THE THERMOELECTRIC EFFICIENCY OF QUANTUM DOTS IN INAS/INP

NANOWIRES

by

ERIC A. HOFFMANN

A DISSERTATION

Presented to the Department of Physics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2009

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Eric Hoffmann

Title:

"The Thermoelectric Efficiency of Quantum Dots in InAs/InP Nanowires"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Physics by:

Stephen Kevan, Chairperson, Physics
Heiner Linke, Member, Physics
Roger Haydock, Member, Physics
Stephen Hsu, Member, Physics
David Johnson, Outside Member, Chemistry

and Richard Linton, Vice President for Research and Graduate Studies/Dean of the Graduate School for the University of Oregon.

December 12, 2009

Original approval signatures are on file with the Graduate School and the University of Oregon Libraries.

An Abstract of the Dissertation of

Eric A. Hoffmann   for the degree of   Doctor of Philosophy

in the Department of Physics   to be taken   December 2009

Title: THE THERMOELECTRIC EFFICIENCY OF QUANTUM DOTS

IN INAS/INP NANOWIRES

Approved: _____

Dr. Heiner Linke

State of the art semiconductor materials engineering provides the possibility to fabricate devices on the lower end of the mesoscopic scale and confine only a handful of electrons to a region of space. When the thermal energy is reduced below the energetic quantum level spacing, the confined electrons assume energy levels akin to the core-shell structure of natural atoms. Such "artificial atoms", also known as quantum dots, can be loaded with electrons, one-by-one, and subsequently unloaded using source and drain electrical contacts. As such, quantum dots are uniquely tunable platforms for performing quantum transport and quantum control experiments. Voltage-biased electron transport through quantum dots has been studied extensively. Far less attention has been given to thermoelectric effects in quantum dots, that is, electron transport induced by a temperature gradient.

This dissertation focuses on the efficiency of direct thermal-to-electric energy conversion in InAs/InP quantum dots embedded in nanowires. The efficiency of thermoelectric heat engines is bounded by the same maximum efficiency as cyclic heat engines, namely, by Carnot efficiency. The efficiency of bulk thermoelectric materials suffers from their inability to transport charge carriers selectively based on energy. Owing to their three-dimensional momentum quantization, quantum dots operate as electron energy filters—a property which can be harnessed to minimize entropy production and therefore maximize efficiency. This research was motivated by the possibility to realize experimentally a thermodynamic heat engine operating with near-Carnot efficiency using the unique behavior of quantum dots.

To this end, a microscopic heating scheme for the application of a temperature difference across a quantum dot was developed in conjunction with a novel quantum-dot thermometry technique used for quantifying the magnitude of the applied temperature difference. While pursuing high-efficiency thermoelectric performance, many mesoscopic thermoelectric effects were observed and studied, including Coulomb-blockade thermovoltage oscillations, thermoelectric power generation, and strong nonlinear behavior. In the end, a quantum-dot-based thermoelectric heat engine was achieved and demonstrated an electronic efficiency of up to 95% Carnot efficiency.

CURRICULUM VITAE

NAME OF AUTHOR: Eric A. Hoffmann

PLACE OF BIRTH: Rock Springs, Wyoming, USA

DATE OF BIRTH: April 28, 1982

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
University of Puget Sound, Tacoma, Washington

DEGREES AWARDED:

Doctor of Philosophy in Physics, 2009, University of Oregon
Bachelor of Science in Physics, 2004, University of Puget Sound
Bachelor of Science in Mathematics, 2004, University of Puget Sound

AREAS OF SPECIAL INTEREST:

Low-temperature Physics
Mesoscopic Thermoelectric Effects
Nonequilibrium Effects in Mesoscopic Systems

PROFESSIONAL EXPERIENCE:

Graduate Research Assistant,
    University of Oregon, 2009–2005
Graduate Teaching Fellow,
    University of Oregon, 2004–2005

GRANTS, AWARDS AND HONORS:

NSF IGERT Fellowship, Univerisity of Oregon 2007–2008
NSF IGERT Fellowship, Univerisity of Oregon 2006–2007
NSF IGERT Fellowship, Univerisity of Oregon 2005–2006

PUBLICATIONS:

E. A. Hoffmann, H. A. Nilsson, J. E. Matthews, N. Nakpathomkun, A. I. Persson, L. Samuelson, and H. Linke, Nano Letters **9**, 779 (2009).

E. A. Hoffmann and H. Linke, Journal of Low-Temperature Physics **154**, 161 (2009).

E. A. Hoffmann, N. Nakpathomkun, A. I. Persson, H. A. Nilsson, L. Samuelson, and H. Linke, Appl. Phys. Lett. **91**, 252114 (2007).

E. A. Hoffmann, N. Nakpathomkun, A. I. Persson, H. A. Nilsson, L. Samuelson, and H. Linke, Physica E: Low-Dimensional Systems & Nanostructures **40**, 1605 (2008).

M. F. O'Dwyer, H. Linke, E. Hoffmann, T. E. Humphrey, R. A. Lewis and C. Zhang, IEEE Cat. No.: 06TH8931C, 228 (2006).

TABLE OF CONTENTS

LIST OF FIGURES

Figure   Page

CHAPTER I

INTRODUCTION

<u>Thermoelectrics Background</u>

A material is thermoelectric if charge carriers preferentially migrate through the material in response to a temperature difference applied across the material. In the linear regime, the total current through such a material, I, is the sum of the biased current and the thermally induced current,

$$I = GV + G_{\text{th}}\Delta T, \tag{I.1}$$

where G is the isothermal conductance, V is the bias voltage, and ΔT is the magnitude of the applied temperature difference. The thermodynamic coefficient, G_{th}, relates ΔT to electric current. Most thermoelectric measurements are actually voltage measurements performed under open-circuit conditions when $I = 0$. In this way, the magnitude of V_{th} is equal to the bias voltage required to stop the total current flow

$$V_{\text{th}} \equiv \lim_{I=0} V = -\frac{G_{\text{th}}}{G}\Delta T. \tag{I.2}$$

So V_{th} is simply the voltage created solely by ΔT (see Fig. 1.1a). Therefore,

it is natural to define the corresponding Seebeck coefficient (also known as the thermopower), S, which is the coefficient of proportionality between V_{th} and ΔT,

$$S \equiv \frac{V_{\mathrm{th}}}{\Delta T} = -\frac{G_{\mathrm{th}}}{G}. \tag{I.3}$$

The Seebeck coefficient is a physical property of the thermoelectric material and is the thermal analog of electrical conductance.

Thermocouples

Perhaps the most common application and best example of the thermoelectric effect is the thermocouple, where a calibrated thermoelectric material is used to determine an unknown temperature, T_1, relative to a known temperature T_2. To eliminate heat leaks and additional thermovoltages, thermocouples employ two thermoelectric leads, a and b, with differing Seebeck coefficients, S_{a} and S_{b}, respectively. The most common thermocouple, owing to its low cost and wide temperature range, is the type K thermocouple composed of Ni-Cr and Ni-Al alloy leads (see Fig. 1.1b). The voltage difference between the thermocouple leads, V, is the difference between the two thermovoltages (V_{a} and V_{b}) established in each arm

$$V = V_{\mathrm{a}} - V_{\mathrm{b}} = S_{\mathrm{a}}\Delta T - S_{\mathrm{b}}\Delta T = (S_{\mathrm{a}} - S_{\mathrm{b}})\,\Delta T = S_{\mathrm{eff}}\Delta T,$$

where S_{eff} is the effective Seebeck coefficient of the thermocouple and $\Delta T = T_2 - T_1$ is the temperature difference. For the thermocouple to function, it is necessary that the

leads are dissimilar so that $S_a \neq S_b$. In most applications, the known temperature, T_2, is room temperature, and the calibration of the thermocouple instrument is made under this assumption. The sensitivity of the type K thermocouple shown in Fig. 1.1b is only approximately 41 $\mu V/^\circ C$ [1]. So in fact, thermovoltages are relatively small, meaning the thermoelectric effect, is not particularly strong.
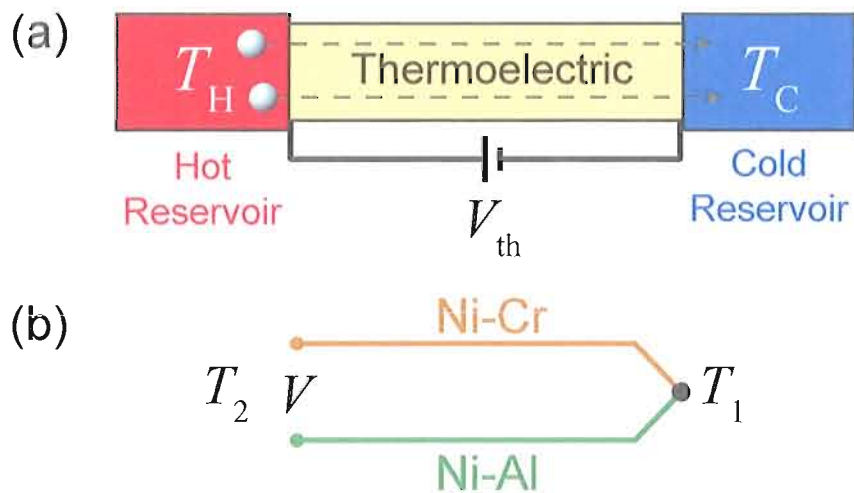
Figure 1.1. **(a)** A schematic demonstrating the thermoelectric effect. Charge carriers traverse a thermoelectric material (or device) in response to a thermal gradient supplied by hot and cold reservoirs at temperatures T_H and T_C, respectively. Subsequent charge build-up creates an open-circuit voltage, defined as the thermovoltage, $V_{th} = S\Delta T$, where $\Delta T = T_H - T_C$, and S is the Seebeck coefficient of the material (see Eq. (I.3)). **(b)** The thermocouple is the most common application of the thermoelectric effect. Here a type K thermocouple using Ni-Cr and Ni-Al alloy thermoelectric leads measures the unknown temperature T_1 relative to known temperature T_2. The voltage V established between the two thermocouple leads is proportional to $T_2 - T_1$. See text for additional thermocouple discussion and Fig. 1.3 for other applications of the thermoelectric effect.

Onsager's Reciprocal Relations Applied to Thermoelectric Coefficients

In the language of thermodynamics, the charge current, I, and heat current, Q, flowing through a thermoelectric material are both thermodynamic currents created in response to two thermodynamic forces, which are derivatives of two thermodynamic potentials. The Onsager formulation relates thermodynamic currents to their thermodynamic potentials via thermodynamic coefficients, such as those appearing in Eq. (I.1). Onsager's famous reciprocal relations express the symmetry between the off-diagonal thermodynamic coefficients of a system in which multiple thermodynamic potentials exist simultaneously [2]. The Onsager relations were first applied to thermoelectrics in 1948 [3], but the treatment is now common in textbooks [4].

In a thermoelectric system, the thermodynamic currents I and Q can be related to the thermodynamic forces V and ΔT by,

$$\begin{bmatrix} I \\ Q \end{bmatrix} = \begin{bmatrix} G & G_{\text{th}} \\ M & K \end{bmatrix} \begin{bmatrix} V \\ \Delta T \end{bmatrix}, \tag{I.4}$$

where G and G_{th} are the coefficients appearing in Eq. (I.1), and M and K are the analogous heat-current thermodynamic coefficients. Unfortunately, V and ΔT are not the proper thermodynamic forces of this system, and the relationship between M and G_{th} is not immediately provided by the Onsager relations. Nonetheless, the system's proper thermodynamic forces can be found [3], and they provide the corresponding

Onsager relation given by [4],

$$M = -G_{\text{th}}T. \tag{I.5}$$

Using this result and Eq. (I.3) for the Seebeck coefficient, M can be written

$$M = -G_{\text{th}}T = GST. \tag{I.6}$$

The heat flux portion of Eq. (I.4) can be understood better if this relation is expressed with I as the dependent variable [4],

$$\begin{bmatrix} V \\ Q \end{bmatrix} = \begin{bmatrix} R & S \\ \Pi & -\kappa_{\text{el}} \end{bmatrix} \begin{bmatrix} I \\ \Delta T \end{bmatrix}, \tag{I.7}$$

where $R = 1/G$ is resistance, S is the Seebeck coefficient, Π is the Peltier coefficient, and κ_{el} is the electrical contribution to the total thermal conductance. The real advantage of Eq. (I.7) is that the thermodynamic coefficients Π and κ_{el} are very familiar. That is, κ_{el} dictates the amount of heat, Q, which flows in response to a thermal difference, ΔT, in the absence of electrical current,

$$-\kappa_{\text{el}} \equiv \lim_{I=0} \frac{Q}{\Delta T}. \tag{I.8}$$

Similarly, Π, the Peltier coefficient, dictates the amount of heat, Q, carried by an

electrical current, I, in the absence of a temperature gradient,

$$\Pi \equiv \lim_{\Delta T = 0} \frac{Q}{I}. \tag{I.9}$$

The corresponding definition of S has already been presented in Eq. (I.3), and the definition of R needs no explanation. On the other hand, any physical intuition of coefficients M and K is remote. In some sense, Eq. (I.4) is the theorist's equation and Eq. (I.7) is the experimentalist's equation. The primary utility of M and K is the relationship between M and G_{th}—afforded by the Onsager relation (Eq. (I.5))—and the ease with which G, G_{th}, and K can be expressed analytically (see Eq. I.24).

Furthermore, it can be shown that knowledge of G, G_{th}, and K is equivalent to knowledge of R, Π, and κ_{el}. Inserting the expression for V from Eq. (I.7) into the expression for Q in Eq. (I.4) gives

$$\begin{aligned}
Q &= MV + K\Delta T \\
&= M\left(RI + S\Delta T\right) + K\Delta T \\
&= MRI + \left(K + MS\right)\Delta T \\
&= \Pi I - \kappa_{\mathrm{el}}\Delta T, \tag{I.10}
\end{aligned}$$

where the last equality follows from Eq. (I.7). Now the coefficients Π and κ_{el} are

easily identified:

$$\Pi = MR = \frac{M}{G} = -\frac{G_{\text{th}}}{G}T = ST \tag{I.11a}$$

$$-\kappa_{\text{el}} = K + MS = K + S^2GT, \tag{I.11b}$$

where Eq. (I.6) has been used. Therefore, the thermodynamic quantities which describe all thermoelectric processes can be efficiently managed using the Onsager relations. While the above calculations are specific to thermoelectrics, Onsager's theorem provides a very general and unified description of both thermal Brownian and thermoelectric heat engines using the language of thermodynamic forces [5].

Phonons

From both theoretical and experimental perspectives, the effect of phonons on the electronic properties of thermoelectrics is a very challenging problem. Phonons constantly traverse the thermoelectric, in both directions, but on average, the phonons carry heat from the hot reservoir to the cold reservoir. This heat leak spoils the temperature gradient for the electrons and reduces the performance of the thermoelectric. If both electrons and phonons are considered, then the total thermal conductance, κ, is

$$\kappa = \kappa_{\text{el}} + \kappa_{\text{ph}}, \tag{I.12}$$

where κ_{ph} is the phonon contribution to the thermal conductance and can be much larger than κ_{el}. The research presented here focuses on understanding only the

electronic aspects of thermoelectrics in hopes to improve overall thermoelectric performance. Future research must merge insights regarding electronic optimization with phonon suppression techniques in order to achieve a high-performance thermoelectric.

Introduction to Particle Heat Engines

A particle heat engine converts heat energy into useful output work using a sustained flow of particles. Specifically, a thermoelectric particle heat engine transfers heat, Q, between hot and cold reservoirs via the flow of charge carriers through a thermoelectric material, whereby useful work, W, is extracted as each charge carrier does work against the electric field, $\vec{E}$, which is created by the thermovoltage, V_{th}. Thermoelectrics have no moving parts and require little or no maintenance throughout their lifetime. Unfortunately, their poor efficiency limits their usefulness. A diagram of a (thermoelectric) particle heat engine is shown in Fig. 1.2.

In analogy to traditional heat engines, the present state-of-the-art particle heat engines have developed only as far as the steam engine. Throughout most of their existence, improvements in particle heat engines have been the results of efforts in material engineering. The thermoelectrics field is groping to understand the physical limitations which have so-far prevented the development of scalable, efficient thermoelectric materials. See, for example, Ref. [6] for a very honest evaluation of the progress in the thermoelectrics field. Although thermoelectric phenomena were discovered nearly 200 years ago, the fundamental understanding of particle heat
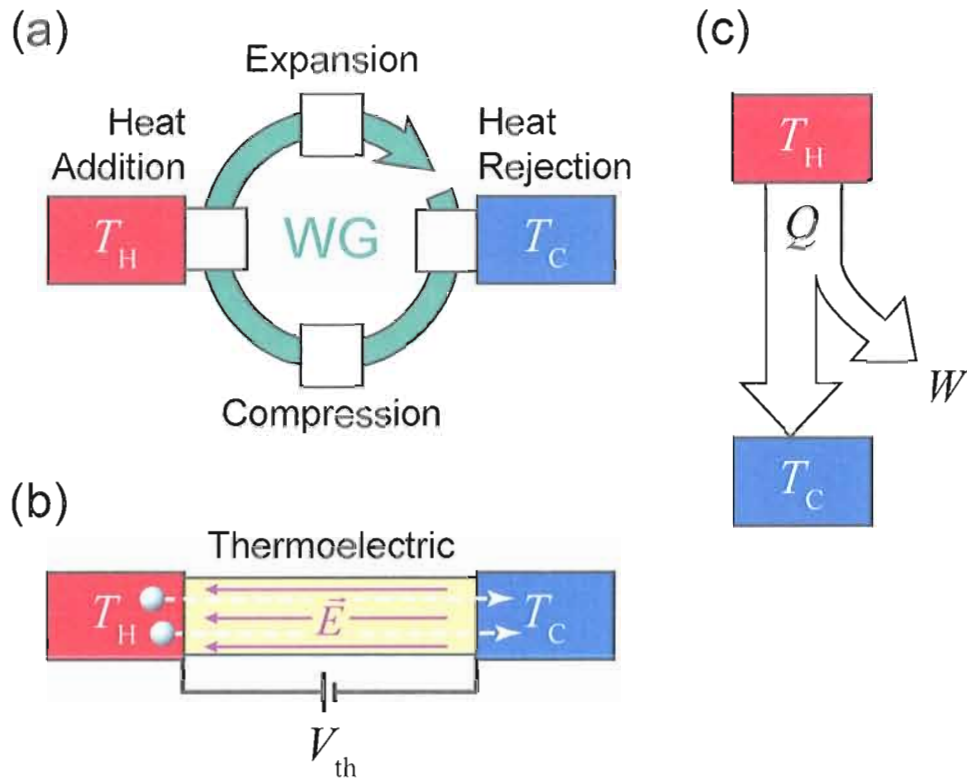
Figure 1.2. Schematics of cyclic and particle heat engines. (**a**) Cyclic heat engines use thermodynamic processes to extract (mechanical) work, W, from the heat, Q, which is transfered (per cycle) via a working gas (WG) from the hot bath at temperature, T_H, to the cold bath at temperature, T_C. To operate with Carnot efficiency, the thermodynamic cycle of the cyclic heat engine must be reversible. (**b**) A particle heat engine, in which a thermoelectric material is in constant contact with both temperature baths. Charge carriers constantly and irreversibly transport heat through the thermoelectric material as they migrate from the hot reservoir to the cold reservoir. The electric field, $\vec{E}$, and associated thermovoltage, V_{th}, are established by the asymmetric charge build-up in the reservoirs. The charge carriers do work against $\vec{E}$ on their way to the cold bath; this is useful (electrical) work, W, extracted from the inter-reservoir heat flow, Q. (**c**) All heat engines can be represented abstractly as work extraction mechanisms. An amount of useful work, W, is extracted from an amount of heat, Q, flowing from the hot to the cold reservoir. The cyclic heat engine extracts mechanical work while the particle heat engine extracts electrical work. The difference $Q - W$ is energy sacrificed in order to maintain non-negative net entropy production, as required by the second law of thermodynamics.

engines lags behind that of cyclic heat engines. As a result, particle heat engines do not contribute to society's general power generation and refrigeration needs and are instead reserved for niche applications. The lessons taught to us by the cyclic heat engine—the concept of entropy and the second law—can be used to understand and improve the particle heat engine.

<div align="center">The Electronic Efficiency of Particle Heat Engines</div>

In the early 19th century, hoping to improve the understanding of steam engines, Sadi Carnot investigated the physical limitations of cyclic heat engines. In one full cycle of such an engine, useful work, W, is extracted from input heat, Q, flowing from a reservoir at temperature T_{H} to a colder reservoir at temperature T_{C}, where $T_{\mathrm{H}} \geq T_{\mathrm{C}}$ (see Fig. 1.2). What Carnot discovered [7] was that the efficiency of this extraction is bounded,

$$\eta = \frac{W}{Q} \leq 1 - \frac{T_{\mathrm{C}}}{T_{\mathrm{H}}}. \tag{I.13}$$

The maximum efficiency—the so-called Carnot efficiency, η_{C}—is achieved when the cycle operates reversibly and the net entropy production is zero. Reversibility is the hallmark of Carnot efficiency and has become a mantra in thermodynamics.

The initial outlook on the efficiency of particle heat engines from the viewpoint of reversibility seems bleak. Particle heat engines are always in direct thermal contact to both the heat bath and the cold bath; irreversible heat flow through this permanent connection is unavoidable. And so it appears that thermoelectrics
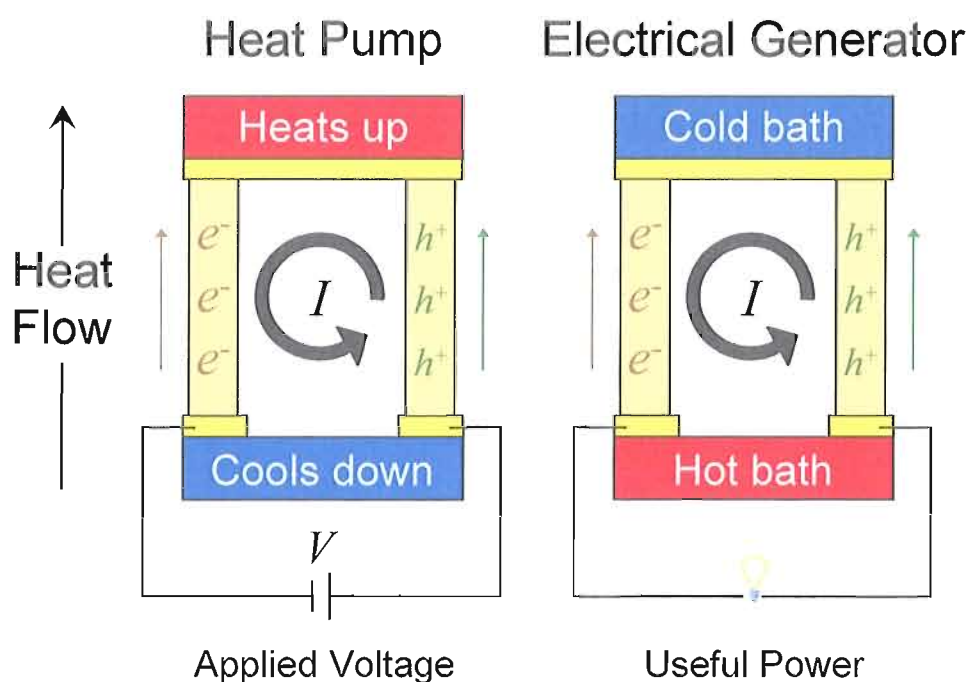
Figure 1.3. Schematics of a thermoelectric heat pump and thermoelectric power generator. Any thermoelectric can in principle be used either as a heat pump or an electric power generator. For simplicity of design, two thermoelectric materials—one n-type and one p-type—are placed in parallel so as to conserve the net charge and create a current loop. For the heat pump, an applied electric voltage, V, maintains a flow of electric charge through the device. Both the electrons and holes create thermal gradients across their respective n-type and p-type thermoelectric materials. In the case of the electric power generator, the hot and cold baths induce electron and hole transport in the n-type and p-type thermoelectric materials, respectively. These transported charge carriers create a net electric current, which can be used to power an electric load. In both applications, the heat flow is upward and the current is counter-clockwise.

cannot operate reversibly and therefore cannot achieve Carnot efficiency. What this analysis is missing, however, is the effect of "loading". The presence of a single dissipative heat flux does not preclude reversibility so long as there exists a second heat flux which is equal in magnitude and opposite in sign so that net entropy production is zero. Moreover, this opposing flux need not be in response to a thermal gradient, but can result from a separate thermodynamic force. Loading is simply the presence of such a force. In a thermoelectric system, the loading is provided by a gradient in the electrochemical potential via a voltage bias, namely the thermovoltage. Therefore, the obvious theoretical tool for analyzing the efficiency of thermoelectric heat engines is the thermodynamics of Fermi gases (statistical mechanics). Through this theoretical framework, the interplay between thermal gradients and gradients in the electrochemical potential becomes natural. This theory is presented in the next section and then applied specifically to thermoelectric efficiency.

Fermi-Dirac Statistics and Thermoelectrics

Most introductions to thermoelectrics exclude Fermi statistics and instead rely on the Drude model (see, for example, Ref. [8]), which ignores electron spin (among other things). Electron spin and the resulting Fermi-Dirac distribution is essential for understanding the thermodynamics of thermoelectrics. Moreover, all the experiments described in this dissertation where performed at low temperature (< 10 K), where Fermi statistics are indispensable for describing the transport of charge carriers.

Therefore, it is worthwhile to review Fermi-Dirac statistics with special attention applied to thermoelectrics.

In the context of thermoelectrics, it is natural to assume that the electron-gas reservoirs are well behaved so that they have well-defined temperatures and electrochemical potentials, regarded as the thermodynamic state variables appearing in the Fermi-Dirac distributions

$$f_{\mathrm{H,C}}^{-1}(\varepsilon) = 1 + \exp\left[\xi_{\mathrm{H,C}}\right], \tag{I.14}$$

where the arguments, $\xi_{\mathrm{H,C}}$, in the exponential function are

$$\xi_{\mathrm{H,C}} = \frac{\varepsilon - \mu_{\mathrm{H,C}}}{kT_{\mathrm{H,C}}} = \frac{\varepsilon - (\mu \mp eV/2)}{kT_{\mathrm{H,C}}} = \frac{\varepsilon - \mu \pm eV/2}{kT_{\mathrm{H,C}}}, \tag{I.15}$$

where ε is the electron energy, $T_{\mathrm{H,C}}$ are the respective hot and cold temperatures, μ is controlled by the gate voltage, and V is the bias voltage, which is assumed to be applied symmetrically across the device. Thus, a positive bias voltage raises and lowers the electrochemical potentials of the cold and hot reservoirs by an equal amount, respectively.

Because Fermi-Dirac distributions will appear often throughout this work, it is convenient to introduce a simple graphical representation, as shown in Fig. 1.4. Migrating away from device schematics such as those in Fig. 1.3 toward an abstract theoretical picture is also very beneficial for understanding electron transport in

thermoelectrics. Akin to band-energy diagrams, the energy of electrons within a thermoelectric device paints a clear picture of the underlying physical mechanisms. Fig. 1.5b shows how a thermoelectric pathway between two thermal baths can be considered an energy-dependent conduit for charge carriers. An assumption is made that the thermal baths are ideal. That is, their temperature is constant, and there is no electrical contact resistance between the baths and the thermoelectric. A second assumption is that charge transport occurs at a single energy. That is, the charge carrier travels ballistically through the thermoelectric, which is a common assumption in mesoscopic systems such as the one used in this research. These assumptions do not restrict complexity or limit physical relevance [9]. In this theoretical framework, all the interesting physics take place within the thermoelectric material itself; a specific model can be made arbitrarily simple or complex, depending on the physical system being considered.

Now that the Fermi-Dirac distributions have been established, the sign convention of the Seebeck coefficient, S, can be discussed. Because G has the opposite sign of G_{th} in Eq. (I.3), the thermally induced current, the thermocurrent, $I_{th} = G_{th}\Delta T$, flows "backwards" relative to the thermovoltage, V_{th},

$$I_{th} = G_{th}\Delta T = -GV_{th}. \tag{I.16}$$

This can create some confusion regarding the sign of V_{th} and therefore S. Indeed,
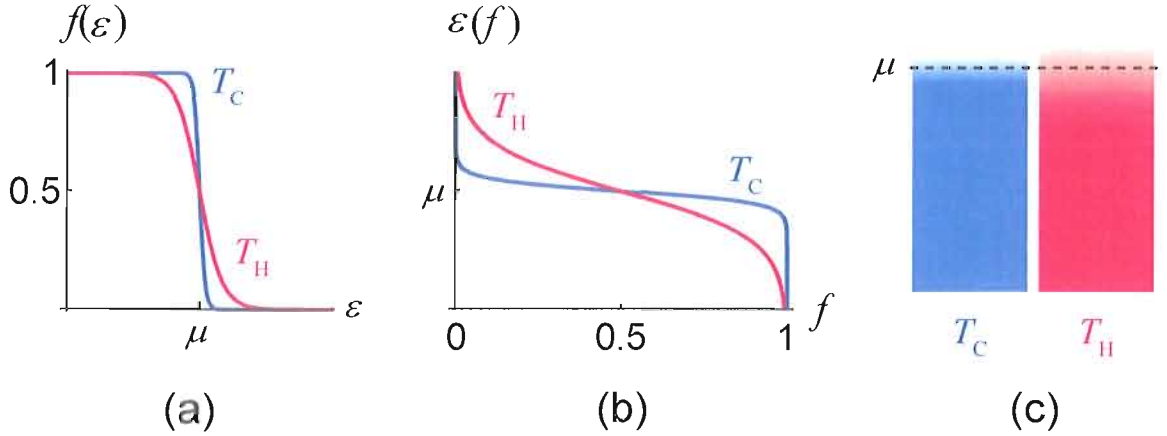
Figure 1.4. The Fermi-Dirac distributions of hot and cold electron reservoirs. **(a)** The functions $f_{H,C}(\varepsilon)$ (Eq. (I.14)) plotted as a function of electron energy, ε, at two different temperatures demonstrate two different thermal smearing at the electrochemical potential, $\mu = \mu_H = \mu_C$. **(b)** The inverse of (a), that is, ε as a function of f. **(c)** A graphical representation of (b), which uses color gradients to represent the thermal smearing at $\mu_{H,C}$. Hot and cold baths are always colored red and blue, respectively.

the sign convention of S is not consistent in the literature. In the spirit of electron transport from a solid state physics perspective, it is most natural to regard S as negative when transport (current) is electron-like and positive when hole-like. This implies that V_{th} must be measured in the opposite sense as I_{th}. This is achieved experimentally by measuring V_{th} as the potential difference from drain to source, that is, from cold reservoir to hot reservoir. This sign convention is made clear in Fig. 1.5.

Thermoelectric Heat Pumps at Carnot Efficiency

The objective of this section is to find the maximum (electronic) efficiency of an ideal thermoelectric heat engine in its most fundamental incarnation: the ballistic
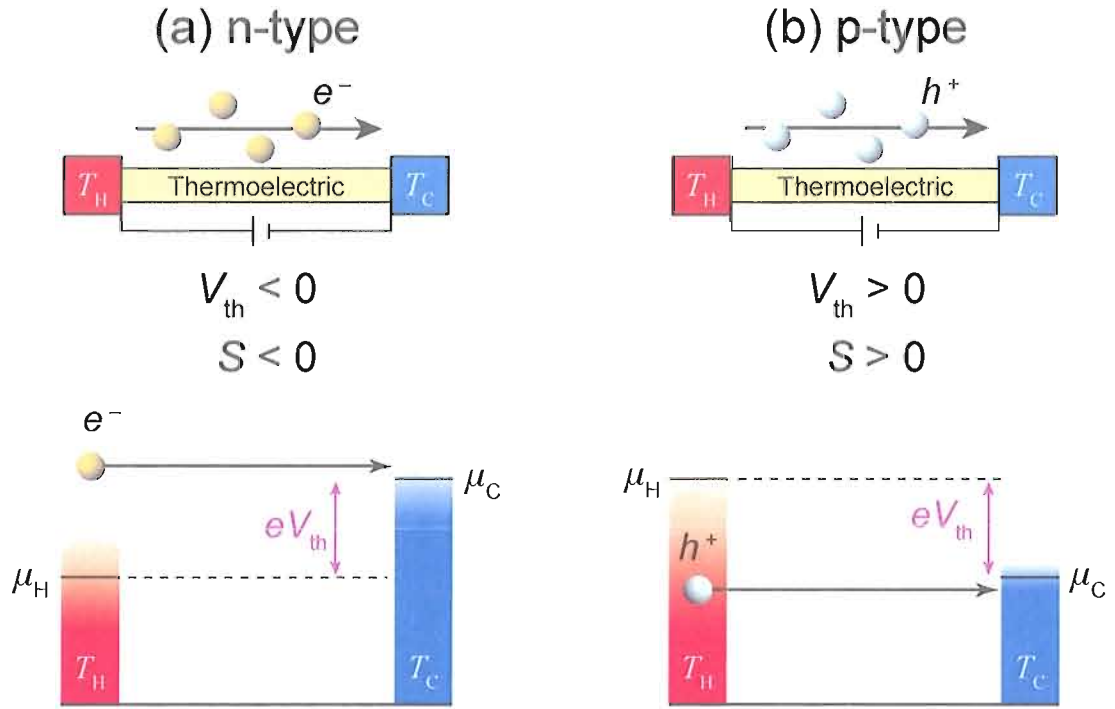
Figure 1.5. Charge transport in n- and p-type thermoelectrics. The signs of the thermovoltage, V_{th} and Seebeck coefficient, S, are determined by the direction of electrical current relative to the temperature gradient. *Top*: **(a)** Electrons (e^-) flow from hot to cold in the n-type thermoelectric, so that $V_{th} < 0$ and $S < 0$. **(b)** Holes (h^+) flow from hot to cold in the p-type thermoelectric, so that $V_{th} > 0$ and $S > 0$. *Bottom*: **(a)** A thermally excited electron traverses the thermoelectric device and fills an unoccupied electron state in the cold Fermi-Dirac distribution. **(b)** A hole created deep in the hot Fermi-Dirac distribution traverses the thermoelectric device and fills an unoccupied hole state in the cold Fermi-Dirac distribution. Because thermally induced charge current flows in "opposition" to the thermovoltage (see Eq. (I.16)), the thermovoltage is related to the electrochemical potentials $(\mu_{H,C})$ by $eV_{th} = \mu_H - \mu_C$.

transport (transport without energy loss) of a single electron at energy, ε, from the hot electron reservoir to the cold electron reservoir through the thermoelectric material (see Fig. 1.6). This quick efficiency derivation closely follows the original theory [10].

In the hot reservoir, the energy, $\varepsilon - \mu_H$, is the average amount of thermal energy required to excite an equilibrium electron at the hot electrochemical potential, μ_H, to energy ε. This amount of energy, $\varepsilon - \mu_H$, is subsequently removed from the hot reservoir when the electron leaves. When the electron enters the destination cold electron reservoir, its free energy dissipates into the surrounding electron sea through a thermalization process (via electron-electron interactions). Therefore, the electron deposits work into the destination electron reservoir, as well as an associated amount of entropy. To quantify this, upon arrival in the cold reservoir, the electron will on average cool to the cold electrochemical potential, μ_C, thereby depositing an amount of work, $\varepsilon - \mu_C$, into the cold reservoir. The respective amounts of work, δQ_H and δQ_C, removed from and added to the hot and cold reservoirs are therefore

$$\delta Q_H = -\left(\varepsilon - \mu_H\right) \tag{I.17a}$$

$$\delta Q_C = \varepsilon - \mu_C. \tag{I.17b}$$

Note that the overall minus sign in δQ_H reflects that the outgoing electron removed

energy from the hot reservoir. The associated changes in entropy are

$$\delta S_{\mathrm{H}} = \frac{\Delta Q_{\mathrm{H}}}{T_{\mathrm{H}}} = -\frac{\varepsilon - \mu_{\mathrm{H}}}{T_{\mathrm{H}}}$$

$$\delta S_{\mathrm{C}} = \frac{\Delta Q_{\mathrm{C}}}{T_{\mathrm{C}}} = \frac{\varepsilon - \mu_{\mathrm{C}}}{T_{\mathrm{C}}}.$$

The net change in entropy of the system, ΔS, is the sum of these two entropies

$$
\begin{aligned}
\Delta S &= \delta S_{\mathrm{C}} + \delta S_{\mathrm{H}} \\
&= \frac{\varepsilon - \mu_{\mathrm{C}}}{T_{\mathrm{C}}} - \frac{\varepsilon - \mu_{\mathrm{H}}}{T_{\mathrm{H}}} \\
&= \frac{(\varepsilon - \mu_{\mathrm{C}})\, T_{\mathrm{H}}}{T_{\mathrm{C}} T_{\mathrm{H}}} - \frac{(\varepsilon - \mu_{\mathrm{H}})\, T_{\mathrm{C}}}{T_{\mathrm{H}} T_{\mathrm{C}}} \\
&= \frac{(\varepsilon - \mu_{\mathrm{C}})\, T_{\mathrm{H}} - (\varepsilon - \mu_{\mathrm{H}})\, T_{\mathrm{C}}}{T_{\mathrm{H}} T_{\mathrm{C}}} \\
&= \frac{(T_{\mathrm{H}} - T_{\mathrm{C}})\, \varepsilon - \mu_{\mathrm{C}} T_{\mathrm{H}} + \mu_{\mathrm{H}} T_{\mathrm{C}}}{T_{\mathrm{H}} T_{\mathrm{C}}} \\
&= \frac{T_{\mathrm{H}} - T_{\mathrm{C}}}{T_{\mathrm{H}} T_{\mathrm{C}}} \left(\varepsilon - \frac{\mu_{\mathrm{C}} T_{\mathrm{H}} - \mu_{\mathrm{H}} T_{\mathrm{C}}}{T_{\mathrm{H}} - T_{\mathrm{C}}} \right).
\end{aligned}
\tag{I.18}
$$

The last equality is only guaranteed if $T_{\mathrm{H}} \neq T_{\mathrm{C}}$. If instead $T_{\mathrm{H}} = T_{\mathrm{C}}$, then $\Delta S = 0$ only when $\mu_{\mathrm{H}} = \mu_{\mathrm{C}}$, which is of course the trivial solution. Otherwise, $\Delta S = 0$ when transport proceeds at the "adiabatic energy", E_{A}, this is, when

$$\varepsilon = E_{\mathrm{A}} \equiv \frac{\mu_{\mathrm{C}} T_{\mathrm{H}} - \mu_{\mathrm{H}} T_{\mathrm{C}}}{T_{\mathrm{H}} - T_{\mathrm{C}}}. \tag{I.19}$$

Eq. (I.18) states that the net entropy production is negative for electrons with energy

$\varepsilon < E_A$, indicating that on average electrons do not flow in this direction (hot to cold) at these energies . When electrons flow in the opposite direction (cold to hot) the sign of ΔS changes. So electrons with energy $\varepsilon < E_A$ must flow from the cold reservoir to the hot reservoir. Therefore, Eq. (I.18) determines at which energy ΔS is zero and also the direction of electron flow as a function of energy. These conclusions are corroborated by using probability arguments based on the Fermi-Dirac distributions of the hot and cold reservoirs (see Fig. 1.6c).

Fig. 1.6 illustrates three possible situations for charge transport between two electron reservoirs. The first case (Fig. 1.6a) is familiar biased transport under isothermal conditions in which an applied bias voltage, V, creates a difference in electrochemical potentials, $\mu_C - \mu_H = eV$. The electron moves from right to left, thermalizing in the destination electron reservoir. Note that the direction of particle flux in Fig. 1.6a is in opposition to the particle flux in Fig. 1.6b in which thermally induced electron transport occurs between the two electron reservoirs that are at equal electrochemical potential, $\mu_H = \mu_C$, but unequal temperature, $T_H \neq T_C$. In this way, a difference in electrochemical potentials can create a thermodynamic force which opposes the thermodynamic force due to a temperature gradient. The loading condition required to achieve reversibility is satisfied when these forces are equal in magnitude. Fig. 1.6c demonstrates this situation in which transport at energy E_A is reversible. Note as well that the hot and cold Fermi-Dirac distributions cross at E_A,
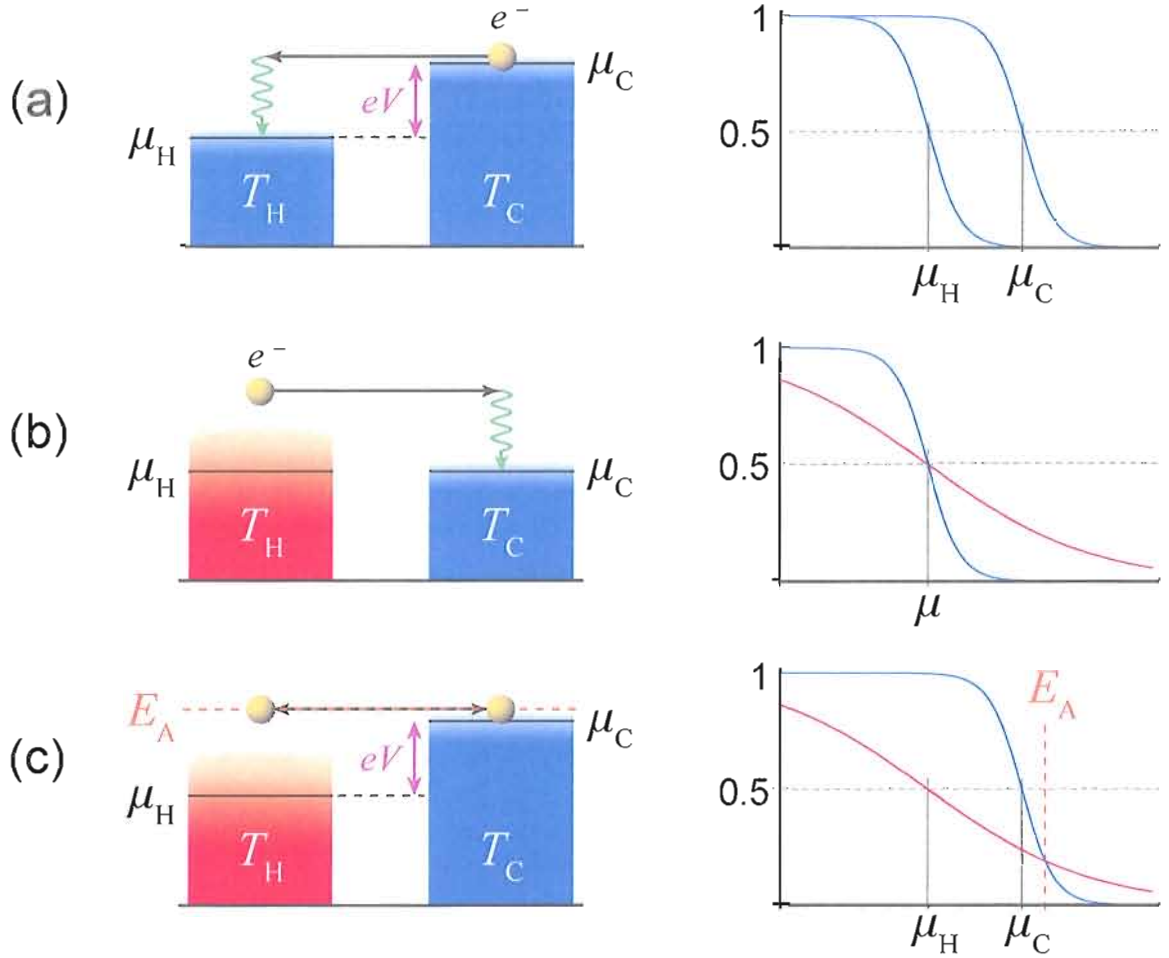
Figure 1.6. Ballistic transport between two electron reservoirs. **(a)** Voltage-baised transport ($\mu_H \neq \mu_C$) under isothermal conditions ($T_H = T_C$). The difference in electrochemical potentials, $\mu_C - \mu_H = eV$, allows a single electron to flow from the cold reservoir to the hot reservoir. Upon arrival it thermalizes, doing work on the hot reservoir. The Fermi-Dirac distributions of the reservoirs are shown on the right. **(b)** Thermally induced transport ($T_H \neq T_C$ and $\mu_H = \mu_C = \mu$) removes work from the hot reservoir and deposits work into the cold reservoir (see Eq. (I.17)). The corresponding Fermi-Dirac distributions cross at μ. **(c)** Under the influence of both a thermal bias ($T_H \neq T_C$) and voltage bias ($\mu_H \neq \mu_C$), the system is allowed to operate reversibly at the adiabatic energy, E_A, with Carnot efficiency (see Eq. (I.19)). Note that E_A is the energy at which the two Fermi-Dirac functions cross. Therefore, electrons with $\varepsilon > E_A$ flow from hot to cold and in the opposite direction when $\varepsilon < E_A$. Electrons at E_A can flow in both directions, indicating reversible transport.

meaning that an electron with energy E_A is just as likely to be in the hot reservoir as the drain reservoir, and both the heat current and the charge current at E_A stalls.

It is now possible to calculate the efficiency achieved by the reversible particle heat engine, which is the ratio of useful work, W, to the input heat, Q. For a thermoelectric system, the useful work emerges as the difference in electrochemical potentials, $W = |\mu_C - \mu_H|$, whereas input heat, Q, is simply the magnitude of energy extracted from the hot reservoir during electron excitation (see Eq. (I.17a)) so that $Q = |\delta Q_H| = \varepsilon - \mu_H$. The efficiency is therefore

$$\eta = \frac{W}{Q} = \frac{\mu_C - \mu_H}{\varepsilon - \mu_H}.$$

Calculate of the efficiency in the adiabatic limit where $\varepsilon = E_A$ is straightforward. Using Eq. (I.19) for E_A provides

$$
\begin{aligned}
\eta &= \frac{\mu_C - \mu_H}{(\mu_C T_H - \mu_H T_C) / (T_H - T_C) - \mu_H} \\
&= \frac{(\mu_C - \mu_H)(T_H - T_C)}{(\mu_C T_H - \mu_H T_C) - (T_H - T_C)\mu_H} \\
&= \frac{(\mu_C - \mu_H)(T_H - T_C)}{(\mu_C - \mu_H)T_H + (\mu_H - \mu_H)T_C} \\
&= \frac{\mu_C - \mu_H}{\mu_C - \mu_H}\frac{T_H - T_C}{T_H} \\
&= 1 - \frac{T_C}{T_H} \\
&= \eta_C.
\end{aligned}
$$

This proves that in principle a particle heat engine can indeed operate reversibly, and when doing so, operates at Carnot efficiency.

Thermoelectrics in Reduced Dimensions

The generally poor thermoelectric performance of bulk homogeneous materials can be partly attributed to entropy production by electron transport and partly due to phonon heat leaks. When entropy-producing electron transport is allowed, thermoelectric efficiency suffers. The fundamental challenge is to discourage phonon flow through the thermoelectric lattice while at the same time encouraging electron flow near the adiabatic limit (see Eq. (I.19)). Thus electronic optimization for the purpose of thermoelectric efficiency requires energy selectivity, that is, the ability to filter electrons based on their energy. Bulk materials cannot meet this challenge. However, through band engineering or density of states (DOS) engineering, electron transport can be modified in an energy-selective manner to enhance efficiency. This has been shown theoretically [10–14] and experimentally [15, 16]. It is widely accepted in the field of thermoelectrics that nanoscale materials are uniquely suited for electronic optimization while simultaneously reducing phononic heat losses [17]. As is true for many solid-state energy applications [18], nanotechnology is the future of thermoelectrics.

