
Projektseminar Beschleunigertechnik: Ausarbeitung

Iterative Optimierung eines Hammerstein-Modells und Code-Design für ein
Barrier-Bucket-System

Artem Moskalew, Jonas Christ, Maximilian Nolte
Darmstadt, August 2018



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Inhaltsverzeichnis

1	Abstract	2
2	Einführung	3
2.1	Modell und Konvention	3
2.1.1	Hammerstein Modell	4
2.2	Motivation	4
2.3	Aufgabenstellung	4
3	Vorgehen	5
3.1	Vorhandene Implementierung	5
3.1.1	Python	5
3.1.2	MATLAB	5
3.2	Funktionalität reproduzieren	5
3.3	Funktionalität erweitern	6
3.3.1	Erste Optimierungs Idee	6
3.4	Gerätekommunikation	7
3.5	Kritische Punkte	8
4	Code-Design	10
4.1	Motivation	10
4.2	Aufbau	10
4.3	Methodik	11
4.3.1	Test Driven Development	11
4.3.2	Mock-System	12
5	Iterative Optimierung des Hammerstein-Modells	13
5.1	Optimierung der linearen Übertragungsfunktion H	13
5.1.1	Umgang mit Fehlerquellen	15
5.1.2	Auswertung	16
5.2	Optimierung der nichtlinearen Kennlinie K	19
5.2.1	Erster Ansatz	19
5.2.2	Zweiter Ansatz	22
6	Fazit	23
7	Ausblick	24
7.1	Impressionen zum Code	24
7.2	Ausblick: Optimierung	24
7.3	— Gerätekomm—	25
8	Anhang	27

1 Abstract

Für die sinnvolle Nutzung der Barrier-Bucket-Kavität an den Speicherringen der GSI Darmstadt ist ein qualitativ hochwertiger Einzelsinus als Ausgangsspannung am Gap unabdingbar. Modelliert man den Messaufbau durch ein Hammerstein-Modell mit einer nichtlinearen Kennlinie im Zeitbereich und einer linearen Übertragungsfunktion im Frequenzbereich, so verlagert sich die Schwierigkeit auf das hinreichend genaue Bestimmen dieser Komponenten. Diese lassen sich zwar einzeln über automatisierte Messungen erfassen, sind dabei aber nicht unabhängig voneinander. Um die Modellierung zu verbessern liegt eine iterative Anpassung der beiden Bausteine nahe.

In einem ersten Ansatz zur Optimierung werden für Kennlinie und Übertragungsfunktion getrennte Algorithmen genutzt. Es zeigt sich am simulierten idealen Hammerstein-Modell, dass dieser Ansatz die richtige Richtung einschlägt. Für eine merkliche Verbesserung der Modellierung auch am realen System sind jedoch noch eine Reihe von Hürden zu nehmen. Diese sind teilweise auf Diskretisierung und Interpolation zurückzuführen. Jedoch liegen gerade bei der nichtlinearen Kennlinie auch konzeptionelle Probleme bezüglich des Definitionsbereichs vor, die eine genauere Betrachtung nicht nur für die Optimierung erfordern.

Unverzichtbare Grundlage der hier verfolgten Ansätze sowie Sprungbrett für die weitere Arbeit an einer Optimierung ist das Design einer Programmstruktur, die an den Kontext eines Messaufbaus angepasst ist. Als außerordentlich hilfreich erweist sich ein modularer Aufbau von Funktionalitäten. Zusätzlich bietet sich ein Programmier-Konzept an, das auf periodische Tests der Funktionalitäten ausgelegt ist. Ergänzend hierzu vereinfacht sich die Erweiterung der Funktionalität durch die Möglichkeit der Simulation des mathematischen Modells erheblich.

2 Einführung

Mit den Barrier Bucket (BB) RF Systemen können am neu entstehenden Synchrotron SIS100 oder im Experimentier Speicherring (ESR) am GSI Helmholtzzentrum viele longitudinale Manipulationen am Teilchenstrahl vorgenommen werden. Der dazu notwendige Spannungspuls hat die Form wie in Abb. (2.1) dargestellt. Wenn die Wiederholfrequenz des Spannungspulses gleich der Umlauffrequenz ist, so wird im Phasenraum eine stationäre Potential Barriere erstellt. Wenn die Wiederholfrequenz nicht gleich der Umlauffrequenz ist verschiebt sich die Potential Barriere im Phasenraum und es entstehen Bunches mit unterschiedlicher Länge.

Der Anspruch an diese Systeme liegt darin eine hohe Qualität des Impulses am Gap der Kavität zu erzeugen, damit keine unerwünscht Microbunches entstehen. Dafür müssen die Nachschwinger nach dem Einzelsinus kleiner als 2,5% von \hat{U}_{BB} sein.

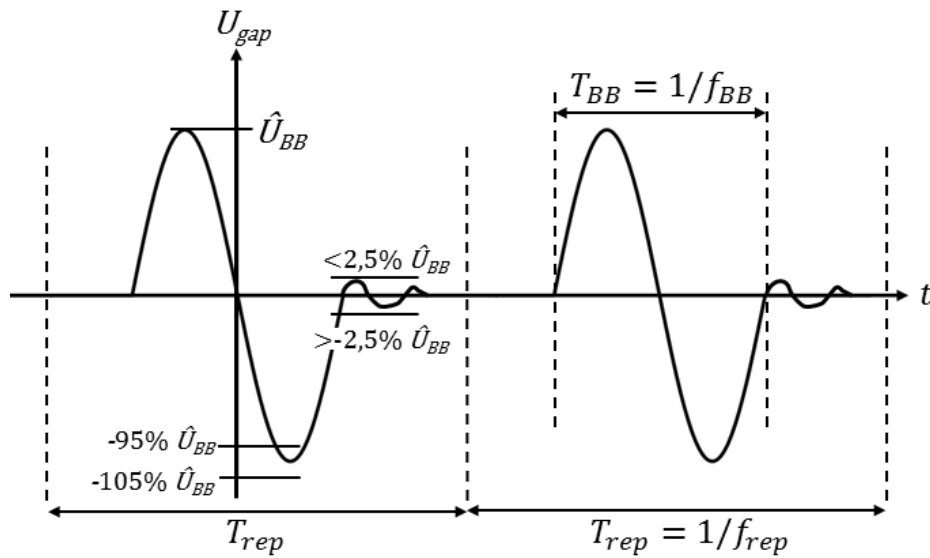


Abbildung 2.1: Ausgangssignal

Dabei wird im Folgenden f_{rep} für die Wiederholfrequenz des Spannungssignals verwendet und die Barrier Bucket Frequenz f_{BB} bestimmt die Breite der Potential Barriere. Für die Messungen in diesem Projektseminar wurden $f_{rep} = 900\text{kHz}$ und $f_{BB} = 5\text{MHz}$ verwendet.

2.1 Versuchsaufbau und Modell

Der Prototyp des ESR BB besteht aus einem Funktionsgenerator (Keysight 3600A series 2-channel AWG), einem Verstärker (AR1000A225) und der Breitband Ringkern Kavität. Der Versuchsaufbau ist in Abb. (2.2) gezeigt.

Dabei kann angenommen werden, dass sich das System bis $\hat{U}_{BB} = 550\text{V}$ annähernd linear verhält und durch die Übertragungsfunktion \underline{H} beschrieben werden kann. Eine mathematische Modellierung ist ebenfalls in Abb. (2.2) gegeben, bei der zur linearen Übertragungsfunktion \underline{H} noch eine nichtlineare Kennlinie enthalten ist. Die Größen werden im nächsten Abschnitt erklärt.

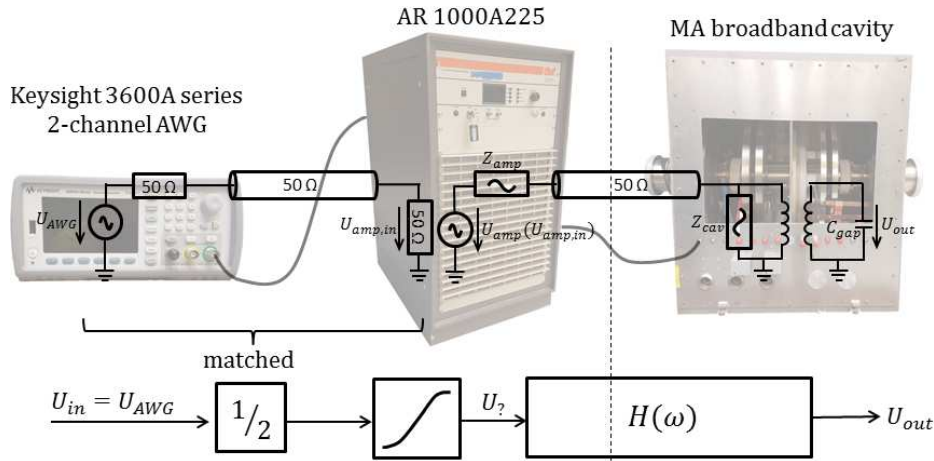


Abbildung 2.2: Versuchsaufbau und Modell

2.1.1 Hammerstein Modell

In Abb. (2.3) sind alle von uns verwendeten Größen dargestellt. U_{in} ist die Eingangsspannung vom Funktionsgenerator. Der erste Block in Abb. (2.3) stellt die nichtlineare Modellierung dar. Die Spannung U_{γ} ist eine rein virtuelle Größe und kann wie in Glg. (2.1) über die Koeffizienten a_n berechnet werden. Bei der Potenzreihe hatte $N = 3$ in der MATLAB Implementierung schon gute Ergebnisse geliefert und wurde von uns auch weiterhin so verwendet. Die nichtlinearen Kennlinie im Folgenden nur noch als K bezeichnet wird als Look-Up Tabelle in unserem Programm hinterlegt. Die Spannung U_{out} ist die Gap Spannung in der Kavität.

