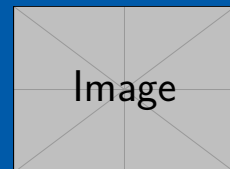# Development of Laser Systems and Spectroscopy of Highly Charged Ions

Zur Erlangung des Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.)
Vorgelegte Dissertation von Patrick Baus aus Mannheim
Tag der Einreichung: 1. März 2023, Tag der Prüfung: 1. März 2023

1. Gutachten: Prof. Dr. Gerhard Birkl
Foobar

Physics Department
Institut für Angewandte Physik
APQ

Development of Laser Systems and Spectroscopy of Highly Charged Ions

Submitted doctoral thesis by Patrick Baus

1. Review: Prof. Dr. Gerhard Birkl

Date of submission: March 1, 2023
Date of thesis defense: March 1, 2023

Foobar

1. Preparation

"Begin at the beginning," the King said gravely, "and go on till you come to the end: then stop."

— Lewis Carroll, *Alice in Wonderland*

1.1. Grounding and Shielding

Add parts from "references
Grounding and Shielding.pdf"

1.2. Laser System

1.2.1. Requirements Laser System

One purpose of the laser system is to be used for the spectroscopy of highly charged ions in a Penning trap. For example, an interesting transition of Ar^{13+} can be found at $\lambda = (441.255\,75 \pm 0.000\,17)$ nm [75] with a lifetime of (9.573 ± 0.006) ms [61], which corresponds to natural linewidth of $\Gamma \approx 2\pi \times (16.63 \pm 0.01)$ Hz. While this linewidth is fairly small, there is substantial doppler broadening at $4\,\mathrm{K}$ of

$$\Delta\nu(\lambda = 441\,\mathrm{nm}, T = 4\,\mathrm{K}, m = 39.948\,\mathrm{u}) = \frac{2}{\lambda}\sqrt{2\ln 2 \frac{k_B T}{m}} \approx 2\pi \times 150\,\mathrm{MHz}\,. \tag{1.1}$$

1.3. Laser Current Driver

Laser diodes are current driven devices, because

$$P_{out} \propto I \, ,$$

and the diode current I approximately follows the Shockley equation [105]

$$I = I_0 \left(e^{\frac{qV_d}{k_B T}} - 1 \right) . \tag{1.2}$$

k_B is Boltzmann constant, T the temperature, q the electron charge and V_d the diode voltage. The exponential dependence of the current on the supply voltage calls for a current source to drive a laser diode safely, without risking thermal damage.

The primary function of a laser driver is to provide a stable, but user adjustable, current. The user adjustable current can typically be modulated at frequencies up to serveral MHz to shape the frequency and ampltitude of the laser beam. Additional features like current and voltage limits aid in protecting the expensive laser diodes. This section deals with the design challenges of such a device used for high precission laser spectroscopy. First the design requirements are derived and then technical specifications are developed.

The focus of this work many lies on two types of laser diodes, indium gallium nitride (InGaN) and aluminium gallium arsenide (AlGaAs), but is not limited to those two types. The former material is used for blue laser didoes at around $450\,\mathrm{nm}$, but also cover up to green wavelengths, and the latter for near-infrared laser diodes at $780\,\mathrm{nm}$, both wavelengths used for experiments in this group.

The design requirements are split into four parts, that need to be discussed. The ambient environment, the diode voltage and current requirements, the modulation bandwith and finally the noise specifications.

1.3.1. Design Goals: Ambient Environment

The laser driver is to be used in a clean laboratory environment. Typical lab temperatures are in the range of $20\,°\text{C}$–$25\,°\text{C}$ and ware mostly met in our labs, before improvements were implemented as part of this work. Humidity is only controlled with dehumidifiers and therefore in the range of $15\,\%\text{rH}$–$60\,\%\text{rH}$. The air is typically filtered using H14 HEPA filters. Figure 1.1 shows a typical $1\,\text{d}$ span of the lab temperature as it was found at the start of this project.
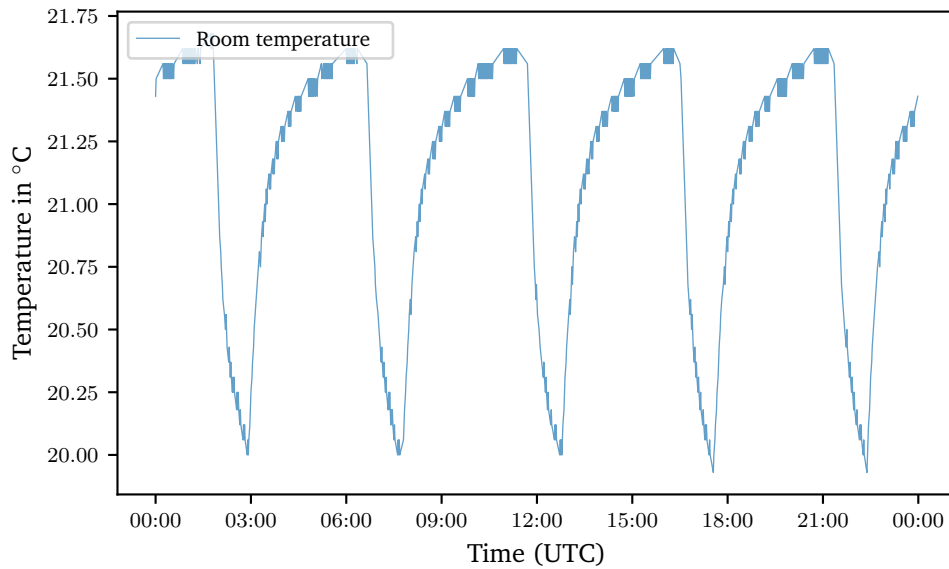

Figure 1.1.: Temperature in Lab 011 on 2016-11-26.

As it can be seen there are strong oszillations of the temperature as a result of the on–off air conditioning temperature controller. The commercial controller used back then was realized using an IMI Heimeier EMO T Valve [34], which is a 2 step valve. Altough this solution was later replaced by a custom design described in section **??**, these type controllers are found in many other labs and temperature swings of $2\,\text{K}$ must therefore be expected.

These expected environmental parameters can now be used to estimate the design requirements for the laser driver. The more demanding laser system is the $450\,\text{nm}$ system [13] required for the spectroscopy of highly charged ions [76] at GSI. This system was found to be more susceptible to changes of the drive current since the wavelength selective filter element was far broader in comparison to a $780\,\text{nm}$ system [77]. This laser is stable over regions of tens of μA and requires a maximum drive current of $145\,\text{mA}$ [88].

From these thoughts, the requirements for the driver can be inferred. It should be able to supply at least $150\,\text{mA}$ and stay well within $10\,\mu\text{A}$ over the whole environmental range. For a worst-case assumption a tolerance of 3σ (99.7 %) must be met [110].

The environmental parameters that mostly affect current sources are temperature and humidity. Air pressure is typically a matter of concern for high voltage systems [50] and secondary to consider for this design as it is a low voltage system ($\leq 48\,\text{V}$). Air pressure effects are also the most expensive to test for, as a pressure chamber is required. While humidity does affect electronics due to corrosion and also indirectly because the epoxy resin used in the FR-4 PCBs and component moulding is hygroscopic and the absorbed humidty leads to swelling and mechanical stress. This effect is very slow at ambient temperature and can easily take days to

show [52]. This parameter is therefore handled via the long-term stability and not specified separately.

Given environmental conditions, the relative coefficients can be calculated. This estimation assumes a minimum setpoint resolution of 2 steps within the mode-hop-free region of the laser and calculates the $99.7\,\%$ confidence interval. The steps are given in table 1.1:

Property	Value	Result
Stable range	$10\,\mu A$	$10\,\mu A$
2 steps of resolution	$\div 2$	$5\,\mu A$
1σ	$\div 2$	$2.5\,\mu A$
Maximum output	$150\,mA$	$17\,\mu A/A$
Temperature range	$5\,K$	$3\,\mu A/(A\,K)$
Worst case (3σ)	$\div 3$	$1\,\mu A/(A\,K)$

Table 1.1.: Estimated requirement for the temperature coefficient of the laser driver.

While the requirements look moderate at first sight, doing a quick estimation leeds to a temperature coeffiction of $1\,\mu A/(A\,K)$, preferably better than that when using a higher output driver – a rather formidable specification for a current source.

Regarding the long-term stability, a $30\,d$ figure can be estimated. One may be inclined to call for a drift, that is smaller than the stable range, but this would be short sighted, as there are other factors to consider. The laser including the external resonator has its own figure of merit regarding the spectral drift rate. Talvitie et al. [113] reported a drift of $2.9\,MHz/h$, which was attributed either to the external resonator itself, the piezo or the collimation lens. It is most likely, that this drift was caused by mechanical changes of the external resonator as it defines the output mode of the laser. The mechanical drift limits the required stability of the current source considerably, as a typical frequency change of the internal resontor with the current of $3\,MHz/\mu A$ [126] can be assumed. The (linear) ageing drift of the external resonator over $30\,d$ is equivalent to a $720\,\mu A$ drift over the same period. For the electronics, the drift is assumed to follow an Arrhenius-like equation resulting from stress induced during manufacturing. This may eventually change to a slow linear drift after several months of relaxation. The coefficient may either be a positive or negative and leads to

Property	Value	Result
Ageing drift limit	$720\,\mu A$	$720\,\mu A$
1σ	$\div 2$	$360\,\mu A$
Maximum output	$150\,mA$	$2400\,\mu A/A$
Worst case (3σ)	$\div 3$	$800\,\mu A/A$

Table 1.2.: Estimated requirement for the long-term stability of the laser driver.

From these number, it straightforward to see, that the long-term stability of a laser driver is less important than the short-term temperature coefficient since the limiting factor is the mechanical construction of the laser. This necessitates an atomic reference for long-term stability and to compensate for accoustic resonances of the external resonator. Regarding the choice of suitable devices, the tight specification of the temperature coefficient most likely leads to a choice of components, that will pass these long-term criteria as well, alleviating the burden of proof a bit as long-term drift specifications are hard to come by since they literally

need solid time to come by and cannot be extrapolated from high temperature burn-in tests [127].

This leads to the following design specifications regarding the stabilty of the current driver:

> **Current controller stability specifications**
>
> - Temperature range $20\,°\text{C}$–$25\,°\text{C}$
>
> - **Temperature coefficient $\leq 1\,\mathbf{\mu A/(A\,K)}$**
>
> - Humidty (non-condensing) $\leq 75\,\%\text{rH}$
>
> - Humidty coefficient not specified, but included in the long-term drift
>
> - Maximium altitude not specified
>
> - Long-term drift over $30\,\text{d}$ $\leq 800\,\mu\text{A/A}$

1.3.2. Design Goals: Current Source

The change in output current caused by load impedance should be an order of magnitude less than the drift specification to ensure a negligible effect compared to the drift over time. The load resistance presented by the laser diodes most commonly used in our experiments ranges from $50\,\Omega$ [88] to $30\,\Omega$ [5] and $10\,\Omega$–$15\,\Omega$ for 780 nm laser diode [42, 60]. It can therefore be estimated as

$$\frac{R_{load}}{R_{out}} = \frac{I_{set}}{I_{out}} - 1 \leq 6.7\,\mu\text{A/A}$$

$$R_{out} \geq \frac{50\,\Omega}{6.7\,\mu\text{A/A}} = 7.5\,\text{M}\Omega \tag{1.3}$$

An output impedance of more than $7.5\,\text{M}\Omega$ for slowly changing loads is a fairly moderate requirement and can typically be realised using a high precision control loop with an operational amplifier, but another important aspect is the compliance voltage of the current source.

The compliance voltage is the maximum voltage the current source can apply to the load and is another non-ideal component of a real current source. The required voltage strongly depends on the type of laser diode used. The near-inrared laser diodes discussed above have an operating voltage of $1.5\,\text{V}$–$3\,\text{V}$, while the Osram PL 450B blue laser diode is specified for $5.5\,\text{V}$–$7\,\text{V}$. The $7\,\text{V}$ required by the Osram laser diode is fairly high for a Fabry–Perot laser diode and has proven difficult in the past [13] as most laser current driver available are designed for the much forward voltage of the near infrad laser diodes. Even higher voltages of around $12\,\text{V}$–$15\,\text{V}$ are required for quantum cascade lasers, but these are currently neither used nor is their use planned in our experiments.

The maximum output current of the laser driver currently required for laser diodes used in our group is $250\,\text{mA}$ for the Thorlabs L785H1 [60]. Therefore a maximum output current of $300\,\text{mA}$ is considered sufficient.

The current noise of the laser driver can be estimated from the laser linewidth sought as the laser frequency is sensitive to the injection current. At low frequencies, about $-3\,\text{MHz/}\mu\text{A}$ can be attributed to the thermal expansion of the internal resonator of the diode due resistive heating [126]. Above $1\,\text{MHz}$ this effect starts declining and exposes the change of the refractive index due the presence of charge carriers. This second effect is an order of magnitude weaker. Since the frequency sensitivity to current variations of the laser diode drops with higher frequencies the most important range is from DC to $100\,\text{kHz}$.

To estimate the linewidth requirement, it is important to look at the experimental setup. While the spectroscopy of Ar^{13+} at $4\,\text{K}$ is limited to around $150\,\text{MHz}$ as shown on page 3, the quantum computing experiments in our group have more stringent needs. It was shown in [22, 96, 113], that with reasonable expense a passive linewidth of less than $100\,\text{kHz}$ can be achieved. Using the frequency sensitivity to a current modulation of laser diodes $100\,\text{kHz}$ translates to a current noise of $30\,\text{nA}_{\text{rms}}$ from $1\,\text{Hz}$ to $100\,\text{kHz}$. The lower $1\,\text{Hz}$ limit is chosen fairly arbitrary, but the presence of $\frac{1}{f}$-noise inhibits a definition down to DC. There should be negligible amounts noise below $1\,\text{Hz}$ compared to the upper $100\,\text{kHz}$ though.

The final aspect of the current source, that needs to be specified, is the bandwidth of the current steering input. The bandwidth in these terms define a reasonably flat ($\leq 3\,\text{dB}$) response. As it was discussed above, beyond a frequency of $1\,\text{MHz}$, the frequency sensitivity of the laser diode to current modulation drops by an order of magnitude, altering the transfer function and introducing new challenges for control loops. Therefore a bandwidth of $1\,\text{MHz}$ or more is considered sufficient.

Above $1\,\mathrm{MHz}$ it is recommended to either use more dedicated solutions like the direct modulation at the laser head presented in [90] or switch to acousto-optic modulators (AOMs) or electro-optic modulators (EOMs).

This leads to the following requirements regarding the current source of the laser driver:

<div style="border: 2px solid red; border-radius: 8px;">

Current source specifications

- Maximum output current $300\,\mathrm{mA}$, optionally $500\,\mathrm{mA}$

- **Compliance voltage $\geq 8\,\mathrm{V}$**

- Output impedance $\geq 7.5\,\mathrm{M\Omega}$ at low frequencies (close to DC)

- **Current noise $\leq 30\,\mathrm{nA_{rms}}$ from DC to $100\,\mathrm{kHz}$**

- $3\,\mathrm{dB}$-bandwidth of the modulation source $\geq 1\,\mathrm{MHz}$

</div>

1.3.3. Design Goals: User Interface and Form Factor

The user interface must be remote controllable, as the Penning trap and the laser system is spacially separated with the laser system being located in a special laser lab for environmental and safety reasons. The spatial separation is about $30\,\text{m}$. Ideally this remote interface is computer controlled to give full access to all features of the laser system. USB or Ethernet is preferred as this does not require extra hardware in the lab.

Regarding the application programming interface (API) support for both Labview and Python, with a strong tendency to Pyhton is favoured. The reason is, that most of the group has switched from Labview to labscript suite [111] on Python to run the experiments.

The local interface must be accessible without a computer to allow simple adjustment of the parameters while on the bench.

The form factor should allow integration into standard 19-inch racks to allow simple transportation from the experiment site at GSI to the university for testing and calibration.

> **Current source user interface and form factor**
>
> - Remote computer interface required
>
> - Python and optionally Labview drivers
>
> - Rack mountable form factor preferred

1.4. LabKraken

1.4.1. Design Goals

LabKraken is a designed to be a asynchronous, resilient data acquisition suite, that scales to thousands of sensors and accross different networks.

1.4.2. Hardware

1.4.3. Software Architecture

LabKraken needs to scale to thousands of sensors, which need to be served concurrently. This problem is commonly refered to as the C10K problem as dubbed by Dan Kegel back in 1999 [58] and refers to serving $10\,000$ concurrent connections via network sockets. While today millions of concurrent connections can be handled by servers, handling $10\,000$ can still be challenging, especially, if the data sources are heterogeneous as is typical for sensor networks of different sensors from different manufacturers.

In order to meet the design goals, an asynchronous architecture was chosen and several different architectures were implemented over time. All in all four complete rewrites of the software were made to arrive at the architecture presented here. The reason for the rewrites is mostly historic and can be explained by the history of the programming language Python, which was used to write the code. The first first version was written for Python 2.6 and exclusively supported sensors made Tinkerforge. In 2015, Python 3.5 was released, which supported a new syntax for asynchronous coroutines. The software was rewritten from scratch to support this new syntax, because it made the code a lot more verbose and easier to follow. With the release of Python 3.7 in 2018 asynchronous generator expressions where mature enough to be used in productions and the programm was again rewritten to use the new syntax. In 2021 a new approach was taken and the programm was once more rewritten with a functional programming style. I will discuss each approach in the next sections to highlight the improvements, that were made over time. Each of these sections discusses the same programm, but written in different styles to show the differences.

Threaded Design

The first version of LabKraken used a threaded design approach, because the original libraries of the Tinkerforge sensors are built around threads. The following simplified example shows some code to connect to a temperature sensor over the network and read its data.

```python
ipcon = IPConnection()
devices = dict()

# Callback function for temperature callback
def cb_temperature(temperature):
    print("Temperature: " + str(temperature/100.0) + " degC")

def cb_connected(connect_reason)::
    ipcon.enumerate()

def cb_disconnected(disconnect_reason)::
    log_reason(disconnect_reason)

def cb_enumerate(uid, *_args):
```

```python
    if uid == OUR_KNOWN_DEVICE:
        dev = BrickletTemperatureV2(uid, ipcon)
        # Register temperature callback to function cb_temperature
        dev.register_callback(dev.CALLBACK_TEMPERATURE, cb_temperature)
        dev.set_temperature_callback_configuration(1000, False, "x", 0, 0)
        devives[uid] = dev

if __name__ == "__main__":
    ipcon.connect(HOST, PORT)  # blocking call
    # Register Enumerate Callback
    ipcon.register_callback(IPConnection.CALLBACK_ENUMERATE, cb_enumerate)
    ipcon.register_callback(IPConnection.CALLBACK_CALLBACK_CONNECTED,
                                        cb_connected)
```

Device Identifiers

Every sensor network needs device identifiers. Preferably those identifiers should be unique. Typically a device has some kind of internal indetifier. Here are a few examples of the sensors used in our network:

Device Type	Identifiers	Example
GPIB (SCPI)	*IDN?* returns $manufacturer,$name,$serial,$revision	
Tinkerforge	Each sensor has a base58 encoded integer device id	QE9 (163684)
Labnode	Universal Unique Identifier (UUID)	cc2f2159-e2fb-4ed9-8021-7771890b37ad

As it can be seen above, these identifiers do not guarantee to uniquely identify a device within a network. The Tinkerforge id is the weakest, as it is a 32 bit integer (4.294.967.295 options), which might easily collide with another id from a different manufacturer. The tinkerforge id is presented as a base58 encoded string. An encoder/decoder example can be found in the TinkerforgeAsync library [14].

The id string returned by a SCPI device is slightly better, but again does not guarantee uniqueness. As it is shown in the example the same device might return a different id defpending on its settings. This typically done by manufacturers for compatibility reasons.

The only reasonably unique id is the universal unique identifier (UUID) or globally unique identifier (GUID), as dubbed by Microsoft, used in the Labnodes. Their id can be used for networks with participant numbers going into the millions.

Calculating the probability of a collision between two random UUIDs is called the birthday problem [122] in probability theory. A randomly generated version 4 UUID of variant 1 as defined in RFC 4122 [66] has 122 bit of entropy, that is out of 128 bit, 4 bit are reserved for the UUID version and 2 bit for the variant. This gives the probability of at least one collision in n devices out of $M = 2^{122}$ possibilities:

$$p(n) = 1 - 1 \cdot \left(1 - \frac{1}{M}\right) \cdot \left(1 - \frac{2}{M}\right) \ldots \left(1 - \frac{n-1}{M}\right)$$

$$= 1 - \prod_{k=1}^{n-1} \left(1 - \frac{k}{M}\right) \tag{1.4}$$

Using the Taylor series $e^x = 1 + x \ldots$, assuming $n \ll M$ and approximating we can simplify this to:

$$
\begin{aligned}
p(n) &\approx 1 - \left(e^{\frac{-1}{M}} \cdot e^{\frac{-2}{M}} \ldots e^{\frac{-(n-1)}{M}} \right) \\
&\approx 1 - \left(e^{\frac{-n(n-1)/2}{M}} \right) \\
&\approx 1 - \left(1 - \frac{n^2}{2M} \right) = \frac{n^2}{2M}
\end{aligned}
\tag{1.5}
$$

For one million devices, this gives a probability of about 2×10^{-25}, which is negligible.

In the Kraken implementation, all devices, except for the Labnodes, will be mapped to UUIDs using the underlying configuration database. It is up to the user to ensure the uniqueness of the non-UUID ids reported by the devices to ensure proper mapping.

Limitations

There is one inherent limitation to the ethernet bus for instrumentation. The ethernet bus is inherently asynchronous and multiple controllers can talk to the device at the same time. Not only that, but different processes within the same controller can talk to the same device. This makes deterministic statements about the device state challenging.

While it is impossible to rule out the possibility of multiple controllers on a network, care was taken to synchronize the workers within Kraken.

1.4.4. Databases

Cardinality

- TimescaleDB vs Influx

- Example Sensors vs. Experiment

1.5. Short Introduction to Control Theory

This section will give a very brief introduction into some basic concepts of control theory. Many systems require control over one or more process variables. For example, temperature control of a room or a device, or creating a programmable current from a voltage. All of this requires control over a process and is established trough feedback, which allows a controller to sense the state of the system.

The focus of this section is narrowed down to the concept of feedback and control with regard of developing and understanding PID controllers for temperature control. In the following sections, first, a model for the system and its controller will be developed, then, using the model, tuning of the control parameters using different tuning algorithms will be discussed.

1.5.1. Introduction to the Transfer Function and the Laplace Domain

There are two types of systems: open- and closed-loop systems. A system is called open loop, if the output of a system does not feed back to its input as in figure 1.2a. On the other hand, if the output influces the input of the system via feedback, it is called a closed-loop system, as shown in figure 1.2b. Although feedback can be treated in static systems, it is more useful to treat it in dynamic systems, either in the time-domain or the frequency-domain. $G(s)$ is called the transfer function of the system, while $U(s)$ is the input, $Y(s)$ is the output, β is the feedback parameter or feedback fraction. In this section, upper case letters denote functions in the Laplace domain, while lower case letters are referring to functions in the time domain.

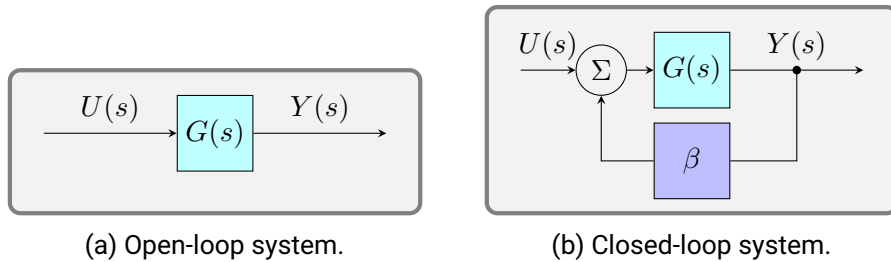(a) Open-loop system.　　　(b) Closed-loop system.

Figure 1.2.: Block diagram of a closed- and an open-loop system.

It is convenient to express the transfer function as its Laplace transform for a number of reasons shown below. The unilateral Laplace transform is defided as:

$$\mathscr{L}\left(f(t)\right) = F(s) = \int_0^\infty f(t)e^{-st}\,dt. \tag{1.6}$$

with $f : \mathbb{R}^+ \to \mathbb{R}$, that is integrable and grows no faster than $e^{s_0 t}$ for $s_0 \in \mathbb{R}$. The latter attribute is important for deriving the rules of differentiation and integration.

To understand the benefits of using the Laplace representation of the transfer function, a few useful properties should be discussed. First of all, the Laplace transform is linear:

$$\begin{aligned}
\mathscr{L}\left(a \cdot f(t) + b \cdot g(t)\right) &= \int_0^\infty (a \cdot f(t) + b \cdot g(t))e^{-st}\,dt \\
&= a\int_0^\infty f(t)e^{-st}\,dt + b\int_0^\infty g(t)e^{-st}\,dt \\
&= a\mathscr{L}\left(f(t)\right) + b\mathscr{L}\left(g(t)\right)
\end{aligned} \tag{1.7}$$

Another interesting property is the derivative and integral of a function f:

$$\mathscr{L}\left(\frac{df}{dt}\right) = \int_0^\infty \underbrace{f'(t)}_{v'(t)}\underbrace{e^{-st}}_{u(t)}\,dt$$

$$= \left[e^{-st}f(t)\right]_0^\infty - \int_0^\infty (-s)f'(t)\,dt$$

$$= -f(0) + s\int_0^\infty f'(t)\,dt$$

$$= sF(s) - f(0) \tag{1.8}$$

$$\mathscr{L}\left(\int_0^t f(\tau)\,d\tau\right) = \int_0^\infty \left(\int_0^t f(\tau)\,d\tau e^{-st}\right)dt$$

$$= \int_0^\infty \underbrace{e^{-st}}_{v'(t)}\underbrace{\int_0^t f(t)\,d\tau}_{u(t)}\,dt$$

$$= \left[\frac{-1}{s}e^{-st}\int_0^t f(t)\,d\tau\right]_0^\infty - \int_0^\infty \frac{-1}{s}e^{-s\tau}f(\tau)\,d\tau$$

$$= 0 + \frac{1}{s}\int_0^\infty e^{-s\tau}f(\tau)\,d\tau$$

$$= \frac{1}{s}F(s) \tag{1.9}$$

If the initial state $f(0)$ can be chosen to be 0, the differentiation becomes a simple multiplication by s, while the integration becomes a division by s. Finally, the most important aspect is, that it is possible to give a simple relation between the input $u(t)$ and the ouput $y(t)$ of a system. The relationship between input and the ouput of a system as shown in figure 1.2a is given by the convolution, see e.g. [10]. Assuming the system has an initial state of 0 for $t < 0$, hence $u(t < 0) = 0$ and $g(t < 0) = 0$, one can calculate:

$$y(t) = (u * g)(t) = \int_0^\infty u(\tau)g(t-\tau)\,d\tau \tag{1.10}$$

Applying the Laplace transform, greatly simplifies this:

$$Y(s) = \int_0^\infty e^{-st}y(t)\,dt$$

$$\stackrel{1.10}{=} \int_0^\infty \underbrace{e^{-st}}_{e^{-s(t-\tau)}e^{-s\tau}}\int_0^\infty u(\tau)g(t-\tau)\,d\tau\,dt$$

$$= \int_0^\infty \int_0^t e^{-s(t-\tau)}e^{-s\tau}g(t-\tau)u(\tau)\,d\tau\,dt$$

$$= \int_0^\infty e^{-s\tau}u(\tau)\,d\tau \int_0^\infty e^{-st}g(t)\,dt$$

$$= U(s) \cdot G(s) \tag{1.11}$$

This formula is a lot simpler than the convolution of $u(t)$ and $g(t)$, therefore the use of the Laplace transform has become very popular in control theory.

Having derived some the most useful properties, it is interesting to look at a few function, that are heavily used in control theory, like a function delayed by the time interval θ. To demonstrate its properties, let $f(t - \theta)$ be

$$g(t) := \begin{cases} f(t - \theta), & t \geq \theta \\ 0, & t < \theta \end{cases}. \tag{1.12}$$

The reason for this definition is, that the system must be causal. This means, it is impossible to get data from the future $(t < \theta)$. To satisfy this requirement, any constant other than 0 may be chosen as well, as is done in later in section 1.5.4, when determining tuning parameters and fitting experimental data to a model. An example of such a time delayed function $g(t)$ is shown in figure 1.3b.

(a) Original signal $f(t)$.
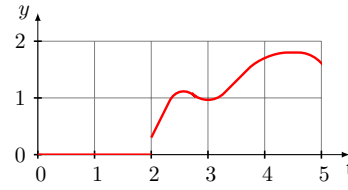
(b) Delayed signal $f(t - 2)$.

The Laplace transform of a delayed signal $g(t)$ can be calculated as follows:

$$\mathscr{L}(g(t)) = \int_0^\infty g(t)e^{-st}\, dt$$

$$\overset{1.12}{=} \int_\theta^\infty f(t - \theta)e^{-st}\, dt$$

$$\overset{\tau := t - \theta}{=} \int_0^\infty f(\tau)e^{-s(\tau + \theta)}\, d\tau$$

$$= e^{-s\theta} \int_0^\infty f(\tau)e^{-s\tau}\, d\tau$$

$$= e^{-s\theta} F(s) \tag{1.13}$$

To satisfy the causaulity requirement in the time domain, the Heaviside function $H(t)$ can be used to give a more concise representation of $g(t)$:

$$\mathscr{L}(f(t - \theta)H(t - \theta)) = e^{-s\theta}F(s) \tag{1.14}$$

Lastly, the Laplace transform of e^{at} is given, which is commonly used in differential equations:

$$\mathscr{L}(e^{at}) = \int_0^\infty e^{(a-s)t}\, dt = \frac{1}{a - s}\left[e^{(a-s)t}\right]_0^\infty = \frac{1}{s - a} \tag{1.15}$$

Using these tools, it is possible calculate the transfer function of a closed-loop temperature controller. This is done in the next section.

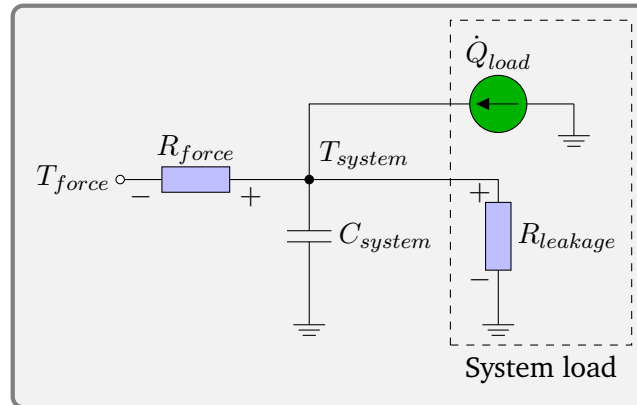## 1.5.2. A Model for Temperature Control

Figure 1.4.: Simple temperature model of a generic system.

In order to describe a closed-loop system using a transfer function $G(s)$, one has to first create a model for the process and the controller involved. This section will derive the simple, but very useful first-order model with dead-time. This model can be derived from the idea, that the system at temperature T_{system} has a thermal capacitance C_{system}, an influx of heat $\dot{Q}_{load}$ from a thermal load and a controller removing heat from the system through a heat exchanger with a resistance of R_{force}. Additionally, there is some leakage through the walls of the system to the ambient environment via $R_{leakage}$. This analogy of thermodynamics with electrodynamics allows to create the model shown in figure 1.4. Since this model is to be used for a temperature controller, more simplifications can be made and a so-called small-signal model can be developed as opposed to the large signal model shown above. The small-signal model is an approximation around a working point, that is valid for small excursions around it, similar to a Taylor approximation. The small signal model can be used calculate the system response to small changes of the controller output in order to estimate the controll parameters.

Using the small signal approach, the system response can be split into a constant and a dynamic part – the 0th and 1st order of the Taylor approximation. In order to simplify the system shown in figure 1.4 an assumption can made, that the system load $\dot{Q}_{load}$ and the flux through $R_{leakage}$ is *reasonably stable*. *Reasonably stable* means that it can be treated as small deviations and additionally any changes are within the bandwidth of the controller and well suppressed. This allows to treat them as (almost) constant effects, which result in an offset applied to the output of the controller. This leaves only the room with its heat capacity and the heat exchanger as the dominant factors in the dynamic model shown in figure 1.5. Here T_{force} and T_{system} were replaced by T_{in} and T_{out} for better readability:
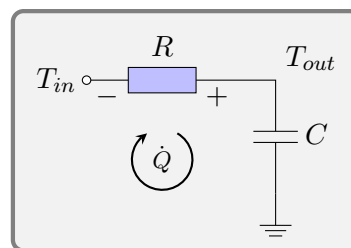

Figure 1.5.: Simplifications of the temperature model of a room lead to this first order model.

This is the classic RC circuit and exploiting the analogy of thermodynamics and electrodynamics again, using Kirchhoff's second law, following the arrow in figure 1.5, one finds:

$$\sum T_i = 0$$

$$T_{in}(t) - \dot{Q}(t)R - \frac{1}{C} \int \dot{Q}(t) \, dt = 0 \tag{1.16}$$

Taking the Laplace transform, applying equation 1.9, solving for $\dot{Q}(s)$ and using $T_{out} = \frac{1}{sC}\dot{Q}(s)$ to replace $\dot{Q}$, equation 1.16 can be written as:

$$T_{in}(s) - \dot{Q}(s)R - \frac{1}{sC}\dot{Q}(s) = 0$$

$$\dot{Q}(s) = \frac{T_{in}(s)}{R - \frac{1}{sC}} = \frac{T_{out}}{\frac{1}{sC}}$$

This allows to calculate the transfer function of the process P using:

$$P(s) = \frac{T_{out}}{T_{in}} = \frac{\frac{1}{sC}}{R - \frac{1}{sC}}$$

$$= \frac{1}{sRC + 1}$$

$$= \frac{1}{1 + s\tau} = \frac{K}{1 + s\tau} \tag{1.17}$$

with the system gain K and the time constant τ. In case of the RC circuit, the gain is 1, but other systems may have a gain factor of $K \neq 1$, so it is included here for the sake of generality.

Equation 1.17 is called the transfer function of a first-order model, because its origin is a differential equation of first order. This model describes homogeneous systems, like a room very well, as can be seen in section **??**, but in order to derive the transfer function including the controller and the sensor some more work is required derive the sensor transfer function.

Expanding on figure 1.2a and equation 1.10 the open-loop transfer function of the process and its sensor becomes:

$$G(s) = P(s) \cdot S(s) \tag{1.18}$$

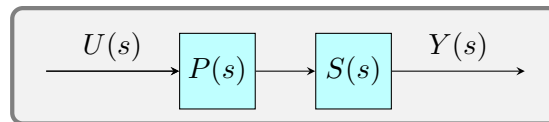and the block diagram changes to

Figure 1.6.: Open-loop system with sensor.

The transfer function of the sensor, given an ideal linear transducer, can be modeled as a delay line with delay θ and $f(t - \theta) = H(t - \theta)$. A gain of 1 is assumed here, because any system gain can already be included in the parameter K. Using equation 1.13, $S(s)$ can be written as

$$S(s) = e^{-\theta s}. \tag{1.19}$$

The full system model including the time delay now is:

$$G(s) = \frac{K}{1 + s\tau} e^{-\theta s} \tag{1.20}$$

This is called a first-order plus dead-time model (FOPDT) or first-order plus time-delay model (FOPTD). To fit experimental data to this model it is more convenient to transform the transfer function 1.20 into the time domain. To have a meaningful result, an input $U(s)$ is required, because $G(s)$ is only a transformation. In principal any function can do, but a step function is typically used, for example by Ziegler et al. [129] and many others [10, 79, 87, 107, 108, 109, 121]. It is both simple to calculate and to apply to a real system in form of a controller output change. Using equations 1.13 and 1.15, the Heaviside $H(t)$ step function transforms as

$$\mathscr{L}\left(u(t)\right) = U(s) = \mathscr{L}\left(\Delta u H(t)\right) = \frac{\Delta u}{s} \tag{1.21}$$

with the step size Δu. The output $Y(s)$ can then be calculated analytically.

$$
\begin{aligned}
Y(s) &= U(s) \cdot G(s) \\
&= \frac{\Delta u}{s} \frac{K}{1 + s\tau} e^{-\theta s} \\
&= K\Delta u \frac{1}{s(1 + s\tau)} e^{-\theta s} \\
&= K\Delta u \left(\frac{1}{s} - \frac{\tau}{s\tau + 1}\right) e^{-\theta s} \\
&= K\Delta u \left(\frac{1}{s} - \frac{1}{s + \frac{1}{\tau}}\right) e^{-\theta s}
\end{aligned}
\tag{1.22}
$$

To derive $y(t)$, the inverse Laplace transform of $Y(s)$ is required. Unfortunately, this is not as simple as the Laplace transform. Fortunately though, the required equations were already derived in equations 1.9 and 1.15. Now, making sure causaulity is guaranteed as shown in 1.14, the simple first order model can be transformed back into the time domain.

$$
\begin{aligned}
y(t) &= \mathscr{L}^{-1}\left(Y(s)\right) \\
&= K\Delta u \mathscr{L}^{-1}\left(\frac{1}{s} e^{-\theta s}\right) - K\mathscr{L}^{-1}\left(\frac{1}{s + \frac{1}{\tau}} e^{-\theta s}\right) \\
&\stackrel{1.15}{=} K\Delta u \cdot 1 \cdot H(t - \theta) - \left(e^{-\frac{t-\theta}{\tau}}\right) H(t - \theta) \\
&= K\Delta u \left(1 - e^{-\frac{t-\theta}{\tau}}\right) H(t - \theta)
\end{aligned}
\tag{1.23}
$$

The time domain solution of the FOPDT model can now be used extract the parameters τ, θ and K from a real physical system.

The procedure can be summarized from the above as follows. The controller must be set to a constant output and the room must be given time reach equilibrium. Once the temperature has settled, an output step of Δu is applied. The system will respond after a time delay and then follow an exponential function. A simulation of the step response applied to a first-order model with time delay is shown in figure 1.7. The gain is $K = 1$. The solid black line shows the response of the transfer function, including the system and the sensor. The dashed lines show the individual components, the Heaviside function governing the delay and the exponential

term of the system. The controller output step $\Delta u = 1$ is applied at $t = 0$ and not shown explicitely. From figure 1.7, it can be clearly seen, that the sensor does not register a change until the time delay θ has passed and the Heaviside function changes from 0 to 1. Then the system responds with an exponential decay towards 1.
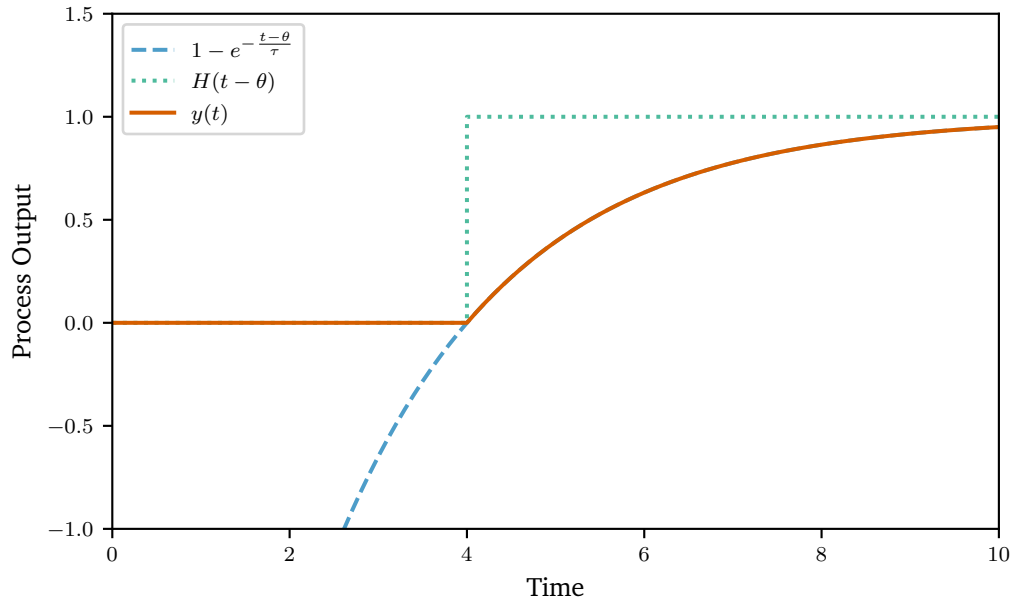

Figure 1.7.: Time domain plot of a first-order plus dead time model showing individual components of the model and the composite function $y(t)$. Model parameters used: $K = \Delta u = 1$, $\tau = 2$, $\theta = 4$.

So far, only open-loop systems were discussed. Using the FOPDT model, the system parameteres can now be extracted from a real system using a fit to the time domain reaction of such a system to a step input. Having extracted the system parameters, the next thing to do, is to design a controller around the system and close the loop to realize a controlled system. This is shown in the next section.

1.5.3. PID Controller Basics

While there are many different types of controllers, like the bang–bang controller utilized in the original lab temperature controller, which turns on at a certain threshold and turns off at another threshold and produced in the saw-tooth shaped room temperature curve shown in figure 1.1, a continuous control system is desired to keep fluctions to a minimum. The most commonly used controller type for non-integrating systems is the proportional–integral–derivative (PID) controller [30]. A non-integrating system is a system without memory, that does not depend on previous inputs. Given the same input, a non-integrating will always return to the same steady state. The advantage of applying a PID controller is, that the controller does not need any special knowledge about the system model. A universal PID is simple to implement and can be tuned to control a wide range of systems. While there are many different variations of the PID algorithm [98], this section only introduces the basic, parallel, PID controller commonly used in digital implementation and deals with some of the shortcomings in practical applications.

In order to extend the FOPDT system, derived in the previous section 1.5.2, with the PID controller, one must move to a closed-loop system. Adding to figure 1.2b and inserting a new control block into the transfer function yields figure 1.8.
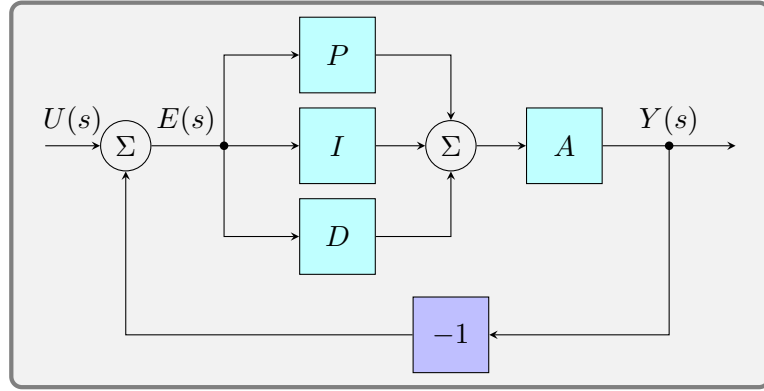

Figure 1.8.: Closed-loop system with a PID controller.