Current thermoelectrics research is in general focused on either phonon reduction or electrical optimization. Phonon reduction in atomically layered superlattice

structures [19] is a good example of nanoscale thermoelectric engineering. However, electrical optimization in such systems has proved more difficult. Of the vast multitude of attempts to increase thermoelectric performance in layered materials (primarily via phonon reduction), few have been successful (see Fig. 1.9). Research on phonon reduction in thin-film systems will certainly influence the future of thermoelectrics.
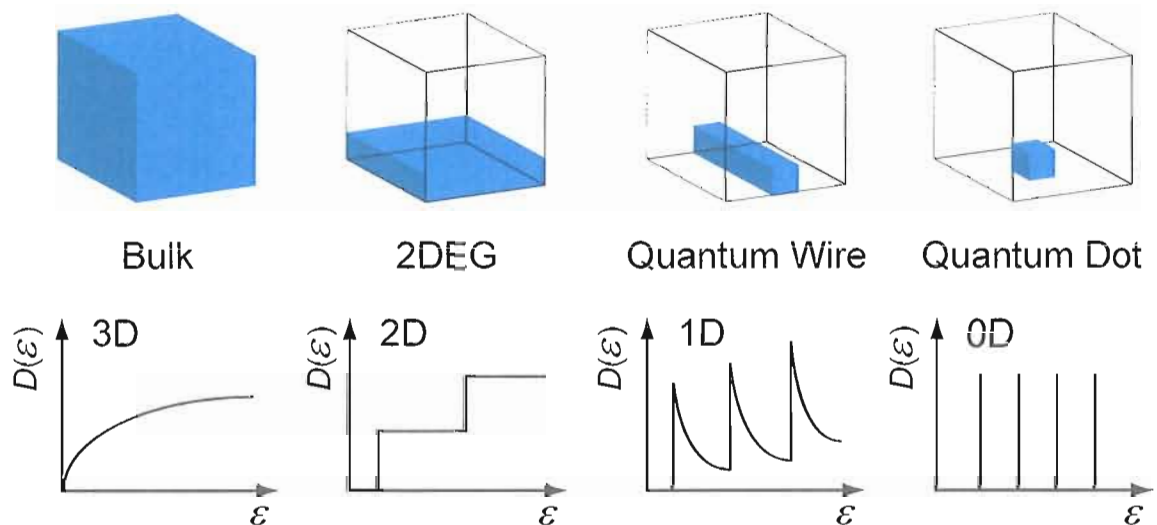
Figure 1.7. The charge-carrier density of states (DOS) in different dimensionalities. As material is chiseled away, a bulk material transitions into a two-dimensional electron gas (2DEG), than a quantum wire, and finally a quantum dot. The associated DOS, $D(\varepsilon)$, as a function of energy, ε, drastically changes as the dimensionality is reduced. The sharp DOS modulations of 1D and 0D systems can be harnessed to enhance the efficiency of nanostructure-based thermoelectric materials.

The research presented here focuses primarily on electronic optimization by means of electron energy filtering. A very simple way to filter electrons by energy is to reduce the dimensionality of the system thereby reducing the charge carrier DOS.

The (electronic) dimensionality of a system is reduced when one or more of its spacial dimensions is made smaller then the Fermi wavelength of the conduction electrons. Fig. 1.7 illustrates how the DOS changes as the dimensionality of the system is reduced from three-dimensional (3D) down to zero-dimensional (0D). Band-gap engineering and reduced dimensionality offer the possibility to place strong energy restrictions on the charge carriers.

In that direction, nanowires are being intensely explored as possible high-performance thermoelectrics motivated by their reduced dimensionality and low thermal conductivity owing (in part) to their large area-to-volume ratio [20–23]. On the other hand, phonon drag has been credited with increasing the thermoelectric performance of silicon nanowires [23]. Another approach toward energy filtering is the use of a single nanoscale heterostructure barrier embedded in a bulk material, which passively selects hot electrons for emission resulting in evaporative cooling [24]. A myriad of other complex nanoscale materials are currently being explored for thermoelectric applications; see Ref. [25] for a review of such systems. For the remainder of this work, the focus is turned toward 1D and 0D systems which operate as electron energy filters.

The Landauer Formalism

The Landauer formalism describes ballistic transport in one-dimensional (1D) constrictions. The original paper [26] introduced a new interpretation of transport physics in which carrier flow (subject to *incoherent* scattering) gives way to an

asymmetric charge accumulation and therefore the electric field. This contrasts the notion that the applied field is the source of carrier transport and resulting charge current. Later Landauer [27] reported a more general relationship between transmission and conductance, and the original Landauer formula appearing in this text related the two-terminal, zero-temperature electrical conductance of a 1D channel to its transmission probability. Later a multi-terminal formula was derived by Büttiker [28] (the Landauer-Büttiker formula). The ubiquitous Landauer formalism of today predicts the current produced when charge carriers flow through a 1D channel in a quantum mechanically coherent and elastic fashion. In this thesis, the Landauer formalism is the primary theoretical tool with which experiments will be analyzed.

The Landauer formula states that the two-terminal, finite-temperature current (per conduction mode) through a 1D channel is [29, 30]

$$I = -\frac{2e}{h} \int \left[f_{\mathrm{H}}\left(\varepsilon\right) - f_{\mathrm{C}}\left(\varepsilon\right)\right] \tau\left(\varepsilon\right) d\varepsilon, \tag{I.20}$$

where $\tau\left(\varepsilon\right)$ is the energy-dependent quantum transmission function of the channel, and $f_{\mathrm{H,C}}$ are the Fermi-Dirac distributions in the source and drain reservoir, respectively (see Eq. (I.14). Note that the sign convention chosen here for I provides negative current when excess electrons in the hot source are transported into the cold drain (as is measured experimentally).

The electrical conductance, G, of the channel is easily calculated using Eq. (I.20)

$$
\begin{aligned}
G &= \frac{dI}{dV} \\
&= -\frac{2e}{h} \int \left[\frac{df_{\mathrm{H}}}{dV} - \frac{df_{\mathrm{C}}}{dV} \right] \tau\left(\varepsilon\right) d\varepsilon \\
&= -\frac{2e}{h} \int \left[\frac{d\xi_{\mathrm{H}}}{dV} \frac{df_{\mathrm{H}}}{d\xi_{\mathrm{H}}} - \frac{d\xi_{\mathrm{C}}}{dV} \frac{df_{\mathrm{C}}}{d\xi_{\mathrm{C}}} \right] \tau\left(\varepsilon\right) d\varepsilon \\
&= -\frac{2e}{h} \int \left[\frac{e}{2kT_{\mathrm{H}}} \frac{df_{\mathrm{H}}}{d\xi_{\mathrm{H}}} + \frac{e}{2kT_{\mathrm{C}}} \frac{df_{\mathrm{C}}}{d\xi_{\mathrm{C}}} \right] \tau\left(\varepsilon\right) d\varepsilon \\
&= -\frac{2e^2}{h} \int \left[\frac{1}{2kT_{\mathrm{H}}} \frac{df_{\mathrm{H}}}{d\xi_{\mathrm{H}}} + \frac{1}{2kT_{\mathrm{C}}} \frac{df_{\mathrm{C}}}{d\xi_{\mathrm{C}}} \right] \tau\left(\varepsilon\right) d\varepsilon, \\
&= -\frac{2e^2}{h} \int \left[\frac{1}{2kT_{\mathrm{H}}} \frac{df_{\mathrm{H}}}{d\xi_{\mathrm{H}}} + \frac{1}{2kT_{\mathrm{C}}} \frac{df_{\mathrm{C}}}{d\xi_{\mathrm{C}}} \right] \tau\left(\varepsilon\right) d\varepsilon
\end{aligned}
$$

where Eq. (I.15) has been used to calculate the derivatives $d\xi_{\mathrm{H,C}}/dV$. It is convenient to define a kernel, Θ, of the form

$$
\Theta_n = \frac{1}{2} \left(\frac{k}{e} \right)^n \left[\frac{\xi_{\mathrm{H}}^n}{kT_{\mathrm{H}}} \frac{df_{\mathrm{H}}}{d\xi_{\mathrm{H}}} + \frac{\xi_{\mathrm{C}}^n}{kT_{\mathrm{C}}} \frac{df_{\mathrm{C}}}{d\xi_{\mathrm{C}}} \right], \tag{I.21}
$$

where $n = 0, 1, 2$. Then the conductance can be written

$$
G = -\frac{2e^2}{h} \int \Theta_0 \tau\left(\varepsilon\right) d\varepsilon. \tag{I.22}
$$

The thermodynamic coefficients, G_{th} and K (see Eqs. (I.4)), can be written [31–33],

$$G_{\text{th}} = -\frac{e^2}{h} \int \Theta_1 \tau\left(\varepsilon\right) d\varepsilon \qquad\qquad \text{(I.23a)}$$

$$K = T\frac{e^2}{h} \int \Theta_2 \tau\left(\varepsilon\right) d\varepsilon. \qquad\qquad \text{(I.23b)}$$

Near equilibrium, in the case of small bias and small temperature difference ($\Delta T/T \ll$ 1 and $eV \ll kT$), the temperatures are approximately equal, $T_{\text{H}} \approx T_{\text{C}} = T$ and $V \approx 0$. In this case, $\xi_{\text{H}} = \xi_{\text{C}} = \xi$ and $f_{\text{H}} = f_{\text{C}} = f$. In this regime, Θ_n becomes

$$\lim_{\substack{\Delta T=0 \\ V=0}} \Theta_n = \left(\frac{k}{e}\right)^n \frac{\xi^n}{kT}\frac{df}{d\xi}$$

$$= \left(\frac{k}{e}\right)^n \frac{\xi^n}{kT}\frac{d\varepsilon}{d\xi}\frac{df}{d\varepsilon}$$

$$= \left(\frac{k}{e}\right)^n \frac{\xi^n}{kT}kT\frac{df}{d\varepsilon}$$

$$= \left(\frac{k}{e}\right)^n \xi^n\frac{df}{d\varepsilon}.$$

Inserting this approximate kernel into Eqs. (I.22) and (I.23) and performing Sommerfeld expansions (see Appendix C of Ref. [8]) of the integrals gives

$$G \approx \frac{2e^2}{h}\tau\left(\mu\right) \qquad\qquad \text{(I.24a)}$$

$$G_{\text{th}} \approx \frac{2e^2}{h}eL_0T \left.\frac{d\tau\left(\varepsilon\right)}{d\varepsilon}\right|_{\varepsilon=\mu} \qquad\qquad \text{(I.24b)}$$

$$K \approx -\frac{2e^2}{h}L_0T\tau\left(\mu\right) = -L_0TG, \qquad\qquad \text{(I.24c)}$$

where $L_0 = \pi^2/3 \left(k/e\right)^2$ is the Lorenz number. Using the above and Eq. (I.3), the Seebeck coefficient can be written

$$
\begin{aligned}
S &= -\frac{G_{\text{th}}}{G} \\
&\approx -eL_0 T \frac{\left.\frac{d\tau(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=\mu}}{\tau\left(\mu\right)} \\
&= -eL_0 T \frac{\left.\frac{2e^2}{h}\frac{d\tau(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=\mu}}{\frac{2e^2}{h}\tau\left(\mu\right)} \\
&= -eL_0 T \frac{\left.\frac{dG(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=\mu}}{G\left(\mu\right)} \\
&= -eL_0 T \left.\frac{d}{d\varepsilon}\ln G\left(\varepsilon\right)\right|_{\varepsilon=\mu},
\end{aligned}
\tag{I.25}
$$

which is in perfect agreement with the Mott formula [34]. Next, using Eqs. (I.11b) and (I.24c), the thermal conductance can be written,

$$
\kappa_{\text{el}} = -K - S^2 G T \approx L_0 T G - S^2 G T = \left(L_0 - S^2\right) G T \approx L_0 G T, \tag{I.26}
$$

where the last approximation assumes $S^2 \ll L_0$, which is provided by Eq. (I.25) when $\tau\left(\varepsilon\right)$ is a smoothly varying function.

The relationship between κ_{el} and G given in Eq. (I.26) is known as the Wiedemann-Franz law [8, 35]. However, L_0, as it appears in the Wiedemann-Franz law, is not universal but depends on the dimensionality of the system [36, 37]. Moreover, theory predicts violations of the Wiedemann-Franz law in the Coulomb-blockade regime due

to strong energy-dependent processes [38]. Deviations of this nature are not restricted to the Wiedemann-Franz law. In fact, all the thermoelectric coefficients provided by the Sommerfeld expansions in Eq. (I.24) might be violated if $\tau(\varepsilon)$ is highly modulated.

The ZT Figure of Merit

In the field of thermoelectrics, device performance is almost exclusively measured in terms of the ZT figure of merit. Therefore, to relate the research presented here with other results in the field, it is necessary to introduce this quantity and show its relation to efficiency. The derivation is performed by treating the thermoelectric system as a electronic circuit.

The thermoelectric device in Fig. 1.1a with resistance R (conductance $G = 1/R$) does useful work when it is connected in series with an electric load of resistance R_L. In this case, the current, I, flowing through the system is

$$I = \frac{V_\mathrm{th}}{R + R_\mathrm{L}} = \frac{S(T_\mathrm{H} - T_\mathrm{C})}{R + R_\mathrm{L}}. \tag{I.27}$$

The efficiency of the system is the ratio [39]

$$\eta = \frac{W}{Q_\Sigma} = \frac{I^2 R_\mathrm{L}}{Q + Q_\mathrm{J}}, \tag{I.28}$$

where Q is the thermodynamic heat flux in Eqs. (I.7) and (I.10). Using Eq. (I.11a)

this heat flux can be written

$$Q = \Pi I - \kappa \Delta T = S T_{\mathrm{H}} I + \kappa \left(T_{\mathrm{H}} - T_{\mathrm{C}} \right),$$

where the total thermal conductance $\kappa = \kappa_{\mathrm{el}} + \kappa_{\mathrm{ph}}$ has been used rather than κ_{el} alone in order to include phononic heat losses. Note that T_{H} is used, that is, $\Pi = T_{\mathrm{H}} I$ because the heat flux is leaving the hot reservoir.

Q_{J} in Eq. (I.28) is due to Joule heating and is the heat flux *returned* to the hot reservoir by Joule heating in the thermoelectric material, making the sign of Q_{J} negative. Only half of the Joule heating is delivered to the hot reservoir, while the other half is delivered to the cold reservoir. Thus the magnitude of Q_{J} is only half the total Joule heating making $Q_{\mathrm{J}} = -\dfrac{1}{2} I^2 R$.

Putting these heat fluxes into Eq. (I.28) gives

$$\begin{aligned}
\eta &= \frac{I^2 R_{\mathrm{L}}}{S T_{\mathrm{H}} I + \kappa \left(T_{\mathrm{H}} - T_{\mathrm{C}} \right) - \frac{1}{2} I^2 R} \\[2mm]
&= \frac{R_{\mathrm{L}}/R}{S T_{\mathrm{H}}/Ir + \kappa \left(T_{\mathrm{H}} - T_{\mathrm{C}} \right)/I^2 R - 1/2} \\[2mm]
&= \frac{T_{\mathrm{H}} - T_{\mathrm{C}}}{T_{\mathrm{H}}} \frac{R_{\mathrm{L}}/R}{\frac{S(T_{\mathrm{H}} - T_{\mathrm{C}})}{Ir} + \frac{\kappa}{I^2 R} \frac{(T_{\mathrm{H}} - T_{\mathrm{C}})^2}{T_{\mathrm{H}}} - \frac{1}{2} \frac{T_{\mathrm{H}} - T_{\mathrm{C}}}{T_{\mathrm{H}}}}
\end{aligned}$$

The ratio $\eta_{\mathrm{C}} = \left(T_{\mathrm{H}} - T_{\mathrm{C}} \right)/T_{\mathrm{H}}$ is Carnot efficiency, and η can be expressed relative to

carnot efficiency: η/η_C. Using Eq. (I.27) to eliminate I gives

$$\eta/\eta_\mathrm{C} = \frac{R_\mathrm{L}/R}{\frac{R_\mathrm{L}+R}{R} + \frac{\kappa}{S^2 T_\mathrm{H}} \frac{(R_\mathrm{L}+R)^2}{R} - \frac{1}{2}\frac{T_\mathrm{H}-T_\mathrm{C}}{T_\mathrm{H}}}$$

$$= \frac{R_\mathrm{L}/R}{R_\mathrm{L}/R + 1 + \frac{\kappa r}{S^2 T_\mathrm{H}}\left(R_\mathrm{L}/R + 1\right)^2 - \frac{1}{2}\frac{T_\mathrm{H}-T_\mathrm{C}}{T_\mathrm{H}}}.$$

Defining the ratios $m \equiv R_\mathrm{L}/R$ and $Z \equiv S^2/\kappa r = S^2 G/\kappa$, the (relative) efficiency becomes

$$\eta/\eta_\mathrm{C} = \frac{m}{(m+1) + \frac{(m+1)^2}{Z T_\mathrm{H}} - \frac{1}{2}\frac{T_\mathrm{H}-T_\mathrm{C}}{T_\mathrm{H}}}. \tag{I.29}$$

The maximum efficiency is reached when m is optimized. The best value of m is found by solving the condition

$$0 = \frac{d}{dm}\eta/\eta_\mathrm{C} = \frac{1}{(m+1) + \frac{(m+1)^2}{Z T_\mathrm{H}} - \frac{1}{2}\frac{T_\mathrm{H}-T_\mathrm{C}}{T_\mathrm{H}}}\left(1 - \frac{m\left(1 + 2\frac{m+1}{Z T_\mathrm{H}}\right)}{(m+1) + \frac{(m+1)^2}{Z T_\mathrm{H}} - \frac{1}{2}\frac{T_\mathrm{H}-T_\mathrm{C}}{T_\mathrm{H}}}\right).$$

After some algebra, the optimal value of m, denoted M, is given by

$$M = \sqrt{1 + Z\bar{T}},$$

where $\bar{T} = (T_\mathrm{H} + T_\mathrm{C})/2$. Feeding M into Eq. (I.29) provides the maximum relative efficiency,

$$\eta/\eta_\mathrm{C} = \frac{M}{(M+1) + \frac{(M+1)^2}{Z T_\mathrm{H}} - \frac{1}{2}\frac{T_\mathrm{H}-T_\mathrm{C}}{T_\mathrm{H}}} = \frac{M-1}{M + T_\mathrm{C}/T_\mathrm{H}}. \tag{I.30}$$

It's clear that thermoelectric efficiency limits to Carnot efficiency as $Z\bar{T}$ (and

therefore M) limits to infinity. And η/η_{C} is monotonic in $Z\bar{T}$, as shown in Fig. 1.8. Therefore, maximizing $Z\bar{T}$ is equivalent to maximizing efficiency. Experimentally, it is not convenient to measure the current produced by a thermoelectric under a load resistance R_{L}, which satisfies the condition $R_{\mathrm{L}}/R = M = \sqrt{1 + Z\bar{T}}$, because $Z\bar{T}$ is not known. Since the better the $Z\bar{T}$, the better the efficiency, the standard experimental shortcut it to simply measure $Z\bar{T}$, which is given by,

$$Z\bar{T} = \frac{S^2 G \bar{T}}{\kappa} = \frac{S^2 G \bar{T}}{\kappa_{\mathrm{el}} + \kappa_{\mathrm{ph}}}. \tag{I.31}$$

This parameter is know as the "ZT figure of merit" and is determined by measuring the Seebeck coefficient and conductances separately. Sometimes a bipolar thermal conductance term, κ_{bip}, is also included in κ to account for the thermal diffusion of electron–hole pairs not contributing to net charge transport [40].

The ZT figure of merit has been rooted so deeply into the thermoelectrics field that it has become not just a metric for performance, but a starting point for innovation. A common approach to thermoelectric research is to maximize S while minimizing κ. However, the underlying electronic system couples S, G, and κ_{el} making it difficult, if not impossible, to increase ZT by addressing any one parameter alone [41]. While record ZT values are occasionally reported (see Fig. 1.9), the devices boasting the highest laboratory ZT values are often not reproducible [18], let alone commercially viable.
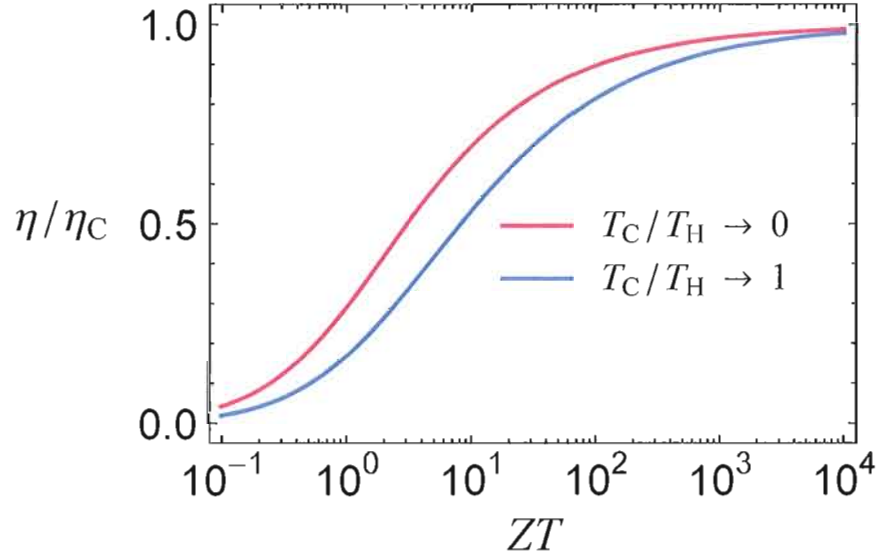
Figure 1.8. A log-linear plot of efficiency relative to Carnot efficiency, η/η_{C}, as a function of the ZT figure of merit (see Eq. (I.30)). η/η_{C} is also influence by the ratio $T_{\mathrm{C}}/T_{\mathrm{H}}$. At every value of ZT, η/η_{C} is bounded above in the limit $T_{\mathrm{C}}/T_{\mathrm{H}} \to 0$ ($T_{\mathrm{C}} \ll T_{\mathrm{H}}$) and bounded below in the limit $T_{\mathrm{C}}/T_{\mathrm{H}} \to 1$ ($T_{\mathrm{C}} \approx T_{\mathrm{H}}$). All experiments operate somewhere in the gray shaded region.

In order to compete with existing thermal-to-electrical energy conversion technologies, thermoelectric efficiency must increase above 25%–30%, which is roughly $\eta/\eta_{\mathrm{C}} = 1/2$ at the intended temperature. Therefore, according to Fig. 1.8, the ZT of thermoelectrics must be increased to above 3 (or even higher if various loss mechanisms are taken into account) before they become viable alternatives to conventional heat engines. Fig. 1.9 highlights the most prominent ZT values reported in the past decade, which are all below 3. Clearly the field of thermoelectrics still has room for improvement. All the recent high-ZT publications focus on either nanowires
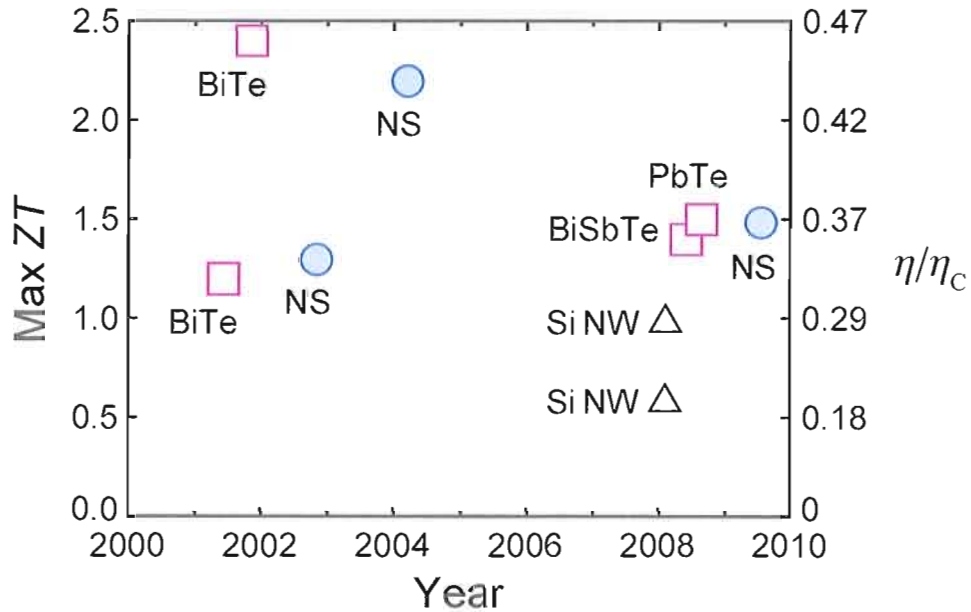
Figure 1.9. High-profile journal articles reporting high ZT values in the past decade. Thermoelectrics need a $ZT > 3$ to compete with conventional energy conversion technologies. The squares are various thin-film Te alloys; the triangles are Si nanowires (NW); and the circles are nanometer-structured (NS) bulk materials, that is, materials engineered with underlaying nanoscale morphology. The scale on the right, η/η_{C}, is the approximate best-case efficiency relative to Carnot efficiency calculated using Eq. (I.30). The ZT values (in chronological order) are from Refs. [16, 22, 23, 41–46].

or bulk materials with underlaying nanometer-scaled features, illustrating the trend toward thermoelectrics based on nanotechnology.

Realizing Carnot Efficiency Experimentally

With the background, theory, and motivation firmly established, the intention of this thesis can be clearly understood. The physical understanding of thermally induced electron transport from the viewpoint of efficiency remains relatively

unexplored. Quantum dots are highly tunable electron energy filters capable of efficient thermal-to-electrical energy conversion. The goal of this research is to use a nanowire-based quantum dot to explore thermoelectric efficiency optimization.

<div align="center">Nanowires</div>

Modern fabrication and synthesis techniques have opened doors into the mesoscopic world offering a myriad of semiconductor-based electronic quantum dot designs. Historically, the most popular architecture is the two-dimensional electron gas (2DEG) in which electrons are confined to a thin plane defined during epitaxial growth. Quantum dots are defined in 2DEGs either electrostatically with external metallic gates or by chemical etching. The next most common quantum dot system is the nanowire, which is gaining popularity as nanowire fabrication techniques constantly emerge and evolve. Quantum dots can be defined in nanowires electrostatically using finger gates on top of an insulating dielectric similar to what is done in 2DEG systems. Alternatively, a quantum dot can be defined in a nanowire by embedding two tunnel barriers during fabrication. This fabrication entails sandwiching two thin semiconductor slices (< 10 nm) into a host semiconductor material that has a lower band gap (see Fig. 1.10a). If small enough, the region between the two barriers becomes a quantum dot, and the entire package is referred to as a double-barrier quantum dot.

For thermoelectric applications, the heterostructure quantum dot is far superior to a quantum dot defined in either a 2DEG or by finger gates in a nanowire. The

main reason being scalability; heterostructure nanowires can be grown in parallel in arbitrary numbers. A "forest" of heterostructure nanowires protruding from a growth substrate could be capped and operated as a thermoelectric. Geometry forbids the parallelization of 2DEG quantum dots, and addressing arrays of homogeneous nanowires with finger gates is impractical, if not impossible. Fig. 1.10 depicts a possible design of a thermoelectric device based on heterostructure nanowires.
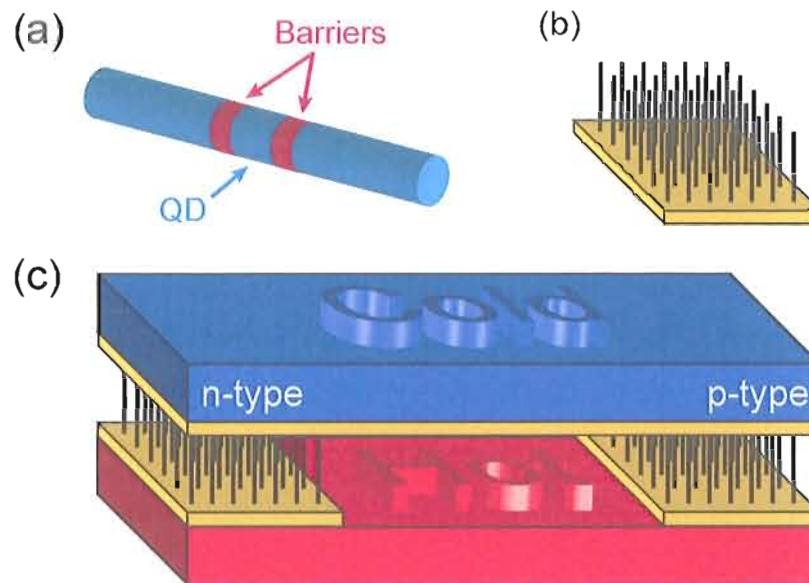
Figure 1.10. (a) Two tunnel barriers embedded in a host semiconductor nanowire. The result is a heterostructure nanowire defining a double-barrier quantum dot. (b) An array or "forest" of heterostructure nanowires standing on their growth wafer. (c) A schematic of a prototype thermoelectric device based on heterostructure nanowires arrays. The left array is made using n-type nanowires and the right array with p-type. In this way, the nanowire-based device operates in the same way as the conventional bulk thermoelectric devices shown in Fig. 1.3.

Outlook

Before building a prototype device like the one shown in Fig. 1.10c, the efficiency

of single heterostructure can be studied in detail. This research is crucial to the understanding of nanoscale thermoelectrics and will guide future research and design. This dissertation demonstrates that heterostructure InAs/InP nanowires operate as electron energy filters resulting in a nanoscale thermoelectric device with an electronic efficiency up to 95% Carnot efficiency manifested experimentally in (electronic) ZT values greater than 1000.

CHAPTER II

FABRICATION AND EXPERIMENTAL METHODS

Nanowire Growth and Properties

The InAs/InP heterostructure nanowires used in this research were grown epitaxially by collaborators at Lund University. The word epitaxy was coined to describe growth in which species are deposited onto a surface and allowed to diffuse to a preferred (energy reducing) crystallization site. In this way, epitaxial growth is a bottom-up approach in which the growth species are like building blocks, as opposed to top-down approaches, such as etching or milling, where large objects are made smaller.

The nanowires from Lund are grown using chemical beam epitaxy (CBE) [47], which is similiar to molecular beam epitaxy (MBE) [48]. In both techniques, growth species are injected into a vacuum chamber where they impinge on the growth substrate and crystallize. In CBE, the group-III and -V source materials are metalorganics, which get injected into the growth chamber similar to an MBE technique. The respective precursors Lund uses for the As, P, and In growth species are the metalorganics tertiary-butyl-arsine (TBA), tertiary-butyl-phosphine (TBP), and tri-methyl-indium (TMI). During injection, the group-V metalorganics

are decomposed or thermally cracked with a heated filament. As a result, As_2, As_4, AsH_3, P_2, H_2, and various hydrocarbons enter the growth chamber, while other material remains at the filament. Uncracked TMI is injected into the growth chamber, that is, as indium and three carbon groups, $In(CH_3)_3$. The $In(CH_3)_3$ can diffuse around on the growth substrate, and while doing so, it decomposes by losing one carbon group at a time, transitioning through $In(CH_3)_2$ and $In(CH_3)_1$ until it is only In.

InAs/InP Heterostructure Nanowire Growth

Nanowire growth was first described in the 1960s [49], and today the resulting vapor-liquid-solid (VLS) growth model remains the main model for understanding nanowire growth, though other similar mechanisms have been suggested [50, 51]. Before growth, small ($<$ 100 nm) metallic particles are deposited onto a growth substrate. The essence of this model as it stands today is that the metallic particles seed nanowire growth by providing a preferred site for crystallization at the interface between the particle and the growing nanowire. The various growth species are collected from the surrounding resources. While the $In(CH_3)_3$ species is free to diffuse and find its way to the growth site, the As_x and the P_x have a negligible diffusion lengths and must impinge directly onto the seed particle in order to contribute to growth. The crystal forms atomic layer by atomic layer, and in doing so, lifts the seed particle off the growth substrate. The resulting nanowire grows up from the substrate underneath each seed particle until the input of source material ceases.
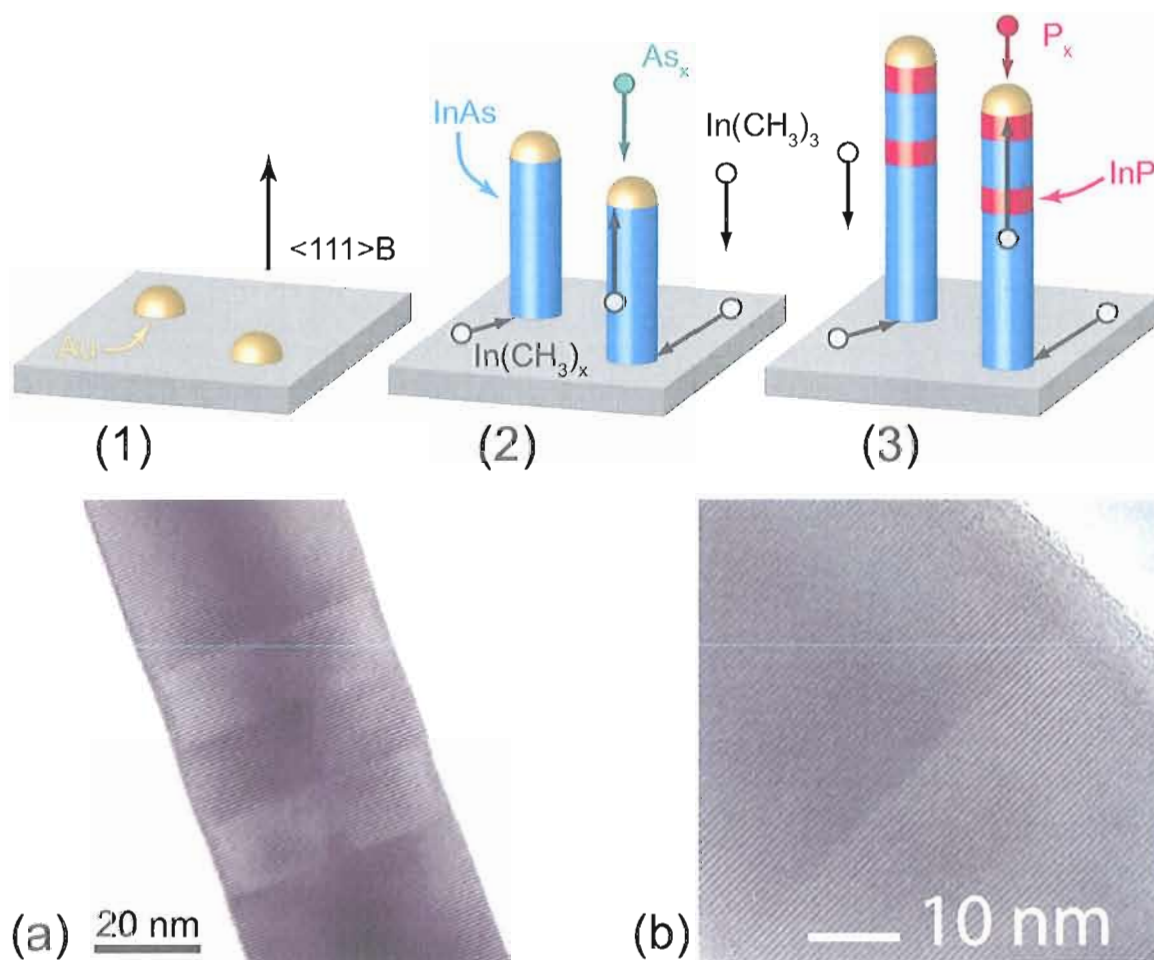
Figure 2.1. Heterostructure InAs/InP nanowires are grown by **(1)** depositing Au seed particles (approx 50 nm in diameter) onto an InAs growth wafer. The InAs crystal grows in the $\langle 111 \rangle$B direction normal to the surface. **(2)** As_x and $In(CH_3)_3$ sources are introduced into the growth chamber. Decomposing $In(CH_3)_x$ diffuses around eventually delivering In to the Au seed particle. Only via direct impingement can As_x find the seed particle. In and As crystallize below the Au seed particle and form an InAs wurtzite crystal. **(3)** Alternating between As_x and P_x during the growth process creates a heterostructure InAs/InP nanowire. **(a)** A TEM image of an InAs/InP heterostructure nanowire. **(b)** A high-resolution TEM image demonstrating a very sharp InAs/InP interface. (Images courtesy of Dr. Jakob Wagner, Lund University.)

Homogenous nanowires result from using a constant supply of a particular growth species. Heterostructure nanowires can be achieved by alternating growth species. The nanowires exhibit n-type semiconductor behavior, and it is believed that the hydrocarbons present during growth act as unintentional dopants. Fig. 2.1 depicts CBE nanowire growth as explained by the metal-particle-seeded growth model. More details regarding the CBE growth of nanowires at Lund can be found in Refs. [52, 53].

The InAs/InP heterostructure used in this research were grown on a $\langle 111 \rangle$B InAs substrate using gold seed particles deposited using an aerosol technique or electron beam lithography (EBL). The resulting nanowires have a hexagonal cross-section and a wurtzite crystal structure. The truly remarkable property of nanowires is their extreme geometry which reduces the strain between dissimilar materials and prevents lattice defects, dislocations, voids, etc. As a result, the interface between the InAs and InP portions of an InAs/InP heterostructure nanowire is nearly atomically sharp, as shown in Fig. 2.1b.

Sample Design and Processing

Sample devices were processed by removing InAs/InP nanowires (typically 50 nm diameter and roughly 1 μm long) from their growth substrate and deposited onto an n-doped Si wafer capped with a 100 nm electrically insulating SiO_x layer (see Fig. 2.2). The nanowires are then located using scattered light provided by a dark-field setup

in an optical microscope. Once the relative positions of the nanowire are known, the sample substrate and nanowires are covered with PMMA resist.

At this point the EBL pattern is designed manually to each wire, and then the sample is exposed in the EBL writer. After EBL exposure, the PMMA is developed and then the sample is given an ozone etch to remove resist residues before metallization. InAs is particularly convenient for performing electrical measurements because the Fermi energies of metals and InAs are similar enough to offer an Ohmic electrical junction between the two materials. Nickel is used as the base layer of the electric contacts because it adheres well to the sample substrate and provides a good adhesion layer for gold. To decrease lead resistance, a final metallic layer of gold is deposited on top of the nickel. However, a native oxide forms on the surface of InAs nanowires apon exposure to atmospheric gases, which is unavoidable during the fabrication process. In order to remove this native oxide, the sample is passivated in a solution of NH_4S_x and water. Immediately following the passivation process, Ni/Au electrical leads are contacted to the InAs ends of the nanowire by first evaporating 25 nm of nickel followed by 90 nm of gold. Finally, the PMMA resist is removed, and the sample is cleaned and dried. A contacted nanowire is shown in Fig. 2.13. For further growth and fabrication details, see Ref. [52–55].

Considerations for Thermoelectric-specific Devices

Nanowire samples intended for thermoelectric measurements must have two abilities: the ability 1) to apply a temperature gradient along the longitudinal axis
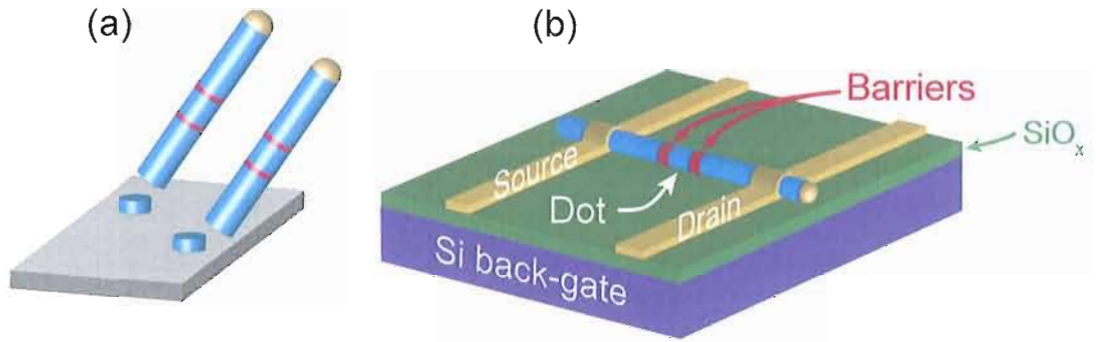
Figure 2.2. Fabrication of a nanowire-based quantum-dot device. (a) Heterostructure nanowires are removed from their growth substrate and placed on (b) a sample wafer that includes an electrically insulating SiO_x capping layer on top of n-doped silicon (not to scale). The Au/Ni source and drain contacts are defined by EBL and subsequent metallization. The n-doped silicon wafer serves as a global back-gate. Altogether the double-barrier quantum-dot device operates as a single-electron transistor.

of the nanowire and 2) to measure electrical voltages and currents in the nanowire. More specifically, it is the electron gas in either the source or drain contact which must be heated relative to the opposite contact. The approach presented here is to heat the electron gas directly by applying a heating current through the metallic source contact (see Fig. 2.13). This guarantees that heated electrons are injected into the nanowire, regardless of electron-phonon coupling. From a fabrication point of view, the extra EBL design and alignment required to achieve this heating geometry is minimal. However, this approach requires a considerable amount of attention in the electronic setup because the heating circuitry and measurement circuitry are one and the same. Details of this method are explained later in Section II.4.4.

As thermoelectric experiments become more demanding, future devices might

require more care regarding heat loss mechanisms. For example, improved devices might suspend the nanowire to reduce heat flow into the substrate and use superconducting leads to stop heat flow through the electrical contacts.

Theoretical Considerations

This section details the background theory of quantum dots, specifically Coulomb blockade as explained by the constant interaction model. Included is a discussion regarding the quantum transmission function of a double-barrier quantum dot. Details are provided about the theoretical modelling of differential conductance data and an experimental approach for measuring the energetic width of transmission peaks.

Quantum Dots, Coulomb Blockade, and the Constant Interaction Model

If electrons are confined to a sufficiently small region (see Fig. 2.3), then their behavior is dominated by two effects. First, quantum confinement in all three dimensions can be achieved by reducing the dimensions of the region until they are comparable to the Fermi wavelength of the electrons. For this reason, quantum confinement is most easily achieved in semiconductor systems where typically small effective mass (relative to metallic systems) increases the Fermi wavelength and therefore loosens the dimensionality requirements from nanometers to tens of nanometers. Once the electrons are confined, they become localized to within a small "dot" in real space, hence the name "quantum dot". Due to the resulting discrete energy spectrum, quantum dots behave in many ways as artificial atoms. This analogy is so pervasive that much of the atomic physics vernacular has become common in the quantum dot field. For example, even though confinement in quantum dots is

not produced by a central force and, therefore, the electrons do not have spherical-harmonic wavefunctions, it is common to speak of quantum dot "orbitals".
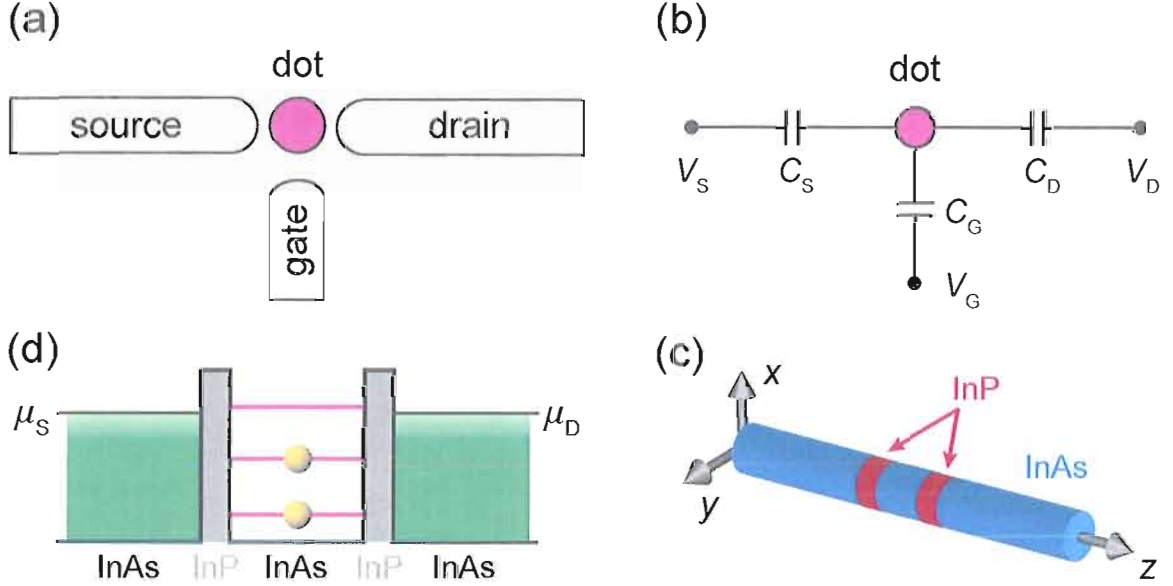
Figure 2.3. (a) A schematic of a quantum dot connected to source and drain contacts via tunnel junctions, which allow electrons to tunnel between the dot and either the source or the drain. The gate is capacitively coupled to the dot and offers continuous external control of the dot's energy. (b) A capacitive circuit equivalent to the quantum dot system used to derive Eq. (II.1). The capacitances between the dot and leads are $C_{S,D}$ and between the dot and gate is C_G. (c) Two InP barriers embedded in an InAs nanowire create a double-barrier quantum dot. A nanowire tens of nanometers in diameter quantizes electron momentum in the x- and y- directions while the barriers quantize electron momentum in the z-direction. (d) The resulting energy landscape of the double-barrier quantum dot. The InP barriers have a larger band gap than the host InAs. In addition to the number of electrons in the dot, the electrochemical potential of the dot is sensitive to the charge in the source and drain electron reservoirs. This charge is quantified by the electrochemical potentials, $\mu_{S,D} = -eV_{S,D}$, appearing in Eq. (II.1).

Second, Coulomb repulsion, although safely ignored in most systems, is a classical effect caused by the Coulomb forces among interacting electrons and can be

comparable to or even larger than quantum effects and therefore cannot be neglected. For each electron added to the dot, there is an associated energy cost, and the tunneling of electrons from the source or drain reservoirs into the dot is prevented if these energy requirements are not fulfilled; this phenomenon is called Coulomb blockade. For quantum dots on the scale of one micron, the quantum energy spacing is very small compared to the Coulombic energy, and the electron transport properties of such systems are entirely dominated by Coulomb blockade.

The first comprehensive overview of Coulomb blockade [56] remains a very worthwhile resource. For examples of the vast variety of quantum dot systems, fabrication techniques, experimental measurements, and an updated theoretical treatment of quantum dots, including electron-spin effects, see Ref. [57]. Here it is enough to review briefly the physics relevant to nanowire quantum dots and thermoelectrics. The most widely used theoretical treatment of quantum dots is the constant interaction model, which explains the phenomenon of Coulomb blockade using an energetic approach based on two assumptions.

First, the total of all the Coulomb interactions provides an effective total capacitance, parametrized by C_Σ, which is constant as a function of electron number. The Coulomb interactions to consider are those among the electrons in the dot in addition to those between the electrons in the dot and the external environment. The charge in the dot is $-e\left(N - N_0\right)$, where N is the number of electrons, and N_0 accounts for the net static charge trapped in or on the semiconductor material

surrounding the quantum dot. The external charges include those in the source, drain, and gate. The respective charges in these reservoirs are quantified by $C_G V_G$, $C_S V_S$, and $C_D V_D$, where C_i are the capacitances between the dot and each of the three reservoirs, and the V_i are the voltages applied to the leads and gate (see Fig. 2.3b). The total charge is $\sum_i Q_i = -e(N - N_0) + C_G V_G + C_S V_S + C_D V_D$, and therefore, the total capacitive energy of the dot is $\left(\sum_i Q_i\right)^2 / 2C_\Sigma$, where $C_\Sigma \equiv C_G + C_S + C_D$ is the total capacitance. Therefore, the energy in the dot changes when N changes, but C_Σ remains *constant*, which is why this model is called the constant interaction model.

The second assumption of this model is that the quantum energy spectrum is unaffected by N. That is, the Nth electron placed in the dot assumes the Nth quantum orbital, and all orbitals remain static as electrons are added. Under these assumptions, the total energy $U(N)$ of a dot with N electrons is simply the sum of the capacitive energy and the quantum energies,

$$
\begin{aligned}
U(N) &= \frac{\left(\sum_i Q_i\right)^2}{2C_\Sigma} + \sum_{n=1}^{N} E_n \\
&= \frac{1}{2C_\Sigma}\left[-e(N - N_0) + C_G V_G + C_S V_S + C_D V_D\right]^2 + \sum_{n=1}^{N} E_n,
\end{aligned}
$$

where E_n is the energy of the nth quantum energy level.

By definition, the electrochemical potential of the dot, $\mu(N)$, is the change in

energy generated by adding one electron, that is,

$$\mu(N) \equiv U(N) - U(N-1)$$

$$= \frac{e^2}{2C_\Sigma} \left[(N - N_0)^2 - ((N-1) - N_0)^2 \right] - \frac{e}{C_\Sigma} (C_G V_G + C_S V_S + C_D V_D) + E_N$$

$$= \frac{e^2}{2C_\Sigma} \left[2(N - N_0) - 1 \right] - \frac{e}{C_\Sigma} (C_G V_G + C_S V_S + C_D V_D) + E_N$$

$$= E_C \left(N - N_0 - \frac{1}{2} \right) - \frac{E_C}{e} (C_G V_G + C_S V_S + C_D V_D) + E_N, \qquad \text{(II.1)}$$

where $E_C \equiv e^2/C_\Sigma$ is the charging energy. The first two terms arise from classical Coulomb interaction whereas the last term is a quantum effect. The first and last terms are quantized by N. The second term is made continuous by the ability to control the externally applied voltages continuously, and so $\mu(N)$ can be tuned continuously. Nonetheless, experimental measurements will reflect charge quantization as single electrons hop on and off the dot thereby changing N and $\mu(N)$ discontinuously.