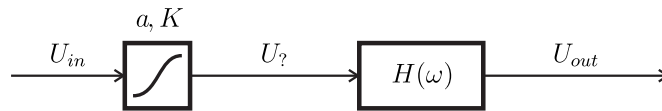


Abbildung 2.3: Hammerstein Modell

$$U_{\gamma}(t) = \sum_{n=1}^N a_n [U_{in}(t)]^n \quad U_{out}(t) = \mathcal{F}^{-1} \{ \underline{H}(\omega) \cdot \underline{U}_{in}(\omega) \} \quad (2.1)$$

2.2 Motivation

Dieses Modell wurde bis auf die Berechnung von K schon erfolgreich in Python implementiert, dabei sei auf das Projektseminar [2] verwiesen. Deshalb lag unsere Motivation darin dieses Modell in Python zu vervollständigen, um dann einen möglichen Optimierungsansatz aufzustellen.

2.3 Aufgabenstellung

Im Rahmen dieses Projektseminars sollte eine iterative Optimierung der Vorverzerrung auf Basis einer Hammersteinmodellierung in Python implementiert werden. Dabei sollte das Tool die nichtlineare Kennlinie K und die Übertragungsfunktion \underline{H} wechselseitig optimieren, wobei eine optimale Gewichtung der Parameter zur Optimierung noch nicht erreicht werden musste.

3 Vorgehen

In diesem Kapitel werden die wichtigsten Punkte in der Weiterentwicklung des vorhandenen Tools erklärt. Dabei wird die Methodik vorgestellt mit der der vorhandenen Weg fortgesetzt wurde im Hinblick darauf, dass andere Personen daran anschließen können.

3.1 Vorhandene Implementierung

Das Projekt setzt an 2 Punkten an. Zum einen das Projektseminar [2], in dem der Grundstein in Python gelegt wurde und zum anderen eine funktionierende Implementierung des nichtlinearen Bestandteils des Modells aus Abb. (2.3) nach der Vorstellung aus [3] in MATLAB.

3.1.1 Python

Das vorhandene Python Tool kann mit einem Pseudorauschen die Übertragungsfunktion des Systems im linearen Bereich bestimmen. Diese Funktionalität wurde vollständig übernommen. Darüber hinaus waren folgende Funktionalitäten implementiert:

- \underline{H} auf das Spektrum von $U_{out,ideal}$ anzuwenden.
- Signale an das AWG zu senden und vom Oszilloskop zu messen.
- aus gegebenen Koeffizienten a_n eine Kennlinie K als Look-Up Tabelle zu berechnen.

3.1.2 MATLAB

Als zweiten Ansatzpunkt hatten wir funktionierende Methoden aus der MATLAB Implementierung des nichtlinearen Teils mit folgenden Funktionalitäten:

- \underline{H} auf das Spektrum von $U_{out,ideal}$ anzuwenden.
- a_n für die nichtlineare Kennlinie zu berechnen.
- K als Look-Up Tabelle aufzustellen.
- ein vorverzerrtes Eingangssignal über K zu berechnen.

3.2 Funktionalität reproduzieren

Als ersten wurde versucht die vorhandene Implementierung zu überblicken und letztendlich beide in Python zusammenzuführen. Dabei wurde eine starke Kopplung innerhalb der vorhandenen Implementierungen festgestellt, was die Fehlersuche und Fortführung erschwerte. Ein paar Funktionen hatten teilweise auch mehrere logische Aufgaben, die für eine bessere Verständlichkeit getrennt werden sollten. Mit dem Code wurden aber schon gute Ergebnisse erzielt und das erste Ziel war, den Code in eine neue Form zu überführen, ohne die Funktionalität zu beeinträchtigen.

Neues Design

Das in Abb. (2.3) dargestellte Modell wird in zwei Blöcke gegliedert und so wurde ein Code-Design gewählt, das diese Blöcke repräsentiert. Dafür bietet sich ein modulares Design an, das in Kap. 4 genauer erklärt ist und dabei möglichst selbsterklärend dieses Modell abbildet. Dieses Design wurde auch in Hinblick auf die Team Arbeit gewählt, damit an mehreren logisch getrennten Abschnitten gleichzeitig gearbeitet werden kann.

Dafür wurde eine Github Repository erstellt, um eine Versionen Kontrolle vorzunehmen.

Tests

Aus vorherigen Messungen und den Ergebnissen aus beiden Implementierungen konnten Unit-Tests erstellt werden, die die geforderte Funktionalität beschreiben und somit die Ergebnisse aus dem neuen Design mit schon bestehenden Ergebnissen evaluieren.

Ziel war es den Code umzustrukturieren mit der Bedingung, dass die Funktionalität nicht verändert wird. Die bestehende Implementierung wurde in mehrere logische Abschnitte unterteilt, bei denen jeder Abschnitt für sich die geforderten Daten erzeugen sollte. Dabei sollte die funktionierenden Teile aus MATLAB in Python überführt werden.

Fehlersuche

Die Möglichkeiten der Fehlersuche soll an zwei Beispielen erklärt werden. In der Python Implementierung aus dem Projektseminar [2] war schon ein erster Entwurf zur Berechnung der Koeffizienten a_n gegeben, der aber noch keine guten Werte berechnete. Die Ergebnisse aus der MATLAB Implementierung wurden in den Unit Test importiert, sodass in Python mehrere Möglichkeiten getestet wurden, bis die Ergebnisse mit den schon bestehenden Werten übereinstimmen.

Das zweite Beispiel ist die Berechnung des Eingangssignals über K . Dabei wurden während einer Messung Daten gespeichert, die zu einem Programmabbruch führten. Diese Daten wurden in einem Unit Test zusammengeführt, damit der Fehler im Programm ohne Messaufbau behoben werden konnte. Einzelheiten dazu im Abschnitt (5.2). Des weiteren wird durch das neue Design eine strukturierte Fehlersucher ermöglicht. Wenn eine Veränderung an einem Block vorgenommen wird, so lässt sich diese isoliert betrachten und sollte keinen Einfluss auf die Logik anderer Blöcke haben.

3.3 Funktionalität erweitern

Nachdem die Funktionalitäten beider Implementierungen in Python zusammengeführt wurden konnten im nächsten Schritt Erweiterungen der vorhandenen Funktionalität vorgenommen werden. Dabei stellte das neue Design mit seinem modularen Aufbau eine gute Grundlage dar, um an den logisch getrennten Aufbau anzuschließen und schon vorhandene Funktionalität zu nutzen.

3.3.1 Erste Optimierungs Idee

Als nächster Schritt wurde ein erster Optimierungsansatz implementiert, wie er in der Aufgabenstellung aus Abschnitt (2.3) vorgesehen war. Genauere Erläuterungen folgen in Kap. 5. Der Ansatz wurde mit zwei Routinen getestet und mit dem RF-Tool zur Güteberechnung des Ausgangssignals evaluiert.

Das Tool befand sich zu diesem Zeitpunkt noch in der Entwicklung und Testphase, sodass erst einmal herausgefunden werden musste, wie das Tool die Daten verarbeiten kann.

Bei der Nutzung des Programms ist eine längere Wartezeit für die Geräte aufgefallen, die bei kleineren

Durchlaufen nicht so ins Gewicht fällt aber im Hinblick auf mehr Iterationsschritten noch optimiert werden sollte. Dazu werden im folgenden Abschnitt Ansätze vorgestellt.

3.4 Gerätekommunikation

Um Messergebnisse sinnvoll für die Verarbeitung in Python aufzunehmen, ist eine effiziente Kommunikation zwischen den Messgeräten und dem angeschlossenen Computer essentiell. Besonders relevant ist hier die zeitliche Abstimmung zwischen Computer und Gerät sowie die bestmögliche Nutzung der Genauigkeiten der verwendeten Geräte. Hierbei wurden die zu Beginn vorliegenden Implementierungen für die Gerätekommunikation erweitert, wobei vor allem

1. die Optimierung der Laufzeit beim Schreiben und Lesen von Gerätewerten sowie
2. die Anpassung der Darstellung des Oszilloskops an die gemessenen Daten für eine höhere Genauigkeit

Ziele der Anpassungen waren.