The error signal $E(s)$ used by the PID controller is the difference between the setpoint and the control parameter, in this case the room temperature. The transfer function of the PID controller can be split into three parts. A proportional part, that is proportional to the error representing the present, an integral part, that is proportional to the accumulated error, representing the past, and a derivative part, that is proportional to the change in the error, extrapolating into the future. Analytically, it can be written as

$$c(t) = k_p e(t) + k_i \int_0^t e(\tau)\, d\tau + k_d \frac{\mathrm{d}e(t)}{\mathrm{d}t} \tag{1.24}$$

$$C(s) = k_p + k_i \frac{1}{s} + k_d s \,. \tag{1.25}$$

The following discussion will mostly focus on equation 1.24, because, the time-domain equation is the one, that can be implemented in software. As hinted above, there are a few shortcommings with the classic PID equation, when used in a real system, which requires dynamic changes of the PID parameters, e.g. the setpoint or k_i.

The first problem to be addressed is occuring, when changing the PID parameter k_i. Assuming a settled system without external disturbances, the output is fully determined by the integrator

value, because the error is zero. Now, when k_i is changed, the output immediate changes, due to the change of the integral term. This is unintended. To fix this, the integral term must changed to

$$k_i \int_0^t e(\tau)\, d\tau \Rightarrow \int_0^t k_i(\tau)e(\tau)\, d\tau \,. \tag{1.26}$$

This way, when adjusting k_i, its new value is applied to future error values only and there is no sudden kick.

The next issue is called derivative kick. When looking at the derivative part of equation 1.24, it can be seen that when instantly changing the setpoint, as in a step function, $\frac{de(t)}{dt} \to \infty$. This behaviour is not intended and to fix this, the derivative part can be modified as follows.

$$\begin{aligned}
\frac{de(t)}{dt} &= \frac{d\,(u(t) - y(t))}{dt} \\
&= \underbrace{\frac{du(t)}{dt}}_{\to \infty} - \frac{dy(t)}{dt} \\
&= -\frac{dy(t)}{dt}
\end{aligned} \tag{1.27}$$

The new derivative term is equal to the unmodified one, except in the case of setpoint changes. Removing the setpoint from the equation, the controller behaves as intended. This solution is sometimes called *derivative on measurement* as opposed to *derivative on error*.

The derivative term is again the cause of another problem, the final problem to be discussed. Assuming a noisy input and by chance there is a very short input spike due to noise. The differential in the derivative term will again be sent to very high values, pushing the output away from the correct value, forcing the controller to rebalance.
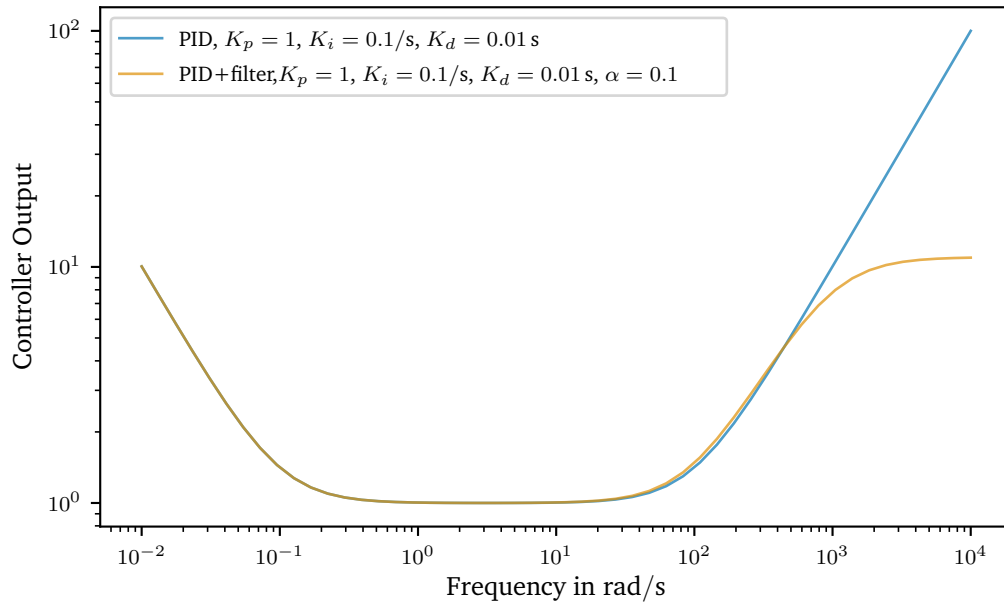

Figure 1.9.: Magnitude plot over frequency of the PID controller transfer function. Both the ideal PID controller and the PID controller with a filtered derivative are shown.

To further discuss the problem and the solution it is best to visit the frequency domain and visualize the transfer function of the PID controller as shown in figure 1.9. The ideal PID controller without filtering of the derivative can be seen to display a very strong response to low frequecy inputs. This is due to the integral action, which removes any (constant) offset. It needs to have infinite gain at DC to push the offset to zero. In reality this is limited by the input noise. Then follows a plateau, with a magnitude of k_p for the proportinal term and finally the differential gain starts growing in magnitude and is ever growing with rising frequency, just as expected.

With some knowledge about the process or the sensor it is possible to define an upper frequecy, above which, inputs become unrealistic and must therefore be unwanted noise. By filtering the derivative term with a first order filter causes it to roll off and its gain becomes constant as shown in figure 1.9. By adding the filter the PID controller tranfer function changes to

$$C(s) = k_p + k_i \frac{1}{s} + \frac{k_d s}{1 + s\alpha k_d} \ . \tag{1.28}$$

Typically α is in the range of 0.05–0.2 [98, p. 129].

Another alternative is to filter the whole input. Depending on the filter cutoff, there is not much difference to equation 1.28, because the filter will not touch the proportional and integral part of the transfer function if both are well within its passband.

From figure 1.9 it also be seen, why in some publications, the gain k_p is applied to all three terms and k_i and k_d are replaced with T_i and T_d to accomodate for that.

$$C(s) = k_p \left(1 \frac{1}{T_i s} + \frac{T_d s}{1 + s\alpha T_d} \right) \tag{1.29}$$

Using this form allows to shift the curve up and down keeping its shape instead of just the k_p part, thus changing the corner frequencies. The alternative form is only given here for the sake of completeness. The author uses the ideal form shown in equation 1.25 with the parameters k_p, k_i and k_d wherever possible.

This concludes the discussion of the PID controller and should introduce the basic terms. It now begs the question how the controller interacts with the system and how to derive the optimal PID parameters from a given system or model. The next section discusses controller tuning rules and their effect on the system performance.

1.5.4. PID Tuning Rules

While there are many PID tuning rules to be found in literature, their application depends on the underlying system and the desired system response. This section will discuss several of the proposed solutions and compare them to the authors use case. It aims to give a simple method to determine decent PI(D) parameters for the applications found in the lab. Among the methods discussed are the most classic set of tuning rules developed by Ziegler et al. [129], an improved version by Skogestad [109], that promises better performance for non-integrating systems. These rules all include simple instruction to extract the neccesary parameters using pen and paper. Using a computer and fitting algorithms, the bar for *simple* has been raised considerably, so more complex approaches can be undertaken, that extract more parameters from the system. Using these additional parameter more precise control is promised by Åström et al. [10, 11] with a method called AMIGO. Finally, it is possible to shape the control loop to result in a desired transfer function. This technique is mostly used in motor control [11, 98] and also requires the model parameters.

All of these rules will be compared against a demo model of a room to explain the details. It is a first order model with delay, that was derived in equation 1.20. The discussion is limited to the FOPDT model only, because the systems treated in this work could be modelled very well using this equation. Higher order models are discussed in more details for example in [11, 98, 109] should the reader encounter such system and feel the need to extract the model parameters.

$$G(s) = \frac{Ke^{-\theta s}}{1 + s\tau} \tag{1.30}$$

The following parameters were extracted from Lab 011, using the techniques shown in section 1.5.2 using equation 1.23. The details are discussed in section **??**. The system gain K was scaled to the full scale output (4095 bit) of the controller, hence the strange unit $\mathrm{K\,bit\,bit^{-1}}$.

Gain K	Lag τ	Delay θ
$13.07\,\mathrm{K\,bit\,bit^{-1}}$	$395\,\mathrm{s}$	$187\,\mathrm{s}$

Before detailing the tuning parameters, the loop shaping method shall be explained first, because it cannot only be used to derive custom rules, but was also used to create the SIMC rules proposed by Skogestad. The aim of this method is to derive a controller, that shapes the model in such a way, that a desired system response to setpoint changes is achieved. A general closed-loop system with a controller C and a system G is shown in figure 1.10. This will be used as a basis to find the required controller for a desired transfer function $\frac{Y(s)}{U(s)}$.
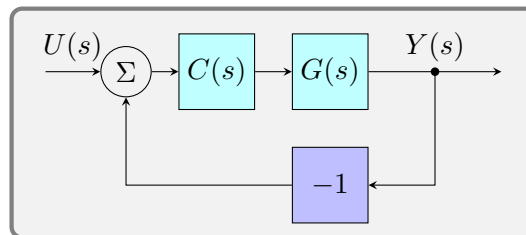

Figure 1.10.: Closed-loop system G with a controller C.

Starting with the transfer function of the controlled system, made up of the controller and the system, most experimenters would, at least in a feaverish dream, prefer a transfer function of the following divine form

$$\frac{Y(s)}{U(s)} = 1 \,,$$

but unfortunately life is more profane and there is no controller, that will always (and with warp speed) force a system to a certain setpoint. One may therefore settle for the second-best choice, a first-order low pass with a slow roll-off and a small delay, which must be added, to ensure causality. One therefore arrives at

$$\frac{Y(s)}{U(s)} = \frac{e^{-\theta s}}{1 + s\tau_c} \,, \tag{1.31}$$

where τ_c is the closed-loop time constant and a measure for the aggressiveness of the controller. A small τ_c results in a more aggressive controller with faster response.

From figure 1.10 the closed-loop transfer function is found to be

$$\frac{Y(s)}{U(s)} = \frac{C(s)G(s)}{C(s)G(s) + 1}$$

$$\Rightarrow C(s) = \frac{1}{G(s)} \frac{1}{\frac{Y(s)}{U(s)} - 1}$$

Using the desired tranfer function 1.31 yields

$$C(s) = \frac{1}{G(s)} \frac{e^{-\theta s}}{s\tau_c + 1 - \underbrace{e^{-\theta s}}_{\approx 1 - \theta s}} \tag{1.32}$$

$$\approx \frac{1}{G(s)} \frac{e^{-\theta s}}{s(\tau_c + \theta)} \,. \tag{1.33}$$

$e^{-\theta s}$ was approximated using a first-order Taylor expansion. Finally, substituting the system equation 1.30 results in

$$C(s) = \frac{1}{K} \frac{s\tau + 1}{(\tau_c + \theta)s}$$

$$= \underbrace{\frac{1}{K} \frac{\tau}{\tau_c + \theta}}_{k_p} + \underbrace{\frac{1}{K} \frac{1}{\tau_c + \theta}}_{k_i} \frac{1}{s} \,. \tag{1.34}$$

This is a PI controller with $k_p = \frac{1}{K} \frac{\tau}{\tau_c + \theta}$ and $k_i = \frac{1}{K} \frac{1}{\tau_c + \theta}$. From these calculations, it can be seen, that a first-order model can be fully treated using a PI controller. Second-order (and higher order) models typically neccesitate a PID or more sophisticated controller for optimal control. The problems discussed in this work mainly focus in temperature control of (mostly) homogeneous objects, so the focus lies on the PI controller for most the remaining section, but the ideas and simulations can similarily be applied to the PID controller as well. Any caveats to be expected when treating a PID instead of a PI controller will be mentioned.

Using the loop shaping technique, it is fairly easy to derive custom rules in case the model parameters can be extracted. As it was said above, one such loop-shaped tuning rule is the SIMC ruleset and the authors give rules for an ample variety of different models and also

Tuning Rule	k_p	T_i	T_d	Source
Z-N PI	$\frac{0.9\tau}{K\theta}$	$\frac{\theta}{0.3}$	–	[129]
Z-N PID	$\frac{1.2\tau}{K\theta}$	2θ	$\frac{\theta}{2}$	[129]
SIMC PI	$\frac{\tau}{K(\tau_c+\theta)}$	$\min\left(\tau, 4(\tau_c+\theta)\right)$	–	[109]
SIMC PID	$\frac{\tau_1}{K(\tau_c+\theta)}$	$\min\left(\tau_1, 4(\tau_c+\theta)\right)$	τ_2	[109]
AMIGO PI	$\frac{0.15}{K} + \left(0.35 - \frac{\tau\theta}{(\tau+\theta)^2}\right)\frac{\tau}{K\theta}$	$0.35\theta + \frac{13\tau^2\theta}{\tau^2+12\tau\theta+7\theta^2}$	–	[11, p. 228]
AMIGO PID	$\frac{1}{K}\left(0.2 + 0.45\frac{\tau}{\theta}\right)$	$\frac{0.4\theta+0.8\tau}{\theta+0.1\tau}\theta$	$\frac{0.5\tau\theta}{0.3\theta+\tau}$	[11, p. 233]

Table 1.3.: PI/PID parameters for different tuning rules. The PI controllers assume a first-order model, the PID rules are required when dealing with a second-order model.

investigate the parameter choice regarding stability, load and setpoint disturbances. Before attempting a custom approach, it is therefore recommended to check [109] for an appropriate set of rules for more complex models in order to save time and effort.

For reasons of brevity, in table 1.3, the PID parameters are given as k_p, T_i and T_d as introduced in equation 1.29. k_i and k_d can be calculated from

$$k_i = \frac{k_p}{T_i}$$

$$k_d = k_p T_d\,.$$

Regarding the SIMC PI/PID algrorithm, Skogestad [109] and [121, ch. 5] suggests using $\tau_c = \theta$ for "*tightest possible subject to maintaining smooth control*". Following this recommendation, the minimum can be calculated from the parameters of this example as $\min\left(\tau, 4(\tau_c+\theta)\right) = \min\left(\tau, 8\theta\right) = \tau$.

Using the rules above, the full system can be simulated now. This was done using Python. The simulation source code can be found in data/simulations/sim_pid_controller.py as part of the online suplemental material [16]. The simulation can be used to model arbitrary PI(D) controller and arbitrary models can be used as well. It allows to compare different settings before applying them to a real system. It also considereably shortens deployment times, because especially for systems with long timescales, it becomes difficult to test several parameter sets on the fly, so using a simulation can reduce this time to a few minutes instead of hours.

The simulation emulates the PID controller developed for the lab temperature controller. By default is has a sampling rate of $1\,\mathrm{Hz}$. The simulation will apply a setpoint change of $1\,\mathrm{K}$ when $10\,\mathrm{s}$ into the simulation. After the simulation it will plot the time domain response of the controlled system. The setpoint change in this scenario is very similar to the load disturbances, that are expected. Typically a noise source is used herer instead, but in contrast to the statistical noise, that could be used to test for disturbance rejection, the situation in the labs are different and cannot be moddeled with stationary noise. While there is some noise coming from the sensor and the lab, the major disturbances are usually caused by experimenters instead of the lab itself. These are events like a device being switched on or off for an extended period of time, longer, than the controller needs to settle. This is equivalent to a setpoint change in terms of the error term in equation 1.24, since there is no difference in the error term between a setpoint and a process variable change. Do note, this is not true for the PID controller, whose derivative term directly works on the measurement (or process variable) as this was explicitly implemented above. For PID controllers, there is therefore a difference between the setpoint change behaviour and system noise rejection. This must be kept in mind and tested accordingly.

Simulating the model above and using the PI parameters derived from table 1.3, gives the plot shown in figure 1.11.
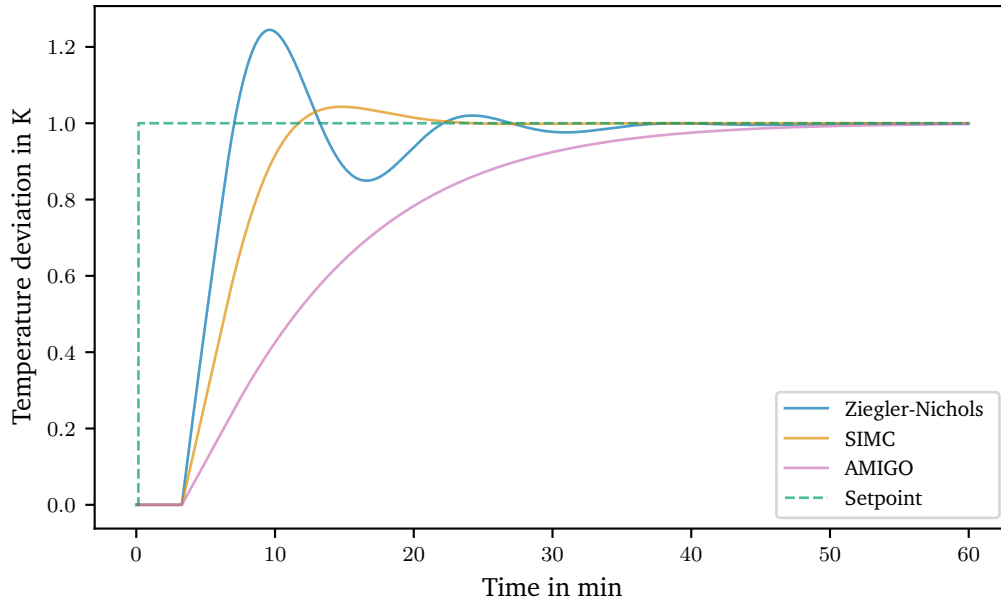

Figure 1.11.: Different PI Controllers tuned with parameter derived using the following methods: Ziegler-Nichols, SIMC and AMIGO. The system model is the FOPTD model for room 011.

As it can be seen in figure 1.11, the Ziegler-Nichols tuning rule produces a very aggressive PI controller, that shows quite a bit ringing, which is undesired for this application. The AMIGO rules are rather conservative, but do not produce any overshoot. The SIMC rules have proven the most useful for this application so far. This experience is in line with the results from Liebmann [69], who tested different PID tuning algorithms for their viability for temperature control in the labs discussed here.

To conclude, several PID tuning rules were presented and using a Python simulation tool it is possible test a set of PID parameters before implementation. Using an example, the different tuning rules were applied to a model for a real lab and the SIMC tuning rules were found to give the best results for this application. The reader should now be able to extract the model parameters from physical system and have the tools to choose an optimal set of tuning parameters for the PID controller. Further reading recommendations are for a broad overview [98], and for more details [11].

1.6. Noise and Allan Deviation

The Allan variance [6] $\sigma_A^2(\tau)$ is a two-sample variance and used as a measure of stability. The Allan deviation $\sigma_A(\tau)$ is the square root of the variance. Originally, the Allan variance was used to quantify the performance of oscillators, namely the frequency stability, but it can be used to evaluate any quantity. In order to define the Allan variance, a few terms need to be defined first. A single measurement value of a time series $y(t)$ can be written as

$$\bar{y}_k(t) = \frac{1}{\tau} \int_{t_k}^{t_k+\tau} y(t)\, dt. \tag{1.35}$$

This is the k-th measurement with a measurement time or integration time τ. The latter term is frequently used for DMMs. t_k is the start of the k-th sampling inverval including the dead time θ

$$t_{k+1} = t_k + T \tag{1.36}$$

with

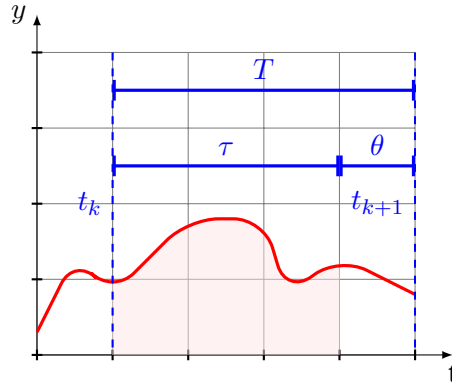$$T := \tau + \theta. \tag{1.37}$$

Figure 1.12.: Measurement interval according to equation 1.35. The shaded region is the signal acquisition period.

Using this, the deviation over N samples is defined as [6, 12]

$$\sigma_y^2(N, T, \tau) = \left\langle \frac{1}{N-1} \left(\sum_{k=0}^{N-1} \bar{y}_k^2(t) - \frac{1}{N} \left(\sum_{k=0}^{N-1} \bar{y}_k(t) \right)^2 \right) \right\rangle \tag{1.38}$$

The $\langle\,\rangle$ denotes the (infinite time) average over all measurands y_k or, simply put, the expected value.

The Allan variance is a special case of this definition with zero dead-time ($\theta = 0$) and only 2 samples:

$$\sigma_A^2(\tau) = \sigma_A^2(N = 2, T = \tau, \tau) \tag{1.39}$$

$$= \left\langle \frac{(\bar{y}_{k+1} - \bar{y}_k)^2}{2} \right\rangle \tag{1.40}$$

It can be shown [12], that 1.41 is indeed more useful than $\sigma_A^2(N \to \infty, T = \tau, \tau)$, because $\sigma_A^2(N = 2, T = \tau, \tau)$ converges for processes, that do not have a convergent $\sigma_A^2(N \to \infty, T = \tau, \tau)$.

In practice, no experiment can take an infinite number of samples, so typically the Allan variance must be estimated using a number of samples m:

$$\sigma_A^2(\tau) \approx \frac{1}{m} \sum_{k=1}^{m} \frac{\left(\bar{y}_{k+1} - \bar{y}_k\right)^2}{2} \tag{1.41}$$

This estimation can lead to artifacts in the results as discussed later. In order to derive the Allan variance from a set of data points, the different values of τ are usually obtained by averaging over a number of samples as there is no dead time (by definition of the Allan variance).

Additionally, the Allan variance is mathematically related to the two-sided power spectral density $S_y(f)$ [12]:

$$\sigma_A^2(\tau) = 2 \int_0^\infty S_y(f) \frac{\sin^4(\pi f \tau)}{(\pi f \tau)^2} \, df \tag{1.42}$$

and therefore all processes, that can be observed in the power spectral density plot can also be seen in the allan deviation. The inverse transform, however, is not always possible as shown by Greenhall [44].

Distinguishing different noise processes using the Allan deviation will be elaborated in the next section.

1.6.1. Identifying Noise in Allan Deviation Plots

It was already mentioned by Allan in [6], that types of noise, whose spectral density follows a power law

$$S(f) = h_\alpha \cdot f^\alpha \tag{1.43}$$

can be easily identified in the Allan deviation plot. The constant h_α is called the power (intensity) coefficient. The most common types of noise encountered in experimental data and their representations can be found in table 1.4, which serves as a summary of this section. Since those types of noise are present in any measurement or electronic device, it warants a further discussion to understand their root causes and ideas to minimize them. While not a type of noise, linear drift can also be easily identified in the Allan deviation plot. It is therefore included in table 1.4 as well.

Amplitude noise type	Power-law coefficient α	Allan variance σ_A^2
White noise	0	$\frac{1}{2} h_0 \tau^{-1}$ [7]
Flicker noise	-1	$2 \ln 2 \, h_{-1} \tau^0$ [7]
Random walk noise	-2	$\frac{3}{2} \pi^2 h_{-2} \tau^1$ [7]
Burst noise	0 and -2	$y_{RMS}^2 \frac{\bar{\tau}^2}{\tau^2} \left(4e^{-\frac{\tau}{\bar{\tau}}} - e^{-\frac{2\tau}{\bar{\tau}}} + 2\frac{\tau}{\bar{\tau}} - 3 \right)$
Drift	$-$	$\frac{1}{2} D^2 \tau^2$ [45]

Table 1.4.: Power law representations using the Allan variance.

In order to arrive at a good understanding of the features seen in an Allan deviation plot, this section will provide the reader with examples of each type of noise and the corresponding

time domain, power spectral density and Allan deviation plot. Since a complete overview is not available in current literature, all required mathematical descriptions and simulation tools will be discussed here. The simulations were done using Python and the source code is linked to in the discussions. The files are found in the online supplemental material found at [16]. Using these scripts all the graphs shown can be recreated and explored further.

White Noise

White noise is probably the most common type of noise found in measurement data. Johnson noise found in resistors, caused by the random fluctuation of the charge carriers, is one example of mostly white noise up to a bandwidth of $100\,\text{MHz}$, from where on quantum corrections are required [38]. Amplifiers also tend to have a white noise spectrum at higher frequencies.

For the latter reason, white noise typically makes up for a considerabe amount of noise in measurements, unless one works at very low frequencies. White noise is a series of uncorrelated random events and therefore characterised by a uniform power spectral density, which means there is the same power in a given bandwidth at all frequencies up to infinity. White noise therefore has infinite power (variance). In reality a measurement is always limited in bandwidth and hence the above property of a constant power spectral density only holds within that bandwidth. Those bandlimited samples of white noise thus have a finite variance. Since white noise is so common, a few properties should be mentioned. One such property is, that the variance σ_{x+y}^2 of two uncorrelated variables x and y adds as:

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + \underbrace{2\,\text{Cov}(x,y)}_{\text{uncorrelated}\,=\,0} = \sigma_x^2 + \sigma_y^2 \tag{1.44}$$

This results in simple addition rules for variances from different sources, but it must be stressed here, that this property is only valid for uncorrelated sources like white noise, although it is usually incorrectly applied to all measurements, which unfortunately obscures rather than clarifies the uncertainties involved.

In order to demonstrate the effect of white noise in Allan deviation plots, it was simulated using the excellent *AllanTools* library [124]. The noise generator is based on the work of Kasdin et al. [54]. The full Python program code is published online [16] and found in data/simulations/sim_allan_variance.py. To allow better comparison, all noise densities are normalized to give an Allan deviation of $\sigma_A(\tau_0) = 1$, with τ_0 being the smallest time interval.

Figure 1.13 shows a sample of white noise in its three different forms. Figure 1.13a is the time series representation, from which the power spectral density was calculated and is shown in figure 1.13b. The dashed line shows the expectation value of the power spectral density and the Allan deviation.

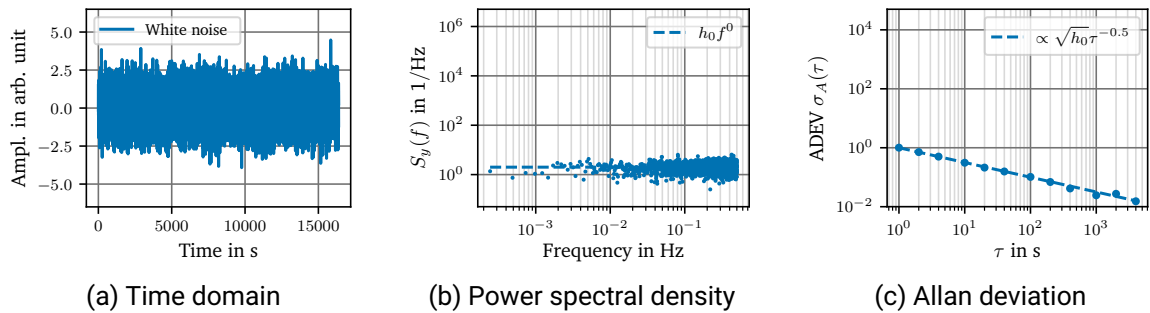(a) Time domain   (b) Power spectral density   (c) Allan deviation

Figure 1.13.: Different representations of white noise.

From this simulation, several features can be observed. First of all, the power spectral density is flat and constant with $h_0 = 2$, which is in accordance with table 1.4 and the normalization mentioned earlier. Figure 1.13c shows the typical $\tau^{-\frac{1}{2}}$ dependence of white noise in the Allan deviation plot. This immediately explains, why filtering white noise scales with $\frac{1}{\sqrt{n}}$ with n being the number of samples averaged.

Burst Noise

Burst noise, popcorn noise, or sometimes referred to as random telegraph signal is a random bi-stable change in a signal and is caused by a generation recombination processes. This, for example, happens in semiconductors if there is a site, that can trap an electron for a prolonged period of time and then randomly release it. Imporities causing lattice defects are discussed in this context [23, 56, 57, 91]. Such latttice defects can also be introduced by ion implantation during doping. Fortunately, this type of noise has become less prevalent in modern manufacturing processes, because the quality of the semiconductors has improved. But if a trap site is located very close to an important structure, for example a high precision Zener diode, its effect might be so strong, that it can be clearly seen.

The discussion is split into two parts. First the power spectral density is calculated and then the Allan variance is calculated using that result.

The spectral density of burst noise caused by a single trap site was derived in [74] by Machlup. The author used the autocorrelation function of the burst noise signal and applied the Wiener-Khinchin (Wiener-Хи́нчин) theorem, which connects the autocorrelation function with the power spectral density. A more detailed derivation can be found in [128], in this paper the preconditions, like stationarity of the process, are also discussed. The burst noise signal consists of two energie levels, called 0 and 1, split by Δy. Multiple burst noise signals can be superimposed in a real device. This would then result in mutiple levels, but they can be treated separately. The measurement interval over an even number of transitions, so that one ends in the same state as the measurement has started, is the time T. The mean lifetime of the levels is called $\bar{\tau}_0$ and $\bar{\tau}_1$:

$$\bar{\tau}_0 \approx \frac{1}{N}\sum_i^N \tau_{0,i} \qquad \bar{\tau}_1 \approx \frac{1}{N}\sum_i^N \tau_{1,i} \tag{1.45}$$

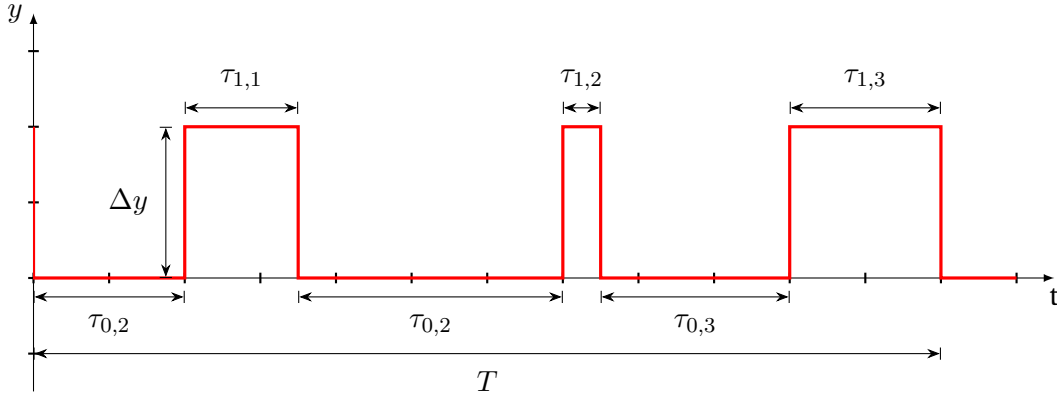Figure 1.14 shows a burst noise signal along with the definitions above.

Figure 1.14.: A random burst noise signal.

Using these definitions, one can then derive [74]:

$$R_{xx}(T) = \Delta y^2 \cdot \frac{\bar{\tau}_1 \bar{\tau}_0 e^{-\left(\frac{1}{\bar{\tau}_1} + \frac{1}{\bar{\tau}_0}\right)T}}{\left(\bar{\tau}_1 + \bar{\tau}_0\right)^2} \quad \text{and} \tag{1.46}$$

$$S(\omega) = 4R_{xx}(0)\frac{\frac{1}{\bar{\tau}_1} + \frac{1}{\bar{\tau}_0}}{\left(\frac{1}{\bar{\tau}_1} + \frac{1}{\bar{\tau}_0}\right)^2 + \omega^2} \qquad \omega > 0. \tag{1.47}$$

Note, that the power spectral density is the one-sided version, hence an additional factor of 2 is included. The d.c. term was ommitted here and can usually be neglected, because it is not relevant for calculating the power spectral density as it only contributes a single peak at $\omega = 0$. Using the following definitions of the average time constant and the duty cycle

$$\frac{1}{\bar{\tau}} = \frac{1}{\bar{\tau}_1} + \frac{1}{\bar{\tau}_0} \quad \text{and} \tag{1.48}$$

$$D_i = \frac{\bar{\tau}_i}{\bar{\tau}_1 + \bar{\tau}_0} \quad i \in \{0; 1\} \tag{1.49}$$

equations 1.46 and 1.47 can be rewritten to give a more intuitive form.

$$R_{xx}(T) = \Delta y^2 D_1 D_0 \, e^{-\left(\frac{1}{\bar{\tau}_1} + \frac{1}{\bar{\tau}_0}\right)T} \tag{1.50}$$

$$S(\omega) = 4R_{xx}(0)\frac{\bar{\tau}}{1 + \omega^2\bar{\tau}^2} \tag{1.51}$$

The special case $\bar{\tau}_0 = \bar{\tau}_1$ with $D_i = \frac{1}{2}$ is the previously mentioned case of random telegraph noise.

$R_{xx}(0)$ can be identified as the mean squared value of y:

$$y_{RMS} = \sqrt{R_{xx}(0)}. \tag{1.52}$$

Equation 1.51 is a Lorentzian function and from this, it can be easily seen, that a single trap site has a power spectral density, which is proportional to $\frac{1}{f^2}$ at high frequencies and is flat at low frequencies.

With the spectral density in hand, it is now possible to calculate the Allan variance as it was done by Van Vliet et al. in [119] for the classic example of random telegraph noise where $\bar{\tau}_1 = \bar{\tau}_0$. Do note, that table I given by Van Vliet et al. shows the total number of events instead of the instantationous number of events typically given. Hence, their notation must be multiplied by $\frac{1}{\tau^2}$ (or $\frac{1}{T^2}$ in their notation). For the generic case with $\bar{\tau}_1$, $\bar{\tau}_0$ and the definition of $\bar{\tau}$ given in equation 1.48 one finds for the Allan variance of burst noise:

$$\sigma_A^2(\tau) = R_{xx}(0)\frac{\bar{\tau}^2}{\tau^2}\left(4e^{-\frac{\tau}{\bar{\tau}}} - e^{-\frac{2\tau}{\bar{\tau}}} + 2\frac{\tau}{\bar{\tau}} - 3\right) \tag{1.53}$$

Having arrived at equations 1.51 and 1.53 of the power spectral density and Allan variance, it it now possible to model it. For this purpose, parts of the Python library *qtt* [33] was used. The algorithm written by Eendebak et al. implements continous-time Markov chains to simulate the burst noise signal. The result can be see in figure 1.15. For these simulations one time constant, namely the lifetime of the lower state $\bar{\tau}_0$ was held constant, while the lifetime of the upper state was varied to show the effect of different $\bar{\tau}$. By looking at the time domain in figure 1.15a it can be seen, that the maximum average number of state changes can be observed, when $\bar{\tau}_1 = \bar{\tau}_0$. If $\bar{\tau}_1 > \bar{\tau}_0$ the system will favour the upper, while if $\bar{\tau}_1 < \bar{\tau}_0$ it will favour the lower state instead. This explains why the noise is strongest for random telegraph noise when $\bar{\tau}_1 = \bar{\tau}_0$, which can also be seen in power spectral density in figure 1.15b. Looking at the Allan deviation in figure 1.15c confirms this, but also shows another interesting implication as it shows an obvious maximum. If the application allows a choice over the sampling interval τ, the effect of the burst noise can mitigated by staying well clear of the maximum.

The small deviation from the analytical solution in figure 1.15c suggesting an upwards trend at large τ is a typical so called end-of-data error. As it was discussed above, the Allan deviation

can only be estimated given a limited number of samples using equation 1.41 and going to longer τ means there are fewer samples to average over.

The burst noise equations can used to gain further insight into other types of noise. The first one is Shot noise, which is commonly found in photodetectors and lasers. Here, electrons or photons are created at discrete intervals resulting in an instantationous signal. This means, that the lifetime of the upper level is very short in comparison to the lower level ($\tau_1 \ll \tau_0$) equation 1.47 becomes:

$$S_{Shot}(\omega) = S_{\tau_1 \ll \tau_0}(\omega) = 4\Delta y^2 \frac{\tau_1}{\tau_0} \frac{\frac{1}{\tau_1}}{\left(\frac{1}{\tau_1}\right)^2 + \omega^2}$$

$$= 4\Delta y^2 \frac{1}{\tau_0} \frac{1}{\frac{1}{\tau_1^2} + \omega^2} \tag{1.54}$$

$$\overset{\omega \ll 1/\tau_0}{\approx} 4\Delta y^2 \frac{\tau_1^2}{\tau_0} = \text{const.} \tag{1.55}$$

For the typical case, a very large number of such events happen. When not counting single events, but rather a stream, the relation $\omega \ll 1/\tau_0$ is valid and hence the results is a white spectrum as $S_{Shot}(\omega)$ is constant with respect to ω — just as observed in photodetectors and lasers.
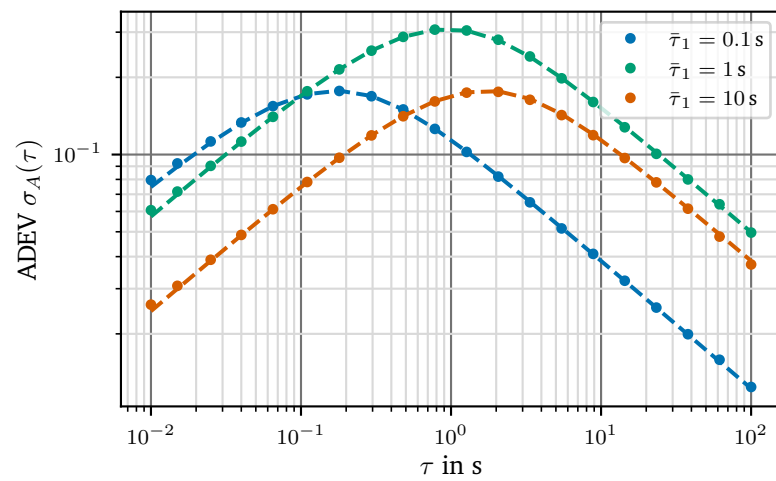
The other interesting case is a case, where many trap sites with different time constants are contributing to the noise. This can change the shape of the spectrum from f^{-2} to f^{-1} and is discussed in the next section.
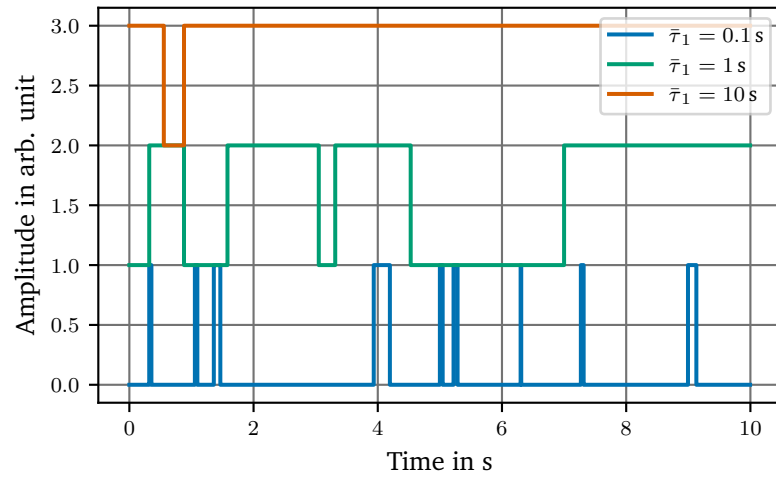
(a) Time domain
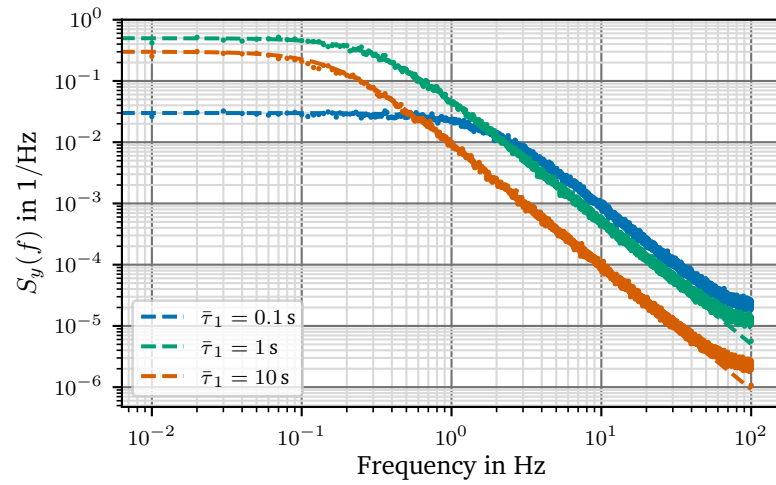
(b) Power spectral density

(c) Allan deviation

Figure 1.15.: Different representations of burst noise for different $\bar{\tau}_1$ and fixed $\bar{\tau}_0 = 1\,\mathrm{s}$.

Flicker Noise

Flicker noise is also called $\frac{1}{f}$-noise and it can be observed in many naturally occuring phenomenen. Its origin is not clear, although there have been many explanations. An overview can be found in [36, 82, 89]. This work concentrates on flicker noise in electronic devices. In thick-film resistors, for example, it was shown to extend over at least 6 decades without any visible flattening [86]. In transistors, flicker noise is caused by the existance of generation-recombination noise or burst noise discussed in the previous section [36]. If there are many uncorrelated trap sites, that contribute to the total noise, the envelope of the noise spectral density changes from $\frac{1}{f^2}$ to $\frac{1}{f^1}$ as shown in figure 1.16
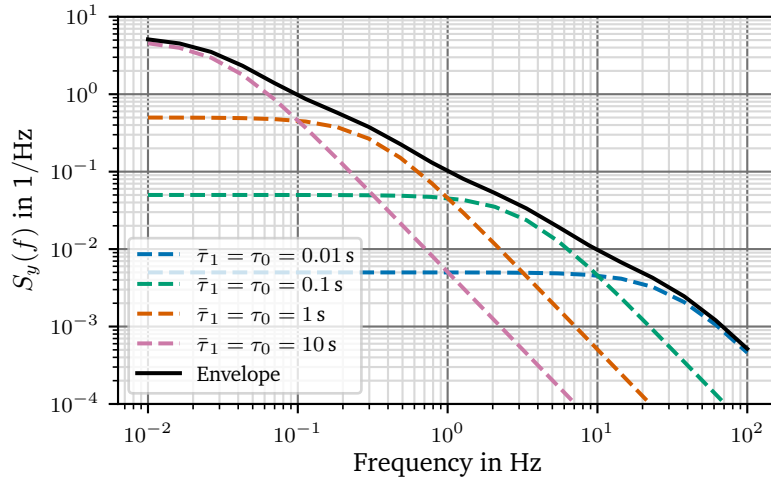

Figure 1.16.: Multiple overlapping Lorentzian noise sources forming a $\frac{1}{f}$-like shape.

