning, and characterization*. Applied Physics Letters **94**, 261903 (2009).
[191] K. Yoshikawa, K. Matsuda, and Y. Kanemitsu. *Exciton Transport in Suspended Single Carbon Nanotubes Studied by Photoluminescence Imaging Spectroscopy*. Journal of Physical Chemistry C **114**, 4353 (2010).
[192] F. Wang, Y. Wu, M. S. Hybertsen, and T. F. Heinz. *Auger recombination of excitons in one-dimensional systems*. Physical Review B **73** (2006).
[193] R. D. Mehlenbacher, M. Y. Wu, M. Grechko, J. E. Laaser, M. S. Arnold, and M. T. Zanni. *Photoexcitation Dynamics of Coupled Semiconducting Carbon Nanotube Thin Films*. Nano Letters **13**, 1495 (2013).
[194] T. Dürkop, S. A. Getty, E. Cobas, and M. S. Fuhrer. *Extraordinary Mobility in Semiconducting Carbon Nanotubes*. Nano Letters **4**, 35 (2003).
[195] S. Wang, Z. Liang, B. Wang, and C. Zhang. *Statistical characterization of single-wall carbon nanotube length distribution*. Nanotechnology **17**, 634 (2006).
[196] X. Sun, T. Chen, Z. Yang, and H. Peng. *The Alignment of Carbon Nanotubes: An Effective Route To Extend Their Excellent Properties to Macroscopic Scale*. Accounts of Chemical Research **46**, 539 (2012).
[197] M. Y. Wu, R. M. Jacobberger, and M. S. Arnold. *Design length scales for carbon nanotube photoabsorber based photovoltaic materials and devices*. Journal of Applied Physics **113** (2013).
[198] D. O. Bellisario, R. M. Jain, Z. Ulissi, and M. S. Strano. *Deterministic modelling of carbon nanotube near-infrared solar cells*. Energy & Environmental Science (2014).

[199] L. Maillaud, C. Zakri, I. Ly, A. Pénicaud, and P. Poulin. *Conductivity of transparent electrodes made from interacting nanotubes*. Applied Physics Letters **103** (2013).
[200] Z. H. Ni, H. M. Wang, J. Kasim, H. M. Fan, T. Yu, Y. H. Wu, Y. P. Feng, and Z. X. Shen. *Graphene Thickness Determination Using Reflection and Contrast Spectroscopy*. Nano Letters **7**, 2758 (2007).
[201] T. K. Cherukuri, D. A. Tsyboulski, and R. B. Weisman. *Length- and Defect-Dependent Fluorescence Efficiencies of Individual Single-Walled Carbon Nanotubes*. Acs Nano **6**, 843 (2012).
[202] A. Hagen and T. Hertel. *Quantitative Analysis of Optical Spectra from Individual Single-Wall Carbon Nanotubes*. Nano Letters **3**, 383 (2003).
[203] M. Barkelid, G. A. Steele, and V. Zwiller. *Probing Optical Transitions in Individual Carbon Nanotubes Using Polarized Photocurrent Spectroscopy*. Nano Letters **12**, 5649 (2012).
[204] Approximation Methods, http://farside.ph.utexas.edu/teaching/qm/lectures/node48.html (2013).
[205] N. Wang, Z. K. Tang, G. D. Li, and J. S. Chen. *Materials science: Single-walled 4 A carbon nanotube arrays*. Nature **408**, 50 (2000).
[206] C. Manzoni, A. Gambetta, E. Menna, M. Meneghetti, G. Lanzani, and G. Cerullo. *Intersubband exciton relaxation dynamics in single-walled carbon nanotubes*. Physical Review Letters **94** (2005).
[207] L. Huang, H. N. Pedrosa, and T. D. Krauss. *Ultrafast Ground-State Recovery of Single-Walled Carbon Nanotubes*. Physical Review Letters **93**, 017403 (2004).
[208] J. J. Crochet, J. G. Duque, J. H. Werner, and S. K. Doorn. *Photoluminescence imaging of electronic-impurity-induced exciton quenching in single-walled carbon nanotubes*. Nature Nanotechnology **7**, 126 (2012).
[209] E. Durgun, S. Dag, S. Ciraci, and O. Gulseren. *Energetics and electronic structures of individual atoms adsorbed on carbon nanotubes*. Journal of Physical Chemistry B **108**, 575 (2004).
[210] D. L. Duong, I. H. Lee, K. K. Kim, J. Kong, S. M. Lee, and Y. H. Lee. *Carbon Nanotube Doping Mechanism in a Salt Solution and Hygroscopic Effect: Density Functional Theory*. Acs Nano **4**, 5430 (2010).
[211] F. Wang, G. Dukovic, E. Knoesel, L. E. Brus, and T. F. Heinz. *Observation of rapid Auger recombination in optically excited semiconducting carbon nanotubes*. Physical Review B **70** (2004).
[212] K. Matsuda, T. Inoue, Y. Murakami, S. Maruyama, and Y. Kanemitsu. *Exciton dephasing and multiexciton recombinations in a single carbon nanotube*. Physical Review B **77** (2008).
[213] A. Rajan, M. S. Strano, D. A. Heller, T. Hertel, and K. Schulten. *Length-dependent optical effects in single walled carbon nanotubes*. Journal of Physical Chemistry B **112**, 6211 (2008).
[214] A. R. T. Nugraha, R. Saito, K. Sato, P. T. Araujo, A. Jorio, and M. S. Dresselhaus. *Dielectric constant model for environmental effects on the exciton energies of single wall carbon nanotubes*. Applied Physics Letters **97**, 091905 (2010).