Die Fernsteuerung von Messgeräten, der `Remote`-Modus, wird standardmäßig durch das Visa-Protokoll und im vorliegenden Fall durch die Implementierung von National Instruments, NIVisa, geregelt. Die Einbindung in Python wird durch die Erweiterung PyVisa ermöglicht. Übertragen werden Befehle in Form standardisierter Befehlsstrukturen, den SCPI (Standard Commands for Programmable Instruments). Dem inneren Aufbau von Messgeräten zugrunde liegt der IEEE-488 Standard, sodass unabhängig vom konkreten Gerät oder der Art der Verbindung eine Reihe von Befehlen existiert, die gemäß IEEE-488.2 gebräuchliche Standards bieten. Diese Befehle sind in der Form `*COMMAND` gehalten [5, S. 224 ff.]. Weiterhin bieten die Geräte spezifische Befehle zur Manipulation der Einstellungen, die in großer Analogie zur direkten Benutzeroberfläche formuliert sind. Diese werden nach Kontext gegliedert in sogenannte Subsysteme oder Command Groups.

Es wird ein Keysight 33622A als Signalgenerator (AWG) und ein Oszilloskop der TDS 5000 Serie von Tektronix genutzt.

Laufzeitoptimierung

Im vorliegenden Messaufbau ist es notwendig, beim Konfigurieren des AWG alle Einstellungen zurückzusetzen und anschließend neu zu setzen. Dabei werden Befehle nach dem First-Come-First-Serve Prinzip aus dem internen Speicher abgearbeitet. Dies führt aufgrund begrenzter Speicherkapazität und Verarbeitungsgeschwindigkeit bei einer Reihe von Befehlen zur Notwendigkeit, das Senden von Befehlen im Programm mit dem Arbeitsstatus des Gerätes zu synchronisieren. Die Größenordnung dieser Wartezeiten schwankt dabei stark. So dauert ein `Reset` des AWG bis zu 5 Sekunden und das Schreiben eines Arbiträrsignals benötigt ebenso mehr als eine Sekunde. Das Oszilloskop hingegen verarbeitet ohne im Kontext dieses Aufbaus messbare Verzögerung. Verbunden damit, dass das Oszilloskop auf absehbare Zeit durch das neuere MAUI von Teledyne LeCroy ersetzt wird, wird im Folgenden nicht weiter auf die Laufzeit dieses Gerätes eingegangen. Durch diese Feststellung konnte die Laufzeit je Lese-Aufruf bereits um ungefähr 15 Sekunden reduziert werden. Bei den zur Sicherheit implementierten Warte-Befehlen handelt es sich um Status-Abfragen ähnlich den nachfolgend für das AWG erläuterten.

In der zu Beginn vorliegenden Implementierung lagen pauschale Wartezeiten nach einigen als besonders zeitintensiv erwarteten Befehlen sowohl in der Ansteuerung des AWG als auch des Oszilloskops vor. Insgesamt blieb die Verbesserung der Zugriffszeiten auf das AWG erfolglos. Es stellt nach wie vor das Nadelöhr des Programms dar mit etwa 25 Sekunden Laufzeit je Aufruf. Es wurden eine Reihe von IEEE-488 Standardbefehlen für diesen Zweck ausprobiert, nachfolgend mit den wichtigsten Eigenschaften:

-
1. `*BUSY?` ist kein für das AWG definierter Standardbefehl. Dieser würde 0 zurückgeben, wenn das Gerät gerade nicht mit Ausführen eines Befehls beschäftigt ist.
 2. `*WAI` ist in erster Linie ein interner Befehl, zuerst alle Befehle abzuarbeiten, bevor neue angegangen werden. Dies geschieht jedoch ohne Rückgabe an den aufrufenden Code.
 3. `*OPC?` zwingt das aufrufende Programm zum Warten auf Beenden aller Befehle im AWG und führt zu einem `VisaTimeoutError` im AWG, wenn zu lange auf die Rückgabe gewartet werden muss [7, S. 9]. Dieser Fehler konnte auch nicht im Programm aufgefangen werden, sodass er folgenlos für den Ablauf geblieben wäre.
 4. `*OPC` setzt das zugehörige Bit `Standard-Event-Register` des Gerätes, sobald alle Befehle abgearbeitet sind. Im Unterschied zu `*OPC?` wird damit jedoch nicht auch das aufrufende Programm geblockt. Jedoch muss mit einer Statusabfrage über `*ESR?` dieses Register ausgelesen werden. Die dafür benötigte Schleife konnte jedoch nicht derart über PyVisa implementiert werden, dass tatsächlich mit Sicherheit alle vorigen Befehle beendet wurden.

Als besonders erfolgversprechend erschien der letzte Ansatz 4, der jedoch auch zeitlich der letzte war. Mit einer Schleife über `*ESR?` könnte das zugehörige Bit erneut ausgelesen und eine sehr kleine pauschale Wartezeit im Programm eingebaut werden, bevor die nächste Abfrage stattfindet. Hier könnte es sich um ein Resultat der Codierung über PyVisa handeln, da hierbei mehrere Arten, einen Befehl an das Gerät zu senden, zur Verfügung stehen.

Weiterhin wurde die pauschale Wartezeit an der im Datenblatt [6, S. 21] für die Verbindung via USB angegebenen, gemessenen Ladedauer eines Arbiträrsignals orientiert. Dieser Richtwert von 1.25 s stellte sich jedoch als zu gering für den hier vorliegenden Fall heraus.

Letztlich sei erwähnt, dass die Abfrage der Fehler im AWG mittels `SYSTEM:ERROR?` möglich ist und durch das dabei erfolgte Löschen der Fehler aus dem internen Speicher nicht nur die Abfrage sondern auch das Rücksetzen des Fehlerstatus ermöglicht. Dies ist insbesondere für das Debugging vorteilhaft, wenn Befehle ausprobiert werden und die gegebenenfalls auftretenden Fehler gehandhabt werden müssen [5, S. 454]. Dies vermeidet das unter Umständen mehrmalige Trennen und Wiederherstellen der Verbindung zum AWG sowie das Löschen der Fehler an der Benutzeroberfläche.

Anpassung der horizontalen Auflösung am Oszilloskop

Um die Auflösung des Oszilloskops ausnutzen zu können, ist die Anpassung der horizontalen wie auch vertikalen Auflösung an das zu messende Signal notwendig. Während die horizontale Auflösung durch die Wiederholfrequenz des Pulses mit f_{BB} sich nahezu unabhängig des betrachteten Signals setzen lässt, muss die vertikale Skalierung an das Signal angepasst werden. Um dies etwa während einer Optimierung mit wechselnden Spannungsamplituden zu ermöglichen, sind zwei direkt aufeinander folgende Messungen notwendig. Dabei wird aus der ersten Messung die Amplitude des Signals entnommen und die Skalierung für die zweite, dann genauere Messung, dahingehend angepasst, dass das Signal bildfüllender dargestellt wird. Dieser Gewinn an Genauigkeit in der Messung stellte sich insbesondere für die Optimierung der linearen Übertragungsfunktion H als notwendig heraus.

3.5 Kritische Punkte

—— sollten wir an andere Stelle setzen — Infolge der Auseinandersetzung mit dem anfangs vorliegenden Code sowie im Zuge der Ausarbeitung sind einige Punkte aufgetreten, an deren Bedeutung die implementierte Funktionalität teilweise essentiell hing, die jedoch im Zuge der Implementierung für Probleme sorgten. Diese seien aufgrund ihrer unterschiedlichen Natur hier stichpunktartig erläutert:

-
- Bei diskreten Zeitsignalen ist zu beachten, dass die Periodendauer T nicht durch die Differenz der Zeitwerte des letzten und des ersten Wertes, sondern noch zuzüglich eines Zeitschritts zu berechnen ist, also $T = \text{time}[\text{end}] - \text{time}[0] + \text{delta_time}$, womit gelten sollte $T = N * \text{delta_time}$. Dies wirkt sich insbesondere auch auf die FFT aus, wenn die zugehörige Frequenzachse berechnet werden muss.

4 Code-Design

Die Aufgabe des Projektseminars war ein bestehendes Tool weiter zu entwickeln. Es wurde nach einer Methode gesucht an die vorhandene Funktionalität anzuschließen und fortzusetzen. Gleiches wird wahrscheinlich mit dieser Projektarbeit passieren. Deshalb wird in diesem Kapitel die Namenskonvention und die allgemeine Struktur unseres Design vorgestellt.

4.1 Motivation

Ziel ist es ein Design aufzustellen, das mit seinen Funktionen und seiner gesamten Dokumentation für sich spricht. Des weiteren wird eine lose Kopplung der einzelnen Blöcke angestrebt, damit Bestandteile ohne die Logik anderer Funktionen verstanden und bearbeitet werden können. Dies verbessert zum einen die Arbeitsaufteilung im Team zum anderen aber auch die Verständlichkeit des gesamten Programms. Ohne ein modulares Design ist eine Optimierung der einzelnen Bestandteile nur schwer umzusetzen.

4.2 Aufbau

Der Aufbau des Programms orientiert sich an dem vorgegebenen Modell aus Abb. (2.3).

Namenskonvention

In der nachfolgenden Tab. (4.1) sind die verwendeten Präfixe erklärt.