Given that no trap site can store an electron indefinetely, the number of trap sites N with a certain time constant $\frac{1}{2}\bar{\tau} = \bar{\tau}_0 = \bar{\tau}_1$ must decline for longer time scales. Assuming N is inversely proportional to the time constant $\bar{\tau}$

$$N(\tau) \propto \frac{1}{\bar{\tau}}\,, \tag{1.56}$$

which can be motivated if the trapping process is thermally activated [32] and using equation 1.51 from the previous section, multiplying the weight function 1.56 and integrating over all possible storage times gives:

$$
\begin{aligned}
S(\omega) &= \lim_{t \to \infty} \int_0^t N(\bar{\tau})\, 4R_{xx}(0) \frac{\bar{\tau}}{1 + \omega^2 \bar{\tau}^2}\, d\bar{\tau} \\
&\overset{\bar{\tau}_0 = \bar{\tau}_1}{=} 4R_{xx}(0)\, C_N \lim_{t \to \infty} \int_0^t \frac{1}{1 + \omega^2 \bar{\tau}^2}\, d\bar{\tau} \\
&= \frac{4R_{xx}(0)\, C_N}{\omega} \lim_{t \to \infty} \arctan \bar{\tau}\omega \Big|_{\bar{\tau}=0}^t \\
&= \frac{4R_{xx}(0)\, C_N}{\omega} \cdot \frac{\pi}{2} \\
&= \frac{2\pi R_{xx}(0)\, C_N}{\omega}
\end{aligned}
\tag{1.57}
$$

$$S(f) = h_{-1} f^{-1} \tag{1.58}$$

C_N is the proportionality constant of 1.56 and h_{-1} is the power coefficient introduced in 1.43. This shows, that for a large number of distributed trap sites, a noise spectrum of f^{-1} is found.

Using equation 1.42, the Allan variance can be calculated from the power spectral density:

$$\sigma_A^2(\tau) = 2h_{-1} \int_0^\infty \frac{1}{f} \frac{\sin^4(\pi f \tau)}{(\pi f \tau)^2} \, df$$
$$= 2 \ln 2 \, h_{-1} \tag{1.59}$$

Again, using the *AllanTools* library [124], flicker noise was simulated to give an impresion of its properties.

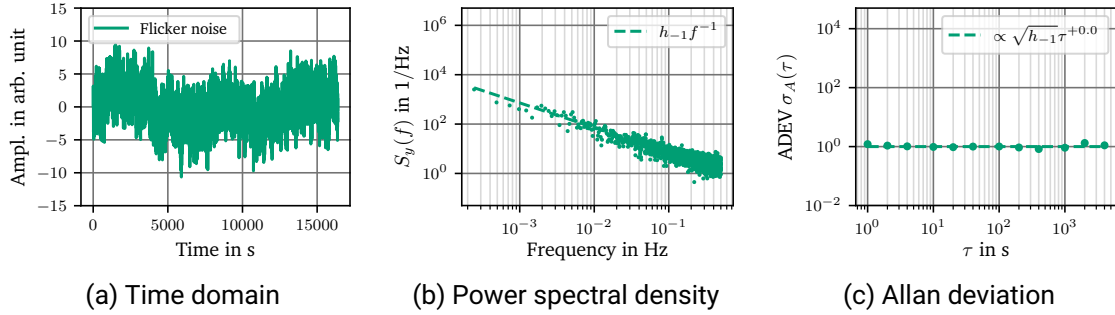(a) Time domain   (b) Power spectral density   (c) Allan deviation

Figure 1.17.: Different representations of flicker noise.

While it is not immediately evident from the power spectral density, the Allan deviation plot explains very well, why additional filtering does not affect flicker noise. No matter how long the integration time, the variance will still be the same.

The small wiggles at longer τ are typical end-of-data errors caused by spectral leakage, because there are insufficient samples to average over [49]. As it was discussed above, the Allan deviation can only be estimated using equation 1.41 given a limited number of samples. Therefore, at $\frac{\tau}{2}$ there are only 2 samples left, so there is no averaging possible to improve the estimate of the Allan deviation, which causes the oscillations at low frequencies or large τ.

As a last remark, a commonly used definition in combination with flicker noise is the corner frequency f_c. The corner frequency appears in situations, where there is both flicker and white noise present. It is the crossover point in frequency, where the flicker noise is equal compared to the white noise.

$$f_c = \frac{h_{-1}}{h_0} \tag{1.60}$$

It can be graphically extracted from the power spectral density plot by drawing a line trough the flicker noise and the white noise and finding the intersection. This can be seen in figure 1.22 on page 42. The corner frequency can be found where the horizontal dashed blue and green line meet.

Random Walk

Random walk noise can be attributed to environmental factors such as temperature [120] and diffusion processes, the latter contributing to the ageing effect seen in semiconductors. It is a process, where in each time step the change is randomly determined to be either a positve or negative step with equal probability and a fixed step size. Its mean is

$$\langle y_n \rangle = \langle e_1 + e_2 + \ldots e_n \rangle = \underbrace{\langle e_1 \rangle}_{=0} + \langle e_2 \rangle + \cdots + \langle e_n \rangle = 0 \,, \tag{1.61}$$

but its variance

$$\sigma_y^2 = \langle y_n^2 \rangle - \underbrace{\langle y_n \rangle}_{=0} = \sigma_{e_1}^2 + \sigma_{e_2}^2 + \ldots \sigma_{e_n}^2 = n\sigma_e^2 \tag{1.62}$$

goes with n (or t). It therefore not a stationary process as can also be seen in figure 1.18c.

The power spectral density can be calculated [12, 54] to

$$S(f) = h_{-2} \frac{1}{f^2} \tag{1.63}$$

and the Allan deviation can again be calculated from the spectral density

$$\sigma_A^2(\tau) = 2h_{-2} \int_0^\infty \frac{1}{f^2} \frac{\sin^4(\pi f \tau)}{(\pi f \tau)^2} \, df$$
$$= \frac{2}{3}\pi^2 h_{-2}\,\tau \tag{1.64}$$

The *AllanTools* library [124] can then be used to simulate the random walk.

(a) Time domain   (b) Power spectral density   (c) Allan deviation
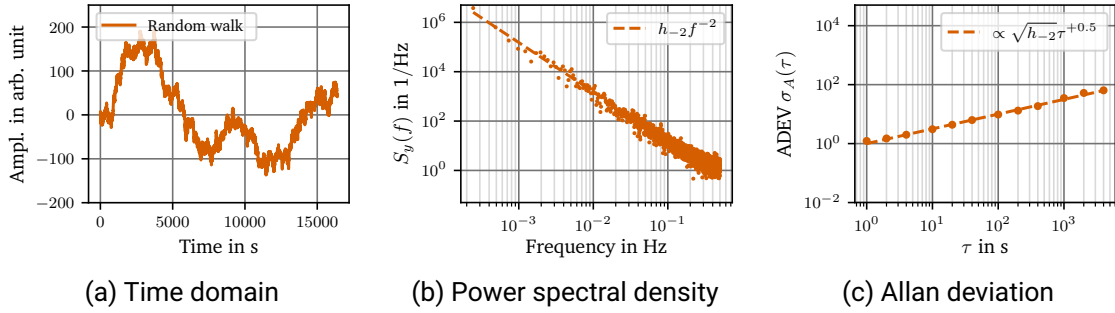
Figure 1.18.: Different representations of random walk noise.

Drift

Finally, the last feature of the Allan deviation plot, that needs to be discussed is drift. Drift happens at very long time scales and describes a linear dependence of the measurand on time. This is also part of the ageing effect. Greenhall discussed the effect of drift [45] on the Allan variance and found the following relationship:

$$\sigma_A^2(\tau) = \frac{D^2}{2}\tau^2 \tag{1.65}$$

with slope of the drift D.

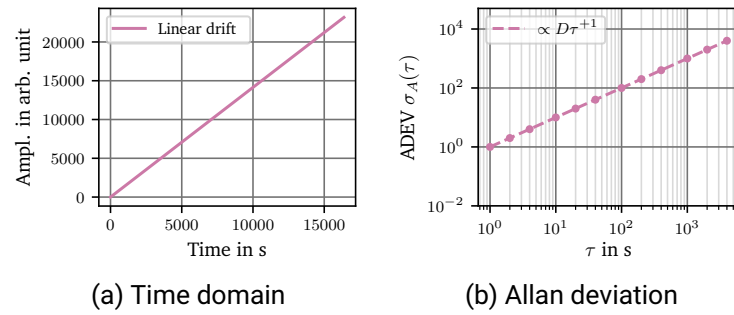(a) Time domain   (b) Allan deviation

Figure 1.19.: Different representations of linear drift.

Dead Time

The coefficients given here in the previous examples were derived using the assumption, that all sampling in a measurement are continous with a dead time $\theta = 0$. Unfortunately, measurements sometime have a dead time, that is non negligible. This problem was extensively discussed by Barnes et al. [12]. Dawkins et al. even developed special models to account for the algorithms of modern frequency counters [27]. While some frequency counters support gapless measurements, the situation is entirely different for digitizers and digital multimeters. Several settings commonly used affect the dead time, which can be considerable. It is therefore important to discuss typical measurement settings for voltmeters to estimate the errors that arise from those settings. The focus of this discussion lies on the dead time introduced by digital multimeters, but the application is not limited to this field.

The most commonly used settings, that affect the dead time of a voltmeter are auto-zeroing and line synchronization. Auto-zeroing is done by adding additional measurements to the normal input integration cycle. To correct for the zero offset drift a zero measurement is added where the ADC is switched to the low terminal. Additionally, some devices add a reading of the reference voltage to correct for gain errors. The implementation details and type of measurements are manufacturer dependent.

The other setting, that can be enabled in voltmeters, is the line synchronization to increase the noise rejection of the instrument. This setting synchronizes the start of a measurement to the zero crossing of the power line. Depending on the instrument, this might cause a delay of one power line cycle (PLC) after each measurment if the instrument is not capable of processing the previous measurement while at the same time recording another one.

A simple measurement with dead time is shown in figure 1.12 on page 28. That model assumes, that the dead time is constant and is always added after the actual integration time τ. This is rarely true for real measurement data as many devices and even ADCs use internal averaging and auto-zeroing to produce a measurement. The actual dead time is therefore spread over the whole measurement and not limited to the end of the measurement. An example is the Keysight 3458A DMM, which automatically switches to averaging when selecting integration times greater than $10\,\text{PLC}$. The reason is simple, for longer integration times, more and more flicker noise starts contributing to the measurement. The measurement is therefore split into single measurements of $10\,\text{PLC}$ and using auto-zeroing the flicker noise is suppressed. This is discussed in more detail as an example in section 1.7. The mathematical problem of a distributed dead time was already noted by Allan [7] and it is distinctively different from the calculations made by Barnes et al. for a single dead time at the end of the measurement. The exact mathematical treatment is complex and is beyond the scope of this work, especially considering, that auto-zeroing does a lot more than just adding dead time at the end of the measurement. Fortunately using a few assumptions the problem can be greatly simplified.

An interesting observation can be made for white noise. Since it is uncorrelated, it makes no difference whether it is sampled in full, or only partially, therefore the Allan deviation for a white noise process with or without dead time is the same:

$$\sigma^2(N, T, \tau) = \sigma^2(N = 2, T = \tau, \tau) = \sigma_A^2(\tau)\frac{1}{2}h_0\tau^{-1} \qquad (1.66)$$

Consequently, if the dead time is added at a frequency high enough, so that the input amplifier output is dominated by white noise, the dead time will have no influence on the Allan variance.

Finally, Barnes et al. [12] notes that for measurement durations or averaging times $T \gg T_0$, the Allan variance with respect to T shows an asymptotic behaviour of $\sigma_A^2(T) \to \sigma_A^2(\tau)$.

1.6.2. Example

Using the results from the previous sections, it is possible to simulate a typical measurement sample containing white noise, flicker noise and random walk behaviour. The simulation was written in Python using the *AllanTools* library [124] to generate the time domain data, which was then converted to a power spectrum using the algorithm of Welch [125]. The Allan deviation was calculated using the *AllanTools*. The full Python source is available at [16] and found in data/simulations/sim_allan_variance_example.py. The time domain data shown here was downsampled from 2^{25} data points to 2000 points for faster plotting, using the Largest-Triangle-Three-Buckets (LTTB) algorithm created by Steinarsson [112]. The downsampling algorithm chosen is optimal for this application, because it aims to visually keep the result the same, by favouring parts of the data, where there is more dynamics. The only difference noticable to the author is, that the edges of the white noise plot are a little rougher. The full data set can be obtained using the source code given above if one desires. The power spectrum and the Allan deviation were always calculated from the full dataset. The data of the power spectrum was additionally binned to be evenly spaced on a logarithmic scale. This considerably reduced the high frequency noise and made the plot easier to read while not negatively impacting the shape.

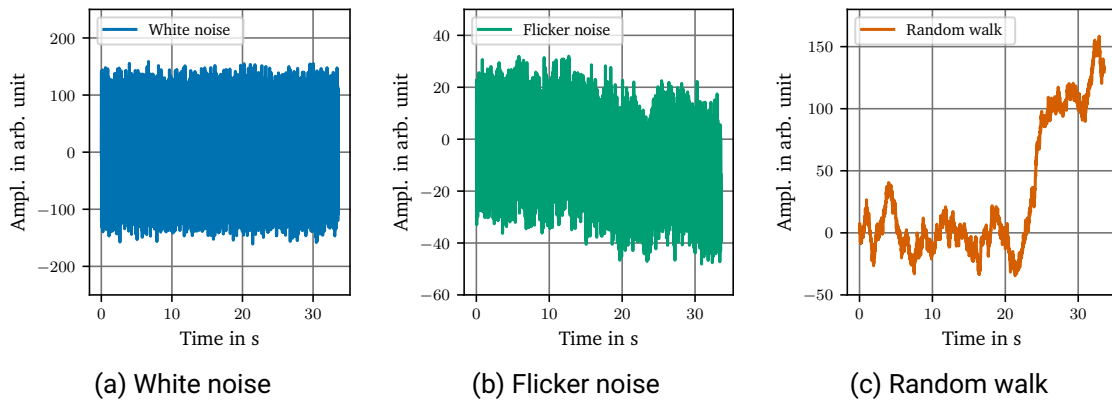(a) White noise   (b) Flicker noise   (c) Random walk

Figure 1.20.: Three sperate noise components, that were summed together to simulate a typical noise source.

The three time series shown in figure 1.20 were sequentially generated using a fixed seed for the random number generator to ensure repeatability as long as the order of creation is kept the same. For generting the noise, the algorithm presented by Kasdin et al. [54], implemented in the *AllanTools* library was used. The noise strength parameters were deliberately chosen in such a way, that both the white noise and the random walk part have more noise power than the flicker noise. This allows to distinguish them in the plots at both extremes of the frequency scale. Finally, the three types of noise data were summed together to give the combined signal, which is shown in figure 1.21, again downsampled using LTTB. The summed series clearly shows the white noise content and it is possible to deduce some flicker or random walk noise, but it is highly obscured due to the amount of white noise. Using only the time domain plot makes it very hard to distinguish the type of noise present, let alone estimate the individual noise power of the three sources. Therefore, a different analysis tool is called for.

A common approach to identify noise sources is the power spectrum. It is easily accessible, even in real-time, using spectrum analyzers and, utilizing the computational power of modern computers, large time-domain data sets can be converted making this the method of choice
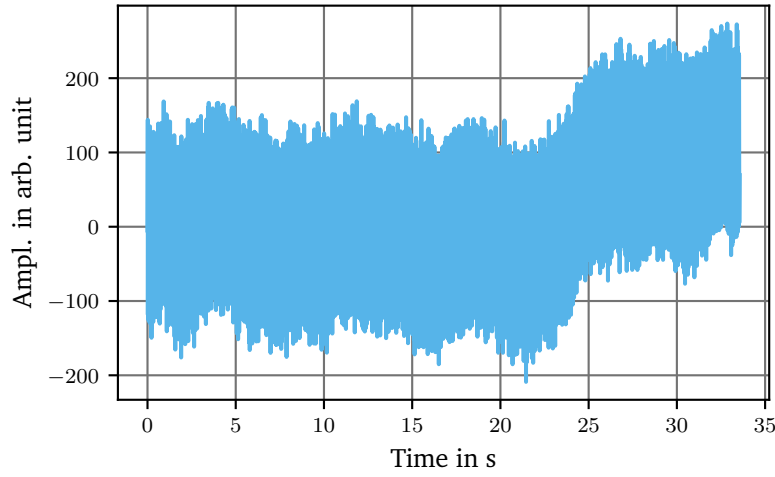
Figure 1.21.: A simulated time series containing white noise, flicker noise and random walk behaviour.

in the lab. The power spectrum of figure 1.21 is shown in figure 1.22. It allows to clearly separate the white noise part from the other f^α components. The dashed lines representing the individual components were plotted using the h_α values calculated from the input parameters of the simulation. The noise spectral density h_0 of the white noise signal can be easily extracted even by hand without resorting to a fit. This yields $h_0 = 2 \times 10^{-3}$/Hz. h_{-1} and h_{-2} can be exctracted as well using a fit to

$$S(f) = \sum_{\alpha=-2}^{0} h_\alpha f^\alpha \,. \tag{1.67}$$

The noise corner frequency f_c can either be calculated from h_0 and h_{-1} using equation 1.60 or determined graphically by contructing a tangent with a slope of -1 to the spectral density. From the intersection of the blue h_0 line and the green h_{-1} line the corner frequency is found to be $f_c \approx 1.8$ kHz.
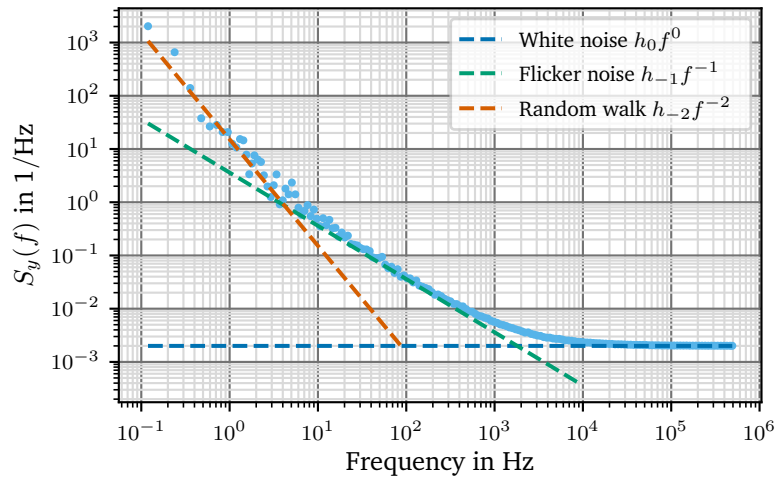

Figure 1.22.: A simulated power spectrum containing white noise, flicker noise and random walk behaviour.

To get an even better representation of the individual noise contributions, the Allan variance or Allan deviation can be used. The Allan deviation plot shown in figure 1.23 gives very clean results and all noise components can be clearly identified. The individual components were plotted using dashed lines as well.
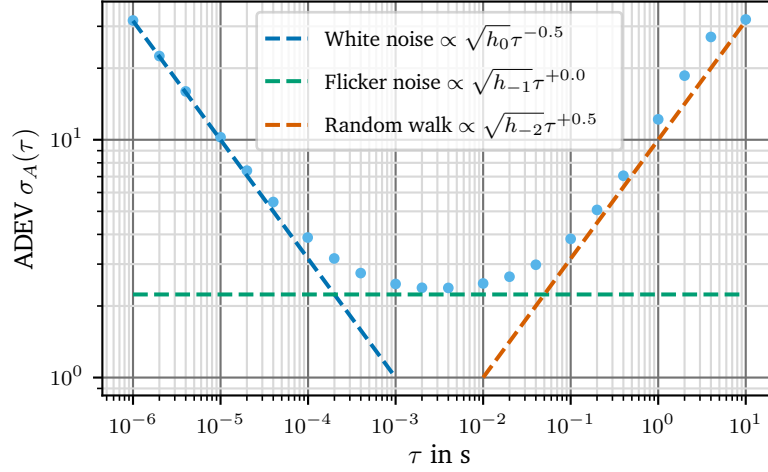

Figure 1.23.: A simulated Allan deviation containing white noise, flicker noise and random walk behaviour.

The Allan variance was calculated using the overlapping Allan variance algorithm [95] and only Allan deviation values for frequency vales of $(1, 2, 4)$ per decade were plotted. The overlapping Allan variance gives a better confidence at longer intervals or lower frequencies, allowing to identify very low frequency noise like the random walk shown here. Reference [95] also gives a very good comparison of other algorithms to identify even more noise types in data sets, like phase noise. Plotting only three values per decade improves the clarity of the plot, because at longer taus, even though the overlapping Allan variance is used, some oscillations inevitably show up. Using fewer values of τ causes less distractions in this case. From the figure 1.23, the Allan deviation of the flicker noise can be estimated from the flat minimum to be around 2.3 or $\sqrt{5}$. Using table 1.4 the Allan variance can be converted to

$$h_{-1} = \frac{5}{2 \ln 2} \approx 3.6$$

Using the previously found h_0, this corner frequency is calculated using equation 1.60 to be:

$$f_c = \frac{5}{2 \times 10^{-3}/\text{Hz} \cdot 2 \ln 2} \approx 1.8\,\text{kHz}$$

This is obviously the same result as the one from the geometric approach above.

This concludes the examples of different noise types. The reader should now able identify different types of noise in meansurement data and have learnt to appreciate the value the Allen variance brings to the table. An example was presented, that applied all techniques shown in this section to extract information about the noise sources in a dataset. Additionally, Python source code is provided to further explore the topic.

1.7. Autozeroing

Autozeroing (AZ), sometimes called zero-drift or dynamic offset compensation, is such an important concept, that it must be discussed in its own right. The need for autozeroing comes from the typical behaviour of amplifiers. Every amplifier has some offset, be it small or large, and especially at high gains, this offset becomes a problem for high precision measurements. To make matters worse, this offset is not stable over time and drifts with both time and temperature. It can therefore not be calibrated out once, it must be permanently adjusted during operation, depending on environmental conditions. This procedure is called autozeroing.

There are many different ways to implement autozeroing and regarding operational amplifiers a good overview can be found in [48]. As an example, the autozero cycle for the Keithley `Model 2002` and the Keysight 3458A Mulimeter is shown in figure 1.24. Keithley uses a more complex and slower algorithm, while HP implemented a simpler, but faster algorithm. The most simple (digital) approach is to regularly switch the input from the signal to zero, take a reading, then subtract this reading from all subsequent readings until a new zero reading is taken. An alternative approach adds another measurement of the reference voltage to apply a gain correction as well. This is done by the Keithley `Model 2002` and works very well to suppress gain drift in the input amplifier due to temperature changes, but increases the time between samples by another 50%. This is the reason, why on the `Model 2002`, there is another mode to strech the time between autozero cycles, but the price to pay is a non-uniform sampling rate. This makes post-processing a lot more complicated as most alorithms assume a (near) constant sampling rate. The Keysight 3458A calculates gain corrections only during the manual autocalibration routine to maintain a higher throughput.
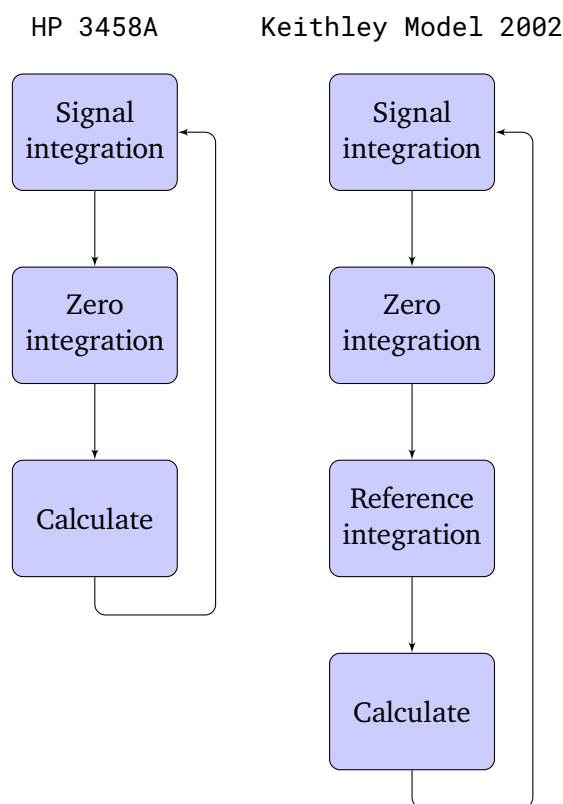
HP 3458A   Keithley Model 2002

Figure 1.24.: Auto-zero phases of the Keysight 3458A and Keithley `Model 2002`.

1.7.1. Offset-Nulling

Offset-nulling is the most basic approach to autozeroing. It aims to remove the offset drift of an amplifier. Especially at high gains, the offset, which is multiplied by the gain, can be substantial. In order to explain how offset-nulling works and how it shapes the spectrum, it is best to discuss it based on an example. While this technique can also be found in many integrated circuits, it is more noticable in DMMs, because is a switchable option. Therefore, the example data set simulated is based on the parameters of the aforementioned Keysight `3458A` multimeter. The corner frequency and the white noise floor is modeled after the $10\,\mathrm{V}$ range of the `3458A` [62, 64] with the values given below. Do note that both references [62, 64] contain a typographical error. The corner frequency of the noise floor is erroneously given as $0.5\,\mathrm{Hz}$, but should be $1.5\,\mathrm{Hz}$. This can be seen in figure 2.35 in [62, p. 116], where the noise spectral density is plotted and it was also confirmed with the author [63]. The data used in this section is generated using the Python *AllanTools* library [124] and the simulation source code can be found in data/simulations/sim_autozero.py as part of the online suplemental material [16].
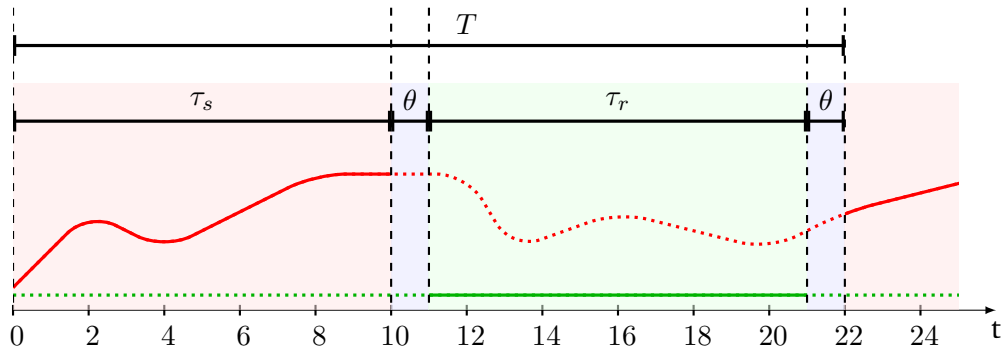

Figure 1.25.: Integration sequences of the offset-nulling algorithm. Solid lines denote sampled data. Red is the input signal, green is the zero reading and blue is the dead time required for switching inputs.

For this simulation, a noise-free and arbitrarily chosen $10\,\mathrm{V}$ input is assumed to be sampled by the device at a sampling rate of $10\,\mathrm{PLC}$ at $50\,\mathrm{Hz}$, the same rate discussed previously on page 40. As it will be shown, the actual mean value of the input signal has no bearing on the outcome of the calculation when considering offset-nulling, but its value must be considered for other types of autozeroing as discussed in section 1.7.2 and is included here only for the sake of completeness.

Figure 1.25 shows the individual sequences of the offset-nulling algorithm. First, the source is sampled for $\tau_s = 10\,\mathrm{PLC}$, then the input is switched to the LO terminal. While this operation is very fast and takes less than $1\,\mathrm{ms}$ [94], if the instrument is synchronized to the line frequency the zero measurement will nontheless be delayed until the next zero crossing, hence the dead-time $\theta = 1\,\mathrm{PLC}$. Finally, the zero reference is measured for another $\tau_r = 10\,\mathrm{PLC}$ and then the instrument switches back to the HI terminal.

The data is simulated in the following way. First, two sets of noise data are generated, a white noise spectrum with a noise spectral density of $165\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ and a flicker noise spectrum with an intesity scaled to result in a final spectrum with a corner frequency of $1.5\,\mathrm{Hz}$. The required flicker noise intensity is calculated using equation 1.60. To get a good low frequency estimate, $2^{20} \approx 10^6$ values were generated. Finally, the two noise data sets are summed with the noise-free input source to give the final result. Other effects, such as power-line

hum are neglected in this simple simulation, because it would needlessly overcomplicate the example and limit the educational value. The same goes for higher order random-walk f^{-2} noise components, which can be introduced by temperature fluctions and other environmental effects and would be present in a real measurement.
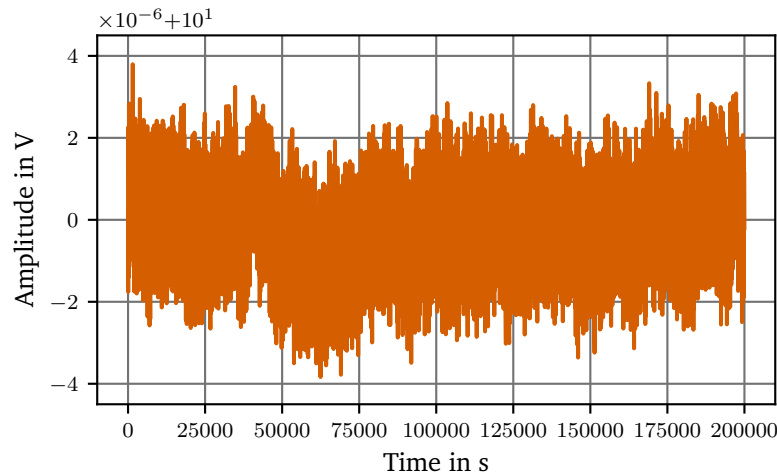

Figure 1.26.: Time series data with white noise and flicker noise.

The time domain plot of the simulation is shown in figure 1.26. The white noise component is clearly visible, while the f^{-1} flicker noise can be recognized, but its strength can hardly be estimated. It was already shown in section 1.6.2, that different types of noise have different frequency components and can be be distinguished in the frequency domain, which leads to the next approach.
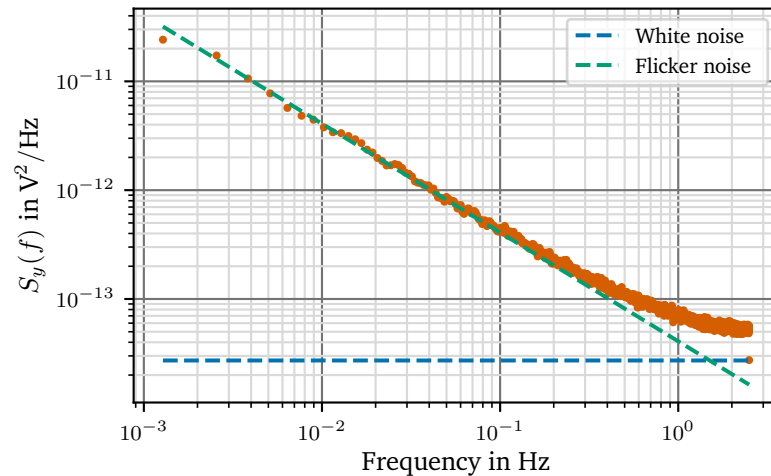
Figure 1.27.: Simulated power spectrum of a Keysight 3458A containing white noise and flicker noise. The line frequency is $50\,\mathrm{Hz}$.

The noise power spectral density shown in figure 1.27 is calculated from the time series and confirms the flicker and white noise content. The theoretical white noise floor is shown as a horizontal dashed blue line and the flicker noise as a dashed green line. The $1.5\,\mathrm{Hz}$ corner frequency, which is defined as the intersection between the f^{-1} noise and the white noise floor

easily identified using the those lines. It is evident, that the $5\,\mathrm{Hz}$ sampling frequency with a $2.5\,\mathrm{Hz}$ bandwidth does not allow the spectral density to fully settle to the noise floor.

From the power spectral density is can be seeen, that higher frequencies have a significantly lower noise spectral density than the low frequencies. It is therefore most beneficial, to do measurements at higher frequencies. To discuss the optimal measurement interval, the Allan deviation is an excellent tool.
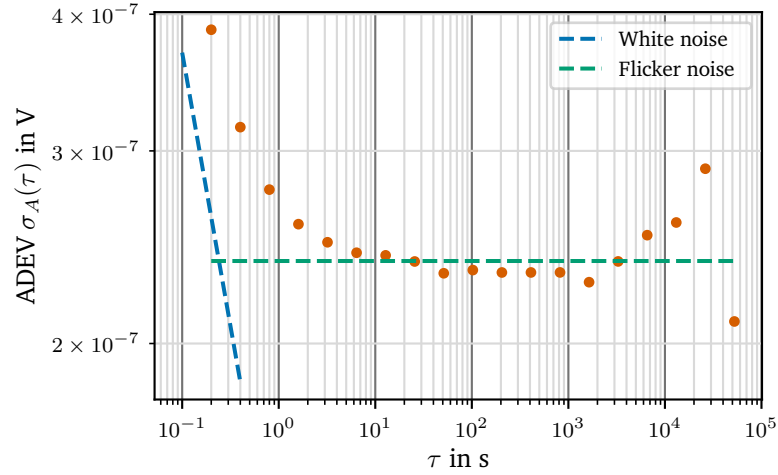

Figure 1.28.: Simulated Allan deviation of the input amplifier of a Keysight 3458A containing white noise and flicker noise. The line frequency is $50\,\mathrm{Hz}$.

The Allan deviation it plotted in figure 1.28 and shows two distinct regions. Short τ display an asymptotic behaviour towards white noise with a $\tau^{-0.5}$ dependence and at longer τ the constant flicker noise region can be identified. At very long τ typical end-of-data oscillations can be seen, which are the result of the limited confidence of the Allan deviation estimator as previously discussed and can be safely ignored. The Allan deviation clearly demonstrates the performance of the device at longer integration times and it is obvious, that beyond an integration time of about $1\,\mathrm{s}$ or $50\,\mathrm{PLC}$ no additional information can be extracted from the measurement and the variance is constant. This leads to the need for autozeroing to remove the flicker noise. It can be shown [97], that subtracting a reference measurement from the actual measurement data removes all correlated effects. Since flicker noise is autocorrelated, it can be removed by subtracting a zero measurement.

To demonstrate autozeroing, two cases will be discussed. Going back to figure 1.25 it can be seen, that between switching inputs, a dead time θ is added. For a first discussion, this dead time neglected and then the effect of adding a dead time is discussed in a next step.

Using figure 1.28 it was shown, that integrating over flicker noise, does not reduce the variance. In order to have as little flicker noise content in the final measurement value, it is clear, that the autozeroing should be done as fast possible to keep the flicker noise content out. This allows to calculate the expected variance of the autozeroed measurement. The noise of the input measurement x and the reference measurement y are the same, because in this model the only noise source comes from the input amplifier, as the input signal is assumed to be noise-free. The zero level is, by definition, noise-free. As dicussed above, the autozero interval is chosen, so that its variance is dominated by white noise The variance σ^2 of the combined measurement of $x - y$ can then be calculated using equation 1.44:

$$\sigma^2_{x-y} = \sigma^2_x + \sigma^2_y \tag{1.68}$$

By subtracting the zero reading, the amplifier noise is effectively added twice to the final result, once for the input measurement and once for the zero measurement. Additional noise from the input signal noise would simply be added to this as it is uncorrelated as well.

Do note, that the number of samples is now half the number before applying autozeroing. This leads to an interesting effect. Imagine a data set containing only white noise with a variance σ^2. Removing half the samples, obviously does not change the variance as white noise is not correlated, but subtracting the samples is effectively decimating the data set and since the sampling rate is halfed, the Nyquist band is halfed as well. Unfortunately the input noise bandwdith stays the same. The second Nyquist band is then folded back into the first, thus doubling the noise power density.

To conclude, it is expected, that the variance doubles and the power spectral density quadruples!

These considerations can be compared to the simulated data. Applying the autozeroing algorithm to the simulated data set, the constant $10\,\mathrm{V}$ input signal was nulled for every odd value and then the residual noise was subtracted from the signal value. The result in the time domain is shown in figure 1.29.
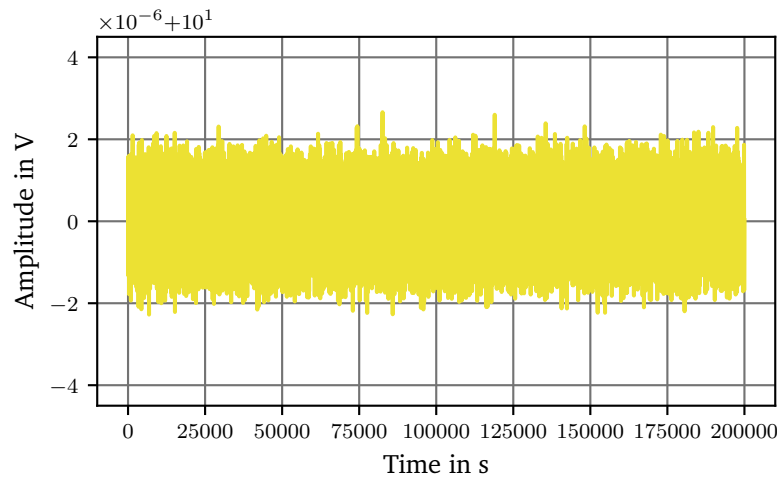

Figure 1.29.: Simulated measurement with auto-zeroing applied.

When comparing to figure 1.26 it is immediately evident, that the f^{-1} component is no longer present. The difference in white noise strength is difficult to compare and it must be turned gain to the power spectral density. When calculating the spectral density it is important to remember, that the sampling rate is now halved. The result is shown in figure 1.30 along with dashed lines showing the noise content prior to applying the autozero algorithm as before in figure 1.27.

The power spectral density in figure 1.30 confirms an increase in the white noise power as discussed above and it can be determined that the white noise power $\sqrt{h_{-1}}$ has increased from $165\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ to $489\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$, an increase by a factor of $\sqrt{8.8}$, which is more than estimated from equation 1.68, including the factor of 2 for the decimation, which gauged the increase of $\sqrt{h_{-1}}$ to be $\sqrt{4}$. The reason for the additional noise was already mentioned above. There is still some substantial f^{-1} noise present at the autozero frequency of $5\,\mathrm{Hz}$. This type of noise is not uncorrelated and therefore the covariance is not zero, hence equation 1.44 does not strictly hold and additional correlated noise is leaking into the result. The hypothesis can be confirmed, by increasing the sampling frequency by an order magnitude. Doing this, the white

noise floor of the autozero measurement now only increases by a factor of $\sqrt{4.5}$, which is close to the expected factor of $\sqrt{4}$. The means, that the autozeroing frequency should be at least a decade above the noise corner frequency to be most effective.

Figure 1.30.: Simulated power spectrum of a Keysight 3458A with autozeroing applied. The dashed lines denote the noise present prior to applying the autozero algorithm. The line frequency is $50\,\mathrm{Hz}$.

Nonetheless, down to very low frequencies, f^{-1} noise is effectively suppressed and the spectral density is almost perfectly flat. For reference, the dashed lines show the noise content that was in the dataset prior to autozeroing, which is less white noise, but far more flicker noise.

The Allan deviation plot in figure 1.31 also confirms, that white noise is the only component and shows a $\tau^{-\frac{1}{2}}$ dependence for the full range of integration times.
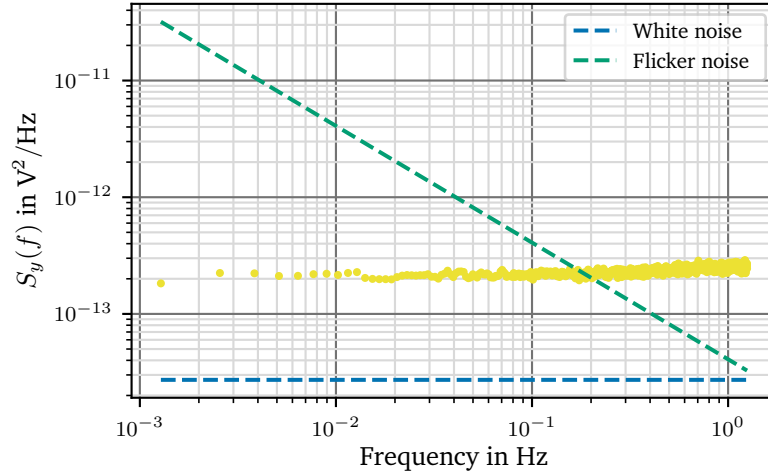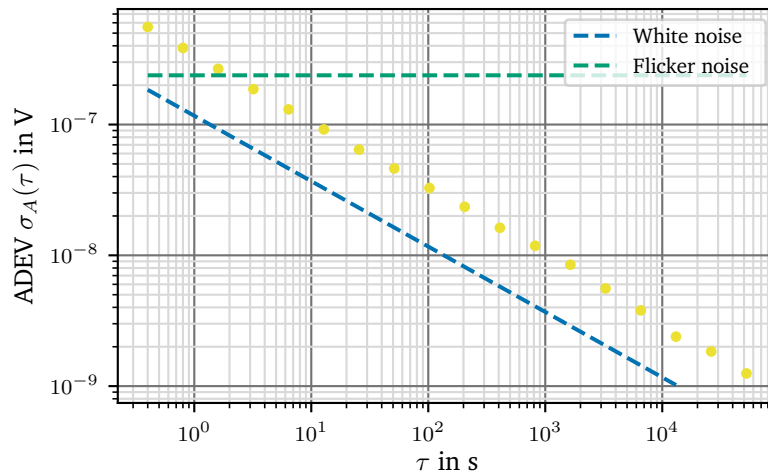

Figure 1.31.: Simulated Allan deviation of a Keysight 3458A with autozeroing applied. The dashed lines denote the deviation prior to applying the autozero algorithm. The line frequency is $50\,\mathrm{Hz}$.

From this plot it can be seen, that for measurement times longer than about $2\,\mathrm{s}$ or $100\,\mathrm{PLC}$

autozeroing has a clear benefit over a measurement without autozeroing. It must be noted though, that judging from this simulation, the device would reach a noise floor of $0.01\,\mu V/V$ only at integration times of slighly more than $10\,s$, while the datasheet claims $2\,s$. Do note, that this simulation is for the $10\,V$ range of the DMM and therefore $0.01\,\mu V/V$ is $0.1\,\mu V_{RMS}$. It is therefore likely, that the noise parameters of a real device are be better than the numbers used in the simulation. Additionally, the datasheet likely refers to an instrument, that is synced to a $60\,Hz$ power line frequency which shifts the sampling frequency up by $20\,\%$ and, as discussed, reduces the noise floor, because more noise content is white noise at the autozero interval. In this simulation the $0.01\,\mu V/V$ ($0.1\,\mu V_{RMS}$) noise level would be reached at exactly $10\,s$ when using a line frequency of $60\,Hz$. For the purpose of demonstrating the autozeroing algorithms these subtleties are irrelavant.

For the comparison of different intregration times before applying autozeroing figure 1.32 can be consulted. Using the Allan deviation makes it is very simple to compare noise figures for identical measurement times τ, but different integration times before autozeroing is applied. The simulation source code can be found in data/simulations/sim_optimal_autozero.py as part of the online suplemental material [16].

Figure 1.32.: Allan deviation for different integration times before applying the AZ algorithm. Dead time $\theta = 0\,s$. The dashed line denotes the Allan variance without autozeroing. The line frequency is $50\,Hz$.

It can be seen, that with increasing integration times before applying the AZ algorithm more uncertainty is accumulated due to the f^{-1} content, which cannot be filtered. As a result, after removing the f^{-1} content using autozeroing more time is required for filtering until the same Allan deviation can be reached. From these simulations it can be said, that when there is negligible dead time θ involved when switching the inputs it is advantageous to switch to switch early, while white noise is still dominating the noise content.

