Temperature Measurement Using a Transmon Device

Henrik Lievonen

School of Science

Bachelor's thesis
Espoo 29.9.2020

Supervisor

Sorin Paraoanu

Advisor

Andrey Lebedev

Aalto University
School of Science

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the bachelor's thesis

Author Henrik Lievonen

Title Temperature Measurement Using a Transmon Device

Degree programme Teknillistieteellinen kandidaattiohjelma

Major Teknillinen fysiikka **Code of major** SCI3028

Teacher in charge Sorin Paraoanu

Advisor Andrey Lebedev

Date 29.9.2020 **Number of pages** 26+8 **Language** English

Abstract

I have developed a method for measuring temperature in range of tens to hundreds of millikelvins. The method uses a quantum mechanical tarnsmon device as the thermometer. The method is based on the fact that at the thermodynamical equilibrium the state of a quantum mechanical system is distributed according to the Boltzmann distribution. By fitting the Boltzmann distribution to the measured state distribution, one can determine the temperature of the system.

The state distribution of the system is measured by applying a measurement pulse and measuring the reflected pulse. The measurement pulse collapses the quantum mechanical state to one of the eigenstates of the Hamiltonian of the system. The energy level of the eigenstate determines the properties of the reflected measurement pulse.

In practice the measured pulse is very weak and noisy. To improve the signal-to-noise ratio, the measurement is repeated multiple times and the response signals are averaged. Each state of the system produces a different reflected pulse, and averaging multiple measurements loses the information of the distribution of those pulses and, by proxy, the states. To recover the information of the state distribution, the system is prepared in different states before the measurement. Because we know how the preparation affects the state, we can solve the state distribution from this set of measurements. We can then use the Boltzmann distribution to solve for the temperature of the sample.

The method has been developed using a transmon qutrit, but it works also for other kinds of quantum mechanical systems. The method produces plausible results for the temperatures, but not all sources of errors could be determined. More measurements are needed to unravel these error sources.

Keywords Quantum mechanics, Transmon, Metrology, Temperature, Thermometer, Qutrit

Aalto University
School of Science

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi
Tekniikan kandidaatintyön tiivistelmä

Tekijä Henrik Lievonen

Työn nimi Temperature Measurement Using a Transmon Device

Koulutusohjelma Teknillistieteellinen kandidaattiohjelma

Pääaine Teknillinen fysiikka **Pääaineen koodi** SCI3028

Vastuuopettaja Sorin Paraoanu

Työn ohjaaja Andrey Lebedev

Päivämäärä 29.9.2020 **Sivumäärä** 26+8 **Kieli** Englanti

Tiivistelmä

Olen kehittänyt kvanttimekaniikkaan perustuvan menetelmän lämpötilan mittaamiseen. Menetelmän avulla voidaan mitata lämpötiloja kymmenien ja satojen kelvinien kokoluokassa. Menetelmä perustuu siihen, että kvanttimekaanisen järjestelmän miehitystodennäköisyydet ovat tasapainotilassa Boltzmann-jakautuneet. Mitattuihin miehitystodennäköisyyksiin voidaan sovittaa Boltzmannin jakauma, mistä saadaan määritettyä järjestelmän lämpötila.

Järjestelmän miehitystila mitataan kohdistamalla järjestelmään mittauspulssi ja mittaamalla mittauspulssin heijastus järjestelmästä. Mittauspulssi romauttaa järjestelmän kvanttimekaanisen tilan yhteen sen Hamiltonin operaattorin ominaistiloista. Ominaistilan energia määrää millaisen heijastuksen järjestelmä tuottaa mittauspulssille.

Käytännössä mitattava heijastunut pulssi on hyvin heikkotehoinen ja kohinainen, joten signaali-kohinasuhteen parantamiseksi mittaus toistetaan useasti ja signaalit keskiarvoistetaan. Keskiarvoistamalla menetetään tieto järjestelmän miehitystodennäköisyyksien jakaumasta. Miehitysjakauma voidaan kuitenkin selvittää valmistelemalla järjestelmä erilaisiin tiloihin ennen mittausta, jolloin jokaisesta eri alkutilasta suoritetusta mittauksesta syntyy erilainen keskiarvoistettu heijastussignaali. Tästä syntyvän yhtälöryhmän ratkaisuna saadaan järjestelmän miehitystodennäköisyyksien suhde, josta saadaan edelleen ratkaistua lämpötila Boltzmannin jakauman avulla.

Mittausmenetelmä on kehitetty käyttämällä transmoni-kutrittia, mutta se yleistyy myös muille kvanttimekaanisille laitteille. Menetelmä tuottaa uskottavia tuloksia lämpötilalle, mutta kaikkia häiriölähteitä ei pystytty selvittämään tutkimuksessa. Näiden häiriölähteiden selvittämiseksi tarvitaan lisää mittauksia.

Avainsanat Kvanttimekaniikka, Transmoni, Metrologia, Lämpötila, Lämpömittari, Kutritti

Contents

Symbols and abbreviations

Symbols

k	Boltzmann constant
ρ	Charge density
q	Charge of a Cooper pair
E_C	Charging energy of transmon
$\langle\phi\|$	Conjugate wave function
I	Electric current
E_J	Energy of Josephson junction
$\|\phi_i\rangle$	Hamiltonian basis wave function
Φ	Magnetic flux
$\mathbf{A}$	Magnetic vector potential
m	Mass
x_0, x_1, x_2, y_0, y_1, y_2	Measurement signal
$\Delta\varphi_a$	Phase difference along path a
$\mathbf{J}$	Probability current
P	Probability density
$\bar{h}$	Reduced Planck's constant
φ_i	Response signal corresponding to $\|\phi_i\rangle$
θ, ν	Phase
T	Temperature
ω_{01}	Transition frequency between ground state and first exited state
$\|\phi\rangle$	Wave function

Operators

$\nabla\cdot$	Divergence
$\dot{f}$	First time derivative of f
∇	Gradient
$\hat{H}$	Hamiltonian operator
$\int_a$	Integral along path a
$\hat{n}$	Number of Cooper pairs transferred over Josephson junction
$\frac{\partial f}{\partial t}$	Partial differential of f w.r.t. t
$\hat{\varphi}$	Phase difference between superconductors
Var	Variance

Abbreviations

CPB	Cooper pair box
JJ	Josephson junction
SQUID	Superconducting quantum interference device
Transmon	Transmission-line shunted plasma oscillation qubit

1 Introduction

Quantum mechanics have gained a lot of attention in the media in the recent years, mainly due to the ground-breaking results in the quantum computing like achieving quantum supremacy [1]. Many of these results are direct consequences of other advancements with the quantum mechanics, like qubit designs with longer decoherence times [2]. These advancements are applicable to a wide range of other applications too. One such application is introduced in this thesis: a thermometer based on a transmon qutrit.

The thermometer developed in this thesis can be used to measure temperatures in the range of dozens to hundreds of millikelvins. Most quantum experiments and devices, like quantum computers, operate at these temperatures as higher temperatures would cause the quantum state to decay too quickly. The method in this thesis has been developed and tested using a transmon qutrit, but it should be applicable to be used with other quantum devices. This makes it possible to measure the temperature of many quantum systems with little to no additional hardware, allowing one to estimate the decoherence of the system.

The method presented in this thesis works by measuring the state distribution of a fully decohered, three-state transmon system. The state distribution follows the Boltzmann distribution whose temperature dependency is used to infer the temperature of the system. Only the energies of the states need to be known for the method to work.

The method has been implemented using Matlab, and the implementation has been used to evaluate the results by comparing the determined temperature at different driving frequencies of the transmon device. The measurements in this study were done at Aalto University by Andrey Lebedev using a transmon device described in [3].

Section 2 explains the theoretical background for the study and gives a short description on how a transmon device works. Section 3 introduces the measurement setup and the data analysis methods. Section 4 presents the results, and Section 5 draws the conclusions of the method. Appendix B presents a derivation for Deming regression without an intercept parameter.

2 Background

The quantum mechanical device used in this study is a *transmon* qutrit [2]. The transmon device is based on the Cooper pair box qubit, but it has been engineered to have a longer dephasing time than a traditional charge qubit. At the heart of a transmon device is a *Josephson junction* (JJ) [4] consisting of two superconducting wires separated by a thing layer of non-superconducting material like a metal or an insulator. When two Josephson junctions are connected in parallel, they show interference behavior and form a *superconducting quantum interference device* (SQUID) [5]. In the rest of this section, I introduce the theory behind these concepts to the level needed to understand the measurement setup and method developed in this study.

2.1 Superconductivity

To fully understand how a Josephson junction works, one needs to have at least a basic understanding of how materials transfer from having a resistance to become superconductive. Electrical current in a material is exerted by a potential difference inside the material; negative charges move towards the higher potential and positive charges towards the lower potential. In solids the atomic nuclei form a rigid lattice, leaving only the freely moving valence electrons to conduct current. During their transport in the solid, electrons scatter from the stationary nuclei by exchanging their energy with the vibrations of the lattice. These scattering events are the microscopic reason for the macroscopic effect we call *electrical resistance*.

An electron in a solid attracts the positively charged nuclei of the lattice the tiniest of amount. The attracted positive charges raise the potential around the electron, causing other electrons to be attracted closer to the electron. It happens that the electrons have a lower energy level when they are attracted in this way, compared to the electrons being far apart, and thus the electrons favor to be paired up like this. At sufficiently low temperatures, this energy difference is large enough that the kinetic energy of the electrons can't break the pair apart, and the state is bound. This kind of pair of electrons is called a Cooper pair. [6]

Electrons have spin $\frac{1}{2}$ and are thus fermions. The Cooper pair, on the other hand, is made of two electrons and thus has spin of either 0 or 1, making the Cooper pair a boson. Bosons follow the Bose-Einstein statistics, and especially are not subject to the Pauli exclusion principle, which allows multiple Cooper pairs to exist in the same quantum state. At low temperatures, the bosons actually tend the same lowest energy state, and getting them to any other state requires energy. This means that most of the Cooper pairs move in the same direction once they are started, and continue moving until an external potential is applied. Resistance causing scattering from atomic nuclei is unlikely due to Bose-Einstein statistics. This phenomenon of resistance-free electric conduction is called superconductivity.

Let's consider a simply connected superconductor as a quantum mechanical device with wave function $|\phi(\mathbf{r}, t)\rangle$. The probability of finding the Cooper pair at any given

location is given by

$$P(\mathbf{r}, t) = \langle \phi(\mathbf{r}, t) | \phi(\mathbf{r}, t) \rangle \, .$$

It shows [7] that the change of this probability can actually be written using probability current $\mathbf{J}$:

$$\frac{\partial P}{\partial t} = -\nabla \cdot \mathbf{J} \tag{1}$$

For a large number of number of Cooper pairs, this probability current density actually becomes the electric current density.