Präfix	Erklärung
adjust	Anpassung einer Größe in einem Iterationsschritt
compute	Berechnung einer Größe, ohne Kommunikation der Geräte
determine	Bestimmung einer Größe mit Kommunikation der Geräte
evaluate	Routinen mit Bewertung der Ergebnisse mit idealen Ergebnissen
generate	Erstellung eines Datensatzes
get	Aus der Implementierung von [2] übernommen für get_H
loop	Eine Iterierungsschleife im Optimierungsalgorithmus
measure	Messung einer Größe

Tabelle 4.1: Erklärung der Präfixe von Methoden

Ordnerstruktur

In Tab. (4.2) sind die Inhalte der verwendeten Ordner aufgelistet.

Ordner	Inhalt
blocks	Implementierung der abstrakten Schritte im Modell
classes	ADTs siehe 4.2
data	Alle Ergebnisse
helpers	Hilfsfunktionen nach [9]
tests	Alle Klassen mit Unit Tests
tools	RF-Tools

Tabelle 4.2: Namen der Ordner

Abstract Data Types

Bestimmte Datensätze wurden als eigene abstrakte Datentypen (ADTs) implementiert, siehe [10, Kap. 6.1]. Daraus ergeben sich folgende Vorteile:

- Allgemeine Funktionalität wird in der Klasse implementiert
- Typüberprüfung wird ermöglicht
- Implementierung generischer Klassen¹ beispielsweise `calculate_error()` ermöglicht, das heißt die Implementierung ist unabhängig von

Für die Klasse `signal_class()` ergeben sich speziell noch weitere Vorteile:

- Vermeidung von ungewollten Manipulationen an Datensätzen, die nur schwer zu debuggen sind
- Das Originalsignal ist immer gespeichert
 - dadurch Vermeidung der Fortpflanzung von Interpolationsfehlern durch Resampling

4.3 Methodik

Für diesen Typ Aufgabenstellung konnte viel Programmierarbeit ohne den Versuchsaufbau vorgenommen werden, sodass viele Fehler schon vor dem eigentlichen Testlauf am System entdeckt und behoben werden konnten. Dafür werden zwei Möglichkeiten vorgestellt, die diese Fehlersuche ermöglicht und erleichtert haben.

4.3.1 Test Driven Development

Nach der Methode TDD werden vor dem eigentlichen Programmieren Tests geschrieben, die mit der folgenden Implementierung der Funktionalität bestanden werden müssen. Diese Tests geben ein gutes Feedback darüber, ob eine Methode gut implementiert wurde. Die folgenden Vorteile waren besonders wichtig in der Entwicklung des Codes für diese Aufgabe:

- Testen von kleinen Abschnitten erzwingt einen modularen Aufbau
- Tests dokumentieren Funktionalität
- Refactoring² wird ermöglicht
- Einfachere Fehlersuche auch ohne Versuchsaufbau möglich

¹ Implementierungen, die unabhängig vom übergebenen Datentyp sind.

² Änderung der Struktur ohne Veränderung der Funktionalität.

-
- Reproduktion von Messvorgängen, die Fehler aufgedeckt haben
 - Arbeitsteilung an logisch getrennten Abschnitten ermöglicht
 - Vergleich zwischen der Implementierung aus MATLAB und der aus Python

Eine große Schwierigkeit dieser Methode liegt darin, gute Tests zu formulieren, die auch alle möglichen Fälle abdecken. Das kann einige Zeit in Anspruch nehmen, lohnt sich aber auf längere Sicht

4.3.2 Mock-System

Das implementierte Mock System stellt ein typisches Beispiel für ein mock object nach [12] dar und simuliert das getestete reale System nach einem bestimmten Modell. In diesem Fall wird zur Simulation ein ideales Hammerstein Modell als Grundlage verwendet, weil dieses Modell verwendet wird, um das reale System abzubilden. Dieses Mock-System testet die Möglichkeiten der implementierten Methoden ein Hammerstein Modell darzustellen und zu optimieren.

Mit diesem System konnten auch Methoden getestet werden, die eigentlich mit den Geräten kommunizieren und somit nicht gut ohne Versuchsaufbau getestet werden können. Der Ablauf sieht wie folgt aus:

- Das Eingangssignal wird nicht an das AWG übergeben sondern an die Simulation
- Verzerrung über eine nichtlineare Kennlinie K
- Berechnung des Ausgangssignals über die Übertragungsfunktion H
- Das Ausgangssignal wird wieder an die eigentliche Methode übergeben

Die Optimierungsansätze aus Kap. 5 wurden bereits in der Simulation des idealen Systems getestet und es wurde festgestellt, dass diese für das ideale Modell tatsächlich zu einer Verbesserung führen. Die Frage dabei ist aber, inwieweit sich der Versuchsaufbau durch ein ideales Hammerstein-Modell beschreiben lässt und sich damit die Verbesserung auf den realen Aufbau übertragen lässt. Je nach Ergebnissen der Übertragung auf das reale System lassen sich Aussagen über die Grenzen des Modells für diesen Aufbau treffen.

Mit dem Testen an der Simulation können Aufgaben wie Debugging, Refactoring und Erweiterungen der Funktion auch ohne den Versuchsaufbau durchgeführt werden.

5 Iterative Optimierung des Hammerstein-Modells

Ziel der Optimierung von Übertragungsfunktion \underline{H} und Kennlinie K mit ihren Parametern a ist die Minimierung des Fehlers zwischen idealem und gemessenem Ausgangssignal, $U_{out,id}$ und $U_{out,meas}$. Die Minimierung des relativen Fehlers ist also gegeben durch

$$\min f(t) = \min \left(\frac{U_{out,meas} - U_{out,id}}{U_{out,id}} \right) = \min \left(\frac{U_{out,meas}}{U_{out,id}} - 1 \right). \quad (5.1)$$

Für das verwendete Hammerstein-Modell liegt die in [?] vorgeschlagene getrennte, iterative Optimierung von \underline{H} und K nahe. Die Auswertung der Qualität des Einzelsinus erfolgt dabei durch das RF-Tool von — Zitat RF-Tool — mit Entwicklungsstand vom — — unter Verwendung des als Q_{Gesamt} 1 geführten Qualitätswerts ¹.

5.1 Optimierung der linearen Übertragungsfunktion H

Die Optimierung von $\underline{H}(\omega)$ beruht auf der Annahme, dass sich Glg. (5.1) auf die Betragsspektren des berechneten und des gemessenen Ausgangssignals, $\underline{U}_{out,id}(\omega)$ und $\underline{U}_{out,meas}(\omega)$ fortsetzen lässt mit

$$f_{abs}(\omega) := \frac{\text{abs}(\underline{U}_{out,meas}(\omega))}{\text{abs}(\underline{U}_{out,id}(\omega))} - 1. \quad (5.2)$$

Ist im Betragsspektrum des gemessenen Signals eine Frequenz mit halbem Betrag verglichen mit dem idealen Signal vertreten, wird dies entsprechend der Linearität der Übertragungsfunktion dahingehend gedeutet, dass die Verstärkung von \underline{H} bei dieser Frequenz um einen Faktor 2 zu gering ist. Iterativ mit einer Schrittweite σ_H ausgeführt, folgt für den i -ten Schritt

$$\text{abs}(\underline{H}^{i+1}) = \text{abs}(\underline{H}^i) \cdot (1 - \sigma_H^i f_{abs}^i) \quad (5.3)$$

für $\sigma_H^i \in [0, 1]$ und $\underline{U}_{out,meas}^i$ in f_{abs}^i als gemessenem Ausgangssignal für das mit \underline{H}^i berechnete Eingangssignal ². Würde allerdings Glg. (5.3) mit komplexen Zahlen und nicht allein den Beträgen ausgeführt, würde auch die Phase der -1 beachtet und folglich die durch σ_H skalierte komplexe Zahl wesentlich verändert. Also muss für das Phasenspektrum eine andere Optimierung erfolgen. Eine Möglichkeit hierfür wäre die simple Anpassung der Phase $\arg(\underline{H}) = \varphi_H$ mit

$$\varphi_H^{i+1} = \varphi_H^i - \sigma_\varphi^i \left(\arg(\underline{U}_{out,meas}) - \arg(\underline{U}_{out,id}) \right) \quad (5.4)$$

mit $\sigma_\varphi^i \in [0, 1]$. Diese Anpassung der Phase wurde jedoch nur kurzen Tests unterzogen und anschließend nicht weiter verfolgt. Es hat sich die Signalform des Ausgangssignals unproportional stärker verändert, als dies nur im Falle der Betrags-Anpassung der Fall war. Vermutlich liegt dies an dem aus dem Ausgangssignal gewonnenen Phasengang, der in wesentlich größerem Maße vom idealen Phasengang abweicht als im Betragsspektrum. In Abb. (5.1) sind Betrag und Phase der durch FFT erhaltenen Spektren für gemessenes und ideales Ausgangssignal vor Durchführung einer Optimierung dargestellt. Insbesondere