The addition energy, E_{add}, is the energy difference between two consecutive electrochemical potentials,

$$E_{\text{add}}(N) \equiv \mu(N) - \mu(N-1) = E_C + \Delta E(N), \qquad \text{(II.2)}$$

where $\Delta E(N) = E_N - E_{N-1}$ is the difference between the Nth and $(N-1)$th quantum energies. $\Delta E(N) = 0$ when the Nth and $(N-1)$th levels are degenerate

(either spin degenerate or geometrically degenerate). In this case, $E_{\text{add}} = E_{\text{C}}$ and the definition of the charging energy becomes clear: it is the energy difference between successive electrochemical potential levels due to charging effects alone. When a quantum dot is cooled so that $E_{\text{add}} < kT$, each $\mu(N)$ energy resonance can be resolved by differential conductance measurements (see, for example, Fig. 2.8).

A complete quantitative description of $\mu(N)$ requires calculating the quantum confinement along the axis of the double-barrier structure as well as the lateral confinement provided by the nanowires itself. The InAs nanowires used in these experiments have a hexagonal cross-section, which is a challenging boundary condition for wavefunction calculations. However, it is reasonable to approximate the nanowires as cylindrical and solve the corresponding 3D Schrödinger equation. Once the quantum eigenenergies are found, the charging energy can be estimated based on the nanowire geometry [58, 59], and the total energy spacing (quantum plus classical) can be calculated [60]. Fig. 2.4 shows the results of such a calculation performed numerically using MATLAB.

One last consideration is to determine the relationship between the gate voltage, V_{G}, and $\mu(N)$. Conveniently, Eq. (II.1) reveals the (linear) response of $\mu(N)$ to V_{G},

$$\frac{d\mu(N)}{dV_{\text{G}}} = -\frac{E_{\text{C}}}{e}C_{\text{G}} = -\frac{C_{\text{G}}}{C_{\Sigma}}e = -\alpha e,$$
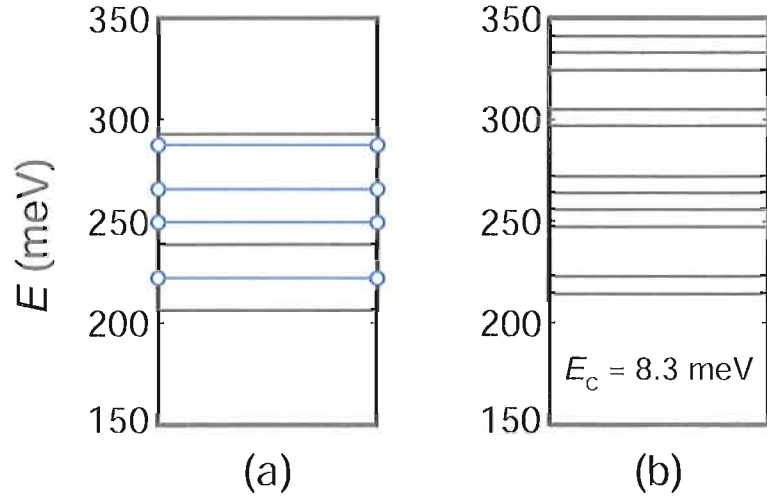
Figure 2.4. **(a)** The quantum energy levels of a double-barrier structure embedded in a 55 nm diameter nanowire with 10 nm barriers and a 7 nm barrier-separation distance. This is the result of a full 3D calculation. The black, single-lined energy levels are two-fold spin degenerate, while the blue, dumbbell-shaped energy levels are four-fold degenerate, owing to spin degeneracy and geometric degeneracy. **(b)** The resulting net energy levels after the classic charging energy, $E_C = 8.3$ meV, has been included.

where

$$\alpha \equiv C_G/C_\Sigma \tag{II.3}$$

is the dimensionless "level arm" of the capacitively coupled gate voltage. Therefore, V_G can be converted to $\mu(N)$ using α. This conversion is very convenient because often experimental measurements made as a function of V_G but must be expressed in energy in order to compare with theory or with other energy scales of the system. The procedure for determining α experimentally is presented in Section II.4.3.

Quantum Transport Through a Double-Barrier Structure

Electron transport through a double-barrier semiconductor structure is analogous to the propagation of light through a Fabry-Pérot interferometer. In each apparatus, electrons or photons with particular energies have near perfect transmission due to the wave nature of both systems. These particular energies are the resonant energies of the systems. This leads to a series of peaks in transmission as a function of energy.

In a Fabry-Pérot interferometer, the energetic spacing between resonant energies is periodic and is influenced only by the separation between the interferometer mirrors. Predicting the resonant energies of a semiconductor double-barrier structure is more difficult for many reasons. Solving the governing equation, the Schrödinger equation, and the associated boundary conditions requires finding solutions to transcendental equations.   Moreover, the physical parameters, such as the barrier widths and separation distance, usually cannot be measured with an accuracy of more than 1–2 nm, which is less than required for quantitative predictions of resonance positions (see Fig. 2.5). Fortunately, the position of energy resonances are easily measured by gate-voltage spectroscopy, where the back-gate scans the resonances of the double-barrier structure through the source and drain electrochemical potentials.   Each resonance is observed by applying a small ac bias and measuring the subsequent peaks in differential conductance as a function of gate voltage (see Fig. 2.8). A more difficult task is to determine the energetic width of each energy resonance, which

is distinct from the width of resonances seen in current or differential conductance measurements.

The Fabry-Pérot interferometer also produces peaks of a particular width, which is related to the so-called finesse of the interferometer. The quality of the mirrors (described by their reflection coefficient) and the alignment of the setup (the degree to which the mirrors are parallel and the angle of incidence of the photon) dictates the sharpness of the transmission resonances. Similarly, double-barrier structures produce resonances of finite width owing to their quality. The alignment of incoming electrons and the physical alignment of energetic barriers is uncontrollable in a double-barrier device. However, the reflection coefficient of a double-barrier structure is very sensitive to the geometry, band gap difference, and effective masses of the semiconductor materials. Therefore, the experimentalist has much control over the resulting resonance width of a double-barrier structure during device fabrication. Indeed, devices with electrostatically defined, tunable barriers can provide tunable resonance widths, but with undesired effects on the positioning and spacing of the resonances.

It would be inappropriate in this text to derive the quantum mechanical wave treatment of a double-barrier structure, because it has already been done elsewhere and more completely than necessary here. The process involves calculating the coefficients of all the wavefunctions in their various regions of the double-barrier structure by considering the boundary conditions. The most important outcome of

such an exercise is the quantum transmission function. From Ref. [30], the quantum

transmission probability of a symmetric double-barrier structure is,

$$Tr\left(\varepsilon, b, d\right) = \frac{T\left(\varepsilon, b\right)^2}{T\left(\varepsilon, b\right)^2 + 4R\left(\varepsilon, b\right)\cos^2\left(kd - \theta\right)} \tag{II.4}$$

where ε is electron energy, b is the width of the barriers, and d is the distance

between the barriers. $T\left(\varepsilon, b\right)$ and $R\left(\varepsilon, b\right)$ are the transmission and reflection

coefficients defined by the ratio of the incoming to outgoing and incoming to returning

wavefunctions, respectively, and are a function of barrier thickness,

$$T\left(\varepsilon, b\right) = \left(1 + \left(\frac{k^2 + \gamma^2}{2k\gamma}\right)^2 \sinh^2\left(\gamma b\right)\right)^{-1} \tag{II.5}$$

$$R\left(\varepsilon, b\right) = 1 - T\left(\varepsilon, b\right). \tag{II.6}$$

Here k and γ are the wavenumbers of the electron outside and inside of the barriers,

respectively,

$$k^2 = \frac{2m^*}{\hbar^2}\varepsilon$$

$$\gamma^2 = \frac{2m^*}{\hbar^2}\left(E_0 - \varepsilon\right),$$

where E_0 is the barrier height ($E_0 = 740$ meV for InAs/InP heterostructure

nanowires) and m^* is the effective mass. This theory assumes the barriers and

host material have equal effective masses, which only approximates InAs/InP heterostructure nanowires because $m^* = 0.023m$ for InAs and $m^* = 0.08m$ for InP.

The most interesting term in Eq. II.4, $\cos^2(kd - \theta)$, dictates at which energies the electron achieves constructive interference and therefore an energy resonance. Its phase factor is

$$\theta = -\arctan\left[\frac{k^2 - \gamma^2}{2k\gamma}\tanh(\gamma b)\right] + kb.$$

$Tr(\varepsilon, b, d) = 1$ at resonance when $\cos^2(kd - \theta) = 0$. Near such a resonance, a Taylor expansion of the $\cos^2(kd - \theta)$ term in Eq. II.4 provides a transmission function of the form [30],

$$Tr(\varepsilon, b, d) \approx \tau(\varepsilon) = \frac{(\Gamma_n/2)^2}{(\Gamma_n/2)^2 + (\varepsilon - \varepsilon_n)^2}, \tag{II.7}$$

where Γ_n is the full width at half maximum (FWHM) of this Lorentzian-shaped transmission peak centered at resonant energy ε_n. Note this equation is only strictly valid for ε near ε_n. It is particularly convenient to include a unitless amplitude in $\tau(\varepsilon)$ when doing numerical calculations to account for other loss mechanisms in the system. It will be assumed that all Γ_n are approximately equal so that $\Gamma_n = \Gamma$. A theoretical estimation of Γ that ignores the classical Coulomb interaction among electrons is given by [30],

$$\Gamma = \sqrt{\frac{2\hbar^2\varepsilon_0}{m^*d^2}\frac{T^2}{R}}, \tag{II.8}$$

where T and R are defined by Eqs. (II.5) and (II.6), respectively. Γ becomes a function of b and d through T and R, as shown in Fig. 2.5. The charging energy is not reflected
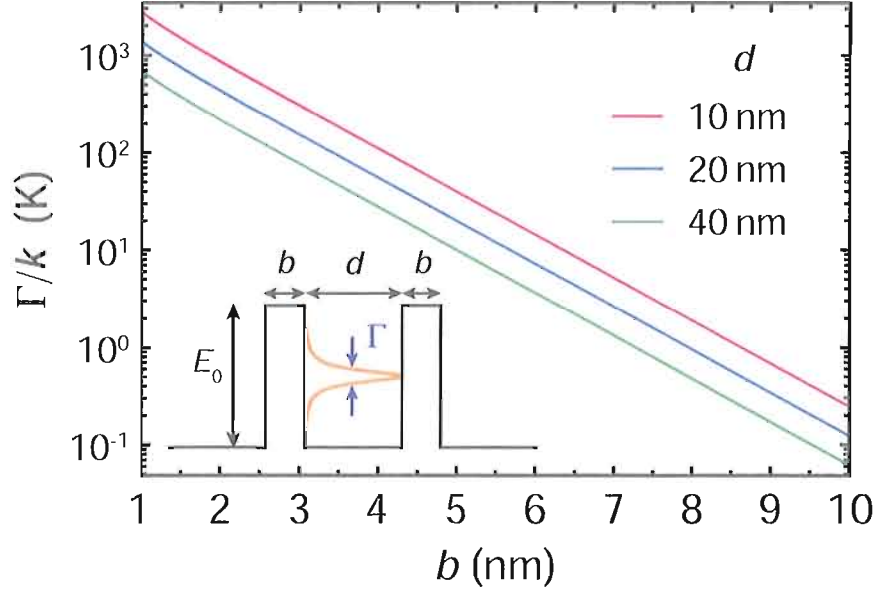
Figure 2.5. A linear-log plot of the Lorentzian FWHM, Γ, calculated numerically using Eq. (II.8). Γ/k is plotted as a function of the barrier width, b, for three different barrier separation values, $d = 10$, 20, and 40 nm (from top to bottom). Γ can be tuned by orders of magnitude by changing b by only a few nanometers. Unfortunately, if b is measured with a $\sim$2 nm uncertainty, the uncertainty in Γ will be an order of magnitude. If b is sufficiently large, $\Gamma \ll kT$ at cryogenic temperatures ($T < 10$ K). *Inset*: A labelled illustration of a double-barrier structure with an energetic barrier height E_0 and the associated Lorentzian-shaped quantum transmission function of width Γ. $E_0 = 740$ meV for an InAs/InP heterostructure.

in Eq. (II.8), and therefore the Γ predicted by this equation is too large owing to the

Coulomb "squeezing" of resonances into a smaller region of energy space. Therefore,

Eq. (II.8) must be considered an upper limit on Γ, and in practice, a quantitative

value of Γ can only be found using experimental data. Nonetheless, Fig. 2.5 illustrates

that Γ is very sensitive to the barrier width and barrier separation making it possible

to engineer double-barrier quantum-dot devices with energy resonances which are

very sharp, that is, $\Gamma \ll kT$.

Fitting Differential Conductance Data

The strategy is to derive an analytical expression for the differential conductance, $G = dI/dV$, which can be used to fit experimental data. In such a fit, Γ is assumed to be much smaller than kT so that the transmission function can be approximated as a Dirac-delta function. Fortunately, in many applications the precise value of Γ is unimportant, only its value in comparison to kT is important. If the differential conductance data can be satisfactorily duplicated by the approximate equation, then it is fair to conclude that the differential conductance peak results from a thermally-broadened Dirac-delta resonance.

Because of the Lorentzian $\tau(\varepsilon)$ (Eq. (II.7)) appearing in the integrand of the Landauer formula (Eq. (I.20)), the expression for G has no known analytical solution. Therefore, as recourse, an approximation must be made. In the limit of $\Gamma \ll kT$, the transmission function $\tau(\varepsilon)$ takes the form of a Dirac-delta function, $\tau(\varepsilon) = \gamma_0 \delta(\varepsilon - \varepsilon_0)$, where γ_0 is its energetic "strength". Under this assumption, G reduces to

$$G = -\frac{1}{kT}\frac{e^2}{h}\gamma_0 \left(\frac{\partial f_{\mathrm{S}}}{\partial \xi_{\mathrm{S}}} + \frac{\partial f_{\mathrm{D}}}{\partial \xi_{\mathrm{D}}} \right),$$

where $\partial f_{\mathrm{S,D}}/\partial \xi_{\mathrm{S,D}}$ are evaluated at $\varepsilon = \varepsilon_0$. Because γ_0 is unknown, the overall prefactor ultimately becomes a fit parameter, A, the amplitude of the conductance

peak. Taylor expanding the above expression to second-order in V gives

$$G = A \operatorname{sech}^2 \left(\frac{e\alpha(V_\mathrm{G} - V_0)}{2kT} \right) \left[1 + \left(\frac{eV}{4kT} \right)^2 \left(2 - 3 \operatorname{sech}^2 \left(\frac{e\alpha(V_\mathrm{G} - V_0)}{2kT} \right) \right) \right] , \quad (\mathrm{II.9})$$

where $V_0 = \varepsilon_0/\alpha e$ is the gate voltage at which the differential conductance peak is centered. A complete derivation of this equation is provided in Appendix A. When $V = 0$, the result limits to the zeroth-order Taylor expansion, which has been derived elsewhere, see for example Ref. [61]. The importance of including the second-order term is to determine the extent to which V broadens differential conductance peaks. Usually this broadening is quite small. However, this extra effort improves the restrictions which can be placed on Γ, which is after-all the main purpose of fitting differential conductance peaks.

Measurement of the Quantum Transmission Width

Comparing the measured performance of quantum-dot thermoelectric devices to theory requires knowing the actual value of Γ because thermoelectric transport processes are very energy sensitive. It is possible to estimate Γ using experimental differential conductance data in conjunction with numerical integration techniques. However, such fits have only been able to place upper-limts on Γ.

The essence is to numerically fit measured data with the isothermal version of

Eq. (I.22) given by,

$$G = \frac{dI}{dV} = -A \int \left(\frac{df_S}{d\xi_S} + \frac{df_D}{d\xi_D} \right) \tau\left(\varepsilon\right) d\varepsilon,$$

where the overall amplitude A is a fit parameter, the arguments in the Fermi-Dirac distributions are $\xi_{S,D} = \left(\varepsilon - \alpha e V_G \pm e V/2\right)/kT$, and $\tau\left(\varepsilon\right)$ is given in Eq. (II.7). The fitting algorithm searches the three-dimensional parameter space of A, Γ, and ε_0 to find the best fit of the data. This is done by minimizing the penalty function,

$$\chi = \sum_i \left(X_i - Y_i\right)^2,$$

where X_i is the ith measured data point and Y_i is the ith value predicted by the fit. ε_0 is a very stiff parameter—that is, χ responds quickly to changes in ε_0—and its optimal value is found very easily. A is also fairly stiff, but is forced to change as Γ changes. Unfortunately, Γ is not very stiff, in part, because $\Gamma \ll kT$. This ultimately limits the accuracy to which Γ can be determined.

The Lorentzian tails decrease very slowly $(1/\varepsilon)$, but thankfully, the Fermi-Dirac distributions fall off exponentially $\left(e^{-\varepsilon/kT}\right)$. The energy range of the numerical integration should be chosen based on the temperature, say $20kT$ to be sure to catch the whole Fermi window. Sufficient partitioning of the energy range into segments $\delta\varepsilon$ in width must be done to insure a reasonable sampling of the Lorentzian resonance. Partitioning on the scale of $\delta\varepsilon = \Gamma/10$ is sufficient. Suppose that $\Gamma = kT/10$, then

the partitioning is $\delta\varepsilon = \Gamma/10 = kT/100$. The number of terms, N, in the numerical integration sum is roughly,

$$N \simeq \frac{\text{energy range}}{\text{division size}} = \frac{20kT}{\delta\varepsilon} = \frac{20kT}{kT/100} = 2,000.$$

Note that this is the requirement for a single data point for a given gate voltage. Fitting a single differential conductance peak could requires calculating 100 such data points. Now consider doing this calculation for several values of A, Γ, and ε_0 while searching the 3D parameter space. This simple fitting algorithm quickly becomes very time consuming.

Time issues can be overcome with a faster computer and greater patience. Ultimately these this fits are limited by the fact that $\Gamma \ll kT$ at the experimental temperatures considered here. In this regime, thermal broadening dominates the outcome of G, and the Γ dependence is too weak to quantify accurately the value of Γ. The numerical fits place constraints on Γ which are no better than what the analytical fit using Eq. (II.9) provides. On the other hand, the assumption that $\Gamma = kT/10$ could be wrong; Γ could be smaller. Therefore, decreasing $\delta\varepsilon$ by another order of magnitude (or more) might help resolve features in $\tau(\varepsilon)$ which otherwise go unnoticed. Unfortunately, this would require an order of magnitude increase in computation time.

Experimental Measurements at Low Temperature

All experiments were performed in ^{3}He and ^{4}He cryogenic refrigerators at temperatures between 240 mK to 10 K, where the temperatures reported here are electron temperatures, rather than the lattice or cryostat temperatures.

Measurement Schemes under Isothermal Conditions

Electronic characterization is be done by performing rudimentary current-voltage (IV) measurements in the absence of an applied temperature gradient. Differential conductance measurements are best performed using an ac bias voltage so that lock-in amplification techniques can be employed to improve the signal-to-noise ratio. Lock-in measurements of electrical current are differential in nature because the lock-in senses the change in current created by a change in applied voltage. The applied ac voltage, δV, is made very small—ideally much smaller than the thermal energy kT—resulting in a very small change in current, δI. In this way, the ratio $\delta I / \delta V$ is a very good approximation of the true derivative dI/dV because δV is very small. A single Standford Research Systems (SRS) Model SRS830 lock-in amplifier can measure differential conductance at zero dc bias voltage. The measurement of Coulomb-blockade diamonds requires dc voltage biasing so that $\delta I / \delta V$ can be measured as a function of bias voltage. For this purpose, a dc voltage source (a Yokogawa Model 7651) was used in conjunction with an SRS830 lock-in amplifier, and the ac and dc sources were combined using a voltage adder. The most simple voltage adder

is a parallel combination of two voltage dividers joined at their ground resistor, as shown in Fig. 2.6. So long as the resistance to ground, R_G, is much less than the load resistance, R_L, the division factor of the voltage divider is, $V_{out}/V_{in} = R_G/(R_{in} + R_G)$, where R_{in} is the input resistance. The combined ac + dc output of the voltage divider is applied across the nanowire during differential conductance measurements.
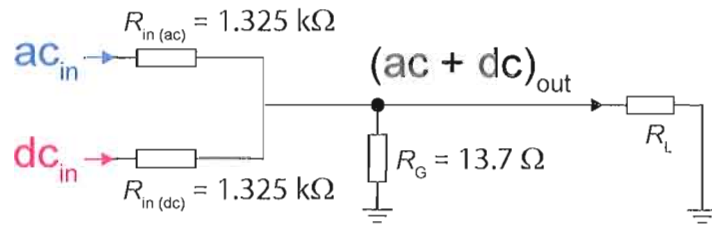
Figure 2.6. An ac + dc voltage adder for floating an ac lock-in voltage on top of the dc bias voltage. Although the circuit is symmetric for the ac and dc portions of the adder, the dc voltage source (Yokogawa Model 7651) can provide the ac current an alternate path to ground, which is appreciable if R_{dc} is comparable to R_G. The resulting division of this divider is $R_G/(R_G + R_{in}) = 0.01$ for both the ac and dc components. The ac source frequency was primarily tuned to 42 Hz, but the performance of the circuit is not acutely sensitive to frequency.

The ac + dc current flowing through the nanowire is measured by attaching an SRS Model SR570 low-noise current preamplifier to the drain contact. This preamp supplies an electrical ground (via a 10 kΩ input impedance) for the circuit and produces a voltage proportional to the input current. The preamp ac and dc output voltages are measured by digital multimeter and lock-in amplifier, respectively. A preamp sensitivity of 10 nA/V was chosen to balance amplification (for improved signal-to-noise), frequency attenuation (for ac measurements), and input impedance. This experimental setup is shown in Fig. 2.7.
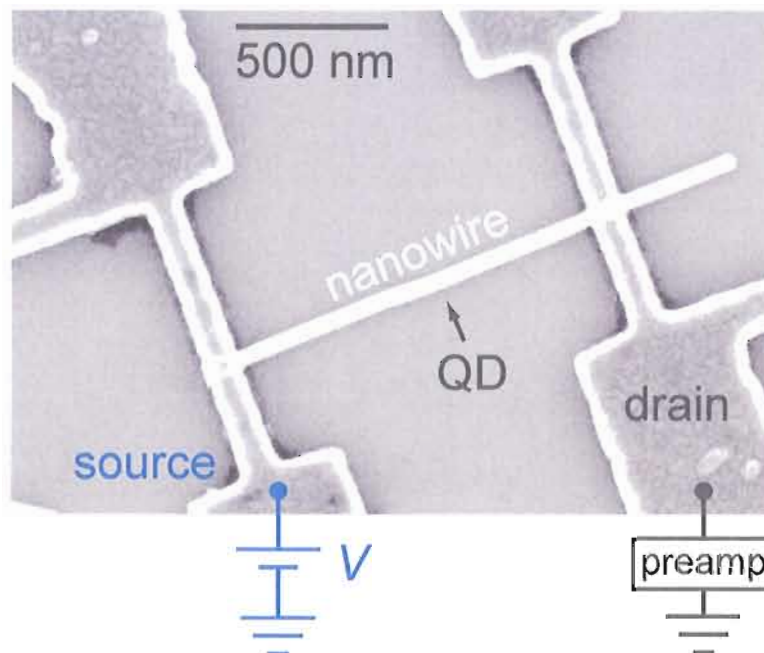
Figure 2.7. An SEM image of an InAs/InP heterostructure nanowire with source and drain contacts. The indicated quantum dot (QD) embedded in the nanowire is not resolved by the SEM. For isothermal IV measurements, the voltage adder in Fig. 2.6 baises the nanowire with an ac + dc voltage. The current preamp provides electrical grounding and an amplified ac + dc output voltage proportional to input current. The other terminals are floating.

The first and most simple measurement performed with this setup is current as a function of gate voltage with a fixed bias voltage, as shown in Fig. 2.8. The result is a series of peaks occurring at gate voltages corresponding to the energy resonances of the quantum dot; this is proof that the sample is functioning as desired. Additional measurements are required to convert gate voltage to energy via α, but once they are complete, differential conductance data can be fit using Eq. (II.9), as shown in the inset of Fig. 2.8. The agreement between data and theory demonstrates that $\Gamma \ll kT$ when $T \geq 550$ mK.
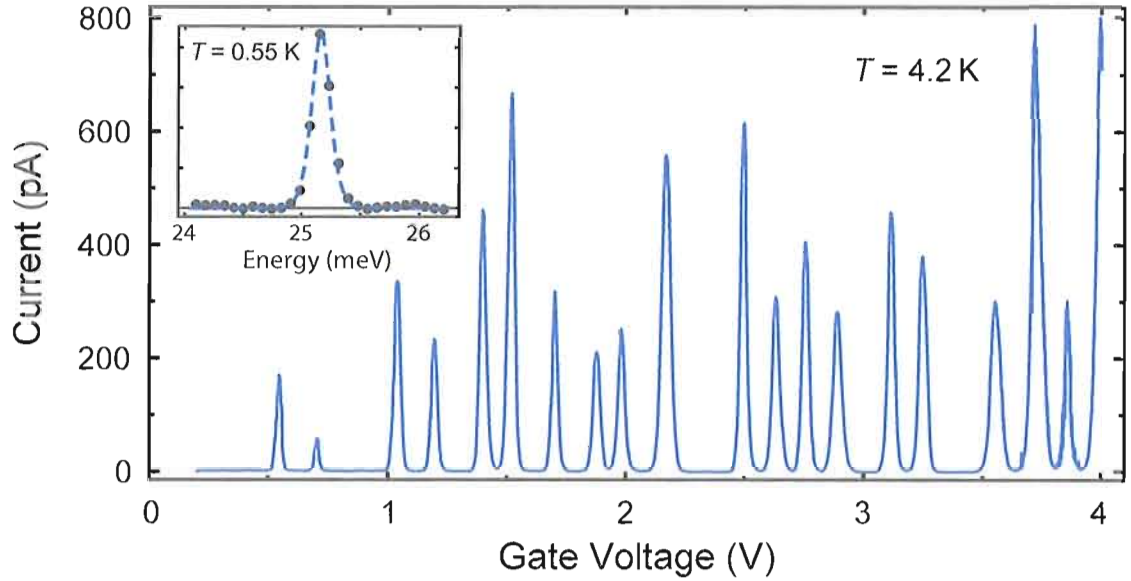
Figure 2.8. Current measured as a function of gate voltage shows peaks at each energy resonance of the quantum dot. The background temperature is $T = 4.2$ K. Once α is measured, gate voltage can be converted to an energy scale. *Inset*: Differential conductance at $T = 0.55$ K has been fit using Eq. (II.9) demonstrating that $\Gamma \ll kT$ whenever $T \geq 0.55$ K.

Coulomb-Blockade Spectroscopy

At low temperatures (< 10 K), where kT is less than the level spacing of the quantum dot, the device will operate in the Coulomb-blockade regime. (Theory regarding Coulomb blockade is in Section II.3.1.) Sweeping the bias and gate voltages while measuring differential conductance via a small ac bias provides the Coulomb-blockade diamonds as shown in Fig. 2.9. This Coulomb-blockade spectroscopy is essential for mapping the important working points of the system. For example, the diamonds are used to find suitable locations for quantum-dot thermometry (see Chapter III) and nonlinear experiments (see Chapter VI). In addition, the aspect
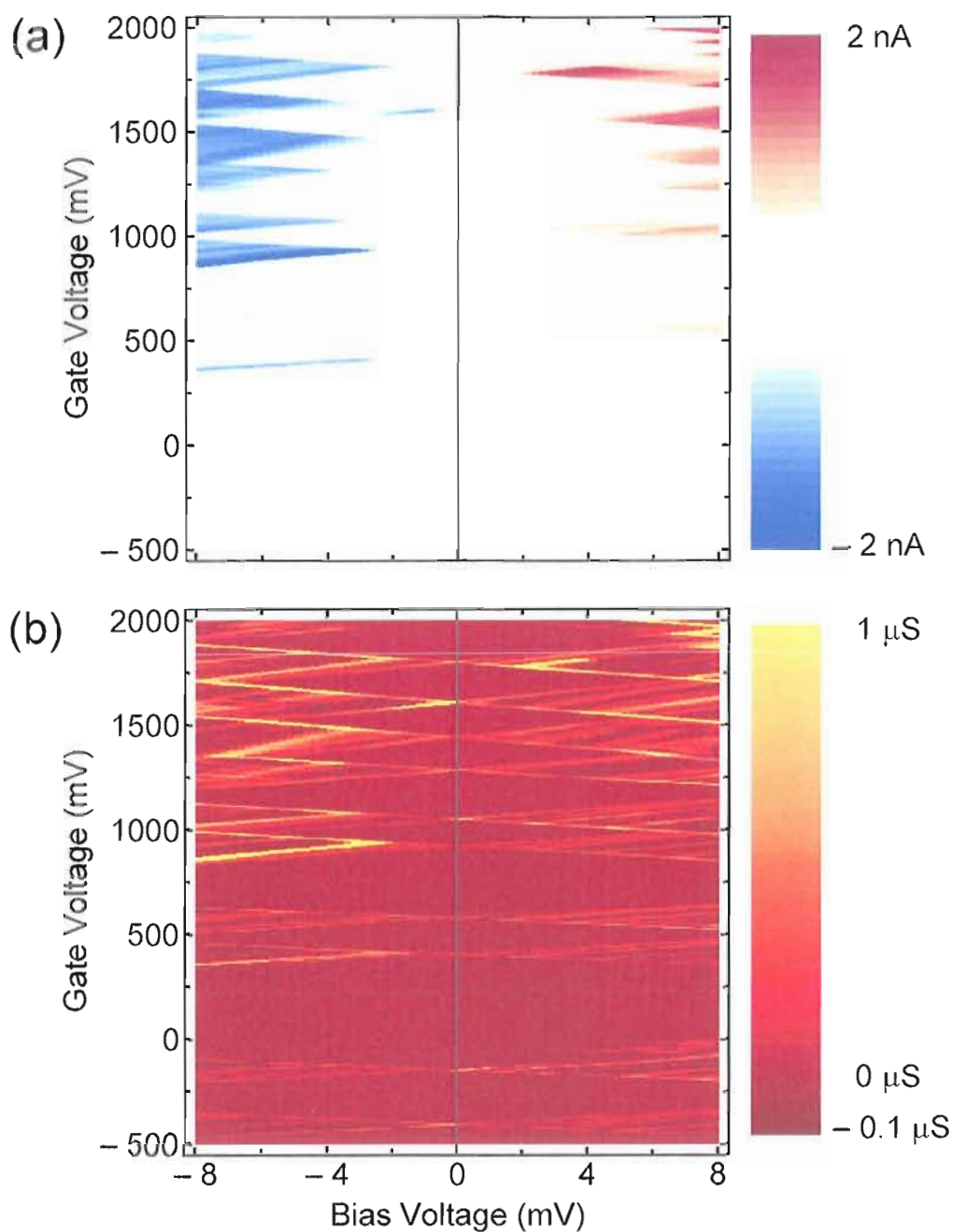
Figure 2.9. Coulomb-blockade measurements at $T = 550$ mK. **(a)** Current and **(b)** differential conductance were measured simultaneously as a function of bias voltage and gate voltage. The advantages of differential conductance are: no offset current, sharper features, better signal-to-noise with lock-in techniques, and excited states and regions of negative differential conductance are easily identified.

ratio of the diamonds reflects the value of α, as shown in the next section. Next, the energetic positioning of the quantum dot ground state resonances can be quantified, and the existence of excited states outside the diamonds can also be studied [57]. And last, the diamonds show regions of negative differential conductance (owing to phonons, blocking states, etc.) beyond the simple picture of quantum-dot transport. These regions are avoided for thermometry measurements, but are of interest for nonlinear measurements.

<div align="center">Measuring α, E_{add}, and E_{C}</div>

The slopes of the Coulomb-blockade diamonds are determined by the various capacitances (C_{S}, C_{D}, and C_{G}) of the system to the quantum dot. The "lever arm" of the gate, $\alpha = C_{\text{G}}/C_{\Sigma}$, quantifies the influence of the gate voltage on the electrochemical potential of the dot. Quantitative analysis of the Coulomb-blockade diamonds, made possible by the constant interaction model (see Section II.3.1), reveals the value of α. In particular, the sum of the slopes of the Coulomb-blockade diamonds (see Fig. 2.10) is a direct measurement of α. The derivation of this observation follows next.

When measuring differential conductance, the drain is connected to the current preamp so that $\mu_{\text{D}} = -eV_{\text{D}} = 0$. Meanwhile, a small, constant ac bias is applied across the nanowire as the bias voltage and gate voltage map out the parameter space. On the boundary of any Coulomb-blockade diamond, the electrochemical potential of the dot, μ, is aligned with either the source or drain electrochemical potential, μ_{S} or μ_{D},

respectively. If the source is aligned to the dot, then $\mu = \mu_S = -eV$, where V is the bias voltage. Inserting the conditions $V_D = 0$, $V_S = V$, and $\mu = -eV$ into Eq. (II.1) gives,

$$-eV = E_C \left(N - N_0 - \frac{1}{2} \right) - \frac{E_C}{e} \left(C_G V_G + C_S V \right) + E_N.$$

Isolating the gate voltage gives,

$$V_G = \left(\frac{C_D + C_G}{C_G} \right) V + \frac{e}{C_G} \left(N - N_0 - \frac{1}{2} \right) + \frac{C_\Sigma E_N}{e C_G}.$$

Therefore, the slope of the "source" portions of a Coulomb-blockade diamond is

$$S_S \equiv \frac{dV_G}{dV} = \frac{C_G + C_D}{C_G}. \tag{II.10}$$

In the opposite case, when the drain is aligned to the dot, the conditions are $\mu = \mu_D = 0$ and $V_S = V$, and Eq. (II.1) gives,

$$0 = E_C \left(N - N_0 - \frac{1}{2} \right) - \frac{E_C}{e} \left(C_G V_G + C_S V \right) + E_N,$$

And here isolating the gate voltage gives,

$$V_G = -\frac{C_S}{C_G} V + \frac{e}{C_G} \left(N - N_0 - \frac{1}{2} \right) + \frac{C_\Sigma E_N}{e C_G}.$$

So the slope along the "drain"portions of a Coulomb-blockade diamond is,

$$S_{\mathrm{D}} \equiv \frac{dV_{\mathrm{G}}}{dV} = -\frac{C_{\mathrm{S}}}{C_{\mathrm{G}}}. \tag{II.11}$$

Note that a perfectly symmetric Coulomb-blockade diamond requires $|S_1| = |S_2|$, which is *not* satisfied by $C_{\mathrm{S}} = C_{\mathrm{D}}$ alone, but rather by the condition $C_{\mathrm{S}} = C_{\mathrm{D}} + C_{\mathrm{G}}$. However, when $C_{\mathrm{G}} \ll C_{\mathrm{D}}$, which is typical for a $\sim$10 nm quantum dot in a $\sim$50 nm nanowire system [62], then $S_{\mathrm{S}} \approx C_{\mathrm{D}}/C_{\mathrm{G}}$. If as well $C_{\mathrm{S}} \approx C_{\mathrm{D}}$, then it is true that $S_{\mathrm{S}} \approx S_{\mathrm{D}}$,and the diamond will be nearly symmetric.

If the magnitude of these two slopes (Eqs. (II.10) and (II.11)) are summed, then

$$|S_{\mathrm{S}}| + |S_{\mathrm{D}}| = \frac{C_{\mathrm{G}} + C_{\mathrm{D}}}{C_{\mathrm{G}}} + \frac{C_{\mathrm{S}}}{C_{\mathrm{G}}} = \frac{C_{\mathrm{G}} + C_{\mathrm{D}} + C_{\mathrm{S}}}{C_{\mathrm{G}}} = \frac{C_{\Sigma}}{C_{\mathrm{G}}} = \alpha^{-1} \tag{II.12}$$

where the last equality follows from Eq. (II.3). Therefore, α can be calculated using measured values of S_{S} and S_{D}, and Fig. 2.10 demonstrates such a calculation.

Once α has been measured, a slice down the center of the Coulomb-blockade diamond ($V = 0$) as a function of V_{G} can be converted into energy revealing the energetic level spacing. Fig. 2.11 is such a plot.

The addition energy, E_{add}, as a function of filling number, N (see Eq. (II.2)) is determined from Fig. 2.11 by measuring the peak-to-peak energies. The result is plotted in Fig. 2.12. The even fillings are due solely to the charging energy as they split spin-degenerate energy levels. By taking an average of $E_{\mathrm{add}}\,(N_{\mathrm{even}})$, in Fig. 2.12,
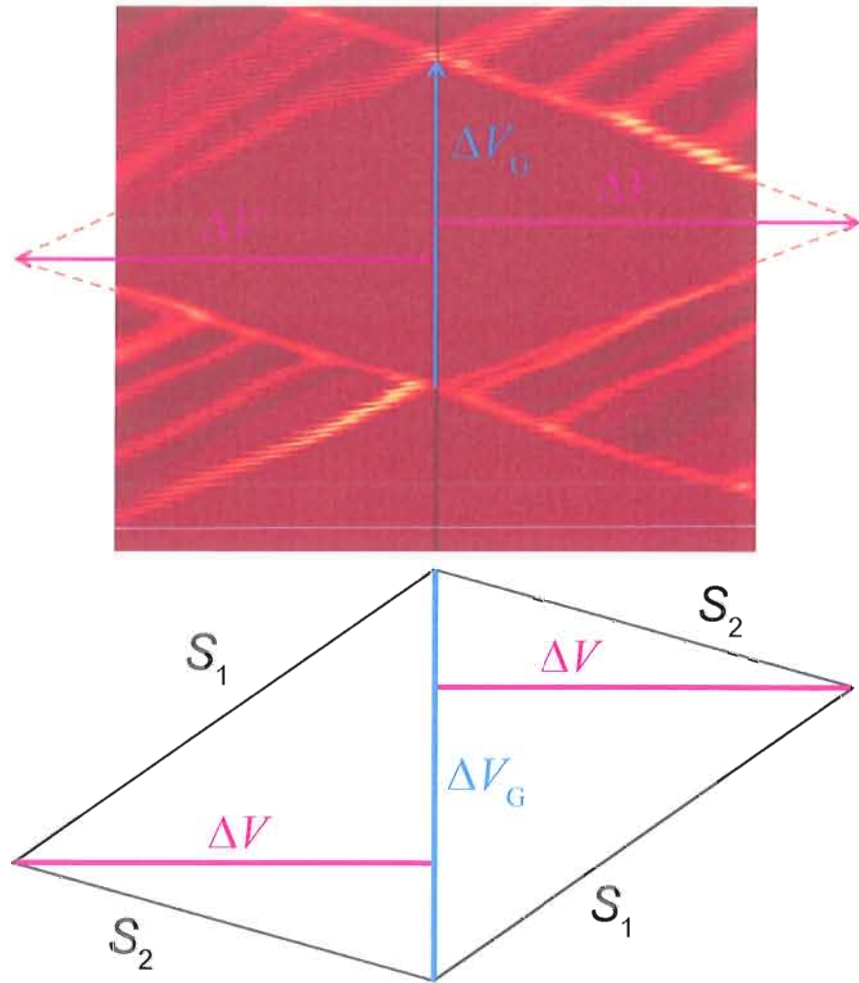
Figure 2.10. A section of a measured Coulomb-blockade diamond and a corresponding diamond sketch. Because this diamond is asymmetric (as most are), the relationship between the slopes of the diamond, S_1 and S_2, and the voltages, ΔV_G and ΔV, is not simple. Nonetheless, S_1 and S_2 can be measured, and they provide a measurement of α according to Eq. (II.12). For this diamond $\alpha = 0.071 \pm 0.008$.
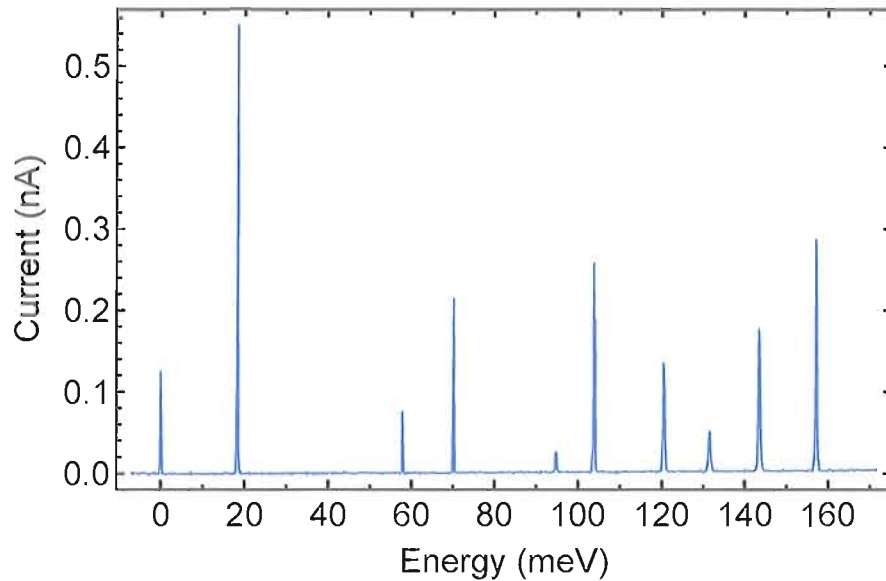
Figure 2.11. A vertical slice of the Coulomb-blockade diamond of Fig. 2.9a near zero bias. Here the dc current is plotted as a function of dot energy, which has been converted from gate voltage via α. (See Fig. 2.10 and associated text for details on measuring α.) The positions of these peaks are used to determine $E_{\text{add}}(N)$ (see Fig. 2.12).

a charging energy of $E_C = 12.9$ meV is found. (The $N = 2$ filling number was not included in the average because this filling number is an outlier in all samples, probably due to non-constant capacitance.)

Applying Temperature Differences Across Mesoscopic Devices

For thermoelectric experiments, a temperature difference must be established across the quantum dot. One approach is to heat the phonons in the underlying lattice with a nearby, separate metallic heating channel, with the result of heating the electron gas in the device via electron-phonon coupling [20–23, 63, 64]. This has the advantage of decoupling the heating electronics and the measurement electronics,
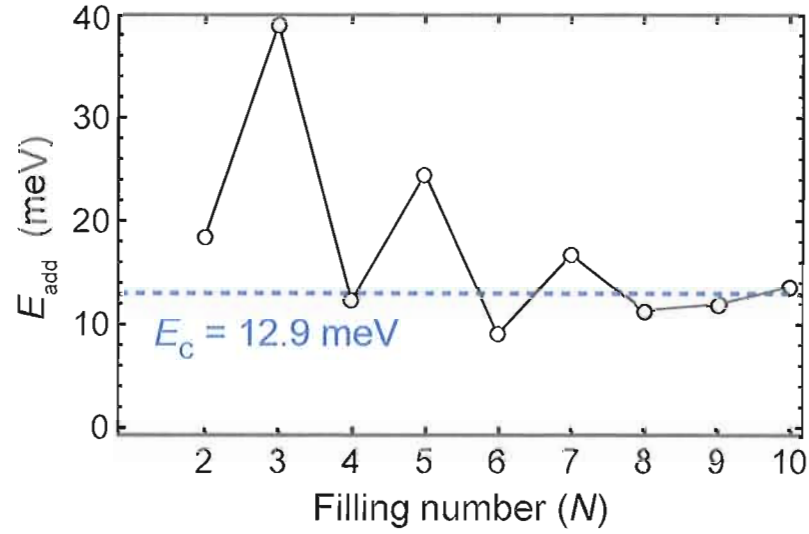
Figure 2.12. The addition energy, E_{add}, of a quantum dot for the first 10 electrons. These data have been collected via the conductance peaks in Fig. 2.11. The average value of the even fillings provides the charging energy, $E_C = 12.9$ meV. Note that at $T = 10$ K, $kT = 860\ \mu\text{eV} \ll E_C$ demonstrating that nanowire-based quantum dot have energy resonances which are very well separated in energy.

but relies on electron-phonon coupling that is weak at low temperatures (< 4 K). Relatively large heating currents and phonon temperatures might be required to raise the temperature of the electron gas in such a device. Besides adding electrical noise to the system and increasing the heating load on the cryostat, this technique creates hot phonons that can pull electrons to one end of the sample and produce a phantom thermovoltage, the so-called phonon-drag thermovoltage [23, 65].

Direct-electron heating is also commonplace in mesoscopic thermoelectric experiments for heating an electron gas [65–68]. This technique uses an electric current to heat an electron gas via Joule heating. This is the approach used in these experiments.

To accomplish this in a nanowire sample, the source and drain electrical contacts to the nanowire are made with three terminals arranged in a 'T'-shape (see Fig. 2.13a). A heating current, I_H, is supplied between the top two terminals, which heats the top bar of the 'T'-shaped contact via Joule heating. The top portion of the 'T', the heating channel, is intentionally made thin to increase resistance and the effect of Joule heating. The bottom terminal is used for electrical balancing. The heating channel is placed over the nanowire, and I_H heats the electron gas in the heating channel and ultimately in the nanowire, as shown in Fig. 2.13a.

The challenge of this setup is that the connection for heating and electrical biasing is one and the same. Therefore, the heating and electrical biasing circuitry must be combined. This is achieved by using a home-built op-amp heating circuit, which has independent heating and biasing voltage inputs and produces two output voltages for each end of the heating channel. The voltage probe assists in tuning the output voltages relative to one another.

The heating box splits the input heating voltage into two op-amps, one of which is inverting. This produces two output heating voltages, $v_\pm$, which are equal in magnitude and opposite in sign. These two voltages are applied across the heating channel (see Fig. 2.13a). Manual trimming built into the heating circuit offers the ability to tune $v_\pm$ relative to one another so that $v = v_+ + v_- = 0$ at the nanowire. Therefore, a finite heating current, I_H, flows through the heating channel without electrically biasing the nanowire. The other function of the heating circuit is to float
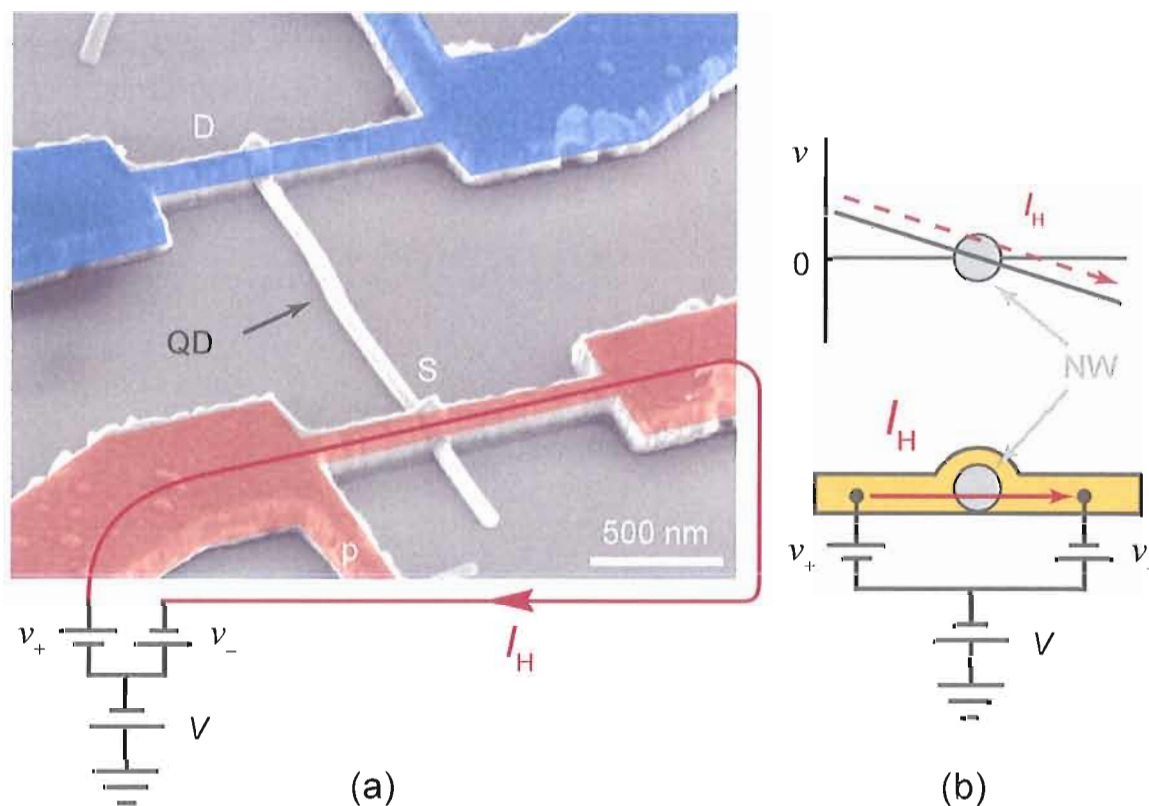
Figure 2.13. (a) An SEM image of a nanowire with source (S) and drain (D) contacts and an embedded quantum dot (QD), which is not resolved at this magnification. The voltage, V, biases the nanowire electrically. Via Joule heating, the heating current, I_H, biases the nanowire thermally by raising the temperature of the electron gas in the narrow heating channel of the source contact. Current measurements are performed by connecting a current preamp to the drain, the same as under isothermal conditions (see Fig. 2.7). Thermovoltage measurements are made using a differential voltage preamplifier connected to both source and drain (see Fig. 2.14). **(b)** A side-view of the source heating channel and the ac bias voltage, v, in the channel. The voltage probe (p) assists in tuning the heating voltages, $v_{\pm}$, relative to one another so that $v = v_+ + v_- = 0$ at the nanowire. In this way, I_H does not bias the nanowire electrically.