Now, because $\mathbf{J}$ is actually the electric current density, we know that P must be proportional to the charge density ρ, and thus $|\phi\rangle$ must be proportional to $\sqrt{\rho}$. We can thus write

$$|\phi\rangle = \sqrt{\rho} e^{i\theta}$$

for some phase θ. Using this notation, the current density $\mathbf{J}$ can now be written as

$$\mathbf{J} = \frac{\hbar}{m} \left(\nabla\theta - \frac{q}{\hbar}\mathbf{A} \right) \rho.$$

2.2 Josephson junction

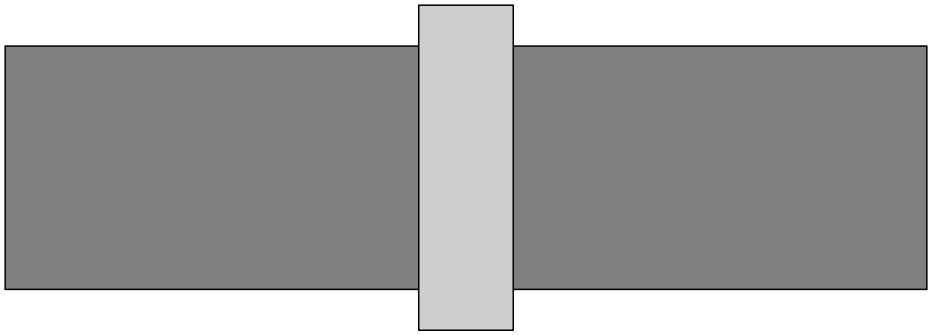


Figure 1: A Josephson junction. The dark gray parts are superconductors which are separated by the light gray insulator.

The basis for many charge qubit based quantum devices is a Josephson junction. Josephson junction is made of two blobs of superconductive material separated by a thin strip of insulator or a non-superconductive material. The basic structure of a Josephson junction is shown in Figure 1. The thin strip of insulator makes it possible for the Cooper pairs to tunnel from one side of the junction to the other. The following derivation of the properties of Josephson junction is based on [7].

The contacts between the superconductive pieces of a Josephson junction are so small that let's approximate them as point-like. Each side of the junction has a wave function, $|\phi_1\rangle$ and $|\phi_2\rangle$, describing its behavior. As they are connected via a

the junction, their time evolution is described by

$$i\hbar\frac{\partial\left|\phi_1\right\rangle}{\partial t} = U_1\left|\phi_1\right\rangle + K\left|\phi_2\right\rangle$$
$$i\hbar\frac{\partial\left|\phi_2\right\rangle}{\partial t} = U_2\left|\phi_2\right\rangle + K\left|\phi_1\right\rangle$$

where U_i is the energy of the lowest state and K_i is a cross term coupling the superconducting regions together via tunneling Cooper pairs. If we apply a voltage V across the junction, we can denote $U_1 - U_2 = qV$ where q is the charge of a Cooper pair. This way we can write

$$\begin{aligned} i\hbar\frac{\partial\left|\phi_1\right\rangle}{\partial t} &= \frac{qV}{2}\left|\phi_1\right\rangle + K\left|\phi_2\right\rangle \\ i\hbar\frac{\partial\left|\phi_2\right\rangle}{\partial t} &= -\frac{qV}{2}\left|\phi_2\right\rangle + K\left|\phi_1\right\rangle . \end{aligned} \tag{2}$$

Because $|\phi_1\rangle$ and $|\phi_2\rangle$ are wave functions of superconductor, we may write them as

$$|\phi_i\rangle = \sqrt{\rho_i}e^{i\theta_i}$$

where ρ_i is the charge density of the superconductor, and θ_i is the phase of $|\phi_i\rangle$. Writing the wave functions in the equation (2) in this form allows us to denote $\theta_2 - \theta_1 = \delta$. By separating real and imaginary parts from each others, we can solve

$$\begin{aligned} \dot{\rho}_1 &= \frac{2K}{\hbar}\sqrt{\rho_2\rho_1}\sin\delta \\ \dot{\rho}_2 &= -\frac{2K}{\hbar}\sqrt{\rho_2\rho_1}\sin\delta \\ \dot{\theta}_1 &= -\frac{K}{\hbar}\sqrt{\frac{\rho_2}{\rho_1}}\cos\delta - \frac{qV}{2\hbar} \\ \dot{\theta}_2 &= -\frac{K}{\hbar}\sqrt{\frac{\rho_1}{\rho_2}}\cos\delta + \frac{qV}{2\hbar}. \end{aligned}$$

These equations describe how the wave functions would start to change if there was no voltage across the junction. Because $I = \frac{\partial\rho}{\partial t} = \dot{\rho}$, we can write

$$I = \frac{2K}{\hbar}\sqrt{\rho_2\rho_1}\sin\delta.$$

By remembering that $\theta_2 - \theta_1 = \delta$, we can write

$$\dot{\delta} = \dot{\theta}_2 - \dot{\theta}_1 = \frac{qV}{\hbar}$$

and thus

$$\delta(t) = \delta_0 + \frac{q}{\hbar}\int_{t_0}^{t} V(t)dt.$$

Let's finally denote $I_0 = \frac{2K}{\hbar}\sqrt{\rho_2\rho_1}$ to gain

$$I = I_0 \sin\delta \tag{3}$$

$$\delta(t) = \delta_0 + \frac{q}{\hbar}\int V(t)d(t) \tag{4}$$

which are the equations describing the physics of a Josephson junction.

The main takeaway from these equations is that when there is a non-zero DC voltage across the junction, the current oscillates with frequency $\frac{q}{h}V$, and when there is no voltage across the junction, there is some current flowing through the junction with an amplitude between $-I_0$ and I_0.

2.3 SQUID

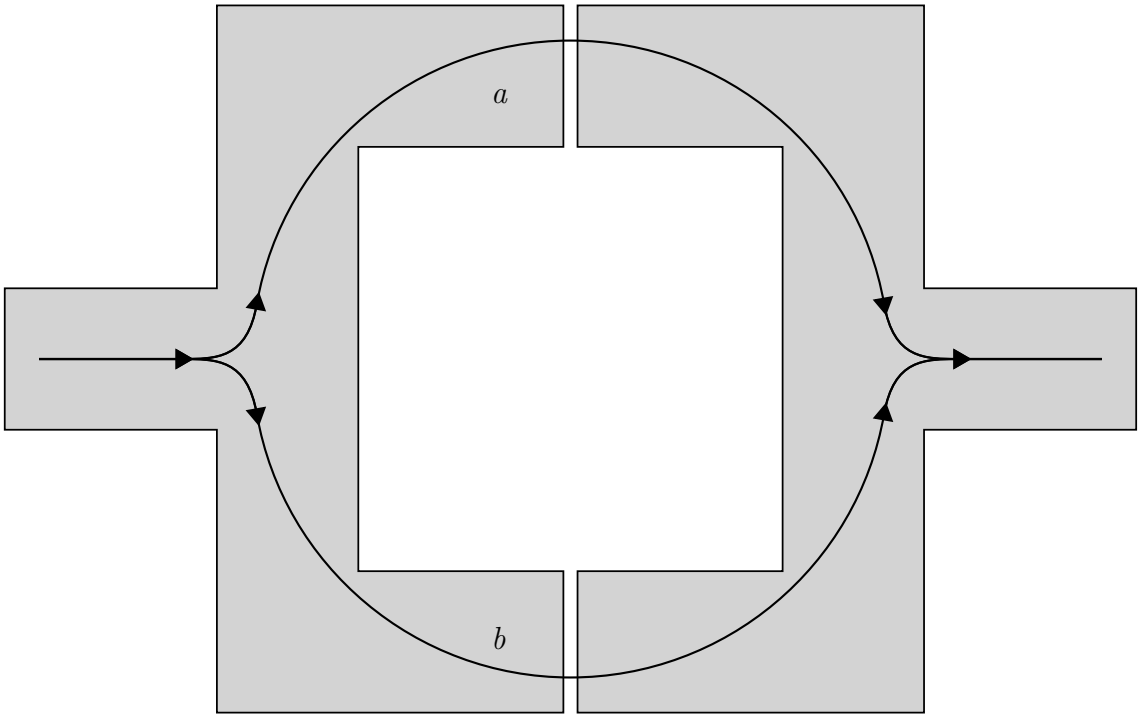


Figure 2: A superconducting quantum interference device, SQUID. The gaps in the ring are the two Josephson junctions. Two integration paths, a and b, are marked in the picture. Each one goes through one of the Josephson junctions.

Two Josephson junctions can be combined in parallel to form a *superconducting quantum interference devices*, abbreviated SQUID [5]. The name comes from the fact that the two Josephson junctions show behavior similar to the Young's double slit interference experiment [8]. Figure 2 shows the structure of a SQUID loop.

Let's consider the current flowing through the SQUID along two paths: path a goes through the upper Josephson junction and path b goes through the lower one. Let's also apply a magnetic potential $\mathbf{A}$ to the device. Now the phase difference

along the path a is

$$\Delta\varphi_a = \delta_a + \frac{q}{\hbar}\int_a \mathbf{A}\cdot d\mathbf{s}$$

and along path b

$$\Delta\varphi_b = \delta_b + \frac{q}{\hbar}\int_b \mathbf{A}\cdot d\mathbf{s}.$$

The δ term comes from the Josephson junction, and the integral term comes from the equation (1) which equates the phase θ and the vector potential $\mathbf{A}$ when the current density around the loop $\mathbf{J} = 0$.

The phase difference $\Delta\varphi$ must be the same along both paths a and b. Thus when these two equations are subtracted from each others, we get

$$\delta_b - \delta_a = \frac{q}{\hbar}\oint \mathbf{A}\cdot d\mathbf{s} = \frac{q}{\hbar}\Phi$$

as the vector potential along a closed loop is equal to the magnetic flux enclosed by the loop. By denoting

$$\delta_a = \delta_0 - \frac{q}{2\hbar}\Phi \qquad\qquad \delta_b = \delta_0 + \frac{q}{2\hbar}\Phi$$

and using equation (3) we get the total current through the SQUID loop

$$\begin{aligned} I &= I_0\left(\sin\left(\delta_0 - \frac{q}{2\hbar}\Phi\right) + \sin\left(\delta_0 + \frac{q}{2\hbar}\Phi\right)\right) \\ &= 2I_0 \sin\delta_0 \cos\frac{q\Phi}{2\hbar}. \end{aligned}$$

The base phase across the junction δ_0 can't be directly controlled, but it is known that $\sin\delta_0$ can reach a maximum value of 1. The equation also shows that the current through the SQUID loop varies as a function of the flux Φ through the loop. This makes SQUID loops useful in high-sensitivity magnetometers, and they can be made to have achieve sensitivity of 0.1 pT [5].