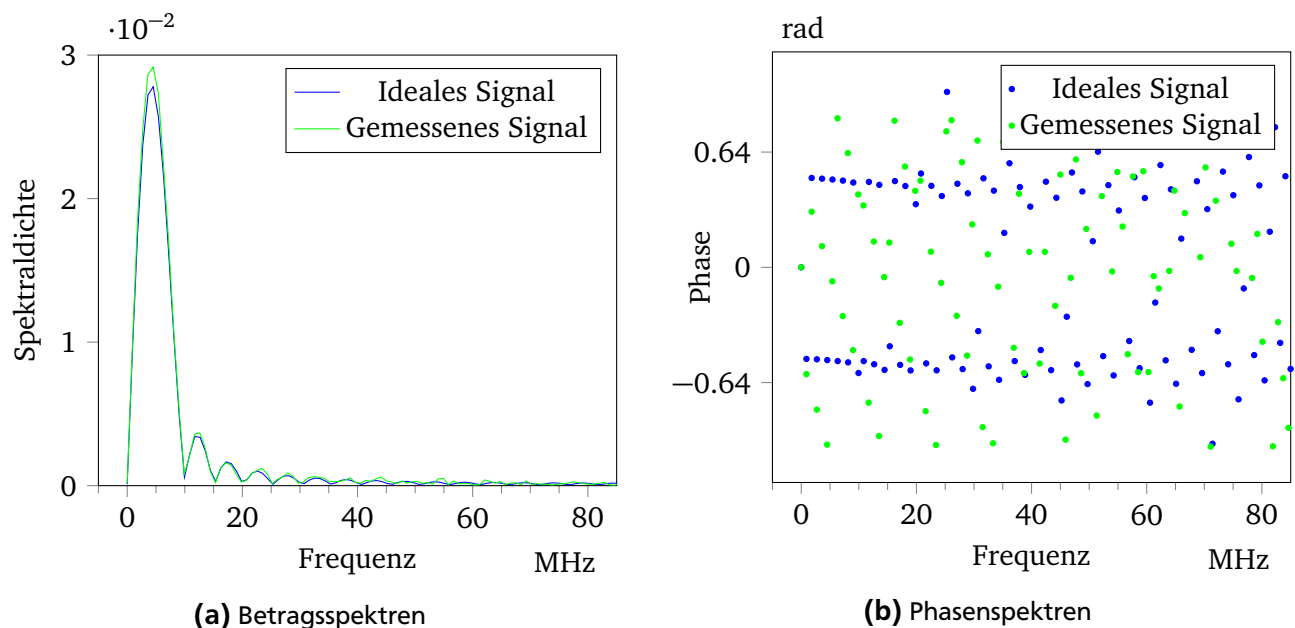


Abbildung 5.1: Spektrum des Einzelsinus-Signals, berechnet und gemessen mit je 109 Punkten

illustriert Abb. (5.1b) die bei gemessenem Signal auftretende Streuung der Phase.

Neben den für kontinuierliche Funktionen problemlos definierbaren iterativen Zuweisungen ergeben sich in Messung und diskreter Ausführung jedoch Fehlerquellen. Problematisch sind insbesondere solche, die in Glg. (5.3) durch das Betragsverhältnis der Ausgangssignale verstärkt werden. Unterscheiden sich die Spektren hier um einen großen Faktor, resultiert dies in einer großen Anpassung der Übertragungsfunktion für die betreffende Frequenz. Dies ist folglich insbesondere bei kleinen Beträgen der Spektren problematisch, wenn Ungenauigkeiten und Störeinflüsse betrachtet werden.

Besondere Störeinflüsse ergeben sich also durch

- Rauschen: Weißes Rauschen macht sich in allen Frequenzen bemerkbar mit kritischem Einfluss bei geringer Spektraldichte des Signals.
- Diskretisierungsfehler: Die FFT bedingt eine begrenzte Auflösung, in den Spektren von \underline{H} und den gemessenen Signalen und liegt insbesondere im Allgemeinen an unterschiedlichen Frequenzen und mit unterschiedlich vielen Punkten vor.
- Interpolationsfehler: Die (hier lineare) Interpolation der Spektren zur Auswertung von f_{abs} an den Frequenzen von \underline{H} kann insbesondere den Einfluss oben genannter Punkte verstärken.

Weiterhin zeigt sich auch in der geringen Stützstellenzahl in Abb. (5.1) bereits eine erste konzeptuelle Problematik des Vorgehens. Der Frequenzabstand zwischen zwei Werten der FFT ist stets mit der Wiederholfrequenz f_{rep} gegeben und lässt sich somit nicht durch eine höhere Auflösung der Messgeräte verbessern, der betrachtete Bereich bis 80 MHz nicht besser auflösen. Folglich setzt sich diese Ungenauigkeit auch auf die Optimierung der Kennlinie fort. Insbesondere relevant wird dies, da die Kennlinie mit nahezu der doppelten Anzahl an Werten erstellt wird und somit der Interpolationsfehler ungleich größer wird als bei ähnlicher Anzahl Stützstellen.

¹ Hierauf beziehen sich alle weiteren Angaben zur Qualität des Signals. Eine intensivere Befassung mit dem Tool hat nicht stattgefunden.

² Nachfolgend wird aus Gründen der Übersichtlichkeit f_{abs} statt dem länglichen Bruch genutzt

5.1.1 Umgang mit Fehlerquellen

Ignorieren kleiner Beträge im Spektrum

Um Rauscheinflüsse und Probleme durch Nulldurchgänge zu dämpfen, wurde ein erster intuitiver Ansatz vorgenommen: Bei den Betragsspektren der in f_{abs} eingehenden Signale, des gemessenen und idealisierten Spannungssignals, wurden alle Anteile, die verglichen mit dem Maximalwert des betreffenden Spektrums besonders klein sind, auf einen vorgegebenen Wert, im Folgenden Default-Wert genannt, gesetzt. Dies führt an den betroffenen Frequenzen zu $f_{\text{abs}} = 0$ und damit keiner Änderung von \underline{H} . Dies bedeutet also, dass alle Einträge des Betragsspektrums von $U_{\text{out,ideal}}$ mit weniger als zum Beispiel 5 ‰ der maximalen Amplitude auf den Default-Wert gesetzt werden. Insbesondere werden auch die Einträge an den Frequenzen zurückgesetzt, die im Spektrum von $U_{\text{out,meas}}$ klein gegen das zugehörige Maximum sind.

Zu beachten bei letzterem Punkt ist die notwendige Rundung, wenn die Einträge der FFT an unterschiedlichen Frequenzen vorliegen.

Mit Beschränkung auf 3 ‰ und dem globalen Minimum beider Spektren als Default-Wert werden insbesondere Einträge in den höheren Frequenzen des Spektrums 5.1a beeinflusst. Erst ab etwa 5 ‰ werden alle Frequenzen ab ungefähr 60 MHz auf den Default-Wert gesetzt. Mit dieser Methodik könnten folglich in erster Linie massive Korrekturen an hohen Frequenzen von \underline{H} verhindert. Verglichen mit der nachfolgend beschriebenen Anpassung hat dieses Vorgehen aufgrund der beschriebenen Bandbreite einen niedrigeren Einfluss auf die Qualität des Ausgangssignals.

Ignorieren großer Korrektur-Terme

Ein zweiter, sehr grober Ansatz liegt in der Beschränkung von f_{abs} auf Werte unterhalb einer vorgegebenen Schwelle. Zugrunde liegt die Annahme, dass die gerade an Nulldurchgängen des Spektrums sowie bei vielen hohen Frequenzen auftretenden großen Werte durch die in obiger Aufzählung genannten Fehlerquellen entstehen. Hier bedeutet dies insbesondere, dass die Diskretisierung die Nulldurchgänge nicht korrekt darstellen kann. Die Interpolation auf Frequenzen von \underline{H} ist dann aufgrund der großen Sprünge von Werten in direkter Umgebung der problematischen Frequenzen mit großer Ungenauigkeit behaftet. Dies kann zu den beschriebenen, großen Korrektur-Termen in f_{abs} führen.

Listing 5.1: Pseudocode zur Veranschaulichung der Anpassung des Korrekturterms

```
rms_orig = root_mean_square( f_abs )
f_abs_to_use = f_abs[ where( abs(f_abs) >= 0.02 * rms_orig ) ]
rms_mod = root_mean_square( f_abs_to_use )
idx_to_clear = f_abs[ where( abs(f_abs) >= rms_mod ) ]
f_abs[ ix_to_clear ] = 0
```

Vereinfacht bedeutet der verfolgte Ansatz, ausnehmend große Werte von f_{abs} als unrealistisch abzutun. Eine Pseudo-Implementierung findet sich in Code (refcode:opt.H.pseudoRMS), um die nachfolgende Erläuterung zu illustrieren. In der vorgenommenen Implementierung wurde f_{abs} an den ausgewählten Frequenzen beliebig auf 0 gesetzt, also keine Anpassung bei diesen Frequenzen ermöglicht. Als Grenze genutzt wurde ein modifizierter Effektivwert, nachfolgend mit RMS (Root Mean Square) bezeichnet. Der reine RMS von f_{abs} unterliegt der Problematik, eine unproportional große Gewichtung von kleinen Einträgen zu enthalten.