$v_\pm$ above ground on top of the input bias voltage, V. A voltage adder (see Fig. 2.6) can be attached to the bias voltage input of the heating circuit to give V ac and dc components. The resulting voltage outputs from the heating circuit are $V + v_\pm$, and the manner in which they connect to the sample is shown in Fig. 2.13b. The op-amp circuit is provided in Appendix B.

For all experiments, a low-frequency (17 Hz) ac heating current was used. The benefits of using an ac heating current are two fold. First, lock-in amplification techniques are used for data acquisition thereby increasing the signal-to-noise ratio. Second, the resulting temperature gradient is frequency doubled, making thermally induced electron transport very distinct in frequency space. This statement is motivated by the following exercise.

The temperature change, ΔT_{H}, of the electron gas in the heated source contact is proportional to the heating power, $P = I_{\mathrm{H}} V = I_{\mathrm{H}}^2 R$, where heating current, I_{H}, can be described as $I_{\mathrm{H}} = I_0 \cos(\omega t)$. Thus the functional form of ΔT_{H} is

$$\Delta T_{\mathrm{H}} \propto P \propto I_{\mathrm{H}}^2 \propto \cos^2(\omega t) \propto \cos(2\omega t). \tag{II.13}$$

Any currents or voltages at frequency 2ω can only be attributed to thermal effects because all other processes are linear and therefore have frequency ω.

In thermoelectric experiments using an ac heating current at frequency ω,

according to Eqs. (I.1) and (I.16), the net (time-dependent) current is

$$I\left(t\right) = GV + Gv\cos\left(\omega t\right) + G_{\text{th}}\Delta T, \qquad\qquad\text{(II.14)}$$

where V is the bias voltage and the term $v\cos\left(\omega t\right)$ accounts for any voltage noise at frequency ω. This can be written

$$I\left(t\right) = GV + Gv\cos\left(\omega t\right) + I_0\cos\left(2\omega t\right),$$

where I_0 is the amplitude of the ac thermocurrent, $I_{\text{th}} = I_0\cos\left(2\omega t\right) = G_{\text{th}}\Delta T$. (A phase factor of $\pi/2$ has been omitted for clarity.) Measuring the second harmonic relative to a reference signal is easily accomplished using a commercially available lock-in amplifier. The amplitude I_0 is found by integrating $I\left(t\right)$ with $\cos\left(2\omega t\right)$,

$$\frac{\omega}{\pi}\int_0^{2\pi/\omega} I\left(t\right)\cos\left(2\omega t\right)dt = \frac{\omega}{\pi}\int_0^{2\pi/\omega} \left[GV + Gv\cos\left(\omega t\right) + I_0\cos\left(2\omega t\right)\right]\cos\left(2\omega t\right)dt$$

$$= 0 + 0 + I_0\frac{\omega}{\pi}\int_0^{2\pi/\omega}\cos\left(2\omega t\right)\cos\left(2\omega t\right)dt$$

$$= I_0$$

Therefore, to measure I_0, the lock-in amplifier simply produces a periodic signal of frequency 2ω and averages over one period. In particular, the measurement is completely insensitive to $v\cos\left(\omega t\right)$ because that term integrates to zero. A second

lock-in amplifier operating at frequency ω can measure Gv and ignore I_0. See Ref. [69] for a pedagogical introduction to the lock-in amplification technique.

The two largest sources of noise at frequency ω are electrical biasing by I_H that is not completely nulled in the heating circuit and inductance between cryostat leads. These noise sources do not interfere with the detection of signals at frequency 2ω. Thus ac heating offers many experimental advantages.

<div align="center">Thermovoltage Measurement Technique</div>

By definition (see Eq. (I.2)), V_{th} must be measured in the absence of electrical current. That is, a thermovoltage measurement is an open-circuit measurement. To achieve this, the voltage measurement instrument must have a large input impedance. The instrument used here (a Fempto DLPVA-100-F-D Series) is a low-noise voltage preamplifier with an FET input stage that provides an input impedance of 1 TΩ. The setup is shown in Fig. 2.14.

The system is made more complex by the back-gate used to tune the quantum dot's energy resonances. When measuring current, nanowires's drain contact is grounded (via the current preamp) and the source contact is pinned to the bias voltage (see Fig. 2.13). In this configuration, the influence of the back-gate on the source and drain ends of the nanowire is miniscule because InAs semiconductor material is pinned to the metallic source and drain contacts.

In a true open-circuit thermovoltage measurement, the cold drain contact is disconnected, and the hot source contact remains connected to the heating circuit
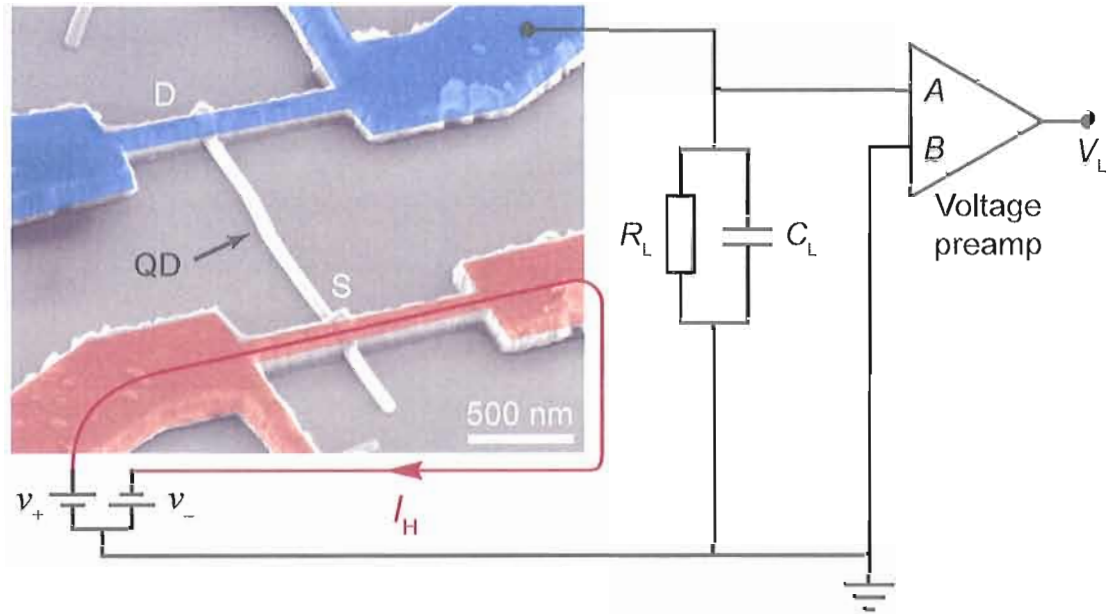
Figure 2.14. A schematic of the experimental setup for measurements of V_{th}. Heating is the same as in Fig. 2.13. Gating effects prevent true open-circuit measurement of V_{th}. This is mitigated by using a load resistance, R_{L}. As a result, the voltage across inputs A and B of the differential voltage preamplifier is V_{L}, that is, the voltage across R_{L}. The true thermovoltage is $V_{\mathrm{th}} = V_{\mathrm{L}}\left(1 + R/R_{\mathrm{L}}\right)$ (see Eq. (II.15)). The load capacitance, C_{L}, is provided by the coaxial voltage cables.

at voltage, V_{S}. In this case, the drain is allowed to float, and the back-gate can bend the semiconductor bands in the drain end of the nanowire. This influences the electrochemical potential in the drain and therefore the drain voltage, V_{D}. Meanwhile, the source electrochemical potential is held constant by V_{S}. A change in V_{D} created by a change in gate voltage biases the nanowire. This bias will eventually create current flow through the quantum dot as the system attempts to establish equilibrium. As a result, the open-circuit voltage waxes and wanes as a function of gate voltage, and the characteristics of this process depend on the properties of the quantum dot. In fact,

the slopes of the waxing and waning are the same as the slopes of Coulomb-blockade diamonds (see Appendix C). The main result is shown theoretically in Fig. 2.15 and experimentally in Fig. 2.16.

The behavior of the open-circuit voltage can be explained by considering the capacitance among the gate, source, drain, and dot. During open-circuit voltage measurements, the source electrochemical potential, μ_S, is held constant at the voltage of the source contact. The drain electrochemical potential, μ_D, is determined by two factors: 1) the amount of charge in the drain and 2) the voltage of the back-gate, V_G. The electrochemical potential of the dot, μ, is influenced by V_G, μ_S, and μ_D (see Eq. (II.1)).

At all gate voltages, either μ_S or μ_D is aligned with a quantum dot resonance. This is guaranteed by carrier flow leaking through the quantum dot and by the capacitive coupling between quantum dot and μ_D. Because the drain portion of the nanowire is much larger than the dot, V_G couples more strongly to μ_D than it to μ—meaning that μ_D moves faster than μ when V_G changes. When sweeping through two resonances at energies ε_1 and ε_2, the open circuit voltage, $eV_{OC} = \mu_S - \mu_D$, follows the unique path in the V_{OC}–V_G plain where either $\varepsilon_1 = \mu_D$ or $\varepsilon_2 = \mu_S$. This is shown in Fig. 2.15. The uphill slope of dV_{OC}/dV_G is proportional to C_G because as V_G pulls the dot down, V_{OC} increases. On the other hand, μ_S is "holding" the dot up via the source-to-dot capacitance, C_S, and, therefore, the slope is inversely proportional to C_S. To lowest order, the uphill slope is C_G/C_S, and the exact slope is $C_G/(C_G + C_S)$. V_{OC}
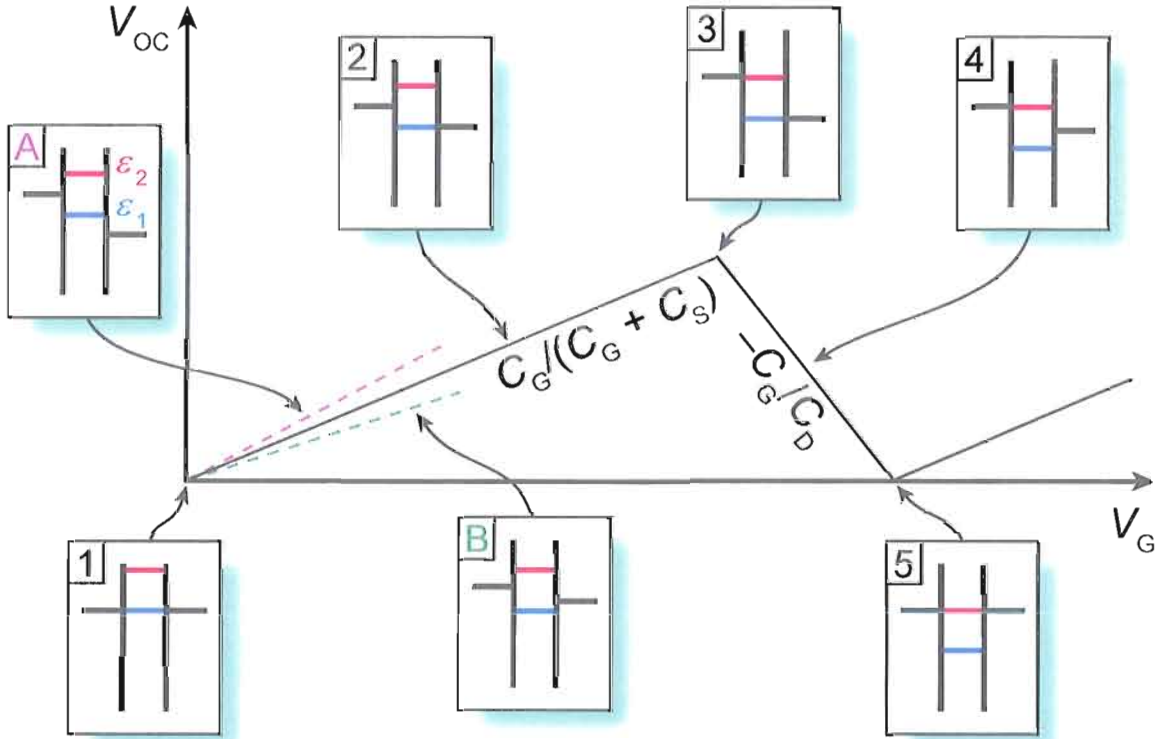
Figure 2.15. A schematic demonstrating the behavior of open-circuit voltage, $eV_{OC} = \mu_S - \mu_D$, as a function of gate voltage, V_G. The panels depict the source (S) and drain (D) electrochemical potentials, μ_S and μ_D, respectfully, as well as two quantum-dot energy resonances, ε_1 and ε_2. Panels 1–6 describe the behavior at key points in the diagram while panels A and B highlight forbidden behavior. See the main text for additional details. **(1)** $\mu_S = \mu_D = \varepsilon_1$ when $V_{OC} = 0$. **(A)** When V_G increases, μ_D decreases faster than ε_1, but electron flow prevents this. **(B)** The situation is forbidden because μ_D decreases faster than ε_1, and thus V_{OC} cannot follow the green dashed line. **(2)** V_{OC} follows the solid black line where $\varepsilon_1 = \mu_D$. The slope here is $C_G/(C_G + C_S)$, as shown. V_{OC} continues to climb until **(3)** when ε_1 aligns with μ_S. Electrons start to flow into the drain through ε_1. **(4)** Electrons fill the drain, lifting μ_D and decreasing V_{OC}. The rate at which V_{OC} decreases depends on the ratio C_G/C_D. **(5)** The system has returned to $V_{OC} = 0$, $\varepsilon_2 = \mu_S = \mu_D$, and the process will repeat sweeping through the resonances at ε_3, ε_4, etc. The derivations of the slopes are provided in Appendix C.

continues to increase until the transition into the next energy resonance. Here the slope dV_{OC}/dV_G changes direction as electrons flow from source to drain thereby lifting μ_D and decreasing V_{OC}. The larger C_G, the harder V_G pulls down the quantum dot and the faster V_{OC} decreases. μ_D opposes this change via the drain-to-dot capacitance, C_D, and the larger the C_D the slower μ decreases. The rate at which V_{OC} decreases depends on the ratio C_G/C_D. The complete derivations of the slopes are provided in Appendix C and the process is illustrated in Fig. 2.15.
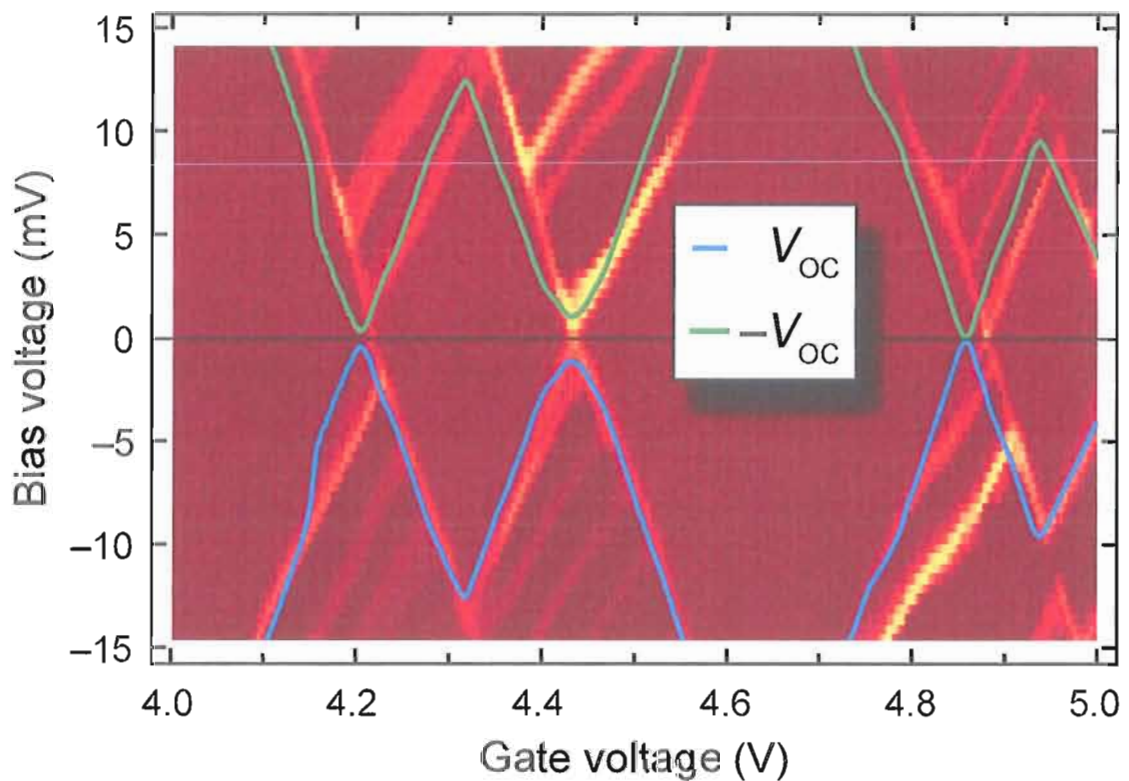
Figure 2.16. Open-circuit voltage, V_{OC}, plotted on top of Coulomb-blockade diamonds. Both V_{OC} and $-V_{OC}$ follow the Coulomb-blockade diamonds, just as the constant interaction model predicts. See Fig. (2.15) and text for the theoretical explanation of this effect. Note that bias voltage and V_{OC} have opposite sign as they are measured oppositely.

This behavior makes true open-circuit thermovoltage measurements impossible. To mitigate this problem, the load resistance, R_L, must be placed between the drain and ground, as shown in Fig. 2.14. This drastically reduces the open-circuit voltage induced by the back-gate by providing an electrical path to ground. The small residual open-circuit voltage owing to the back-gate is measured under isothermal conditions and subtracted from thermovoltage data. For ac measurements, one must also consider that the capacitance in the coaxial cables provides a load capacitance, C_L, connected in parallel with R_L.

This is not an open-circuit measurement, so the measured voltage is not thermovoltage. What is really measured is V_L, the voltage across the load resistor, R_L (see Fig. 2.14). Therefore, this measurement is only a lower-bound of the true thermovoltage. The system is analogous to a battery with an internal resistance, R, running a load with resistance R_L. Therefore, the true thermovoltage is,

$$V_{th} = I\left(R_L + R\right) = \frac{V_L}{R_L}\left(R_L + R\right) = V_L\left(1 + \frac{R}{R_L}\right), \qquad (\text{II.15})$$

where $R = 1/G$ is the resistance of the nanowire. Thus the scaling factor, ϕ, between the measured voltage, V_L, and true thermovoltage, V_{th}, is

$$\phi \equiv \frac{V_{th}}{V_L} = \left(1 + \frac{R}{R_L}\right). \qquad (\text{II.16})$$

The situation is made more complicated because G is not constant, but rather a

function of gate voltage. However, if $R_L \gg R$, then $V_{th} \approx V_L$ at all gate voltages, and no additional measurements are necessary. Alternatively, the effective resistance of the device can be found using power measurements. For details, see Chapter V.

CHAPTER III

QUANTUM-DOT THERMOMETRY

Quantitative evaluation of a quantum-dot-based thermoelectric requires the application of a known temperature difference across the device. The heating technique used for this purpose is discussed in Section II.4.4. This chapter focuses solely on quantum-dot thermometry, which was developed specifically for measuring the electron temperature differences across a quantum dot.

Introduction

Mesoscopic devices exhibit a range of unique thermal phenomena including the quantization of thermal conductance [70] and nonlinear [71] and energy-modulated [72] thermovoltage. Understanding mesoscopic thermal phenomena and applications of such phenomena require knowledge of the absolute ambient temperature as well as knowledge about the applied temperature difference that exists across length scales that are smaller than typical thermometry techniques can resolve.

A challenge when performing a low-temperature, quantitative, mesoscopic thermal experiment is the measurement of the temperature difference across the small device under conditions where electronic and lattice temperatures can be very different. In this context, such thermometry is often accomplished using the thermovoltage of

a QPC to measure temperature differences within a two-dimensional electron gas [65, 70, 72–74]. While thermal experiments using micrometer-length nanotubes and nanowires often exploit the temperature dependence of local resistors to determine the lattice temperature difference across the device [20, 21, 23, 63, 64, 75].

Quantum-dot thermometry [76–79] is a technique which exploits temperature-dependent electron transport through a two-terminal quantum dot in order to measure directly the temperature of the electrons entering the dot. At the very heart of this technique lies the assumption that each energy resonance of the quantum dot has the ability to sample separately the energy-dependent probability distribution of the electron gas at either the source or drain side of the quantum dot. For this assumption to hold, the thermometry technique requires that the energy difference between neighboring energy resonances, $\Delta\varepsilon$, is larger than the thermal energy, kT.

One way to fulfill this requirement is to use a quantum dot in the Coulomb-blockade regime, that is, with a charging energy, $E_{\mathrm{C}} \gg kT$. In very small quantum dots, quantum-confinement effects yields additional level separation. This energy requirement is easily achieved (even near, for example, $T = 10$ K) using a nanowire in which a sufficiently small quantum dot has been defined ($\sim$10 nm) [62]. The dot could be defined via physical barriers created with heterostructure growth or via electrostatically depleted barriers created by finger gates. The dimensions of quantum dots in two-dimensional electron gas devices are typically on the order of hundreds or thousands of nanometers and therefore quantum-dot thermometry with such a

device would be limited to much lower temperatures. Other devices, for example, metallic islands, which rely completely on Coulomb blockade, rather than quantum confinement, to create single transmission resonances could also be used for quantum-dot thermometry.

Many primary and secondary thermometers are available for low-temperature physics applications [80], some with higher precision than is possible with quantum-dot thermometry. However, quantum-dot thermometry provides a measurement of electron temperature difference across the quantum dot using the dot itself as the thermometer, making possible an all-in-one experimental device.

Because the quantum dot's energy resonance is effectively the tool which measures the thermal envelope of the electron gas Fermi-Dirac distribution, the energetic width of the resonance, Γ, (see Eqs. (II.7) and (II.8)) influences the outcome of the measurement. The original proposal [77] considered a limit in which $\Gamma \gg kT$. Later the technique was extended [79] into the limit where $\Gamma \ll kT$. Quantum-dot thermometry is introduced best via a unified theoretical treatment [78] discussing these two opposite regimes and the intermediate regime which separates the two.

To apply a temperature difference across the dot, an ac heating current, I_H, modulates the temperature of an ohmic contact, taken here to be the source contact, via Joule heating (see Section II.4.4). As a result, the temperature of the electron gas in the source contact rises ΔT_{SC} with respect to the unperturbed background temperature, T, thereby establishing an unknown temperature profile along the

nanowire. As a result, the electron gas near the source and drain sides of the quantum dot are elevated to $T_{\mathrm{H,C}} = \Delta T_{\mathrm{H,C}} + T$. The quantities of interest are the electronic temperature rises, $\Delta T_{\mathrm{H,C}} = T_{\mathrm{H,C}} - T$, in the immediate vicinity of the quantum dot (see Fig. 3.1).
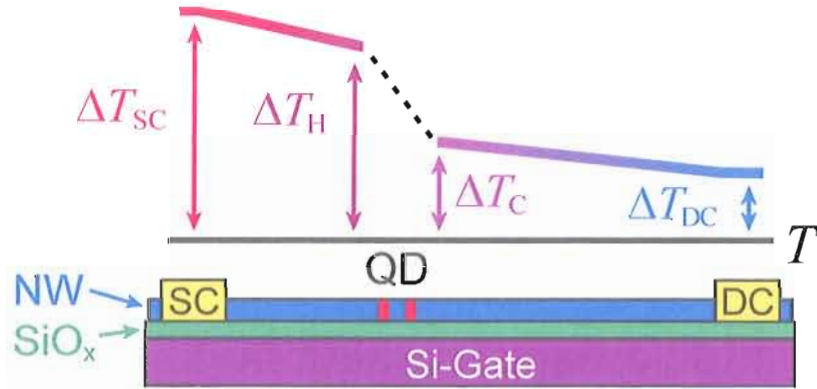
Figure 3.1. A schematic (not to scale) of a heated nanowire (NW) with embedded quantum dot (QD), hot source contact (SC), and cold drain contact (DC). The quantum dot is not necessarily centered between contacts. A heating current in the source contact creates the illustrated electron temperature gradient along the nanowire. In the metallic source and drain contacts, the electron gas temperature rises by $\Delta T_{\mathrm{SC,DC}}$, respectively, above the cryostat temperature, T. In the nanowire, the electron gas temperature rises by $\Delta T_{\mathrm{H,C}}$ near the source and drain sides of the quantum dot, respectively ($\Delta T_{\mathrm{SC}} \geq \Delta T_{\mathrm{H}} \geq \Delta T_{\mathrm{C}} \geq \Delta T_{\mathrm{DC}}$).

During thermometry measurements, the nanowire is biased so that electrons flow predominantly from either the source or drain contact, and the energy-dependent transmission probability, $\tau(\varepsilon)$, samples the energy distribution of the incoming electrons (see Fig. 3.2). As a result, any electric current—a physically measurable quantity—carries with it information about the temperature of the electron gas precisely at the quantum dot. In the following, it is shown that $\Delta T_{\mathrm{H,C}}$ can be extracted

via a comparison of two measured quantities: the differential thermocurrent, I_{th}, and

the second differential conductance, $G_2 \equiv \partial^2 I/\partial V^2$.
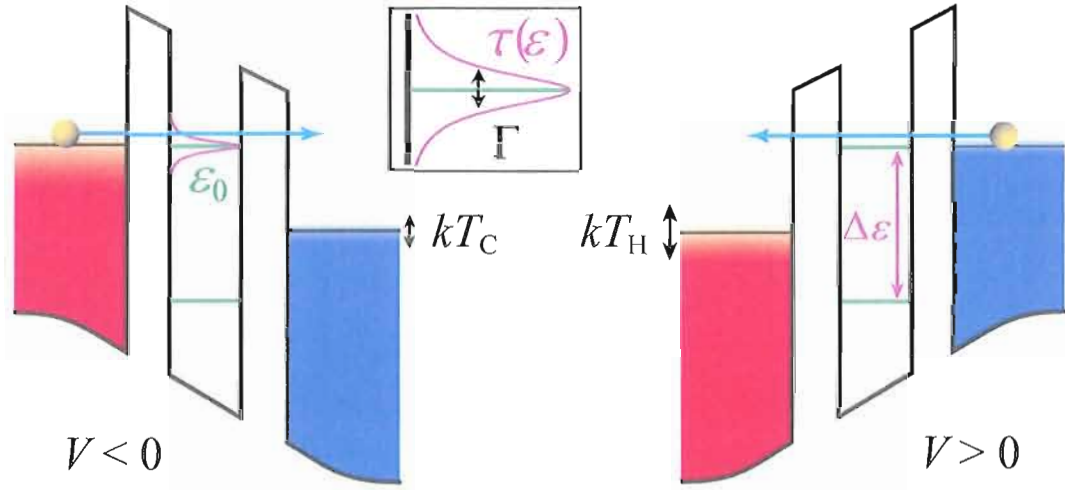
Figure 3.2. The band diagram of a quantum dot subjected to an applied voltage bias and temperature difference and the resulting source and drain electron gas Fermi-Dirac distributions. An energy of $\Delta\varepsilon$ separates the dot's energy resonances, and $\Delta\varepsilon \gg kT_{H,C}$. *Inset*: Each transmission resonance can be described by a Lorentzian function with an energetic width of Γ (see Eq. (II.8)). Quantum-dot thermometry has two regimes of operation in which either $\Gamma \ll kT$ or $\Gamma \gg kT$ and requires that $\Gamma \ll \Delta\varepsilon$ in both regimes. *Left*: While negatively biased, a transmission resonance at energy ε_0 samples only the source electron distribution. *Right*: The opposite case under positive bias.

The Theory of Quantum-dot Thermometry

When the device is connected to source and drain contacts, the applied temperature difference produces a thermocurrent, I_{th}, that is, a thermally induced net electron flow. Measurements of I_{th} in conjunction with conductance spectroscopy are used to

deduce the temperature rises $\Delta T_{\mathrm{H,C}}$ separately. To measure $\Delta T_{\mathrm{H,C}}$ separately, the nanowire is biased to sample hot source and cold drain electrons individually.

By tuning the bias and gate voltages, respectively, one can adjust V and μ in order to satisfy one of two conditions: $\varepsilon_0 = \mu_{\mathrm{H,C}} = \mu \mp eV/2$, therefore centering either f_{H} or f_{C} at ε_0 (see Eq. (I.14)). In terms of applied bias voltage, these two conditions are,

$$V_{\mathrm{H,C}}^0 \equiv \mp 2\left(\varepsilon_0 - \mu\right)/e. \tag{III.1}$$

Tuning to bias voltages in a range near this condition allows the transmission function to sample only the overlapping Fermi-Dirac distribution because the opposite distribution is several kT away from any energy resonance provided that the energetic spacing of the quantum dot is sufficiently larger than kT (see Fig. 3.2). In such a resonant bias configuration, the total current flowing through the dot, I, can be written using the Landauer formula (see Eq. (I.20)),

$$I_{\mathrm{H,C}} = \mp \frac{2e}{h} \int f_{\mathrm{H,C}}\left(\varepsilon\right) \tau\left(\varepsilon\right) d\varepsilon,$$

where the subscript denotes which Fermi-Dirac distribution is contributing to the current. In linear response, thermocurrent can be written,

$$I_{\mathrm{th}} = \Delta T \frac{\partial I}{\partial T}.$$

When I_{th} is evaluated near a resonant bias condition,

$$I_{\text{th}}|_{\text{H,C}} = \Delta T_{\text{H,C}} \frac{\partial I_{\text{H,C}}}{\partial T_{\text{H,C}}} = \Delta T_{\text{H,C}} \frac{2e}{h} \int H_{\text{H,C}}\left(\varepsilon\right)\tau\left(\varepsilon\right)d\varepsilon, \qquad \text{(III.2)}$$

where

$$H_{\text{H,C}} = \mp \frac{\partial f_{\text{H,C}}}{\partial T_{\text{H,C}}} = \mp \frac{\partial f_{\text{H,C}}}{\partial \xi_{\text{H,C}}}\frac{\partial \xi_{\text{H,C}}}{\partial T_{\text{H,C}}} = \pm \frac{\partial f_{\text{H,C}}}{\partial \xi_{\text{H,C}}}\frac{\xi_{\text{H,C}}}{T_{\text{H,C}}}. \qquad \text{(III.3)}$$

Similarly, the second differential conductance can be written,

$$G_2|_{\text{H,C}} = \left.\frac{\partial^2 I}{\partial V^2}\right|_{\text{H,C}} = -\frac{e^2}{8k^2}\frac{1}{T_{\text{H,C}}}\frac{2e}{h}\int_{-\infty}^{\infty} K_{\text{H,C}}\left(\varepsilon\right)\tau\left(\varepsilon\right)d\varepsilon, \qquad \text{(III.4)}$$

where

$$
\begin{aligned}
K_{\text{H,C}} &= \pm - \frac{e^2}{8k^2}T_{\text{H,C}}\frac{\partial^2 f_{\text{H,C}}}{\partial V^2} \\
&= \pm - \frac{e^2}{8k^2}T_{\text{H,C}}\frac{\partial \xi_{\text{H,C}}}{\partial V}\frac{\partial}{\partial \xi_{\text{H,C}}}\frac{\partial \xi_{\text{H,C}}}{\partial V}\frac{\partial f_{\text{H,C}}}{\partial \xi_{\text{H,C}}} \\
&= \pm - \frac{e^2}{8k^2}T_{\text{H,C}}\left(\frac{\partial \xi_{\text{H,C}}}{\partial V}\right)^2\frac{\partial^2 f_{\text{H,C}}}{\partial \xi_{\text{H,C}}^2} \\
&= \pm - \frac{e^2}{8k^2}T_{\text{H,C}}\left(\frac{e}{2kT_{\text{H,C}}}\right)^2\frac{\partial^2 f_{\text{H,C}}}{\partial \xi_{\text{H,C}}^2} \\
&= \pm \frac{2}{T_{\text{H,C}}}\frac{\partial^2 f_{\text{H,C}}}{\partial \xi_{\text{H,C}}^2}. \qquad \text{(III.5)}
\end{aligned}
$$

Note that the derivative $\partial f_{H,C}/\partial \xi_{H,C}$ can be written.

$$
\begin{aligned}
\frac{\partial f_{H,C}}{\partial \xi_{H,C}} &= -\frac{e^{\xi_{H,C}}}{(1+e^{\xi_{H,C}})^2} \\
&= \frac{1-1-e^{\xi_{H,C}}}{(1+e^{\xi_{H,C}})^2} \\
&= \frac{1}{(1+e^{\xi_{H,C}})^2} - \frac{1+e^{\xi_{H,C}}}{(1+e^{\xi_{H,C}})^2} \\
&= \frac{1}{(1+e^{\xi_{H,C}})^2} - \frac{1}{1+e^{\xi_{H,C}}} \\
&= f_{H,C}^2 - f_{H,C}.
\end{aligned} \tag{III.6}
$$

Therefore, the second derivative $\partial^2 f_{H,C}/\partial \xi_{H,C}^2$ is

$$
\begin{aligned}
\frac{\partial^2 f_{H,C}}{\partial \xi_{H,C}^2} &= \frac{\partial}{\partial \xi_{H,C}}\left(f_{H,C}^2 - f_{H,C}\right) \\
&= 2 f_{H,C}\frac{\partial f_{H,C}}{\partial \xi_{H,C}} - \frac{\partial f_{H,C}}{\partial \xi_{H,C}} \\
&= \frac{\partial f_{H,C}}{\partial \xi_{H,C}}\left(2 f_{H,C} - 1\right).
\end{aligned} \tag{III.7}
$$

Inserting this second derivative into the expression for $K_{H,C}$ in Eq. (III.5) provides,

$$
\begin{aligned}
K_{H,C} &= \pm\frac{2}{T_{H,C}}\frac{\partial f_{H,C}}{\partial \xi_{H,C}}\left(2 f_{H,C} - 1\right) \\
&= \pm 2\frac{\partial f_{H,C}}{\partial \xi_{H,C}}\frac{2 f_{H,C} - 1}{T_{H,C}}.
\end{aligned} \tag{III.8}
$$

The similarity between $H_{H,C}$ and $K_{H,C}$ as written in Eqs. (III.3) and (III.8), and shown

in Fig. 3.3, is the main motivation for choosing the second differential conductance, $G_2 = \partial^2 I/\partial V^2$, rather than the orthodox differential conductance, $G = \partial I/\partial V$. This fortunate compatibility between $H_{\text{H,C}}$ and $K_{\text{H,C}}$ provides the mathematical foundation on which this thermometry technique rests. In fact, the ratio

$$R \equiv \frac{I_{\text{th}}}{G_2} \qquad \text{(III.9)}$$

is enough to determine ΔT_{H} and ΔT_{C} when the bias voltage and gate energy have been tuned so that the dot's energy resonance overlaps with either the source or drain Fermi-Dirac distribution, respectively.

Fig. 3.3 shows $H_{\text{H,C}}$ and $K_{\text{H,C}}$ and three transmission functions with different values of Γ plotted as a function of ε for fixed μ and V. Measurements of I_{th} and G_2 reflect the convolution of $H_{\text{H,C}}$ and $K_{\text{H,C}}$ with $\tau(\varepsilon)$, respectively, and therefore the position and width of the transmission function influences to what degree I_{th} and G_2 differ. Unaware of an analytical solution to either Eq. (III.2) or Eq. (III.4), analytic progress continues by considering two separate regimes using two separate approximations.

In the narrow-resonance regime, $\tau(\varepsilon)$ is very sharp, that is, $\Gamma \ll kT$, and the contributing gray area in Fig. 3.3 is made very small; so small, in fact, that $R_{\text{H,C}}$ depends primarily on the ratio of $H_{\text{H,C}}$ and $K_{\text{H,C}}$ at ε_0. A transmission function in the form of a Dirac delta function provides precisely this value and leads to an
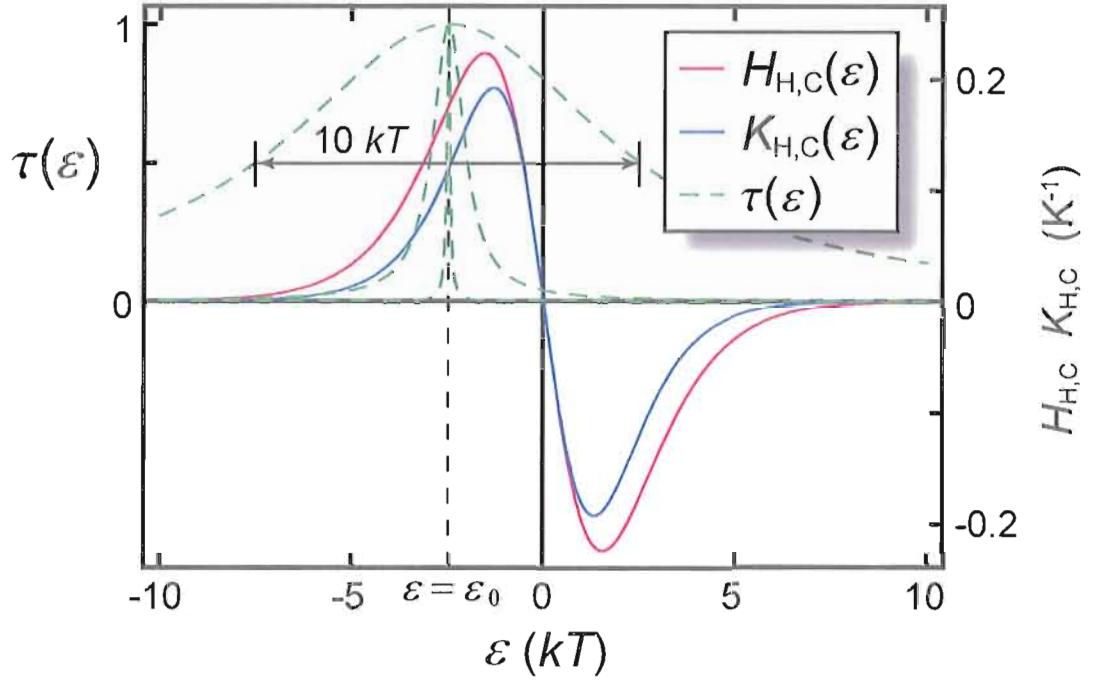
Figure 3.3. The functions $H_{\mathrm{H,C}}(\varepsilon)$, $K_{\mathrm{H,C}}(\varepsilon)$, and $\tau(\varepsilon)$—Eqs. (III.3), (III.8), and (II.7)—plotted as a function of ε for fixed μ and V. The three Lorentzian transmission functions, $\tau(\varepsilon)$, are centered at ε_0 and have widths $\Gamma = 10\ kT$, $1\ kT$, and $0.1\ kT$. The disparity between $H_{\mathrm{II,C}}$ and $K_{\mathrm{II,C}}$, illustrated by the gray region, evolves into disparity between $I_{\mathrm{th}}(\varepsilon)$ and $G_2(\varepsilon)$ during integration with $\tau(\varepsilon)$. When $\Gamma \ll kT$ or when $\Gamma \gg kT$, this disparity can be described analytically—allowing one to determine $\Delta T_{\mathrm{H,C}}$. The intermediate regime, where $\Gamma \approx kT$, escapes analytical description, and $\Delta T_{\mathrm{H,C}}$ cannot be determined.

analytical limit of $R_{\mathrm{H,C}}$ near $V_{\mathrm{H,C}}^0$. With this analytical expression in hand, one can

extract $\Delta T_{\mathrm{H,C}}$ using standard statistical data analysis, as will be described in the

next section The hallmark of this regime is the existence of very sharp transmission

resonances that produce pronounced R curves, which look like the one shown in

Fig. 3.4a.

In the broad-resonance regime, $\tau(\varepsilon)$ is very wide, that is, $\Gamma \gg kT$, and most of
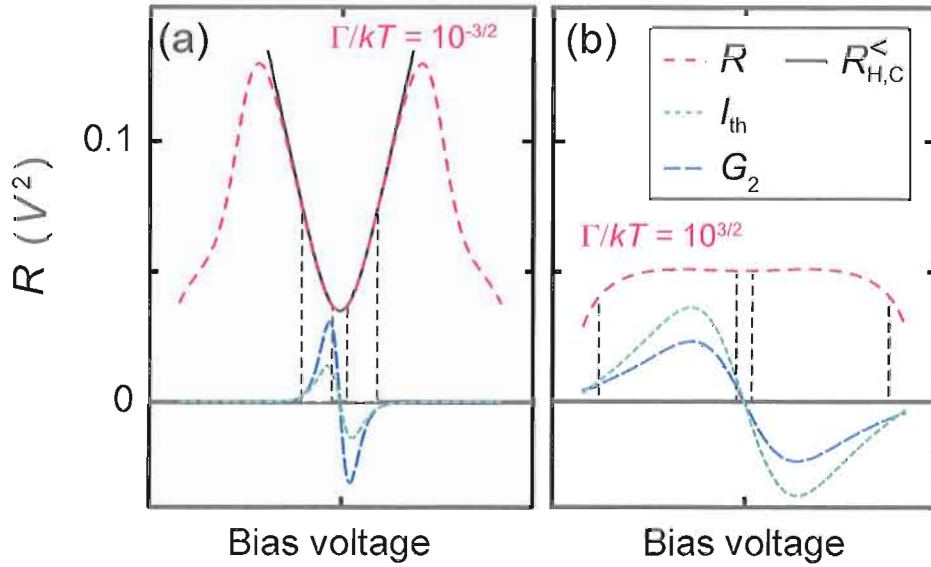
Figure 3.4. Numerical calculations of I_{th}, G_2, and their ratio, $R = I_{th}/G_2$, as a function of bias voltage. For these calculations, $T = 1$ K, $\Delta T/T = 2/5$, and two values of Γ are chosen so that Γ/kT is $10^{3/2}$ and $10^{-3/2}$. The I_{th} and G_2 curves and their corresponding shaded regions show the bias voltage range from which experimental data would be collected. (a) The solid line $R_{H,C}^<$, Eq. (III.12), is the analytical approximation of R and, in the $\Gamma \ll kT$ regime, agrees with the numerically calculated R. (b) R in the $\Gamma \gg kT$ regime becomes very flat when I_{th} and G_2 peak. When $\Gamma \approx kT$, R forms neither a suitable trough nor a suitable plateau limiting the accuracy to which $\Delta T_{H,D}$ can be determined.

the gray area in Fig. 3.3 contributes nearly equally to R when ε_0 is near the center

of $H_{H,C}$ and $K_{H,C}$. Once this contribution has been calibrated numerically, values of

$R_{H,C}$ in a voltage range near $V_{H,C}^0$ predict the values of ΔT directly. In this regime,

R forms plateaus that are not sharply delineated and suitable for data averaging over

a range of bias voltages; such a plateau is shown in Fig. 3.4b. The two regimes of

operation, their limits, and the intermediate regime which separates them will be

discussed in detail.

In this derivation, an assumption was made that the voltage drops equally across the two barriers that define the quantum dot and, therefore, implicitly assumed that the two barriers are identical and couple to the electron gas equally. In practice, the barriers are not exactly identical. In the case of asymmetric coupling, the resonances at V_H^0 and V_C^0 will have different heights and widths, even under isothermal conditions, owing to the different tunneling rates. However, quantum-dot thermometry is not adversely affected by asymmetric coupling because the ratio R is calculated *locally* at each resonance and is therefore immune to differences *between* resonances. Evidence of asymmetric coupling in the data is best demonstrated by asymmetry in resonance position. If coupling is equal, then $|V_H^0| = |V_C^0|$, but $|V_H^0| \neq |V_C^0|$ when the coupling is not equal. Therefore, the primary effect of coupling asymmetry is skewed Coulomb-blockade diamonds.

<center>The Narrow-resonance Regime ($\Gamma \ll kT$)</center>

In the following, $\tau(\varepsilon)$ is assumed to have such a narrow energy width that it can be approximated as a Dirac delta function centered at ε_0, that is, $\tau(\varepsilon) = A\delta(\varepsilon - \varepsilon_0)$, where A is an unknown energy amplitude. Inserting this expression for $\tau(\varepsilon)$ into Eq. (III.2) and integrating gives

$$I_{\text{th}}|_{\text{H,C}} = \Delta T_{\text{H,C}} A \frac{2e}{h} H_{\text{H,C}}\bigg|_{\varepsilon=\varepsilon_0}. \tag{III.10}$$

This alone is not sufficient to determine $\Delta T_{\text{H,C}}$ because A is unknown. This is

why a second measurement—the second differential conductance, G_2—is necessary; it provides an energy scale (eV) and depends on $\tau(\varepsilon)$ in the same way as I_{th}. Under the same bias conditions as before and after inserting $\tau(\varepsilon) = A\delta(\varepsilon - \varepsilon_0)$ into Eq. (III.4) and integrating, G_2 becomes

$$G_2|_{\text{H,C}} = -\frac{e^2}{8k^2}\frac{1}{T_{\text{H,C}}}A\frac{2e}{h}K_{\text{H,C}}\bigg|_{\varepsilon=\varepsilon_0}. \tag{III.11}$$

In this narrow-resonance regime, where $\Gamma \ll kT$, R is approximated by $R^<$, the ratio of Eq. (III.10) to Eq. (III.11)

$$R_{\text{H,C}} \approx R^<_{\text{H,C}} = -\Delta T_{\text{H,C}}\frac{8k^2}{e^2}T_{\text{H,C}}\frac{H_{\text{H,C}}}{K_{\text{H,C}}}\bigg|_{\varepsilon=\varepsilon_0} = -\Delta T_{\text{H,C}}\frac{4k^2}{e^2}T_{\text{H,C}}\frac{\xi_{\text{H,C}}}{2f_{\text{H,C}}-1}\bigg|_{\varepsilon=\varepsilon_0},$$

and A drops out. By expressing $R^<_{\text{H,C}}$ in bias voltage $V^0_{\text{H,C}} = \mp 2(\varepsilon_0 - \mu)/e$, rather than energy ε_0, $R^<_{\text{H,C}}$ becomes

$$R^<_{\text{H,C}} = \Delta T_{\text{H,C}}\frac{2k}{e}\left(V - V^0_{\text{H,C}}\right)\coth\left(\frac{e}{4k}\frac{V - V^0_{\text{H,C}}}{\Delta T_{\text{H,C}} + T}\right). \tag{III.12}$$

Note that $R^<_{\text{H,C}}$ is only valid when either the source or drain electrochemical potential is near a resonance of the quantum dot, but not both, and when $\Gamma \ll kT$.

Fig. 3.4a shows a numerical simulation of the measured ratio $R = I_{\text{th}}/G_2$ and its analytical approximation, $R^<_{\text{H,C}}$, as a function of bias voltage with $\Gamma/kT = 10^{-3/2}$. In this $\Gamma \ll kT$ limit, the analytic $R^<_{\text{H,C}}$ agrees with R. In fact, the region over which

they agree includes bias voltages where the values of I_{th} and G_2 are the largest (the gray shaded areas in Fig. 3.4a) and, therefore, where R should have the best signal-to-noise ratio in experiments. This demonstrates that experimental values of $\Delta T_{\text{H,C}}$ can be determined by comparing $R^<_{\text{H,C}}$ to raw R data. See Section III.4 for details on data analysis.