2.4 Transmon

The quantum mechanical device used in this study is a *transmission-line shunted plasma oscillation qubit*, or a *transmon* for short [2]. It is one of the most studied quantum devices due to its long coherence time which is needed for creating efficient quantum computers [9]. The long coherence times make transmon devices useful also for other quantum applications, and one of such applications is measuring the temperature of a given sample.

A transmon device is made of a SQUID loop or a single Josephson junction shunted by a large capacitor. The basic structure is shown in Figure 3. This structure is similar to a Cooper pair box qubit (CPB) design, and their Hamiltonians have actually identical form [2]

$$\hat{H} = 4E_C(\hat{n} - n_g)^2 - E_J\cos\hat{\varphi}. \tag{5}$$

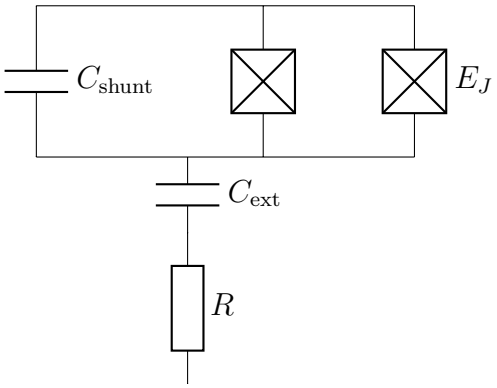


Figure 3: Equivalent circuit of Transmon device. The SQUID loop consists of the two Josephson junctions (the boxes with crosses) on the right. The shunting capacitor C_{shunt} is on the left. The circuit is capacitively connected to the surroundings, shown here as C_{ext}. Based on circuit diagram in [3].

Here $\hat{n}$ denotes the number of Cooper pairs transferred over the Josephson junction, and $\hat{\varphi}$ denotes the phase difference between superconductors. E_C is the charging energy of the transmon, and E_J is the energy of the Jospehson junctions. The charging energy E_C can be lowered by increasing the capacitance of the shunting capacitor C_{shunt}.

A typical CPB qubit is operated at $E_J \approx E_C$, while a transmon is operated in the transmon regime $E_J \gg E_C$. Figure 4 shows the three lowest eigenenergies of (5) for different values of E_J/E_C. The small charging energy caused by the large shunting capacitor smoothens the energy bands of the transmon and makes it less susceptible to fluctuations of charge. This also means that the device can't be driven by variating the charge, but microwaves are needed to excite the system into other states. The transition frequency of the transmon device can be controlled by varying the magnetic flux through the SQUID loop.

The energy levels of transmon are anharmonic, meaning that the transition frequency between adjacent energy levels becomes smaller as the energy increases. The difference between the 0-1 transition frequency, ω_{01}, and the 1-2 transition frequency, ω_{12}, is $\omega_{01} - \omega_{12} = E_C/h$. This quantity is called anharmonicity of the transmon. It is important that the anharmonicity is non-zero because it allows us to differentiate between the 0-1 transition and 1-2 transition when driving the state of the transmon.

In this study, a transmon device is used as a three-state quantum mechanical system, *qutrit*, in contrast to a two-state system, *qubit*. The state of the qutrit can be presented as a superpositions of the three lowermost eigenstates $|\phi_i\rangle$ of the Hamiltonian operator $\hat{H}$. Any pure state $|\phi\rangle$ of the transmon device can thus be written in the form

$$|\phi\rangle = \alpha_0 |\phi_0\rangle + \alpha_1 |\phi_1\rangle + \alpha_2 |\phi_2\rangle$$

where α_i are the probability amplitudes of the states such that $|\alpha_i|^2 = p_i$ and $p_0 + p_1 + p_2 = 1$.

The state of the transmon device can be altered using so-called Rabi oscillations.

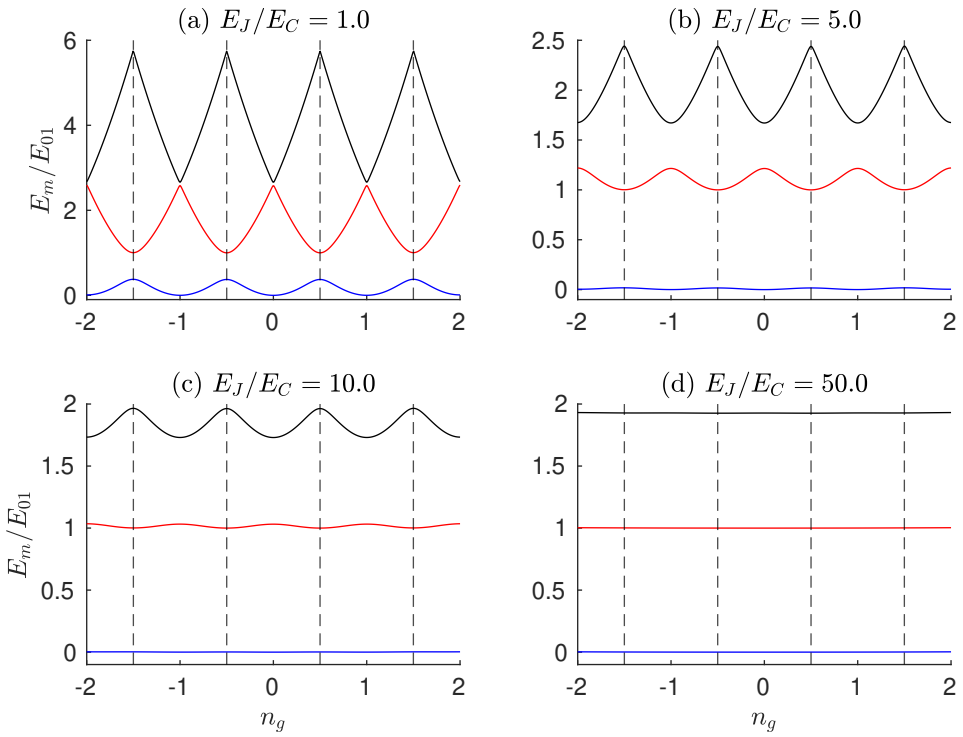


Figure 4: Three lowest eigenenergies of (5) plotted against the offset charge n_g for different values of E_J/E_C. The plots are normalized such that the zero point of energy is at the lowest point of $m = 0$ level, and the energy of the first exited state is 1 at the degeneracy point $n_g = 1/2$.

When a tuned microwave field with frequency ω_d close to the state transition frequency ω_{01} is applied to the sample, the probability amplitudes α_0 and α_1 start to oscillate. By selecting length of the microwave pulse correctly, the probability amplitudes α_0 and α_1 of a pure state can be swapped such that the new state of the transmon is

$$|\phi'\rangle = \alpha_1 |\phi_0\rangle + \alpha_0 |\phi_1\rangle + \alpha_2 |\phi_2\rangle .$$

This kind of pulse is called a π_{01} pulse, and a similar pulse π_{12} exists also for the 1-2 transition. For more complete introduction to Rabi oscillations with transmon, see [3]

3 Methods

It is well known that the state of a quantum mechanical system at the thermodynamic equilibrium is distributed according to the Boltzmann distribution

$$p_i \propto \exp(-E_i/kT) \tag{6}$$

where p_i is the occupation probability of the given state, E_i is the energy of the i^{th} state, k is the Boltzmann constant, and T is the temperature of the system. When the energy of the system is measured, the state collapses to one of the basis states $|\phi_i\rangle$ of the Hamiltonian (5), in proportion to the Boltzmann distribution. An interesting consequence of this, and the fact that the Boltzmann distribution is dependent on the temperature of the system, is that if we could experimentally measure the state distribution p_i of the system, we could use that to determine its temperature T.

3.1 Measurement setup

A transmon, described in Section 2.4, is used as the quantum mechanical system in this study. The transition frequency of the transmon can be adjusted using the magnitude of magnetic flux passing through the SQUID loop. This allows repeating the measurement with multiple transition frequencies and to compare the results between measurements. The temperature measured at different transition frequencies should be identical.

At the start and between the measurements, the transmon sample is left to relax to its equilibrium state which can be written using (6) as

$$\rho = p_0 |\phi_0\rangle \langle\phi_0| + p_1 |\phi_1\rangle \langle\phi_1| + p_2 |\phi_2\rangle \langle\phi_2| . \tag{7}$$

The time given for relaxation is many times longer than the average coherence time of the used transmon device to ensure that the system reaches its equilibrium state. Once the equilibrium state has been reached, the state may be permuted into another using the π pulses. For example applying π_{01} pulse to the state above produces

$$p_1 |\phi_0\rangle \langle\phi_0| + p_0 |\phi_1\rangle \langle\phi_1| + p_2 |\phi_2\rangle \langle\phi_2| .$$

Up to three pulses are used to reach the wanted state, after which the state is measured. The measurement is done by sending a 2 µs measurement pulse and measuring the response.

Let's denote the response pulse corresponding to basis state $|\phi_i\rangle$ as φ_i. The act of measuring the state collapses the quantum state into one of its basis states $|\phi_i\rangle$, and thus the response signal is one of φ_i. To reduce the noise produces by the amplification of the measurement signal, the measurement is repeated 60 000 times. The averaged response signal has response signals φ_i of the basis states in the same proportion as the transmon device is populating its basis states $|\phi_i\rangle$. Thus the measurement of (7) would give

$$x_0(t) = p_0\varphi_0(t) + p_1\varphi_1(t) + p_2\varphi_2(t).$$

For simplicity the dependency of the signals on time t won't be explicitly written from this point onwards.

The equilibrium state shown in (7) can be permuted into six different states before the measurement by applying different sequences of π pulses. Table 1 show the π pulse sequences and the signals they produce. The x_i and y_i signals differ only in that y_i signal has one additional π_{12} pulse applied before other pulses.

Table 1: Different π pulse sequences, the corresponding states, and the averaged signals signals they produce.