Idealerweise enthält f_{abs} mit jeder Iteration kleinere Einträge als zuvor. Es würden also bei Nutzung des reinen RMS unter Umständen mit zunehmender Schrittzahl zunehmend mehr Werte in f_{abs} ignoriert

- was der Optimierung entsprechende Grenzen setzt. In Kombination mit den im vorigen Abschnitt erläuterten Anpassungen wäre die Problematik unumgänglich, da Frequenzen, die explizit nicht bei der Anpassung berücksichtigt werden sollen, den reinen RMS-Wert beeinflussen. Folglich muss der RMS modifiziert werden. Hier wurden zur Berechnung des modifizierten RMS nur die Werte einbezogen, die mehr als beliebig gewählte 2% des reinen RMS betragen. Es handelt sich also bei der vorgenommenen Anpassung um eine sehr grobe und größtenteils willkürliche Wahl der Parameter, die zu Zwecken der Illustration jedoch brauchbare Ergebnisse liefert.

Für einen Eindruck der Tragweite der RMS-Beschränkung bietet sich die Betrachtung von f_{abs} in Abb. (5.2) an. Klar zu erkennen sind die Ausreißer bei den meisten Frequenzen der originalen Nulldurchgänge. Auch die bei höheren Frequenzen auftretenden größeren Fehlerterme fallen auf. Die Korrekturen sind an den auf 0 gezogenen Werten zu erkennen. Der Unterschied zwischen dem originalen und dem modifizierten RMS liegt hier bei ungefähr 0.04, also nahezu 10%. Für die Einordnung der Größenordnung des Korrekturterms sei nochmals betont, dass aufgrund der Anpassung Glg. (5.3) ein Korrekturterm von ± 0.5 bei dem im Folgenden standardmäßig verwendeten $\sigma_H = 1/2$ zu einer Anpassung in \underline{H} in Höhe eines Viertels des ursprünglichen Wertes führt.

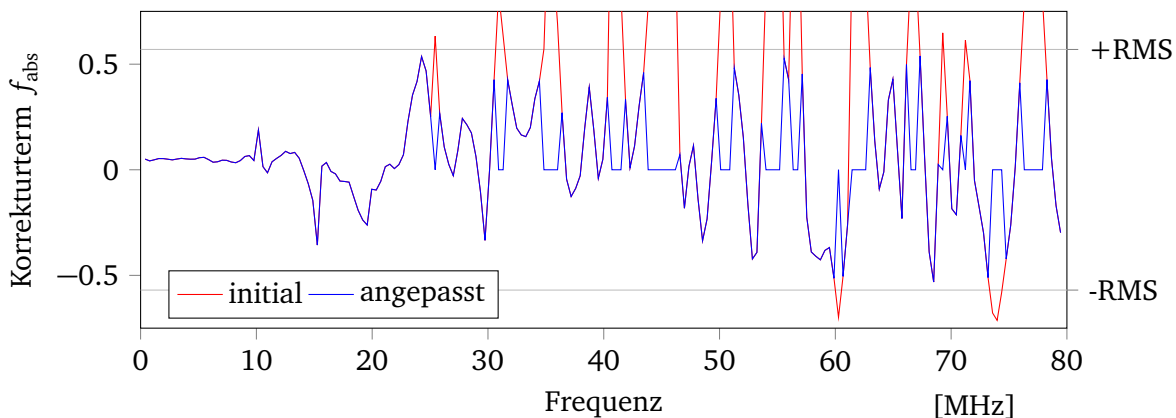


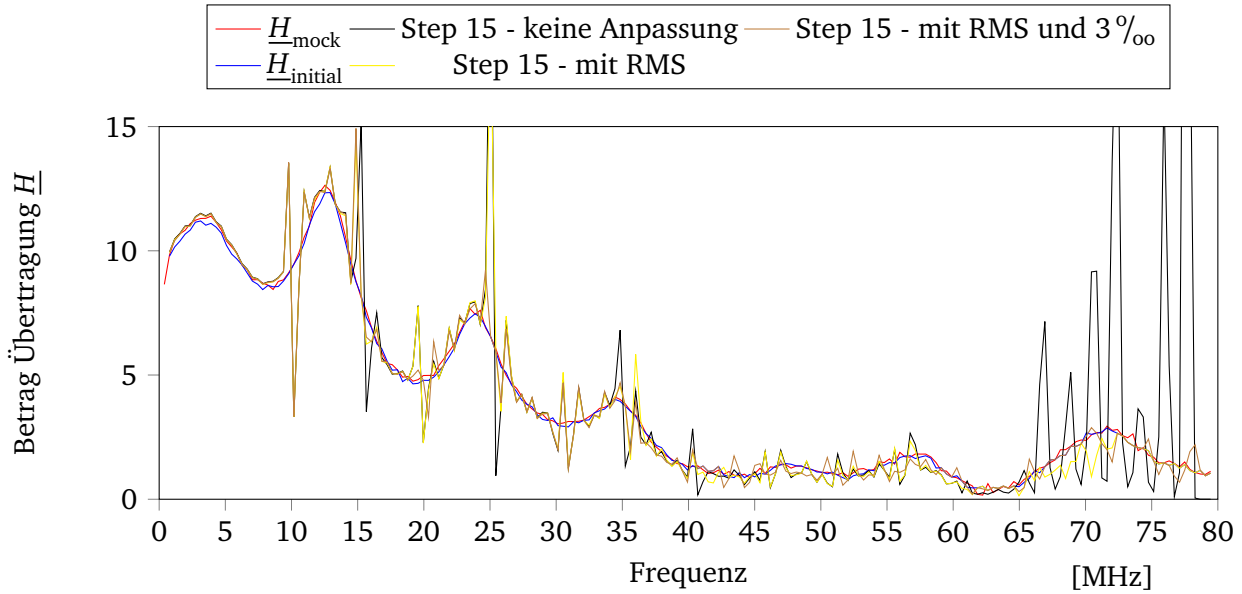
Abbildung 5.2: Korrekturterm in initialer und in angepasster Form mit RMS-Korrektur, am Messaufbau

5.1.2 Auswertung

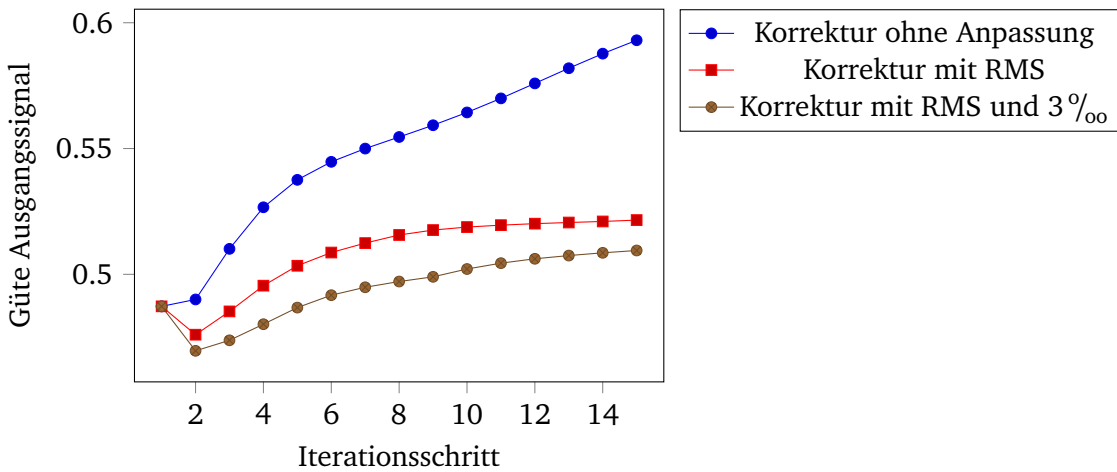
Eine erste Illustration bietet die Anwendung des beschriebenen Vorgehens auf das mock -System. Hier lässt sich testen, ob unter der Annahme einer vorgegebenen Übertragungsfunktion $\underline{H}_{\text{mock}}$ im Sinne des Hammerstein-Modells und unter Vernachlässigung der Kennlinie (entspricht einer als ideal berechneten Kennlinie) eine Verbesserung des Signals erreicht wird und wie sich die Fehlerquellen mit Ausnahme von Rauschen auswirken. Das Ergebnis dieses Tests ist eindeutig und anschaulich in Abb. (5.3a) aufgetragen für jeweils 15 Iterationen. Es lässt sich feststellen, dass eine Anpassung sowohl der Nulldurchgänge als auch bei hohen Frequenzen unbedingt notwendig ist und dies durch die Beschränkung großer Korrekturterme angegangen werden kann.³ Auch zeigen sich in direkter Umgebung der Nulldurchgänge noch Frequenzen, an denen die Korrekturterme zwar geringer als der modifizierte RMS, jedoch immer noch ungenau durch die Interpolation aufgelöst sind. Diese Fehler sind ausschlaggebend dafür, dass sich die Qualität des erhaltenen Ausgangssignals über die Iterationen verschlechtert.⁴ Nichtsdestotrotz nähert sich die Übertragungsfunktion außerhalb der problematischen Nulldurchgänge gerade in niedrigeren Frequenzen sehr gut an $\underline{H}_{\text{mock}}$ an.