Finally, the case of a non-negligible dead time shall be treated. When the dead time has to be considered, it is clear, that the autozero frequency cannot be arbitrarily increased, because in increasing proportion of sampling time is lost to the dead time. This effective loss in sampling time then increases the noise spectral density due to aliasing as discussed above. To show this effect, the simulation above is modified to include a dead time of $1\,PLC$ as detailed in figure 1.25. The measurement sequence now includes a dead time after switching each input. Schieder, R. et al. proposes [97] splitting the measurement interval in two and instead of

measuring HI-LO-HI-LO, to measure HI-LO-LO-HI. This scheme is a mixed bag, because the f^{-1} noise is correlated and its autocorrelation function decays with e^{-t}, therefore constantly changing the order of subtracted samples is not as efficient in removing the noise as the normal autozero procedure. Only when the dead time is large in comparison to the measurement time, this method yields an advantage. Therefore only the simplest case of a HI-LO-HI-LO measurement is treated here. The Allan deviation for different integration times is evaluated in same way as in figure 1.32. The results are shown in figure 1.33.

Figure 1.33 demonstrates, that the effectiveness of the AZ scheme no longer increases with an increasing switching frequency and there is an optimal autozero interval. For the parameters chosen for this simulation ($f_c = 1.5\,\mathrm{Hz}$ and $165\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$), $5\,\mathrm{PLC}$ at $50\,\mathrm{Hz}$ is the optimal interval. If the corner frequency is shifted to a lower frequency, the optimum shifts more towards $10\,\mathrm{PLC}$. The same goes for a higher line frequency of $60\,\mathrm{Hz}$. This explains, why HP chose $10\,\mathrm{PLC}$ as the maximum integration time. For integration times higher than that, software averaging is used, therefore delivering the perfomance shown in figure 1.33 along the $10\,\mathrm{PLC}$ line.
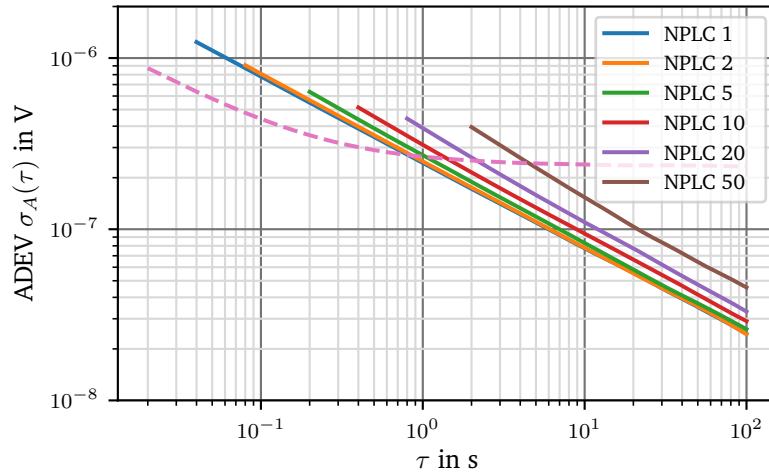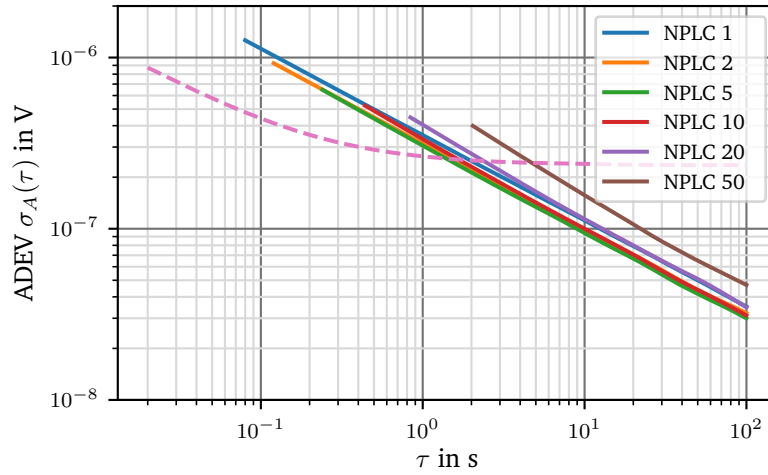

Figure 1.33.: Allan deviation for different integration times before applying the AZ algorithm. Dead time $\theta = 1\,\mathrm{s}$. The dashed line denotes the Allan variance without AZ. The line frequency is $50\,\mathrm{Hz}$.

It should be stressed here, that the dead time is not be the only factor to consider when choosing the autozero interval. For example, in case of an amplifier, switching the input also adds an error current due to the charge injection of the switching transistors. This may negatively impact the measurement of a high impedance source. These additional drawbacks are implementation specific and must considered during the design phase.

1.7.2. Gain Correction

The effect of the gain correction, where the input value x is scaled by a scaling factor y to adjust the gain error, can be calculated, assuming white noise, as follows:

$$\sigma_{x \cdot y}^2 = \langle x^2 y^2 \rangle - \langle xy \rangle^2$$

$$= \langle x^2 \rangle \langle y^2 \rangle + \underbrace{2 \, \mathrm{Cov} \left(x^2, y^2 \right)}_{\text{uncorrelated} = 0} - \left(\langle x \rangle \langle y \rangle + \underbrace{2 \, \mathrm{Cov} \left(x, y \right)}_{= 0} \right)^2$$

$$= \left(\sigma_x^2 + \langle x \rangle^2 \right) \cdot \left(\sigma_y^2 + \langle y \rangle^2 \right) - \langle x \rangle \langle y \rangle$$

$$= \sigma_x^2 \sigma_y^2 + \sigma_x^2 \langle y \rangle^2 + \sigma_y^2 \langle x \rangle^2 \tag{1.69}$$

With respect to the gain correction, equation 1.69 can be further reduced. The scaling factor is derived from the reference voltage V_{ref} and normalized using $\frac{V_{ref,meas}}{V_{ref}}$. The expected value, therefore is $\langle y \rangle \approx 1$, as the ADC full scale gain should not drift much. Furthermore, σ_y^2 is scaled by the constant $1/V_{ref}$ and $\sigma_x^2 \sigma_y^2 \ll \sigma_x^2$. The latter should be true for any measurement of significance.

$$\sigma_{x \cdot y}^2 \approx \sigma_x^2 + \sigma_y^2 \langle x \rangle^2 \tag{1.70}$$

The gain correction noise therefore behaves similar to the offset correction case, except, that it scales with the input voltage x and has no effect with a shorted input, while fully introducing its additional noise when a full scale input is applied.

1.8. Current Sources

Throughout this work the concept of current sources is widely used, for example section 1.3 discusses a current source to drive laser diodes and the temperature controller discussed in section 1.9 uses a current source to measure the resistance of a temperature sensitive resistor. While there are many more use cases, this section will limit the discussion to a few examples used by the devices presented in this work. Namely, this is a unidirectional transconductance amplifier with an operational-amplifier in conjunction with a field-effect transistor and a bidirectional Howland current pump invented by Bradford Howland in 1962 and first published in 1964 by Sheingold [103]. The discussion will start with the properties of the ideal current source and, based on that, develop a more accurate model. The models developed typically represent the static, time-independent case unless explicitly stated. First, the unidirectional current source is treated, then the bidirectional Howland current pump is discussed.

1.8.1. Current Sink and Current Source

The question whether to use a current source or a current sink is elemental for the design of a laser driver. Figure 1.34 shows different configuration of current sinks and sources with respect to the laser diode.

(a) Source with grounded LD.

(b) Source with floating LD.

(c) Sink with grounded LD.

(d) Sink with floating LD.
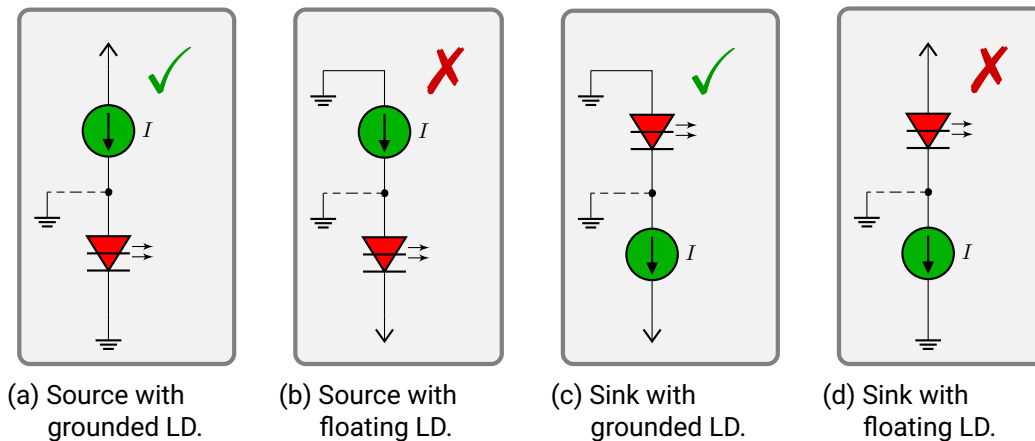
Figure 1.34.: Different configurations of current sinks and sources with respect to the laser diode. A green checkmark denotes a fail-safe configuration when accidentally shorting one or more pins of the diode to the laser chassis, illustrated by a dashed connection.

The most practical configuration depends on the laser diode and safety aspects in terms of protecting the laser diode. The protection of the laser diode is discussed first. The laser resonator is assumed grounded in the setup. This is not the design case, but incorrect assembly can facilitate this condition. While not intended, there are numerous ways to also accidently short the diode to ground and since there are no immediate consequences arrising from it, when the controller is disconnected, it might easily be overlooked. This blunder should not bear the risk of destroying an expensive laser diode. To ensure this, a configuration where the laser diode is shorted out, instead of the current source or sink, must be chosen. That way, the laser diode is automatically removed from the circuit in case of an error condition. Choosing between a current sink and a current source is more subtle. If the other shell of the laser

diode is connected to the anode, a current sink can be considered, to keep the can at ground potential. This is not an issue with our laser design though, because the laser diode mount is floating. Another aspect is the electronics side. A current source is typically implemented using p-channel field-effect transistors, while current sinks are using n-channel transistors and additionally the input of a current source is referenced to the positive supply, while the sink is referenced to the negative supply. Using the negative supply as a reference for control signals brings more challenges than vice versa, because typically integrated components like digital-to-analog converters prefer working with positive voltages and would need additional support to be floated to a negative reference. This makes a current source simpler to implement in this scenario and this work focusses on the current source, but in principle all methods that will be discussed can be applied to a current sink as well.

1.8.2. Ideal Current Source

The ideal current source as shown in figure 1.35 has two major properties besides the output current I_{out}, the output impedance R_{out} and the compliance voltage, which are best understood when looking at the two equivalent representations of a current source separately. On the left in figure 1.35a, the Norton representation can be seen. Norton's theorem reduces any linear circuit to a current source, shown in green, with a parallel resistance R_{out}, usually called output resistance or impedance. On the right, the Thévenin representation can be see, which simplifies a circuit as a voltage source, also shown in green, with a series resistance.

(a) Norton representation.
(b) Thévenin representation.
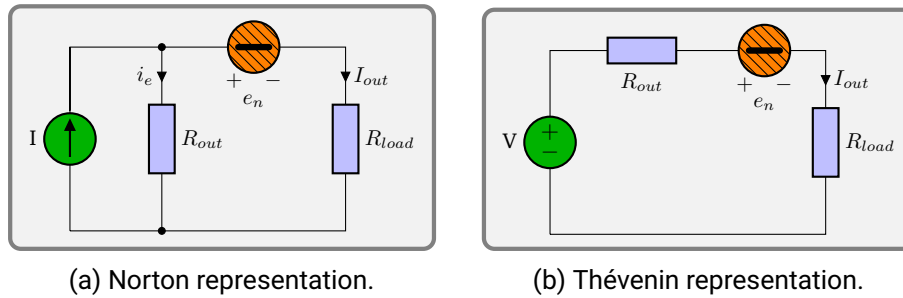
Figure 1.35.: An ideal current source with output impedance R_{out} and noise e_n.

First, the output impedance is discussed. Ideally, R_{out} is infinite and all current is forced to flow through the load. Given a finite output impedance leads to a decreased accuracy of I_{out}, because it is influenced by the load impedance as

$$I_{out} = I_{set} \cdot \frac{R_{out}}{R_{load} + R_{out}} \,. \tag{1.71}$$

In addition to a decreased accuracy, inserting a noise voltage source between the current source and the load as shown in figure 1.35 in orange, has the same effect as a changing load resistance and due to the finite output impedance R_{out}, any voltage noise e_n translates to current noise i_n through the load as

$$i_n = \frac{e_n}{R_{load} + R_{out}} \approx \frac{e_n}{R_{out}} \,, \tag{1.72}$$

again making a high output impedance desirable to supress noise sources between the current source and the load.

Going to figure 1.35b of a current source in Thévenin representation allows to discuss the compliance voltage property. As it was said above, the output impedance of an ideal current source is infinite and so is the maximum output voltage of said current source. A finite output impedance immediately implies a finite supply voltage to keep the current to a finite limit, which dictates a maximum output voltage. This is called the compliance voltage.

1.8.3. The Field-Effect Transistor Current Source

Given the limited supply voltage of a real current source drives the need for a resistive element that has a finite resitance and infinite, or very high, frequency dependent dynamic impedance to react to load changes. One such pass element, having these properties, is a field-effect-transisistor (FET). A junction-gate field-effect transistor (JFET) or metal–oxide–semiconductor field-effect transistor (MOSFET) can be used either as a current source or sink, depending on its doping. A p-channel FET, which uses a positve doping of the channel, is a current source, while an n-channel FET works as a current sink. This discussion is focussing on the p-channel FET with MOSFETs at its center, because it covers the bulk of the laser current driver design in section 1.3.

(a) P-Channel JFET. (b) P-Channel MOSFET.
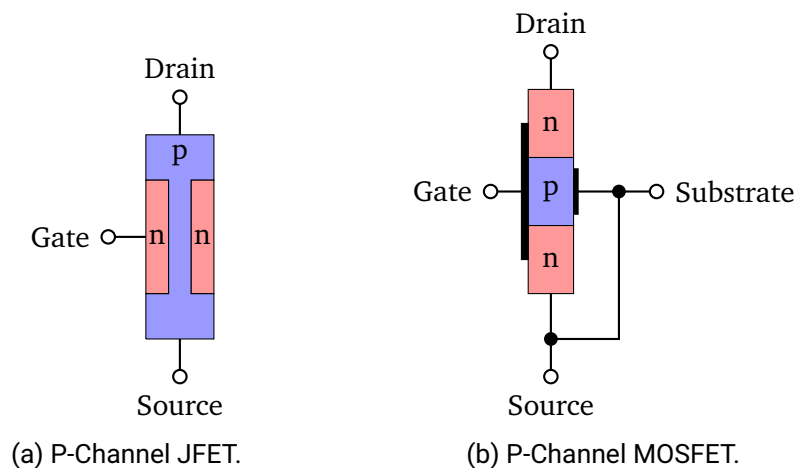
Figure 1.36.: The simplified semiconductor structure of a JFET and a MOSFET.

The difference between a JFET and a MOSFET, is the gate structure as illustrated in figure 1.36. While a MOSFET has an insulated gate, the JFET does not. This reduces the gate leakage current, typically by about three orders of magnitude, and allows to forward bias the device since there is no diode, resulting in larger current handling capacity. So for low currents up to a few mA or low noise applications, JFETS are preferred, while MOSFETs can handle several hundred ampere. The same mathematical approach can be applied to both types of FETs though. The other difference between a JFET and a MOSFET is the fact that JFETs are only available as depletion-mode (normally-on) devices, while MOSFETs are available as both depletion and enhancement (normally-off) devices. The reason is the gate structure as mentioned above. An enhancement-mode device does not conduct, when the gate-to-source voltage $V_{GS} = 0\,\text{V}$, so V_{GS} must be decreased or the junction enhanced for the device to allow conduction. This is not possible with an uninsulated gate like a simple n-p junction of a JFET, which would then start conducting. A p-channel depletion-mode device on the other hand conducts at $V_{GS} = 0\,\text{V}$ and V_{GS} must be increased and the junction depleted to reduce the

current, which is possible with the uninsulated gate, because the n-p junction is reverse biased. The annotated circuit symbol and the quantities used to discuss the device properties are shown in figure 1.37.

(a) P-channel JFET.
(b) P-channel MOSFET.
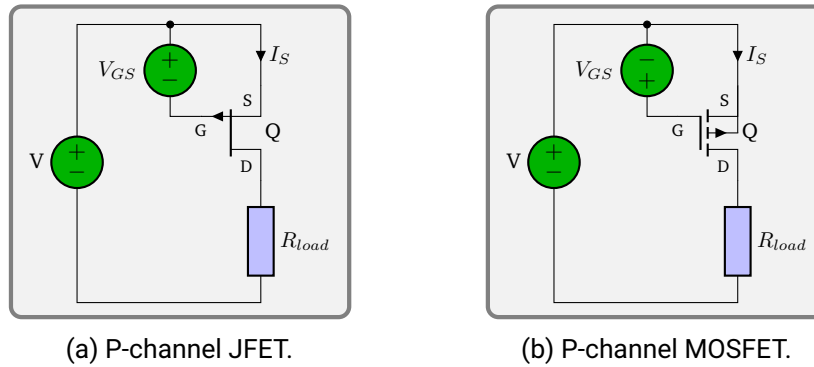
Figure 1.37.: Basic p-channel FET circuit.

A p-channel FET has its source (S) connected to the positive supply and the drain (D) is connected to a more negative voltage, typically the load. For the MOSFET the gate (G) is biased below the source to allow conduction. The source is usually connected to the substrate for solitary devices as shown in figure 1.36b. This will be assumed in all further discussions and the consequences of a substrate, that is biased differently are omitted here. The interested reader may look up these details in [8].

As it was hinted above, if appropriately biased, a FET can be considered a voltage controlled current source. This property can be seen in figure 1.38.
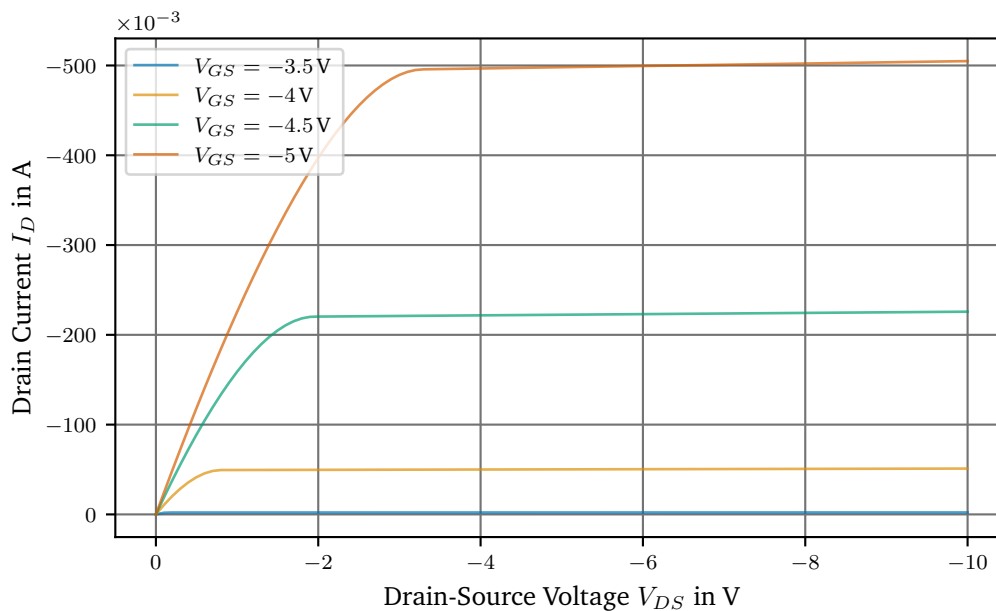
Figure 1.38.: Simulated drain current for different gate bias voltages of an `IRF9610` p-channel MOSFET.

Figure 1.38 shows the current I_D flowing out of the drain of a p-channel MOSFET over

the drain-to-source voltage V_{DS} that is applied accross the FET. For illustrative purposes an example p-channel MOSFET was chosen and its *Simulation Program with Integrated Circuit Emphasis* (SPICE) model [18, 47] was used to generate the data, yet the overall shape is the same for all FETs. For more information on modelling MOSFETs in SPICE, [24, p. 442] can be consulted. There are two regions, the first region, where $V_{DS} > V_{GS} - V_{th}$, demonstrates an almost linear correlation of the channel current and the voltage across the device. This is called the ohmic region, where the MOSFET behaves much like a (gate-) voltage controlled resistor and can be described [106] as

$$I_{D,ohmic} = \underbrace{\kappa(V_{GS} - V_{th})V_{DS}}_{\text{ohmic}} - \underbrace{\frac{1}{2}\kappa V_{DS}^2}_{\text{pinch off}} . \tag{1.73}$$

For small voltages V_{DS} the output current is proportional to the applied voltage V_{DS} accross the channel, just like a normal resistor, giving rise to its name ohmic region. As the voltage increases further I_D starts leveling off, because V_{DS} starts affecting the channel conductivity. The channel is slowly getting pinched off at one end and becomes tapered. The reason is, that the voltage V_{DS} is dropped accross the length of the channel. This voltage drop is linear with V_{DS}, resulting in a $-V_{DS}^2$ dependency of the current, reducing the conductivity of the channel. V_{th} is called the threshold voltage of a MOSFET or pinch-off voltage V_p in case of a JFET and is the voltage at which a current starts flowing.

The parameter κ is a device specific parameter and depends on process parameters and the geometry of the device.

$$\kappa = \kappa' \frac{W}{L} = \mu C_{ox} \frac{W}{L} \tag{1.74}$$

μ is the electron mobility, which is about $1350\,\text{cm}^2/\text{V}$ for n-channel MOSFETs and about $540\,\text{cm}^2/\text{V}$ for p-channel MOSFETs [101]. C_{ox} is the gate-oxide capacitance per unit area and determined by the thickness t_{ox} of the silicon dioxide layer of the gate

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \approx \frac{3.9 \cdot \epsilon_0}{t_{ox}} \approx \frac{3.45 \times 10^{-11}\,\text{F/m}}{t_{ox}} , \tag{1.75}$$

W is the width of the channel, and L is the length of the channel.

The letter κ is used here instead of the usual k as it is used by Sedra et al. [101] to avoid confusion with the Boltzmann constant k_B. Unfortunately, κ is not well controlled [48], because it is not just determined by the size, but also the doping of the material. While the size of the structure can be well controlled to within a few nm using lithography masks, the doping is a matter of temperature and time in a diffusion furnace. The ohmic mode of operation is, for example, used in switches or linear voltage regulators to control the output voltage of the regulator, forming a low impedance voltage source and not the desired current source. This brings up the next region to discuss.

Once the voltage V_{DS} has reached $V_{GS} - V_{th}$, the channel is fully pinched off, any further increase in V_{DS} will not lead to an increase in I_D, in other words the output resistance becomes infinite. The MOSFET is said to be pinched-off or in saturation. In practice there still is a small influence of V_{DS} on the channel. While the depth can no longer decrease as its length is 0 at one end already, the channel will retract a small amount in length with increasing V_{DS}. This is taken into account by the factor λ, called channel-length modulation. The drain current in saturation can now be described [106] as

$$I_{D,sat} = \underbrace{\frac{1}{2}\kappa\left(V_{GS} - V_{th}\right)^2}_{\text{ideal FET}}(1 + \lambda V_{DS}) . \tag{1.76}$$

The parameter λ is the first order Taylor expansion of the length dependence of κ and typically is small and on the order of 0.01–$0.05\,\mathrm{V}^{-1}$ for p-channel MOSFETs [93, p. 23]. It mainly depends on the length of the channel to which it is inversely proportional, since the channel length defines the slope of the tapered channel. Sometimes the value $\frac{1}{\lambda}$ is also referred to as the Early voltage V_A. It is noteworthy, that more modern processes choose a smaller channel length to reduce the on-state resistance of the MOSFET, because the main application of a MOSFET nowadays is as a switch. The reduced channel length makes the MOSFET more susceptible to the channel length modulation effect. This will be discussed in more detail in section 1.8.7, when choosing a suitable MOSFET.

Going back to figure 1.38 the effect of the channel-length modulation can be seen as a small slope of I_D in the saturation region.

Combining the previous equations, the FET drain current behaviour can be summed up as

$$I_D = \begin{cases} 0 & \text{if } V_{GS} - V_{th} < 0 \\ \kappa(V_{GS} - V_{th})V_{DS} - \frac{1}{2}\kappa V_{DS}^2 & \text{if } V_{GS} - V_{th} >= 0 \text{ and } V_{DS} < V_{GS} - V_{th} \\ \frac{1}{2}\kappa\left(V_{GS} - V_{th}\right)^2(1 + \lambda V_{DS}) & \text{if } V_{GS} - V_{th} >= 0 \text{ and } V_{DS} \geq V_{GS} - V_{th} \end{cases} \tag{1.77}$$

The saturation region is the region of interest for building a high output impedance current source, because for a wide range of V_{DS}, the current remains almost constant and can be adjusted using the gate voltage V_{GS}. As a reminder, for the p-channel MOSFET, all voltages are reversed. V_{GS}, V_{th}, V_{DS}, κ and I_D are negative. Some datasheets therefore only give the magnitude of those quantities. The important aspect to remember, is that for the p-channel enhancement-mode MOSFET the gate must be biased negative with respect to the source pin by a least the threshold voltage ($V_{GS} < V_{th}$ or $|V_{GS}| > |V_{th}|$) to turn the transistor on and allow current to flow.

Before proceeding to the precision current source in section 1.8.4, the concept of conductance and transconductance must be explored. The transconductance describes the relationship of the input voltage with the output currrent. The conductance is a measure for how well current flows from input to output. The transconductance g_m and the channel conductance g_{DS} are defined as

$$g_{m,sat} \coloneqq \left.\frac{\partial I_{D,sat}}{\partial V_{GS}}\right|_{V_{DS}=const} = \kappa\left(V_{GS} - V_{th}\right)\left(1 + \lambda V_{DS}\right), \tag{1.78}$$

$$= \sqrt{2\kappa I_D\left(1 + \lambda V_{DS}\right)} \approx \sqrt{2\kappa I_D} \tag{1.79}$$

$$g_{DS,sat} \coloneqq \left.\frac{\partial I_{D,sat}}{\partial V_{DS}}\right|_{V_{GS}=const} = \frac{1}{2}\kappa\left(V_{GS} - V_{th}\right)^2\lambda \tag{1.80}$$

$$= \frac{I_D}{\frac{1}{\lambda} + V_{DS}} = \frac{1}{R_o} \approx I_D\lambda\,. \tag{1.81}$$

The transconductance g_m, as a measure of the current gain with respect to the gate-source voltage of the MOSFET, is proportional to the square root of the drain current I_D. The inverse of the channel conductance g_{DS} is called output resistance R_o and discussed below. Typically the V_{DS} term in the denominator of the output resistance in equation 1.81 can be neglected.

The meaning of g_m and g_{GS} can be best understood, when looking at a mathematical model of the MOSFET. These models come in varying complexity and either as a large-signal or small-signal model. Only the latter is used here. The small-signal model, is a first-order Taylor approximation around the working point, for a constant gate-source voltage V_{GS} and constant

drain-source V_{DS}, hence both g_m and g_{GS} are constants.

$$I_D \approx \frac{\partial I_D}{\partial V_{GS}}\Delta V_{GS} + \frac{\partial I_D}{\partial V_{DS}}\Delta V_{DS} \tag{1.82}$$

$$= g_m\Delta V_{GS} + g_{DS}\Delta V_{DS} \tag{1.83}$$

$$= g_m v_{GS} + \frac{1}{R_o}v_{DS} = i_D \tag{1.84}$$

The lower case letters denote the variables of the small-signal model as they only change very little compared to the working point parameters. From 1.84 it can be seen, that the g_{DS} term adds to the output current and is proportional to v_{DS}. Comparing with figure 1.35a, this the proportionality constant can be identified as $\frac{1}{R_o}$ like proposed above. Just like the ideal current source in figure 1.35, the model can be given in the Norton or Thévenin representation both shown in figure 1.39.

(a) Small signal model of a saturated MOSFET including the output resistance. The output resistance models the channel-length modulation as given by equation 1.84.

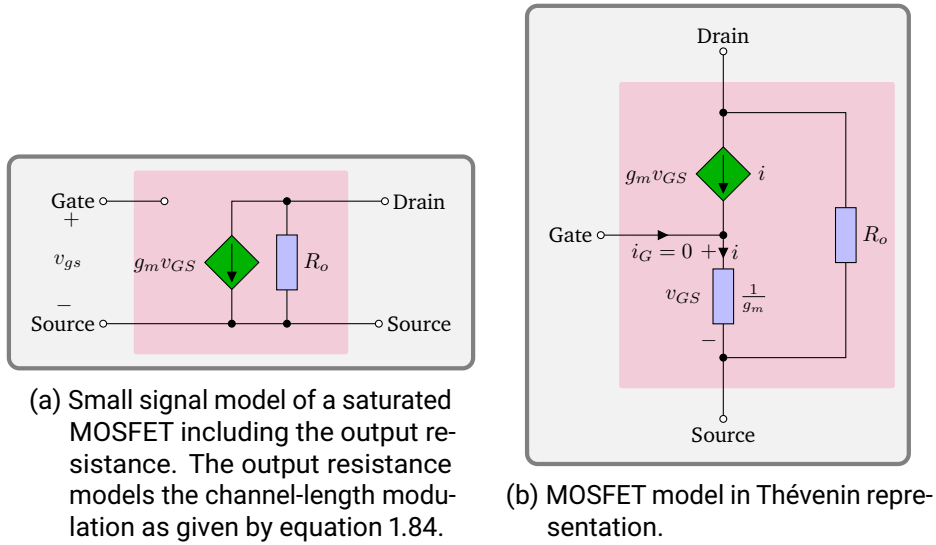(b) MOSFET model in Thévenin representation.

Figure 1.39.: Equivalent MOSFET models in Norton and Thévenin representations.

A detailed graphic derivation of the Thévenin representation can be found in [101]. The Thévenin representation will prove especially valuable, when treating circuits with a resistance in the source leg. The small-signal model now shows, that the output impedance is dependent on the channel-length modulation λ and v_{DS}. Typically, $\frac{1}{\lambda} \gg v_{DS}$, so λ is the most important factor governing the output impedance of a MOSFET.

To give an example of the output impedance of a MOSFET, parameters were taken from the aforementioned SPICE model of the IRF9610. Do note, that these parameters of a model are tuned to match certain operating conditions by their creators and only present an estimation of the real MOSFET. Using the example parameters from table 1.9, $I_D = 250\,\text{mA}$, $\lambda = 4\,\text{mV}^{-1}$, $V_{DS} = 3.5\,\text{V}$ equation 1.81 yields

$$R_{out} = R_o\left(I_D = 250\,\text{mA}, \lambda = 4\,\text{mV}^{-1}\right) = 1014\,\Omega \overset{V_{DS}=0}{\approx} 1\,\text{k}\Omega\,, \tag{1.85}$$

which is not very convincing as a current source. The insignificant impact of V_{DS} on the output impedance can be seen when dropping the V_{DS} term, which leads to an output impedance of $1\,\text{k}\Omega$. Usually, in textbooks, this dependence is therefore neglected. To improve R_{out}, the focus thus lies on the λ dependence. The model derived from equation 1.84 can be used to do so, leading to the precision current source presented next.

1.8.4. Precision Current Source

In the previous section 1.8.3 it was shown in equation 1.84, that the output impedance of a MOSFET depends on the channel-length modulation λ and is too low for practical purposes. On the quest to improve the output impedance of the MOSFET circuit in figure 1.39a, the most obvious solution would be to simply add a source resistor R_S into the circuit as shown in in figure 1.40a. At first glance, this may seem to only add a series resistance to R_o, but the attempt is more interesting and will lead to an even better solution.

(a) P-channel MOSFET with source resistor R_S to improve the output impedance R_{out}.

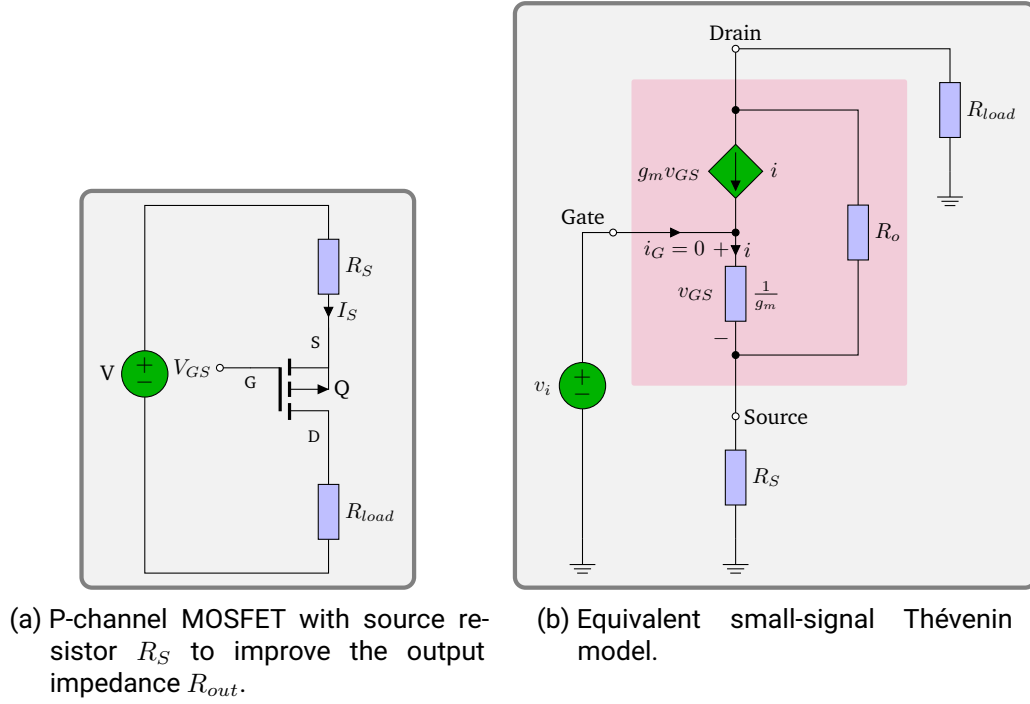(b) Equivalent small-signal Thévenin model.

Figure 1.40.: Circuit of a MOSFET with source degeneration resistor and equivalent Thévenin model.

Before calculating the output impedance, we shall have a look at v_{GS} and the input signal v_i derived from it. With the introduction of the source resistor R_S, v_i no longer equals v_{GS}, because $\frac{1}{g_m}$ now forms a voltage divider with R_S and it follows

$$v_{GS} = v_i \frac{\frac{1}{g_m}}{R_S + \frac{1}{g_m}} = v_i \frac{1}{1 + g_m R_S} \ . \tag{1.86}$$

This implies a reduction in gain, by the factor $\frac{1}{1+R_S g_m}$ compared to the previously discussed approach. The cause of this reduction is negative feedback. To understand this, imagine, that with a constant v_i and hence a constant current I_D flowing, a changing load resistance is trying to modulate I_D. Any increase in I_D will cause the voltage across R_S to rise, reducing v_{GS}, because v_i is still constant. The decreasing v_{GS} will then reduce I_D, thus introducing negative feedback. Having realized there is negative feedback present, it can be postulated, that the reduction in input sensitivity, or effective transconductance, will be passed on to the output impedance. This very interesting relationship will now be derived.

To calculate the output impedance, figure 1.40b can be simplified by grounding v_i, because there is no AC component as there is no current flowing through the insulated MOSFET gate

and is not modulated. The load R_{load} resistance must is replaced by an AC test voltage v_{load} to modulate I_D. These changes result in the small signal model shown in figure 1.41. As a sidenote, this configuration is also called a common-gate amplifier.
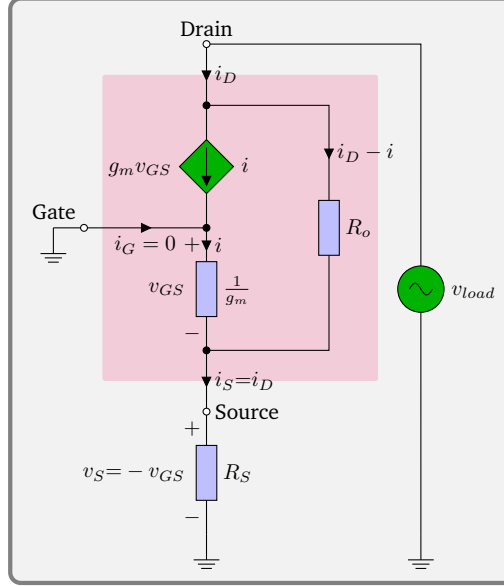

Figure 1.41.: Small signal model of the common-gate amplifier with source resistance R_S.

The (dynamic) output impedance is given by

$$R_{out,cg} = \frac{v_{load}}{i_D} \, , \tag{1.87}$$

with $i_D = i_S$, since there is no gate current. v_{load} can easily be calculated by looking at figure 1.41 and is the total voltage across R_o and R_S. v_{GS} can also be found, because the gate is grounded. With the resistance $\frac{1}{g_m}$ at one end, the voltage at the source pin must be $-v_{GS}$.

$$
\begin{aligned}
v_{load} &= (i_D - i) \, R_o + i_S R_S \\
&= (i_D - g_m v_{gs}) \, R_o + i_D R_S \\
&= (i_D + g_m i_D R_S) \, R_o + i_D R_S
\end{aligned}
\tag{1.88}
$$

Using equations 1.87 and 1.88 gives

$$R_{out,cg} = (1 + g_m R_S) \, R_o + R_S \tag{1.89}$$

for the output impedance.

This result is interesting, as it can be be immediately seen, that the output impedance scales very quickly with the transconductance g_m and R_S. As it was already speculated above, the reduction in the transconductance $\frac{1}{1+g_m R_S}$ of the MOSFET is transfered to the output impedance, which is increasing by the inverse of the loss in transconductance.

Going back to the quest for an increased output impedance, it is appararent, that increasing R_S quickly raises the output impedance, as it scales with $gm_m R_o$, but it would come at the cost of a significantly reduced compliance voltage. So, other means need to be explored. As we have seen, the scale factor $gm_m R_o$ is explained by feedback and this brings up another solution.
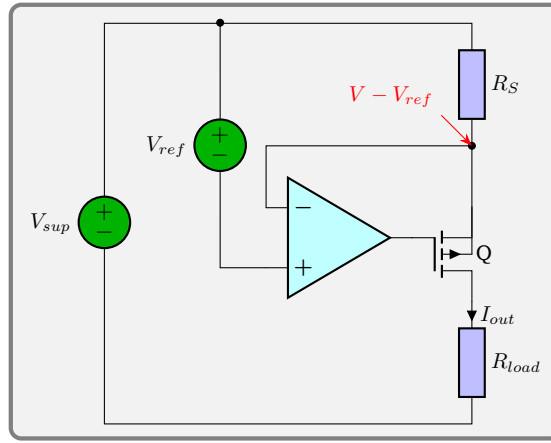
Figure 1.42.: Transconductance amplifier with a p-channel MOSFET.

The amount of feedback can be increased further using an operational amplifier (op-amp) as shown in figure 1.42.

The output impedance of this transconductance amplifier is amplified by the open-loop gain of the op-amp as shown in appendix A.3, while the transfer function greatly simplified and found to be

$$R_{out} \approx A_{ol} \left(g_m R_o R_S + R_o + R_S \right)$$
$$I_{out} \approx \frac{V_{ref}}{R_S} \tag{1.90}$$

In addition to the increased output impedance, the current $I_D = I_{out}$ can now steered by adjusting V_{ref} and is, given sufficient loop gain of the op-amp, no longer dependent on the MOSFET, but rather only on the sense resistor R_S.

This has the added benefit, that it is possible to leverage the tight accuracy and precision of a resistor, over the poor specifications of a MOSFET. Resistors can be manufactured with tolerances of less than $100\,\mu\Omega/\Omega$, which is orders of magnitude better than FETs, which can be matched to low % values with patience.

Using the example parameters from table 1.9, the output impedance in staturation can now be calculated again for $I_{out} = 250\,\mathrm{mA}$ and the ideal `IRF9610` model with the addition of an idealized `AD797` op-amp using the worst-case specifications.

$$R_{out} \approx 2\,\mathrm{V\,\mu V^{-1}} \left(0.64\,\mathrm{S} \cdot 1014\,\Omega \cdot 30\,\Omega + 1014\,\Omega + 30\,\Omega \right) \approx 40\,\mathrm{G\Omega} \tag{1.91}$$

From these consideration, it can be seen, that the open-loop gain and the unity-gain band-width of the op-amp essentially determine the properties of the current source, given that $R_{id} \gg R_S$ and $R_o \gg R_S$. This will be important for selecting an operational amplifier later.

The next section will focus on the MOSFET and discuss the compliance voltage of the current source, which was only briefly touched during the introdution. It will give rise to criteria for selecting a MOSFET for the precision current source.

1.8.5. Compliance Voltage

The compliance voltage of a current source is the maximum voltage it can output to maintain the requested output current. For an ideal current source, the compliance voltage is infinite, but it is obviously limited in the physical world.

The precision current source discussed in section 1.8.4 has several limiting factors of the compliance voltage, which shall be discussed now. The compliance voltage is taxed most at the maximum output current $I_{out,max}$. So for the following discussion, the output is always treated as set to maximum.

Looking at figure 1.42 of the precision current source it is immediately evident, that the output voltage can be calculated by subtracting the voltage accross the source resistor V_{R_S} and the MOSFET V_{DS} from the supply voltage V_{sup}

$$V_{out} = V_{sup} - V_{R_S} - V_{DS} = V_{sup} - V_{ref} - V_{DS}\,.$$

The voltage V_{R_S} is given by equation 1.90 and equal to the setpoint voltage and hence given by the system parameters. This leads to the question of the minimum working point voltage V_{DS} at $I_{out,max}$. As a reminder, from equation 1.77 and figure 1.38 one can see, that the drain current is almost constant over V_{DS} in the saturation region and in the ohmic region is proportional to V_{DS}. The transition point from the ohmic region to the saturation region is at $V_{DS} = V_{GS} - V_{th}$ and putting this into equation 1.77 yields for the drain current

$$I_D = \frac{1}{2}\kappa V_{DS}^2 \left(1 + \lambda V_{DS}\right)$$

$$\Rightarrow V_{DS} \approx \sqrt{\frac{2I_D}{\kappa}} \tag{1.92}$$

$$\approx 784\,\mathrm{mV} \tag{1.93}$$

The latter result was calculated using the example parameters from table 1.9. At this point it can already be postulated, that the MOSFET will severly change in its function as a current source for $V_{DS} < 0.78\,\mathrm{V}$. To quantify this, one has to look at the output impedance of the transconductance amplifier once again. In the last section, the output impedance was only treated for the staturation region, but this time, R_{out} must must be considered over a wide range of V_{DS}, thus not only in the saturation region, but also in the ohmic region. Instead of using the small-signal model as before, which assumed only small changes of V_{DS}, a large-signal model must applied, which also includes the non-linear nature of the piece-wise defined equation 1.77 of the drain current.