Pulse sequence	State	Signal
no pulse	$p_0 \left\|\phi_0\right\rangle \left\langle\phi_0\right\| + p_1 \left\|\phi_1\right\rangle \left\langle\phi_1\right\| + p_2 \left\|\phi_2\right\rangle \left\langle\phi_2\right\|$	$x_0 = p_0\varphi_0 + p_1\varphi_1 + p_2\varphi_2$
π_{01}	$p_1 \left\|\phi_0\right\rangle \left\langle\phi_0\right\| + p_0 \left\|\phi_1\right\rangle \left\langle\phi_1\right\| + p_2 \left\|\phi_2\right\rangle \left\langle\phi_2\right\|$	$x_1 = p_1\varphi_0 + p_0\varphi_1 + p_2\varphi_2$
$\pi_{01}\pi_{12}$	$p_1 \left\|\phi_0\right\rangle \left\langle\phi_0\right\| + p_2 \left\|\phi_1\right\rangle \left\langle\phi_1\right\| + p_0 \left\|\phi_2\right\rangle \left\langle\phi_2\right\|$	$x_2 = p_1\varphi_0 + p_2\varphi_1 + p_0\varphi_2$
π_{12}	$p_0 \left\|\phi_0\right\rangle \left\langle\phi_0\right\| + p_2 \left\|\phi_1\right\rangle \left\langle\phi_1\right\| + p_1 \left\|\phi_2\right\rangle \left\langle\phi_2\right\|$	$y_0 = p_0\varphi_0 + p_2\varphi_1 + p_1\varphi_2$
$\pi_{12}\pi_{01}$	$p_2 \left\|\phi_0\right\rangle \left\langle\phi_0\right\| + p_0 \left\|\phi_1\right\rangle \left\langle\phi_1\right\| + p_1 \left\|\phi_2\right\rangle \left\langle\phi_2\right\|$	$y_1 = p_2\varphi_0 + p_0\varphi_1 + p_1\varphi_2$
$\pi_{12}\pi_{01}\pi_{12}$	$p_2 \left\|\phi_0\right\rangle \left\langle\phi_0\right\| + p_1 \left\|\phi_1\right\rangle \left\langle\phi_1\right\| + p_0 \left\|\phi_2\right\rangle \left\langle\phi_2\right\|$	$y_2 = p_2\varphi_0 + p_1\varphi_1 + p_0\varphi_2$

Figure 5 shows the measured signals in one of the measurement. The signal of the equilibrium state, x_1, is shown as is while the other measurements are shown relative to that signal. This is done to visualize the small differences between the signals which would be otherwise difficult to see. These differences have amplitude of only 2 % to 3 % of the total signal amplitude.

The data used in this study is measured at 41 different frequencies between 5.54 GHz and 6.22 GHz. The measurements are done in three batches, each batch in decreasing order. The first batch is from 5.74 GHz to 5.59 GHz, the second is from 6.22 GHz to 5.78 GHz, and the last is from 5.58 GHz to 5.54 GHz. The measurement is repeated 16 times at each frequency.

Sampling rate of 1 GHz is used in the measurements. The response signal is measured for little over 2 µs giving 2043 samples for each measurement. Only samples 60 to 500 are used in the data analysis due to the first around 50 first samples being close to 0, and because the state being measured relaxes with the time and thus the last samples have more error than the first ones.

3.2 Data analysis methods

Once the averaged signals described in Table 1 are measured, they need to be analyzed in order to get the temperature of the sample. To do this, we need to first solve for the occupation probabilities p_i of the states and then use the Boltzmann distribution (6) to convert those probabilities into temperatures.

Solving the temperature directly from these equations is difficult due to the unknown response signals φ_i of each basis state $|\phi_i\rangle$, so instead of solving the probabilities directly, we examine the differences of the measurement signals, for

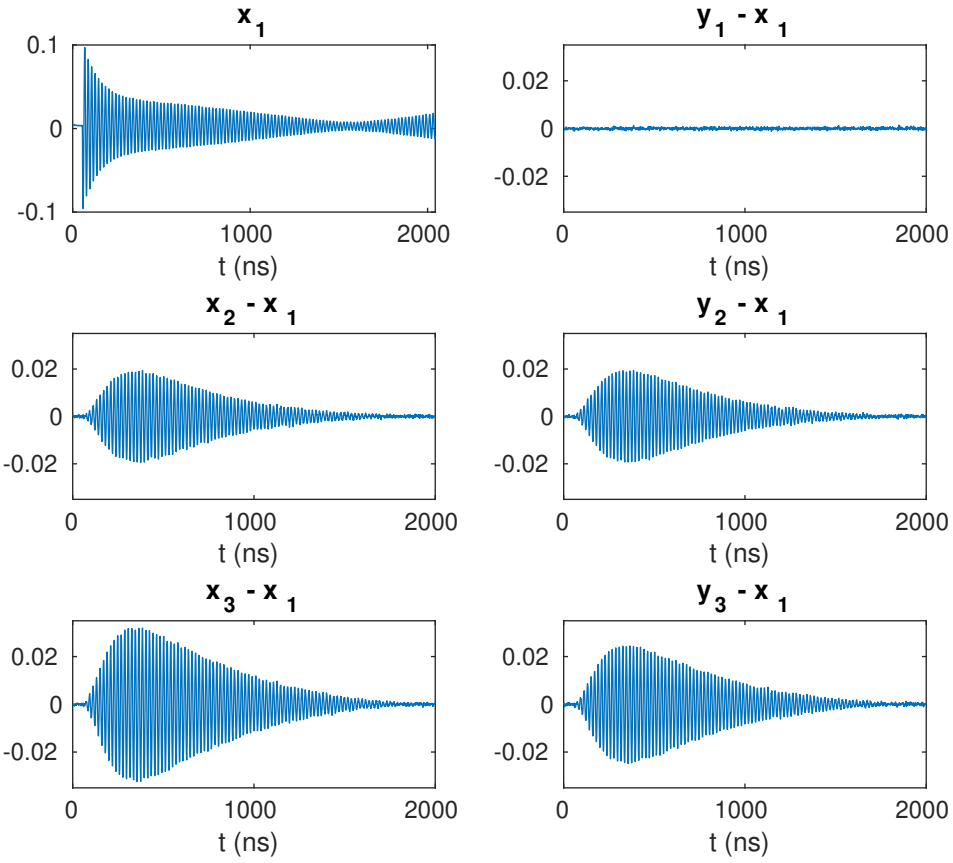


Figure 5: Examples of measurement signals. The first plot shows one of the measured signals (x_1). The rest of the plots show how the other measurement signals differ from the first one. This is done to better show the small differences between the measured signals.

example

$$\begin{aligned} x_2 - y_2 &= (p_1 - p_2)\,(\varphi_0 - \varphi_1) \\ x_0 - x_1 &= (p_0 - p_1)\,(\varphi_0 - \varphi_1) \\ y_0 - y_1 &= (p_0 - p_2)\,(\varphi_0 - \varphi_1)\,. \end{aligned}$$

These equations have the same difference of the unknown response signals $\varphi_0 - \varphi_1$ as factor on the right hand side. This hints us to divide the equations with each others to dispose the unknown response signals of the basis states, leaving behind only the

measured signals x_i, y_i, and the probabilities p_i we desires to solve:

$$\begin{aligned}\frac{p_1 - p_2}{p_0 - p_1} &= \frac{x_2 - y_2}{x_0 - x_1} \\ \frac{p_1 - p_2}{p_0 - p_2} &= \frac{x_2 - y_2}{y_0 - y_1} \\ \frac{p_0 - p_1}{p_0 - p_2} &= \frac{x_0 - x_1}{y_0 - y_1}.\end{aligned} \tag{8}$$

We can write these quotients of differences of signals for other pairs of signals and replace the p_i in the equation with the Boltzmann distribution (6), noting that the constant factor cancels out, to get

$$\begin{aligned}\frac{p_1 - p_2}{p_0 - p_1} &= \frac{\exp(\omega_{01}/kT) - \exp(\omega_{02}/kT)}{1 - \exp(\omega_{01}/kT)} = \frac{x_2 - y_2}{x_0 - x_1} = \frac{x_1 - y_1}{y_0 - x_2} = \frac{x_0 - y_0}{y_1 - y_2} = a \\ \frac{p_1 - p_2}{p_0 - p_2} &= \frac{\exp(\omega_{01}/kT) - \exp(\omega_{02}/kT)}{1 - \exp(\omega_{02}/kT)} = \frac{x_2 - y_2}{y_0 - y_1} = \frac{x_1 - y_1}{x_0 - y_2} = \frac{x_0 - y_0}{x_1 - x_2} = b \\ \frac{p_0 - p_1}{p_0 - p_2} &= \frac{1 - \exp(\omega_{01}/kT)}{1 - \exp(\omega_{02}/kT)} = \frac{x_0 - x_1}{y_0 - y_1} = \frac{y_0 - x_2}{x_0 - y_2} = \frac{y_1 - y_2}{x_1 - x_2} = c.\end{aligned} \tag{9}$$

In reality the measurement signals are noisy, so computing the fraction and expecting to get a single quotient as a result would be in vain. Instead, we can think of the fraction as a linear relation between the differences of the signals, and the quotient as the factor of that relation. When thought in this way, the first equation of (8) can be written as

$$x_2 - y_2 = \underbrace{\frac{p_1 - p_2}{p_0 - p_1}}_{a}(x_0 - x_1)$$

where a is the factor of linear relation. Now, to solve for a, we need fit a linear regression. In this case we can't use *simple linear regression* as it expects there to be dependent and independent variables where the independent variable is assumed to be noise free. In our case this is not true as both of the sides have measurement signals with some amount of noise and neither of them can be thought of as the independent or dependent variable.

To correctly fit the factor in the model, we need to use *Deming regression* [10]. Deming regression allows us to fit a linear model between two sets of data where both data sets have a known, but possibly different, amount of Gaussian noise. The Deming regression is symmetrical in a sense that we could swap the signals on both sides of the equation and still get the same regression, though with fitting factor a^{-1}. More about the Deming regression and derivation is presented in Appendix B.

The ratio of the variance of the noises, δ, is needed for the Deming regression. The special case where both signals have the same amount of noise, and thus $\delta = 1$, is called *orthogonal regression*. In this study both signals being fitted to are differences of two measurement signals and we can thus assume that they have the same amount of noise. Thus we use orthogonal regression to fit the coefficient.

One challenge with Deming regression is that it doesn't provide an estimate for the error in the fitted variable. To get an estimate, we need to rely on more general error estimation methods, like *jackknifing* [11] which is used in this study. Jackknifing is a resampling method for estimating variance of a given parameter. It works by systematically leaving out one of the samples and computing the parameter for the rest of the samples. The variance of the parameter can then be computed using

$$\mathrm{Var}(a) = \frac{n-1}{n}\sum_{i=1}^{n}(a_i - a)^2$$

where a_i is the value of the parameter when sample i has been left out and a is the value for the parameter when all n data samples are being considered. We can use Student's t-distribution to get the 95 % confidence interval for the actual value of the parameter:

$$\mathrm{ConfidenceIntervalSize} = \mathrm{Var}(a)F^{-1}\left(1 - \frac{\alpha}{2}, n-1\right)$$

where $\alpha = 1 - 95\,\%$.