³ Die Auswirkungen der zusätzlichen Nutzung einer Beschränkung auf 3‰ verbessern die Übertragungsfunktion optisch nicht wesentlich.

⁴ Je geringer der Wert des Qualitätskriteriums, desto ähnlicher ist das Signal einem idealen Einzelsinus.



(a) Entwicklung des Betrags der Übertragungsfunktion des mock -Systems mit und ohne Anpassung großer Korrekturterme

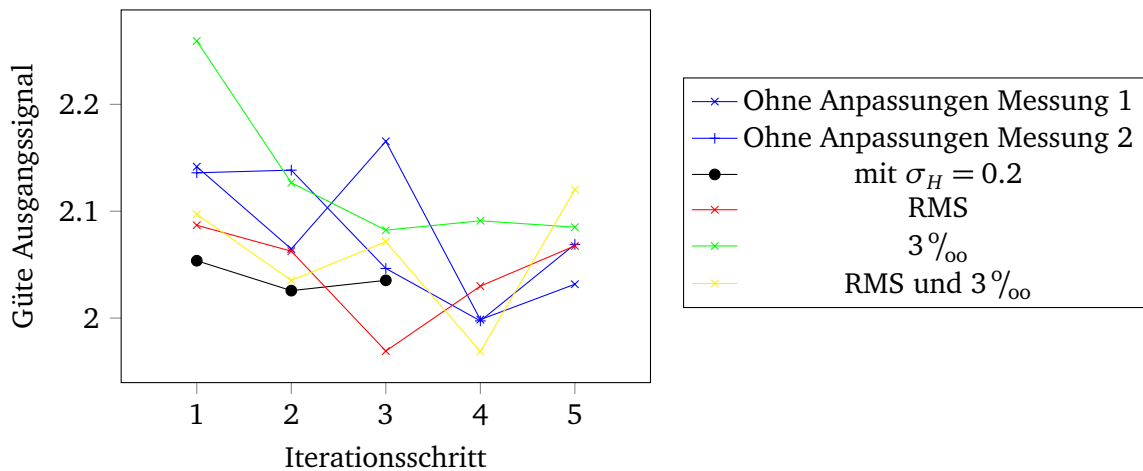


(b) Entwicklung des Qualitätswertes des Ausgangssignals über 15 Iterationen, mit und ohne Anpassung des Korrekturterms

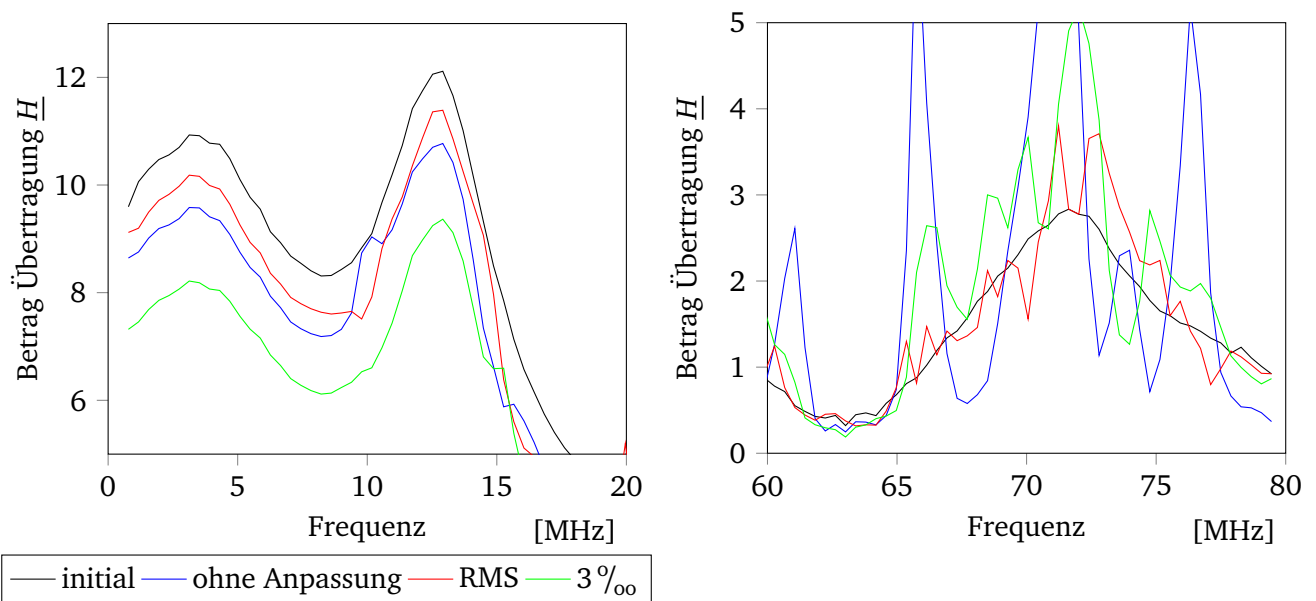
Abbildung 5.3: Anwendung der Optimierung von \underline{H} mit und ohne Anpassung des Korrekturterms auf das mock -System bei $\sigma_H = 1/2$

Nutzt man die Optimierung für \underline{H} im nahezu linearen Bereich der Kennlinie K , übertragen sich die prinzipiellen Erkenntnisse aus dem mock -System auf die Kavität und es lässt sich auch der Einfluss von Rauschen auf die Optimierung feststellen. Genutzt wurde hierzu und im Folgenden eine Kennlinie, die mit einer Peak-to-Peak-Spannung des Eingangssignals von 0.6 V erzeugt wurde. Das Ausgangssignal wurde in der Optimierung mit $V_{pp} = 0.6\text{ V}$ angesetzt, was zu etwa 60 mV Peak-to-Peak-Spannung am Eingang zurückgerechnet wird.⁵ Klar erkennbar sind wieder die Problematik der Nulldurchgänge und die Auflösung hoher Frequenzen. Einen Überblick über die Variation der Parameter der Optimierung bietet Abb. (5.4).

⁵ Hier sei auf die unter beschriebene Problematik der nichtlinearen Vorverzerrung bei großen Unterschieden zwischen der Amplitude des momentan betrachteten Signals und des zur Berechnung von K genutzten hingewiesen. Diese Erkenntnis lag zum Zeitpunkt der hier betrachteten Messung noch nicht in dieser Schärfe vor.



(a) Entwicklung der Qualität des Ausgangssignals bei unterschiedlicher Parameterwahl



(b) Einfluss unterschiedlicher Parameterwahl auf Entwicklung der Übertragungsfunktion bei hohen Frequenzen, nach 5 Iterationen

Abbildung 5.4: Einfluss unterschiedlicher Parameterwahl auf Entwicklung der Qualität des Ausgangssignals und die Übertragungsfunktion

Klar erkennbar ist aus Abb. (5.4a) vor allem, dass der Einfluss zufälliger Schwankungen auf die Qualität des Signals enorm ist, man vergleiche hierzu nur die unter identischen Voraussetzungen und um wenige Minuten verzögert entstandenen Messungen mit \times und $+$. Weiterhin lässt sich im Vergleich mit Abb. (5.3b) die wenig verwunderliche Aussage treffen, dass die Qualität aller Ausgangssignale wesentlich schlechter ist, als dies beim `mock`-System der Fall war. Diese nicht eindeutige Änderung der Qualität ist insofern verwunderlich, als dass auch die starke Anpassung über die Iteration bei in \underline{H} stark verstärkten niedrigen Frequenzen nicht nennenswert auf die Qualität auswirkt. Und letztlich ist über die hier vorgenommenen Iterationsschritte keine eindeutige Verbesserung des Ausgangssignals zu erkennen. Da gleichzeitig jedoch auch keine enorme Verschlechterung eintritt und der Rauscheinfluss nicht beziffert werden kann, ist eine qualitative Aussage über die einzelne Optimierung von \underline{H} mit den vorliegenden Daten für die vorgenommenen Anpassungen nicht zu treffen. Festhalten jedoch lässt sich auch bei Messungen am Messaufbau, dass die vorgenommenen Anpassungen und insbesondere die Beschränkung mit

einem modifizierten RMS wesentlich weniger als abwegig angesehene Korrekturterme erlauben und so eine kleinschrittigere Anpassung möglich machen.