$R^<_{\text{H,C}}$ matches the top three R curves in Fig. 3.5 corresponding to Γ/kT values of 10^{-1}, $10^{-1/2}$, and 1. It is surprising that when $\Gamma/kT = 1$, which is outside of the intended range of this regime, $R^<_{\text{H,C}}$ reproduces R well. The $\Gamma/kT > 1$ curves in Fig. 3.5 are not well-predicted by Eq. (III.12), that is, $R^<_{\text{H,C}}$ and R disagree (not shown), indicating that this method is not useful for determining $\Delta T_{\text{H,C}}$ when $\Gamma > kT$.

The broad-resonance regime ($\Gamma \gg kT$)

The essence of this broad-resonance method is that $H_{\text{H,C}}$ and $K_{\text{H,C}}$ (see Eqs. (III.3) and (III.8)) are very similar. It is mathematically true that

$$H_{\text{H,C}} = \frac{\partial f_{\text{H,C}}}{\partial \xi_{\text{H,C}}} \frac{\xi_{\text{H,C}}}{T_{\text{H,C}}} \approx -2 \frac{\partial f_{\text{H,C}}}{\partial \xi_{\text{H,C}}} \frac{2f_{\text{H,C}} - 1}{T_{\text{H,C}}} = K_{\text{H,C}}. \qquad \text{(III.13)}$$

This approximation is valid for all $\xi_{\text{H,C}}$, because $2f_{\text{H,C}} - 1$ limits to $-\xi_{\text{H,C}}/2$ when $\xi_{\text{H,C}}$ is small, and $\partial f_{\text{H,C}}/\partial \xi_{\text{H,C}}$ goes to zero in all other cases. Therefore it is possible
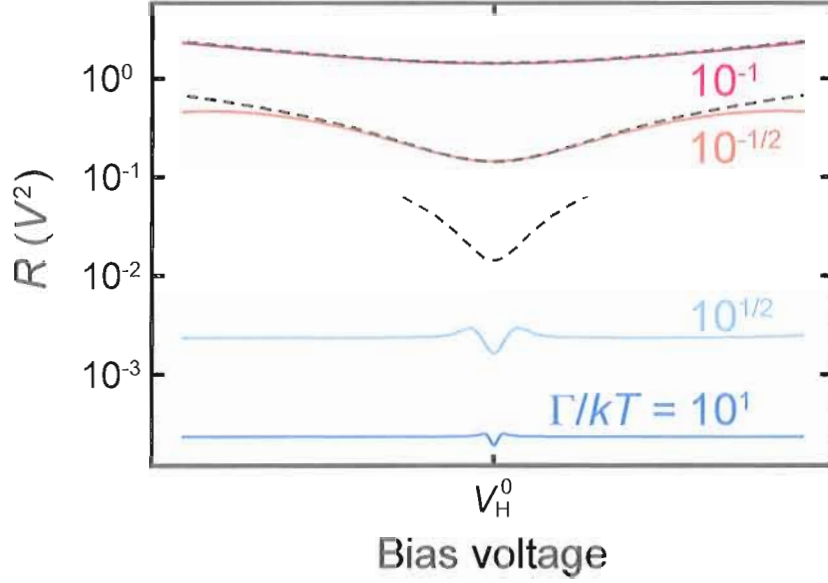
Figure 3.5. A linear-log plot of R versus bias voltage numerically calculated for $\Gamma = 43$ μeV $\cong 500$ mK and $\Delta T_{\mathrm{H}}/T = 3/5$. Each curve was calculated at a different temperature to produce the five labelled values of Γ/kT. The dashed lines following the top three R curves are analytical estimates of R predicted by Eq. (III.12). Even when $\Gamma/kT = 1$, the two agree in a region which is large enough to extract $\Delta T_{\mathrm{H,C}}$. But as Γ/kT increases, R begins to form plateaus, and the narrow-resonance regime breaks down when $\Gamma/kT \geq 10^{1/2}$.

to write

$$I_{\mathrm{th}}|_{\mathrm{H,C}} = \Delta T_{\mathrm{H,C}} \frac{2e}{h} \int_{-\infty}^{\infty} H_{\mathrm{H,C}} \tau\left(\varepsilon\right) d\varepsilon \approx \Delta T_{\mathrm{H,C}} \frac{2e}{h} \int_{-\infty}^{\infty} K_{\mathrm{H,C}} \tau\left(\varepsilon\right) d\varepsilon \qquad \text{(III.14)}$$

by substituting $K_{\mathrm{H,C}} \approx H_{\mathrm{H,C}}$ in Eq. (III.2). In this broad-resonance regime, where $\Gamma \gg kT$, Eq. (III.14) is divided by Eq. (III.4) and R is approximated as

$$R \approx \Delta T_{\mathrm{H,C}} \frac{T_{\mathrm{H,C}}}{\Lambda_{\mathrm{H,C}}} \frac{4k^2}{e^2}, \qquad \text{(III.15)}$$

where $\Lambda_{H,C}$ is introduced as a unitless scaling factor of order 1, which must be calculated numerically and corrects for the error associated with equating $H_{H,C}$ and $K_{H,C}$. (Calculated values of $\Lambda_{H,C}$ appear in Fig. 3.9.) Solving for $\Delta T_{H,C}$ in Eq. (III.15) yields

$$\Delta T_{H,C} = \frac{1}{2}\sqrt{T^2 + \Lambda_{H,C}\frac{e^2}{k^2}R} - \frac{T}{2}, \qquad \text{(III.16)}$$

which shows that ΔT_H and ΔT_C can be obtained from measured values of R and knowledge of T and $\Lambda_{H,C}$. Note that this equation is only valid at bias and gate voltages that provide an overlap between $\tau(\varepsilon)$ and either f_H or f_C and when $\Gamma \gg kT$. Fig. 3.4b shows a numerical simulation of R with $\Gamma/kT = 10^{3/2}$. At bias voltages near $V_{H,C}^0$, R forms a plateau as it limits to the value predicted by Eq. (III.16). This plateau is suitable for extracting $\Delta T_{H,C}$ experimentally by feeding into Eq. (III.16) data points from areas where R forms plateaus and averaging the resulting values.

Fig. 3.6 shows three ΔT_H curves calculated by feeding numerically simulated values of R into Eq. (III.16). The two bottom curves, corresponding to $\Gamma/kT = 10^{1/2}$ and 10, form plateaus that limit to $\Delta T_H = 3/5T$ thereby predicting the correct values. Although the tails of the $\Gamma/kT = 10^0$ curve do indeed lie near the correct value of ΔT_H, the center of this curve—where experimental data would be collected—does not form a suitable plateau necessary for extracting ΔT_H.

Fig. 3.7 is a contour plot of temperature rise predicted by Eq. (III.16) as a function of bias voltage and gate voltage. Although this simulation was performed in the broad-resonance regime, the figure demonstrates how ΔT_H and ΔT_C can be measured
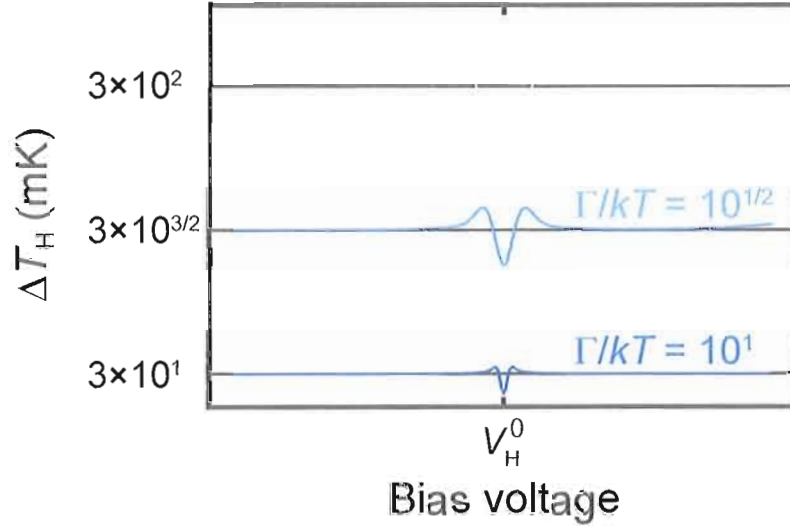
Figure 3.6. A linear-log plot of ΔT_{H} versus bias voltage calculated using Eq. (III.16) and the bottom three R curves shown in Fig. 3.5. Each curve was numerically calculated with $\Gamma = 43\ \mu$eV $\cong 500$ mK, $\Delta T_{\mathrm{H}}/T = 3/5$, and different T in order to produce the labeled values of Γ/kT. Because $\Delta T_{\mathrm{H}}/T$ is constant, ΔT_{H} in these calculations is 3×10^{2} mK, $3 \times 10^{3/2}$ mK, and 3×10^{1} mK. See Fig. 3.9 for the values of Λ_{H} used in these calculations. In practice, the broad-resonance regime can predict ΔT only when R forms plateaus *near* $V_{\mathrm{H,C}}^{0}$, where data is collected, requiring $\Gamma/kT \geq 10$.

separately by performing measurements in different regions of the bias-voltage–gate-voltage plane, regardless of the regime of operation.

Fig. 3.8 is a plot of numerically simulated $\Delta T_{\mathrm{H,C}}$ as a function of ΔT_{SC} demonstrating the effectiveness of the broad-resonance regime. The simulated data points agree well with the true values of $\Delta T_{\mathrm{H,C}}$ up to nearly $\Delta T_{\mathrm{H}}/T = 10$. This is a surprisingly extreme limit considering the method assumes linear response. In practice, experiments [79] will always be well below this limit ($\Delta T_{\mathrm{H}}/T \lesssim 2$) insuring that the method produces accurate values of $\Delta T_{\mathrm{H,C}}$. In addition, Fig. 3.8 shows that
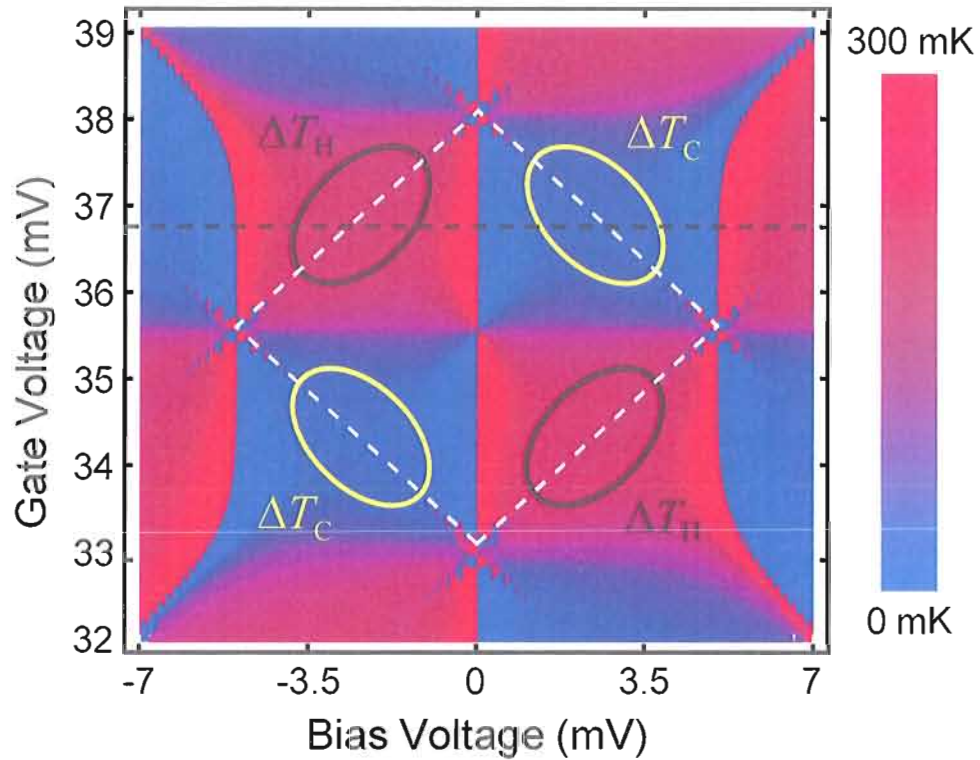
Figure 3.7. A contour plot of temperature rise as a function of bias and gate voltage for a single Coulomb-blockade diamond (white dashed line) *as predicted* by quantum-dot thermometry (see Eq. (III.16)). In the regions indicated, either the source or drain electrochemical potential, but not both, is within a few kT of a resonant energy of the dot. The theoretical assumptions are fulfilled in these regions, and the simulation produces temperature plateaus predicting the correct $\Delta T_{H,C}$. In other regions, the reported temperature is meaningless. The horizontal dashed line indicates a candidate gate voltage for quantum-dot thermometry where both ΔT_H and ΔT_C can be found by sweeping bias voltage. In this model, $\Gamma = 0.5$ meV, the energy resonances are spaced by 5 meV, $T_{SC} = 300$ mK, $T_H = 0.8 T_{SC} = 240$ mK, $T_C = 0.2 T_{SC} = 60$ mK, $T = 230$ mK, and $\Lambda = 0.304$.

choosing a constant value of Λ over a large range of Γ results in an error of less than 1% demonstrating that Λ is insensitive to Γ.
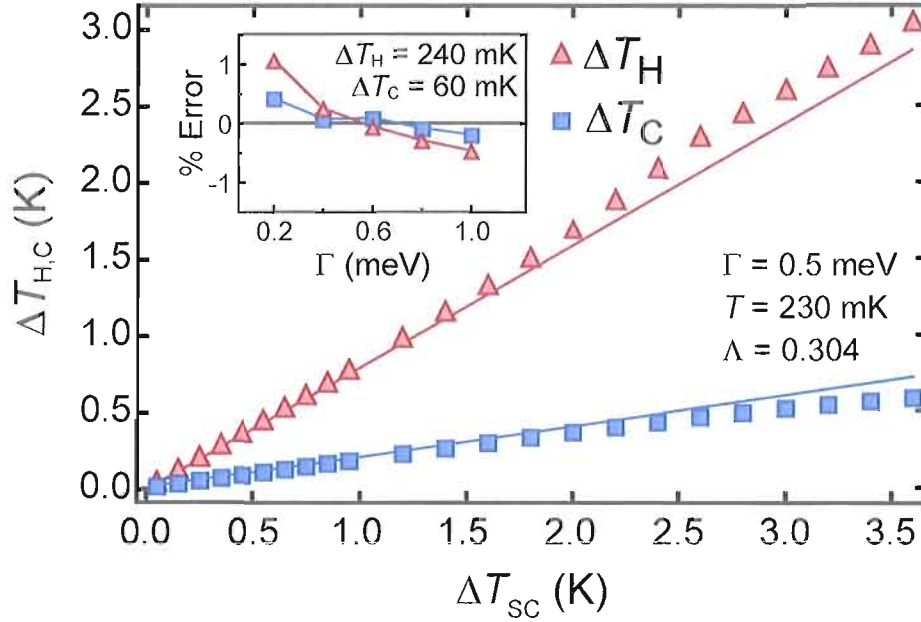
Figure 3.8. Numerically simulated temperature rises, $\Delta T_{H,C}$, as a function of the source contact temperature, ΔT_{SC}, determined using Eq. (III.16). The calculated values agree well with the true values (solid lines) up to nearly $\Delta T_H/T = 10$, which is well beyond the intended range of use. Here $\Gamma = 0.5$ meV, $T = 230$ mK, and $\Lambda = 0.304$. Inset: The percent error of the technique as a function of Γ, which is within 1% over an order of magnitude in Γ. Here $\Delta T_H = 240$ mK, $\Delta T_C = 60$ mK, and $\Lambda = 0.304$. The inset and main panel demonstrate that a single value of Λ can be used over a large range of $\Delta T_{H,C}$ and Γ.

The numerical search for $\Lambda_{H,C}$ is done by assessing the quality of this approximation at a range of bias voltages near $V_{H,C}^0$. This process is done numerically, and the result of these calculations are shown in Fig. 3.9. This calculation is the main result of this proceeding and will prove useful for anyone who employs quantum-dot thermometry in the regime where $\Gamma \gg kT$.

T (K)	Λ_H	Λ_C
0.05	0.304	0.304
0.158	0.304	0.304
0.5	0.305	0.305
1.58	0.310	0.343
5	0.360	0.420
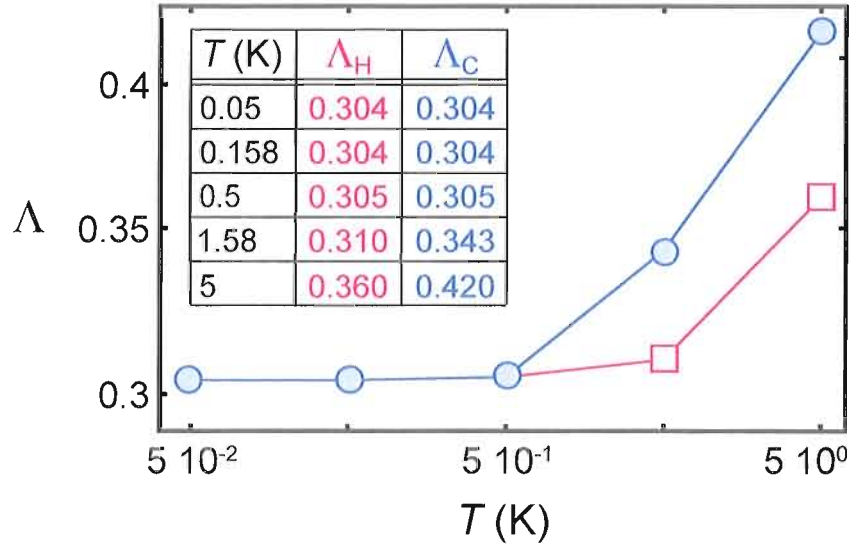
Figure 3.9. A log-linear plot of Λ_H (squares) and Λ_C (circles) as a function of background temperature T. These $\Lambda_{H,C}$ values were calculated using $\Gamma/kT = 10^{3/2}$, $\Delta T_H/T = 3/5$, and $\Delta T_C/T = 2/5$ by searching for values of $\Lambda_{H,C}$ which predicted $\Delta T_{H,C}$ most accurately. Λ_H and Λ_C are identical and nearly constant over the first order of magnitude in T, but take on separate values at higher temperatures. However, the broad-resonance regime is most likely not applicable in the high temperature range because the $\Gamma \gg kT$ condition is almost certainly not met. For example, at $T = 5$ K, in order to guarantee that $\Gamma \gg kT$, Γ must be on the order of 50 K $\cong$ 4 meV, which is an unreasonable value for the quantum dot devices we are considering. In fact, Γ will most likely be much smaller than kT at $T = 5$ K. Therefore, when the background temperature is on the order of a few Kelvin, the narrow-resonance regime ($\Gamma \ll kT$) will most likely be the preferred thermometry technique.

The intermediate regime ($\Gamma \approx kT$)

In the intermediate regime, where $1 < \Gamma/kT < 10$, R varies dramatically with bias voltage, but this behavior cannot be predicted analytically, making it less suitable than the other regimes for predicting accurate values of $\Delta T_{H,C}$. If raw R data produces plateaus at a given temperature T, then all that remains is choosing the appropriate $\Lambda_{H,C}$ from Fig. 3.9 and calculating $\Delta T_{H,C}$ via Eq. (III.16). On the other hand, if raw R

data produces troughs *and* the relationship between Γ and kT is unknown, then one cannot know if the troughs will predict the correct $\Delta T_{H,C}$ because, in the intermediate regime, $R_{H,C}^<$ predicts curves which are inconsistent with R. In such a situation, this puzzle could be solved by placing an upper bound on Γ, for example, by fitting zero-bias differential conductance peaks [61]. However, $R_{H,C}^<$ is always within an order of magnitude of R in the regions where I_{th} and G_2 are measurable. Furthermore, though plateaus in R do not predict the correct values $\Delta T_{H,C}$ in the intermediate regime, calculations made via Eq. (III.16) straddle the correct $\Delta T_{H,C}$ and deviate from this value by much less than an order of magnitude. (See, for example, the top curve of Fig. 3.6.) Therefore, *any* experimental R data is capable of predicting $\Delta T_{H,C}$ to well within an order of magnitude.

Conclusion

The two regimes of operation of quantum-dot thermometry, the narrow-resonance and wide-resonance regimes, provide thermometry in the temperature range where low-temperature quantum-dot experiments are typically performed. The narrow-resonance regime was established theoretically by assuming that $\Gamma \ll kT$, but in practice the regime is very robust and has the potential to provide temperature measurements even when Γ approaches kT, depending on the accuracy tolerance and background temperature of the experiment.

The broad-resonance regime is applicable at low temperatures (< 200 mK) where kT is much smaller than typical intrinsic transmission widths. Despite resulting

from an integration involving $\tau(\varepsilon)$ and $f_{\text{H,C}}$, $\Lambda_{\text{H,C}}$ has no Γ or $\Delta T_{\text{H,C}}$ dependence and depends only weakly on T (see Figs. 3.8 and 3.9), provided that $\Gamma \gg kT$. The wide-resonance regime can predict values of $\Delta T_{\text{H,C}}$ whenever $\Gamma/kT > 10$ and becomes more accurate as Γ/kT increases as the plateaus in R become wider. In the intermediate regime, when $1 < \Gamma/kT < 10$, neither method is suitable for finding accurate values of $\Delta T_{\text{H,C}}$. However, a combination of the two methods could estimate $\Delta T_{\text{H,C}}$ to well within an order of magnitude.

In practice, the design and realization of a quantum dot that has a suitable value of Γ/kT is straightforward. Regardless of the regime of operation, producing isolated, single energy resonances is the greatest experimental challenge for quantum-dot thermometry. This is a fundamental issue because, when many interacting electrons are confined in a 3D potential, their excitation spectrum becomes very complex (see for example Ref. [81]), and energy resonances are infrequently isolated. Therefore, quantum-dot thermometry benefits from operating the dot in the few-electron regime.

Experimental Results

The most challenging aspect of quantum-dot thermometry is finding a single gate voltage (or small gate voltage range) with ideal behavior. That is, a location where the quantum dot is not influenced by second-order effects such as excited states, phonons, and cotunneling that can lead to negative differential conductance and other abnormal behavior. Coulomb-blockade diamonds (see for example Fig. 2.9b) assist in

finding candidate locations for quantum-dot thermometry, and then high-resolution bias voltage sweeps pinpoint gate voltages providing the best data. Ideal behavior as a function of bias voltage is shown in Fig. 3.10a, which is a numerical simulation. Once a proper gate voltage has been found, data collection can proceed. During quantum-dot thermometry, currents are measured, and therefore a SRS Model SR570 low-noise current preamplifier was attached to the drain contact. In order to measure I_{th} and G_2 simualtaneously, an ac + dc bias voltage was applied to the nanowire in addition to an ac heating current, I_{H}. This input was acheived by using a voltage adder and heating circuit described in Section II.4.4. The ac bias voltage frequency was 42 Hz, and I_{H} was at a frequency of 17 Hz. An example experimental setup is shown in Fig. 2.13, and example experimental data measured as a function of bias voltage is shown in Fig. 3.10b. During experiments the differential conductance, $G = dI/dV$, is measured using a lock-in amplifier, and G_2 is found by taking a numerical derivative of G, $G_2 = d^2I/dV^2 = dG/dV$. The numerical derivative proceedure is outlined in Appendix D.

Data Analysis and Temperature Determination

Once data collection is complete, the analysis proceedure begins by determining the value of Γ relative to kT. All the quantum-dot nanowire devices used in these thermoelectric experiments have relatively wide tunnel barriers and therefore narrow
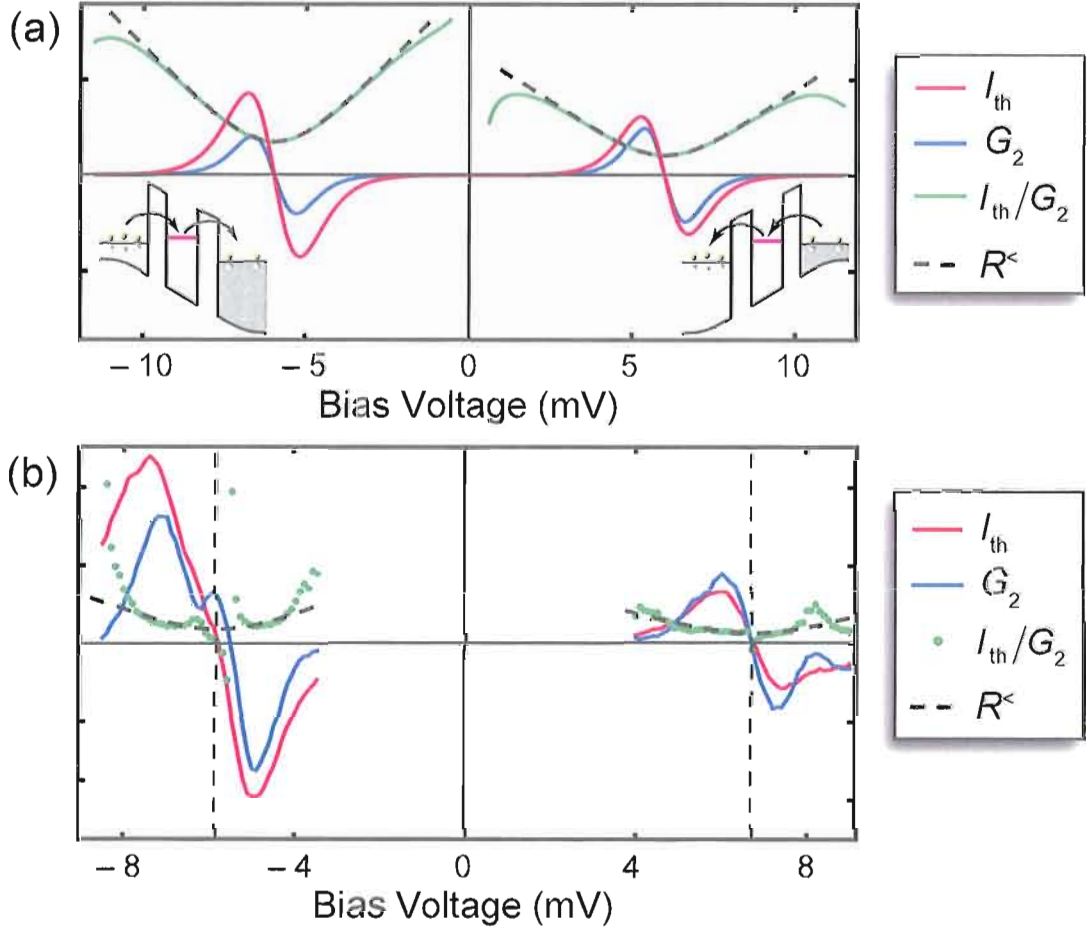
Figure 3.10. (a) Modeled thermocurrent, I_{th}, (red), second differential conductance, G_2, (blue), and their ratio I_{th}/G_2 (green) as a function of bias voltage calculated using $\Gamma = 10~\mu\text{eV} \ll kT = 190~\mu\text{eV} = 2.2$ K. Therefore, the system is in the narrow-resonance regime. Near a quantum-dot resonance (see insets), both I_{th} and G_2 have a zero-crossing, and here the analytical expression $R_{\text{H,C}}^{<}$ (dashed lines, see Eq. (III.12)) agrees with the ratio I_{th}/G_2. *Insets*: When biased negatively, heated electrons flow through the dot, while cold electrons flow when positively biased. (b) Experimental data of I_{th} (red), G_2 (blue), their ratio (green dots), and the fit, $R_{\text{H,C}}^{<}$, (dashed) as a function of bias voltage. For these data, $I_{\text{H}} = 150~\mu\text{A}$ and $T = 2.2$ K, and the fitting function $R_{\text{H,C}}^{<}$ provided the experimental values $\Delta T_{\text{s}} = 230$ mK and $\Delta T_{\text{d}} = 160$ mK.

transmission resonances providing $\Gamma \ll kT$. Data analysis is carried out under the narrow-resonance regime (see Section III.2.1).

Near the zero-crossings of I_{th} and G_2 data, the ratio $R = I_{th}/G_2$ suffers from numerical division-by-zero errors that create data outliers making statistical analyses difficult. Rather than manually excluding these data points, the data can be binned into a histogram. In this way, outliers automatically fall left and right of the histogram center but do not significanlty affect the mean.
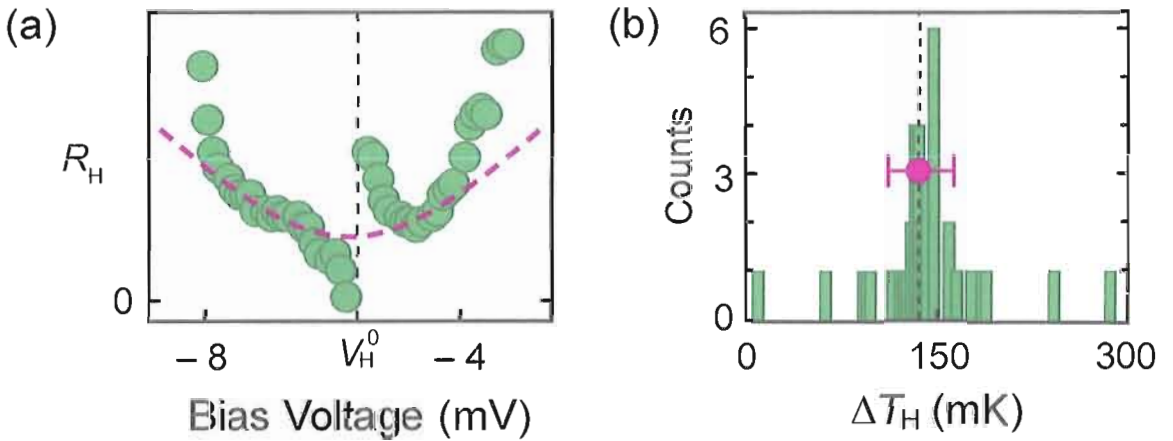
Figure 3.11. (a) Raw R_H data measured as a function of bias voltage at $I_H = 115$ μA and $T = 2.94$ K. The center of the data, $V_H^0 = -5.75$, is where I_{th} and G_2 are zero. The (violet) dashed line is Eq. (III.16) with the parameters V_H^0 and the mean ΔT_H from (b) a histogram of ΔT_H data obtained by binning ΔT_H values found by numerically solving Eq. (III.16) using the data shown in (a). The dot indicates the mean of the data, $\Delta T_H = 140$ mK, and the error bars encompass 67% of the histogram data. In general, the error bars are not symmetric about the mean. Data outliers in (a) appear in the tails of the histogram with little effect on the mean ΔT_H.

In order to bin the data into a temperature histogram, each raw $R_{H,C}$ data point from a given bias voltage sweep along with the zero-crossing bias voltage values, $V_{H,C}^0$,

are fed into Eq. (III.16). This transcendental function is then solved numerically to find the cooresponding $\Delta T_{H,C}$. For example, a series of ΔT_H values are plotted as a function of bias voltage in Fig. 3.11a. The $\Delta T_{H,C}$ are then binned to form a histogram. The histogram is then used to calculated the mean $\Delta T_{H,C}$ of the data set. Error bars are placed on the mean by finding the temperture range neccessary to encompass 67% of the data. The error bars are typically asymmetric about the mean. An example histogram is shown in Fig. 3.11b. Once the mean $\Delta T_{H,C}$ has been found, its value can be used to plot Eq. (III.16) on top of the original $R_{H,C}$ data. The resulting functional fit should agree well with the $R_{H,C}$ data providing insurance that the chosen mean $\Delta T_{H,C}$ is resonable. Such a fit is included in Fig. 3.11a. Once a series of data has been collected and analyzed, it can be plotted as a function of heating current as shown in Fig. 3.12.

Quantum-dot thermometry is an excellent experimental tool for measuring 1D thermoelectric effects as well as exploring thermal phenomena in mesoscopic systems. For example, quantum-dot thermometry has been used to investigate electon-phonon interaction in nanowires [79]. Possible future applications of quantum-dot thermometry include studies on phonon drag, thermal managment, electron-phonon interaction, and of course thermoelectrics.
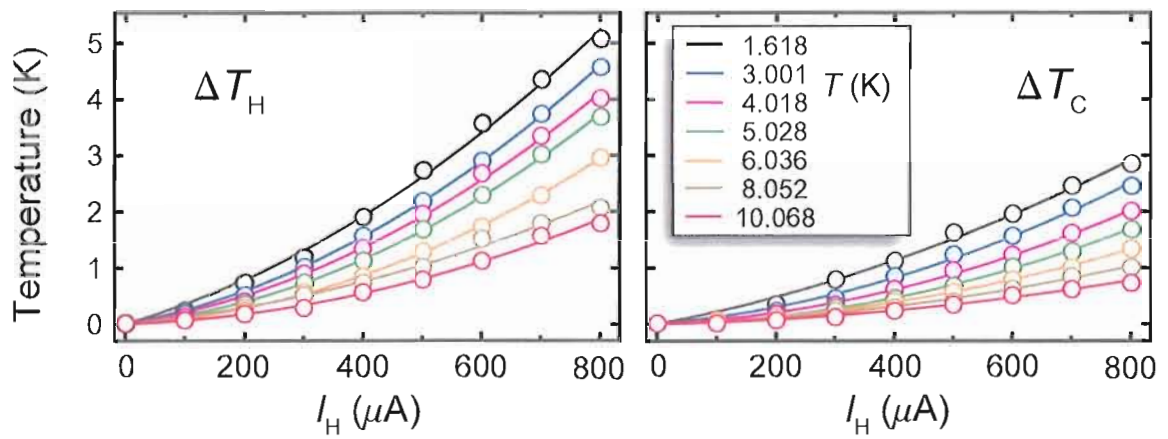
Figure 3.12. Experimental ΔT_H and ΔT_C data (circles) plotted as a function of I_H and measured using quantum-dot thermometry. The solid lines are second-order polynomial fits of the data. The data was collected at the seven base electron temperaures, T, listed in the legend.

CHAPTER IV

THERMOPOWER OF QUANTUM DOTS

The thermoelectric efficiency of a quantum dot device operating at a particular energy level depends strongly on the energetic width of the transmission resonance, Γ [10]. In particular, an infinitely small Γ provides the theoretical maximal efficiency, Carnot efficiency, as demonstrated in Section I.2.3. The efficiency of a real device, with a finite Γ, can be measured using the ZT figure of merit (see Section I.4). Because the thermopower, S, plays a key role in ZT, this chapter focuses solely on the measurement of the thermovoltage and thermopower of quantum dots and how these measurements compare to theory.

Quantum-dot Thermopower Measurements

Quantum dots operate as very sensitive energy filters, and their thermopower is quite sensitive to the energetic difference between a given energy resonance and the source and drain electrochemical potentials. This energy difference can be tuned experimentally using the gate to shift the quantum dot resonances up or down in energy. Therefore, the three experimental variables in quantum-dot thermopower measurements are 1) the background temperature, T; 2) the temperature difference,

ΔT; and 3) the gate voltage, V_G, which can be converted to an energy scale using the gate lever arm, α (see Section II.3.1).

A single-electron transistor (SET) quantum dot in the Coulomb-blockade regime demonstrates thermopower oscillations as a function of gate voltage. This phenomenon was first measured in the early '90s [71, 72, 82] in 2DEG systems. Since then, it has lead to several other interesting 2DEG-based experiments (most notably Refs. [67, 68, 83, 84]). These experiments focused primarily on the extraction of information about the systems' mobile carriers not offered by conventional transport measurements. For example, thermopower is more sensitive to the particular transport mechanism than conductance [85] and can be used to identify the dominant transport regime [84]. Along these lines, recent theoretical [86–88] and experimental [73] work has been carried out to understand better the sequential tunneling and cotunneling transport regimes in quantum dots as well as the interplay between the two regimes.

In the meantime, carbon nanotubes have also demonstrated thermopower oscillations [63, 64]. These experiments investigated the possibility of using carbon nanotubes as thermoelectric devices. Although the large thermal conductivity of carbon nanotubes [89] produces a rather small ZT [20] making them unattractive as thermoelectric devices. The research presented here is the first time Coulomb-blockade thermopower oscillations have been observed in nanowire-based quantum dots. Fig. 4.1 is an example of thermopower oscillations as a function of energy
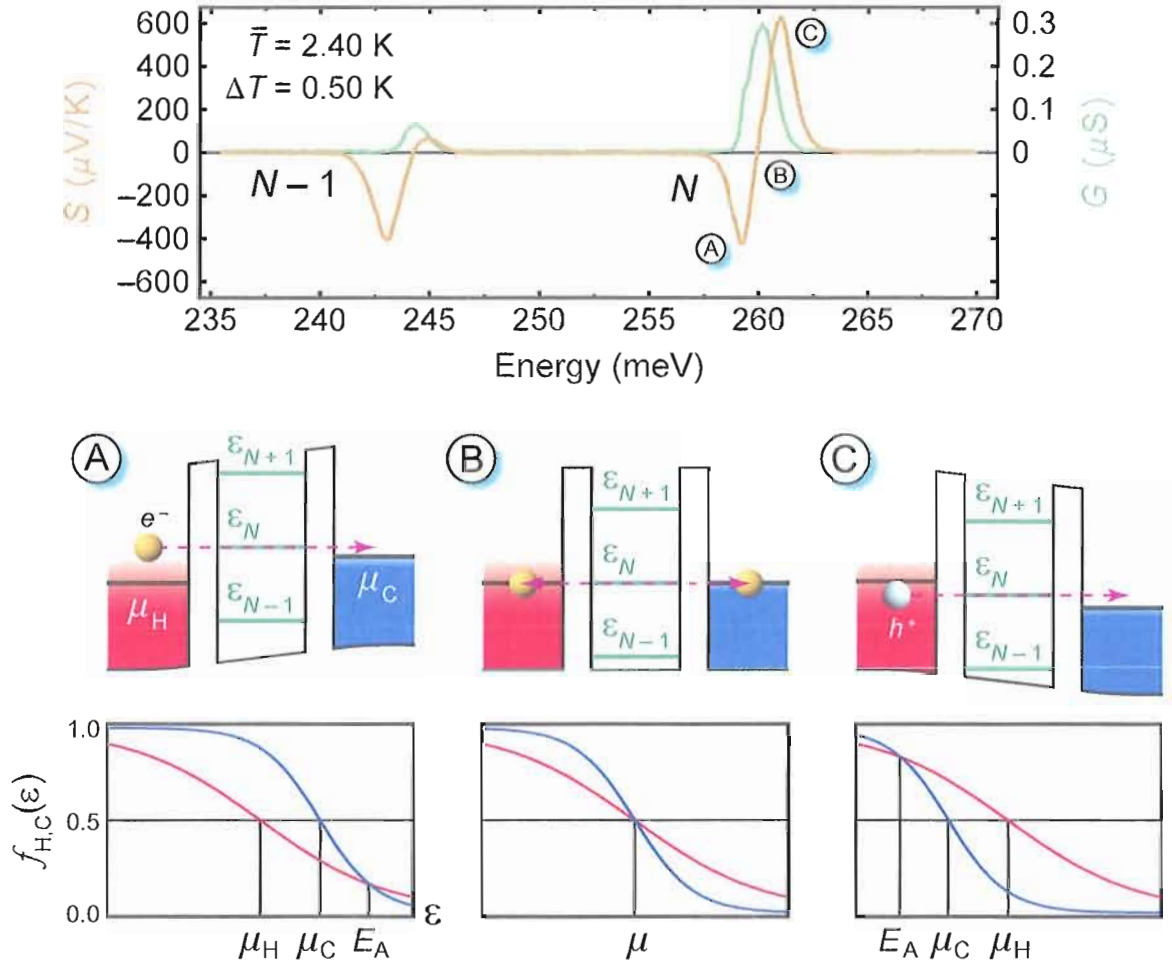
Figure 4.1. Thermopower, S, and conductance, G, as a function of quantum-dot energy. Peaks in G identify at which energy quantum-dot resonances approach the hot and cold electrochemical potentials, $\mu_{\mathrm{H,C}}$. Here, S experiences oscillations when the $(N-1)$th and Nth energy resonances lift the Coulomb blockade. Device operation at the (A), (B), and (C) working points is explained below. *Panels*: **(A)** The resonance at ε_N is tuned above μ_{H} allowing hot electrons to flow from source to drain thereby increasing μ_{C}. Here $S < 0$ and the system behaves as an n-type thermoelectric device. **(B)** Thermally excited electrons (and holes) are Coulomb blockaded and $S = 0$. This is the unique energy where $\mu_{\mathrm{H}} = \varepsilon_N = \mu_{\mathrm{C}} = E_{\mathrm{A}} = \mu$. **(C)** As ε_N is pulled below μ_{H}, excited holes flow from hot to cold decreasing μ_{C} and making $S > 0$. This is p-type thermoelectric behavior. At *all* working points, equilibrium is established when the hot and cold Fermi-Dirac distributions, $f_{\mathrm{H,C}}(\varepsilon)$, cross at E_{A}. This equilibrium is facilitated by transport through the Nth quantum dot resonance so that $E_{\mathrm{A}} = \varepsilon_N$.

(controlled experimentally using the gate). See the figure text for a qualitative explanation of thermopower oscillations in quantum dots.
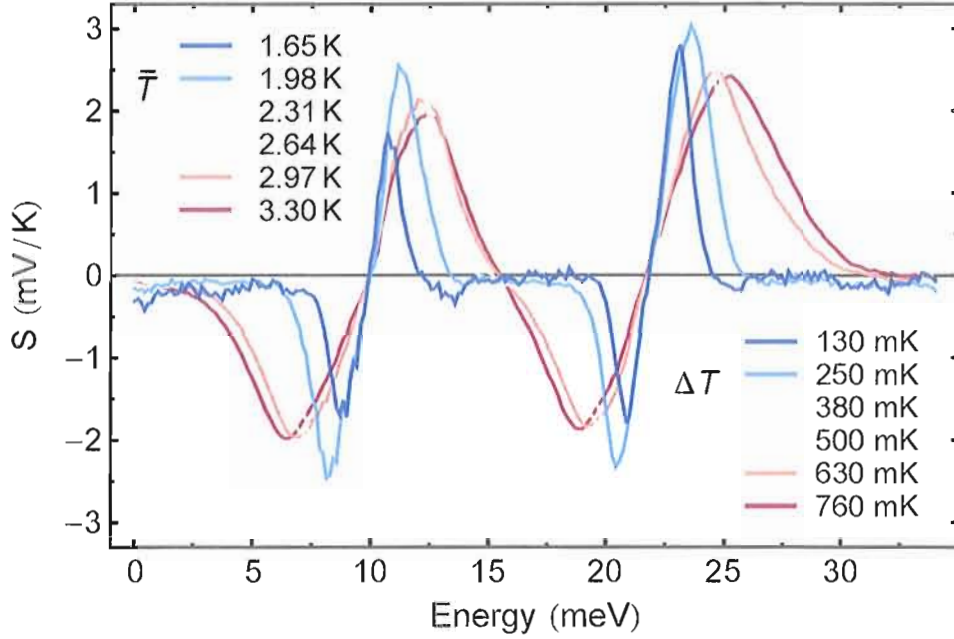
Figure 4.2. Thermopower, S, as a function of energy at the indicated temperature differences, $\Delta T = T_H - T_C$, and average temperatures, $\bar{T} = (T_H + T_C)/2$. This measurement was performed by repeatedly sweeping the gate voltage using six different heating currents, which supply ΔT as well as raise $\bar{T}$. The slopes of S at each resonance reveal information about the underlying transport mechanisms (see Fig. IV.1).

Thermopower was measurements at many different values of ΔT and T because these data help to understand thermopower oscillations and quantum-dot thermo-electrics in general. These measurements are easily performed by measuring S at many different heating currents thereby raising both T_H and T_C. Measurements of this kind are presented in Fig. 4.2.

Thermopower Slope dS/dE

S crosses zero at every quantum dot resonance, and near these zero-crossings, S is linear in energy, E. All the theoretical literature on this subject agrees that the qualitative behavior of this slope is $dS/dE = b/e\bar{T}$. But the quantitative value of the coefficient, b, is not universal. Multiple derivations [85, 86, 88, 90] predict $b = 1/2$ in the sequential tunneling regime in the limit where the quantum spacing, ΔE, is very small, that is, $\Delta E \ll kT \ll E_C$ (see Eq. (II.2)). This is often the case in 2DEG-based quantum dots on the scale of 1 μm. In the limit where $\Delta E > E_C \gg kT$, a value of $b = 1$ has been predicted [91]. This is the regime of nanowire-based quantum dots.

Unfortunately, these theoretical derivations do not take into account subtle experimental issues, and their relevance to experiments must be carefully scrutinized. The slope dS/dE near resonance is generally derived under the implicit assumption that the electrochemical potentials of the leads, $\mu_{H,C}$, do not change; only the quantum dot electrochemical potential, μ, is allowed to change. Experimentally, it is impossible to measure a *change* in S if μ_H and μ_C do not change because $S \propto \mu_H - \mu_C$.

Furthermore, any changes in μ_H and μ_C will perturb μ capacitively via the source and drain tunnel barriers. This creates feedback among μ, μ_H, and μ_C. V_G shifts μ with the strength of the lever arm, α, according to $\mu = -\alpha e V_G$, but only if $\mu_{H,C}$ are *constant*. But, in thermovoltage measurements, $\mu_{H,C}$ must be allowed to change as V_G changes, and the resulting change in μ becomes more complex. Therefore, to

model experiments correctly, derivations of dS/dE must consider changes in $\mu_{\mathrm{H,C}}$ and the interplay among $\mu_{\mathrm{H,C}}$ and μ.

Two derivations of dS/dE are included in Appendix E. One derivation uses the constant interaction model (CIM) (see Section II.3.1), while the other derivation uses the Mott formula (see Eq. (I.25)) to produce two different expressions for dS/dE. The results are summarized here,

$$\text{CIM:} \qquad \left(\frac{dS}{dE}\right)_{\mathrm{CIM}} = \frac{2}{e\overline{T}} \qquad\qquad \text{(IV.1a)}$$

$$\text{Mott:} \qquad \left(\frac{dS}{dE}\right)_{\mathrm{M}} = \frac{1}{2}\left(\frac{e}{k}\right)^2 L_0 \frac{1}{e\overline{T}}. \qquad \text{(IV.1b)}$$

The CIM-based derivation attempts to account for changes in μ owing to changes in $\mu_{\mathrm{H,C}}$ and assumes ballistic transport occurring exclusively at the adiabatic limit, that is, at energy E_{A} (see Eq. (I.19)). This derivation is appealing due to the success of the CIM as applied to Coulomb-blockade diamonds and open-circuit voltage measurements (see Section II.3.1 and Appendix C, respectively). The derivation from the Mott formula uses the Landauer equation to predict dS/dE from the analytical approximation of a differential conductance peak in the limit of $\Gamma \ll kT$, $\Delta T \ll kT$, and $eV \ll kT$ (see Eq. (II.9)). This derivation is trustworthy due to the ability of Eq. (II.9) to predict differential conduction peaks correctly (see, for example, the inset of Fig. 2.8). Also, this derivation is appealing because it generates a slope proportional to the Lorenz number, which has been carefully studied in relation to

thermal conductance [36] and the Wiedemann-Franz law [38]. In particular, it has been shown theoretically that the Lorenz number can be enhanced by sequential tunneling [90] and by cotunneling [38].

Inelastic cotunneling involves the transport of two electrons [92] and allows excited states to contribute to charge current [93, 94] and thermovoltage [87]. Inelastic cotunneling is a two-step transport process in which an electron leaves the dot at the same time an electron *of a different energy* enters the dot. See Fig. 4.3 for a brief illustration of inelastic cotunneling. Such transport involves the Nth-level ground state at energy ε_N and its mth excited state at energy $\varepsilon_{N,m} = \varepsilon_N + \Delta$. During transport, energy Δ is given to the quantum dot. When cotunneling events occur during voltage-biased current measurements, the bias voltage, V, is the energy source that supplies Δ. Therefore, inelastic cotunneling is not observed until $eV \geq \Delta$ [93, 94]. In thermoelectric measurements, the thermal energy supplies Δ.