For the sake of simplicity, a SPICE simulation of figure 1.42 was carried out in LTSpice [81]. Solving this analytically, bears no educational value over the numerical solution shown below as will be seen. Additionally, the SPICE simulation also offers the opportunity to add additional, parasitic elements to the model to evaluate their effect, for example, the capacitive nature of the MOSFET gate.

The simulation itself is numerically fairly challenging and the typical approaches will lead to the limits of the numerical precision. To make the simulation feasable, the large-signal model is broken down into several small segments. For each these segments, the small-signal models at its respective working point is evaluated and then the result joined back together to reconstruct the large-signal model sought. How this is done in detail, is shown in appendix A.4 as it is beyond the scope of this section. The final result was calculated for two different frequencies, one frequency was deliberately chosen so low (1 μHz), that it is well below the dominant pole

of the op-amp, meaning, that the full open-loop gain applies and the other frequency chosen was $1\,\text{MHz}$, were the gain had dropped to $10\,\text{V/V}$. This is shown in figure 1.43.
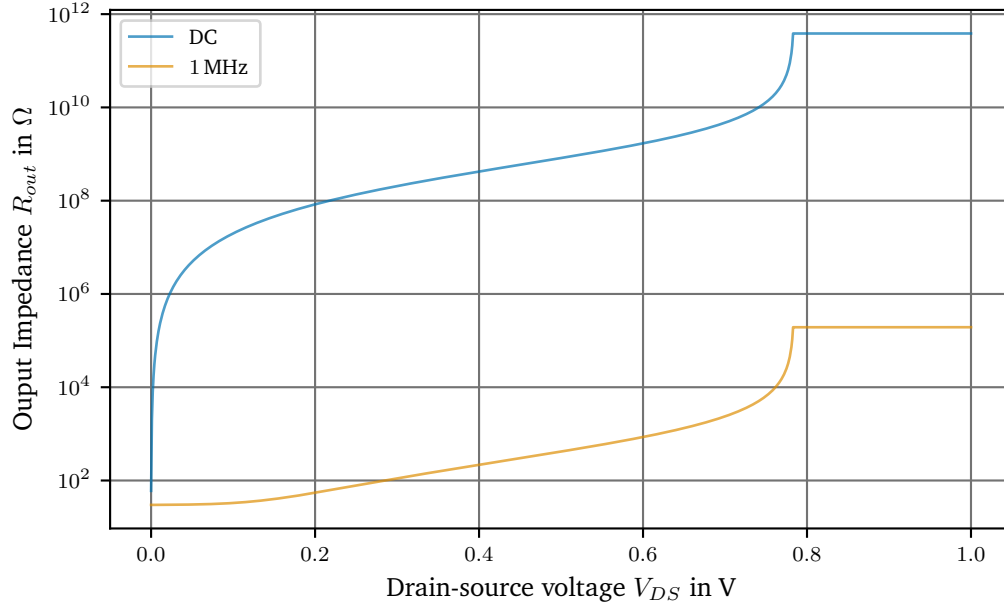

Figure 1.43.: Simulated output impedance for the precision current source from figure 1.42 at DC and $1\,\text{MHz}$ over the drain-source voltage.

Looking at figure 1.43 clearly shows the effect of entering the ohmic region of the MOSFET. Over a range of about $100\,\text{mV}$ below the $0.78\,\text{V}$ calculated above, the output impedance drops by two orders of magnitude and then keeps dropping at an exponential rate with decreasing V_{DS}. The same effect applies to the output impedance at $1\,\text{MHz}$, although the starting value is around $200\,\text{k}\Omega$ due to the reduced gain from the op-amp at $1\,\text{MHz}$. It can also be seen, that R_{out} levels off at $30\,\Omega$, the value of the sense resistor.

This overall effect of leaving the saturation region is so drastic, that the compliance voltage must be defined in such a way, that the MOSFET remains in saturation and this leads to

$$V_{comp} = V_{sup} - V_{ref} - \sqrt{\frac{2I_D}{\kappa}}.$$ (1.94)

Now turning to the supply voltage, it is limited by the op-amp which must drive the gate of the MOSFET all the way up to the supply to turn off the current source. The reference voltage is, unless one divides it down, which is a delicate matter, dictated by the reference chosen. This, unfortunately, leaves only little room for the MOSFET and it must be carefully chosen not limit the compliance voltage too much.

At this point a fallacy, the author has observed multiple times must addressed. In order to address the limited compliance voltage, one may be tempted to use multiple MOSFETs in parallel to divide the current between the MOSFETs and thereby reduce the voltage that needs to be dropped across the FET proportional to $\frac{1}{\sqrt{N}}$, where N is the number of MOSFETs paralleled.

Imagine the following modified circuit of the precision current source shown in figure 1.44 with two MOSFETs in parallel. For clarity the gate resistors required are not included.
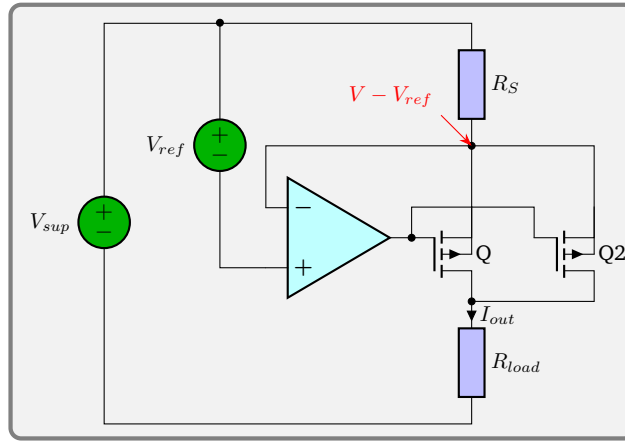
Figure 1.44.: Transconductance amplifier with two p-channel MOSFETs in parallel.

While at first, this seems like a solution to the limited V_{DS}, it is not that easy and a bad idea for number of reasons given here. The first reason is, MOSFET specifactions are very loose, notably the threshold voltage V_{th}, the transconductance g_m and the capacitances, but the latter is of little concern here. Paralleling MOSFETs works well under certain conditions, when using the MOSFETs as a switch, not as a current source. It seems to be a common misunderstanding, that MOSFETs are imune to thermal runaway. This is true, when using them as a switch, fully turned on and in the ohmic region. In this case, there a two effects occuring, the first is, that the (absolute) value of V_{th} decreases with temperature, thus increasing I_D and the second effect is, $R_{DS,on}$ is rising with temperature [19]. But here, the MOSFET is operating in pinch-off and not the ohmic region, $R_{DS,on}$ has no influence on the current, therefore, the only effect at work is the decreasing V_{th}, so depending on how bad the imbalance in V_{th} of the paralled MOSFETs is, one MOSFET will gobble up most of the current and power. Adding source resistors, can compensate for this by pushing down the source voltage as the current goes up. This will then reduce V_{GS}. The size of the resistor depends on the transconductance g_m and the temperature coefficent of V_{GS}, which is around 1.5–$2\,\mathrm{mV/K}$ [40]. Unfortunately, $1\,\Omega$ or $2\,\Omega$, will already eat up, most of the benefits gained in compliance voltage as will be shown below. A detailed analyis of paralleling MOSFETs can be found in [39].

The second reason, why paralleling MOSFETs is a bad idea can be found, when remembering equation 1.77. We know that the transition from the undesirable ohmic region to the saturation region is

$$V_{DS} \geq V_{GS} - V_{th} \tag{1.95}$$

Looking at 1.44, we see that V_{GS} is set by the op-amp and is the same for both MOSFETs, but V_{th} is device specific and according to the datasheet of our example IRF9610 [51] V_{th} values can show a spread of as much as -2 to $-4\,\mathrm{V}$, although [85] suggests, that MOSFETs from the same reel show a spread of only $\pm 125\,\mathrm{mV}$ of V_{th} within the same batch for consecutive devices. The $125\,\mathrm{mV}$ was found for the BUK7S1R5-40H [20], which was sampled in this report. The number given in the report is for 3σ and assuming the datasheet values are also referring to 3σ, the value found in the report is about twice as good as the datasheet value of 2.4–$3.6\,\mathrm{V}$. Assuming similar numbers for IRF9610 MOSFET used in our examples, this leads to $\pm 208\,\mathrm{mV}$ for the IRF9610, again applying 3σ. Using this number, a Monte Carlo simulation (not quite, because the dice were biased to yield a Gaussian distribution) was run using LTSpice, simulating the circuit shown in figure 1.44 and also the original circuit using only one MOSFET. The

current source was set to $250\,\text{mA}$ as per table 1.9. The load voltage was set to

$$V_{DS,parallel} = \sqrt{\frac{2\frac{I_d}{2}}{\kappa}} \approx 555\,\text{mV}\,, V_{DS,single} = \sqrt{\frac{2I_d}{\kappa}} \approx 784\,\text{mV}\,.$$

$\frac{I_D}{2}$ was used for the parallel configuration to show the effect assuming perfect current sharing between the MOSFETs. Additionally a configuration with an increased safety margin of $1\sigma = 70\,\text{mV}$ added to $V_{DS,parallel}$ was investigated. 4000 samples were drawn and the spread of the output impedance was calculated for each circuit. The results are shown as a histogram in figure 1.45. The counts give the number of cases for each bin of the output imdedance.
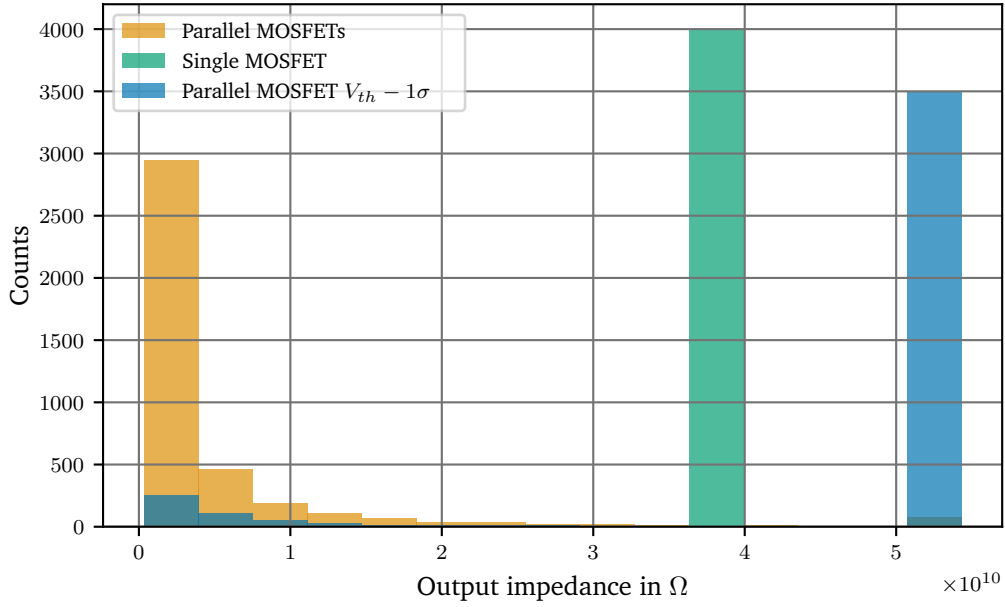

Figure 1.45.: Results of a Monte Carlo simulation of the output impedance for different configurations of MOSFETs.

Unsurprisingly, there is no variance of the output impedance in the single MOSFET case in accordance to what we have learnt in appendix A.3. The op-amp gain simply supresses all device properties of the MOSFET. The slight variation of g_m for different samples was not simulated, because this variation stems from the variation of κ and goes as $\frac{1}{\sqrt{\kappa}}$, so its effect is not as pronounced as the threshold.

In case of two mosfets, the output impedance varies over an order of magnitude from about 1.8–$52\,\text{G}\Omega$. Even when increasing the drain-source voltage by $1\sigma = 70\,\text{mV}$ to $625\,\text{mV}$ on average, the spread is still an order of magnitude. Only when increasing V_{DS} to around $700\,\text{mV}$, the situation stabilizes, but then the net gain from this measure has shrunk to a meager $84\,\text{mV}$. We can see from this simuation, that the system-to-system spread becomes very unstable in tough situations. This instability can also be brought into the system by temperature effects as V_{th} is temperature dependent as discussed above. This is a designers nightmare, because the product is no longer interchangeable in situations of high load currents and load impedances. Additionally, it may suffer from thermal runaway unless each individual MOSFET is laid out to carry the full current. As a final remark: Do not parallel MOSFETs in saturation, ever.

1.8.6. Noise Sources

The fundamentals of different types of noise were already introduced in section 1.6. Here, a subset of these noise types is treated. It is expected, that the dominant noise observed in this circuit is $\frac{1}{f}$-noise at low frequencies and white wideband-noise. All noise components will be converted to the so-called input referred notation to make the noise sources comparable. This can be easily understood, when looking at two amplifiers with different gain. If both of them add a fixed amount of noise to the output signal, the absolute amount of noise may be the same, but the signal to noise ratio shows a different picture. To compare these amplifiers it is useful to divide the noise by the transfer function (gain) of the amplifier. This is called input-referred noise, since it treats the noise in relation to the input signal. Additionally, when calculating noise figures, the noise bandwidth is always considered to be $1\,\mathrm{Hz}$.
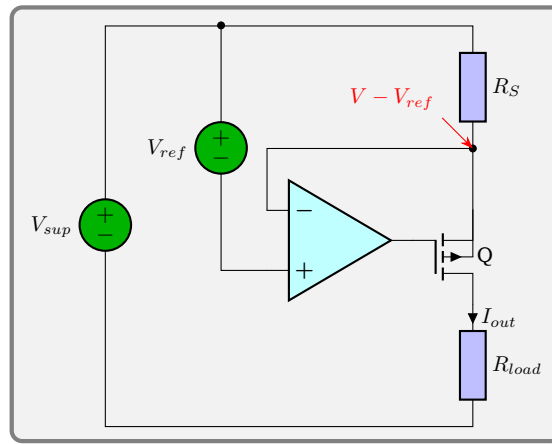

Figure 1.46.: Transconductance amplifier with a p-channel MOSFET. Repeated from page 62.

Noise sources are ubiquitous in the circuit in figure 1.42 on page 62, repeated here as figure 1.46 for clarity. The resistor R_S, the MOSFET, the op-amp, the setpoint voltage V_{ref} and the supply voltage V_{sup} can all contribute noise to the output current. Fortunately, some of those noise contributions are either very small or are well suppressed in this design, so each component must be briefly discussed to see if this is the case.

Starting with the supply voltage V_{sup}, it can be seen, that any change of this voltage affects the string R_S-Q-R_{load}. From equation A.8, we know that if the op-amp gain is high (true within the bandwidth of the op-amp) all disturbances of the voltage accross R_S will be suppressed and the output current is only defined by the reference input and R_S. Looking closer, the supply noise is present at the inverting and non-inverting input of the op-amp with the same magnitude. If there is no current flowing into the op-amp pins, which is true for low frequecies, the noise is affecting both pins equally and it will be suppressed by the common-mode-rejection ratio (CMRR) of the device. Fortunately, this is a strong quality of precision op-amps and values of more than $1\,\mathrm{\mu V\,V^{-1}}$ are not uncommon. The op-amp will therefore take care of the supply noise at low frequencies. At high frequencies the parasitic capacitance of the input pins and the reduced gain and CMRR come into play, reducing the CMRR and the gain also drops going to high frequencies. To take care of this, it is therefore prudent to filter the supply for high frequency noise.

The next noise source is the reference voltage. The reference is directly connected to the input and its noise dictates most of the circuit noise. While the high-frequency noise can again be filtered to some extend, the low frequency noise, which is mostly $\frac{1}{f}$-noise can not be filtered

as was shown in section 1.6.1, so it must be kept low from the start and the reference selected for low flicker noise.

The MOSFET as a noise source is considered in appendix A.5 and the interested reader may find the derivation of the MOSFET noise component there. The two types of noise that need to be considered are the flicker noise of the MOSFET and its wideband thermal noise as calculated in equation A.20

$$i_n = \sqrt{\underbrace{4k_B T \frac{2}{3} g_m}_{\text{thermal}} + \underbrace{\frac{K_f I_D}{C_{ox} L^2} \frac{1}{f}}_{\text{flicker}}} \, .$$

To calculate the input referred noise, which shows that the MOSFET noise will be suppressed by the op-amp, the current noise needs to be divided by the open-loop gain derived as equation A.7

$$e_{n,FET} = \frac{i_n}{A_f} = \frac{\sqrt{4k_B T \frac{2}{3} g_m + \frac{K_f I_D}{C_{ox} L^2} \frac{1}{f}}}{\frac{A_{op}}{R_S} \frac{g_m(R_o||R_S||R_{id})}{g_m(R_o||R_S||R_{id})+1}} \, . \tag{1.96}$$

Looking at the parameters from table 1.9, we find $(R_o||R_S||R_{id}) \approx R_S$ and e_n can be simplified to

$$e_{n,FET} \approx \frac{\sqrt{4k_B T \frac{2}{3} g_m + \frac{K_f I_D}{C_{ox} L^2} \frac{1}{f}}}{A_1 \frac{1}{R_S + \frac{1}{g_m}}}$$

$$\approx \frac{R_S + \frac{1}{g_m}}{A_1} \sqrt{4k_B T \frac{2}{3} g_m + \frac{K_f I_D}{C_{ox} L^2} \frac{1}{f}}$$

$$\overset{A_1 \to \infty}{=} 0$$

Unless the MOSFET transconductance g_m or the gain of the op-amp A_1 become very small, the noise of the MOSFET is very well suppressed. This means, that if the wideband thermal noise contribution is small (it is, see A.5) and the flicker noise corner frequency is within the bandwdith of the op-amp, the noise contribution from the MOSFET can be neglected.

The noise contribution from the sense resistor R_S is the (approximated) Johnson–Nyquist noise, which when transformed to its Norton representation can be written as a current noise

$$i_{n,R} = \sqrt{\frac{4k_B T}{R_S}} \, . \tag{1.97}$$

Additionally, it was shown, that depending on the material of the resistive element, a flicker noise component can also be present. This is especially prevalent in carbon and tick-film resistors [37, 86]. While thin-film resistors are less noisy, their performance varies greatly between different models [102], so their make and model must be carefully selected for the application. Foil and wirewound resistors were shown to perform best and have almost no flicker noise [17, 102]. Using a high quality resistor, the flicker noise can be neglected and only the thermal noise must be taken into account.

The sense resistor is part of the feedback network and therefore it contributes fully to the noise of the transimpedance amplifier. Input referred, the current noise must be divided by the closed-loop gain A_f given by A.2.

$$e_{n,R} = i_{n,R} \cdot \beta \approx i_n \cdot R_S = \sqrt{4k_B T R_S} \tag{1.98}$$

The final component to be discussed is the operational amplifier. Although the op-amp is a rather complex device, its noise can be modeled by a small number of noise sources. This noise model of the op-amp is shown in figure 1.47.
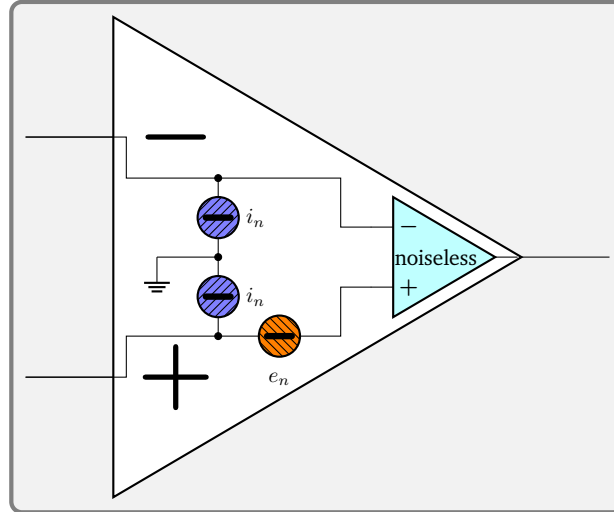
Figure 1.47.: Noise model of the operational amplifier.

In figure 1.47 we can see, that there are three noise sources required to treat the op-amp. The input voltage noise e_n and two input current noise sources i_n. The current noise noise sources i_n are assumed to be mostly uncorrelated. This assumption will lead to an upper bound as can be seen from figure 1.48, which shows the the input differential amplifier, that is the first stage of a typical bipolar op-amp.
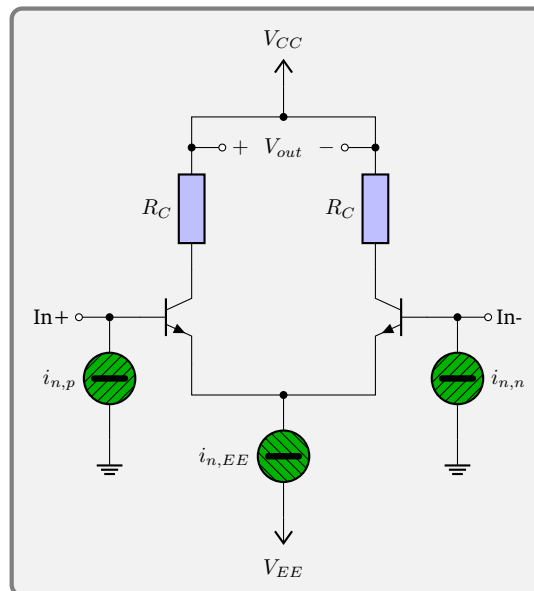
Figure 1.48.: Bipolar op-amp input stage with noise sources.

Of the three noise sources $i_{n,p}$ and $i_{n,n}$ are uncorrelated, because it is the input bias current of the individual transistors, and only the effect of $i_{n,EE}$ is correlated, because the current of

the emitter bias current souce is equally distributed between the two input transistors. Since effects of equal magnitude and sign cancel out due to the differential nature of the input stage, correlated effects are suppressed. An equal magnitude can be assumed, because the gain of the two transistors is well matched, due to their close proximity on the semiconductor die. Therefore assuming all noise is uncorrelated presents an upper bound. A more detailed analysis can be found in [43], if interested. Due to the matching of the transistors, the magnitude $i_{n,p}$ and $i_{n,n}$ are also closely matched, hence in our model, they are assumed equal and only called i_n.

As we have done before in this section by referring all noise sources to the input, the same was done by the manufacturer and the noise figures given in the datasheet are typically input referred and combine all noise sources in the op-amp. The two current noise sources can not be combined with the input voltage noise, because they depend on the external impedances connected to the op-amp, so this final step must be done by the circuit designer. Given the complete circuit as in figure 1.46 it is possible to calculate the full noise contribution of the op-amp as

$$e_{n,op} = \sqrt{e_n^2 + e_{n,+}^2 + e_{n,-}^2}\,, \tag{1.99}$$

assuming the noise sources are uncorrelated. The input referred noise $e_{n,-}$ of the inverting input can be calculated in a similar fashion as e_{n_R} in equation 1.98. It is likewise part of the feedback network and must therefore be divided by the closed-loop gain A_f as before.

$$e_{n,-} \approx i_n \cdot R_S \tag{1.100}$$

The current noise of the non-inverting input can be translated by looking at its input impedance. This is be determined by the filter of the reference voltage, which is required to remove the high frequency noise as discussed above. Assuming an RC-filter of first order, the output impedance can be calculated from the transfer function of the low-pass filter, derived in equation 1.17

$$R_{out,filt} = R_{filt} \cdot A = \frac{R_{filt}}{1 + sR_{filt}C}$$

$$\lim_{s \to 0} R_{out,filt} = R_{filt}$$

$$\lim_{s \to \infty} R_{out,filt} = 0\,.$$

$$e_{n,-} \approx \frac{i_n R_{filt}}{1 + sR_{filt}C} \tag{1.101}$$

Looking at the output impedance of the filter, it can be seen, that for high frequecies, the output impedance goes to 0, while for low frequencies it is R_{filt}. If the filter corner frequency $\omega_0 = \frac{1}{RC}$ is close to or at the flicker noise corner frequency of the reference voltage, it means that there is almost no wideband current noise contribution as well. Only the $\frac{1}{f}$ component of the op-amp current noise multiplied with R_{filt} is left. This should be lower than the reference noise to have negligible impact and must be kept in mind when selecting an op-amp. This leads to the total noise of the op amp

$$e_{n,op} = \sqrt{e_n^2 + (i_n R_S)^2 + \left| \frac{i_n R_{filt}}{1 + sRC} \right|^2}\,. \tag{1.102}$$

To conclude, table 1.5 is given as a reference for the noise contributions in the low-frequency and also the wideband domain. From this table, it can be seen, that the only wideband-noise

contributors are the reference resistor and the op-amp. The low-frequency contributors are the voltage reference and the op-amp, since they have a strong flicker noise component. A low-noise, precision op-amp typically has far less low frequency noise than a voltage reference and the dominant low frequency contributor remains the voltage reference.

Noise component	Low frequency	Wideband
V_{sup}	≈ 0	≈ 0
MOSFET	≈ 0	≈ 0
V_{ref}	$\sqrt{e_{n,ref}^2 + 4k_BTR_{filt}}$	≈ 0
R_S	$\sqrt{4k_BTR_S}$	$\sqrt{4k_BTR_S}$
Op-amp	$\sqrt{e_n^2 + i_n^2(R_S^2 + R_{filt}^2)}$	$\sqrt{e_n^2 + i_n^2R_S^2}$

Table 1.5.: Input referred noise components of the transimpedance amplifier. Multiply by $\frac{1}{R_S}$ to get the output referred current noise.

To summarize the findings, the most important choices regarding the noise contributions are a good quality metal-foil or wirewound sense resistor R_s, a low noise voltage reference and a low noise op-amp. Regarding the low-noise op-amp it is important to decide between low voltage noise or low current noise. This choice depends on the value of R_S. For typical values of R_S below $1\,\text{k}\Omega$ voltage noise is dominating. The reference should be chosen for low flicker noise.

1.8.7. Component Selection

This section deals with selecting the right components for the precision current source presented in section 1.8.4. The focus lies on the requirements defined in section 1.3, notably tables 1.3.1 and 1.3.2. Most attention will be on the MOSFET, the operational amplifier and the voltage reference. First is the voltage reference, because this will define several parameters down the road. Then, the op-amp is discussed, for which several examples from scientific publications and other alternatives are shown and the best solution is presented. Finally, the selection parameters for the MOSFET will be elaborated. The reader must warned though, that the lineup of p-channel MOSFETs in production is decreasing, with more and more products being discontinued in favor of n-channel MOSFETs and the examples may be outdated.

Numerous laser driver designs can be found in literature [35, 46, 68, 99, 114, 115, 118]. While Libbrecht et al. where not the first to present a design and a similar design can already be found in [21], their design stands out for its simplicity. The designs in literature can be divided into two groups. High power drivers for quantum cascade lasers (QCL) typically featuring a compliance voltage of more than $10\,\mathrm{V}$ and output currents of up to serveral ampere based on the work of Taubman and medium power devices for laser diodes having a lower compliance voltage of around $2\,\mathrm{V}$ and capable of driving a few hundred mA based on the work of Libbrecht et al. The requirements of this work mostly fall into the latter category, except for the compliance voltage, which is targeted to be $\geq 8\,\mathrm{V}$. All these drivers share one common aspect, though, the type of voltage reference. Most laser drivers in literature and commercial products are designed around low-noise, low-drift buried Zener diode voltage references, namely the LM399 [70] or LTZ1000 [72].

Component	Voltage	Temperature coefficent	Stability	Package
LT1021	7 V	2–5 μV/(V K)	15 μV/(V $\sqrt{\mathrm{kh}}$)	SO-8
LT1027	5 V	1–2 μV/(V K)	not specified	SO-8
LM399	7 V	0.3–1 μV/(V K)	8 μV/(V $\sqrt{\mathrm{kh}}$)	TO-46
ADR1399	7 V	0.2–1 μV/(V K)	7 μV/(V $\sqrt{\mathrm{kh}}$)	TO-46
LTZ1000	7.2 V	0.05 μV/(V K)	0.3 μV/(V $\sqrt{\mathrm{kh}}$)	TO-99
ADR1000	6.6 V	<0.2 μV/(V K)	0.2 μV/(V $\sqrt{\mathrm{kh}}$)	TO-99

Table 1.6.: List of buried Zener diodes and selected properties.

The buried types of voltage references are *Zener* diodes, that are created within the bulk silicon using ion implantation. This reduces noise due to surface contamination [104]. These diodes are not true Zener diodes, but called Zeners nonetheless and use a mix of Zener and avalanche breakdown to compensate the temperature coefficient. The Zener effect is the tunneling of electrons through the barier from the valence band to conduction band. It has a negative temperature coefficient, because an increase in temperature reduces the size of the bandgap. Avalanche breakdown, on the other hand describes the mechanism, that free electrons (due to temperature) are accelerated to such energies, that they knock out other electrons, causing a avalanche of electrons. This effect has a positive temperature coefficient, because a higher temperature results in more free carriers, that cause leakage but less avalanches due to the reduced free length of path. While the zero temperature coefficient point is around $5\,\mathrm{V}$, this operating point implies a high succeptibility to changes in the reverse current. So typically the Zener voltage is shifted slightly upwards to result in a net positive coeffcient, which is then compensated by the negative temperature coeffcient of a forward biased diode [104].

This results in the typical Zener diode voltage of around $6.2\,\mathrm{V} + 0.7\,\mathrm{V} = 6.9\,\mathrm{V}$. In comparison to other types of diodes, buried Zeners have the best stability and lowest noise. In order to achieve high stability and low noise, $V_{ref} \approx 7\,\mathrm{V}$ is therefore pretty much set in stone. Table 1.6, lists some commercially available buried Zener diodes. All diodes are manufactured by Analog Devices as they are the sole manufacturer left to produce this kind of diodes.

Choosing a voltage reference can be done according to table 1.3.1. A temperature coefficient of $\leq 1\,\mathrm{\mu A/(A\,K)}$ rules out any non-hermetic unheated voltage reference. Using a hermetic package improves the stability against humidity as the epoxy used for an SO-8 package is hydrophilic and swells when exposed to water vapour causing pressure on the die, resulting in a change of the output voltage. The hermetic voltage references can be divided into two groups, the LM399 and the newer ADR1399 in one group and the LTZ1000 and its newer counterpart ADR1000 in aother. While the LM399 requires very few external components, the external circuit for the LTZ1000 is far more elaborate requiring more parts and space. Additionally, the LTZ1000 is more than four times the price of the LM399 in quantities of 10 at the time of writing. Last but not least, the stability and temperature coefficent of the LTZ1000 cannot be matched by the performance of the sense resistor, so the sense resistor gives a lower bound of about $0.5\,\mathrm{\mu A/(A\,K)}$. Unless the better low frequency noise performance is absolutely required, the LM399 and ADR1399 are the more economical parts. The performance of those two references will be discussed in section 2.1.7.

With the maximum reference voltage of $7\,\mathrm{V}$ known, a sense resitor between $14\,\Omega$ ($500\,\mathrm{mA}$) and $28\,\Omega$ ($250\,\mathrm{mA}$ is required. The Vishay VPR221Z are high power low drift metal foil resistors in a TO-220 package and a solid choice here. The low value resistors, combined with the requirement for a low noise output, limits the choice of op-amps to bipolar low-noise devices or discrete implementations. Table 1.7 lists some choices compiled from the literature sources, which will now be discussed.

Component	Wideband-noise	Low frequency noise	Temperature coefficent
LT1028	$0.85\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$	$35\,\mathrm{nV_{p-p}}$	$0.2\,\mathrm{\mu V/K}$
AD797	$0.9\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$	$50\,\mathrm{nV_{p-p}}$	$0.2\,\mathrm{\mu V/K}$
ADA4898	$0.9\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$	not specified	$1\,\mathrm{\mu V/K}$
ADA4004	$1.8\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$	$150\,\mathrm{nV_{p-p}}$	$0.7\,\mathrm{\mu V/K}$
AD8671	$2.8\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$	$77\,\mathrm{nV_{p-p}}$	$0.3\,\mathrm{\mu V/K}$

Table 1.7.: List of low-noise precision bipolar operational amplifiers with typical performance properties.

The low value of the sense resistor makes a bipolar op-amp the preferred choice, because they have a very low voltage noise and their current noise and input bias current do not interfere with such a low value resistor. While a discrete solution using matched jfets or bipolar transistors may push the input noise even lower, the temperature stability, circuit complexity and again the size speaks against this option, so the disussion will be limited to integrated solutions only. To find a reference point for the choice of op-amp, the thermal noise of the sense resistor must looked at. The $28\,\Omega$ sense resistor has a thermal noise of

$$e_n\,(23\,^\circ\mathrm{C}) = 0.67\,\mathrm{nV}/\sqrt{\mathrm{Hz}}\,.$$

This means, that even the lowest noise op-amp from table 1.7 dominates the wideband-noise. With this said, the component choices made in literate can be discussed. The AD8671 chosen

by [35] only makes sense, because they have chosen a very large filter resistor R_{filt} of $2 \times 3\,\mathrm{k\Omega}$. The ADA4004 was used by Moglabs in the DLC-202, again likely due to the high values of R_{filt} used. The ADA4898 might seem like a good choice at first sight but the very limited (in terms of precision op-amps) open-loop gain of $0.14\,\mathrm{V/\mu V}$ makes this op-amp a cheap, but poor choice. The final choice is between the AD797 and the LT1028, both op-amps have very similar specifications, but there is a peculiarity in the datasheet of the LT1028 [71]. While there is a current noise spectrum, there is no voltage noise spectrum to be found in the datasheet. The author assumes a good deal of specsmanship at this point. The publication by Libbrecht et al. already blames the LT1028 for a noise peak around $400\,\mathrm{kHz}$. This peak is also included in the noise models for the op-amp and was additionally confirmed by the author with a measurement. This peak is the reason why Seck et al. found the LT1028 to be higher noise than the AD797. Additionally to the superior noise performance, the AD797 (B-grade) has excellent specifications overall. The open-loop gain is between 2–$20\,\mathrm{V/\mu V}$, the supply rejection is greater than $1\,\mathrm{\mu V/V}$, the bias current is almost constant between 20–$100\,^{\circ}\mathrm{C}$ and the unity gain bandwidth is around $10\,\mathrm{MHz}$. Finally it does have a very high output drive capability of $50\,\mathrm{mA}$, which allows to drive fairly large MOSFETs. These features make the AD797 the ideal op-amp for those low-value sense resistors, although it puts limits on the maximum filter resistor to limit the low frequency current noise contribution.

Finally, the choice of MOSFETs can be discussed. As it was shown in section 1.8.3 in equation 1.76, the channel length modulation plays an important role in increasing the channel conductance g_{DS} and limiting the output impedance. To reduce the channel length modulation a longer channel is preferred. Manufacturers do not give these numbers, nor the manufacturing process. Older technologies like the planar (lateral) FET is better suited for operating in the saturation region than the modern trench (vertical) FET. Trench MOSFETs are geared towards a low on-state resistance $R_{DS,on}$, which is important for switching applications, but their lower resistance comes from a shorter channel. One of the few planar MOSFETs still available on the market is the HEXFET, which was designed for switching applications, but prooves useful nonetheless as we will see. High voltage MOSFETs also have longer channels than low voltage MOSFETs, so browsing for MOSFETs, that are rated for 60–$100\,\mathrm{V}$ or more can narrow down the candidates. While the output impedance is a factor worth keeping in mind, the most important aspect is, whether the MOSFET can drive the load regarding the compliance voltage. To outline the problem, we can again refer to the example parameters from table 1.9.

Assuming a supply voltage of $15\,\mathrm{V}$ and the AD797 op-amp, the current source supply voltage V_{sup} is then limited to about 11–$12\,\mathrm{V}$, because the AD797 is no rail-to-rail op-amp and its output only swings to within $3\,\mathrm{V}$ of the rail (minimum) and the input is limited to within $2\,\mathrm{V}$ of the rail (minimum). Considering the maximum V_{ref} at full output of $7\,\mathrm{V}$ and a load voltage of $3\,\mathrm{V}$ in case of the L785H1 [60] used as an example in this section leaves only

$$V_{DS,min} = V_{sup} - V_{ref} - V_{load} = (11\text{–}12)\,\mathrm{V} - 7\,\mathrm{V} - 3\,\mathrm{V} = (1\text{–}2)\,\mathrm{V} \qquad (1.103)$$

for the MOSFET – a serious challenge.

To find a suitable MOSFET, one has to consult the *Typical Output Characteristics* graph in the datasheet. Using the maxium output current specifaction it is possible to estimate the minimum drain-source voltage V_{DS} to keep the MOSFET in saturation at the given maxium output current. This again narrows down the list of candidates.

The final aspect is the capacitive nature of the MOSFET gate. This property was brushed in appendix A.5 and the parasitic capacitances can be found in figure A.8. The AD797 can drive fairly large capacitive loads and several hundred pF are possible. It is best to keep the input

capacitance C_{iss} below $500\,\mathrm{pF}$. Do remember the output impedance of the AD797, is about $10\,\Omega$ at $1\,\mathrm{MHz}$ and rising by an order of magnitude at $10\,\mathrm{MHz}$. The $500\,\mathrm{pF}$ results in an impedance of around $300\,\Omega$ dropping by an order of magnitude at $10\,\mathrm{MHz}$, so keeping capacitance low, allows for a higher bandwidth of the current source.

Using these guidelines, searching a MOSFET across a lot of manufactures can still be tedious, but, for example, the distributor Digikey allows filtering and sorting by voltage and input capacitance. The follwing MOSFETs in table are given as an example and can be chosen for their respective current ranges.

MOSFET	Maximum V_{DS}	Input capacitance C_{iss}	Current range
IRF9610	200 V	170 pF	100–250 mA
IRF9Z10	50 V	270 pF	250–500 mA
IRF9Z14	60 V	270 pF	250–500 mA

Table 1.8.: Example MOSFETs for a current source and recommended current ranges.

The current range of the MOSFETs in table 1.8 is given based on the datasheet, making sure, that the MOSFET can be biased into saturation for the estimated minimum V_{DS} according to 1.103. The IRF9Z10 is a lower voltage version of the IRF9Z14 and the IRF9Z14 should be preferred if available. Those MOSFETs starting with *IRF* are all HEXFETs formerly made by International Recitfier, whose MOSFET business was bought by Vishay in 2007.

To summarize the component selection. The ADI AD797 op-amp is a highly recommended choice for being low-low noise with enormous gain, high bandwidth and a strong drive current. The MOSFET to accompany it is the Vishay IRF9Z14 for medium power applications. The reference must be a burried Zener diode and the ADI LM399 or ADR1399 is recommended. Regarding the sense resistors, they must be able to dissipate up to $500\,\mathrm{mA} \cdot 7\,\mathrm{V} = 3.5\,\mathrm{W}$ with minimal drift, making the Vishay VPR221Z a very good choice.

1.8.8. Current Source Example Parameters

Throughout this section, example calculations are performed to give the reader an idea of real-life parameters that can be applied to the theoretical models. These parameters are summarized in table 1.9, including their origin.

Parameter	Value	Source
MOSFET drain current I_D	$250\,\mathrm{mA}$	L785H1 [60]
MOSFET κ	$0.813\,\mathrm{A\,V^{-2}}$	IRF9610 SPICE model [18]
MOSFET channel length modulation λ	$4\,\mathrm{mV^{-1}}$	IRF9610 SPICE model [18]
MOSFET source voltage	$3.5\,\mathrm{V}$–$4\,\mathrm{V}$	section ??
Source/Sense Resistor R_S	$30\,\Omega$ or $50\,\Omega$	section ??
Op-amp differential input impedance R_{id}	$7.5\,\mathrm{k\Omega}$	AD797 [4]
Op-amp open-loop gain A_{ol}	$2\,\mathrm{V\,\mu V^{-1}}$	AD797 [4]
Op-amp gain bandwidth product GBP	$10\,\mathrm{MHz}$	AD797 [4]

Table 1.9.: Parameters used throughout this section and their sources.

1.9. Temperature Controller

1.9.1. Tuning of a PID controller

The number of empirical algorithms to determine a set of PID parameters (k_p, k_i, k_d) are numerous. In this work only the most common algorithms and a few notable exceptions will be presented.

1.9.2. Design

2. Results

"It's still magic even if you know how it's done."

– Sir Terry Pratchett, *A Hat Full of Sky*

2.1. Laser Current Driver

For this project several commercial and publicly available laser current drivers were evaluated for their performance. The following devices were all tested for the requirements listed in 1.3.1 and 1.3.2.

- Moglabs `DLC-202`

- Newport `TLB-6800-LN`

- SISYPH `SMC11 Puy Mary`

- Toptica `DCC 110`

- Vescent `D2-105`

- LQO `LQpr0` [41]

- A driver based on the work of Erickson et al. [35]

As a disclaimer, Moglabs, Vescent and SISYPH provided demo units, free of charge to the author and without any obligations regarding this work. The opinions and measurements in this work are in no way biased by this service. All of these drivers claim low-noise in various comparative forms, but vary in features. The `DLC-202`, the `TLB-6800-LN` and the `D2-105` additionally include a peltier controller. The `DLC-202` and the `TLB-6800-LN` have a modulation source. The `TLB-6800-LN`, the `DCC 110` and the design by Erickson et al. also feature a digital interface.

The drivers were all compared to our requirements. While not all drivers feature a remote accessible interface, their performance was assesed nonetheless to have a broader range of choices. A performance comparison can be found in section **??**. Unfortunately, none the drivers was able to properly drive the high compliance voltage required by the blue laser diode PL 450B of about 6–$7\,$V [88]. As it was discussed in section 1.8.5, the compliance voltage of all laser drivers based on the design of Libbrecht et al. [68] is limited to around 2–$3\,$V at full output (compare 1.103 for details). Since the compliance voltage is rougly proportional to the reciprocal of the output current current, limiting the maximum output current to about 30–$40\,\%$ increases the compliance voltage to the required level. Not only does this limit the choice of drivers, but also requiring a $500\,$mA driver for a $150\,$mA laser diode seems excessive and does not help with the noise requirements, because the output noise of those drivers scales

roughly with I_{max} as detailed in section 1.8.6, since the op-amp noise is the limiting factor. This lead to the decision to design a current source that meets all of our requirements, while surpassing all available alternative and tackling the compliance voltage limit. This design and its individual components are discussed in the following sections. First, the state of the art is presented, then the problems we encountered are outlined and finally our design, that resolves these issues is presented and the caveats and technical challenges are discussed.

2.1.1. The State of the Art in Laser Current Drivers

Prior to this work, all laser drivers for scientific purposes, were more or less strictly following the design proposed by Libbrecht et al. [68]. This design was presented in 1993 and back then, blue laser diodes were were not available and only developed in 1996. See [78] for an interesting historic summary. Finally, the efforts of Isamu Akasaki, Hiroshi Amano and Shuji Nakamura were rewarded with the Nobel Prize in Physics in 2014. The original laser driver design was therefore created for laser diodes requiring a low current and low compliance voltage compared to modern laser diodes. While the design remains useful for many low power near-infrared (NIR) laser diodes, these shortcommings were never addressed or even acknowledged by commercial alternatives. Sadly, the topic of the compliance voltage is usually not even mentioned in the datasheets – the Moglabs DLC-202 and SISYPH SMC11 are notable exceptions, but it is unclear from the datasheet to which version and/or currents of the devices the numbers relate. The Newport TLB-6800-LN is a bit different to the rest of the drivers tested, because it comes with Newport laser heads and reads its configuration data from the laser head. Ours came with a Vantage TLB-7100 and the laser head needs to be connected for it to work. Without some reverse engineering, these drivers can only be used with certain Newport laser heads. The TLB-6800-LN is included in the list of devices anyway to give an idea of its performance in existing systems.