Once we have computed values and confidence intervals for a, b and c in (9), we can use numerical equation solving methods for computing the temperature T of the sample. There are three different equations presented for each of the coefficients a, b and c, giving nine equations in total. This also means that there are nine different ways to compute the temperature, each with different characteristics. In an idealized world, all of these equations would produce the same result for the same set of measurements, but due to imperfections and errors in the measurement setup, some of these equations produce better predictions than others.

3.3 Data characteristics

Figure 6 shows all of the differences between two signals. Most of the signals have a clear and similar envelope, but there are two notable exceptions: $x_1 - y_1$ and $x_0 - y_0$. These difference signals are an order of magnitude weaker than other signals and don't have a clear envelope. This is important as in (9) the signals in the same columns in Figure 6 are compared together and thus they should have the same shape to get a good fit. This is not the case with $x_1 - y_1$ and $x_0 - y_0$, and thus equations where they are present shouldn't be used to determine the temperature.

The fact that the signals which have least structure are $x_1 - y_1$ and $x_0 - y_0$ is not surprising. When writing these differences in terms of the hidden signals φ_i, we get

$$\begin{aligned} x_0 - y_0 &= (p_1 - p_2)(\varphi_1 - \varphi_2) \\ x_1 - y_1 &= (p_1 - p_2)(\varphi_0 - \varphi_2). \end{aligned}$$

At the temperature and frequency scale the measurements are done, $p_0 \approx 90\,\%$, $p_1 \approx 10\,\%$, and $p_2 \approx 1\,\%$. This means that there is only one tenth of the information available for these signals compared to those where p_0 is not canceled out.

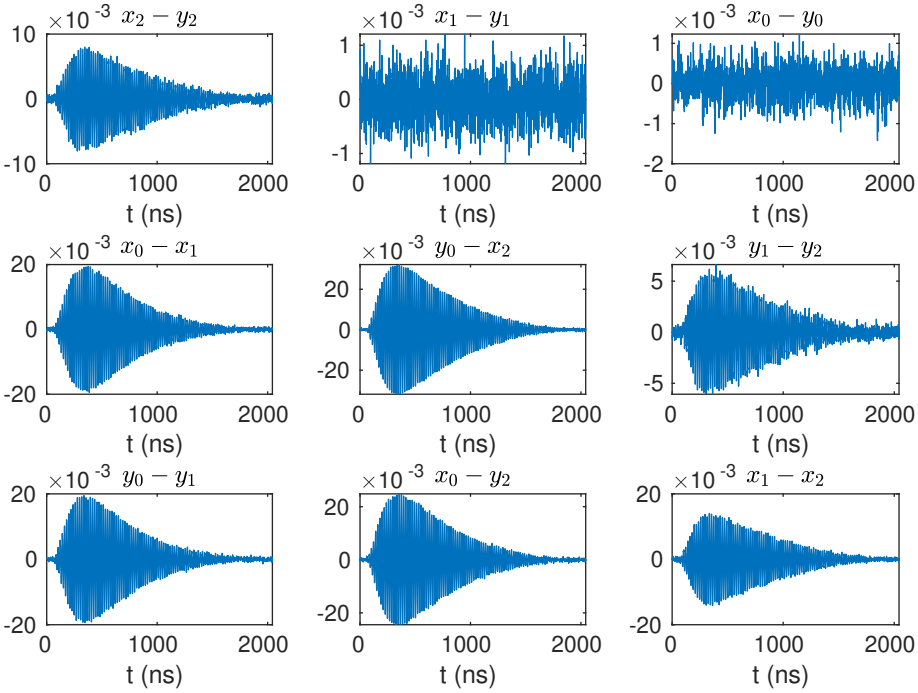


Figure 6: Differences between measured signals.

It is little surprising that $x_2 - y_2$ has such prominent and clear envelope. One possible explanation is that to produce y_2 signal, we must apply three consecutive π pulses to the sample, which takes longer than just one or two pulses needed for other signals. This means that the sample has more time to decay into other states making the readout less accurate.

This all means that using the following set of equations to determine the temperature should provide the best results

$$\begin{aligned}
\frac{\exp(\omega_{01}/kT) - \exp(\omega_{02}/kT)}{1 - \exp(\omega_{01}/kT)} &= \frac{x_2 - y_2}{x_0 - x_1} = a \\
\frac{\exp(\omega_{01}/kT) - \exp(\omega_{02}/kT)}{1 - \exp(\omega_{02}/kT)} &= \frac{x_2 - y_2}{y_0 - y_1} = b \\
\frac{1 - \exp(\omega_{01}/kT)}{1 - \exp(\omega_{02}/kT)} &= \frac{x_0 - x_1}{y_0 - y_1} = \frac{y_0 - x_2}{x_0 - y_2} = \frac{y_1 - y_2}{x_1 - x_2} = c.
\end{aligned}$$

If we also disregards those measurements where y_2 is present due to it having most time to decay, we are left with only the following equation to solve

$$\frac{1 - \exp(\omega_{01}/kT)}{1 - \exp(\omega_{02}/kT)} = \frac{x_0 - x_1}{y_0 - y_1} = c. \tag{10}$$

In the following section, it is actually seen that this equation yields one of the best-looking results, but there are also other pairs of differences of signals that produce similar-looking results.

4 Results

All six states described in Table 1 were constructed and the corresponding signals were measured. The equations (9) were used to determine coefficients a, b and c for all nine relevant pairs of differences of two signals. The coefficients were converted to temperatures by solving the quotients of occupation probabilities in the equations.

Figure 7 shows the coefficients plotted as functions of temperature. The figure shows that the coefficients can have values only in a limited range. When the coefficients are determined from the measurement data, their values are not necessarily limited to this range. In those cases, the temperature can't be determined. Three of the pairs of differences of signals produced invalid values for coefficients at all driving frequencies and one produced invalid values at most of the frequencies.

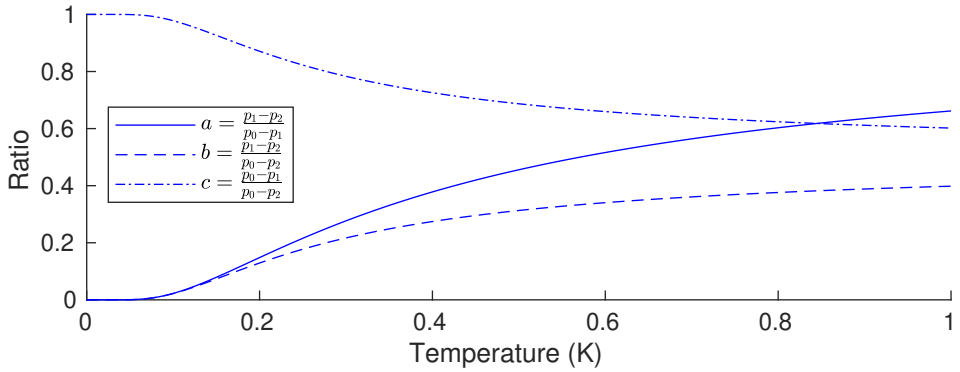


Figure 7: Values of coefficients a, b and c as functions of temperature. The temperature is converted into occupation probabilities p_i using the Boltzmann distribution (6).

Figure 8 shows the temperature determined using (10) which was postulated to be the most accurate. Most of the measurements are between 45 mK and 65 mK, but there are some outliers. The mean of all temperatures is 56 mK. The 95 % confidence interval for most of the measurements is about ±7 mK.

Quotients between signals $x_0 - y_0$ and $x_1 - x_2$, and signals $x_1 - y_1$ and $x_0 - y_2$ show similar results. These are plotted in Figure 9a and Figure 9b, respectively. Both of these figures agree with the temperatures determined using (10) to a reasonable accuracy, though they still produce somewhat different values and error bounds for the temperature. It is noteworthy that the former figure is determined using difference $x_0 - y_0$ and the latter using $x_1 - y_1$, both of which were postulated to produce worse results than other combinations of measurement signals.

Appendix A shows that other pairs of differences of signals actually produce wildly different results than the three presented here. These can be safely ignored as it is known that the temperature of the sample was around 100 mK during the measurements.

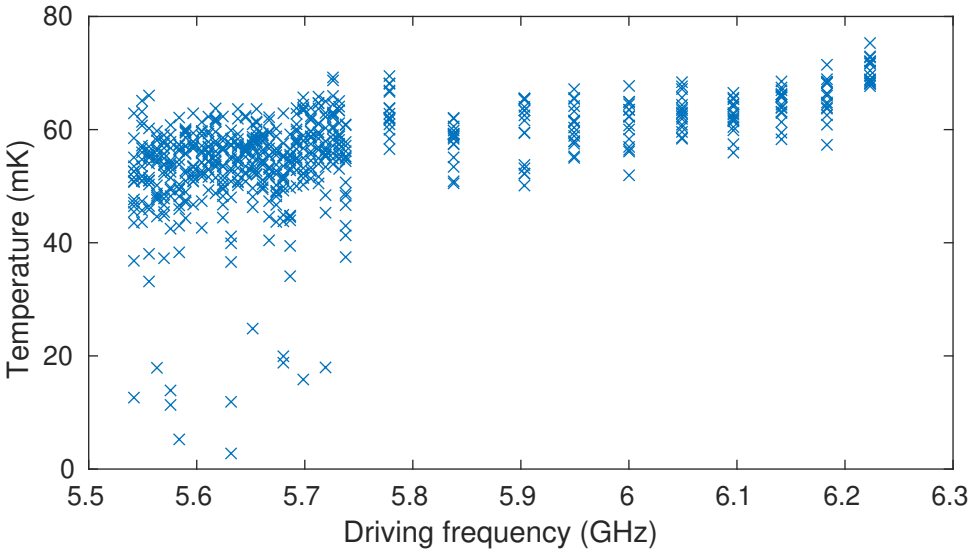
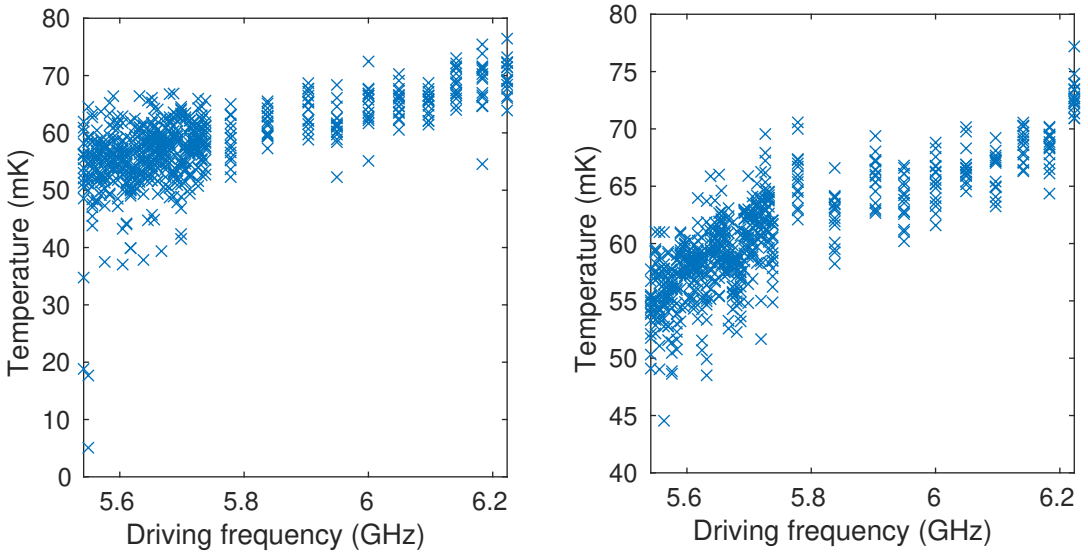