Ref. [38] predicts a 9/5 increase in the Lorenz number due to cotunneling. Inserting the conventional value, $L_0 = \pi^2/3 \, (k/e)^2$, and the cotunneling value, $L_{\mathrm{CT}} = 9/5 L_0$, into Eq. (IV.1b) provides the sequential (S) and cotunneling (CT) results for dS/dE,

$$\left(\frac{dS}{dE}\right)_{\mathrm{S}} = \frac{\pi^2}{6} \frac{1}{e\bar{T}} \tag{IV.2a}$$

$$\left(\frac{dS}{dE}\right)_{\mathrm{CT}} = \frac{9}{5} \frac{\pi^2}{6} \frac{1}{e\bar{\bar{T}}}. \tag{IV.2b}$$
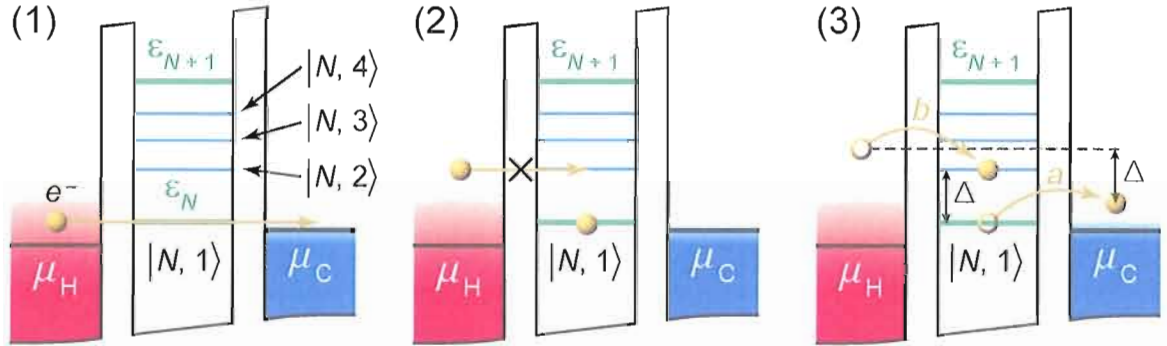
Figure 4.3. Different transport scenarios in a Coulomb-blockaded quantum dot. **(1)** Normal coherent tunneling of an electron occurring via the ground state, $|N, 1\rangle$ at energy ε_N. The mth excited state of the Nth resonance, $|N, m > 1\rangle$, with energy $\varepsilon_{N,m}$ does not contribute to charge transport or thermovoltage, because, for example, **(2)** the electron in the $|N, 1\rangle$ state has Coulomb blockaded the $|N, 2\rangle$ state. **(3)** Inelastic cotunneling. In process a, an $|N, 1\rangle$ electron leaves the dot lifting the Coulomb blockade on the excited states. Process b immediately follows; a thermally excited electron tunnels into $|N, 2\rangle$. This combined a-b process leaves the charge in the dot unchanged, but increases the energy in the dot by $\Delta = \varepsilon_{N,2} - \varepsilon_N$. The difference in energy between the final and initial states in the leads is $-\Delta$, so the total energy is conserved.

It must be said that all the theoretical predictions of dS/dE are only upper-limits because they assume the transmission resonances are infinitely narrow. Real resonances are of finite width and will reduce the magnitude of S and therefore its slope. In addition, other loss mechanisms in the experimental circuitry can decrease the measured value of S.

The various theoretical predictions are compared to experimental results in Fig. 4.4. The data agrees best with Eq. (IV.2b). This result highlights the importance of cotunneling effects in thermoelectric processes and suggests that higher-order processes, like cotunneling, can be explored using thermopower measurements.
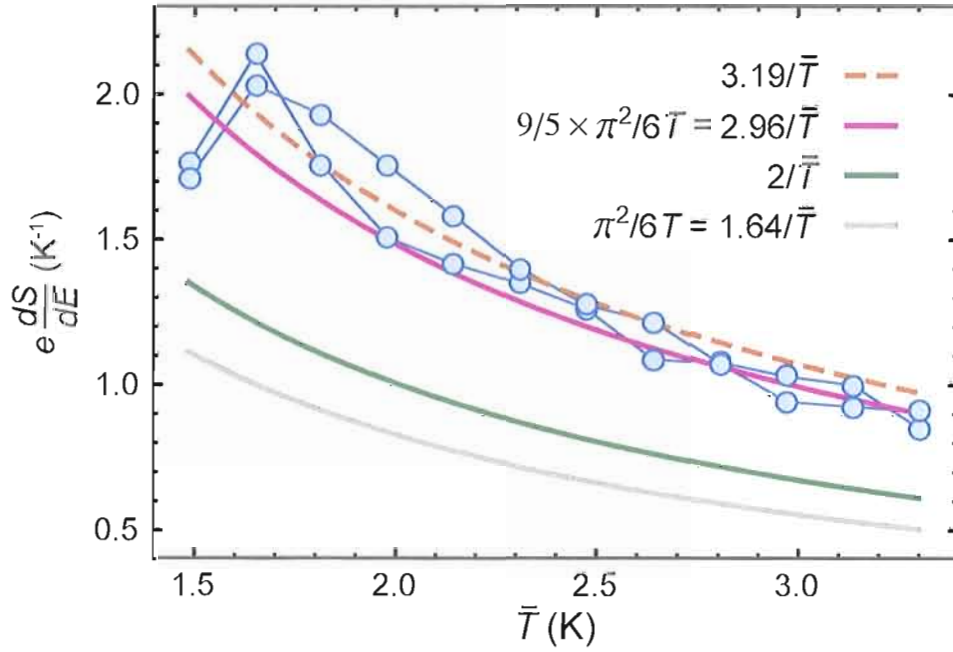
Figure 4.4. Thermopower slopes, dS/dE, versus average temperature, $\bar{T}$. The blue circles are values of dS/dE measured using the two neighboring thermopower oscillations shown in Fig. 4.2. The solid lines are theoretic predictions of the form $edS/dE = b/\bar{T}$ with $b = \pi^2/6$, 2, and $9/5 \times \pi^2/6$ based on Eqs. (IV.2a), (IV.1a), and (IV.2b), respectively. The dashed line is a $b/\bar{T}$ fit of the data with $b = 3.19$.

Thermopower and the Mott Formula

Thermopower can be predicted from conductance via the Mott formula (see Eq. (I.25)),

$$S_{\text{Mott}} = -eL_0\bar{T}\frac{1}{G}\frac{dG}{dE}. \tag{IV.3}$$

This equation requires no fit parameters if ΔT and T are known. Therefore, the Mott formula is a convenient tool for evaluating both the qualitative and quantitative behavior of thermopower measurements. The data in Fig. 4.1 has been used to

compare the measured S to that predicted by S_{Mott}, and the result is shown in

Fig. 4.5. As was done in Section IV.2, the Lorenz number has been increased by

9/5 to improve the agreement between theory and experiment providing additional

evidence that cotunneling plays a role in transport.
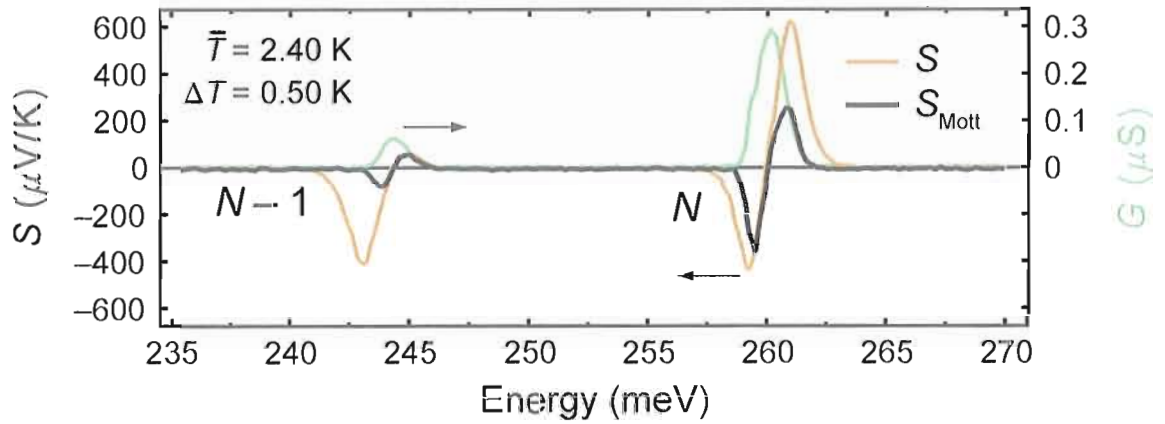
Figure 4.5. G from Fig. 4.1 is used to calculate S_{Mott}, the thermopower predicted by the Mott formula, Eq. (IV.3). The measured thermopower, S, is plotted for comparison. The Lorenz number was multiplied by 9/5 to improve the agreement between S_{Mott} and S. See text for details.

S_{Mott} in Fig. 4.5 does a fair job reproducing the qualitative behavior S and gross

quantitative values, such as approximate slope and peak height. This agreement

between S and S_{Mott} suggests that the measured V_{th} and ΔT values are reasonable.

Of course what the Mott formula cannot do is predict features in S that are not

correlated with features in G. For example, the $(N-1)$th peak of Fig. 4.5 displays

a very symmetric conductance peak, and as a result, S_{Mott} is very symmetric here.

Yet S is very asymmetric. The asymmetry in S at the Nth peak is less pronounced,

but present nonetheless, and unexplained by S_{Mott}. This behavior highlights the fact

that thermopower is much more sensitive to transport idiosyncrasies than electrical current and provides a useful investigation tool.

The best example of what thermopower measurements can offer is shown in Fig. 4.6, where S displays features completely uncorrelated to G. It has been suggested [95] that under special circumstances the thermodynamic coefficients G_{th} and G do not tend toward zero at the same rate, and $S = -G_{th}/G$ can have singularities when G goes to zero faster than G_{th}. These singularities create a signal in measurements of S that are unseen in current (or conductance) measurements. In these instances, it becomes impossible to predict S using G.
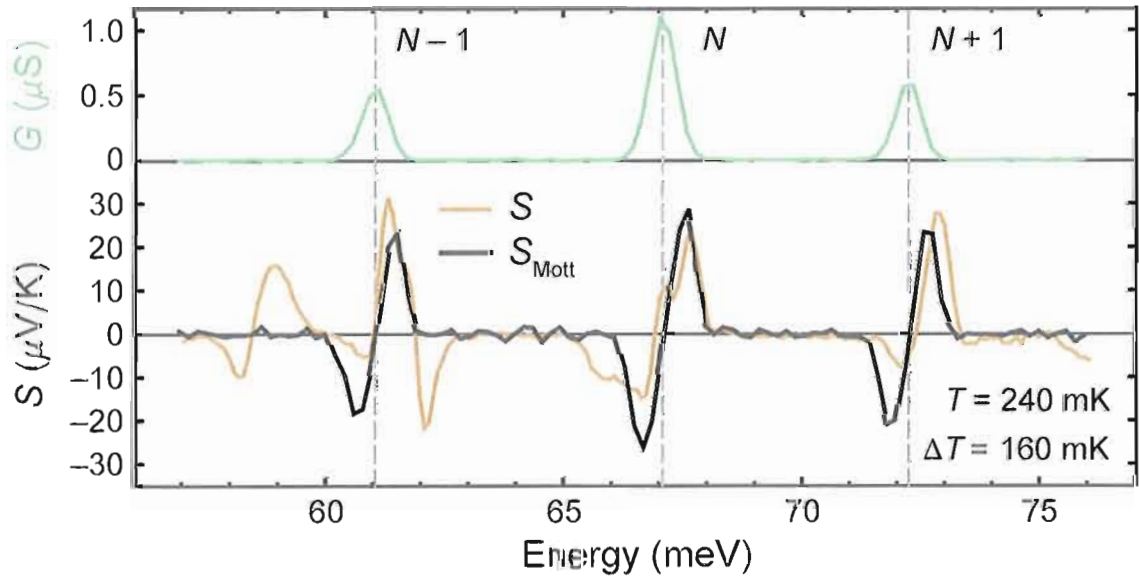
Figure 4.6. Differential conductance data, G, of the $N-1$, N, and $N+1$ conductance peaks is used in conjunction with the Mott formula (see Eq. (IV.3)) to calculate the thermopower, S_{Mott}, plotted alongside the measured value, S. The Mott formula does a fair job reproducing the N and $N+1$ thermopower oscillations. However, much of the behavior of S at the $N-1$ peak is not captured by S_{Mott}. Moreover, the features in the 57–60 meV energy range completely escape description because G is zero here.

One physical mechanism that could cause such a singularity is (destructive) quantum interference [95]. If an electron tunnels through the quantum dot via two separate conduction peaks simultaneously, the electron can interfere with itself destructively, masking its conduction. Because quantum interference requires two closely spaced energy levels, it is similar to cotunneling, which has already been shown to affect thermal transport in Section IV.2. Unfortunately, interference features like those in Fig. 4.6, although repeatable, are exceedingly rare, and a systematic investigation of their origin and behavior has not been possible.

<div align="center">Numeric Modelling of Thermovoltage</div>

Sommerfeld expansions or some other approximation is required to derive the Mott formula (see Section I.3.1). Without approximations, expressions for thermopower become complex rather quickly. The next most basic expression for S can be derived using the Landauer equation in conjunction with Eqs. (I.22) and (I.23a). Near equilibrium, when $\Delta T/\bar{T} \ll 1$ and $eV/k\bar{T} \ll 1$ the thermopower can be written,

$$S = -\frac{G_{\text{th}}}{G} \approx \frac{1}{e\bar{T}} \frac{\int \frac{df}{d\varepsilon} \left(\varepsilon - \mu\right) \tau\left(\varepsilon\right) d\varepsilon}{\int \frac{df}{d\varepsilon} \tau\left(\varepsilon\right) d\varepsilon}. \qquad (\text{IV.4})$$

Although this expression is only valid for small ΔT and V_{th}, it makes no assumptions about the transmission function, and therefore can be used to test the influence of Γ on the thermopower.
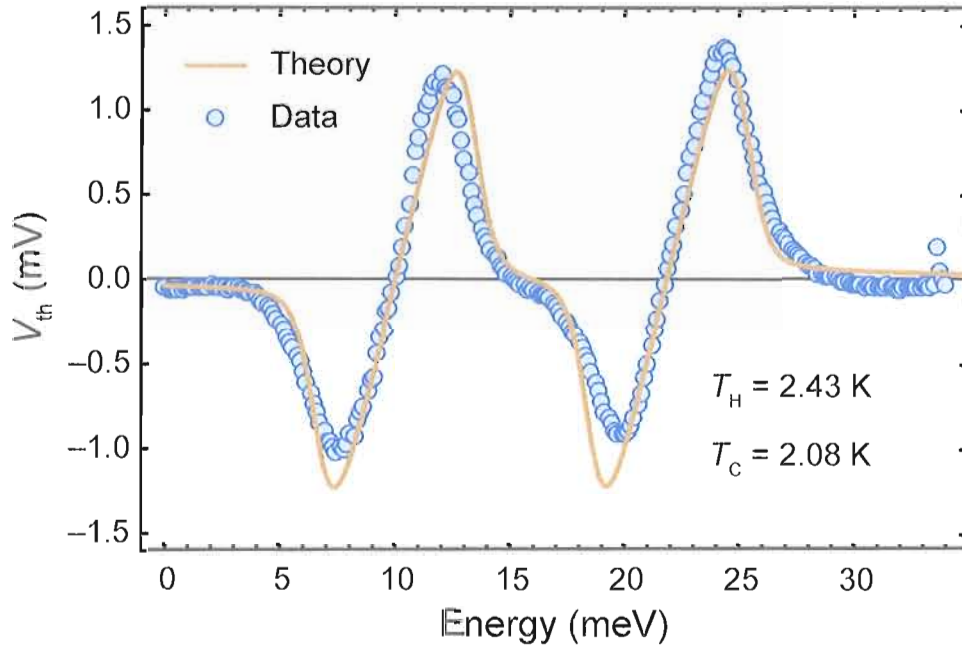
Figure 4.7. Data and a numerical model of thermovoltage as a function of energy. The simulation used $T_{\mathrm{H}}^{\dagger} = 5.02$ K, $T_{\mathrm{C}}^{\dagger} = 2.82$ K, and a Lorentzian transmission function with a FWHM of $\Gamma^{\dagger} = 10$ μeV.

Eq. (IV.4) was used to simulate experimental thermovoltage data by calculating $V_{\mathrm{th}} = S\Delta T$. The result is shown in Fig. 4.7. The model was produced using the parameters, $T_{\mathrm{H}}^{\dagger} = 5.02$ K, $T_{\mathrm{C}}^{\dagger} = 2.82$ K, and $\Gamma^{\dagger} = 10$ μeV (where $\dagger$ denotes model parameters). The experimentally measured temperatures were $T_{\mathrm{H}} = 2.43$ K and $T_{\mathrm{C}} = 2.08$ K. Though the $T_{\mathrm{H,C}}^{\dagger}$ and $T_{\mathrm{H,C}}$ values agree almost within a factor of 2, the values $\Delta T^{\dagger} = 2.2$ K and $\Delta T = 0.34$ K differ by an order of magnitude. This means that $S^{\dagger}$ and S would disagree by an order of magnitude. The value $\Gamma^{\dagger} = 10$ μeV is larger than expected because differential conductance measurements at $T = 550$ mK show that $\Gamma \ll kT = 47$ μeV (see Fig. 2.8). These discrepancies challenge the

validity of the numerical simulation, which can be addressed in the following manner. The average temperature of the model is $\bar{T}^{\dagger} = 3.92$ K, and therefore the two ratios $\Delta T^{\dagger}/\bar{T}^{\dagger} = 0.56$ and $eV_{\text{th}}^{\dagger}/k\bar{T}^{\dagger} = 2.96$ (for $V_{\text{th}}^{\dagger} = 1$ mV) violate the near-equilibrium assumptions, namely that $\Delta T/\bar{T} \ll 1$ and $eV/k\bar{T} \ll 1$. Therefore, the model has been pushed outside of its range of validity. Under these experimental conditions, thermopower cannot be modelled with equations that are only approximate.

Regardless of any approximations, models based on the Landauer formula do not include higher-order effects such as cotunneling and asymmetric tunnelling rates. The data in Figs. 4.4 and 4.5 suggests that cotunneling plays an important role in quantum-dot thermoelectrics, and nearly all measured thermopower oscillations demonstrate some degree of asymmetry. Therefore, more-advanced models are necessary to capture the observed behavior of thermopower oscillations, and further experimental research will assist in the development of such models and advance theoretical understanding.

<u>Conclusion</u>

Much effort was invested into building-up experimental and theoretical techniques for measuring V_{th} and ΔT in order to determine S. The main purpose of this chapter has been to establish that the measured values of S are quantitatively accurate. Measurements of dS/dE and comparison of measured S to Mott formula predictions show that the measured values of V_{th} and ΔT are reasonable if cotunneling effects are

included. Therefore, the measured thermopower values are reliable and can be used to determine the efficiency of a quantum dot.

Along the path toward efficiency measurements, thermopower has proven to be very interesting in its own right as a tool for investigating transport mechanisms. Much is still known and thermopower measurements can provide pathways for further research. The quantum interference and cotunneling effects observed here are only two phenomena which could be studied in a thermally biased quantum dot. Presently, understanding such effects is very important, because they limit the performance of single-electron devices and could play an important role in applications, such as semiconductor-based quantum computing, spintronics, and of course, thermoelectrics.

CHAPTER V

THERMOELECTRIC POWER MEASUREMENTS

<u>Introduction</u>

Power measurements complement efficiency measurements nicely because the relationship between power and efficiency is constrained; optimizing one of these parameters diminishes the other (see, for example, Fig. 7.2b). In the extreme case, a reversible process operating at Carnot efficiency has no preferred direction and operates infinitely slowly providing, therefore, zero output power. In any physical application, a compromise must be made between power and efficiency. For example, a power plant engineer needs only match the demand of the power grid and aims to maximize efficiency while doing so. On the other hand, the mechanic in a Formula 1 pit crew is concerned primary with power and has almost no regard for efficiency. The fundamental physics of this interplay is, in fact, quite rich. Motivated by the power-efficiency tradeoff, device efficiency can be approached from a power optimization perspective.

The topic of efficiency at maximum power is a relatively new interest in the field of thermodynamics, considering the fact that thermodynamic is one of the oldest branches of physics. In 1975, Curzon and Ahlborn found (in the endo-reversible

approximation) that the upper-limit of the efficiency at maximum power, η_P, obeys $\eta_P \leq \eta_{CA} = 1 - \sqrt{T_C/T_H}$ [96], where η_{CA} has become known as the Curzon-Ahlborn (CA) efficiency. This paper has created new thinking about old problems in thermodynamics. For example, the Feynman ratchet [97] has been studied from a power perspective [98] as have similar systems. Recent technical advances allowing the creation of synthetic Brownian motors, molecular motors, quantum ratchets, and quantum dots have lead to a flurry of both theoretical and experimental research on the efficiency and power of finite-size and finite-time processes in small-scale systems. Applying theory derived under ideal conditions to these small-scale, *near-ideal* experimental systems must be done with care. For example, because it is model-specific, the validity and universality of the Curzon-Ahlborn efficiency has been criticized. While the maximum efficiency of all heat engines is bounded by Carnot efficiency, the maximum efficiency at maximum power depends on the details of the heat engine. Theoretical research on the maximum η_P of a quantum dot [99] motivated broader research. As a result, it has been demonstrated [100] that the upper-limit of η_P is universal up to second-order in Carnot efficiency for all heat engines which posses 1) left-right symmetry and 2) linearly related thermodynamic fluxes. That is, all heat engines satisfying these two conditions operate with an η_P bounded according to $\eta_P \leq \eta_C/2 + \eta_C^2/8 + O\left(\eta_C^3\right)$, where η_C is Carnot efficiency. Note that the Curzon-Ahlborn efficiency can be expanded as $\eta_{CA} = \eta_C/2 + \eta_C^2/8 + O\left(\eta_C^3\right)$ and therefore agrees with this result up to second-order in η_C. Investigations of this

flavor highlight the complex relationship between power and efficiency. Quantum dots can be used to measure experimentally the value of η_P and to investigate other aspects of the power-efficiency relationship.

This chapter presents power measurements made using a quantum-dot nanowire device, and for the reasons discussed above, these results are interesting in their own right. In relation to the larger goal of this Ph.D. research, these power measurements serve primarily to support thermovoltage data using a comparison technique described in the next section. The conclusion of this chapter is that power measurements corroborate the measured thermovoltage data in Chapter IV.

Experimental Considerations

Thermoelectric power generation is measured by supplying a temperature gradient, ΔT, across a the device while simultaneously supplying a dc bias voltage, V. Any thermally produced current, that is, dc thermocurrent I_H, which flows through the device has done so by overcoming V. Therefore, the thermoelectric power is $P_{th} = I_{th}V$. Experimentally, however, I_{th} cannot be distinguished from the total dc current, I, and measurement of the thermoelectric power generation must be accomplished in another way. Using Eqs. (I.1) and (I.3), I can be written,

$$I = I_V + I_{th} = GV - GV_{th} = G\left(V - V_{th}\right), \tag{V.1}$$
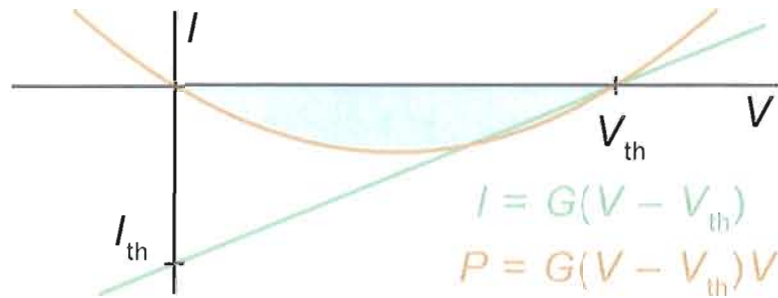
Figure 5.1. Total current, I, from Eq. (V.1) and output power, P, from Eq. (V.2) as a function of bias voltage, V, demonstrate the operation of a generic thermoelectric generator in the linear response regime. By definition (see Eq. (I.2)), the thermovoltage, V_{th}, is the voltage at which $I = 0$. Note that when $V = V_{th}$, $P = 0$ meaning that the system has stalled. Therefore, the $P = 0$ condition at finite V can be used to determine V_{th} (see Fig. 5.2). When $V = 0$, only the thermocurrent, I_{th}, flows, but $P = 0$ here because there is no voltage bias against which work can be done. The blue shaded region of negative P is where power production is accomplished thermally.

where I_V is the current produce by V. The total power, P, of the system is,

$$P = IV = I_V V + I_{th} V = GV^2 - GV_{th}V = G\left(V - V_{th}\right)V. \tag{V.2}$$

Therefore, the P can be determined by measuring I as a function of V and multiplying by V

V_{th} causes current to flow backward against the applied voltage, if and only if V_{th} is greater in magnitude than V and of the same sign (as per the chosen sign convention). This requirement is clearly reflected in Eq. (V.1). More to the point, this condition is also reflected in P because it is negative only when $|V_{th}| > |V|$ and when V_{th} and V have the same sign. Therefore, ΔT creates useful electric work at a

rate $|P|$ when $P < 0$, and V creates work at a rate P when $P > 0$. When $P = 0$, the system stalls ($I = 0$) and no work is performed (see Fig. 5.1).

Experimental Results

In a quantum dot device, power can be measured as a function of both V and gate voltage, V_G, to map out the power at all the working points in the V–V_G plain. The result is 2D plot of Coulomb-blockaded power, as shown in Fig. 5.2. Note that when $P = 0$ and $V \neq 0$, it must be true that $V_{th} = V$, meaning that a power measurement as a function of V is also a thermovoltage measurement. In addition to power, Fig. 5.2 includes thermovoltage measured as a function of gate voltage using the quasi-open-circuit technique (see Section II.4.5). This V_{th} data is plotted on top of the power data and follows the $P = 0$ contour around each lobe of the Coulomb-blockaded power. The qualitative agreement between V_{th} and P is quite striking. For example, the left resonance displays gross asymmetry captured by both P and V_{th}. Their quantitative agreement necessitates multiplying the V_{th} data by a factor of 2.4. This requirement is interpreted as the scaling factor ϕ between the measured V_L and V_{th} (see Section II.4.5).

The effective resistance of the nanowire (which is analogous to the internal resistance of a battery) can be found using Eq. (II.16) for the thermovoltage scaling factor, $\phi = 1 + R/R_L$. From the power measurement, $\phi = 2.4$, and for this setup, the load resistance is $R_L = 47$ MΩ. Therefore, the effective resistance of the nanowire is
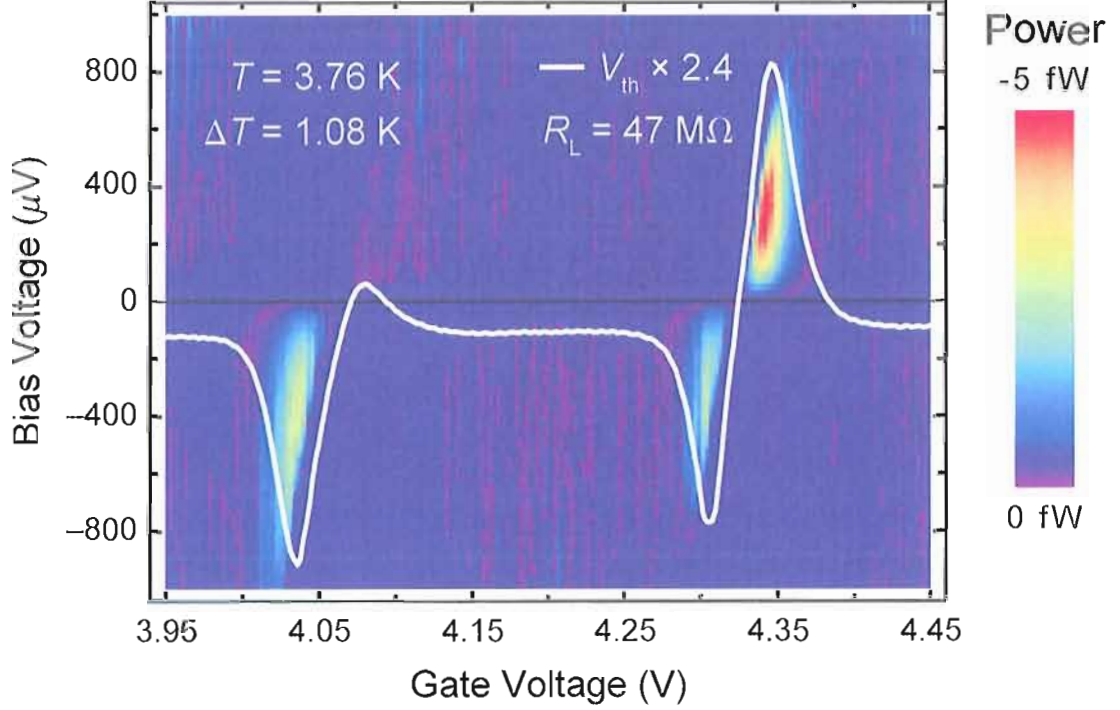
Figure 5.2. Power, P, as a function of bias voltage and gate voltage displays Coulomb-blockade features. The experiment involves measuring, as a function of gate voltage, the net current, I, produced by the nanowire, which has been simultaneously electrically and thermally biased ($\Delta T = 1.08$ K). In order to determine $P = IV$ (see Eq. (V.2)), the measured current is simply multiplied by the applied bias voltage, V. Only negative power is shown ($P \geq 0$ is assigned the same color), because negative P is the electrical power generated thermally (see Fig. 5.1). The solid white line is a trace of V_{th} measured at $\Delta T = 1.08$ mK with a load resistance $R_L = 47$ MΩ using the quasi-open-circuit technique described in Section II.4.5. R_L is *not* used during power measurements. (The effective load resistance used during power measurements is the 10 kΩ input impedence of the low-noise current preamplifier.) For each gate voltage, the finite bias voltage at which $P = 0$ provides a measurment of V_{th} (see Fig. 5.1). Therefore, this power measurement is also a thermovoltage measurement. The quasi-open-circuit V_{th} must be multiplied by a factor of 2.4 to achieve agreement with the power measurement. This $\phi = 2.4$ scaling factor implies that the effective internal resistance of the nanowire is roughly 66 MΩ (see text for details).

$R = 66$ MΩ. In comparison, differential conductance data shows that on resonance these two peaks have a resistance $R \approx 3$ MΩ, while far off resonance (where electrons tunnel through the entire double-barrier structure as if it were a single barrier) $R > 100$ MΩ. Most of the thermovoltage data is collected on the shoulders of the conductance peaks where the effective resistance is between 3 MΩ and 100 MΩ. Therefore, the $R = 66$ MΩ result is reasonable, and the power data supports the quasi-open-circuit thermovoltage measurements.

Conclusion and Outlook

Coulomb-blockaded power has been measured in a quantum dot as a function of bias and gate voltage—perhaps the first measurement of its kind. The power measurements presented here are intended to confirm the quasi-open-circuit thermovoltage measurement technique. The two methods do indeed agree once the internal resistance of the thermoelectric circuit is considered. These findings bolster the confidence with which efficiency results are reported (see Chapter VII).

Power is an interesting phenomena from a thermodynamic perspective and is extremely relevant to device applications. The power of thermoelectric devices remains relatively unexplored, and the fundamental relationship between power and efficiency is underrated. Quantum dots offer a unique experimental perspective on these topics owing to energetic tunability. Coupled with quantum-dot thermometry,

nanowire devices provide an excellent platform for future research on thermoelectric power generation.

CHAPTER VI

NONLINEAR THERMOELECTRICS

Introduction

Mesoscopic systems are easily driven into the nonlinear regime by applying a sufficiently large voltage bias. Nonlinear behavior can be attributed in part to the density of states (DOS) and the quantum transmission function, which are acutely energy dependent. In particular a non-monotonic, modulated transmission function provides similarly modulated, nonlinear IV behavior, in direct violating of Ohm's law. The analogous thermoelectric behavior is a nonlinear relationship between the thermovoltage, V_{th}, and the applied temperature difference ΔT. In the nonlinear thermoelectric regime, Eq. (I.2) is no longer valid, and V_{th} must be described to higher order in ΔT. A power series expansion of V_{th} reads,

$$V_{\text{th}} = S_1 \Delta T + \frac{1}{2} S_2 \left(\Delta T \right)^2 + \frac{1}{3!} S_3 \left(\Delta T \right)^3 + \cdots ,$$

where the S_i coefficients are the ith-order Seebeck coefficients. In the mesoscopic realm, the evaluation of thermoelectric performance must performed with care by considering nonlinear behavior and avoiding assumptions about S and ZT that are

meaningful only if V_{th} is linear in ΔT. This is an important proviso in quantum-dot thermoelectrics, because the sought enhanced thermoelectric performance resulting from abrupt energy features and electron energy filtering requires the transport properties to change on a scale of ΔT, which is same the physical condition that generates nonlinearity.

The mesoscopic thermoelectrics community has addressed nonlinear thermoelectric behavior on few occasions. In fact, the first, and perhaps only, experimental evidence of strong nonlinear behavior in a mesoscopic system was considered an odd curiosity, neither explored nor understood [71]. This research deficit is due in part to the challenge of thermometry and in part to the deficit of suitable mesoscopic devices. Nonlinear experiments might require a relatively large $\Delta T/T$ before the onset of strong nonlinear effects are observed. Often in experiments, attempts to raise ΔT significantly have the side effect of increasing T as well, and $\Delta T/T$ remains relatively small. However, the combination of heating and thermometry techniques developed for thermoelectric efficiency experiments have made nonlinear investigations possible.

As for candidate devices, the system must have sharp, non-monotonic energy features in order to observe strong nonlinearity. Not every mesoscopic system offers this. For example, theory predicts that quantum point contacts (QPCs) only demonstrate weak nonlinear thermoelectric effects due to their monotonic transmission spectrum [101], a conclusion which has been corroborated experimentally [74, 102]. On the other hand, a double-barrier quantum structure offers energy

quantization adequate for producing strong nonlinear thermal effects [103]. A small quantum dot operating in the few-electron regime is the best choice for observing for nonlinearity, because it offers well-spaced energy resonances with relatively few neighboring excited states. Therefore, a quantum-dot nanowire, equipped with contacts for direct electron-gas heating, is very well suited for observing strong nonlinear thermoelectric behavior. In the remainder of this chapter, nonlinear thermovoltage measurements and a numerical model are presented. Quantum-dot thermometry (see Chapter III) was used to determine the temperature difference across the quantum dot. Fig. 3.8 shows that this thermometry technique is valid even with the potentially large $\Delta T/T$ in nonlinear experiments.

Experimental Results

Nonlinear V_{th} as a function of ΔT was measured at several different gate voltages. The qualitative behavior of each V_{th} versus ΔT curve varies significantly depending on the gate voltage. The nonlinearity is very strong, and even complete sign reversals have been measured (see Fig. 6.1). During thermovoltage measurements, the drain and quantum dot electrochemical potentials are allowed to shift, obfuscating which energy scales and physical features are responsible for the nonlinearity. Nonlinear experiments can be simplified if thermocurrent rather than thermovoltage is measured, because thermocurrent measurements reduce the number of degrees of freedom in the system.
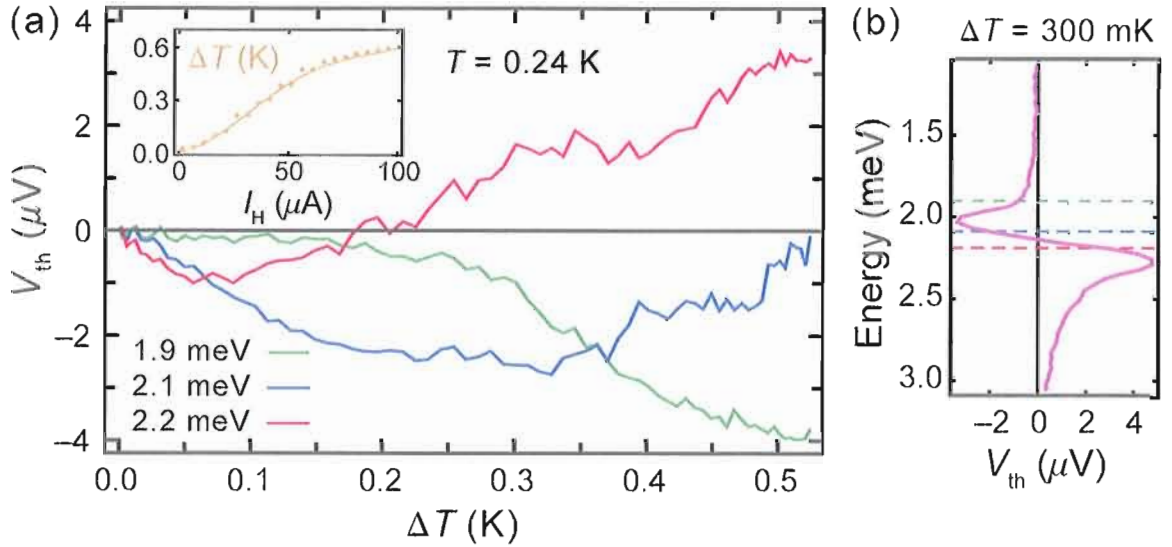
Figure 6.1. (a) ac thermovoltage, V_{th}, measured as a function of temperature difference, ΔT, shows strong nonlinear behavior for $\Delta T/T$ as small as 0.2. The data was measured at the three labelled quantum dot energies, which correspond to the three dashed horizontal lines in (b). *Inset:* ΔT as a function of heating current, I_H, determined using quantum-dot thermometry (see Section III). (b) V_{th} as a function of quantum dot energy with $\Delta T = 300$ mK. The dashed horizontal lines indicate at which energies data in (a) was measured. A quantum dot resonance aligns with the source electrochemical potential near 2.1 meV. See Chapter IV for more information about thermovoltage oscillations.

Measurements of thermocurrent, I_H, are performed with the cold drain contact connected to a low-noise current preamplifier (see Fig. 2.13). A further simplification is to apply a dc bias voltage, V, to the hot source contact and then use the gate voltage to align the hot electrochemical potential, μ_H, with a resonance of the quantum dot. At the same time, the cold electrochemical potential, μ_C, can be kept far from any quantum dot resonance (see Fig. 6.2b). In this dc voltage-biased setup, thermocurrent is distinguished from biased current by using a low-frequency

ac heating current. Therefore, the temperature gradient and resulting thermocurrent are also ac (and frequency doubled). The thermocurrent is measured using a lock-in amplifier referenced to the second-harmonic of the heating current. See Section II.4.4 for more experimental details. Measured thermocurrent and a numerical model are shown in Fig. 6.2.

Numerical Modelling

The system is modelled using the full Landauer equation without any approximations, because, in the nonlinear regime, it cannot be assumed that $\Delta T/T$ and eV/kT are small, as is usually done to calculate I via the Landauer equation. The total current is given by,

$$ I = -\frac{2e}{h} \int \left[f_{\mathrm{H}}\left(\varepsilon, T_{\mathrm{H}}, V_{\mathrm{H}} \right) - f_{\mathrm{C}}\left(\varepsilon, T_{\mathrm{C}}, V_{\mathrm{C}} \right) \right] \tau\left(\varepsilon \right) d\varepsilon, \qquad \text{(VI.1)} $$

where the arguments of the two Fermi-Dirac distributions, $f_{\mathrm{H,C}}$, are $\xi_{\mathrm{H,C}} = \left(\varepsilon - \mu_{\mathrm{H,C}} \right)/kT_{\mathrm{H,C}}$ and the Lorentzian transmission function has an energetic width of Γ. Here $\mu_{\mathrm{H,C}}$ are measured relative to the quantum dot electrochemical potential, μ, so that, $\mu_{\mathrm{H,C}} = \mu - eV_{\mathrm{H,C}}$. When $I = 0$, the system has stalled, and the thermovoltage, $V_{\mathrm{th}} = V_{\mathrm{C}} - V_{\mathrm{H}}$, has been established. In principle, V_{th} can be determined by finding the values of $V_{\mathrm{H,C}}$ which provide $I = 0$. But there is one additional complication. As the voltages $V_{\mathrm{H,C}}$ change, they influence μ capacitively via $C_{\mathrm{S,D}}$, which in turn

influences $\mu_{\mathrm{H,C}}$ thereby creating a feedback mechanism. Modelling experimental data with Eq. (VI.1) becomes impractical once the feedback is folded in because the number of fit parameters becomes large: $V_{\mathrm{H,C}}$, Γ, and $C_{\mathrm{S,D,G}}$ make six. The parameters $V_{\mathrm{H,C}}$ and $C_{\mathrm{S,D,G}}$ are interrelated and could be reduced in number. However, ΔT is large in the nonlinear regime, and the thermally smeared Fermi-Dirac distributions are free to explore large regions of energy space, which likely includes excited states. Describing extra details of the transmission function, such as excited states, blocking states, and cotunneling requires more fit parameters. With so many fit parameters, a model based on Eq. (VI.1) is most likely overdetermined.

During thermocurrent measurements, the bias voltage is fixed, and the only fit parameters necessary for modelling I_{th} are those which describe $\tau(\varepsilon)$. Furthermore, because μ_{C} is not near a resonance of the quantum dot, $f_{\mathrm{C}}(\varepsilon)$ does not influence the current, that is, $\int f_{\mathrm{C}}(\varepsilon)\tau(\varepsilon)\,d\varepsilon = 0$. Measurement of ac thermocurrent is differential in nature because the lock-in measures the rms amplitude of the current as it oscillates between its maximum (at full ΔT) and minimum (at $\Delta T = 0$). Therefore, the (differential) thermocurrent can be written,

$$I_{\mathrm{th}} = -\frac{2e}{h}\int\left[f_{\mathrm{H}}(\varepsilon,T_{\mathrm{H}},V) - f_{\mathrm{H}}(\varepsilon,T,V)\right]\tau(\varepsilon)\,d\varepsilon = -\frac{2e}{h}\int F\tau(\varepsilon)\,d\varepsilon, \qquad \text{(VI.2)}$$

where $F = f_{\mathrm{H}}(\varepsilon,T_{\mathrm{H}},V) - f_{\mathrm{H}}(\varepsilon,T,V)$, and $T_{\mathrm{H}} = \Delta T_{\mathrm{H}} + T$. In order to model excited states not described by a Lorentzian transmission function, $\tau(\varepsilon)$ is found numerically
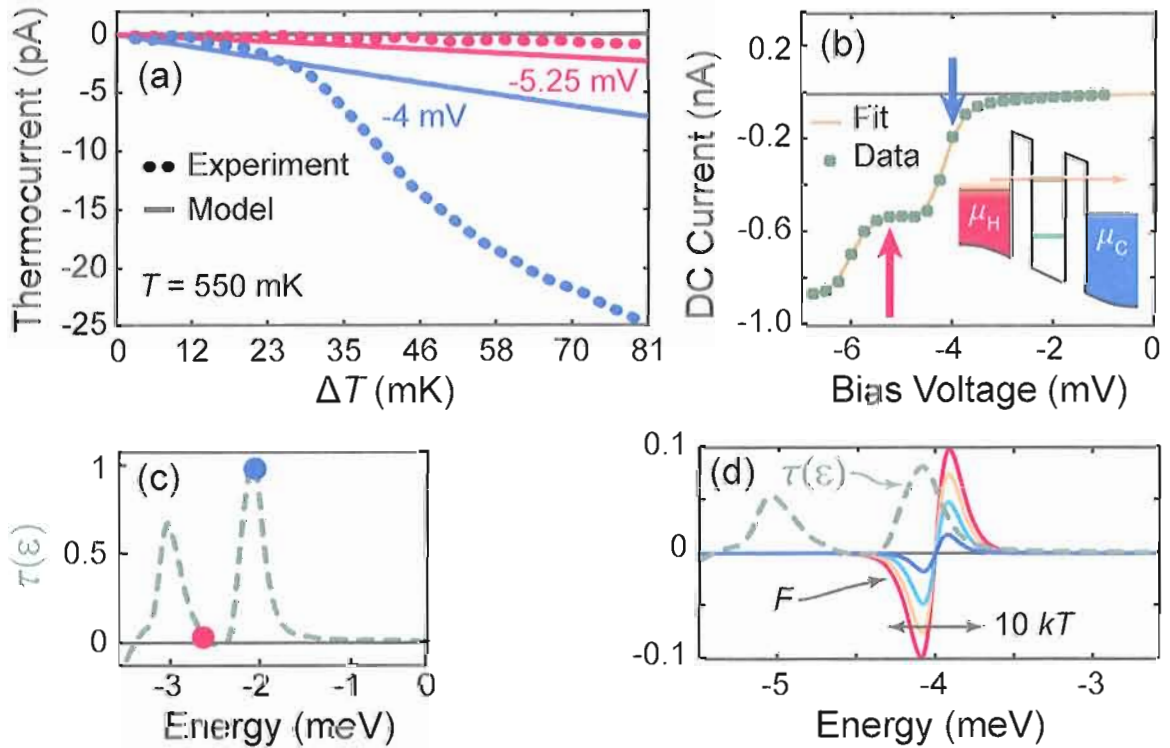
Figure 6.2. (a) Experimental and modelled ac thermocurrent as a function of ΔT. The data was collected at a bias voltage of -4 mV and -5.25 mV, as shown by arrows in (b) Measured dc current as a function of bias voltage was used to find the quantum transmission function, $\tau(\varepsilon)$, in (c). $\tau(\varepsilon)$ was tweaked iteratively until it correctly reproduced the IV curve in (b). Inset: μ_C is biased away from a resonance, and μ_H is brought near a resonance. (c) $\tau(\varepsilon)$ calculated using the data in (b). The red and blue dots indicates the working points for data collection in (a). (Note that $\tau(\varepsilon)$ moves during biasing.) (d) The integrand of Eq. (VI.2) as a function of energy, with F evaluated at $\Delta T = 12$, 35, 58, and 81 mK. When F is centered just off resonance, nonlinear behavior in (a) is enhanced.

based on dc IV data. A numeric $\tau\left(\varepsilon\right)$ function is first approximated then iteratively adjusted until a suitable match between data and simulation is achieved (see b and c of Fig. 6.2). Once $\tau\left(\varepsilon\right)$ has been found, I_{th} is calculated numerically at different values of ΔT in an attempt to reproduce the experimental data.

This model has been used to simulate the data in Fig. 6.2a with mixed success. Data that is weakly nonlinear can be reproduced by Eq. (VI.2), as demonstrated by the -5.25 mV data in Fig. 6.2a. However, the strong nonlinear behavior already observed when $\Delta T/T \geq 0.06$ cannot be duplicated numerically, as shown by the -4 mV data in Fig. 6.2a.

Thermally driving a system into a nonlinear regime requires a ΔT which is large compared to the energy scale over which charge transport has appreciable energy dependence. The strong nonlinear behavior observed as soon as $\Delta T/T \geq 0.06$ suggests that kT is not the relevant energy scale. $\tau\left(\varepsilon\right)$ is the next most likely candidate for establishing an energy scale. However, the excited state in the calculated $\tau\left(\varepsilon\right)$ in Fig. 6.2c has an approximate energetic width of $\delta E = 0.25$ meV, which is equivalent to $5.3\ kT$ at the background temperature of the experiment, $T = 550$ mK. It is no surprise that the model cannot produce nonlinear effects because, at its largest, $k\Delta T/\delta E = 0.03$, implying that $\tau\left(\varepsilon\right)$ is too broad to effect nonlinear behavior in I_{th}. Nonetheless, $\tau\left(\varepsilon\right)$ does produce the correct dc current using the Landauer equation. With $\tau\left(\varepsilon\right)$ eliminated as a possible candidate, the source of the observed nonlinearity is unknown. However, $\tau\left(\varepsilon\right)$ has not been measured directly, only inferred

from IV data. Other effects could influence the behavior of $\tau(\varepsilon)$ and mask its true functional form. For example, $\tau(\varepsilon)$ could be a function of bias voltage, something the model does not include. Or the excited state could in reality be a cluster of very narrow excited states which cannot be resolved by IV measurements. Investigations in those directions might prove fruitful.
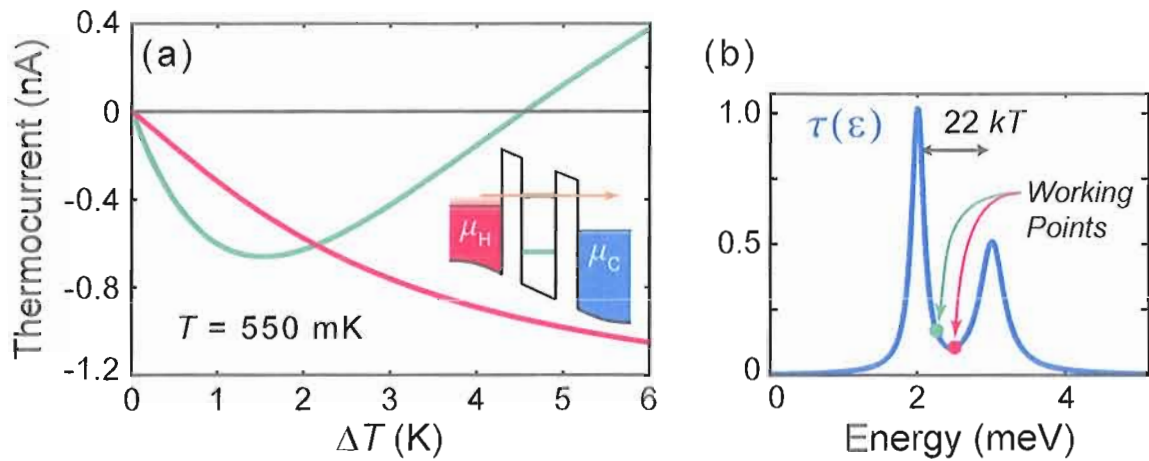
Figure 6.3. **(a)** Modelled ac thermocurrent as a function of ΔT. *Inset:* The system is biased to align μ_H to the quantum-dot resonance. **(b)** A trial quantum transmission function $\tau(\varepsilon)$. The green and red thermocurrent curves where calculated by feeding $\tau(\varepsilon)$ into Eq. (VI.2) and integrating while F was centered at the respective working points indicated by the green and red dots. The system demonstrates strong nonlinear behavior only when $\Delta T/T > 2.5$.