Laser driver	Output current	Compliance voltage	Additional features
Moglabs DLC-202	**100 mA**, 250 mA, 500 mA	3.1 V	TEC, PID, Piezo
Newport TLB-6800-LN		–	TEC, Piezo, Digital
SISYPH SMC11	210 mA, **470 mA**	5 V	
Toptica DCC 110	**100 mA**, 500 mA, 3 A, 5 A	–	
Vescent D2-105	200 mA, **500 mA**	–	TEC
LQO LQpr0	**140 mA**, 400 mA	–	

Table 2.1.: Overview of laser current drivers tested. Marked in bold is the version tested in this work. A dash denotes that no official information is available.

The drivers shown in table 2.1 will now be discussed in a little more detail to familiarize with them.

Starting with the Moglabs DLC-202, which is fully integrated unit, that leaves little functions to be desired, it includes the current driver, a pizo driver, a temperature controller for a thermoelectric cooler (TEC) and finally a lock-in amplifier with a PID controller. It is the most integrated solution, that was tested and brings all features required to set up an ECDL locked

to an atomic transition. Some of its features are accessible via pin headers. The current source can be remotely enabled and the PID controller can be manipulated to allow relocking the laser or deliberately taking it out lock remotely. There is no way directly adjust the output current over the whole range. It also features a broad range of protection mechanisms for the laser diode, for example disconneting the driver in case of a short or open condition. All relavant quantities can be adjusted and read back from the front panel display. The manual is fairly comprehensive and gives a lot of examples to set up a laser system.

The Vescent D2-105 also features more than just the laser driver and includes a TEC controller as well. Unfortunately, adjusting this temperature controller prooved fairly cumbersome, because the driver has to be opened to adjust trimpots inside the driver. When the unit is placed in a rack, this might even be impossible to realize. The laser current is normally adjusted via a 10-turn potentiometer and can also be steered via an external input, but the stability of the current then depends on the external control voltage. It does not feature any protection features like open or short detection, so disconnecting and reconnecting the cable to the laser will most likely damage the laser diode. The display can be switched to show all relavant quantities. Moreover, the manual is not as comprehensive as the manual of the DLC-202, but covers all relavant settings of the driver.

The SISYPH SMC11 does not include any additional features, but covers the most important protection features like an open detection and shuts down the driver accordingly. It is fully rackmountable, unlike the drivers discussed so far. The setpoint of the driver is adjusted via a recessed trimpot using a screwdriver, which has prooven troublesome to adjust in the lab when not directly in front of the unit. The current can be externally adjusted using an input connector, but again this limits the stability to that of the external source. The driver does not have any display and all setpoints must either be adjusted blindly or a volmeter must be attached to the monitoring connector limiting the usefulness in a lab environment. The user manual covers only the basic settings and gives no details regaring the layout of the pin headers or external connectors making it hard to understand without having the device at hand.

The Toptica DCC 110 is also only a current source and is rack mounted. It comes with a separate display module, which connect via the backplane. The setpoint is adjusted via 10-turn potentiometer and can additionally be adjusted via the backplane with an external signal, again limiting the stability of the driver to the external source. The manual is fairly comprehensice and covers all essentials.

The Newport TLB-6800-LN is the only driver, that has a digital interface and supports Standard Commands for Programmable Instruments (SCPI) commands. It incorporates a TEC controller and a piezo controller. Unfortunately, it only works with a limited number of lasers, because it reads some parameters, like the maximum output current from the laser head at startup. The user manual covers all device functions, but gives little detail about the hardware, making this a closed system.

The results of the performance tests conducted prior to building our own solution will be presented on the following pages, including problems typically encountered and their solution. These tests include the stability, current noise and output impedance of the drivers. Not all drivers were put through the full test, if it was already clear that they could not perform in our setup for a laser system for the spectroscopy of highly charged ions driving a blue laser diode.

2.1.2. Laser Driver: Design Concept

In order to interpret the results in sections 2.1.5, 2.1.6, it it helpful to discuss in more detail the design concept of the current state of the art, which is based on the driver design presented by Libbrecht et al. [68]. The design can be split into the four building blocks shown in figure 2.1. A supply voltage input filter, a reference voltage to create the setpoint, a unidirectional current source and some form of bidirectional current source used for modulating the laser current at high frequency.
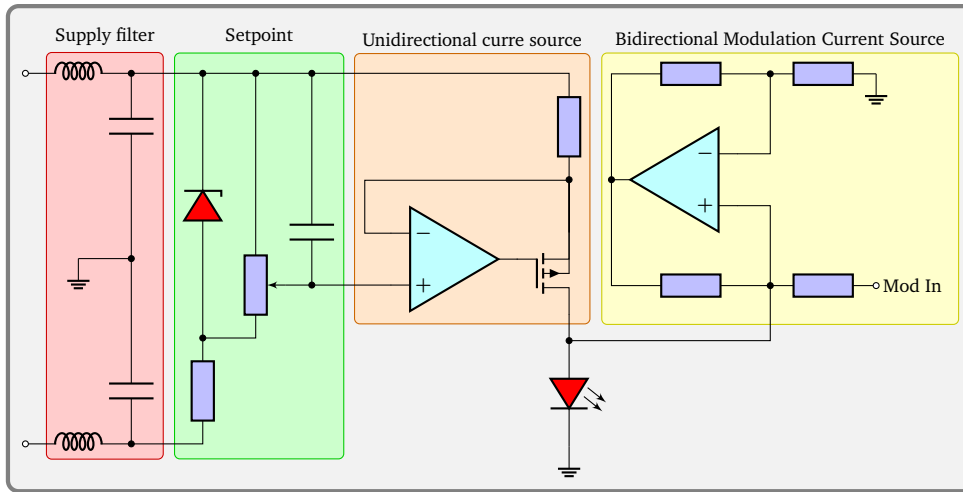

Figure 2.1.: Building blocks of a laser driver based on [68].

The original design is the most straightforward approach and it is possible to reproduce it even on prototype printed circuit boards (PCBs). Erickson et al. [35] replaced the potentiometer with a DAC, but left the other parts untouched. Taubman [114, 115] published some extensive modifications, which not only replaced the reference circuit with a DAC and an LTZ1000 reference, but also added extensive filtering of the supply. The next sections will discuss these different elements separately and give some in insight into the different versions of the elements found in literature. The sections also details problems discoverd and the solution proposed in the design presented here.

2.1.3. Supply Filtering

The supply section of the design by Libbrecht et al. was shown simplified in figure 2.1. The original filteronly consists of a CLC filter or sometimes called π-filter shown in figure 2.2. Do note, that due to the small input capacitance, the filter is basically just an LC-filter.
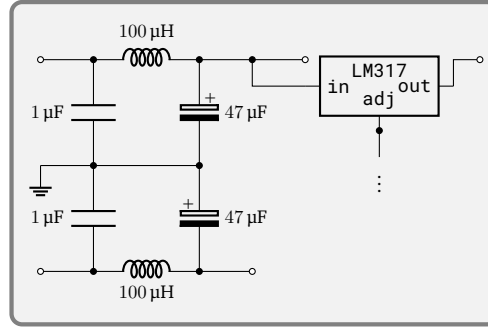

Figure 2.2.: Power supply filter of a laser driver based on [68]. The op-amps are supply by the filtered voltage and the current source is supplied by the LM317.

The LC-filter is best suited for a low impedance source like a power supply, because it has a high input impedance. From the transfer function

$$H(s) = \frac{Z_{out}}{Z_{in}} = \frac{\frac{1}{sC}}{sL + \frac{1}{sC}} = \frac{1}{s^2 LC + 1} = \frac{\frac{1}{LC}}{s^2 + \frac{1}{LC}} = \frac{\frac{1}{LC}}{\left(s + i\frac{1}{\sqrt{LC}}\right)\left(s - i\frac{1}{\sqrt{LC}}\right)}, \quad (2.1)$$

one can deduce, that the passband gain at DC is 1 (obviously) and additionally, there are two complex poles in the imaginary plane at $s = \frac{\pm i}{\sqrt{LC}}$, putting the cutoff frequency of the 2nd order filter at $f_c = 2.3\,\text{kHz}$. Do note, due to the imaginary poles, there is some gain peaking at f_c. Normally, this is damped by the parasitic resistance of the inductor and capacitor. A more detailed analysis follows below, because the cutoff frequency is all that is of interest right now.

In this design, the op-amps are directly driven off the filtered supply rail. Using this information it is possible to estimate the effectiveness of the filter. Using the $30\,\text{nA}_{\text{rms}}$ in a $100\,\text{kHz}$ bandwidth current noise requirement from table 1.3.2, the voltage noise at the sense resistor ($14\,\Omega$, (see section 1.8.7) must be no more than $420\,\text{nV}_{\text{RMS}}$. Now, taking for example a low-noise switch-mode power supply like the Rohde & Schwarz HMP4040 used at CERN [9], which does have fairly pronouced noise at the switching frequency of around $170\,\text{kHz}$ and harmonics. The author measured these glitches to be about $3\,\text{mV}_{\text{pp}}$. The noise will have to go through the filter and the supply rejection ratio (PSRR) of the op-amp. The PSRR of the LT1028 at $170\,\text{kHz}$ is about 10^{-2} under ideal conditions [71] and the filter adds another 10^{-2} when accounting for a $0.5\,\Omega$ series resistance of the output capacitor. The total filtering adds up to about 10^{-4}, which still leaves $21\,\text{nA}_{\text{pp}}$ of ripple on the drive current in a very small bandwidth.

To have a better rejection of such switch-mode noise, the filter must be improved. The paper presented by Taubman [114] shows a brute-force approach. They applied extremely high values for the capacitor C_{LC} of the LC filter of $10\,\text{mF}$ and then put a second filter based on a so-called capacitance multiplier behind it. This implementation is shown in a simplified form in figure 2.3 and briefly discussed now. For a more detailed schematic and part names see [114].

Taubman built this filter for a driver with a driving capacity of $2\,\text{A}$, which limits the size of the inductor that can be used, in order to make up for that, he is forced to use giant capacitors.
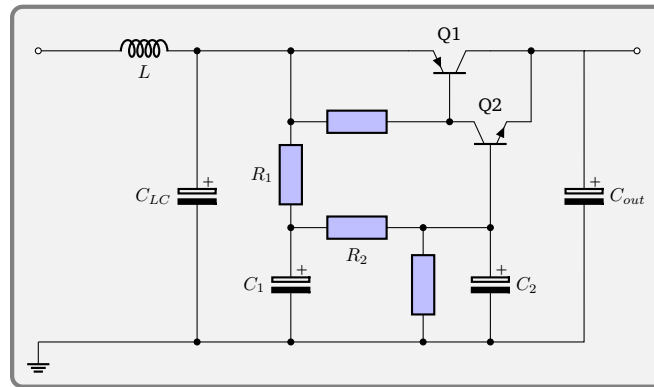
Figure 2.3.: Power supply filter using a capacitance multiplier for a cutoff frequency of $0.5\,\mathrm{Hz}$. This is a simplified schematic based on [114]. Only the positive rail is shown.

The second stage of the filter comprises a capacitance multiplier, which is formed by wrapping a feedback loop around the 2$^{\mathrm{nd}}$ order filter created by $R_1 C_1$ and $R_2 C_2$. This feedback loop removes the main current from the filter resistors, to allow larger values for R_1 and R_2, while maintaining a fairly low output impedance of the filter. The properties of this construction will be discussed now.

As a first note, the circuit presented by Taubman misses a detail, which should included, when handling such capacitances. The circuit must include a reverse polarity protection, or rather a reverse current protection. If the input is shorted by accident, the $10\,\mathrm{mF}$ of capacitance would immediately discharge via the parasitic body diode of the slow start-up transistor, likely vaporizing everything in its path. This could be implemented by adding another transistor to act as a reverse current protection.

The following explanation relies on some basic knowlegde about transistors. A good introduction to transistors is the Transistor Manual [28], some even call the bible of transistors. Armed with some basic knowledge, the capacitance multiplier can be discussed and it must already be said, that the term capacitance multiplier is a bit misleading. It neither multiplies the capacitance, nor does is behave like a real 2$^{\mathrm{nd}}$ order filter. The only thing that is multiplied by the gain of the transitor, is the output capacitance seen by load, making the capacitor look more ideal. Unfortunately, it highly depends on the properties of the transistor(s) and the gain of transistors drops with increasing output current (although it rises with temperature). Make sure to consult the datasheet, which typically gives a plot of the ac current gain h_{fe} vs the collector curren I_C to see at which load current, h_{fe} starts dropping. Another issue is the bandwidth of the circuit, because h_{fe} of any transistor rolls off with frequency. The high frequency resonse of this filter then becomes constant. This limits the suppression at around $1\,\mathrm{kHz}$ (depending on the output current of course) to around 500–1000. It is usefuly at low frequencies though, as it is shown in the publication of Taubman. Another point is the maximum ripple voltage, that can be filtered. The muliplied capacitance does, of course, not store the same amount of energy as a real capacitor. This means the maximum peak-to-peak input ripple is limited to about one diode drop of $0.68\,\mathrm{V}$ from the Collector-Emitter diode. If more ripplle rejection is required an additional resistor from the base of Q_2 to ground like shown in figure 2.3 can be applied. This reduces the output voltage further though. In this design, the current through R_m is sufficient.

As a final remark regarding the capacitance multiplier is the output impedance. The transistor has an output impedance like a diode, so it increases with decreasing current. This means it

will also drop about $1.2\,$V at $500\,$mA and about $0.3\,$V at $1\,$mA. This behaviour must be taken care of by the voltage regulator following the capacitance multiplier. The $2\,$V drop is also not a problem in this use case and even comes in handy. If just the supply rail of the laser diode current is fed through the capacitance multiplier, as it is the most sensitive, and the supply rail for the op-amps is not, then those extra $2\,$V will do not be a problem. In section 1.8.7 it was already mentioned, that for example the AD797 op-amp needs a supply that is $3\,$V above the diode supply. This means, that less voltage needs to be dropped by the linear regulator that follows the filter. To sum it up, the capacitance multiplier behaves like an ordinary RC filter, but with a lower output impedance and only works at low ripple voltages, is limited in the high frequency domain.

The power supply filters applied in this design use a passive LC filter for the negative and positive rail, then a capacitance multiplier on the diode supply. The negative rail is simply mirrored from the positive rail and pnp instead of npn transistors and vice versa are used. The combined filter is shown in figure 2.4. The diode supply and the analog rail, which is taken before the capacitance multiplier, are fed to low noise post-regulators, the LT3045 and its negative counterpart, the LT3094. Both regulators have excellent power supply ripple rejection (PSRR) out to at least $1\,$MHz of more than 10^3. At low frequency the PSRR is even higher and more than 10^5 can be expected. This allows a combined PSRR of better than 10^6 from low to high frequencies, even beyond $1\,$MHz.
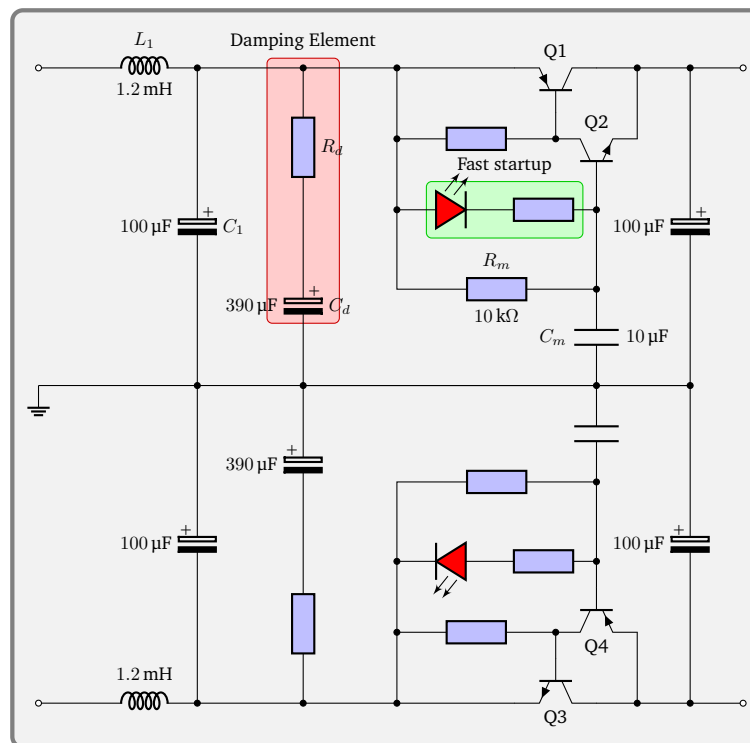

Figure 2.4.: Power supply filter of the digital current driver.

Regarding the filter circuit shown in figure 2.4 a few explaining words on the choice of components are in order before proceeding to the measurement of the PSRR. Going back to equation 2.1, we saw, that the undamped second filter has excessive ringing at the cutoff frequency, because the filter poles are imaginary. To address this, there are several solutions. The most simple one is adding a damping element, either in parallel to the capacitor or in

parallel the inductor. In this case a damping element in parallel to the capacitor was chosen, because using a a damping element parallel to the inductor will degrade the filter performance by making the blocking inductor lossy. Using the arangement shown in figure 2.4, the transfer function can be calculated.

$$H(s) = \frac{Z_{out}}{Z_{in}}$$

$$Z_{out} = (R_d + Z_{C_d}) \,||\, Z_{C_1} = \left(R_d + \frac{1}{sC_d}\right) \,||\, \frac{1}{sC_1} = \left(\left(R_d + \frac{1}{sC_d}\right)^{-1} + sC_1\right)^{-1}$$

$$= \frac{sC_dR_d + 1}{s^2C_1C_dR_d + s\,(C_1C_d)} \tag{2.2}$$

$$Z_{in} = sL_1 + (R_d + Z_{C_d})$$

$$H(s) = \frac{\frac{sC_dR_d+1}{s^2C_1C_dR_d+s(C_1C_d)}}{sL_1 + \frac{sC_dR_d+1}{s^2C_1C_dR_d+s(C_1+C_d)}} = \frac{sC_dR_d + 1}{s^3L_1C_1C_dR_d + s^2L_1\,(C_1 + C_d) + sC_dR_d + 1} \tag{2.3}$$

This is the transfer function of a 3rd order filter. This type of filter was discussed by Middlebrook [80] (reprinted in [25] and [26]). Middlebrook derived, that there is an optimal value for the series resitance R_d given a capacitance C_d and the filter components L_1 and C_1. This optimal value has minimal gain peaking, hence a minimal quality factor Q at the resonance freuency. The existance of such an optimal value can be easily understood from the fact, that if $R_d = \infty$ the resonance frequency is $\omega_0 = \frac{1}{\sqrt{L_1C_1}}$ and in case $R_d = 0$ it is $\omega_1 = \frac{1}{\sqrt{L_1(C_1+C_d)}}$. In between ω_0 and ω_1, there is a lossy zone, where R_d due to its lossy nature reduces Q, but at both ends $Q = \infty$, so there must be a minimum in between. By calculating, the minimum value of the transfer function at the point of resonance, Middlebrook found the following results:

$$R_0 := \sqrt{\frac{L_1}{C_1}} \tag{2.4}$$

$$n := \frac{C_d}{C_1} \Rightarrow C_d = nC_1 \tag{2.5}$$

$$Q_{optimal} = \sqrt{\frac{(4+3n)(2+n)}{2n^2(4+n)}} \tag{2.6}$$

$$R_d = R_0 \cdot Q_{optimal} \tag{2.7}$$

From these equations, it can be seen, that the damping capacitor C_d needs to be fairly large, depending on n. A critically damped system with $Q = 0.5$ would be preferred, but this would require $n \approx 6$ making the C_d prohibitively large. For this filter $n = 4$ was chosen, so making the filter slightly underdamped, so a slight gain peaking at the resonace can be expected. The following componets were chosen. First, a large, low resistance inductor L_1 capable of carrying at least $1\,A$ was chosen. In this case a Coilcraft `MSS1210-125KEB`. High reliability capacitors were chosen to ensure a long lifetime of the device. Choosing capacitors rated with a lifetime of $5000\,h$ at $105\,°C$ gives a expected service life of more than $10\,a$, when assuming an Arrhenius law with a doubling of the lifetime every $10\,K$. Apart from the reliability of the capacitors, there are no special requirements for them as there is little ripple current to be expected. The input power supply is supposed to be a filtered low noise supply and not the unfiltered output of a DC/DC regulator. So it is possible to maximize L_1 and choose a physically smaller C_1 since

board space is limited. This results in the following design values, calculated from equations 2.5 and 2.5, given the components values discussed above.

$$C_1 = 100\,\mu F$$
$$n = 4$$
$$Q_{optimal} \approx 0.61$$
$$R_d = 0.61 \approx R_0 \approx 2\,\Omega$$
$$C_d = 400\,\mu F \approx 390\,\mu F$$
$$f_c \approx 300\,Hz$$

Do note, that R_d does include the equivalent series resistance (ESR) of C_d, so the ESR of the capacitor must be subtracted from the final value of the damping resistor placed on the board. This may even absolve one from the need for a discrete resistor if the ESR of the capacitor is high enough. The transistors chose were a combination of a Toshiba TTA004B/TTC004B and Onsemi BC817-40/BC807-40 for the positive/negative rail. The TTA004B/TTC004B are good up to about $500\,mA$. At this point the gains start dropping. A higher power transistor like the Onsemi D45H8/D44H8 used by Taubman is recommended for Q_1 and Q_3.

Finally, one last part of the capacitance multiplier should be explained. Highlighted in green in figure 2.4 is a fast startup circuit. At startup, the capacitor C_m is discharged and $15\,V$ will be applied, it will then begin to charge with a current of $1.5\,mA$ through the $10\,k\Omega$ resistor. Because the Q_2 is an emitter follower, hence the emitter follows the voltage at the base (minus a diode drop for the base-emitter diode). As a sidenote, when using this kind of circuit, since Q_2 is an emitter follower, all output capacitors, that follow the capacitance multiplier will charge at the same rate as C_m, voltage-wise, this means, that for every $10\,\mu F$ of output capacitance, a current of $1.5\,mA$ will flow. While this not significant at this moment it become so, when looking at the fast startup circuit. Applying the input voltage of $15\,V$ at startup over the LED, a $625\,nm$ Würth Elektronik 150080RS75000, it will start conducting, resulting in a 1.8–$2\,V$ drop. The current flowing into C_m is therefore dependent on the diode series resistor, which was chosen to be $510\,\Omega$, a value not particularly important in this case. So at startup about $25\,mA$ will flow into C_m, which means $2.5\,mA/\mu F$ will flow through Q_1 and Q_2. Assuming roughly $100\,\mu F$ of distributed bypassing capacitance around the board, this is around $500\,mA$. All these values are still well below the damage threshold of the transistors ($2.5\,A$ and $0.5\,A$) and the LED ($30\,mA$), but these values must be kept in mind, when adding larger output capacitors. The fast startup circuit ensures an output voltage of $13\,V$ within $100\,ms$ instead of around $0.5\,s$, reducing the time to boot and leaving more time for self-checks without impacting the user experience.

After choosing the values above, the filter was simulated using LTSpice to assert the validity of the parameters chosen. The simulation was conducted with a load current of $500\,mA$ running through the capacitance multiplier to simulate the worst case. As discussed above, the gain of the transistor h_{fe} drops at higher currents as the transistor saturates. This particularly affects the high frequency behaviour above $10\,kHz$. The source file can be found in source/spice/input_filter_dgdrive.asc. The simulation additionally includes the series resistance and parasitic parallel capacitance of L_1, the latter will induce some ringing at the self resonance frequency of the inductor at $1\,MHz$ and limit the useful attenuation beyond that to around 10^3 due to the capacitive coupling of the conductor windings. At the $170\,kHz$ discussed before, the damping is about the same figure of merit, 10^3.

The suppression is an order of magnitude better than the filter used by Libbrecht et al. and it does not even include the high performance regulators that follow. The transfer function for

both the damped LC filter and the LC filter with the capacitance multiplier in series is plotted in figure 2.5. The self resonance peak at $1\,\text{MHz}$ can be clearly seen and is not damped, but from the output impedance shows shows that, there is enough capacitance present to compensate for this. The output impedance above $1\,\text{MHz}$, it is dominated by the local bypass capacitors and not accurately represented by the simulation. It can be expected to by lower than the simulated results, which do not include those capacitors.
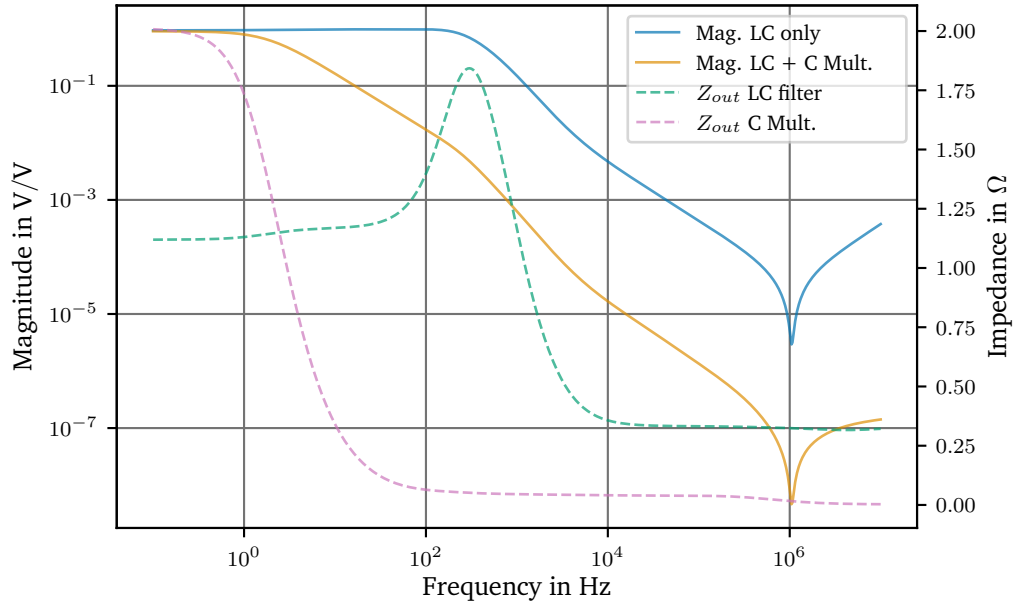

Figure 2.5.: Simulated response of input filter used in the digital current driver. Both magnitude and output impedance of the stages are shown.

At $300\,\text{Hz}$, the LC filter cutoff frequency, the output imedance of the LC filter shows some gain peaking. This is due to the underdamped response chosen and discussed above. This peaking increases the output impeance of $1.2\,\Omega$ in the passband, which is mostly the resistance of the inductor, to a total of $1.8\,\Omega$. This can be easily compensated for by the regulators downstream.

The rejection ratio of the LC filter and the capacitance multiplier is better than 10^3 at at $100\,\text{kHz}$ an above, delivering the performance estimated above. This is expected to keep switch-mode noise away from the laser driver current.

The high rejection ratio of the filter is expected to make the the experimental validation rather challenging, because there are a number of complications that derive from the active nature of the circuit. The capacitance multiplier must be loaded, preferably at the maximum current to show the worst case and additionally, the ripple voltage must be low enough to not saturate the capacitance multiplier.

This requires a highly sensitive VNA, that has a low frequency range. This setup uses an Omicron Lab `Bode 100`, which can measure from $1\,\text{Hz}$ to $50\,\text{MHz}$ with an exceptionally low noise floor of about $180\,\text{nV/Hz}^{0.5}$ [116]. Additionally a Stanford Research `SR560` was used as a pre-amplifier. To apply the ripple voltage to the power supply rails a Picotest `J2123A` negative line injector and a self-designed positive line injector was used. The positive line injector design is available open-source and be found in a Github repository at [15]. This injector is called `PB02`. During the measurement, it was found, that since the expected signal

ist extremely small, capacitive coupling became an issue. This necessitated isolating the VNA from the power supply. The output of the VNA was isolated via a transformer and same was done with the output of the battery powered SR560 going into the VNA. The transformer used for isolating the VNA output, was an injection transformer named PB01. It is center tapped to create an anti-symmetrical output for the injection transformers. The center tap reduces the output amplitude by one half. The details regarding this device and its construction can be found in annex A.6. The transformer used at the output of the SR560 is a Picotest J2123A. Both transformers are injection transformers and not dedicated isolation transformers. The consequences of this subtle detail will become imminent in a moment. The full measurement setup is shown in figure 2.6.
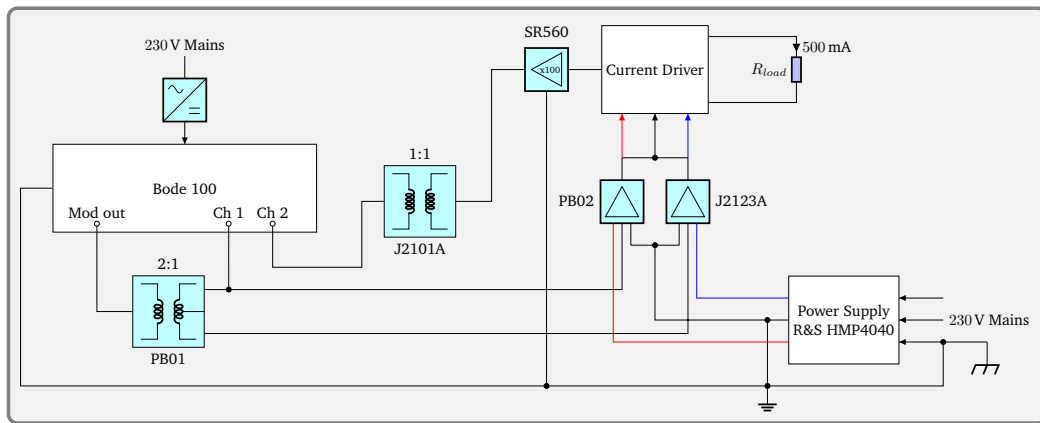

Figure 2.6.: Power and grounding scheme for a low noise measurement of the line filter rejection ratio, minimizing the interfernce of circuit return currents.

2.1.4. TODO: Need title

The author had actually intended to test the stability of the laser drivers first, but the plans were foiled by a misbehaving lab. First tests revealed, that depending on the temperature of cooling water supplied to the overhead air conditioning unit there were temperature fluctuation of up to $2\,\mathrm{K}$ observable in the lab. While the effect is not suprising, the observed temperature fluctuations impose a limit on the observable accuracy, due to the temperature coeficient of the multimeter used to record the data. To illustrate the effect, a sample measurement[1] is shown in figure 2.7. The Keysight 34470A used to record the data is specified at $6\,\mathrm{\mu A/K}$ for a current of $50\,\mathrm{mA}$ on the $100\,\mathrm{mA}$ range.
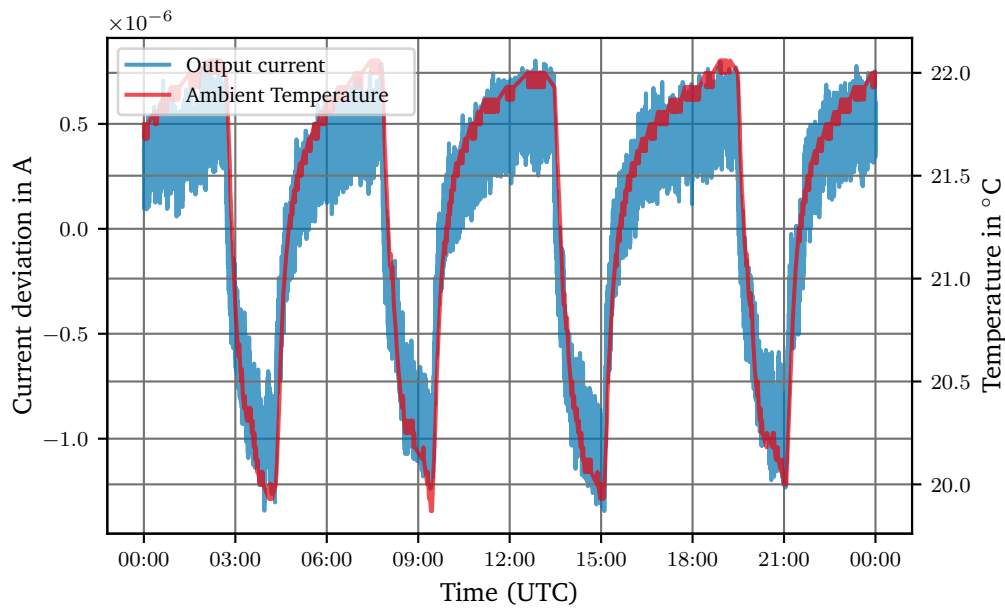

Figure 2.7.: Measurering a $50\,\mathrm{mA}$ current using a Keysight 34470A with changing ambient temperature. The current source is based on the design of Erickson et al.

Admittedly, the example in figure 2.7 does not reflect the correct way to measure current with high stability, but serves as an excellent example to highlight the problem. The room temperature heavily depends on the inner workings of the air conditioning system used in the building and the problem is not always present. Nonetheless, longer measurements over several days necessitated the development of a PID controller module to replace the stock temperaure controller of the air conditioning unit in the lab. This project is described in section 2.2.

In the meantime, measurements, that only required short-term stability were conducted. Measurering the current noise of the drivers is one such measurement.

---

[1] $100\,\mathrm{mA}$ range, $10\,\mathrm{PLC}$, AZERO ON, $f_s = 0.5\,\mathrm{Hz}$

2.1.5. Test Results: Current Noise

The spectral current noise density is a quantity, that is both seemingly trivial to measure and also easy to understand and graphically compare, therefore many devices are emblazoned by such graphs. The upside is, that these numbers can used for reference. Defining the bandwidth of such a measurement is a matter of debate, the available measurement devices and depends of the future use-case of the current driver. We chose an upper frequency of $1\,\text{MHz}$ for two reason, first, to limit the number of amplifiers required. As the noise power rises with the bandwidth (in the best case as $\sqrt{\Delta f}$ for white noise) and impedance matching comes into play, a higher power amplifier is required, a trait that does not bode well with low noise, low frequency frontends. So for frequencies above a few MHz different amplifiers are called for. More amplifiers make the whole measurement more intricate, because the amplifiers are the most critical parts in the whole chain.

The second reason does not root in the lazyness of the researcher, but has a physical origin. Cables used in the lab like RG-58 or RG-223 have a capacitance of about $100\,\text{pF/m}$. With a cable length of around $3\,\text{m}$ resulting in $300\,\text{pF}$, one finds, that at $10\,\text{MHz}$ the impedance seen by the laser diode approaches $50\,\Omega$, not unsurprising, given that the cable impedance is about $50\,\Omega$ and a signal at $10\,\text{MHz}$ has a wavelength of $\lambda \approx 2\,\text{m}$. This is aproaching the quarter-wave rule beyond which one should treat the cable as a transmission line. It is therefore reasonable to limit the noise measurement to $1\,\text{MHz}$, beyond which a design specific implementation including the laser head is called for anyway.

2.1.6. Test Results: Stability

When remotely controlling a laser system, stability of the laser driver is of immediate concern, because uninterrupted operation of the system is a key requirement and if the laser cannot be locked again remotely, it is time consuming, if possible at all, to go to the remote laser lab and readjust the laser current driver. The development of the past years have also shown a greater demand of remote working rendering readjustment unfeasable. To assure the specifications given in 1.3.1, the current drivers were first tested for $24\,\mathrm{h}$ to ensure. This test was refined several times over the course of this work, to reflect the need for a better signal so noise ratio. The first tests were done by feeding the output current of the laser driver into a (calibrated) Keysight 34470A and measuring the output over $24\,\mathrm{h}$. The 34470A was warmed up for $8\,\mathrm{h}$ and so was the laser driver. The ambient temperature and humidity were recorded by the lab monitoring system descibed in section **??**.

2.1.7. Zener Diode Selection

Early tests of the LM399 Zener diode as a reference have confirmed, what the data sheet [70] already suggest in the 'Low Frequency Noise Voltage' plot. There are random bi-stable voltage step changes. This phenomenon is called burst noise or popcorn noise.
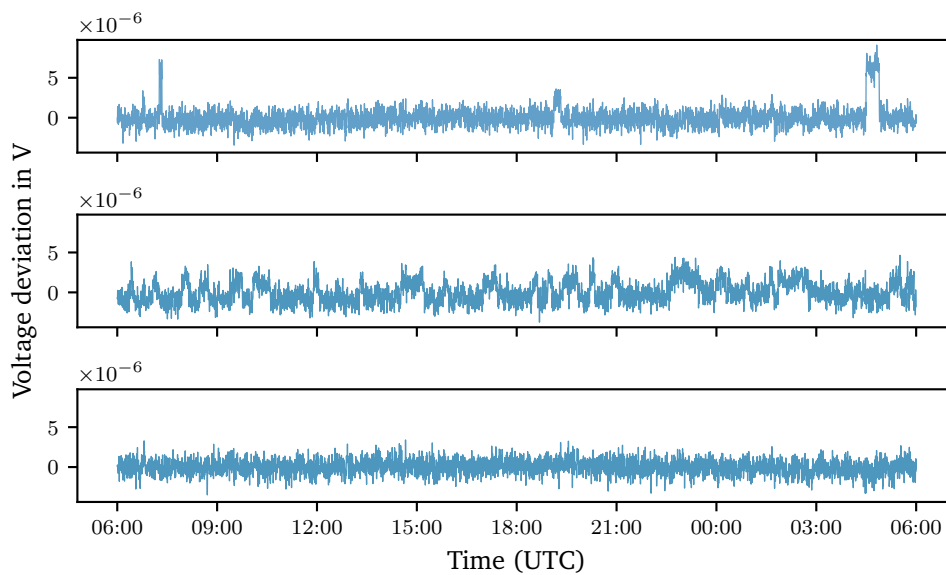

Figure 2.8.: Popcorn noise in different samples of the LM399 over a $24\,\mathrm{h}$ period.

Figure 2.8 shows two samples of the LM399, that exhibit popcorn noise, while the last one does not.

The sources of popcorn noise in semiconductor devices are not yet fully understood, but some sources have been identified. Defects in the semiconductor crystal lattice and contamination of the semiconductor material have been linked to popcorn noise [56]. This problem has improved over the years as manufacturing processes and wafer quality has evolved. Unfortunately the LM399 is built around a process from 1991, as can be seen etched into the die [55].

The popcorn noise caused by defects and contamination can be reduced by lowering the strain on the lattice and removing surface contaminants on the die. This can be can be

achieved using a high-temperature burn-in process. Manufacturers like Fluke and Keysight use similar techniques in their products. Fluke, for example, uses a period of $60\,\mathrm{d}$ burn-in for their references [92].

Fortunately, the LM399 is a heated reference, which regulates its die to $90\,°\mathrm{C}$ when turned on, so it is only required put the diodes in a simple test circuit and wait. The use of a separate test setup instead of the final circuit has both advantages and disadvantages. The disadvantage is, that the Zener diode will subjected to mechanical stress when soldered, this stress will not be removed by the burn-in process as it happens after the testing, when diode is soldered into the final circuit, but this mainly affects the voltage drift properties of the Zener diode and not the popcorn noise. The drift of the diode is also only of secondary concern in our setup, as the drift is mainly caused by the reference resistors used and are typically at least an order of magnitude worse than the the drift of the diode judging by the data sheet [70, 123].

The advantages of testing the Zener diodes separately, on the other hand, are, that more diodes can be tested at the same time, as a special compact test fixture can be used. It is also simpler to remove the diodes from the test fixture, because they be socketed. Therefore for our application a separate test board was used. Building this test setup is detailed in the next sections.

2.1.8. Building a Test Setup for Zener Diodes

There are several ways to measure the popcorn noise of semiconductor devices. The most trivial one is to directly monitor the device in the time-domain. In this case, the Zener voltage can be monitored with a long-scale multimeter. It requires a low noise DMM, that can reliably distinguish between both voltage levels, which are about $4\,\mathrm{\mu V}$ apart. A related option is to use a second reference, whose voltage is similar to the device under test (DUT). Measuring the the voltage difference between the two references, less resolution is required. Directly comparing the difference of two references using a millivolt meter is commonly done when intercomparing primary voltage references. This method, however, increases the measurement noise by a factor of $\sqrt{2}$, if both references produce the same level of uncorrelated noise. The noise of the LM399 with a $100\,\mathrm{PLC}$ integration time ($2\,\mathrm{s}$) is about $1.5\,\mathrm{\mu V_{pp}}$ as can be determined from the data in figure 2.8.

Measuring two references against each other would then result in around $2.1\,\mathrm{\mu V}$ of noise. This make distinguishing the jumps possible, but challenging.

A third option is to use a high-pass filter and an amplifier. Additionally, the signal can be low-pass filtered to remove any excess high frequency noise. This approach also requires less resolution than directly measuring the voltage, because the signal-to-noise-ratio is improved du the amplifier. It is therefore possible to use an off-the-shelf analog-to-digital converter (ADC). One such circuit, along with some examples, is demonstrated in [56, 57]. It must be noted, that due to the high-pass filtering, it not possible to measure slow voltage drifts using this method.

The forth and final option presented here, is approaching the problem in the frequency domain and requires a low-noise amplifier with a low frequency cutoff. As it was already discussed in section 1.6.1, popcorn noise is found to have a frequency dependence of $1/f^2$. This can be used to distinguish it from other random noise processes that show a frequency dependence of $1/f$. A good example of an op-amp, that has excessive burst in comparison to a good sample is given in *The Art of Electronics* [48, p. 478]. Going to frequencies below $10\,\mathrm{Hz}$, one can sort the references by their noise spectrum.

In this work only options one and two were were tested, as it was said above, with options

Figure 2.9.: Voltage noise of an LM399, measured with a bandwidth of $0.1\,\mathrm{Hz}{-}10\,\mathrm{Hz}$.

three and four there is a chicken and egg problem. One needs a number on known good devices to compare other DUTs to. At the start of the evaluation, most of the data available about the LM399 was from the data sheet. Compiling a dataset of the performance of dozens of LM399 is expensive and time consuming and companies typically treat such data as a closely guarded secret.

The next section deals with the choice of multimeter to satisfy the requirements test the Zener diodes according to options one and two, so either directly measuring the output voltage or difference of a known good sample against the DUT.

2.1.9. Choosing a Multimeter for Testing Zener Diodes

The DMM used plays an important role for the test setup. In this section, some of the challenges, that can be encountered will be discussed. The expected amplitude of the popcorn noise is around $0.5\,\mathrm{\mu V/V}$ or $3.5\,\mathrm{\mu V}$ of the output voltage, when considering the $7\,\mathrm{V}$ Zener voltage of the LM399 diode.