Figure 8: Temperature of the sample as a function of driving frequency determined using (10).



(a) Determined using $x_0 - y_0$ and $x_1 - x_2$. (b) Determined using $x_1 - y_1$ and $x_0 - y_2$.

Figure 9: Temperature of the sample as a function of driving frequency determined using two different pairs of differences of measurement signals.

4.1 Frequency dependency of the temperature

The temperature of the sample shouldn't depend on the driving frequency of the transmon device. We used Kolmogorov–Smirnov test [12] to examine whether this is true. We performed the test pairwise on all 41 different driving frequencies of the transmon device for the temperatures determined using (10). The null hypothesis

was that there is no difference between the pairs of measured distributions at different frequencies.

Figure 10 shows the p-value for the rejection of the null hypothesis. The lower the p-value, the more likely it is that the driving frequency affects the distribution of determined temperatures. Cells where $p < 5\,\%$ are shown in bright red. The cells are not evenly spaced.

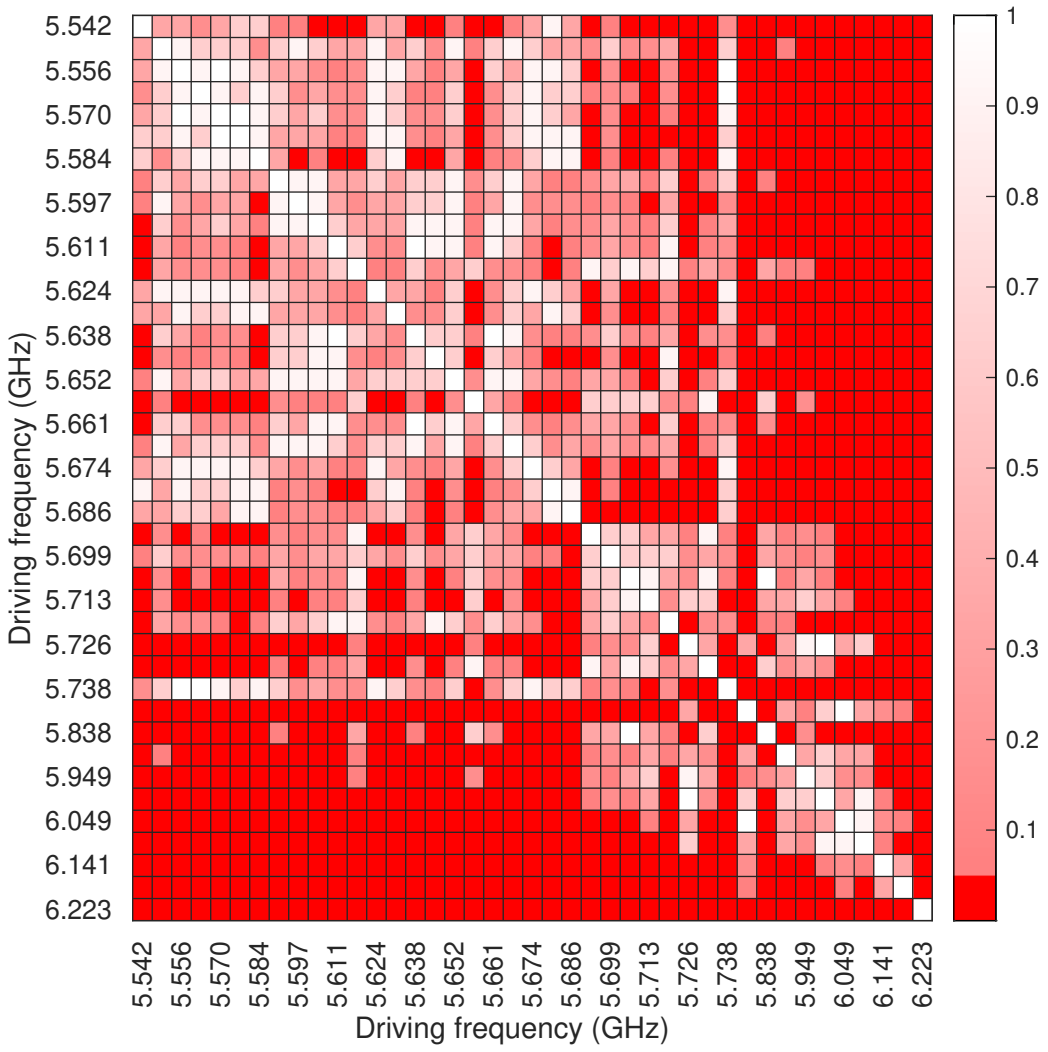


Figure 10: p-values from pairwise Kolmogorov–Smirnov tests run on all samples collected at 41 different driving frequencies. The bright red color shows where $p < 5\,\%$.

The figure shows that the determined temperatures agree with each others at frequencies between 5.542 GHz and 5.738 GHz, and also between 5.778 GHz and 6.183 GHz, with couple outliers. When comparing between these two ranges, the temperatures don't agree that well. It is notable that in the higher range the difference between adjacent measurements is larger than in the lower range, and that is likely a

partial reason why there is so hard discrepancy between the lower and upper ranges.

One thing to note is that the Kolmogorov–Smirnov test tests the similarity of the *distributions* of the determined temperatures, not just the values, so the discrepancy might also be partially due to the different precision characteristics of the method at different driving frequencies. Still, Figure 8 shows a slight rising trend of the temperature as a function of driving frequency, which, together with the results from Kolmogorov–Smirnov test, hint towards a conclusion that the determined temperature somewhat depends on the driving frequency of the transmon device.

5 Conclusion

The method developed in this study produces plausible values for the temperature of the sample. The temperatures have slight dependency on the temperature. More measurements are needed to determine what causes this temperature dependency. The measurements should be performed in a random order of driving frequency to eliminate systematic increase or decrease of the temperature of the sample during the measurements. In this study, the measurements were done in three consecutive batches of measurements in decreasing order during a 72 hour period, which may make the results presented susceptible to systematic bias.

The number of samples that are averaged together should be varied in future studies. Increasing the number of samples lowers the noise of measurements and allows more samples with response signals corresponding to p_1 and p_2 to be recorded. On the downside, increasing number of samples increases the time the measurement takes and allows more random temporal variation of the temperature to occur.

One method worth considering would be to remove the averaging of samples altogether and use a clustering algorithm to cluster the measured response signals into three classes. This way one could compute the occupancy probabilities of the states without using the somewhat complicated method described in this study. This might not be possible in practice as the response signals to different states might differ too little to be distinguishable above the noise level. The raw, unaveraged measurement signals were not available for this study so the viability of this method couldn't be evaluated.

References

[1] Frank Arute et al. “Quantum supremacy using a programmable superconducting processor”. In: *Nature* 574.7779 (2019), pp. 505–510.

[2] Jens Koch et al. “Charge-insensitive qubit design derived from the Cooper pair box”. In: *Physical Review A* 76.4 (2007), p. 042319.

[3] Sergey Danilin. “Experiments with a transmon artificial atom - state manipulation and detection of magnetic fields”. In: Aalto University publication series DOCTORAL DISSERTATIONS, 76/2018 (2018).

[4] BRIAN D Josephson. “The discovery of tunnelling supercurrents”. In: *Reviews of Modern Physics* 46.2 (1974), p. 251.

[5] J. E. Zimmerman and A. H. Silver. “Macroscopic Quantum Interference Effects through Superconducting Point Contacts”. In: *Phys. Rev.* 141 (1 Jan. 1966), pp. 367–375. DOI: 10.1103/PhysRev.141.367.

[6] Leon N Cooper. “Bound electron pairs in a degenerate Fermi gas”. In: *Physical Review* 104.4 (1956), p. 1189.

[7] RP Feynman, RB Leighton, and M Sands. *The Schrödinger equation in a classical context: a seminar on superconductivity (Chapter 21) The Feynman lectures on physics, Vol. III Quantum Mechanics*. 1995.

[8] R. C. Jaklevic et al. “Quantum Interference Effects in Josephson Tunneling”. In: *Phys. Rev. Lett.* 12 (7 Feb. 1964), pp. 159–160. DOI: 10.1103/PhysRevLett.12.159.

[9] G Wendin. “Quantum information processing with superconducting circuits: a review”. In: *Reports on Progress in Physics* 80.10 (2017), p. 106001.

[10] William Edwards Deming. *Statistical adjustment of data*. Wiley, 1943.

[11] Rupert G Miller. “The jackknife-a review”. In: *Biometrika* 61.1 (1974), pp. 1–15.

[12] Nikolai V Smirnov. “Estimate of deviation between empirical distribution functions in two independent samples”. In: *Bulletin Moscow University* 2.2 (1939), pp. 3–16.

[13] Chas. H. Kummell. “Reduction of Observation Equations Which Contain More Than One Observed Quantity”. In: *The Analyst* 6.4 (1879), pp. 97–105. ISSN: 07417918.

[14] Kristian Linnet. “Evaluation of regression procedures for methods comparison studies.” In: *Clinical chemistry* 39.3 (1993), pp. 424–432.