If the model is pushed hard enough, strong nonlinearity is observed (see Fig. 6.3), but only at much larger $\Delta T/T$ than observed experimentally. Quantitatively the model fails to duplicate experiment, but qualitatively the two are very similar. What is clear from the modeling is that asymmetry plays a large role in nonlinearity. If the integrand term F is centered just off resonance, the thermocurrent eventually

makes a complete sign reversal at large enough ΔT. This can be explained by the competition between electron- and hole-dominated transport. For example, at the working point indicated by the green dot in Fig. 6.3b, the peak to the left transports electrons, while the peak to the right transports holes. Starting at small ΔT, the system is in the electron-dominated regime because it is closer to the left peak, and so I_{th} is negative. As ΔT increases, F expands increasing hole transport via the right peak and increasing I_{H} until eventually $I_{\mathrm{H}} = 0$.

Conclusion and Outlook

Several aspects of the system contribute to the observed nonlinear behavior, including sharp energy features, asymmetry, and a competition between electrons and holes. The present thinking is that the transmission function of the quantum dot—which has sharp, non-monotonic features—is the primary source of nonlinearity. Asymmetry helps induce nonlinearity if the energy features are sufficiently sharp, and any symmetry-breaking mechanisms, such as barrier asymmetry, might play a role in the creation of nonlinear behavior. The competition between electrons and holes cannot be achieved with a monotonically increasing transmission function and is enhanced by excited states.

Predicting nonlinear behavior with numeric modelling has been somewhat success-ful, especially for understanding the qualitative features. Although, improvements

can certainly be made toward achieving better agreement between simulation and experiment.

A competition between electrons and holes was used to explain the experimental observation that thermocurrent as a function of ΔT can be non-monotonic and eventually fall to zero. Therefore, as ΔT increases and carriers are given more energy to explore larger regions of the system's energy space, the combined nonlinear effect of multiple small energy features could limit or even quench device efficiency and power generation. In this way, nonlinear effects are most likely detrimental to efficiency and power generation.

CHAPTER VII

THERMOELECTRIC EFFICIENCY MEASUREMENTS

Single-energy charge transport in particle heat engines provides the necessary condition for achieving thermodynamic reversibility and the theoretical maximum efficiency, Carnot efficiency (see Section I.2.3). The best candidate for an experimental realization of this ideal theoretical construct is a quantum dot because it satisfies two necessary conditions. First, quantum dots operated at cryogenic temperatures (< 10 K) have single-electron energy resonances which are separated in energy by much more than the thermal energy (see, for example, Fig. 2.12). This allows each resonance to be addressed individually. Second, quantum dots can, in theory, provide arbitrarily narrow energy resonances and reach the ideal single-energy limit.

A real device will display some degree of intrinsic resonance broadening due to the finite width of its tunnel barriers (see Fig. 2.5). However, the energetic resonance width, Γ, can indeed be made to satisfy the condition that $\Gamma \ll kT$ (see, for example, the inset of Fig. 2.8). Thus, quantum dots provide all the prerequisites for the realization of a highly efficient thermoelectric heat engine. This chapter outlines the technique used to evaluate experimentally the electronic efficiency of a quantum dot using the ZT figure of merit and provides quantitative evidence concluding that a quantum dot can indeed operate very near Carnot efficiency.

<u>Electronic Efficiency $(ZT)_{\mathrm{el}}$</u>

The efficiency of any thermoelectric device is most easily measured by determining the ZT figure of merit (see Section I.4 and Eq. (I.31)). The ZT figure of merit is defined as

$$ Z\bar{T} = \frac{S^2 G \bar{T}}{\kappa_{\mathrm{el}} + \kappa_{\mathrm{ph}}}, $$

where S is thermopower, G is electrical conductance, $\kappa = \kappa_{\mathrm{el}} + \kappa_{\mathrm{ph}}$ is total thermal conductance, and $\bar{T} = (T_{\mathrm{H}} + T_{\mathrm{C}})/2$. These device parameters have strict definitions (see Section I.1), and determining ZT accurately requires measurement techniques that respect these definitions. The thermal terms S, κ_{el}, and κ_{ph} must be measured under open-circuit conditions in which $I = 0$, while G must be measured under isothermal conditions in which $\Delta T = 0$. All the parameters must be measured at the same temperature, namely $\bar{T}$.

In order to evaluate only the electronic performance of a thermoelectric device, phonons are completely neglected. In this case, the *electronic ZT* is defined as

$$ (ZT)_{\mathrm{el}} \equiv \lim_{\kappa_{\mathrm{ph}}=0} Z\bar{T} = \frac{S^2 G \bar{T}}{\kappa_{\mathrm{el}}}. $$

G and κ_{el} are related to one another according to the Wiedemann-Franz law (see Eq. (I.26)) when $\Delta T \ll \bar{T}$ and $eV \ll k\bar{T}$ in the low-temperature and small bias regime. In this regime, the Wiedemann-Franz law states $\kappa_{\mathrm{el}} = L_0 G \bar{T}$ so that $(ZT)_{\mathrm{el}}$

can be further simplified to

$$(ZT)_{\text{el}} = \frac{S^2 G \bar{T}}{L_0 G \bar{T}} = \frac{S^2}{L_0}, \qquad \text{(VII.1)}$$

where the conventional Lorenz number, $L_0 = \pi^2/3 \, (k/e)^2$, has been discussed in Section IV.2. In this way, maximizing the Seebeck coefficient optimizes $(ZT)_{\text{el}}$.

Experimental Results

The prediction for the Lorenz number of 1D ballistic transport through a nondegenerate energy level is $L_{\text{B}} = 3L_0/\pi^2 = (k/e)^2$ [36]. This regime offers the best conditions for achieving highly efficient thermoelectric performance because it provides strong energy filtering. On the other hand, cotunneling can affect thermoelectric performance, for example, by increasing the thermopower (see Section IV.2). While the additional conduction pathways provided by cotunneling increase S and G and the device power factor, $S^2 G$, these pathways should in fact decrease efficiency as they destroy thermodynamic reversibility. Efficiency degradation via cotunneling is reflected in $(ZT)_{\text{el}}$ because cotunneling increases the Lorenz number. The cotunneling Lorenz number is $L_{\text{CT}} = 9/5 L_0 = 3\pi^2/5 \, (k/e)^2$ [38]. In a device with energy levels separated by much more than the thermal energy ($E_{\text{add}} \gg kT$) and very narrow transmission widths ($\Gamma \ll kT$), cotunneling is a very likely candidate for limiting device performance.

It is quite possible that cotunneling is weak and the effective Lorenz number, L_{eff}, falls somewhere between L_{B} and L_{CT} ($L_{\text{B}} \leq L_{\text{eff}} \leq L_{\text{CT}}$). Even more likely is that L_{eff} is not constant, but instead changes as the electrochemical potential of the quantum dot changes. Under both circumstances, neither L_{B} nor L_{CT} can predict the true value of $(ZT)_{\text{el}}$. Therefore, calculations of $(ZT)_{\text{el}}$ using L_{B} can be regarded as an upper-limit, while those using L_{CT} can be regarded as a lower limit.

Fig. 7.1 shows measured values of G, S, S^2, and $(ZT)_{\text{el}}$ as a function of energy across two-quantum dot resonances. $(ZT)_{\text{el}}$ in the figure has been calculated using both L_{B} and L_{CT} to show the disparity between the two predictions. Fig. 7.1d demonstrates the relationship between the ZT figure of merit and the efficiency relative to Carnot efficiency, η/η_{C}, evaluated using experimental parameters according to Eq. (I.30). Note that, in the derivation of this relationship, perfect impedance matching was assumed (see Section I.4 for details). This provides a conversion of $(ZT)_{\text{el}}$ to electronic efficiency. The upper-limit of the measured electronic efficiency of the device is nearly 95% Carnot efficiency. Fig. 7.2d shows the lower-limit of $(ZT)_{\text{el}}$ calculated using L_{CT}. Despite being the lower-limit, its maximum near 88% Carnot efficiency is very respectable. This bounds the efficiency between 88% and 95% of Carnot efficiency demonstrating that the quantum dot operates as a very efficient thermoelectric device.

Fig. 7.2a shows numeric modelling of $(ZT)_{\text{el}}$ as a function of quantum dot energy that is quantitatively similar to the measured $(ZT)_{\text{el}}$. The agreement between theory
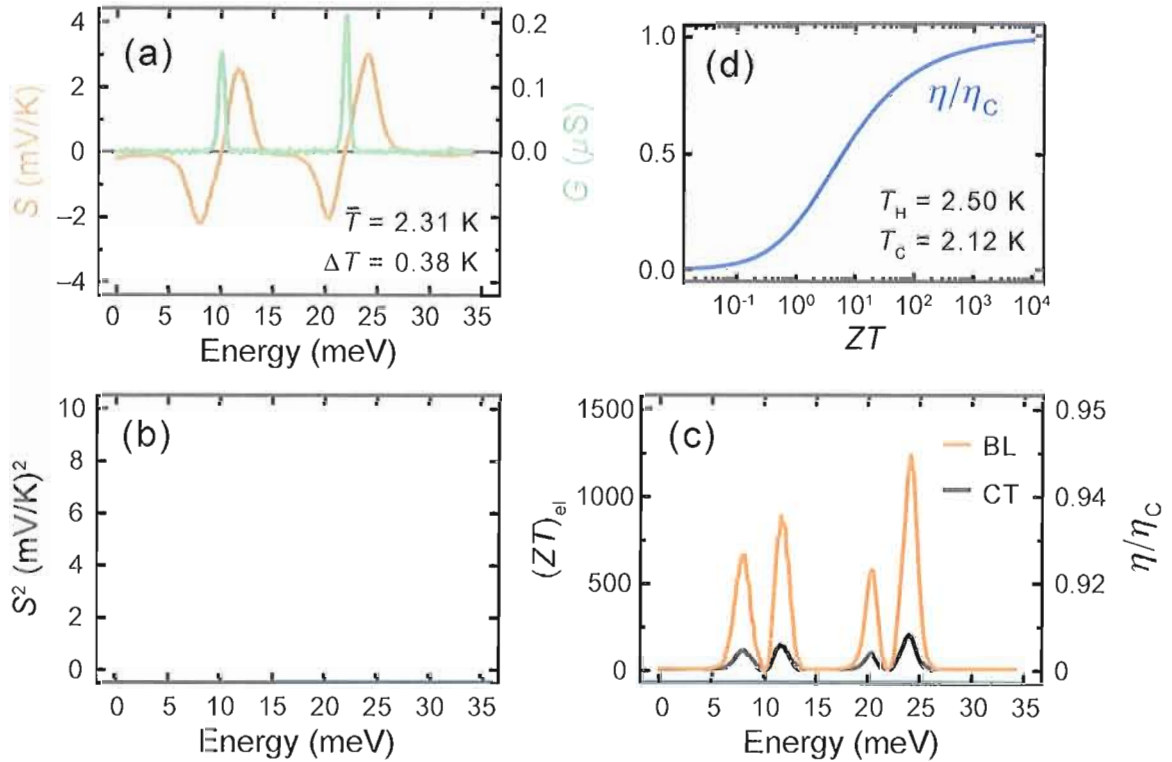
Figure 7.1. The evolution of raw data in (a) concludes with measured efficiency in (c) via the theoretical relation in (d). **(a)** Measured values of S and G as a function of energy. **(b)** The square of S from (a). **(c)** The electronic efficiency, $(ZT)_{\text{el}}$, calculated using S^2 in (b). The two curves result from using two different Lorenz numbers, the ballistic (BL) and the cotunneling (CT) (see text for details). The electronic efficiency relative to Carnot efficiency, η/η_{C}, on the right-hand axis was converted from $(ZT)_{\text{el}}$ values via **(d)**; a log-linear plot of the theoretical expression relating, η/η_{C}, to the ZT figure of merit calculated using Eq. (I.30) with the listed experimental $T_{\text{H,C}}$ values.

and modeling offers insurance that the measured values of $(ZT)_{\text{el}}$ are reasonable. The

next section provides details about the numeric simulations.

Modelling Results

Numeric modeling of quantum dot efficiency can be calculated in two different but

equivalent ways. The first method essentially calculates $(ZT)_{el}$ directly using the full Landauer formula. This is achieved by calculating the thermodynamic coefficients G, S, and κ_{el} using Eqs. (I.11), (I.22), and (I.23) and combining them to calculate $(ZT)_{el} = GS^2\bar{T}/\kappa_{el}$. An example of $(ZT)_{el}$ calculated in this way is presented in Fig. 7.2a at $\bar{T} = 1$ K for $\Gamma/k\bar{T} = 0.1$ and 0.01. The maximum $(ZT)_{el}$ for these two choices of Γ are $(ZT)_{el} \approx 30$ and $(ZT)_{el} \approx 265$, respectively, which corresponds to a relative efficiency of $\eta/\eta_C = 0.71$ and $\eta/\eta_C = 0.89$ (see Fig. 7.2c).

The second method for calculating efficiency does not rely on the ZT figure of merit, which is an advantage over the aforementioned method. The method again uses the Landauer formula, but calculates both efficiency and power directly based on their definitions. The efficiency is the ratio of the output work, W, performed by the total current, I, against the applied voltage, V, to the total heat flow, Q, leaving the hot bath. That is,

$$\eta = \frac{W}{Q} = \frac{IV}{\Pi I + \kappa_{el}\left(T_{\mathrm{H}} - T_{\mathrm{C}}\right)}, \qquad (\mathrm{VII.2})$$

where Q has been provided by Eq. (I.7). Eq. (VII.2) is an expression for electronic efficiency because κ_{ph} is not included. No Joule heating term is included in the total heat flux because the quantum dot operates ballistically, that is, without energy dissipation. The Landauer equation (Eq. (I.20)) can be used to calculate I directly. Calculations of the Π and κ_{el} thermodynamic coefficients, without approximation, are less direct. First, the thermodynamic coefficients G, G_{th}, and K defined via Eq. (I.4)

are calculated using Eqs. (I.22), (I.23a) and (I.23b), respectively. Π and κ_{el} can then be calculated via G, G_{th}, and K according to Eq. (I.11).

Both efficiency in Eq. (VII.2) and power, $P = IV$, are calculated numerically as a function of quantum dot energy, μ, and bias voltage, V. At each working point in the μ–V plain, values of η and P are paired and cataloged. The complete catalog is then plotted in the η–P plain. The resulting η–P phase space is only partially filled, and the filled regions place bounds on what efficiency and power the system can achieve. Efficiency and power calculated in this way are plotted in Fig. 7.2b at $\bar{T} = 1$ K for $\Gamma/k\bar{T} = 0.01$, 0.1, 1, and 10. For each choice of $\Gamma/k\bar{T}$, the quantum dot forms lobes of operation in the η–P plain. Note the tradeoff between efficiency and power, and that Γ dictates whether the device is disposed to efficiency or power, but not both.

The two dissimilar models were used to calculate η at $\bar{T} = 1$ K for a choice of $\Gamma/k\bar{T} = 0.1$ and 0.01, as shown in (a) and (b) of Fig. 7.2. Both models predict that, for these two choices of Γ, the system efficiencies are about 70% and 90% of Carnot efficiency, respectively. More important than the actual value of η is the fact that the two models agree. Calculating η directly from its definition is very trustworthy, and the power versus efficiency plots are interesting and pedagogical in their own right. However, this method requires much more computation time than calculating ZT directly. Furthermore, the η–P plots do not predict η as a function of quantum dot energy and thus cannot be easily compared to experiments. But because the two models agree, direct calculations of ZT and subsequent conversions to η are also
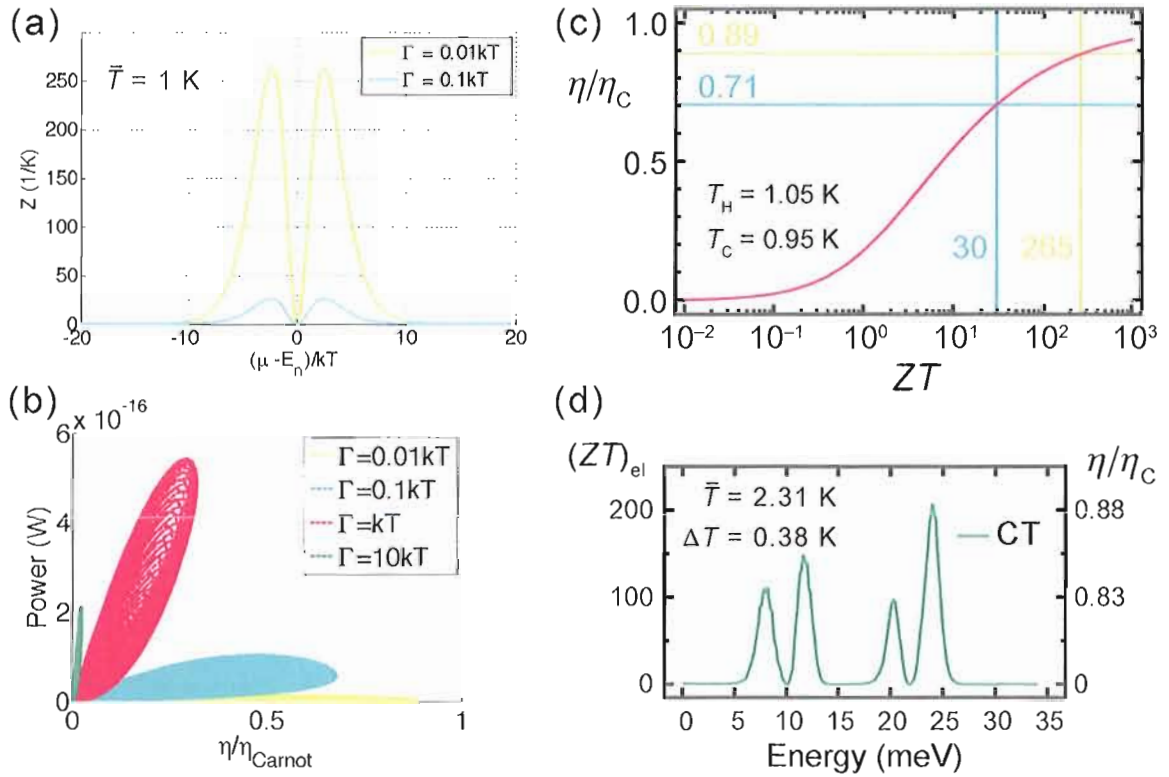
Figure 7.2. (a) $(ZT)_{el}$ calculated numerically as a function of quantum dot energy centered at a conductance peak. The simulation demonstrates very clearly that Γ improves performance. (b) Numerical simulations demonstrating that devices with different Γ cluster within different regions of power–efficiency phase space. The maximum efficiency calculated in this way agrees with the ZT values in (a) if ZT is converted to relative efficiency, η/η_C, using (c) η/η_C as a function of ZT. The vertical and horizontal lines show the conversion of the maximum ZT values in (a). (d) Measured $(ZT)_{el}$ as a function of energy over two resonances determined using L_{CT} represents the lower-bound on $(ZT)_{el}$. This data is very similar to the $\Gamma = 0.01kT$ simulation in (a).

trustworthy. In the same way, measured values of ZT can be considered an accurate and trustworthy measure of η.

Conclusions

The measured $(ZT)_{\text{el}}$ and resulting efficiency support the original theoretical proposition that quantum-dot-based devices provide ideal conditions for high-efficiency thermoelectric performance. The measured thermopower and temperature differences have been carefully collected and scrutinized insuring the reliability of the measured efficiency, and numerical simulations have corroborated these results. Therefore, the observed electronic efficiency between 88% and 95% Carnot efficiency can be trusted as accurate.

This research would not be complete without discussing the full ZT figure of merit of a quantum-dot device. The ZT figure of merit is related to $(ZT)_{\text{el}}$ by,

$$Z\bar{T} = \frac{S^2 G \bar{T}}{\kappa_{\text{el}} + \kappa_{\text{ph}}} = \frac{S^2 G \bar{T}}{\kappa_{\text{el}}} \frac{1}{1 + \kappa_{\text{ph}}/\kappa_{\text{el}}} = (ZT)_{\text{el}} \frac{1}{1 + \kappa_{\text{ph}}/\kappa_{\text{el}}}. \tag{VII.3}$$

Measurement of ZT requires knowledge of κ_{ph}. Collaborators at the University of Texas at Austin have measured κ_{ph} using homogeneous InAs nanowires grown at Lund University. Extrapolating their low-temperature results [104] down to the cryogenic temperatures used in these experiments gives $\kappa_{\text{ph}} \approx 1$ W/mK. This value is an upper-limit on the κ_{ph} of the double-barrier nanowires used in these experiments because,

as phonon modes freeze out, κ_{ph} will decrease faster than the linear extrapolation. The Wiedemann-Franz law provides $\kappa_{\mathrm{el}} = L_0 \sigma T$, where the electrical conductivity, σ, is used to provide unit agreement with the measured κ_{ph}. Using the measured conductance, nanowire dimensions, experimental temperature, and the cotunneling Lorenz number gives $\kappa_{\mathrm{el}} = 2.6 \times 10 \times^{-7}$ W/mK. Putting the measured values of κ_{el} and κ_{ph} into Eq. (VII.3) provides a maximum ZT of 3×10^{-4}. While this ZT value is low, it must be understood that these experiments aimed to optimize $(ZT)_{\mathrm{el}}$, which can actually decrease ZT.

The reduction in ZT stems from strong electron filtering without commensurate phonon filtering. The nanowires used in these experiments were intentionally grown with thick barriers to provide a very small Γ. As a direct result, both G and κ_{el} are small. However, this is of no consequence to $(ZT)_{\mathrm{el}}$ because it depends only on the ratio G/κ_{el}. On the other hand, the ZT figure of merit is sensitive to the ratio G/κ_{el} as well as the ratio $\kappa_{\mathrm{el}}/\kappa_{\mathrm{ph}}$, as demonstrated in Eq. (VII.3). In the limit where $\kappa_{\mathrm{el}}/\kappa_{\mathrm{ph}} \ll 1$, as is the case in these experiments, $ZT \approx (ZT)_{\mathrm{el}} (\kappa_{\mathrm{el}}/\kappa_{\mathrm{ph}})$ and is therefore small. Optimizing the electronic performance and ignoring the phonons has decreased $\kappa_{\mathrm{el}}/\kappa_{\mathrm{ph}}$ and therefore decreased ZT. A larger Γ could increase $\kappa_{\mathrm{el}}/\kappa_{\mathrm{ph}}$ without causing a significant decrease in S^2 thereby optimizing ZT rather than $(ZT)_{\mathrm{el}}$. The marriage of phonon suppression and strong electron filtering will produce a relatively large $\kappa_{\mathrm{el}}/\kappa_{\mathrm{ph}}$ value and result in a large ZT figure of merit. This result highlights the fundamental importance of merging insights regarding electronic

optimization with phonon suppression techniques in order to achieve a well-rounded, high-efficiency thermoelectric device.

CHAPTER VIII

CONCLUSIONS

Research Summary

Cyclic heat engines and thermoelectric heat engines are bounded by the same maximum efficiency, namely, Carnot efficiency, which requires a thermodynamically reversible process. The classic Carnot cycle achieves reversibility by assuming a perfectly insulated working gas and infinitely slow compression and expansion. Thermoelectric heat pumps suffer from a constant, entropy-producing, parasitic heat flow from the hot bath to the cold bath. Achieving maximal efficiency in such systems is accomplished in a fundamentally different fashion than in cyclic heat engines. The entropy produced by the constant heat flux flowing through a thermoelectric can be nulled by an equal and opposite entropy produced by charge flux. Such entropy cancellation is possible at only one specific energy, the so-called adiabatic energy [10]. Therefore, to realize (electronic) Carnot efficiency in a thermoelectric device, energy-specific electron transport must be achieved.

Quantum dots have an energy-filtering capability uniquely suited for realizing (near) Carnot efficiency. The goal of this Ph.D. research has been to measure quantitatively the electronic efficiency of a quantum dot operating as an energy filter

and demonstrate its ability to achieve near Carnot efficiency. Quantitative efficiency analysis required the application and subsequent measurement of a temperature difference across a distance of only 25 nm. This prerequisite was fulfilled by developing novel heating and thermometry techniques. The quantum-dot thermometry technique has proven particularly convenient because it measures electron temperature, and uses the same quantum dot that is used as the thermoelectric generator, thereby making additional thermometer components unnecessary. Together the heating and thermometry techniques have supplied and determined temperature differences of up to 2 K across a 25 nm distance, which is equivalent to an 8000 K temperature drop across the width of a human hair.

Besides electronic efficiency, other relevant quantum-dot thermoelectric behavior has been studied throughout the course of this research. Cotunneling has been and continues to be a troublesome topic in quantum dot transport as it threatens to decrease many performance aspects of these very sensitive devices, such as spin coherence times, quantum lifetimes, and thermoelectric efficiency. Thermovoltage oscillations offer insights into the quantitative behavior of cotunneling and how it affects thermally induced charge transport. The measured power of the quantum-dot nanowire used in this work (roughly 50 nm in diameter) was 5 fW at 3.76 K, which gives a power density of about 2.4 W/m^2. Note that this nanowire was optimized for efficiency, not power, and so this result must be interpreted as a lower-limit of quantum-dot nanowire power generation. Power production should be higher in an

appropriately designed device. Also, power is expected to increase quadratically with temperature. Future power research should bare out both of these predictions. Below 1 K, thermovoltage demonstrated strong nonlinear behavior as a function of ΔT, even when $\Delta T/T$ was as small as 5%. This result suggests that kT is not the relevant energy scale in the system and that instead quantum effects must be considered. The observed nonlinearity is not well understood and is still being investigated.

The electronic efficiency of direct thermal-to-electric energy conversion using a quantum-dot nanowire has been measured and is between 88% and 95% Carnot efficiency. This is the first quantitative efficiency measurement of quantum dots, and it demonstrates that quantum dots can indeed operate very near the fundamental upper-limit of all heat engines, that is, Carnot efficiency. Though a different thermoelectric device design might prove more advantageous than quantum dots, this result elucidates the fundamental solution to optimizing the electronic aspect of all thermoelectric heat engines.

Research Outlook

The number of possible thermal physics topics that can be studied using quantum-dot nanowires is large. The quantum-dot thermometry technique that developed from this research could be used to study a variety of fundamental physics topics concerning low-dimensional heat flow and electron-phonon interactions. The unique thermoelectric effects in quantum dots offer a myriad of possible research

avenues, and nanowire-based devices offer extra flexibility and broaden research horizons. For example, future research regarding the relationship between cotunneling and thermoelectric effects could provide a better understanding of cotunneling in general. The power measurements presented here are only a precursor. A complete study of power optimization and the relationship between power and efficiency is interesting from both a low-temperature physics and a device application perspective. The observed nonlinear thermoelectric behavior threatens to decrease device performance and is not completely understood. These experimental results offer a wonderful opportunity to bolster the theoretical understanding of subtle mesoscopic thermoelectrics effects and perhaps improve device design.

The electronic aspect of quantum dots is becoming well understood—supported by a strong foundation of research. These results will be well complemented by turning attention toward phonons. Presently, not everything is understood about the behavior of phonons in 1D, such as their density of states, interaction with confined electrons, and surface boundary effects. Especially relevant to the field of thermoelectrics are the topics of phonon drag, phononic heat conduction, electron-phonon interaction, and surface scattering. Phonon experiments are challenging, and novel experimental techniques must rise to the occasion. One possible platform for studying these topics is a suspended quantum-dot nanowire, which remains untried and could reveal insights regarding the influence of phonons.

Applications

Thermoelectric devices based on quantum-dot nanowires offer several practical advantages. First, the choice of semiconductor material used to make the quantum dot is irrelevant to its basic thermoelectric operation and ultimate efficiency so long as the quantum dot provides energy resonances separated by more than the background thermal energy. Second, once the device is in operation, gating can be used to toggle device operation between n-type and p-type, thereby offering two devices in one. Nanowires can be grown in arrays offering scalability through parallelization. The gating of many-nanowire arrays is important for several other nanowire applications and is presently under development [105]. Coating nanowires with HfO_2, a dielectric material that has gained a recent surge in popularity, creates an insulating shell on top of which a metallic gate can be deposited [53]. These so-called wrap gates [106, 107] offer the possibility to gate an entire array of nanowires.

The ultimate goal is to achieve high-efficiency operation at room temperature and above. Nanowire superlattices create energy minibands which are separated in energy by more than the room-temperature thermal energy. If the minibands are narrow enough, they can operate as narrow-band energy filters improving thermoelectric performance. Therefore, nanowire superlattice arrays have the potential to become efficient room-temperature thermoelectric devices.

From a phonon standpoint, nanowires have lower phononic thermal conductivity

than their bulk counterparts via the enhancement of phonon surface scattering owing to the large surface-to-volume ratio of nanowires. Furthermore, it is fundamentally possible to promote the InP semiconductor barriers, which define the quantum dot, to vacuum barriers thereby significantly suppressing phonon flow. Therefore, nanowires have phononic advantages in addition to their electronic advantages.

Going beyond nanowires in spatial reduction can be achieved using molecules, and thermoelectric measurements on molecular junctions have already been accomplished [108]. Although molecular materials cannot provide vastly better electron energy quantization or filtering than semiconductor-based quantum dots, small molecular dimensions offer a better packing fraction than nanowire arrays. Thus, a relatively dense molecular array could have a better thermoelectric power density than even the most-dense nanowire array. The Holy Grail in molecular thermoelectrics is a self-assembled, dense, 2D molecular array of arbitrary size. Such massive parallelization could create a very worthwhile device. Because it is universal to all single-energy thermoelectric processes, the research presented here on quantum-dot-based thermoelectrics can help optimize the electronic aspect of molecular systems as well.

APPENDIX A

DIFFERENTIAL CONDUCTANCE DERIVATION

Starting from the Landauer equation (Eq. (I.20)) the differential conductance is

$$
\begin{aligned}
G = \frac{\partial I}{\partial V} &= -\frac{2e}{h} \int_{-\infty}^{\infty} \frac{\partial}{\partial V} \left[f_{\mathrm{H}} - f_{\mathrm{C}} \right] \tau\left(\varepsilon\right) d\varepsilon \\
&= -\frac{2e}{h} \int_{-\infty}^{\infty} \left(\frac{\partial \xi_{\mathrm{H}}}{\partial V} \frac{\partial f_{\mathrm{H}}}{\partial \xi_{\mathrm{H}}} - \frac{\partial \xi_{\mathrm{C}}}{\partial V} \frac{\partial f_{\mathrm{C}}}{\partial \xi_{\mathrm{C}}} \right) \tau\left(\varepsilon\right) d\varepsilon \\
&= -\frac{2e}{h} \int_{-\infty}^{\infty} \frac{e}{kT} \left(\frac{1}{2} \frac{\partial f_{\mathrm{H}}}{\partial \xi_{\mathrm{H}}} + \frac{1}{2} \frac{\partial f_{\mathrm{C}}}{\partial \xi_{\mathrm{C}}} \right) \tau\left(\varepsilon\right) d\varepsilon \\
&= -\frac{1}{kT} \frac{2e^2}{h} \int_{-\infty}^{\infty} \left(\frac{1}{2} \frac{\partial f_{\mathrm{H}}}{\partial \xi_{\mathrm{H}}} + \frac{1}{2} \frac{\partial f_{\mathrm{C}}}{\partial \xi_{\mathrm{C}}} \right) \tau\left(\varepsilon\right) d\varepsilon.
\end{aligned}
$$

In the limit of $\Gamma \ll kT$, the transmission function $\tau\left(\varepsilon\right)$ (Eq. (II.7)) takes the form of a Dirac-delta function, $\tau\left(\varepsilon\right) = E_0 \delta\left(\varepsilon - \varepsilon_0\right)$, where E_0 is its energetic "strength", and G reduces to

$$
\begin{aligned}
G &= -\frac{1}{kT} \frac{2e^2}{h} \int_{-\infty}^{\infty} \left(\frac{1}{2} \frac{\partial f_{\mathrm{H}}}{\partial \xi_{\mathrm{H}}} + \frac{1}{2} \frac{\partial f_{\mathrm{C}}}{\partial \xi_{\mathrm{C}}} \right) E_0 \delta\left(\varepsilon - \varepsilon_0\right) d\varepsilon \\
&= -\frac{1}{kT} \frac{e^2}{h} E_0 \left(\left. \frac{\partial f_{\mathrm{H}}}{\partial \xi_{\mathrm{H}}} \right|_{\varepsilon = \varepsilon_0} + \left. \frac{\partial f_{\mathrm{C}}}{\partial \xi_{\mathrm{C}}} \right|_{\varepsilon = \varepsilon_0} \right).
\end{aligned}
$$

Because E_0 is unknown, the overall prefactor can be condensed into a single coefficient A, which will ultimately become a fit parameter:

$$G = -A \left(\left. \frac{\partial f_{\mathrm{H}}}{\partial \xi_{\mathrm{H}}} \right|_{\varepsilon=\varepsilon_0} + \left. \frac{\partial f_{\mathrm{C}}}{\partial \xi_{\mathrm{C}}} \right|_{\varepsilon=\varepsilon_0} \right). \tag{A.1}$$

Assuming $eV \ll kT$, the above equation can be Taylor expanded. To improve accuracy, the expansion will be carried out to second-order in V. Expanding the derivatives gives

$$
\begin{aligned}
\frac{\partial f_{\mathrm{H,C}}}{\partial \xi_{\mathrm{H,C}}} &\cong \left. \frac{\partial f_{\mathrm{H,C}}}{\partial \xi_{\mathrm{H,C}}} \right|_{V=0} + \left. \frac{\partial}{\partial V} \frac{\partial f_{\mathrm{H,C}}}{\partial \xi_{\mathrm{H,C}}} \right|_{V=0} V + \frac{1}{2} \left. \frac{\partial^2}{\partial V^2} \frac{\partial f_{\mathrm{H,C}}}{\partial \xi_{\mathrm{H,C}}} \right|_{V=0} V^2 \\
&= \left. \frac{\partial f_{\mathrm{H,C}}}{\partial \xi_{\mathrm{H,C}}} \right|_{V=0} + \left. \frac{\partial \xi_{\mathrm{H,C}}}{\partial V} \frac{\partial^2 f_{\mathrm{H,C}}}{\partial \xi_{\mathrm{H,C}}^2} \right|_{V=0} V + \frac{1}{2} \left(\frac{\partial \xi_{\mathrm{H,C}}}{\partial V} \right)^2 \left. \frac{\partial^2}{\partial \xi_{\mathrm{H,C}}^2} \frac{\partial f_{\mathrm{H,C}}}{\partial \xi_{\mathrm{H,C}}} \right|_{V=0} V^2 \\
&= \left. \frac{\partial f_{\mathrm{H,C}}}{\partial \xi_{\mathrm{H,C}}} \right|_{V=0} - \frac{\pm e}{2kT} \left. \frac{\partial^2 f_{\mathrm{H,C}}}{\partial \xi_{\mathrm{H,C}}^2} \right|_{V=0} V + \frac{1}{2} \left(\frac{\partial \xi_{\mathrm{H,C}}}{\partial V} \right)^2 \left. \frac{\partial^3 f_{\mathrm{H,C}}}{\partial \xi_{\mathrm{H,C}}^3} \right|_{V=0} V^2. \tag{A.2}
\end{aligned}
$$

Now each term can be handled separately. The zeroth-order term is

$$\left. \frac{\partial f_{\mathrm{H,C}}}{\partial \xi_{\mathrm{H,C}}} \right|_{V=0} = \left. \frac{e^{\xi_{\mathrm{H,C}}}}{\left(1 + e^{\xi_{\mathrm{H,C}}}\right)^2} \right|_{V=0}. \tag{A.3}$$

Introducing the new term (using Eq. (I.15)) $\xi \equiv \xi_H|_{V=0} = \xi_C|_{V=0} = (\varepsilon - \mu)/kT$, the

above simplifies to

$$
\begin{aligned}
\left.\frac{\partial f_{H,C}}{\partial \xi_{H,C}}\right|_{V=0} &= -\frac{e^\xi}{\left(1 + e^\xi\right)^2} \\
&= -\frac{1}{\left(e^{-\xi/2}\right)^2 \left(1 + e^\xi\right)^2} \\
&= -\frac{1}{\left(e^{\xi/2} + e^{-\xi/2}\right)^2} \\
&= -\frac{1}{4}\frac{1}{\cosh^2\left(\xi/2\right)}.
\end{aligned}
$$

The first-order coefficient is

$$
\begin{aligned}
\left.\frac{\pm e}{2kT}\frac{\partial^2 f_{H,C}}{\partial \xi_{H,C}^2}\right|_{V=0} &= \left.\frac{\mp e}{2kT}\frac{\partial}{\partial \xi_{H,C}}\frac{e^{\xi_{H,C}}}{\left(1 + e^{\xi_{H,C}}\right)^2}\right|_{V=0} \\
&= \left.\frac{\mp e}{2kT}\left(\frac{e^{\xi_{H,C}}}{\left(1 + e^{\xi_{H,C}}\right)^2} - 2\frac{e^{2\xi_{H,C}}}{\left(1 + e^{\xi_{H,C}}\right)^3}\right)\right|_{V=0} \\
&= \left.\frac{\mp e}{2kT}\frac{e^{\xi_{H,C}}\left(1 + e^{\xi_{H,C}}\right) - 2e^{2\xi_{H,C}}}{\left(1 + e^{\xi_{H,C}}\right)^3}\right|_{V=0} \\
&= \left.\frac{\mp e}{2kT}\frac{e^{\xi_{H,C}} - e^{2\xi_{H,C}}}{\left(1 + e^{\xi_{H,C}}\right)^3}\right|_{V=0} \qquad\qquad (A.4) \\
&= \mp\frac{eV}{2kT}\frac{e^\xi - e^{2\xi}}{\left(1 + e^\xi\right)^3} \qquad\qquad\qquad (A.5)
\end{aligned}
$$

The second-order coefficient is

$$
\left.\frac{1}{2}\left(\frac{\partial \xi_{H,C}}{\partial V}\right)^2 \frac{\partial^3 f_{H,C}}{\partial \xi_{H,C}^3}\right|_{V=0} = \left.-\frac{1}{2}\left(\frac{e}{2kT}\right)^2 \frac{\partial_{H,C}}{\partial \xi_{H,C}}\left(\frac{e^{\xi_{H,C}} - e^{2\xi_{H,C}}}{\left(1 + e^{\xi_{H,C}}\right)^3}\right)\right|_{V=0},
$$

where Eq. (A.4) has been used to compute $\partial^3 f_{H,C}/\partial \xi_{H,C}^3$. Therefore,

$$
\begin{aligned}
\frac{1}{2} &\left(\frac{\partial \xi_{H,C}}{\partial V}\right)^2 \left.\frac{\partial^3 f_{H,C}}{\partial \xi_{H,C}^3}\right|_{V=0} \\
&= -\frac{e^2}{8k^2T^2} \left.\left(\frac{e^{\xi_{H,C}} - 2e^{2\xi_{H,C}}}{(1+e^{\xi_{H,C}})^3} - 3\frac{e^{\xi_{H,C}}\left(e^{\xi_{H,C}} - e^{2\xi_{H,C}}\right)}{(1+e^{\xi_{H,C}})^4}\right)\right|_{V=0} \\
&= -\frac{e^2}{8k^2T^2} \left.\frac{\left(e^{\xi_{H,C}} - 2e^{2\xi_{H,C}}\right)\left(1+e^{\xi_{H,C}}\right) - 3e^{\xi_{H,C}}\left(e^{\xi_{H,C}} - e^{2\xi_{H,C}}\right)}{(1+e^{\xi_{H,C}})^4}\right|_{V=0} \\
&= -\frac{e^2}{8k^2T^2} e^{\xi} \frac{1 - e^{\xi} - 2e^{2\xi} - 3e^{\xi} + 3e^{2\xi}}{(1+e^{\xi})^4} \\
&= -\frac{e^2}{8k^2T^2} e^{\xi} \frac{1 - 4e^{\xi} + e^{2\xi}}{(1+e^{\xi})^4} \\
&= -\frac{e^2}{8k^2T^2} e^{-\xi} \frac{1 - 4e^{\xi} + e^{2\xi}}{(e^{-\xi/2})^4 (1+e^{\xi})^4} \\
&= -\frac{e^2}{8k^2T^2} \frac{e^{\xi} + e^{-\xi} - 4}{(e^{\xi/2} + e^{-\xi/2})^4} \\
&= -\frac{e^2}{8k^2T^2} \frac{2\cosh \xi - 4}{2^4 \cosh^4 (\xi/2)} \\
&= -\frac{e^2}{64k^2T^2} \frac{\cosh \xi + 1 - 3}{\cosh^4 (\xi/2)} \\
&= -\frac{e^2}{64k^2T^2} \frac{2\cosh^2 (\xi/2) - 3}{\cosh^4 (\xi/2)}
\end{aligned}
\tag{A.6}
$$

Putting the coefficients (Eqs. (A.3), (A.5), and (A.6) into theTaylor series (Eq. (A.2)) gives,

$$
\frac{\partial f_{H,C}}{\partial \xi_{H,C}} = -\frac{1}{4}\text{sech}^2 (\xi/2) \mp \frac{e}{2kT}\frac{e^{\xi} - e^{2\xi}}{(1+e^{\xi})^3}V + \frac{e^2}{64k^2T^2}\frac{2\cosh^2 (\xi/2) - 3}{\cosh^4 (\xi/2)}V^2.
$$

The differential conductance is therefore

$$
\begin{aligned}
G &= -A \left(\frac{\partial f_{\mathrm{H}}}{\partial \xi_{\mathrm{H}}} + \frac{\partial f_{\mathrm{C}}}{\partial \xi_{\mathrm{C}}} \right) \\
&= A \frac{1}{2} \frac{1}{\cosh^2 (\xi/2)} + A \frac{e^2}{32 k^2 T^2} \frac{2 \cosh^2 (\xi/2) - 3}{\cosh^4 (\xi/2)} V^2 .
\end{aligned}
$$

The first-order terms cancel, which suppresses the effect of the bias voltage on the resulting differential conductance peak width, because the lowest-order, non-zero eV/kT term is quadratic. The resulting differential conductance simplifies to

$$
\begin{aligned}
G &= A \frac{1}{\cosh^2 (\xi/2)} \left[1 + \frac{e^2}{16 k^2 T^2} \frac{2 \cosh^2 (\xi/2) - 3}{\cosh^2 (\xi/2)} V^2 \right] \\
&= A \frac{1}{\cosh^2 (\xi/2)} \left[1 + \frac{e^2}{16 k^2 T^2} \left(2 - \frac{3}{\cosh^2 (\xi/2)} \right) V^2 \right] \\
&= A \operatorname{sech}^2 (\xi/2) \left[1 + \left(\frac{eV}{4kT} \right)^2 \left(2 - 3 \operatorname{sech}^2 (\xi/2) \right) \right] .
\end{aligned}
$$

Expressing this in terms of gate voltage gives the final expression

$$
G = A \operatorname{sech}^2 \left(\frac{e \alpha (V_{\mathrm{G}} - V_0)}{2kT} \right) \left[1 + \left(\frac{eV}{4kT} \right)^2 \left(2 - 3 \operatorname{sech}^2 \left(\frac{e \alpha (V_{\mathrm{G}} - V_0)}{2kT} \right) \right) \right] ,
$$

$$(A.7)$$

where $V_0 = \varepsilon_0 / \alpha e$ is the gate voltage at which the differential conductance peak is centered.

APPENDIX B

OP-AMP HEATING CIRCUIT DIAGRAM
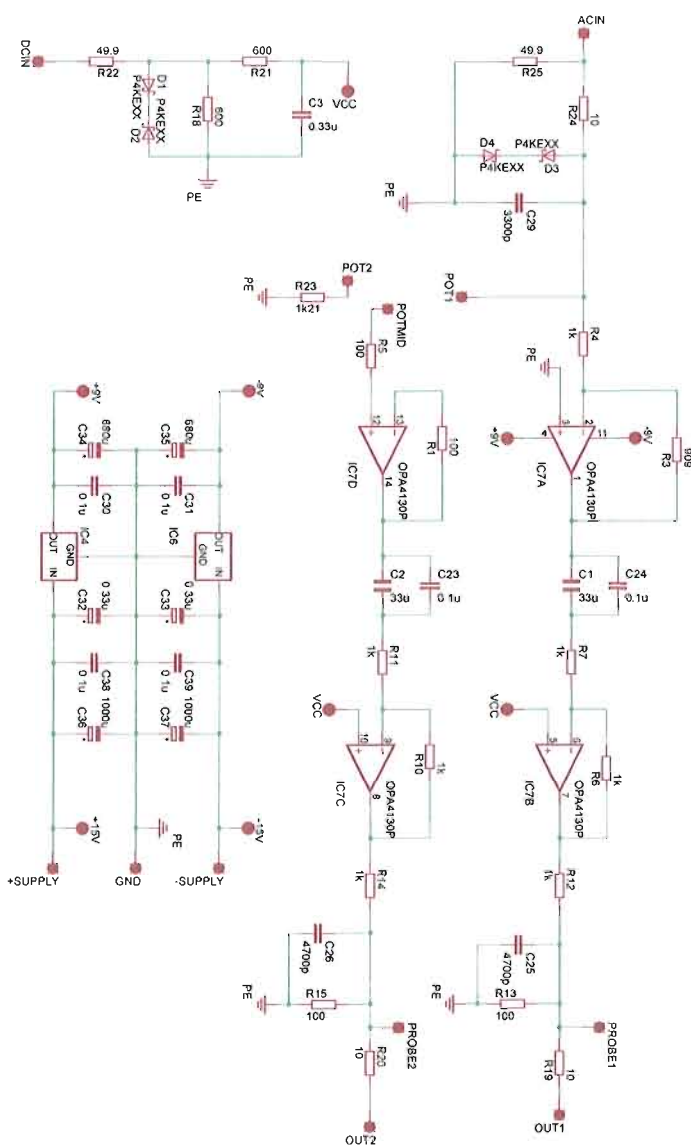
Figure B.1.. The schematic of the op-amp heating circuit.

APPENDIX C

THEORY OF COULOMB-BLOCKADED OPEN-CIRCUIT VOLTAGE

The intention of this Appendix is to derived the slope of the open-circuit voltage, V_{OC}, as it modulates linearly as a function of gate voltage, V_{G}. In this discussion, the source contact is connected to ground making its chemical potential, μ_{S}, and voltage, V_{S}, constant, that is, $\mu_{\mathrm{S}} = -eV_{\mathrm{S}}$. On the other hand, the drain contact is floating and is influenced by the gate voltage, V_{G}, and the addition or removal of charge from ground via the quantum dot and source contact. This is not a thermal effect. In both the theory and the experiments the source and drain contacts have the same temperature. The physical reasoning in the following mathematics is gleaned from Fig. 2.15. The conclusion of this Appendix is that the respective uphill and downhill slopes of V_{OC} as a function of V_{G} (that is, $dV_{\mathrm{OC}}/dV_{\mathrm{G}}$) are the same as the source and drain slopes, S_{S} and S_{D}, of the Coulomb-blockade diamonds given by Eqs. (II.10) and (II.11), provided that $C_{\mathrm{S}} = C_{\mathrm{D}}$, as is the case if the tunnel barriers are the same.