The $7\,\mathrm{V}$ will typically be measured on the $10\,\mathrm{V}$ range. It is not a trivial task, because a signal-to-noise-ratio of $0.35\,\mathrm{\mu V/V}$ or more than $130\,\mathrm{dB}$ is required. This calls for a device, that not only has the required resolution, but also the stability over time and temperature to ensure the measurement will not be distorted by the DMM.

Therefore, a voltmeter with lower noise and a more stable reference, than the DUT is mandatory. This only leaves the class of very low noise 7.5 or 8.5 digit multimeters. These multimeters feature a different type of voltage reference, because the LM399 is not suitable due to its noise. The only Zener diodes that meet those requirements are the Analog Devices LTZ1000 [72], the Motorola SZA263 (out of production) and the Linear Technology (LT) LTFLU-1, a proprietary design by Fluke and LT. The LTZ1000, for example, is specified for a typical noise of $1.2\,\mathrm{\mu V_{pp}}$ in a frequency range of $0.1\,\mathrm{Hz}{-}10\,\mathrm{Hz}$ [72]. Additionally, in comparison to the LM399, those Zener diodes do not suffer from the popcorn noise issue.

The equipment manufacturers typically have a preference for one of those diodes. Keysight utilizes the LTZ1000, Fluke uses the SZA263 (in older devices) or the LTFLU-1 in newer model, while Keithley employs the LTZ1000 in their `Model 2002` and the LTFLU-1 in the newer

DMM7510, because they were bought by Fortive, the same company that owns Fluke. To sum it it up, Keysight uses the LTZ1000 and Fluke/Keithley the LTFLU-1 in their top end meters.

Comparing only 7.5 and 8.5 digit voltmeters, narrows down the choice of multimeters considerably. The market for high-end 8.5 digit DMMs is limited and therefore every device on the market caters for a certain niche. It is therefore prudent to look at their specifications to choose the correct device for this purpose. In table 2.2 a list of popular 8.5 DMMs can be found. Several models included in the table, are already discontinued, but these DMMs can still be acquired on the second-hand market.

Manufacturer	Model	Remarks
Advantest	R6581	Discontinued. Scanner cards available.
Datron/Wavetek	1812	Discontinued. Wavetek was bought by Fluke.
Fluke	8508A	Discontinued. $20\,\text{V}$ range.
Fluke	8588A	In production.
Keithley/Tektronix	2002	In production. Scanner card available. $20\,\text{V}$ range.
Keysight	3458A	In Production.
Solartron	7081	Discontinued. Slow.
Transmille	8104	In Production. External scanner available. Slow.

Table 2.2.: Overview of 8.5 digit multimeters.

While the author has not tested every multimeter in table 2.2, it is possible to judge some of them apriori by their specifications. The Solartron 7081 (also sold as Guildline 9578) is a less optimal choice, because a conversion takes $52\,\text{s}$ for 8.5 digits. The discontinued Fluke 8508A and the Wavetek 1812 multimeter are very similar devices, because Fluke bought Wavetek in 2000 and as a result, the Fluke 8508A is more of an update to the Wavetek 1812 than a new device. They are both in included in the list, because it is very rare to see one of the Fluke devices on the second hand market, while the Wavetek 1812 can be found with a bit of patience. Again they are fairly slow, taking $25\,\text{s}$ for a conversion at 8.5 digits.

The other multimeters are still in production and similar in price, but their field of use is slightly is different. The Fluke 8588A excels at stability and features a modern user interface, whereas the Keysight 3458A is unbeaten in linearity and noise. A detailed comparison of those two meters can be found in the work of Lapuh, Kucera, Kovac, and Voljc [65]. The Keithley Model 2002 focuses on its scanning capability and the Transmille 8104 does have electrometer functions. Unfortunately, the 8104 is also fairly slow at 8.5 digit with conversions taking $4\,\text{s}$ at its fastest setting [2], so it will not be considered.

To narrow it down even further, several 7.5 and 8.5 digit multimeters were tested. The results of those tests will be discussed here to give an impression of the performance of these devices. The tested multimeters are the Keysight 3458A, the Keithley Model 2002, the Keysight 34470A and a Keithley DMM6500. The 3458A was chosen, because it is very fast and already used in section ?? of this work. The Model 2002 was chosen for its internal scanning unit. The 34470A was chosen as a lower-end and cheaper alternative and because it is a fairly low noise device. Finally the DMM6500 is on the list to compare a DMM with an LM399 reference. A Fluke 8588A was not tested, because it was not released at the time of testing and the older model 8508A is considered too slow as mentioned above.

The tests

Two test were run on this selection of devices. The first one was done using a Fluke 5440B calibrator supplying $10\,\text{V}$ to all mulimeters and taking readings over the course of a week.

This data was used to estimate the noise and the stability of the multimeters, including burst noise. The noise of the DMM at $10\,\mathrm{V}$ is typically not found in the datasheet, because the noise performance is usually quoted for shorted inputs, which does not include the internal reference noise. This test allows to check for popcorn noise of the internal reference. The calibrator has a specified output noise of $<1.5\,\mathrm{\mu V}$ within a bandwidth of $0.1\,\mathrm{Hz}$–$10\,\mathrm{Hz}$ at $1\,\mathrm{V}$ and is stable to within $5\,\mathrm{\mu V_{RMS}}$ over $30\,\mathrm{d}$, a specification far superior to the LM399.

The second test was done using a known bad LM399 voltage reference instead of the calibrator. This test was done to see how well a DMM can make out the popcorn noise.

Based on these two tests, a multimeter was chosen for an automated test setup to bin the LM399s.

Test Setup

The tests were done in a stable and monitored lab environment, with a temperature deviation of at most $\Delta T = \pm 0.2\,\mathrm{K}$. All multimeters were connected to the same DUT. Although this might potentially cause interference between the multimeters due to the pump out current spikes caused by the switching interals, no ill effects, like voltage offsets or increased noise, were observed during the setup of the tests. A more detailed discussion of the pump out current of the 3458A can be found in [94].

The three 8.5 and 7.5 digit multimeters were connected using shielded cables, either Pomona 1167-60 or self-made cables. See section **??** for details on the self-made cables. The GUARD terminal of the calibrator was connected to chassis GROUND at the calibrator and then connected to the cable shield. On the 3458A, the shield was connected to the GUARD terminal and the GUARD switch was set to open according to the manual [59]. For the other multimeters, that do not have a GUARD terminal, the shield was left floating at the DMM side. Additionally the `Fluke 5440B`, the `HP 3458A` and the `Keysight 34470A` have an autocalibration routine, which was run once prior to the measurement. The detailed settings used for the DMMs can be found in the appendix A.1 on page 111, a summary ist given in table 2.3 to show the important differences.

DMM	Integration time in NPLC	Conversion time in s
HP 3458A	100	$0\,\mathrm{s}$
Keithley Model 2002	40	$0\,\mathrm{s}$
Keysight 34470A	100	$0\,\mathrm{s}$
Keithley DMM6500	90	$0\,\mathrm{s}$

Table 2.3.: Concise list of differences in the settings used for comparing the DMMs.

All DMMs were configured to have a similar conversion time. This lead to different integration times, which are given in power line cycles at $50\,\mathrm{Hz}$. The `Model 2002` takes considerable longer for a measurement than the Keysight multimeters. The reason is the auto-zero function, which is shown in figure 1.24. The `Model 2002` does three steps when doing auto-zeroing, it measures the signal, the zero point for an offset compensation and also the reference voltage for a gain correction. In comparison, the 3458A only corrects for the offset drift. The gain is adjusted when using the ACAL function. The former auto-zero routine, therefore takes longer by one half, but results in more stable measurements.

These measurements were done by measuring the output voltage of a pre-production version of the reference PCB for the digital current driver. The reference board was kept at $23\,\mathrm{°C}$ in a custom thermal chamber. The chamber is detailed in section **??**. Additionally, a $500\,\mathrm{g}$ bag of

Bentonite desiccant was added to keep the references at a low humidity of around 20 % relative humidity. The reference board inserted into a motherboard holding up to 4 reference modules. The motherboard, also called LM399 breakout board, provides the voltage regulators and the operational amplifier for the kelvin sensed pins of the reference. The multimeter was directly connected to the reference via a DB9 connector, without an other components in between the reference and the DMM like buffers, multiplexers or filters. The DMM itself was exposed to the ambient temperature of the lab. The setup is shown in figure 2.10.
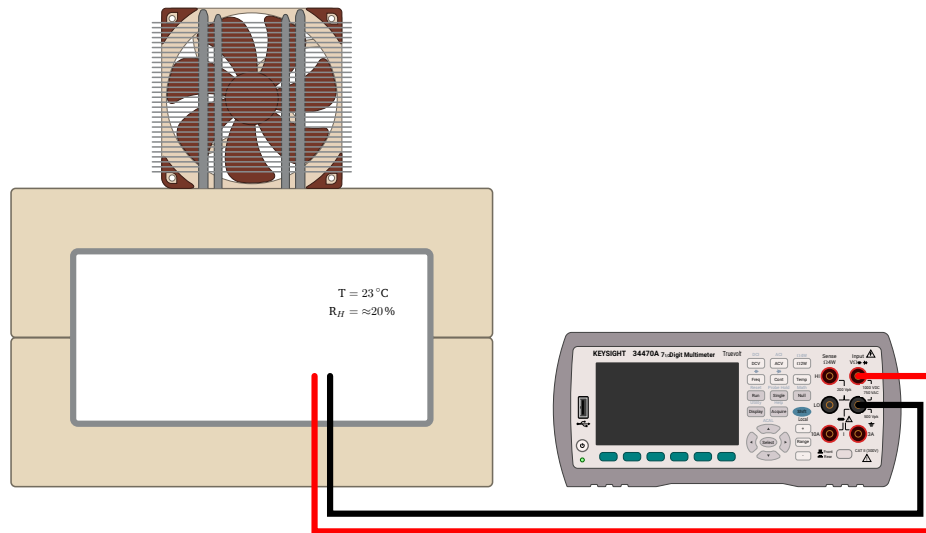

Figure 2.10.: Measurement setup for tesing an LM399 reference board with the `Keysight 34470A`

The reference boards amplify the Zener voltage to $10\,V$, which improves the signal to noise ratio, because it makes use of the full DMM range. The $10\,V$ range is typically the lowest (relative) noise and lowest drift range those multimeter because no internal pre-amplifiers of attenuators are required. It is important to keep the temperature drift of the DMM low or at least predictable, because the device is exposed to the ambient laboratory and not in a temperature controlled environment like the references.

The reference is a negative $10\,\mathrm{V}$ reference that uses a self-biasing technique to derive its $1\,\mathrm{mA}$ Zener current from its own $-10\,\mathrm{V}$ output. The details of this circuit are discussed in section **??**.

With the amplified output the expected burst noise step size of about $0.5\,\mathrm{\mu V/V}$, becomes $5\,\mathrm{\mu V}$. The resolution of the $10\,\mathrm{V}$ range of the `34470A` is $100\,\mathrm{nV}$, but the measurement is not limited by quantization. See section **??** of this work for a detailed characterization.
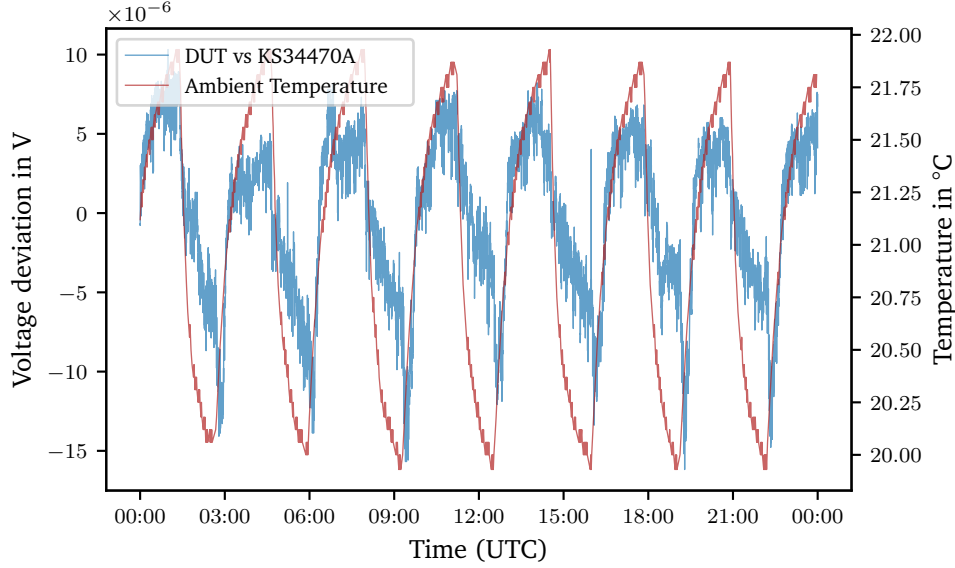

Figure 2.11.: Voltage deviation from the mean voltage of an LM399 negative $-10\,\mathrm{V}$ reference measured with a Keysight 34470A at $100\,\mathrm{PLC}$.

Figure 2.11 shows an example of such measurement. This measurement highlights one the problems encountered during those measurements. From this measurement it is unclear whether the features seen in the graph are only a result of ambient temperature changes due to the cycling of the air conditioning or popcorn noise on top of that. These results hightlight the fact, that sub-ppm measurements not only requires high-end gear, but also a very stable environment. From the data it follows, that the temperature coefficient of the DMM in linear approximation is:

$$\alpha_{\text{34470A}} \approx \frac{6.08\,\mathrm{\mu V} - (-9.30\,\mathrm{\mu V})}{(21.85\,\mathrm{°C} - 19.96\,\mathrm{°C})10\,\mathrm{V}} = 0.86\,\mathrm{\mu V/(V\,K)} \tag{2.8}$$

While the temperature coeeficient is vastly better than the specified $2\,\mathrm{\mu V/(V\,K)}$ [31], it is not low enough for this type of measurement. The multimeter must therefore be kept in a temperature controlled environment. This issue was resolved by replacing the stock air conditioning controller with a custom PID controller as discussed in section **??**. Lastly, the noise floor of the measurement is $1.5\,\mathrm{\mu V_{RMS}}$, resulting in an estimated signal-to-noise ratio (SNR) of about $10\,\mathrm{dB}$, which is suffient to detect the popcorn noise.

While the temperature issue was being worked on, testing of the Zener diodes continued. To work around the temperature drift of the DMM, the amplification of the reference voltage was increased to $15\,\mathrm{V}$, the same voltage required by the digital current driver, and a differential measurement was realized. This measurement was done against a primary $15\,\mathrm{V}$ reference board. To ensure, that any popcorn noise found originates only in the DUT and not the primary

reference used, several reference boards where tested against a `Fluke 5440B`. The 5440B does not exhibit popcorn noise as it uses a different voltage reference ic, namely two Motorola SZA263 in series [1]. Finally, a board that did not show popcorn noise in a period of three days was selected. The serial number of this primary or golden reference is *#1*.

Using this differential technique, the results greatly improved

In order to test a large amount of Zener diodes, and considering the duration of the burn-in process, which can take anything between $100\,\mathrm{h}$–$1000\,\mathrm{h}$, it is necessary to have an automated setup. This consists of a digital multimeter (DMM) a scanner and test board, that holds the Zener diodes and provides the necessary infrastructure for the diodes.

To conclude, we need a high performance DMM, a scanner, and a test fixture. The choices will detailed in the following sections.

2.1.10. A Scanner System for Testing Zener Diodes

As discussed before the diodes need to be tested for $1000\,\mathrm{h}$ and it is not be feasible to test them individually. So a minimum of 10 diodes must be tested at the same time. To keep the system compact, the test setup a scanner to multiplex a single multimeter input. Several commercial options currently available were considered for this project and are shown in table 2.4.

	Keysight		Keithley			Fluke	Rigol
	DAQ973A	34980A	DAQ6510	2750	3706	2680	M300
DMM	6.5	6.5	6.5	6.5	7.5	18 bit	6.5
Channels	3x20	8x40	2x10	5x20	6x60	6x20	5x32
FET	✓	✓	✓	✓	✓	✗	✗
Voltage	120 V	80 V	60 V	60 V	200 V	75 V	300 V
Card	DAQM900A	34925A	7710	7710	3724	2680A-PAI	MC3132
USB	✓	✓	✓	✗	✓	✗	✓
Ethernet	✓	✓	✓	✗	✓	✓	✓
GPIB	✓	✓	✓	✓	✓	✗	✓

Table 2.4.: Overview of scanner mainframes

A recent trend to more compact devices has led major manufacturers to include multimeters in the scanner mainframe creating so called data acquisition units. Legacy devices, that only have switching capabilities are no available. For example Keithley replaced the small desktop switch mainframe `Model 7001` with the `DAQ6510` and Keysight is offering the `DAQ973A`, a scanning 6.5 digit DMM, that accepts extension cards. Unfortunately, for this project, as discussed above, the integrated 6.5 digit multimeter does not add any value.

The simplest option is to go with an 8.5 digit multimeter that already included a scanner option or buy a used `Keithley 7001` from a second-hand dealer. The author has tested both options and the simplicity of only having a single device to connect and program makes the integrated scanner card of the `Model 2002` very attractive.

The scanner card used to multiplex the DMM does have to meet several specifications. The most important aspects are the number channels and the lifetime of the relays. Other factors, such as channel to channel isolation, the contact potential, resistance and maximum voltage is not the limiting factor.

The reason is, that in this case, the voltage is low, there is no ac component involved and the the typical input impedance of high-end multimeters is far more than $100\,\mathrm{G\Omega}$ [3, 67, 83, 94].

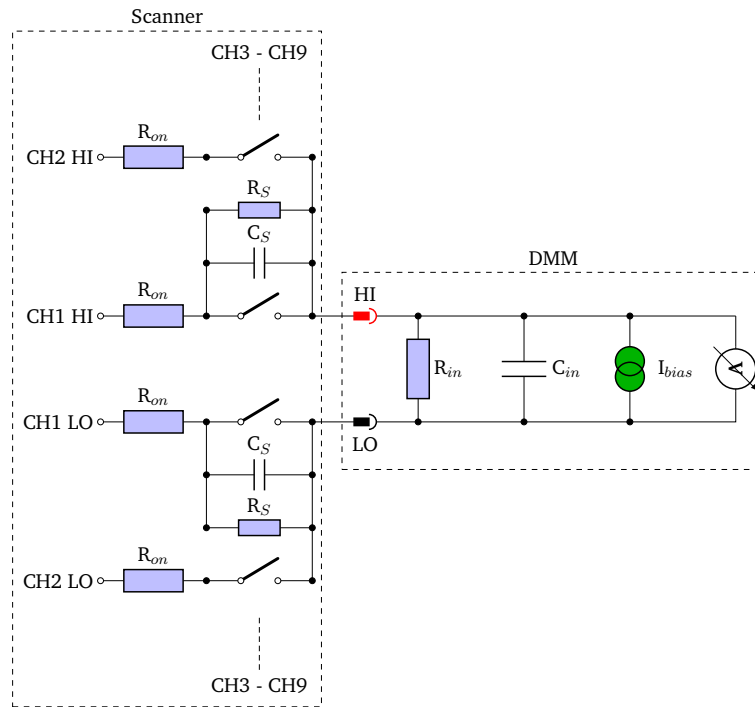In this work the Keithley (now Tektronix) `Model 2002` was chosen for three reasons. It is a

Figure 2.12.: Simplified schematic of the scanner front-end with parasitic elements

very compact system requiring only a half-sized 2U rack in comparison to the other DMMs, that are typically full-sized 2U rack devices. The other two advantages are the integrated scanner card slot, that allows to to fit a 10 channel scanner card and finally the $20\,\text{V}$ range. The latter is interesting for testing the final voltage reference boards, as these have a $15\,\text{V}$ output, which is too much for the $10\,\text{V}$ range of most DMMs, so that testing the voltage reference Printed circuit boards (PCBs) one would have to switch to the $100\,\text{V}$ range and forgo an extra digit of resolution and add more noise.

The test setup consists of a mounting PCB, that holds up to 20 Zener diode. It provides power regulation and a minimal circuit required to support each diode. This circuit is given here:
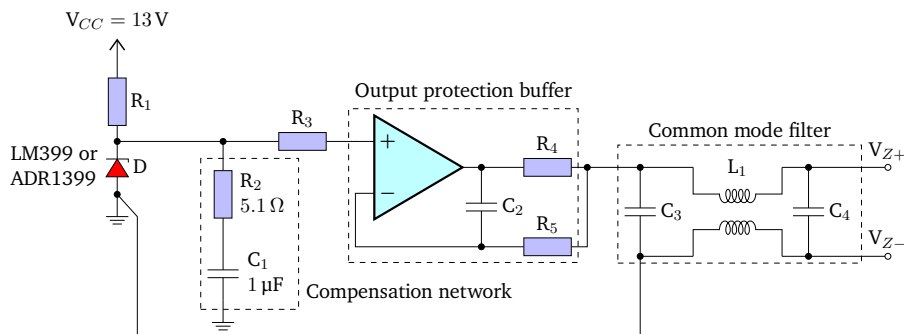

Figure 2.13.: Circuit used for burning in the Zener diodes

The compensation network is required when using the ADR1399, because of its very low dynamic impedance as recommended in the data sheet [84]. It is not strictly required for the LM399, but fitted nonetheless, because there are no downsides to it. This makes the board

compatible with both types of references. Each Zener output is protected using an output buffer, which provides isolation and short circuit protection. Finally there is a common mode filter at the output to suppress high frequency noise via ground loops.

The two key metrics of concern, that need to measured are popcorn noise and drift.
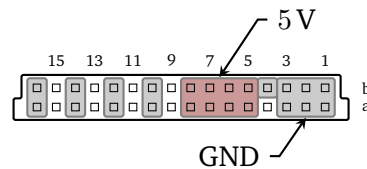
digital multimeter and a scanner card

Figure 2.14.: The extension connector used in several Keithley multimeters

Pin	Function	Cable Colour	Pin	Function	Cable Colour
a1, b1	$+6\,V$–$+20\,V$	brown	6	GND	green/white
a2, b2	PD cathode	red	7	LD Cathode	blue/white
a3, b3	LD case (GND)	red/white	8	LD Anode	blue
a4	PD anode (GND)	red/white	9	LD current	green
a5	$-6\,V$–$-20\,V$	brown/white			

As a sidenote, for the pure entertainment of the author, several batches of LM399 Zener diodes were purchased from non-authorized dealers. Some were marked as refurbished, the others were not marked as such, but clearly were. These so-called refurbished diodes are not to be used in production devices. To entertain and warn the reader a small selection of examples are shown here in figure 2.15. All but one diode, which is shown for comparison, are refurbished.

Figure 2.15.: Refurbished LM399 Zener diodes. From left to right:

As it can be clearly seen, the sellers have gone to some effort to hide the fact, that these diodes have been used before. When a through-hole is soldered to the PCB, its legs will be trimmed to match the PCB thickness. In order to conceal this, the legs need to be extended to their original length. The legs of the LM399 are Kovar, because the LM399 is hermetically sealed with a glass seal and Kovar has the same coefficient of expansion as borosilicate glass. The forgers typically weld steel legs to the Kovar legs and then either gold-plate or tin them, as can be seen in fig. 2.16.

Much to the delight of the author the refurbished diodes prove valuable for educational purposes. As the origin and method of extraction from the original circuit is unknown, but can be imagined to be rather savage, the diodes are typically faulty. They can therefore be used to validate the test setup and demonstrate the popcorn noise found in the LM399. A very good example in shown in fig. 2.17.

TODO: Chinese/Ebay Zeners. Welded legs. Phototts. Decap one of those.

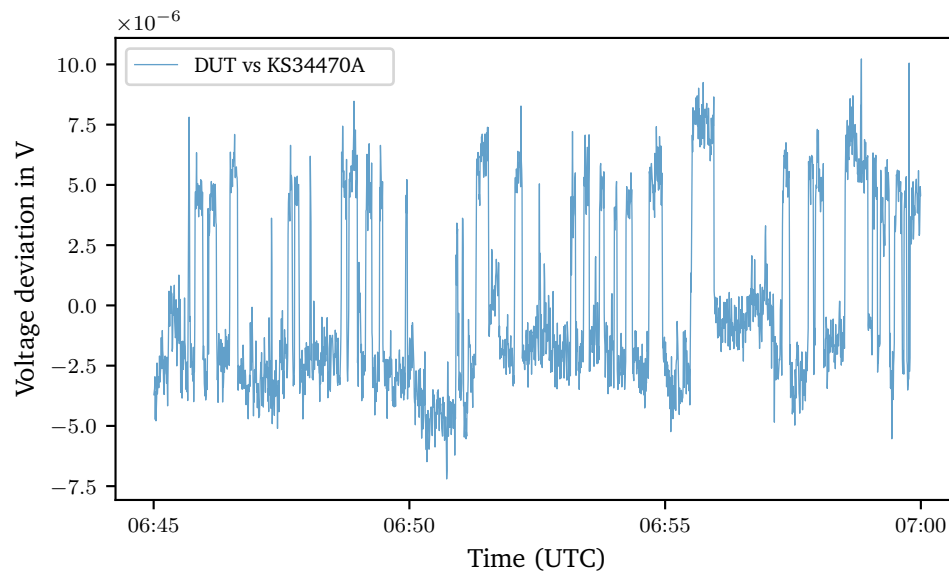Figure 2.16.: Fake steel legs of a refurbished LM399.

Figure 2.17.: Popcorn noise of a refurbished LM399 (#15) over a period of $15\,\mathrm{min}$.

2.2. Lab Temperature Control

2.3. Digital Temperature Controller

Bibliography

[1]   *5440B/AF Direct Voltage Calibrator - Service Manual*. John Fluke Mfg. Co. Inc. Dec. 1986.

[2]   *8100 Series - 8104 Extended Specifications*. 1.1. Transmille Ltd.

[3]   *8588A Reference Multimeter - Product Specifications*. rev. G. Fluke Corporation. Aug. 2022.

[4]   *AD797 Datasheet - Ultralow Distortion, Ultralow Noise Op Amp*. K. Analog Devices, Inc. 2015.

[5]   *ADL-78901TX Datasheet*. 0th ed. Arima Lasers Corporation.

[6]   David W. Allan. "Statistics of atomic frequency standards". In: *Proceedings of the IEEE* 54.2 (1966), pp. 221–230. DOI: 10.1109/PROC.1966.4634.

[7]   David W. Allan. "Should the classical variance be used as a basic measure in standards metrology?" In: *IEEE Transactions on Instrumentation and Measurement* IM-36.2 (1987), pp. 646–654. DOI: 10.1109/TIM.1987.6312761.

[8]   P.E. Allen et al. *CMOS Analog Circuit Design*. Analog Circuit Design: A Tutorial Guide to Applications and Solutions. Oxford University Press, USA, 2011. ISBN: 9780123851857.

[9]   C. Arnaboldi et al. "Very low noise AC/DC power supply systems for large detector arrays". In: *Review of Scientific Instruments* 86.12 (2015), p. 124703. DOI: 10.1063/1. 4936269.

[10]   K.J. Åström and R.M. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, 2010. ISBN: 9781400828739. URL: https:// fbswiki.org/wiki/index.php/Feedback_Systems:_An_Introduction_for_Scientists_and_ Engineers.

[11]   Karl Johan Åström and Tore Hägglund. *Advanced PID Control*. English. ISA - The Instrumentation, Systems and Automation Society, 2006. ISBN: 978-1-55617-942-6.

[12]   James A. Barnes et al. "Characterization of Frequency Stability". In: *IEEE Transactions on Instrumentation and Measurement* IM-20.2 (1971), pp. 105–120. DOI: 10.1109/ TIM.1971.5570702.

[13]   Patrick Baus. "Characterization of a Wide Tunable, Robust, Multi Application Diode Laser for Spectroscopy". MA thesis. Technische Universität Darmstadt, 2014.

[14]   Patrick Baus. *TinkerforgeAsync*. Version 1.2.0. July 2021. URL: https://github.com/ PatrickBaus/TinkerforgeAsync.

[15]   Patrick Baus. *Line Injector*. Version 1.0.1. July 2022. URL: https://github.com/ PatrickBaus/line_injector.

[16]   Patrick Baus. *Simualtion Scripts for the PhD Thesis*. 2023. URL: https://github.com/ PatrickBaus/php_supplemental.

[17] Nikolai Beev. "Measurement of Excess Noise in Thin Film and Metal Foil Resistor Networks". In: *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. 2022, pp. 1–6. DOI: 10.1109/I2MTC48687.2022.9806690.

[18] Alexander Bordodynov. *Bordodynov's Electronics*. http://bordodynov.ltwiki.org/. Accessed: 2022-12-06. 2022.

[19] G. Breglio et al. "Electro-thermal instability in low voltage power MOS: Experimental characterization". In: *11th International Symposium on Power Semiconductor Devices and ICs. ISPSD'99 Proceedings (Cat. No.99CH36312)*. 1999, pp. 233–236. DOI: 10.1109/ISPSD.1999.764106.

[20] *BUK7S1R5 Datasheet - N-channel 40 V, 1.5 mΩ standard level MOSFET in LFPAK88*. Nexperia B.V. Aug. 2002.

[21] Michael S. Cafferty and Eric D. Thompson. "Stable current supply with protection circuits for a lead-salt laser diode". In: *Review of Scientific Instruments* 60.9 (1989), pp. 2896–2901. DOI: 10.1063/1.1140625. eprint: https://doi.org/10.1063/1.1140625. URL: https://doi.org/10.1063/1.1140625.

[22] Eryn C. Cook et al. "High passive-stability diode-laser design for use in atomic-physics experiments". In: *Review of Scientific Instruments* 83.4 (2012), p. 043101. DOI: 10.1063/1.3698003.

[23] K.B. Cook and A.J. Brodersen. "Physical origins of burst noise in transistors". In: *Solid-State Electronics* 14.12 (1971), pp. 1237–1242. ISSN: 0038-1101. DOI: 10.1016/0038-1101(71)90112-2.

[24] B. Cordell. *Designing Audio Power Amplifiers*. 2nd ed. Focal Press. Routledge, 2019. ISBN: 9781138555440.

[25] S. Cuk and R.D. Middlebrook. *Advances in Switched-mode Power Conversion: Modelling, analysis and measurement*. Bd. 1. TESLAco, 1981.

[26] S. Ćuk. *Power Electronics: Modeling, Analysis and Measurements*. Vol. 2. Power Electronics. CreateSpace Independent Publishing Platform, 2015. Chap. 10. ISBN: 9781519513267.

[27] Samuel T. Dawkins, John J. McFerran, and Andre N. Luiten. "Considerations on the measurement of the stability of oscillators with frequency counters". In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 54.5 (2007), pp. 918–925. DOI: 10.1109/TUFFC.2007.337.

[28] General Electric Company. Semiconductor Products Dept. *GE Transistor Manual*. 7th ed. 1964.

[29] A. Van Der Ziel. "Thermal Noise in Field-Effect Transistors". In: *Proceedings of the IRE* 50.8 (1962), pp. 1808–1812. DOI: 10.1109/JRPROC.1962.288221.

[30] Lane Desborough and Randy Miller. "Increasing Customer Value of Industrial Control Performance Monitoring — Honeywell's Experience". In: 2002.

[31] *Digital Multimeters - 34460A, 34461A, 34465A (6½ digit), 34470A (7½ digit)*. Keysight Technologies. June 2022.

[32] P. Dutta and P. M. Horn. "Low-frequency fluctuations in solids: $\frac{1}{f}$ noise". In: *Rev. Mod. Phys.* 53 (3 July 1981), pp. 497–516. DOI: 10.1103/RevModPhys.53.497.

[33] Pieter Eendebak et al. *qtt*. 2022. URL: https://github.com/QuTech-Delft/qtt (visited on 11/03/2022).

[34] *EMO T - Stellantriebe, Thermischer Zweipunkt-Stellantrieb*. IMI Hydronic Engineering Deutschland GmbH. Aug. 2019.

[35] Christopher J. Erickson et al. "An ultrahigh stability, low-noise laser current driver with digital control". In: *Review of Scientific Instruments* 79.7 (2008), p. 073107. DOI: 10.1063/1.2953597.

[36] D. M. Fleetwood. "Origins of 1/f Noise in Electronic Materials and Devices: A Historical Perspective". In: *Noise in Nanoscale Semiconductor Devices*. Ed. by Tibor Grasser. Springer International Publishing, 2020, pp. 1–31. DOI: 10.1007/978-3-030-37500-3_1.

[37] D. M. Fleetwood, T. Postel, and N. Giordano. "Temperature dependence of the 1/f noise of carbon resistors". In: *Journal of Applied Physics* 56.11 (1984), pp. 3256–3260. DOI: 10.1063/1.333845.

[38] Nathan Flowers-Jacobs et al. *The NIST Johnson noise thermometry system for the determination of the Boltzmann constant*. en. Dec. 2017. DOI: 10.6028/jres.122.046. URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=923576.

[39] James B. Forsythe. "Paralleling Of Power MOSFETs For Higher Power Output". In: 1996.

[40] Richard Frey. *New 500V Linear MOSFETs for a 120 kW Active Load*. Application note. Advanced Power Technology, June 2000.

[41] Thorsten Führer. *Ultra-Low Noise Current Controller LQprO*. TU darmstadt, Institut für Angewandte Physik, Laser und Quantenoptik.

[42] *GH0781JA2C Data Sheet*. Sharp Fukuyama Laser Co., Ltd.

[43] Gino Giusi et al. "Full Model and Characterization of Noise in Operational Amplifier". In: *IEEE Transactions on Circuits and Systems I: Regular Papers* 56.1 (2009), pp. 97–102. DOI: 10.1109/TCSI.2008.927011.

[44] C.A. Greenhall. "Spectral ambiguity of Allan variance". In: *IEEE Transactions on Instrumentation and Measurement* 47.3 (1998), pp. 623–627. DOI: 10.1109/19.744312.

[45] Charles Greenhall. "Removal of drift from frequency stability measurements". In: *Telecommunications and Data Acquisition Progress Report* (Nov. 1981).

[46] Stewart M. Hansen. "Precision current control for quantum cascade lasers as flight calibration sources". MA thesis. Utah State University, Jan. 2014. DOI: 10.26076/bb34-2427.

[47] Ian Hegglun. *PAK Project - PA Calcs*. https://paklaunchsite.jimdofree.com/spice-models/. Accessed: 2022-12-06. 2022.

[48] Paul Horowitz and Winfield Hill. *The Art of Electronics*. 3rd ed. Cambridge: Cambridge University Press, 2015.

[49] David Howe. "Interpreting Oscillatory Frequency Stability Plots". en. In: 2002 IEEE Intl. Freq. Cont. Symp , New Orleans, LA, Jan. 2002. URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=105279.

[50] *Generic Standard on Printed Board Design*. Standard. Bannockburn, USA: IPC International Inc., Nov. 2012.

[51] *IRF9610*. C. Vishay Siliconix. Aug. 2021.

[52] K.M.B. Jansen et al. "Effect of temperature and humidity on moisture diffusion in an epoxy moulding compound material". In: *Microelectronics Reliability* 107 (2020), p. 113596. ISSN: 0026-2714. DOI: 10.1016/j.microrel.2020.113596.

[53] J. B. Johnson. "Thermal Agitation of Electricity in Conductors". In: *Physical Review* 32.1 (July 1928), pp. 97–109. DOI: 10.1103/PhysRev.32.97.

[54] N.J. Kasdin and T. Walter. "Discrete simulation of power law noise (for oscillator stability evaluation)". In: *Proceedings of the 1992 IEEE Frequency Control Symposium*. 1992, pp. 274–283. DOI: 10.1109/FREQ.1992.270003.

[55] Richard Kaußler. *Richi´s Lab - Linear Technology LM399*. 2021. URL: https://www.richis-lab.de/REF02.htm (visited on 09/02/2022).

[56] Arthur Kay. *Analysis and Measurement of Intrinsic Noise in Op Amp Circuits, Part VIII: Popcorn Noise*. Technote VIII. Texas Instruments Incorporated, Feb. 2008.

[57] Arthur Kay. *Operational Amplifier Noise: Techniques and Tips for Analyzing and Reducing Noise*. Elsevier Science, 2012. ISBN: 9780080942438.

[58] Dan Kegel. *The C10K problem*. May 1999. URL: https://web.archive.org/web/19990508164301/http://www.kegel.com/c10k.html.

[59] *Keysight 3458A Multimeter - Calibration Manual*. Keysight Technologies. July 2022.

[60] *L785H1 Data Sheet*. A. Thorlabs, Inc. Jan. 2019.

[61] A. Lapierre et al. "Relativistic Electron Correlation, Quantum Electrodynamics, and the Lifetime of the $1s^2 2s^2 2p\ ^2P^o_{3/2}$ Level in Boronlike Argon". In: *Phys. Rev. Lett.* 95 (18 Oct. 2005), p. 183001. DOI: 10.1103/PhysRevLett.95.183001.

[62] R. Lapuh. *Sampling with 3458A, Understanding. Programming, Sampling and Signal Processing*. 1st ed. Ljubljana: Left Right d.o.o., 2018. ISBN: 9789619447604.

[63] Rado Lapuh. personal communication. Nov. 26, 2022.

[64] Rado Lapuh et al. "Keysight 3458A Noise Performance in DCV Sampling Mode". In: *IEEE Transactions on Instrumentation and Measurement* 66 (Apr. 2017). DOI: 10.1109/TIM.2017.2681238.

[65] Rado Lapuh et al. "Fluke 8588A and Keysight 3458A DMM Sampling Performance". In: (May 2022). DOI: 10.48550/arXiv.2205.11321.

[66] Paul J Leach, Michael Mealling, and Rich Salz. "A Universally Unique IDentifier (UUID) URN Namespace". In: *RFC* 4122 (2005), pp. 1–32. URL: https://datatracker.ietf.org/doc/html/rfc4122.

[67] Ivan Lenicek, Damir Ilic, and Roman Malaric. "Determination of digital voltmeter input parameters". In: *2007 IEEE Instrumentation & Measurement Technology Conference IMTC 2007*. 2007, pp. 1–4. DOI: 10.1109/IMTC.2007.379145.

[68] K. G. Libbrecht and J. L. Hall. "A low-noise high-speed diode laser current controller". In: *Review of Scientific Instruments* 64.8 (1993), pp. 2133–2135. DOI: 10.1063/1.1143949.

[69] Tobias Liebmann. "Mikrocontrollergesteuerte Temperaturregelung". Bachelor's Thesis. TU Darmstadt, 2017.

[70] *LM199/LM399, LM199A/LM399A Precision Reference*. C. Analog Devices, Inc. Dec. 2014.

[71]   *LT1028 Datasheet - Ultralow Noise Precision High Speed Op Amps*. D. Analog Devices, Inc. 2015.

[72]   *LTZ1000/LTZ1000A - Ultra Precision Reference*. D. Analog Devices, Inc. Apr. 2012.

[73]   Kent Howard Lundberg. "Noise Sources in Bulk CMOS". In: 2002.

[74]   Stefan Machlup. "Noise in Semiconductors: Spectrum of a Two-Parameter Random Signal". In: *Journal of Applied Physics* 25.3 (1954), pp. 341–343. DOI: 10.1063/1. 1721637.

[75]   Volkhard Mäckel et al. "Laser spectroscopy of highly charged argon at the Heidelberg electron beam ion trap". In: *Physica Scripta Volume T* T156 (Sept. 2013), pp. 4004–. DOI: 10.1088/0031-8949/2013/T156/014004.

[76]   Alexander Martin. "Laser Spectroscopic Investigation of Exotic States in Noble Gases". PhD thesis. Technische Universität Darmstadt, 2017. URL: http://tuprints.ulb.tu-darmstadt.de/6666/.

[77]   Alexander Martin, Patrick Baus, and Gerhard Birkl. "External cavity diode laser setup with two interference filters". In: *Applied Physics B* 122.12 (Dec. 2016). DOI: 10.1007/ s00340-016-6575-9.

[78]   Herbert Paul Maruska and Walden Clark Rhines. "A modern perspective on the history of semiconductor nitride blue light sources". In: *Solid-State Electronics* 111 (2015), pp. 32–41. ISSN: 0038-1101. DOI: 10.1016/j.sse.2015.04.010.

[79]   A.S. McCormack and K.R. Godfrey. "Rule-based autotuning based on frequency domain identification". In: *IEEE Transactions on Control Systems Technology* 6.1 (1998), pp. 43–61. DOI: 10.1109/87.654876.

[80]   R. D. Middlebrook. "Input filter consideration in design and application of switching regulators". In: *IEEE Industry Applications Soc. Arm. Meeting* (Jan. 1976).

[81]   Mike Engelhardt. *LTspice*. Version 17.1.5. Jan. 14, 2023. URL: https://www.analog. com/en/design-center/design-tools-and-calculators/ltspice-simulator.html.

[82]   Edoardo Milotti. "1/f noise: a pedagogical review". In: *arXiv e-prints* (Apr. 2002).

[83]   *Model 2002 - Multimeter Specifications*. M. Keithley Instruments, Inc. Aug. 2022.

[84]   *Oven-Compensated, Buried Zener, 7.05 V Voltage Reference*. A. Analog Devices, Inc. Mar. 2022.

[85]   *Paralleling power MOSFETs in high power applications*. Application note. Nexperia B.V., Sept. 2021.

[86]   B. Pellegrini et al. "$\frac{1}{f^\gamma}$ noise in thick-film resistors as an effect of tunnel and thermally activated emissions, from measures versus frequency and temperature". In: *Phys. Rev. B* 27 (2 Jan. 1983), pp. 1233–1243. DOI: 10.1103/PhysRevB.27.1233.

[87]   David W. Pessen. "A new look at PID-controller tuning". In: *Journal of dynamic systems, measurement, and control* 116.3 (1994), pp. 553–557. DOI: 10.1115/1.2899252.

[88]   *PL 450B Data Sheet (PRELIMINARY)*. 1.2. OSRAM Opto Semiconductors GmbH. Aug. 2021.

[89]   W. H. Press. "Flicker noises in astronomy and elsewhere." In: *Comments on Astrophysics* 7.4 (Jan. 1978), pp. 103–119.

[90]   Tilman Preuschoff et al. "Wideband current modulation of diode lasers for frequency stabilization". In: *Review of Scientific Instruments* 93.6 (2022), p. 063002. DOI: 10.1063/5.0093520.

[91]   Jr. Puckett Jason Niles. "An Electrical and Statistical Study of Burst Noise". PhD thesis. California Institute of Technology, 1971.

[92]   Paul Rako. *Voltage reference drift considerations*. Dec. 2010. URL: https://www.edn.com/voltage-reference-drift-considerations (visited on 09/02/2022).

[93]   B. Razavi. *Design of Analog CMOS Integrated Circuits*. 2nd ed. McGraw-Hill higher education. Tata McGraw-Hill, 2017.

[94]   G. Rietveld. "Accurate determination of the input impedance of digital voltmeters". In: *Science, Measurement and Technology, IEE Proceedings* - 151 (Oct. 2004), pp. 381–383. DOI: 10.1049/ip-smt:20040700.

[95]   William Riley and David Howe. *Handbook of Frequency Stability Analysis*. en. July 2008. URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=50505.

[96]   Sebastian D. Saliba and Robert E. Scholten. "Linewidths below 100 kHz with external cavity diode lasers". In: *Appl. Opt.* 48.36 (Dec. 2009), pp. 6961–6966. DOI: 10.1364/AO.48.006961.