[15] Peter Nagy. *Deming regression*. Retrieved December 4, 2019. 2019. URL: https://www.mathworks.com/matlabcentral/fileexchange/55056-deming-regression.

A Determined temperatures

In the main report, nine equations were found for determining the temperature of the sample

$$\begin{aligned} a &= \frac{x_2 - y_2}{x_0 - x_1} \quad a = \frac{x_1 - y_1}{y_0 - x_2} \quad a = \frac{x_0 - y_0}{y_1 - y_2} \\ b &= \frac{x_2 - y_2}{y_0 - y_1} \quad b = \frac{x_1 - y_1}{x_0 - y_2} \quad b = \frac{x_0 - y_0}{x_1 - x_2} \\ c &= \frac{x_0 - x_1}{y_0 - y_1} \quad c = \frac{y_0 - x_2}{x_0 - y_2} \quad c = \frac{y_1 - y_2}{x_1 - x_2} \end{aligned}$$

where

$$\begin{aligned} a &= \frac{p_1 - p_2}{p_0 - p_1} = \frac{\exp(\omega_{01}/kT) - \exp(\omega_{02}/kT)}{1 - \exp(\omega_{01}/kT)} \\ b &= \frac{p_1 - p_2}{p_0 - p_2} = \frac{\exp(\omega_{01}/kT) - \exp(\omega_{02}/kT)}{1 - \exp(\omega_u 02/kT)} \\ c &= \frac{p_0 - p_1}{p_0 - p_2} = \frac{1 - \exp(\omega_{01}/kT)}{1 - \exp(\omega_{02}/kT)}. \end{aligned}$$

Of these equations, one was deemed to be most likely to give most accurate results:

$$c = \frac{p_0 - p_1}{p_0 - p_2} = \frac{1 - \exp(\omega_{01}/kT)}{1 - \exp(\omega_{02}/kT)} = \frac{x_0 - x_1}{y_0 - y_1}$$

The temperatures determined using this equation are shown in Figure A1.

Two other pairs of differences of measurements also produced similar results:

$$b = \frac{x_1 - y_1}{x_0 - y_2}$$

shown in Figure A2, and

$$b = \frac{x_0 - y_0}{x_1 - x_2}$$

shown in Figure A3. The following pairs of differences of measurements produced differing results

$$a = \frac{x_1 - y_1}{y_0 - x_2} \quad a = \frac{x_0 - y_0}{y_1 - y_2} \quad c = \frac{y_1 - y_2}{x_1 - x_2}.$$

They are shown in Figures A4, A5 and A6 respectively. The last of these produced only barely valid values for coefficient c on some of the measurements and invalid values for the rest. Only invalid values for the coefficients were produced for the following equations:

$$a = \frac{x_2 - y_2}{x_0 - x_1} \quad b = \frac{x_2 - y_2}{y_0 - y_1} \quad c = \frac{y_0 - x_2}{x_0 - y_2}$$

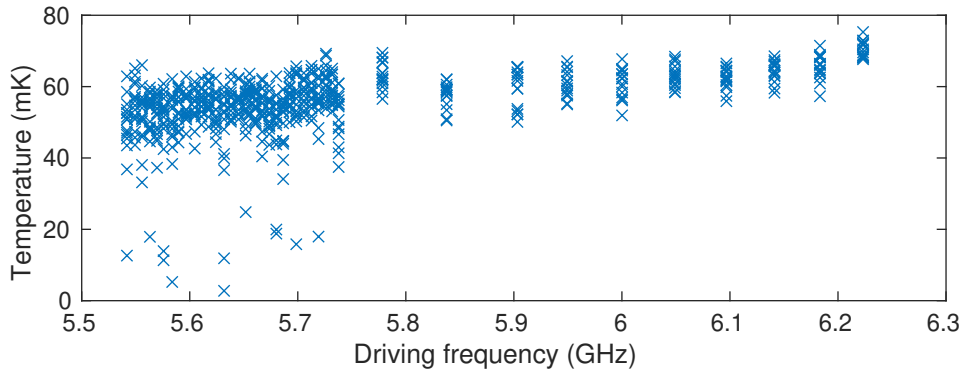


Figure A1: Temperature determined using signals $x_0 - x_1$ and $y_0 - y_1$ as function of driving frequency of the transmon device.

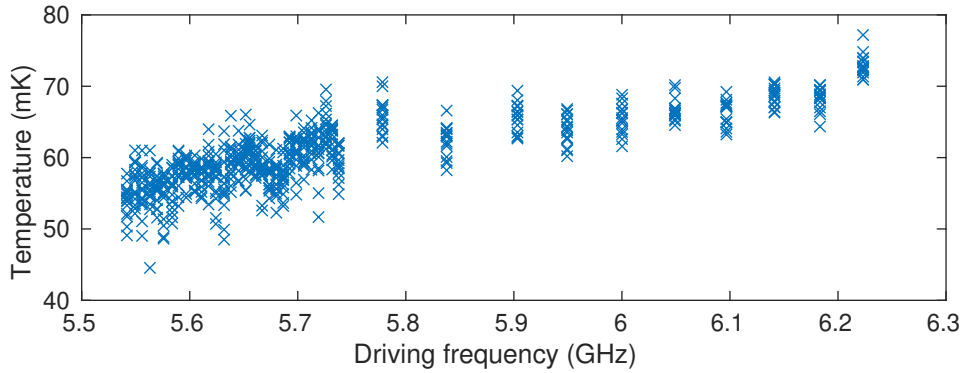


Figure A2: Temperature determined using signals $x_1 - y_1$ and $x_0 - y_2$ as function of driving frequency of the transmon device.

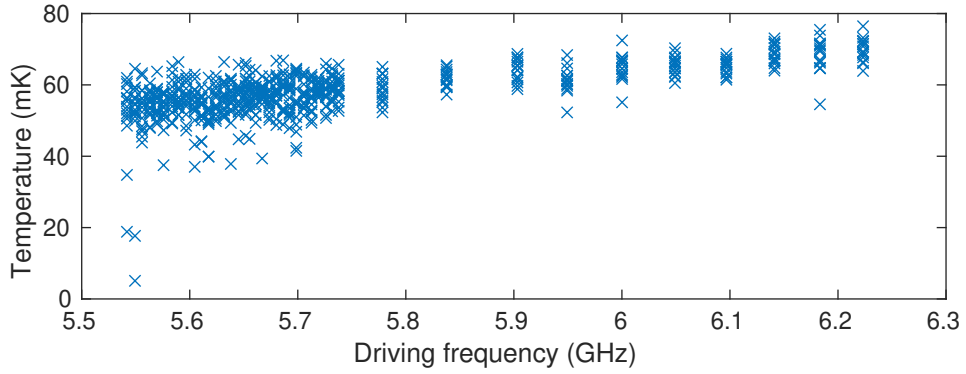


Figure A3: Temperature determined using signals $x_0 - y_0$ and $x_1 - x_2$ as function of driving frequency of the transmon device.

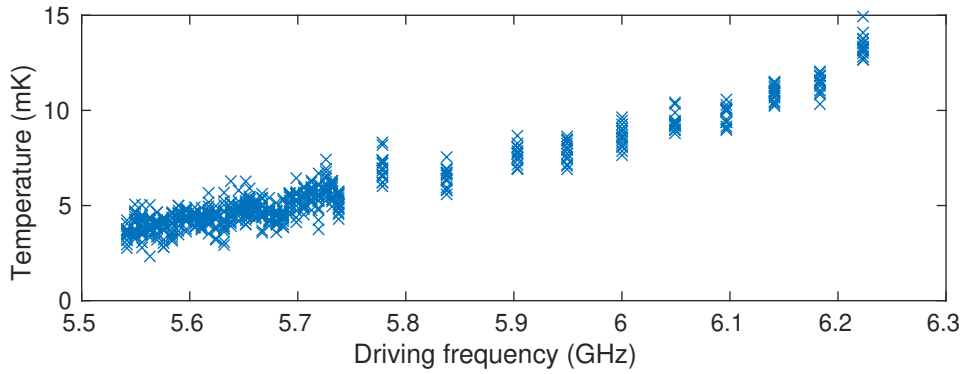


Figure A4: Temperature determined using signals $x_1 - y_1$ and $y_0 - x_2$ as function of driving frequency of the transmon device.

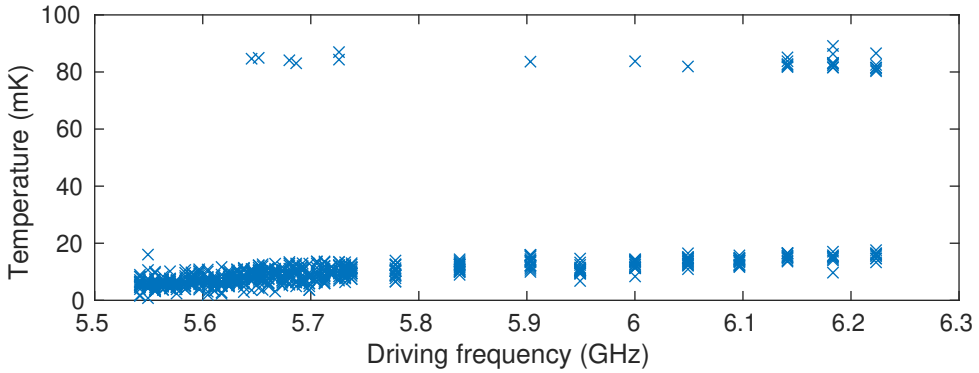


Figure A5: Temperature determined using signals $x_0 - y_0$ and $y_1 - y_2$ as function of driving frequency of the transmon device.

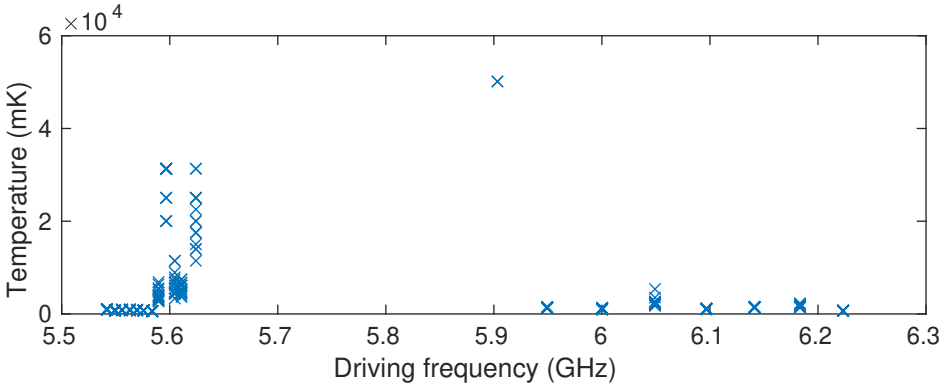


Figure A6: Temperature determined using signals $y_1 - y_2$ and $x_1 - x_2$ as function of driving frequency of the transmon device.