Zero Thermovoltage

When the open-circuit voltage is zero, $V_{\mathrm{OC}} = V_{\mathrm{D}} - V_{\mathrm{S}} = 0$, and therefore $V_{\mathrm{D}} = V_{\mathrm{S}}$. The resonance of the quantum dot is aligned to this energy, that is, $\mu = -eV_{\mathrm{S}} = -eV_{\mathrm{D}}$, because the dot always follows either the source or drain electrochemical potential as

charge carriers flow left and right through the quantum dot maintaining equilibrium. This is shown in panels 1 and 5 of Fig. 2.15. According to Eq. (II.1), under Coulomb blockade it is true that,

$$\mu\left(N\right) = -eV_{\mathrm{S}} = \left(N - N_0 - \frac{1}{2}\right)\frac{e^2}{C_\Sigma} - \frac{e}{C_\Sigma}\left(C_{\mathrm{S}}V_{\mathrm{S}} + C_{\mathrm{D}}V_{\mathrm{S}} + C_{\mathrm{G}}V_{\mathrm{G}}\right) + E_N.$$

Multiplying by C_Σ/eC_{G} gives,

$$-\frac{C_\Sigma}{C_{\mathrm{G}}}V_{\mathrm{S}} = \left(N - N_0 - \frac{1}{2}\right)\frac{e}{C_{\mathrm{G}}} - \frac{1}{C_{\mathrm{G}}}\left(C_{\mathrm{S}}V_{\mathrm{S}} + C_{\mathrm{D}}V_{\mathrm{S}} + C_{\mathrm{G}}V_{\mathrm{G}}\right) + \frac{C_\Sigma E_N}{eC_{\mathrm{G}}}$$

$$\frac{C_{\mathrm{S}} + C_{\mathrm{D}} - C_\Sigma}{C_{\mathrm{G}}}V_{\mathrm{S}} + V_{\mathrm{G}} = \left(N - N_0 - \frac{1}{2}\right)\frac{e}{C_{\mathrm{G}}} + \frac{C_\Sigma E_N}{eC_{\mathrm{G}}}$$

$$-V_{\mathrm{S}} + V_{\mathrm{G}} = \left(N - N_0 - \frac{1}{2}\right)\frac{e}{C_{\mathrm{G}}} + \frac{C_\Sigma E_N}{eC_{\mathrm{G}}}$$

$$V_{\mathrm{G}} = V_{\mathrm{S}} + V_N,$$

where V_N is the gate voltage at which the Nth resonance occurs if $V_{\mathrm{S}} = V_{\mathrm{D}} = 0$ and is given by,

$$V_N = \left(N - N_0 - \frac{1}{2}\right)\frac{e}{C_{\mathrm{G}}} + \frac{C_\Sigma E_N}{eC_{\mathrm{G}}}.$$

Therefore, the effect of the voltage V_{S} at the source contact is to shift the position of the resonance energies relative to V_N. So, without loss of generality, the source voltage can be set to zero. For the remainder of this Appendix it is implicitly assumed that $\mu_{\mathrm{S}} = -eV_{\mathrm{S}} = 0$.

Increasing Thermovoltage

When the open-circuit voltage is increasing, the resonance of the quantum dot, μ, moves together with the drain electrochemical potential, $\mu_D = eV_D$, so that $-eV_D = \mu$. This is shown in panel 2 of Fig. 2.15. Putting this into Eq. (II.1) gives

$$-eV_D = \left(N - \frac{1}{2}\right)\frac{e^2}{C_\Sigma} - \frac{e}{C_\Sigma}(C_D V_D + C_G V_G) + E_N$$

$$-\frac{C_\Sigma}{C_G}V_D = \left(N - \frac{1}{2}\right)\frac{e}{C_G} - \frac{1}{C_G}(C_D V_D + C_G V_G) + \frac{C_\Sigma E_N}{eC_G}$$

$$\frac{C_D - C_\Sigma}{C_G}V_D = \left(N - \frac{1}{2}\right)\frac{e}{C_G} + \frac{C_\Sigma E_N}{eC_G} - V_G$$

$$\frac{C_G + C_S}{C_G}V_D = V_N - V_G$$

$$V_D = \frac{C_G}{C_G + C_S}\left(V_G - V_N\right).$$

Therefore, the uphill slope of the open-circuit voltage, $\delta V_{OC}^+ = dV_{OC}/dV_G$, can be written

$$\delta V_{OC}^+ = \frac{dV_{OC}}{dV_G} = \frac{dV_D - dV_S}{dV_G} = \frac{dV_D}{dV_G} = \frac{C_G}{C_G + C_S}. \tag{C.1}$$

Note that this is the same as the source slope of the Coulomb-blockade diamond (Eq. (II.10)) if $C_S = C_D$. That is, $C_G/(C_G + C_S) = C_G/(C_G + C_D) = S_S^{-1}$. (Note that $S_S = dV_G/dV$ in Eq. (II.10) is defined in a reciprocal manner to the slope here.) When $C_G \ll C_D$, which is typical for a $\sim$10 nm quantum dot in a $\sim$50 nm

nanowire system [62], the uphill slope can be approximated to first-order in $C_\mathrm{G}/C_\mathrm{S}$ as $dV_\mathrm{OC}/dV_\mathrm{G} \approx C_\mathrm{G}/C_\mathrm{S}$. This provides a nice symmetry with Eq. (C.2).

Maximum Thermovoltage

Note that the maximum V_OC occurs when the source and drain electrochemical potentials are aligned with two adjacent energy resonances. This is simply because the lowering of $\mu(N)$ and μ_D stops when the $(N+1)$th energy resonance aligns with μ_S. This is shown in panel 3 of Fig. 2.15. Therefore $\mu(N) = -eV_\mathrm{D}$ and $\mu(N+1) = -eV_\mathrm{S}$ making the open-circuit voltage

$$
\begin{aligned}
eV_\mathrm{OC} &= eV_\mathrm{D} - eV_\mathrm{S} \\
&= -\mu(N) - (-)\mu(N+1) \\
&= \mu(N+1) - \mu(N) \\
&= E_\mathrm{add}(N+1) \\
&= E_\mathrm{C} + \Delta E(N+1),
\end{aligned}
$$

where the last equality follows from Eq. (II.2). Therefore, the peak value of V_OC is an alternative way to measure $E_\mathrm{add}(N+1)$ and thus E_C.

Decreasing Thermovoltage

When the open-circuit voltage is decreasing, the quantum dot follows the source

electrochemical potential, as shown in panel 4 of Fig. 2.15. Mathematically, $\mu = -eV_S = 0$ and from Eq. (II.1),

$$0 = \left(N - \frac{1}{2}\right)\frac{e^2}{C_\Sigma} - \frac{e}{C_\Sigma}\left(C_D V_D + C_G V_G\right) + E_N$$

$$0 = \left(N - \frac{1}{2}\right)\frac{e}{C_G} - \frac{1}{C_G}\left(C_D V_D + C_G V_G\right) + \frac{C_\Sigma E_N}{eC_G} = 0$$

$$\frac{C_D}{C_G}V_D = \left(N - \frac{1}{2}\right)\frac{e}{C_G} + \frac{C_\Sigma E_N}{eC_G} - V_G$$

$$\frac{C_D}{C_G}V_D = V_N - V_G$$

$$V_D = -\frac{C_G}{C_D}\left(V_G - V_N\right).$$

Therefore, the downhill slope of the open-circuit voltage, $\delta V_{OC}^- = dV_{OC}/dV_G$, can be written

$$\delta V_{OC}^- = \frac{dV_{OC}}{dV_G} = \frac{dV_D - dV_S}{dV_G} = \frac{dV_D}{dV_G} = -\frac{C_G}{C_D}. \tag{C.2}$$

If the source and drain are symmetric, then $C_D = C_S$, and this is the same as the drain slope of the Coulomb-blockade diamond (Eq. (II.11)). That is $dV_{OC}/dV_G = -C_G/C_S = S_D^{-1}$. ($S_D$, like S_S, is defined in the reciprocal way.) Note that if the Coulomb-blockade diamonds are symmetric, $C_D - C_S = C_G$ and the uphill and downhill slopes in Eqs. (C.1) and (C.2) are equal.

Measuring α

V_{OC} data carries information regarding α, the system's lever arm, because δV_{OC}^+

and $\delta V_{\mathrm{OC}}^{-}$ are equivalent to the slopes of the Coulomb-blockade diamonds. In particular,

$$\left|\delta V_{\mathrm{OC}}^{+}\right|^{-1} + \left|\delta V_{\mathrm{OC}}^{-}\right|^{-1} = \frac{C_{\Sigma} - C_{\mathrm{D}}}{C_{\mathrm{G}}} + \frac{C_{\mathrm{D}}}{C_{\mathrm{G}}} = \frac{C_{\Sigma}}{C_{\mathrm{G}}} = \alpha^{-1}, \tag{C.3}$$

where the last equality follows from Eq. (II.3). Therefore, the slopes of V_{OC} data provide an alternative to Coulomb-blockade diamonds for measuring α.

APPENDIX D

NUMERICAL DERIVATIVE ROUTINE

Outlined here is a prescription for finding the derivative $df/dx = f'$ given a series of equally spaced data points representing the function $f(x)$. If the data is spaced in the variable x by and amount h, the goal is to express $f'(x)$ at the point x by

$$f'(x) = \frac{1}{h}\left[\sum_{m=1}^{M} c_{-m} f(x - mh) + c_0 + \sum_{n=1}^{N} c_n f(x + nh)\right] \qquad (D.1)$$

In this mathematical form, the problem reduces to finding the "backward" and "forward" coefficients, c_{-m} and c_n, respectively. This method is known as the method of undetermined coefficients [109]. Note that in general $M \neq N$ because the expansion need not be symmetric. For example, it is not possible to look backward beyond the first data point of the data series, and M must be zero to make the numerical derivative at the first data point completely forward looking. Similarly, at the last data point in the series, N must be zero and the formula is backward looking.

The most familiar (symmetric) derivative formula is the 2-point formula,

$$f'(x) = \frac{1}{2h}\left[f(x + h) - f(x - h)\right], \qquad (D.2)$$

making $c_{-1} = -1/2$, $c_1 = 1/2$ and all other $c_i = 0$. The problem with the above

formula is that it only uses two data points, making no attempt to sample more data and possibly additional curvature (see Fig. D.1). Sampling more data can also help mitigate experimental noise.

Finding the $M+N+1$ unknown coefficients, c_i, requires $M+N+1$ linear equations. This is accomplished by using $M + N + 1$ test functions of the form

$$f_{m+n}\left(x\right) = x^{m+n}.$$

The true derivative of each $f_{m+n}\left(x\right)$ is equated to each $f'_{m+n}\left(x\right)$ predicted by Eq. (D.1) providing $M + N + 1$ linear equations. For convenience, all these derivatives are eventually evaluated at $x = 0$. Note this same technique can be used to find second-order (or higher) derivative formulas as well.

The formula derived here is the "symmetric-7" with three forward and three backward coefficients. In this case, when $M = N = 3$, then the $M + N + 1 = 7$ test functions are

$$
\begin{aligned}
f_0\left(x\right) &= 1 \\
f_1\left(x\right) &= x \\
f_2\left(x\right) &= x^2 \\
&\vdots \\
f_6\left(x\right) &= x^6.
\end{aligned}
$$

The vector of true derivatives evaluated at $x = 0$ are

$$\vec{F} = \left. \frac{df_i(x)}{dx} \right|_{x=0} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{D.3}$$

Writing all the derivatives $f'(x)$ predicted by Eq. (D.1) is too tedious. As an example, $f'_1(x)|_{x=0}$ is

$$f'_1(x)|_{x=0} = -3c_{-3} - 2c_{-2} - c_{-1} + c_1 + 2c_2 + 3c_3$$

For brevity, all these functions can be written in matrix form $f'_i(x)|_{x=0} = \mathbf{A}\vec{c}$, where

$$\vec{c} = \begin{bmatrix} c_{-3} \\ c_{-2} \\ c_{-1} \\ c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

and $\mathbf{A}$ is a 7×7 matrix of coefficients.

Now all 7 numerically predicted derivatives are equated to the true derivatives, $f'_i(x)|_{x=0} =$

$df_i(x)/dx|_{x=0}$ .This result is most easily written in matrix notation, $\mathbf{A}\vec{c} = \vec{F}$,

$$
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\
-3 & -2 & -1 & 0 & 1 & 2 & 3 & 1 \\
9 & 4 & 1 & 0 & 1 & 4 & 9 & 0 \\
-27 & -8 & -1 & 0 & 1 & 8 & 27 & 0 \\
81 & 16 & 1 & 0 & 1 & 16 & 81 & 0 \\
-243 & -32 & -1 & 0 & 1 & 32 & 243 & 0 \\
729 & 64 & 1 & 0 & 1 & 64 & 729 & 0
\end{bmatrix}
\begin{bmatrix}
c_{-3} \\
c_{-2} \\
c_{-1} \\
c_0 \\
c_1 \\
c_2 \\
c_3
\end{bmatrix}
=
\begin{bmatrix}
0 \\
1 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}.
$$

Note that all the h^i terms conveniently divide out.

The unknown coefficients $\vec{c}$ are found by multiplying $\mathbf{A}\vec{c} = \vec{F}$ from the left with the inverse matrix $\mathbf{A}^{-1}$,

$$
\vec{c} = \mathbf{A}^{-1}\mathbf{A}\vec{c} = \mathbf{A}^{-1}\vec{F} =
\begin{bmatrix}
-1/60 \\
3/20 \\
-3/4 \\
0 \\
3/4 \\
-3/20 \\
1/60
\end{bmatrix}.
$$

Therefore, the "symmetric-7" derivative formula is

$$
f'(x) = \frac{3}{h}\left[-\frac{f(x-3h)}{180} + \frac{f(x-2h)}{20} - \frac{f(x-h)}{4} \right.
$$
$$
\left. + \frac{f(x+h)}{4} - \frac{f(x+2h)}{20} + \frac{f(x+3h)}{180} \right]. \quad (\text{D.4})
$$

This same technique can be used to find the "forward-7" formula

$$f'(x) = \frac{1}{h}\left[-\frac{6727}{6060}f(x) + \frac{84}{101}f(x+h) + \frac{120}{101}f(x+2h)\right.$$
$$\left. -\frac{520}{303}f(x+3h) + \frac{465}{404}f(x+4h) - \frac{204}{505}f(x+5h) + \frac{6}{101}f(x+6h)\right],$$

and the "backward-7" formula

$$f'(x) = -\frac{1}{h}\left[-\frac{6727}{6060}f(x) + \frac{84}{101}f(x-h) + \frac{120}{101}f(x-2h)\right.$$
$$\left. -\frac{520}{303}f(x-3h) + \frac{465}{404}f(x-4h) - \frac{204}{505}f(x-5h) + \frac{6}{101}f(x-6h)\right].$$

Fig. D.1 compares the performance of the above 7-point formulas to that of the 2-point formulas, that is, Eq. D.2 and its associated backward and forward derivatives.
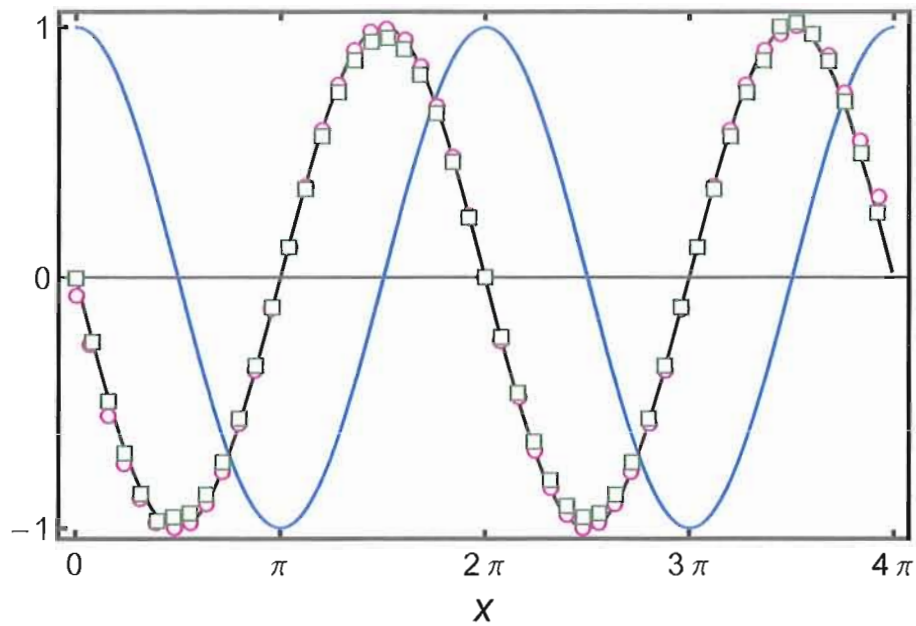
Figure D.1. The function $\cos x$ (blue line) was used to create 50 data points (not shown). The magenta circles and green squares were calculated with the 7-point derivative formulas (for example, Eq. (D.4)) and the 2-point derivative formulas (for example, Eq. (D.2)), respectively. The numerical derivatives are compared to the expected result, that is, $-\sin x$ (black line). The 2-point formula is accurate within 4% while the 7-point formula is accurate within 0.01%. In particular, the 7-point formula outperforms the 2-point formula whenever $\cos x$ approaches zero, which is where curvature is less pronounced and where quantum-dot thermometry is performed (see Section III.3).

APPENDIX E

SLOPE OF THE SEEBECK COEFFICIENT VERSUS ENERGY

The intent of this Appendix is to derive the slope of the Seebeck coefficient near a transmission resonance as a function of quantum dot energy, dS/dE. Two derivations are presented. The first derivation is based on the constant interaction model (CIM) (see Section II.3.1) and the second is based on the Mott formula as derived from the Onsager relations and Landauer equation (see Section I.3.1). Both derivations assume that the width of the transmission resonance, Γ, is very thin, that is, $\Gamma \ll kT$. Therefore, the derived slopes represent a theoretical upper-limit of dS/dE.

Seebeck Slope from the Constant Interaction Model

In isothermal open-circuit voltage measurements the system finds its equilibrium by aligning a source or drain electrochemical potential to a quantum-dot energy resonance (see Appendix C). Under the presence of a temperature gradient, the system no longer finds its equilibrium in this way, but rather by tuning the hot source electrochemical and cold drain electrochemical potential so that their Fermi-Dirac distributions cross at a quantum-dot energy resonance (see Fig. 4.1). The energy at which the two Fermi-Dirac distributions cross is the adiabatic energy, E_A,

given by Eq. (I.19),

$$E_{\mathrm{A}} \equiv \frac{\mu_{\mathrm{C}} T_{\mathrm{H}} - \mu_{\mathrm{H}} T_{\mathrm{C}}}{T_{\mathrm{H}} - T_{\mathrm{C}}}. \tag{E.1}$$

According to Eq. (II.1), if the Nth quantum-dot resonance aligns with E_{A}, then it is true that

$$\mu\left(N\right) = E_{\mathrm{A}} = \left(N - N_0 - \frac{1}{2}\right) \frac{e^2}{C_\Sigma} - \frac{e}{C_\Sigma} \left(C_{\mathrm{S}} V_{\mathrm{S}} + C_{\mathrm{D}} V_{\mathrm{D}} + C_{\mathrm{G}} V_{\mathrm{G}}\right) + E_N. \tag{E.2}$$

Note this is only strictly true if the Lorentzian transmission function of the resonance is infinitely thin, that is, if it is a Dirac-delta function. The equilibrium condition with a broadened transmission function is satisfied when the integrated current sums to zero requiring the Fermi-Dirac distributions to move off the center of the resonance slightly, which is why this derivation is only an upper limit on dS/dE.

The two InP barriers are assumed to be symmetric so that the capacitances are $C_{\mathrm{S}} = C_{\mathrm{D}} = C_{\mathrm{B}}$, where C_{B} is the capacitance of a single barrier. Substituting these terms into and multiplying Eq. (E.2) by C_Σ/eC_{G} gives,

$$
\begin{aligned}
\frac{C_\Sigma}{eC_{\mathrm{G}}} E_{\mathrm{A}} &= \left(N - N_0 - \frac{1}{2}\right) \frac{e}{C_{\mathrm{G}}} - \frac{1}{C_{\mathrm{G}}} \left(C_{\mathrm{B}} V_{\mathrm{S}} + C_{\mathrm{B}} V_{\mathrm{D}} + C_{\mathrm{G}} V_{\mathrm{G}}\right) + \frac{C_\Sigma E_N}{eC_{\mathrm{G}}} \\
&= \left(N - N_0 - \frac{1}{2}\right) \frac{e}{C_{\mathrm{G}}} + \frac{C_\Sigma E_N}{eC_{\mathrm{G}}} - V_{\mathrm{G}} - \frac{C_{\mathrm{B}}}{C_{\mathrm{G}}} \left(V_{\mathrm{S}} + V_{\mathrm{D}}\right) \\
&= V_N - V_{\mathrm{G}} - \frac{C_{\mathrm{B}}}{C_{\mathrm{G}}} V_{\mathrm{S}} - \frac{C_{\mathrm{B}}}{C_{\mathrm{G}}} V_{\mathrm{D}},
\end{aligned} \tag{E.3}
$$

where V_N is the gate voltage at which the Nth resonance occurs when $V_{\mathrm{S}} = V_{\mathrm{D}} = 0$.

V_N is defined as,

$$V_N = \left(N - N_0 - \frac{1}{2} \right) \frac{e}{C_\mathrm{G}} + \frac{C_\Sigma E_N}{e C_\mathrm{G}}.$$

The hot and cold electrochemical potentials, $\mu_{\mathrm{H,C}}$, in Eq. (E.1) are related to the source and drain voltages by, $\mu_\mathrm{H} = -eV_\mathrm{S}$ and $\mu_\mathrm{C} = -eV_\mathrm{D}$. Inserting these terms into Eq. (E.1) and insert that result into Eq. (E.3) gives,

$$-\frac{C_\Sigma}{C_\mathrm{G}} \frac{T_\mathrm{H}}{\Delta T} V_\mathrm{D} + \frac{C_\Sigma}{C_\mathrm{G}} \frac{T_\mathrm{C}}{\Delta T} V_\mathrm{S} = V_N - V_\mathrm{G} - \frac{C_\mathrm{B}}{C_\mathrm{G}} V_\mathrm{S} - \frac{C_\mathrm{B}}{C_\mathrm{G}} V_\mathrm{D}.$$

Collecting the V_S and V_D terms gives,

$$\frac{C_\Sigma}{C_\mathrm{G}} \left(\frac{T_\mathrm{H}}{\Delta T} - \frac{C_\mathrm{B}}{C_\Sigma} \right) V_\mathrm{D} - \frac{C_\Sigma}{C_\mathrm{G}} \left(\frac{T_\mathrm{C}}{\Delta T} + \frac{C_\mathrm{B}}{C_\Sigma} \right) V_\mathrm{S} = V_\mathrm{G} - V_N. \tag{E.4}$$

Now the slope dS/dE can be calculated by,

$$\frac{dS}{dE} = \frac{dV_\mathrm{G}}{dE} \frac{dS}{dV_\mathrm{G}} = \frac{dV_\mathrm{G}}{dE} \frac{dS}{dV_\mathrm{th}} \frac{dV_\mathrm{th}}{dV_\mathrm{G}} = \frac{1}{e} \frac{C_\Sigma}{C_\mathrm{G}} \frac{1}{\Delta T} \frac{dV_\mathrm{th}}{dV_\mathrm{G}}.$$

where the last equality follows from Eq. (II.3) and the definition of the Seebeck coefficient. The thermovoltage is the difference $V_\mathrm{th} = V_\mathrm{D} - V_\mathrm{S}$. Therefore, expanding

out $dV_{\text{th}}/dV_{\text{G}}$ gives,

$$
\begin{aligned}
\frac{dS}{dE} &= \frac{1}{e}\frac{C_{\Sigma}}{C_{\text{G}}}\frac{1}{\Delta T}\frac{d\left(V_{\text{D}}-V_{\text{S}}\right)}{dV_{\text{G}}}\\
&= \frac{1}{e}\frac{C_{\Sigma}}{C_{\text{G}}}\frac{1}{\Delta T}\left[\frac{dV_{\text{D}}}{dV_{\text{G}}}-\frac{dV_{\text{S}}}{dV_{\text{G}}}\right]\\
&= \frac{1}{e}\frac{C_{\Sigma}}{C_{\text{G}}}\frac{1}{\Delta T}\left[\left(\frac{dV_{\text{G}}}{dV_{\text{D}}}\right)^{-1}-\left(\frac{dV_{\text{G}}}{dV_{\text{S}}}\right)^{-1}\right].
\end{aligned}
\tag{E.5}
$$

From Eq. (E.4),

$$
\begin{aligned}
\left(\frac{dV_{\text{G}}}{dV_{\text{D}}}\right)^{-1} &= \frac{C_{\text{G}}}{C_{\Sigma}}\left(\frac{T_{\text{H}}}{\Delta T}-\frac{C_{\text{B}}}{C_{\Sigma}}\right)^{-1}\\
&= \frac{C_{\text{G}}}{C_{\Sigma}}\left(\frac{T_{\text{H}}C_{\Sigma}-\Delta TC_{\text{B}}}{\Delta TC_{\Sigma}}\right)^{-1}\\
&= \frac{C_{\text{G}}\Delta T}{T_{\text{H}}C_{\Sigma}-\Delta TC_{\text{B}}}.
\end{aligned}
$$

And

$$
\begin{aligned}
\left(\frac{dV_{\text{G}}}{dV_{\text{S}}}\right)^{-1} &= -\frac{C_{\text{G}}}{C_{\Sigma}}\left(\frac{T_{\text{C}}}{\Delta T}+\frac{C_{\text{B}}}{C_{\Sigma}}\right)^{-1}\\
&= \frac{C_{\text{G}}}{C_{\Sigma}}\left(\frac{T_{\text{H}}C_{\Sigma}-\Delta TC_{\text{B}}}{\Delta TC_{\Sigma}}\right)^{-1}\\
&= -\frac{C_{\text{G}}\Delta T}{T_{\text{C}}C_{\Sigma}+\Delta TC_{\text{B}}}.
\end{aligned}
$$

Putting these into Eq. (E.5) gives,

$$\frac{dS}{dE} = \frac{1}{e}\frac{C_\Sigma}{C_G}\frac{1}{\Delta T}\left(\frac{C_G\Delta T}{T_H C_\Sigma - \Delta T C_B} + \frac{C_G\Delta T}{T_C C_\Sigma + \Delta T C_B}\right)$$

$$= \frac{C_\Sigma}{e}\left(\frac{1}{T_H C_\Sigma - \Delta T C_B} + \frac{1}{T_C C_\Sigma + \Delta T C_B}\right).$$

By expanding out the total capacitance, $C_\Sigma = C_S + C_D + C_G = 2C_B + C_G$, the above can be written

$$\frac{dS}{dE} = \frac{C_\Sigma}{e}\left(\frac{1}{T_H\left(2C_B + C_G\right) - \left(T_H - T_C\right)C_B} + \frac{1}{T_C\left(2C_B + C_G\right) + \left(T_H - T_C\right)C_B}\right)$$

$$= \frac{C_\Sigma}{e}\left(\frac{1}{T_H\left(C_B + C_G\right) + T_C C_B} + \frac{1}{T_C\left(C_B + C_G\right) + T_H C_B}\right)$$

$$= \frac{C_\Sigma}{e}\left(\frac{1}{\left(T_H + T_C\right)C_B + T_H C_G} + \frac{1}{\left(T_H + T_C\right)C_B + T_C C_G}\right)$$

$$= \frac{C_\Sigma}{e}\left(\frac{1}{2\bar{T}C_B + \left(2\bar{T} - T_C\right)C_G} + \frac{1}{2\bar{T}C_B + \left(2\bar{T} - T_H\right)C_G}\right),$$

where the average temperature, $\bar{T}$, is defined by

$$\bar{T} \equiv \frac{T_H + T_C}{2} = T_C + \frac{\Delta T}{2} = T_H - \frac{\Delta T}{2}. \tag{E.6}$$

Continuing the algebra,

$$
\begin{aligned}
\frac{dS}{dE} &= \frac{C_\Sigma}{e} \left(\frac{1}{\bar{T}\left(2C_\mathrm{B}+C_\mathrm{G}\right)+\left(\bar{T}-T_\mathrm{C}\right)C_\mathrm{G}} + \frac{1}{\bar{T}\left(2C_\mathrm{B}+C_\mathrm{G}\right)+\left(\bar{T}-T_\mathrm{H}\right)C_\mathrm{G}} \right) \\
&= \frac{C_\Sigma}{e} \left(\frac{1}{\bar{T}C_\Sigma + C_\mathrm{G}\Delta T/2} + \frac{1}{\bar{T}C_\Sigma - C_\mathrm{G}\Delta T/2} \right) \\
&= \frac{C_\Sigma}{e} \frac{\bar{T}C_\Sigma - C_\mathrm{G}\Delta T/2 + \bar{T}C_\Sigma + C_\mathrm{G}\Delta T/2}{\left(\bar{T}C_\Sigma + C_\mathrm{G}\Delta T/2\right)\left(\bar{T}C_\Sigma - C_\mathrm{G}\Delta T/2\right)} \\
&= \frac{C_\Sigma}{e} \frac{2\bar{T}C_\Sigma}{\bar{T}^2 C_\Sigma^2 - \left(C_\mathrm{G}\Delta T/2\right)^2} \\
&= \frac{2}{e\bar{T}} \left(1 - \left(\frac{C_\mathrm{G}}{C_\Sigma} \frac{\Delta T}{2\bar{T}} \right)^2 \right)^{-1} \\
&= \frac{2}{e\bar{T}} \left(1 - \left(\alpha \frac{\Delta T}{2\bar{T}} \right)^2 \right)^{-1}.
\end{aligned}
$$

Note that $\alpha \ll 1$ and $\Delta T / 2\bar{T} < 1$ so that $\alpha \Delta T / 2\bar{T} \ll 1$. Therefore, a first-order

Taylor expansion of the above provides

$$
\frac{dS}{dE} \approx \frac{2}{e\bar{T}} = \frac{1}{e}\frac{4}{T_\mathrm{H}+T_\mathrm{C}}. \tag{E.7}
$$

This is the final result of this section.

Seebeck Slope from the Mott Formula

The Mott formula (see Eq. (I.25)) relates thermopower as a function of energy,

$S(E)$, to conductance, $G(E)$, by

$$S(E) = -eL_0 \bar{T} \frac{1}{G(E)} \frac{dG(E)}{dE},$$

where the (conventional) Lorenz number is $L_0 = \pi^2/3\,(k/e)^2$ and the average temperature $\bar{T} = (T_\mathrm{H} + T_\mathrm{C})/2$ has been used to be consistent with the previous section. In the narrow resonance limit ($\Gamma \ll kT$), Eq. (A.7) provides an analytical approximation for $G(E)$,

$$G(E)|_{V=0} = A\,\mathrm{sech}^2\left(\frac{E - E_N}{2k\bar{T}}\right),$$

where E_N is the energy at which the Nth differential conductance peak is centered. Inserting this expression for $G(E)$ into the Mott formula and performing the derivative provides,

$$
\begin{aligned}
S(E) &= -eL_0\bar{T}\cosh^2\left(\frac{E - E_N}{2k\bar{T}}\right) \frac{d}{dE}\mathrm{sech}^2\left(\frac{E - E_N}{2k\bar{T}}\right) \\
&= \frac{eL_0}{k}\cosh^2\left(\frac{E - E_N}{2k\bar{T}}\right) \mathrm{sech}^2\left(\frac{E - E_N}{2k\bar{T}}\right) \tanh\left(\frac{E - E_N}{2k\bar{T}}\right) \\
&= \frac{eL_0}{k}\tanh\left(\frac{E - E_N}{2k\bar{T}}\right).
\end{aligned}
$$

The slope dS/dE on resonance can be calculated as,

$$
\begin{aligned}
\left.\frac{dS}{dE}\right|_{E=E_N} &= \frac{e}{k} L_0 \left.\frac{d}{dE} \tanh\left(\frac{E-E_N}{2k\bar{T}}\right)\right|_{E=E_N} \\
&= \frac{e^2}{2k^2} L_0 \frac{1}{e\bar{T}} \left.\mathrm{sech}^2\left(\frac{E-E_N}{2k\bar{T}}\right)\right|_{E=E_N} \\
&= \frac{1}{2}\left(\frac{e}{k}\right)^2 L_0 \frac{1}{e\bar{T}}.
\end{aligned}
\tag{E.8}
$$

This is the final result of the section.

BIBLIOGRAPHY

[1] G. W. Burns and M. G. Scroger, NIST Spec. Publ. **250-35** (1989).

[2] L. Onsager, Phys. Rev. **37**, 405 (1931), and **38**, 2265 (1931).

[3] H. B. Callen, Phys. Rev. **73**, 1349 (1948).

[4] S. R. de Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (Dover, New York, 1984).

[5] C. van den Broeck, Adv. Chem. Phys. **135**, 189 (2007).

[6] C. B. Vining, Nature Materials **8**, 83 (2009).

[7] S. Carnot, *Réflexions sur la Puissance Motrice du Feu et sur les Machines propres à développer cette Puissance* (Bachelier, Paris, 1824).

[8] N. W. Ashcroft and N. D. Merman, *Solid State Physics* (Saunders College Publishing, New York, 1976).

[9] R. Kim, S. Datta, and M. S. Lundstrom, J. Appl. Phys. **105**, 034506 (2009).

[10] T. E. Humphrey, R. Newbury, R. P. Taylor, and H. Linke, Phys. Rev. Lett. **89**, 116801 (2002).

[11] L. D. Hicks and M. S. Dresselhaus, Phys. Rev. B **47**, 12727 (1993).

[12] L. D. Hicks and M. S. Dresselhaus, Phys. Rev. B **47**, 16631 (1993).

[13] G. D. Mahan and J. O. Sofo, Proc. Natl. Acad. Sci. USA **93**, 7436 (1996).

[14] T. E. Humphrey and H. Linke, Phys. Rev. Lett. **94**, 096601 (2005).

[15] M. Zebarjadi, A. Shakouri, and K. Esfarjani, Phys. Rev. B **74** (2006).

[16] J. P. Heremans, V. Jovovic, E. S. Toberer, A. Saramat, K. Kurosaki, A. Charoenphakdee, S. Yamanaka, and G. J. Snyder, Science **321**, 554 (2008).

[17] A. Majumdar, Science **303**, 777 (2004).

[18] J. Baxter, Z. X. Bian, G. Chen, D. Danielson, M. S. Dresselhaus, A. G. Fedorov, T. S. Fisher, C. W. Jones, E. Maginn, U. Kortshagen, et al., Energy Environ. Sci. **2**, 559 (2009).

[19] C. Chiritescu, D. G. Cahill, N. Nguyen, D. Johnson, A. Bodapati, P. Keblinski, and P. Zschack, Science **315**, 351 (2007).

[20] L. Shi, D. Y. Li, C. H. Yu, W. Y. Jang, D. Kim, Z. Yao, P. Kim, and A. Majumdar, J. Heat Trans. **125**, 881 (2003).

[21] J. H. Seol, A. L. Moore, S. K. Saha, F. Zhou, L. Shi, Q. L. Ye, R. Scheffler, N. Mingo, and T. Yamada, J. Appl. Phys. **101**, 023706 (2007).

[22] A. I. Hochbaum, R. K. Chen, R. D. Delgado, W. J. Liang, E. C. Garnett, M. Najarian, A. Majumdar, and P. D. Yang, Nature **451**, 163 (2008).

[23] A. I. Boukai, Y. Bunimovich, J. Tahir-Kheli, J. K. Yu, W. A. Goddard, and J. R. Heath, Nature **451**, 168 (2008).

[24] A. Shakouri and J. E. Bowers, Appl. Phys. Lett. **71**, 1234 (1997).

[25] G. J. Snyder and E. S. Toberer, Nature Materials **7**, 105 (2008).

[26] R. Landauer, IBM J. Res. Dev **1**, 223 (1957), (see also J. Math. Phys. **37**, 10 (1996)).

[27] R. Landauer, Phil. Mag. **21**, 863 (1970).

[28] M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986).

[29] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1995).

[30] D. K. Ferry and S. K. Goodnick, *Transport in Nanostructures* (Cambridge University Press, Cambridge, England, 1997).

[31] U. Sivan and Y. Imry, Phys. Rev. B **33**, 551 (1986).

[32] P. N. Butcher, J. Phys. Condens. Matt. **2**, 4869 (1990).

[33] H. van Houten, L. W. Molenkamp, C. W. J. Beenakker, and C. T. Foxon, Semicond. Sci. Technol. **7**, B215 (1992), (see also e-print arXiv:cond-mat/0512612).

[34] M. Cutler and N. F. Mott, Phys. Rev. **181**, 1336 (1969).

[35] R. Franz and G. Wiedemann, Annal. der Phys. und Chem. **165**, 497 (1853).

[36] A. Greiner, L. Reggiani, T. Kuhn, and L. Varani, Phys. Rev. Lett. **78**, 1114 (1997).

[37] F. Volklein, H. Reith, T. W. Cornelius, M. Rauber, and R. Neumann, Nanotech. **20**, 8 (2009).

[38] B. Kubala, J. Konig, and J. Pekola, Phys. Rev. Lett. **100**, 066801 (2008).

[39] A. F. Ioffe, *Physics of semiconductors* (Infosearch, London, 1960).

[40] A. V. Golubkov, L. S. Parfen'eva, I. A. Smirnov, H. Misiorek, J. Mucha, and A. Jezowski, Phys. Solid State **42**, 1394 (2000).

[41] R. Venkatasubramanian, E. Siivola, T. Colpitts, and B. O'Quinn, Nature **413**, 597 (2001).

[42] B. Wölfing, C. Kloc, J. Teubner, and E. Bucher, Phys. Rev. Lett. **86**, 4350 (2001).

[43] T. C. Harman, P. J. Taylor, M. P. Walsh, and B. E. LaForge, Science **297**, 2229 (2002).

[44] K. F. Hsu, S. Loo, F. Guo, W. Chen, J. S. Dyck, C. Uher, T. Hogan, E. K. Polychroniadis, and M. G. Kanatzidis, Science **303**, 818 (2004).

[45] B. Poudel, Q. Hao, Y. Ma, Y. C. Lan, A. Minnich, B. Yu, X. Yan, D. Z. Wang, A. Muto, D. Vashaee, et al., Science **320**, 634 (2008).

[46] J. S. Rhyee, K. H. Lee, S. M. Lee, E. Cho, S. I. Kim, E. Lee, Y. S. Kwon, J. H. Shim, and G. Kotliar, Nature **459**, 965 (2009).

[47] W. T. Tsang, Appl. Phys. Lett. **45**, 1234 (1984).

[48] A. Y. Cho, J. Vac. Sci. Technol. **8**, S31 (1971).

[49] R. S. Wagner and W. C. Ellis, Appl. Phys. Lett. **4**, 89 (1964).

[50] A. I. Persson, M. W. Larsson, S. Strenström, B. J. Ohlsson, L. Samuelson, and L. R. Wallenberg, Nature Mat. **3**, 677 (2004).

[51] K. A. Dick, K. Deppert, T. Martensson, B. Mandl, L. Samuelson, and W. Seifert, Nano Lett. **5**, 761 (2005).

[52] A. I. Persson, Ph.d. thesis, Lund University (2005).

[53] L. Fröberg, Ph.d. thesis, Lund University (2008).

[54] C. Thelander, M. T. Björk, M. W. Larsson, A. E. Hansen, L. R. Wallenberg, and L. Samuelson, Solid State Communications **131**, 573 (2004).

[55] M. T. Björk, Ph.d. thesis, Lund University (2004).

[56] H. van Houten, C. W. J. Beenakker, and A. A. M. Staring, in *Single Charge Tunneling*, edited by H. Grabert and M. Devoret (Plenum, New York, 1992), vol. 294 of *NATO Advanced Studies Institutes*, (see also e-print arXiv:cond-mat/0508454).

[57] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, and L. M. K. Vandersypen, Rev. Mod. Phys. **79**, 1217 (2007).

[58] M. T. Björk, B. J. Ohlsson, C. Thelander, A. I. Persson, K. Deppert, L. R. Wallenberg, and L. Samuelson, Appl. Phys. Lett. **81**, 4458 (2002).

[59] C. Thelander, T. Martensson, M. T. Björk, B. J. Ohlsson, M. W. Larsson, L. R. Wallenberg, and L. Samuelson, Appl. Phys. Lett. **83**, 2052 (2003).

[60] M. F. O'Dwyer, H. Linke, E. A. Hoffmann, T. E. Humphrey, R. A. Lewis, and C. Zhang, in *The 25th International Conference on Thermoelectrics* (IEEE, Vienna, Austria, 2006), p. 228.

[61] E. B. Foxman, P. L. McEuen, U. Meirav, N. S. Wingreen, Y. Meir, P. A. Belk, N. R. Belk, M. A. Kastner, and S. J. Wind, Phys. Rev. B **47**, 10020 (1993).

[62] M. T. Björk, C. Thelander, A. E. Hansen, L. E. Jensen, M. W. Larsson, L. R. Wallenberg, and L. Samuelson, Nano Lett. **4**, 1621 (2004).

[63] J. P. Small, K. M. Perez, and P. Kim, Phys. Rev. Lett. **91**, 256801 (2003).

[64] M. C. Llaguno, J. E. Fischer, A. T. Johnson, and J. Hone, Nano Lett. **4**, 45 (2004).

[65] W. E. Chickering, J. P. Eisenstein, and J. L. Reno, Phys. Rev. Lett. **103**, 046807 (2009).

[66] L. W. Molenkamp, H. Vanhouten, C. W. J. Beenakker, R. Eppenga, and C. T. Foxon, Phys. Rev. Lett. **65**, 1052 (1990).

[67] A. S. Dzurak, C. G. Smith, C. H. W. Barnes, M. Pepper, L. MartinMoreno, C. T. Liang, D. A. Ritchie, and G. A. C. Jones, Phys. Rev. B **55**, 10197 (1997).

[68] R. Scheibner, M. König, D. Reuter, A. D. Wieck, C. Gould, H. Buhmann, and L. W. Molenkamp, New Journal of Physics **10**, 8 (2008).

[69] J. H. Scofield, Amer. J. Phys. **62**, 129 (1994).

[70] L. W. Molenkamp, T. Gravier, H. van Houten, O. J. A. Buijk, M. A. A. Mabesoone, and C. T. Foxon, Phys. Rev. Lett. **68**, 3765 (1992).

[71] A. A. M. Staring, L. W. Molenkamp, B. W. Alphenaar, H. Vanhouten, O. J. A. Buyk, M. A. A. Mabesoone, C. W. J. Beenakker, and C. T. Foxon, Europhysics Letters **22**, 57 (1993).

[72] L. W. Molenkamp, A. A. M. Staring, B. W. Alphenaar, H. van Houten, and C. W. J. Beenakker, Semicond. Sci. Technol. **9**, 903 (1994).

[73] R. Scheibner, E. G. Novik, T. Borzenko, M. König, D. Reuter, A. D. Wieck, H. Buhmann, and L. W. Molenkamp, Phys. Rev. B **75**, 041301 (2007).

[74] J. T. Nicholls and O. Chiatti, Journal of Physics: Condensed Matter **20**, 164210 (2008).

[75] J. Hone, I. Ellwood, M. Muno, A. Mizel, M. L. Cohen, A. Zettl, A. G. Rinzler, and R. E. Smalley, Phys. Rev. Lett. **80**, 1042 (1998).

[76] E. A. Hoffmann, N. Nakpathomkun, A. I. Persson, H. Linke, H. A. Nilsson, and L. Samuelson, Appl. Phys. Lett. **91**, 252114 (2007).

[77] E. A. Hoffmann, N. Nakpathomkun, A. I. Persson, H. A. Nilsson, L. Samuelson, and H. Linke, Physica E: Low-dimensional Systems and Nanostructures **40**, 1605 (2007).

[78] E. A. Hoffmann and H. Linke, Journal of Low Temperature Physics **154**, 161 (2009).

[79] E. A. Hoffmann, H. A. Nilsson, J. E. Matthews, N. Nakpathomkun, A. I. Persson, L. Samuelson, and H. Linke, Nano Lett. **9**, 779 (2009).

[80] F. Giazotto, T. T. Heikkilä, A. Luukanen, A. M. Savin, and J. P. Pekola, Rev. Mod. Phys. **78**, 217 (2006).

[81] F. Catara and M. Sambataro, Journal of Physics: Condensed Matter **13**, L705 (2001).

[82] A. S. Dzurak, C. G. Smith, M. Pepper, D. A. Ritchie, J. E. F. Frost, G. A. C. Jones, and D. G. Hasko, Solid State Communications **87**, 1145 (1993).

[83] S. Möller, H. Buhmann, S. F. Godijn, and L. W. Molenkamp, Phys. Rev. Lett. **81**, 5197 (1998).

[84] R. Scheibner, H. Buhmann, D. Reuter, M. N. Kiselev, and L. W. Molenkamp, Phys. Rev. Lett. **95**, 176602 (2005).

[85] C. W. J. Beenakker, Phys. Rev. B **44**, 1646 (1991).

[86] M. Turek and K. A. Matveev, Phys. Rev. B **65**, 9 (2002).

[87] K. A. Matveev and A. V. Andreev, Phys. Rev. B **66**, 15 (2002).

[88] X. Zianni, Phys. Rev. B **78**, 9 (2008).

[89] P. Kim, L. Shi, A. Majumdar, and P. L. McEuen, Phys. Rev. Lett. **87**, 215502 (2001).

[90] X. Zianni, Phys. Rev. B **75**, 12 (2007).

[91] X. Zianni, Phys. Rev. B **78**, 9 (2008), and private communications.

[92] D. V. Averin and Y. V. Nazarov, Phys. Rev. Lett. **65**, 2446 (1990).

[93] S. D. Franceschi, S. Sasaki, J. M. Elzerman, W. G. van der Wiel, S. Tarucha, and L. P. Kouwenhoven, Phys. Rev. Lett. **86**, 878 (2001).

[94] D. Becker and D. Pfannkuche, Phys. Rev. B **77**, 205307 (2008).

[95] T. Nakanishi and T. Kato, Journal of the Physical Society of Japan **76** (2007).

[96] F. L. Curzon and B. Ahlborn, Amer. J. Phys. **43**, 22 (1975).

[97] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics I* (Addison-Wesley, Reading, MA, 1963).

[98] Z. C. Tu, J. Phys. A-Math. Theor. **41**, 6 (2008).

[99] M. Esposito, K. Lindenberg, and C. V. den Broeck, Epl **85**, 5 (2009).

[100] M. Esposito, K. Lindenberg, and C. V. den Broeck, Phys. Rev. Lett. **102**, 4 (2009).

[101] M. A. Cipiloglu, S. Turgut, and M. Tomak, Physica Status Solidi B-Basic Research **241**, 2575 (2004).

[102] A. S. Dzurak, C. G. Smith, L. Martinmoreno, M. Pepper, D. A. Ritchie, G. A. C. Jones, and D. G. Hasko, Journal of Physics-Condensed Matter **5**, 8055 (1993).

[103] J. M. Wang, L. H. Wan, Y. D. Wei, Y. X. Xing, and J. Wang, Mod. Phys. Lett. B **20**, 215 (2006).

[104] F. Zhou, Ph.d. thesis, University of Texas at Austin (2009).

[105] T. Bryllert, L.-E. Wernersson, T. Löwgren, and L. Samuelson, Nanotech. **17**, S227 (2006).

[106] C. Thelander, L. E. Fröberg, C. Rehnstedt, L. Samuelson, and L.-E. Wemersson, IEEE Electron Device Letters **29**, 206 (2008).

[107] C. Thelander, C. Rehnstedt, L. E. Fröberg, E. Lind, T. Martensson, P. Caroff, T. Lowgren, B. J. Ohlsson, L. Samuelson, and L.-E. Wernersson, IEEE Transactions on Electron Devices **55**, 3030 (2008).

[108] P. Reddy, S.-Y. Jang, R. A. Segalman, and A. Majumdar, Science **315**, 1568 (2007).

[109] H. M. Antia, *Numerical Methods for Scientists and Engineers* (Tata McGraw–Hill, New Delhi, 1991), p. 160.