[97]   Schieder, R. and Kramer, C. "Optimization of heterodyne observations using Allan variance measurements". In: *A&A* 373.2 (2001), pp. 746–756. DOI: 10.1051/0004-6361:20010611.

[98]   D.E. Seborg et al. *Process Dynamics and Control*. 4th ed. John Wiley & Sons, 2016.

[99]   Christopher M. Seck et al. "Noise reduction of a Libbrecht–Hall style current driver". In: *Review of Scientific Instruments* 87.6 (2016), p. 064703. DOI: 10.1063/1.4953330.

[100]   *Secrets Of Rf Circuit Design, Third Edition*. Vol. 3. McGraw-Hill/TAB Electronics, Dec. 2000. ISBN: 9780070593855.

[101]   A.S. Sedra and K.C. Smith. *Microelectronic Circuits*. Oxford series in electrical and computer engineering. Oxford University Press, 1998. ISBN: 9780195116632.

[102]   Frank Seifert. *Resistor Current Noise Measurements*. Tech. rep. LIGO- T0900200-v1.

[103]   D.H. Sheingold. "Impedance & admittance transformations using operational amplifiers". In: *Lightning Empiricist* 12.1 (1964), pp. 1–8.

[104]   D.H. Sheingold et al. *Analog-digital Conversion Handbook*. 3rd ed. Analog Devices technical handbooks. Prentice-Hall, 1986. ISBN: 9780130328489.

[105]   W. Shockley. "The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors". In: *Bell System Technical Journal* 28.3 (1949), pp. 435–489. DOI: 10.1002/j.1538-7305.1949.tb03645.x.

[106]   W. Shockley. "A Unipolar "Field-Effect" Transistor". In: *Proceedings of the IRE* 40.11 (1952), pp. 1365–1376. DOI: 10.1109/JRPROC.1952.273964.

[107]   G.J. Silva, A. Datta, and S.P. Bhattacharyya. *PID Controllers for Time-Delay Systems*. Control Engineering. Birkhäuser Boston, 2007. ISBN: 9780817644239. DOI: 10.1007/b138796.

[108]   Guillermo J. Silva, Aniruddha Datta, and S.P. Bhattacharyya. "PI stabilization of first-order systems with time delay". In: *Automatica* 37.12 (2001), pp. 2025–2031. ISSN: 0005-1098. DOI: 10.1016/S0005-1098(01)00165-0.

[109] Sigurd Skogestad. "Simple analytic rules for model reduction and PID controller tuning". In: *Journal of Process Control* 13.4 (2003), pp. 291–309. ISSN: 0959-1524. DOI: 10.1016/S0959-1524(02)00062-8.

[110] W.M. Smith. "Worst case circuit analysis-an overview (electronic parts/circuits tolerance analysis)". In: *Proceedings of 1996 Annual Reliability and Maintainability Symposium*. 1996, pp. 326–334. DOI: 10.1109/RAMS.1996.500682.

[111] P. T. Starkey et al. "A scripted control system for autonomous hardware-timed experiments". In: *Review of Scientific Instruments* 84.8 (2013), p. 085111. DOI: 10.1063/1.4817213.

[112] Sveinn Steinarsson. "Downsampling Time Series for Visual Representation". MA thesis. 2013.

[113] H. Talvitie et al. "Passive frequency and intensity stabilization of extended-cavity diode lasers". In: *Review of Scientific Instruments* 68.1 (1997), pp. 1–7. DOI: 10.1063/1.1147810.

[114] Matthew S. Taubman. "Low-noise high-performance current controllers for quantum cascade lasers". In: *Review of Scientific Instruments* 82.6 (2011), p. 064704. DOI: 10.1063/1.3600602.

[115] Matthew S. Taubman et al. "Precision control of multiple quantum cascade lasers for calibration systems". In: *Review of Scientific Instruments* 85.1 (2014), p. 014704. DOI: 10.1063/1.4861200.

[116] *Technical Data Sheet Bode 100*. 1.3. OMICRON electronics GmbH. 2021.

[117] S. Tedja, J. Van der Spiegel, and H.H. Williams. "Analytical and experimental studies of thermal noise in MOSFET's". In: *IEEE Transactions on Electron Devices* 41.11 (1994), pp. 2069–2075. DOI: 10.1109/16.333824.

[118] Daylin L. Troxel, Christopher J. Erickson, and Dallin S. Durfee. "Note: Updates to an ultra-low noise laser current driver". In: *Review of Scientific Instruments* 82.9 (2011), p. 096101. DOI: 10.1063/1.3630950.

[119] Carel M. Van Vliet and Peter H. Handel. "A new transform theorem for stochastic processes with special application to counting statistics". In: *Physica A: Statistical Mechanics and its Applications* 113.1 (1982), pp. 261–276. ISSN: 0378-4371. DOI: 10.1016/0378-4371(82)90019-X.

[120] Vincent C. Vannicola. "'Modelling and Properties of Modulated RF Signals Perturbed by Oscillator Phase Instabilities and Resulting Spectral Dispersion'". In: *IEEE Trans. Inf. Theory* 29 (1983), p. 631.

[121] R. Vilanova and A. Visioli, eds. *PID Control in the Third Millennium: Lessons Learned and New Approaches*. Advances in Industrial Control. Springer London, 2012. ISBN: 9781447124245. DOI: 10.1007/978-1-4471-2425-2_5.

[122] Richard Von Mises. *Über Aufteilungs-und Besetzungswahrscheinlichkeiten*. 1939.

[123] *VPR221Z (Z-Foil) - Vishay Foil Resistors*. Vishay Precision Group, Inc. Mar. 2010.

[124] Anders E.E. Wallin. *AllanTools*. 2022. URL: https://github.com/aewallin/allantools (visited on 10/17/2022).

[125] P. Welch. "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms". In: *IEEE Transactions on Audio and Electroacoustics* 15.2 (1967), pp. 70–73. DOI: 10.1109/TAU. 1967.1161901.

[126] Carl E. Wieman and Leo Hollberg. "Using diode lasers for atomic physics". In: *Review of Scientific Instruments* 62.1 (1991), pp. 1–20. DOI: 10.1063/1.1142305.

[127] John Wright. *Don't Be Fooled by Voltage Reference Long-Term Drift and Hysteresis*. Tech. rep. Linear Technology Corp., Apr. 2000.

[128] Y. Yamamoto. *Fundamentals of Noise Processes*. 2017. Chap. 9. URL: https://www.nii. ac.jp/qis/first-quantum/e/forStudents/lecture/index.html.

[129] J. G. Ziegler and N. B. Nichols. "Optimum Settings for Automatic Controllers". In: *Journal of Dynamic Systems, Measurement, and Control* 115.2B (June 1993), pp. 220– 222. DOI: 10.1115/1.2899060. eprint: https://asmedigitalcollection.asme.org/ dynamicsystems/article-pdf/115/2B/220/5546571/220_1.pdf.

A. Appendix

A.1. Multimeter Settings for the Comparison Test

All were configured for maximum stability and similar conversion times using the following settings via SCPI. For better readability, all commands are shown unabridged.

HP 3458A

```
PRESET NORM;   # reset the device
TARM HOLD;   # stop readings
BEEP;
OFORMAT ASCII;   # return text
TRIG AUTO;   # trigger when ready
NRDGS 1,AUTO;  # take 1 reading
DCV 10;
AZERO ON;   # enable autozero
NDIG 9;
NPLC 100;
FIXEDZ OFF;   # High input impedance
TARM AUTO;   # enable readings
```

Keithley Model 2002

```
*CLS;   # clear events and errors
*RST;   # reset all settings
*OPC?;   # wait until device is reset
:INITiate:CONTinuous OFF;   # disable continuous initiation
:ABORt;   # place K2002 in idle
:SYSTem:AZERo:STATe ON;   # enable autozero
:SYSTem:AZERo:TYPE SYNChronous;   # azero for every reading
:SYSTem:LSYNc:STATe ON;   # line sync
:SENSe:VOLTage:DC:RANGe:UPPer 20;
:SENSe:VOLTage:DC:DIGits 9;
:SENSe:VOLTage:DC:NPLCycles 10;
:SENSe:VOLTage:DC:AVERage:COUNt 4;   # the averaging length
:SENSe:VOLTage:DC:AVERage:TCONtrol REPeat;   # filter type
:SENSe:VOLTage:DC:AVERage:ADVanced:STATe OFF;
:SENSe:VOLTage:DC:AVERage:STATe ON;   # Enable averaging
:FORMat:DATA REAL,64;   # read data as doubles
:FORMat:ELEMents READing;   # only return the reading
:FORMat:EXPonent HPRecision;   # Scientific notation
:INITiate:CONTinuous ON;   # Enable continuous triggering
```

Keysight 34470A

```
:SYSTem:BEEP;
:ABORt;
*RST;
*CLS;
:CONFigure:VOLTage:DC;
:SENSe:VOLTage:RANGe 10;
:SENSe:VOLTage:ZERO:AUTO ON;   # enable autozero
:SENSe:VOLTage:NPLCycles 100;
:SENSe:VOLTage:IMPEdance:AUTO ON;   # High input impendance
:FORMat:DATA ASCii,9;   # return 9 digits ASCII
```

Keithley DMM6500

```
SYSTem:BEEPer 500, 0.2;
ABORt;
*RST;
*CLS;
:SENSe:FUNCtion:ON "VOLTage:DC";
:SENSe:VOLTage:DC:RANGe:UPPer 10;
:SENSe:VOLTage:DC:LINE:SYNC ON;
:SENSe:VOLTage:DC:AVERage:COUNt 9;   # the averaging length
:SENSe:VOLTage:DC:AVERage:TCONtrol REPeat;   # filter type
:SENSe:VOLTage:AZERo:STATe ON;   # enable autozero
:SENSe:VOLTage:DC:NPLCycles 10;
:SENSe:VOLTage:INPutimpedance AUTO;   # High input impedance
:SENSe:VOLTage:DC:AVERage:STATe ON;   # Enable averaging
:FORMat:DATA ASCii;   # read data as double instead of text
:FORMat:ASCii:PRECision 16;   # return 16 digits ASCII
:DISPlay:VOLTage:DC:DIGits 6;   # set the screen to 6 digits
```

A.2. Multimeter Settings for the Comparison Test

SQL query

```sql
SELECT
  time
  ,data_values [1] humidity  --1st value in the array
  ,data_values [2] temperature --2nd value
FROM (
  SELECT
    bucket as "time"
    ,array_agg("data") as data_values
  FROM (
    SELECT
      time_bucket('6h',"time") AS "bucket"
      ,sensor_id
      ,avg(value) AS "data"
        FROM sensor_data
        WHERE
        time BETWEEN
          '2022-01-01T00:00:00.00Z' AND '2023-01-01T00:00:00Z'
        AND sensor_id IN (
          SELECT id
          FROM sensors
          WHERE label = '011_humidity' OR label = '011_temperature'
            AND enabled
        )
    GROUP BY bucket, sensor_id
    ORDER BY 1
  ) t1
  GROUP BY "bucket"
) t2
```
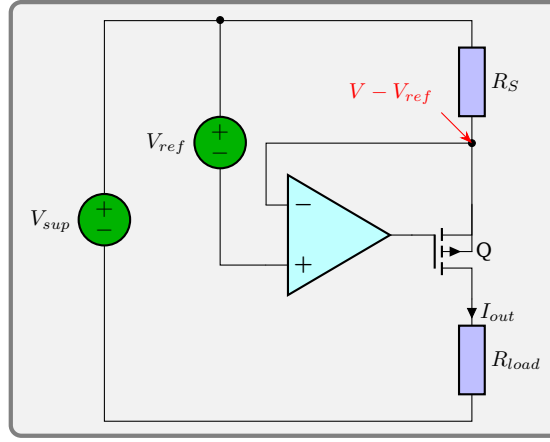
A.3. The Transconductance Amplifier with a MOSFET

Figure A.1.: Transconductance amplifier with a p-channel MOSFET.

The amplifier shown in figure A.1 is a feedback transconductance ampplifier as discussed in [101]. Its transfer function can be derived using the techniques presented in section 1.5.1. As a reminder, the general transfer function is definded as:

$$P(s) = \frac{I_{out}}{V_{ref}} \equiv A_f \,. \tag{A.1}$$

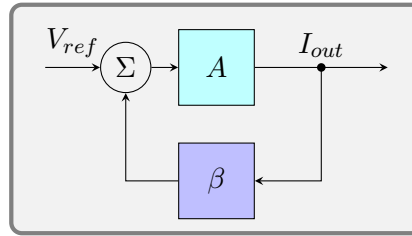The closed-loop transfer function is sometimes also called gain-with-feedback A_f [101] or noise-gain.

Figure A.2.: Block diagram of an amplifier with feedback β and gain A.

For the system shown in figure A.2, the closed-loop gain A_f can be written as

$$A_f = \frac{A}{1 + A\beta} \overset{A \to \infty}{=} \frac{1}{\beta} \,. \tag{A.2}$$

For the ideal transconductance amplifier with infinite open-loop gain A it follows, that the gain is simply reduced by the feedback factor β. For the MOSFET source voltage shown in figure A.1, β can be easily determined by inspection. The ideal op-amp with infinite open-loop gain A_{ol} has the same voltage at the inverting and non-inverting input. This means that below R_S at the source node of the MOSFET denoted in red, the voltage must be $V - V_{ref}$. This implies, that the voltage V_{ref} is dropped across R_S, defining I_{out}. Using equation A.2, β can be calculated

$$A_f = \frac{I_{out}}{V_{ref}} = \frac{\frac{V_{ref}}{R_S}}{V_{ref}} = \frac{1}{R_S} \approx \frac{1}{\beta} \,. \tag{A.3}$$

Calculating the transconductance amplifier gain A requires a little more work and it is useful to switch to the small-signal model of the circuit. To build the small-signal model, a number of simplifications can be applied. In the same same way as it was done for the MOSFET with a source resistor in figure 1.41 on page 61, the AC component of V_{ref} can be set to zero, because it is considered constant and so can the supply voltage V. The load is also considered constant and hence shorted to ground. In order to ground V_{ref}, the non-inverting input of the MOSFET must be disconnected, because there still is the voltage v_id connected to it. The model includes the differential input resistance R_{id} between the inverting and non-inverting input of the op-amp, because for bipolar input op-amps, the differential input resistance can be as low as a few kΩ and must be considered. The common-mode input resistance of the op-amp inputs is typically serveral dozens of MΩ or higher and can be safely neglected. This leads to the small signal model shown in figure A.3. The MOSFET model is the Thévenin model introduced in figure 1.40b on page 60. Do note, that this model is for low frequencies only, as it neglects capacitive effects of the op-amp and mosfet. Capacitors are treated as having infinite impedance in this model.
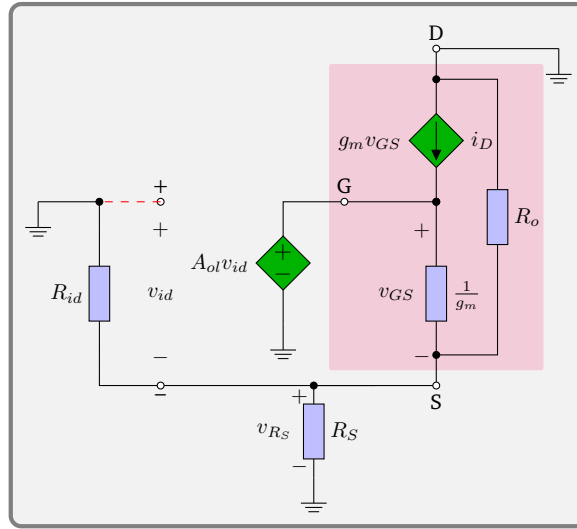

Figure A.3.: Small signal model for a transconductance amplifier with a MOSFET as shown in figure A.1

From the model in figure A.3, the following equations can be extracted in a similar fashion as it was done for the common-gate amplifier and equation 1.88 on page 61.

$$v_{GS} = A_{ol}v_{id} - V_{R_S} \tag{A.4}$$

$$V_{R_S} = i_D \left(R_o||R_S||R_{id}\right) = g_m v_{GS} \left(R_o||R_S||R_{id}\right) \tag{A.5}$$

$$A\beta = \frac{V_{R_S}}{V_{id}} = \frac{g_m v_{GS} \left(R_o||R_S||R_{id}\right)}{\frac{1}{A_{ol}}\left(1 + g_m \left(R_o||R_S||R_{id}\right)\right) v_{GS}}$$

$$= A_{ol} \frac{g_m \left(R_o||R_S||R_{id}\right)}{1 + g_m \left(R_o||R_S||R_{id}\right)} \tag{A.6}$$

Dividing by R_S yields the open-loop gain of the transconductance amplifier, a quantity, that is interesting for calculating the MOSFET noise contribution:

$$A = \frac{A_{ol}}{R_S} \frac{g_m \left(R_o||R_S||R_{id}\right)}{1 + g_m \left(R_o||R_S||R_{id}\right)} \tag{A.7}$$

This leads to the closed-loop transfer function

$$A_f = \frac{A_{ol}}{R_S} \frac{g_m \left(R_o||R_S||R_{id}\right)}{(A_{ol}+1)g_m \left(R_o||R_S||R_{id}\right)+1} \, , \tag{A.8}$$

and finally the output impedance of the transconductance amplifier can be calculated using the output impedance of the common-gate amplifier 1.89, calculated on page 61.

$$\begin{aligned} R_{out} &= (1+A\beta)\, R_{out,cg} \\ &= \left(1 + A_{ol} \frac{g_m \left(R_o||R_S||R_{id}\right)}{1+g_m \left(R_o||R_S||R_{id}\right)}\right) (g_m R_S R_o + R_o + R_S) \\ &\overset{A_{ol}\gg 1}{\approx} A_{ol} \frac{g_m \left(R_o||R_S||R_{id}\right)}{1+g_m \left(R_o||R_S||R_{id}\right)} (g_m R_S R_o + R_o + R_S) \, . \end{aligned} \tag{A.9}$$

Equation A.9 can be simplified for typical applications by approximation of $g_m \left(R_o||R_S||R_{id}\right)$. Using the example parameters for the `IRF9610` in saturation used previously on page 59 and additionally the ADI `AD797` [4] op-amp $g_m \left(R_o||R_S||R_{id}\right)$ with the following parameters

$$I_D = 250\,\mathrm{mA}\,, \lambda = 4\,\mathrm{mV}^{-1}\,, V_{DS} = 3.5\,\mathrm{V}\,, R_S = 30\,\Omega\,,$$
$$R_{id} = 7.5\,\mathrm{k\Omega}\,, \kappa = 0.813\,\mathrm{A\,V}^{-2}\,, A_{ol} = 20\,\mathrm{V\,\mu V}^{-1}$$

one finds

$$R_o = \frac{I_D}{\frac{1}{\lambda}+V_{DS}} = 1014\,\Omega$$
$$g_m = \sqrt{2\kappa I_D \left(1+\lambda V_{DS}\right)} = 0.642\,\mathrm{S}$$
$$g_m \left(R_o||R_S||R_{id}\right) \approx g_m R_S \approx 29.03$$
$$\frac{g_m \left(R_o||R_S||R_{id}\right)}{1+g_m \left(R_o||R_S||R_{id}\right)} \approx 0.97$$

Using typical parameters, it can be seen, that dropping the $\frac{g_m(R_o||R_S||R_{id})}{1+g_m(R_o||R_S||R_{id})}$ term will only lead to error of about $3\,\%$. Given the datasheet uncertainties for the MOSFET related parameters on the order of $50\,\%$–$100\,\%$, it can be safely neglected, leading to the following approximations

$$R_{out} \approx A_{ol} \left(g_m R_o R_S + R_o + R_S\right) \tag{A.10}$$
$$A_f \approx \frac{1}{R_S} \, .$$

The approximation for the output impedance holds true when $g_m R_S \gg 1$, which typically is the case. While R_S might become small, this is compensated by an increase in g_m in our application, because a smaller source resistor implies a higher output current, demanding a MOSFET with a higher transconductance, so the product of $g_m R_S$ remains constant.

It can therefore be said, that the op-amp is simply amplifying the output impedance of the MOSFET with the source resistor and the closed-loop gain is defined entirely by R_S, a very convenient property.

If the model is to be considered at frequencies $\omega > 0$, A_{ol} can be replaced by the first order approximation of the op-amp gain as

$$A_1(\omega) = \frac{A_{ol}}{\sqrt{1+\left(\frac{\omega}{\omega_c}\right)^2}} \, , \tag{A.11}$$

which is valid for most compensated op-amps, which have a dominant pole at $\omega_c \approx 1\,\mathrm{Hz}$.

A.4. Simulating Current Source Properties in LTSpice

This section explains some more advanced concepts of LTSpice [81] to simulate device properties and circuit properties used when working with the current source presented in section 1.8.4. This section does not aim at explaining the basic functions of LTSpice, but rather some special functions. It is left to the interested reader to acquire those basic skills. The example presented here, allows to generate the MOSFET *Typical Output Characteristics* plot found in datasheets, the transconductance of a MOSFET and the (dynamic) output impedance of a current source. The typical output characteristics can be used to compare the model with the datasheet or with measurements taken. Comparing these model parameters with the datasheet can establish confidence, that the simulation results can be transferred to a real circuit.

A.4.1. MOSFET Typical Output Characteristics

The output characterisic is a graph found in all MOSFET datasheets and is shown below in figure A.4.
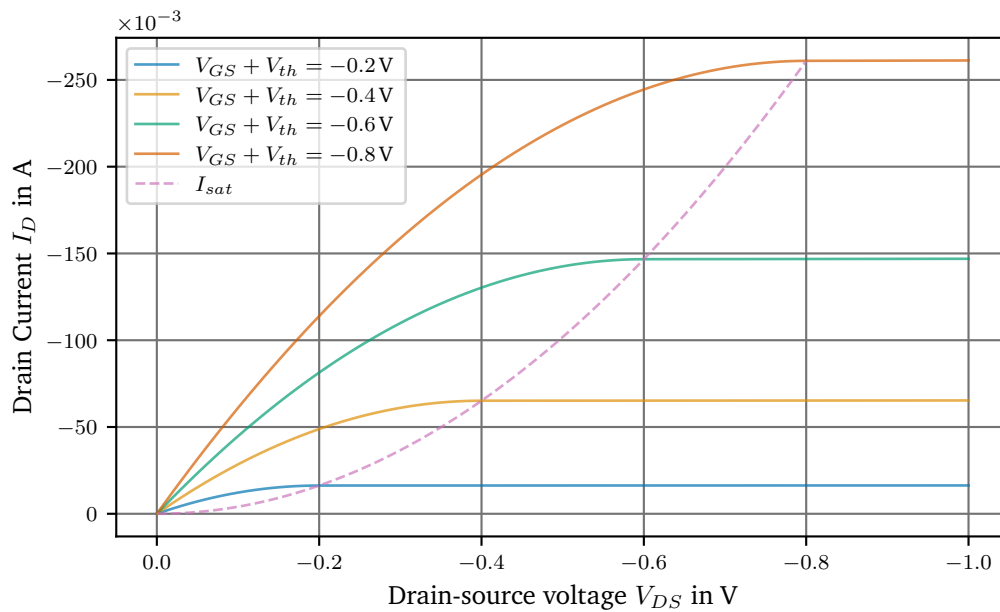

Figure A.4.: Simulated drain current over the drain-source voltage, also called output characteristics of a MOSFET.

Plotting this graph allows to compare the model to the datasheet or the measured values in order to tweak the model. To create this graph the simulation file found in the folder source/spice/mosfet_gm-id.asc as part of this document can be used. The SPICE simulation for the output characteristics of the MOSFET simulate the following circuit shown in figure A.5a.

Do note, that the V_{DS} and V_{GS} are inverted and given as V_{SD} and V_{SG}. The reason is, that the plotter in LTSpice works better with positive numbers to guess the correct scaling of the axis. Figure A.5b shows the same circuit drawn in LTSpice. The MOSFET parameters are entered using the **.model** syntax
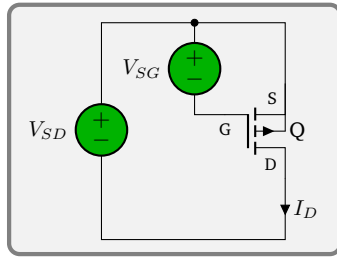
(a) P-channel MOSFET under test.
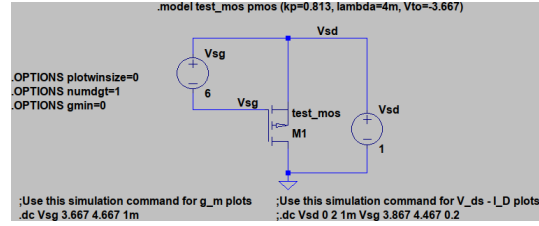
(b) LTSpice model.

Figure A.5.: P-channel MOSFET circuit and its LTSpice model.

> **.model** test_mos pmos (kp=0.813, lambda=4m, Vto=−3.667)

with the parameters $\kappa = 0.813\,\mathrm{A\,V^{-2}}$, $\lambda = 4\,\mathrm{mV^{-1}}$ and $V_{th} = -3.667\,\mathrm{V}$. The options **plotwinsize** and **numdgt** make sure, that LTSpice does not compress the output data and increases the floating point precision. This is important, because I_D spans a large range values. Setting **gmin** to 0 prevents LTSpice from adding small transconductance to every pn-junction, thus changing the MOSFET model. Finally, the most important command is the **.dc** command, which instructs LTSpice to step the voltage sources V_{SD} and V_{SG} to evaluate I_D over V_{SD}. The command

> **.dc** Vsd 0 2 1m Vsg 3.867 4.467 0.2

steps the voltage source V_{SD} from $0\,\mathrm{V} - -2\,\mathrm{V}$ in steps of $10\,\mathrm{mV}$ and for each step of V_{SD}, steps V_{SG} from $0.2\,\mathrm{V} - V_{th}$–$0.8\,\mathrm{V} - V_{th}$ in steps of $200\,\mathrm{mV}$. Plotting

> Id(M1)

results in the plot shown in figure A.4, which can be found in datasheets as the *Typical Output Characteristics* plot. To draw a line in the graph showing the point where the MOSFET enters the saturation region, denoted I_{sat} in figure A.4, as given by equation 1.77, add the following plot command to the graphing window and rescale the axis.

> 0.5*0.813*1A/1V**2*V(vsd)**2

This command must be adjusted for the value of κ and do note, that κ is entered with units of $\mathrm{A/V^2}$ to correctly display the output in A.

A.4.2. MOSFET Transconductance

Another interesting property to plot is the transconductance g_m of the MOSFET. Again, using the same model used previously in figure A.5b and from equation 1.78 we known that g_m is defined as

$$g_m = \left. \frac{\partial I_D}{\partial V_{GS}} \right|_{V_{DS}=const} .$$

To derive g_m, we need to generate values of $I_D(V_{GS})$. This can again be done by stepping V_{GS}

> **.dc** Vsg 3.667 4.667 1m

To produce a smooth plot, the the steps size of V_{SG} was decreased to $1\,\mathrm{mV}$. V_{DS} is now fixed and can be set using the voltage source V_{SD}. The MOSFET is intentionally biased into the saturation region at $V_{DS} = -1\,\mathrm{V}$ as can be seen in figure A.4.

LTSpice is now able to numerically differentiate the data, which can be invoked by by plotting

```
−d(Id(M1))
```

The minus sign comes from the inverted $V_{SG} = -V_{GS}$. To plot g_m over I_D, the formula for g_m given above needs to entered manually into the *Expression Editor* by right clicking the expression label on top of the graph. Finally, the x-axis must be changed to $Id(M1)$, leading to the plot in figure A.6.
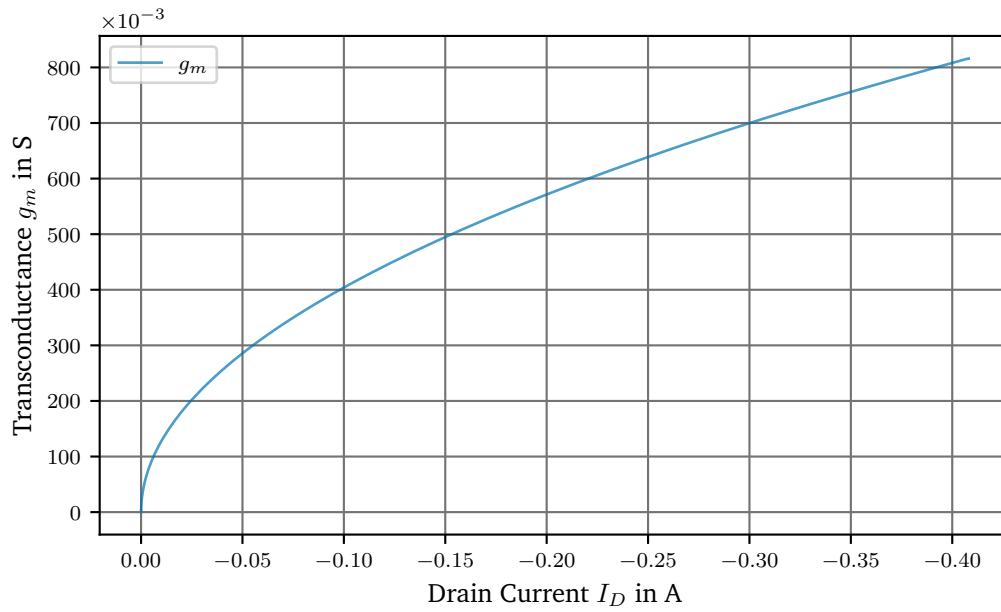

Figure A.6.: Simulated transconductance in saturation at $V_{DS} = -1\,\mathrm{V}$.

As expected from equation 1.78, g_m is proportional to the square root of I_D when the MOSFET is in saturation.

As a sidenote, if the MOSFET model includes gate leakage, this leakage current may influce the calculation of g_m, especially at very low currents. In this case, it it better to plot the positive derivative of the source current $Is(M1)$, which does not include the leakage current.

```
d(Is(M1))
```

A.4.3. Output Impedance

This sections will explain how to calculate the dyncamic output impedance using LTSpice. The example circuit used, is the precision current source from section 1.8.4. The dynamic output impedance was defined in equation 1.81 as the inverse of the conductance leading to

$$R_{out} = \frac{1}{\frac{\partial I_D}{\partial V_{DS}}} \, .$$

Using the technique presented in the previous section, the obvious solution would be to again use the **.dc** sweep command and then numerically differentiate the result. Unfortunately this will lead to disappointing results, because the output impedance in question is very large and the limits of the numerical precision will be reached, nicely demonstrating the boundries of numercial methods. LTSpice allows to increase the numeric precision to double using the option **numdgt**

.**options** numdgt=15

Unfortunately, this only forces LTSpice to interally use the double floating point number format, which does have a precision of 53 bit which means $\log_{10}\left(2^{53}\right) = 15.95$ decimals. So instead of using the large-signal model of the MOSFET, it becomes more convenient to evaluate the small-signal model

$$R_{out} = \frac{v_{load}}{i_D} = \frac{v_{DS}}{i_D}$$

at several different points of V_{DS}, therefore reconstructing the large-signal model from rasterized versions of the small-signal model. For the small-signal model, $v_{DS} = v_{load}$, because the supply voltage and the voltage across the sense resistor can be considered constant, so any change in the voltage across the load must cause the opposite change in the source-drain voltage $v_{SD} = -v_{DS}$.

To run this simulation the small-signal simulation must be used and additionally some commands not available through the graphical user interface need to be entered by hand.

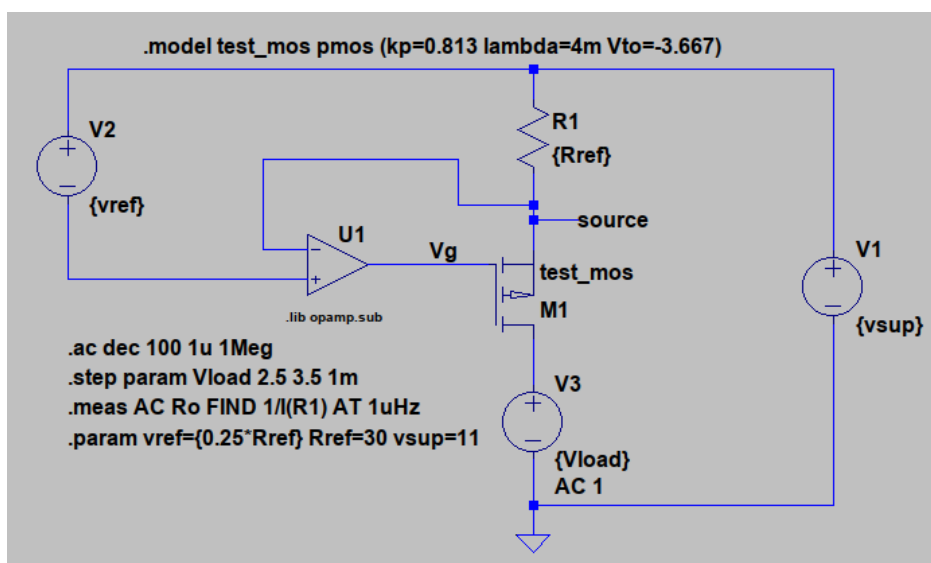The LTSpice simulation is shown in figure A.7 and will now be explored.

Figure A.7.: LTSpice model.

The simulation uses the same MOSFET model as above and adds an ideal op-amp to control the loop. The op-amp model has a open-loop gain of 2×10^6 and a gain-bandwidth product of $10\,\mathrm{MHz}$ as can be approximated from the the datasheet of the AD797 [4] and is also given in table 1.9. This leads to a $3\,\mathrm{dB}$ corner frequency of $5\,\mathrm{Hz}$, which will be interesting later.

To access the small-signal model the **.ac** command is used, because LTSpice uses the small-signal model to calculate the ac response of a circuit at a given working point. The command

```
.ac dec 100 1u 1Meg
```

calculates the ac response from $1\,\mu\text{Hz}$ to $1\,\text{MHz}$ with 100 points/decade. Additionally, as discussed, the load will be stepped, by stepping voltage source in the source leg of the MOSFET. We use a voltage source in this case instead of a resistor, because the AC impedance of a laser diode is typically very small. For the working point, it does not matter whether V_{load} is resistive or not. To step the voltage source, the command

```
.step param Vload 2.5 3.5 1m
```

is used to change V_{load} from $2.5\,\text{V}$–$3.5\,\text{V}$ in steps of $1\,\text{mV}$, which is exactly the maximum V_{DS}, which is $V_{sup} - V_{ref} = 3.5\,\text{V}$. This is done to show the effect of the complete loss of regulation. The last thing to do, is to extract the desired output impedance from the many stepped small-signal simulations. This can be done using the **.meas** command telling LTSpice to save a single value at certain frequency from each step.

```
.meas AC Ro FIND 1/I(R1) AT 1uHz
```

The **.meas** command shown will save the value of $\frac{1}{i_D} = \frac{1}{I(R1)}$ at $1\,\mu\text{Hz}$ to the (error) log file whenever the **.ac** command is run. The value of v_{DS} was already set to $1\,\text{V}_{\text{rms}}$ in the LTSpice simulation as shown in figure A.7, thus $\frac{1\,\text{V}}{I(R1)} = R_{out}$. The current through sense resistor instead of i_D was chosen because it is numerically more stable and since there is no gate current it is the same as i_D. The frequency were the R_{out} is measured was chosen to be well below the corner frequency of the op-gain, which was calculated above to be $5\,\text{Hz}$. This gives the near DC output impedance of the current source.

To plot the values are stored in the log file, click on *View* in the top menu, then *SPICE Error Log*. Now right-click on the error log and select *Plot stepp'ed .meas data*. This will open a new plot window showing the output impedance curve.

Those results are discussed in more detail in section 1.8.5.

A.5. MOSFET Noise Sources

This section gives the reader a quick overlook of the noise sources found in MOSFETs. A good overview of different types of noise in MOSFETs can also be found in [73] and goes beyond the scope presented here.

The MOSFET wideband noise can be attributed to thermal noise in the channel [29]. Der Ziel developed a model for the thermal noise in the saturation region of the MOSFET, while the classic Johnson–Nyquist noise [53] can be used for the ohmic region as it behaves like a voltage controlled resistor. This results in the noise density of

$$i_{n,thermal} = \begin{cases} \sqrt{4k_B T \frac{2}{3} g_m} & \text{saturation} \\ \sqrt{4k_B T g_{DS}} & \text{ohmic} \end{cases} \tag{A.12}$$

Using the example parameters from table 1.9, one finds

$$g_m = \sqrt{2\kappa I_D (1 + \lambda V_{DS})} = 0.642\,\text{S}$$
$$T = 25\,^\circ\text{C}$$
$$i_{n,thermal} \approx 83.9\,\text{pA}/\sqrt{\text{Hz}}, \tag{A.13}$$

the equivalent noise of a resistor $R_D = \frac{3}{2g_m} = 2.3\,\Omega$.

A more detailed analysis, which also points out the limits of the model above can be found in [117].

Additionally the MOSFET also suffers from shot noise due to leakage through the gate, but this can be neglected because this leakage current is very small and even a relatively large current of $1\,\text{mA}$ only produces

$$i_{n,shot}^2 = \sqrt{2eI_D} \tag{A.14}$$
$$\approx 1.8\,\text{pA}/\sqrt{\text{Hz}}. \tag{A.15}$$

Shot noise becomes interesting, when the MOSFET is used well below threshold or at higher frequencies, because, then the parasitic gate-drain capacitor C_{GD} will leak from the input to the ouptut as can be seen in figure A.8. Figure A.8 shows the different parasitic capacitances of a MOSFET.
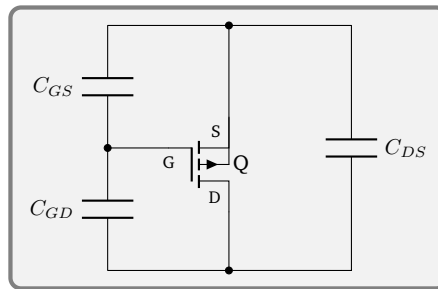

Figure A.8.: Parasitic capacitances of a MOSFET.

These capacitances can also be found in datasheets, although not directly, because there they are defined as

$$C_{iss} = C_{GD} + C_{GS} \quad \text{input capacitance} \tag{A.16}$$
$$C_{oss} = C_{DS} + C_{GD} \quad \text{output capacitance} \tag{A.17}$$
$$C_{rss} = C_{GD} \quad\quad\quad \text{reverse transfer capacitance}. \tag{A.18}$$

Regarding low frequencies, MOSFETs also show strong flicker noise. We know from section 1.6.1, that the sources of flicker noise are not clearly understood, so there are several theories regarding flicker noise models for MOSFETs.

An an empirical model given by [73, 93] can be used to describe the flicker noise as

$$i_{n,flicker} = \sqrt{\frac{K_f I_D}{C_{ox} L^2} \frac{1}{f}} \, . \tag{A.19}$$

This model is presented here, because it is also supported and easy to implement in LTSpice. While the parameter K_f is approximately 2×10^{-10} fC^2/µm^2 [73] for p-channel MOSFETs, the gate width and length W, L are device specific und unfortunately not given by the manufacturers. The typical corner frequency for MOSFETs, though, is between a few hundred kHz and a few dozen MHz depending on the size of the transistor. Larger transistors tend to show lower noise. Hence older processes are prefered in this regard. Given that the noise is uncorrelated, the total noise of the MOSFET in saturation can be written as

$$i_n = \sqrt{4k_B T \frac{2}{3} g_m + \frac{K_f I_D}{C_{ox} L^2} \frac{1}{f}} \tag{A.20}$$

As a reminder, the MOSFET is a (transconductance) amplifier, that takes a voltage at the input and outputs a current. To make the noise figures comparable, the noise is divided by the gain g_m. This called the input referred noise. The input referred (voltage) noise e_n is given by:

$$e_{n,thermal} = \sqrt{4k_B T \frac{2}{3 g_m}} \tag{A.21}$$

$$e_{n,flicker} \overset{1.79}{\approx} \frac{K_f}{2\kappa C_{ox} L^2} \frac{1}{f} \tag{A.22}$$

We can see, that flicker noise is fully determined by process parameters in this model.

A.6. Building an Injection Transformer

Typically devices in the lab are supplied with a positive and a negative voltage – usually ±15 V. This is readily achieved using two floating outputs of a power supply and connecting them in series, then taping off the center as the common voltage around which the ±15 V is centered.

This setup requires a posive and a negative line injector like the positive injector PB02 presented in [15] and the negative Picotest J2123A line injector. When driving these injectors it is desirable to drive them both from a single output of a VNA. The Picotest Bode 100 used for many low frequency applications does not have galvanically isolated inputs and outputs. This can be achieved using a transformer to drive the injectors. Additionally using a transformer it is easy to create two outputs, that are $180°$ out phase.

The Bode 100 covers a frequency range from 1 Hz to 50 MHz. While the whole range is a bit too much to ask, because the low frequency end likely requires an iron core to keep the transformer out of saturation, which then limits the high frequency domain. This transformer aims to cover most of the range, while accepting a limited performance at the corners.

This section does not aim to explain the intricate details of the parasitic effects of different types of transformers, their geometry and materials. The interested reader may look up [100] for those details. The aim of this section is a simple instruction manual to allow the reader to build an affordable alternative to the fairly expensive commercial solutions, with similar performance.

The materials required are:

- A box like the the Hammond 1590B.

- A nanocrystalline ferrite core, a Vacuumschmelze T60006-L2040-W452 or T60006-L2040-W424.

- 3 m of Cat5e Ethernet cable. Preferably FEP insulated like Belden 7928A, but any other will also do.

- 2–3 isolated BNC connectors like the Amphenol 031-10-RFXG1. You will need 3 connectors for the center tapped version and 2 for a 1:1 transformer.

- 1 Cinch Connectivity Solutions 111-2223-001 earthing connector.

- Drills in sizes 6 mm and 9.7 mm.

- Kapton tape

The author used a Vacuumschmelze T60006-L2040-W452, because it was available at the time, but the T60006-L2040-W424 might be a better choice, because of its higher inductance (12.2 µH at 10 kHz vs 101 µH at 10 kHz). The T60006-L2040-W452 has a slightly smaller inner diameter (25 mm vs 32 mm), so less windings will fit onto the core, so this will offset some of the higher inductance of the core, but fewer windings also reduce the inter-winding capacitance due to the shorter cable length.

The target is 46 turns of the twisted pair cable around the core for a T60006-L2040-W452. This should give a tight fit. When center tapping the transformer do make sure to accurately count and then exactly cut one wire in the center. Do not cut the wire in advance, because you will need to leave some overhead at the beginning to leave plenty of room to solder the cable to the BNC connectors.

When done winding the transformer, wrap it with Kapton tape to secure the windings. It is recommended to test it before final assembly. Carfully solder the BNC connectors to the ends and test it with a VNA,. These connectors will later be remove again. Make sure to calibrate the VNA beforehand and when the transformer matches the requirements, it is time to mount it in the box.

The box require one $6\,\text{mm}$ hole for the earthing connector and 3 $9.7\,\text{mm}$ holes for the BNC connectors.