B Deming regression

In physics, and data analysis in general, there is often a need to fit a model to the measured data. Of special interest in this study is a linear model through the origin

$$y \approx \beta x \tag{B1}$$

which is a special case of the more general linear model with an intercept parameter

$$y \approx \beta_1 x + \beta_0. \tag{B2}$$

Both of these models can be fitted using *simple linear regression* scheme taught at the basic-level data science courses. The simple linear regression fit for (B1) is given by

$$\beta = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

The reason why simple linear regression is not sufficient for use in this study is that it minimizes the the error only in y direction. This means that swapping x and y coordinates produce a different fit to the measurement data and thus a different coefficient. This is sufficient when we assume that the measurement error in x direction is negligible. In this study, both x and y variables are differences of two measurement signals and thus both of them have approximately same amount of measurement error in them. Hence we need another way for fitting the model, namely one which minimizes the error of the fit in both x and y direction simultaneously. This difference is visualized in Figure B1, where the black vertical and diagonal lines show the differences being minimized between measurement and model.

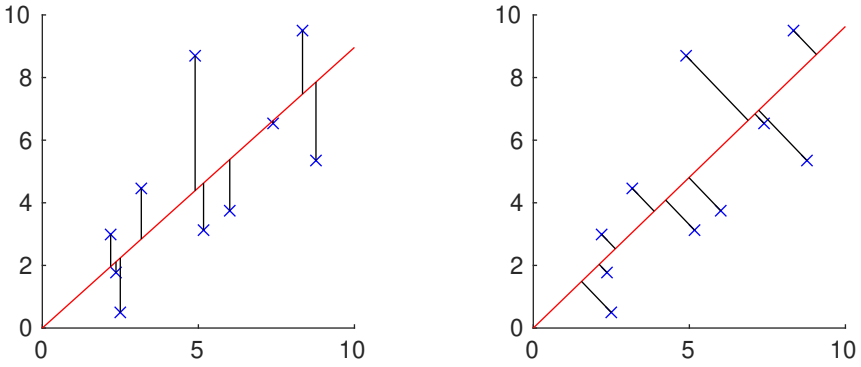


Figure B1: Difference between simple linear regression (left) and Deming regression (right). Blue crosses are the data and red line is the fitted model. Black lines are the distances being minimized.

The original version, first found by Kummell [13], and later popularized by Deming [10], actually fit the model (B2) with intercept parameter. Linnet [14] later showed that the Deming regression is equivalent to maximum likelihood estimate (MLE) of the (B2) model. In the following we will derive the intercept-free Deming regression following in the steps of Peter Nagy [15].

Let (x_i, y_i) be the i$^{\text{th}}$ data sample, and let ξ_i be the optimal x coordinate for the i$^{\text{th}}$ data sample. Let δ be the ratio between variance in x and y directions, $\delta = \frac{\sigma_y^2}{\sigma_x^2}$. This allows us to write $\sigma_x^2 = \sigma^2$ and $\sigma_y^2 = \delta\sigma^2$. Let's start by writing the likelihood formula for (B1)

$$l = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\xi_i)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\delta\sigma^2}} e^{-\frac{(y_i-\beta\xi_i)^2}{2\delta\sigma^2}}. \tag{B3}$$

The logarithm is a monotonic function, and thus the maximum of logarithm of the likelihood is also the maximum of the likelihood. The log-likelihood is

$$L = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{n}{2}\log(2\pi\delta\sigma^2) - \frac{\sum_{i=1}^{n}(x_i-\xi_i)^2}{2\sigma^2} - \frac{\sum_{i=1}^{n}(y_i-\beta\xi_i)^2}{2\delta\sigma^2}. \tag{B4}$$

A necessary but not sufficient condition for the maximum is that the partial derivatives of (B4) w.r.t. ξ_i must be 0

$$0 = \frac{\partial L}{\partial \xi_i} = \frac{2(x_i-\xi_i)}{2\sigma^2} + \frac{2\beta(y_i-\beta\xi_i)}{2\delta\sigma^2}$$

which can be solved for ξ_i to get

$$\xi_i = \frac{\delta x_i + \beta y_i}{\delta + \beta^2}. \tag{B5}$$

Another necessary condition is that the partial derivative w.r.t. β

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} \frac{\xi_i(y_i-\beta\xi_i)}{\delta\sigma^2}$$

must equal to 0. Substituting (B5) into this yields

$$0 = \frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} \frac{\frac{\delta x_i+\beta y_i}{\delta+\beta^2}(y_i - \beta\frac{\delta x_i+\beta y_i}{\delta+\beta^2})}{\delta\sigma^2} = \frac{\sum_{i=1}^{n}(\delta x_i+\beta y_i)(y_i-\beta x_i)}{(\delta+\beta^2)^2\sigma^2}.$$

Clearly $(\delta+\beta^2)^2 > 0$ and $\sigma^2 > 0$. Thus it suffices to equate the denominator to zero. Expanding the terms yields

$$0 = \delta\sum_{i=1}^{n} x_i y_y - \delta\beta\sum_{i=1}^{n} x_i^2 + \beta\sum_{i=1}^{n} y_i^2 - \beta^2\sum_{i=1}^{n} x_i y_i.$$

By noting

$$S_{xx} = \sum_{i=1}^{n} x_i^2 \qquad S_{yy} = \sum_{i=1}^{n} y_i^2 \qquad S_{xy} = \sum_{i=1}^{n} x_i y_i$$

the above equation can be simplified into

$$0 = \delta S_{xy} - \delta\beta S_{xx} + \beta S_{yy} - \beta^2 S_{xy} = -\beta^2 S_{xy} + (S_{yy} - \delta S_{xx})\beta + \delta S_{xy}.$$

This is just a second degree polynomial equation, whose roots are

$$\hat{\beta}_{\pm} = \frac{S_{yy} - \delta S_{xx} \pm \sqrt{(S_{yy} - \delta S_{xx})^2 + 4\delta S_{xy}^2}}{2S_{xy}}. \tag{B6}$$

These roots can correspond to minima, maxima and saddle points of the likelihood. The only thing left to do is to verify which one of these roots is the maximum of the log-likelihood (B4) and thus also of the likelihood (B3). This can be found out by examining the local curvature of the log-likelihood at the roots $\hat{\beta}_{\pm}$, which can be done by examining the second derivative w.r.t. β. The easiest way to get the second derivative is to substitute (B5) into (B4) to turn the log-likelihood into a function of only β and then computing the second derivative. The substitution yields

$$L = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{n}{2}\log(2\pi\delta\sigma^2) - \frac{\delta S_{xx} + S_{yy} - \frac{\delta^2 S_{xx} + \beta^2 S_{yy} + 2\delta\beta S_{xy}}{\delta + \beta^2}}{2\delta\sigma^2}$$

whose second derivative is

$$\frac{d^2 L}{d\beta^2} = \frac{2\beta^3 S_{xy} - 6\delta\beta S_{xy} - 3\beta^2 S_{yy} + \delta S_{yy} - \delta^2 S_{xx} + 3\delta\beta^2 S_{xx}}{\sigma^2(\delta + \beta^2)^3}.$$

As previously, the nominator is strictly positive. This means that the sign of $\frac{d^2 L}{d\beta^2}$ is the same as the sign of the denominator and it suffices to disregard the nominator. Another trick is to reduce the order of the denominator at the roots $\hat{\beta}_{\pm}$ by noting

$$\begin{aligned}
&2\hat{\beta}_{\pm}^3 S_{xy} - 6\delta\hat{\beta}_{\pm} S_{xy} - 3\hat{\beta}_{\pm}^2 S_{yy} + \delta S_{yy} - \delta^2 S_{xx} + 3\delta\hat{\beta}_{\pm}^2 S_{xx} \\
=&2\hat{\beta}_{\pm} \cdot \underbrace{\left(S_{xy}\hat{\beta}_{\pm}^2 + (\delta S_{xx} - Syy)\hat{\beta}_{\pm} - \delta S_{xy}\right)}_{=0} \\
&+ (\delta S_{xx} - S_{yy})\hat{\beta}_{\pm}^2 - 4\delta S_{xy}\hat{\beta}_{\pm} + \delta S_{yy} - \delta^2 S_{xx} \\
=&(\delta S_{xx} - S_{yy})\hat{\beta}_{\pm}^2 - 4\delta S_{xy}\hat{\beta}_{\pm} + \delta S_{yy} - \delta^2 S_{xx}
\end{aligned}$$

as $\hat{\beta}_{\pm}$ are the roots of $S_{xy}\beta^2 + (\delta S_{xx} - Syy)\beta - \delta S_{xy}$ by definition. Substituting in the values of roots $\hat{\beta}_{\pm}$ from (B5) yields

$$\begin{aligned}
&\left(\sigma^2(\delta + \beta^2)^3\right) \left.\frac{dL^2}{d\beta^2}\right|_{\beta=\hat{\beta}_{\pm}} \\
&= \mp\frac{1}{2S_{xy}^2}\left((S_{yy} - \delta S_{xx})^2 + 4\delta S_{xy}^2\right) \cdot \left(\pm(S_{yy} - \delta S_{xx}) + \sqrt{(S_{yy} - \delta S_{xx})^2 + 4\delta S_{xy}^2}\right).
\end{aligned}$$

Clearly

$$\frac{1}{2S_{xy}^2} > 0 \qquad \text{and} \qquad (S_{yy} - \delta S_{xx})^2 + 4\delta S_{xy}^2 > 0.$$

It can also be seen that

$$|S_{yy} - \delta S_{xx}| < \sqrt{(S_{yy} - \delta S_{xx})^2 + 4\delta S_{xy}^2},$$

which means that

$$\pm(S_{yy} - \delta S_{xx}) + \sqrt{(S_{yy} - \delta S_{xx})^2 + 4\delta S_{xy}^2} > 0.$$

As all factors are positive, the whole product is positive and it's sign is determined only by the sign chosen for $\hat{\beta}_{\pm}$. When selecting $\hat{\beta}_{+}$, the second derivative of the log-likelihood, and thus of the likelihood, is negative meaning that $\hat{\beta}_{+}$ is the optimal choice of parameter β. Likewise $\hat{\beta}_{-}$ yields the pessimal choice for the parameter.