5.2 Optimierung der nichtlinearen Kennlinie K

Der Unterschied zur Optimierung von \underline{H} ist, dass diese Optimierung im Zeitbereich statt findet. Deshalb kann Glg. (5.1) zu

$$\Delta U_{\gamma}(t) = U_{\gamma,\text{meas}}(t) - U_{\gamma,\text{ideal}}(t) \quad U_{\gamma,\text{meas}}(t) = \mathcal{F}^{-1} \left\{ \underline{H}^{-1}(\omega) \cdot \underline{U}_{\text{out,meas}}(\omega) \right\} \quad (5.5)$$

geändert werden. Die Funktionen $U_{\gamma,\text{meas}}(t)$ und $U_{\gamma,\text{ideal}}(t)$ werden im mathematischen Modell aus $U_{\text{in}}(t)$ mit einem Polynom gleichen Grades berechnet, deshalb kann die Differenz dieser Funktionen $\Delta U_{\gamma}(t)$ auch als Polynom gleichen Grades aus $U_{\text{in}}(t)$ berechnet werden

$$\Delta U_{\gamma}(t) = \sum_{n=1}^N \tilde{a}_n [U_{\text{in}}(t)]^n \quad (5.6)$$

Die Berechnung der Koeffizienten \tilde{a}_n stellt ebenso ein lineares Optimierungsproblem dar wie schon die Berechnung der Koeffizienten a_n in Glg. (2.1) siehe [3]. Dabei werden M Samples von $\Delta U_{\gamma,i} = \Delta U_{\gamma}(i \cdot \Delta t)$ mit zugehörigen Samples des Eingangssignals $U_{\text{in},i} = U_{\text{in}}(i \cdot \Delta t)$ verglichen. Mit der Potenzreihe aus Glg. (5.5) ergibt sich folgendes Gleichungssystem

$$\begin{pmatrix} U_{\text{in},1} & U_{\text{in},1}^2 & \dots & U_{\text{in},1}^N \\ U_{\text{in},2} & U_{\text{in},2}^2 & \dots & U_{\text{in},2}^N \\ \vdots & \vdots & \ddots & \vdots \\ U_{\text{in},M} & U_{\text{in},M}^2 & \dots & U_{\text{in},M}^N \end{pmatrix} \cdot \begin{pmatrix} \tilde{a}_1 \\ \tilde{a}_2 \\ \vdots \\ \tilde{a}_N \end{pmatrix} = \begin{pmatrix} \Delta U_{\gamma,1} \\ \Delta U_{\gamma,2} \\ \vdots \\ \Delta U_{\gamma,M} \end{pmatrix} \quad (5.7)$$

Dieses Gleichungssystem ist mit normalerweise $M > N$ überbestimmt und wird mit der Methode der kleinsten Quadrate gelöst. Die Koeffizienten \tilde{a}_n werden nun wie folgt zur Anpassung der Koeffizienten a_n verwendet

$$a_n^{i+1} = a_n^i + \sigma_a^i \tilde{a}_n^i \quad (5.8)$$

Für die Schrittweite gilt $\sigma_a^i \in [0, 1]$.

5.2.1 Erster Ansatz

Für die Berechnung der ersten Kennlinie K_0 wurde das ideale Ausgangssignal $U_{\text{out,ideal}}$ mit $V_{pp} = 6\text{V}$ über \underline{H}^{-1} zurückgerechnet und als Eingangssignal verwendet $U_{\text{in,initial}} = U_{\gamma,\text{ideal}}$. Damit ergibt sich für das Eingangssignal ein $V_{pp} \approx 587\text{mV}$, um K_0 für den nichtlinearen Spannungsbereich des Systems zu berechnen. Als erster Ansatz wurde versucht K über das selbe $U_{\gamma,\text{ideal}}$ zu optimieren, mit dem es zuvor zur Berechnung verwendet wurde. Das heißt, dass die beiden Signale die gleiche Amplitude haben. Beim Zurückrechnen über die Kennlinie wurde festgestellt, dass K in diesem Bereich nicht mehr bijektiv ist. Für den Fall, dass das Signal zu groß für den gültigen Bereich ist wird es auf beliebig gewählte 80% des maximal zulässigen Wertes gesetzt. Dabei könnten sich in jedem Iterationsschritt unterschiedliche Amplituden von U_{γ} ergeben. Ziel hierbei ist es den Bereich, in dem K bijektiv ist, schrittweise zu vergrößern.

Ergebnisse und Erkenntnisse

Das Ergebnis ist eine Schwankung der Qualität siehe Abb. (5.5b). Die Kennlinien, zu den markierten Iterationen sind in Abb. (5.5a) eingezeichnet. Zu erkennen ist eine alternierende Anpassung der Kennlinie, die schon in den ersten drei Iterationen angefangen hat. Ergänzend wurden auch die letzten beiden Kennlinien eingezeichnet, um die große Anpassung innerhalb einer Iteration zu verdeutlichen. Ursache dafür könnte sein, dass die Kennlinie in einen nicht mehr sinnvollen Bereich genutzt wurde, um das Eingangssignal zu berechnen. Dadurch könnten zu große Änderungen in einem Iterationsschritt vorgenommen worden sein, die nicht mehr zielführend waren.

Ein weiteres Ergebnis ist auch, dass ein U_{γ} , das über das ideale Ausgangssignal mit $V_{pp} = 6\text{ V}$ bestimmt wird, nur als Eingangssignal sinnvoll ist, um die Kennlinie zu berechnen. Dabei wurde am Ausgang eine Spannung von $V_{pp} = 4.7\text{ V}$ gemessen und daraus folgt, dass für die Optimierung ein U_{γ} verwendet werden sollte, das sich über ein Ausgangssignal mit höchstens diesem Wert für V_{pp} berechnet.

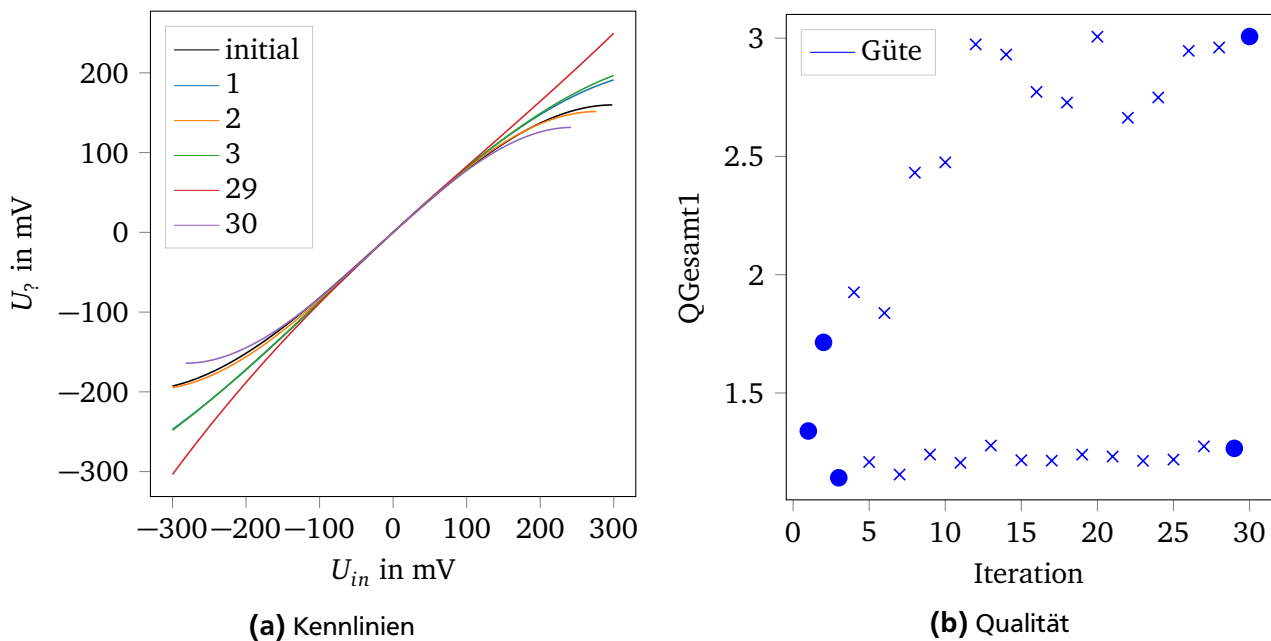


Abbildung 5.5: Erster Ansatz zur Anpassung von K

Grenzen der initial Kennlinie

Das Signal $U_{\gamma,\text{meas}}$ wurde aus dem gemessenen $U_{\text{out,meas}}$ berechnet und ist in Abb. (5.6) eingezeichnet. Die sich daraus ergebende Kennlinie ist in Abb. (5.6a) eingezeichnet. Das dazu gehörige Eingangssignal hat dabei die Form von $U_{\gamma,\text{ideal}}$ mit $V_{pp} = 578\text{ mV}$ und ist damit etwa doppelt so groß wie das in Abb. (5.6b) eingezeichnete $U_{\gamma,\text{ideal}}$.

Wenn man jetzt $U_{\gamma,\text{ideal}}$ mit $V_{pp} = 300\text{ mV}$ über K_0 zurückrechnet, um das erste nichtlinear vorverzernte Eingangssignal zu erhalten, so stellt man fest, dass dies an der Grenze von K_0 stößt, siehe auch das eingezeichnete Maximum von U_{γ} in Abb. (5.6a).

Deshalb kann damit zwar ein vorverzerntes Signal berechnen werden, aber Messungen zeigen, dass die Qualität in diesem Bereich abnimmt.

Eine Ursache dafür könnte sein, dass die Methode der kleinsten Quadrate, mit der die Parameter a_n für K bestimmt werden, die Extrema von U_{γ} nicht genug gewichtet, um in diesem Bereich noch gültig zu sein.

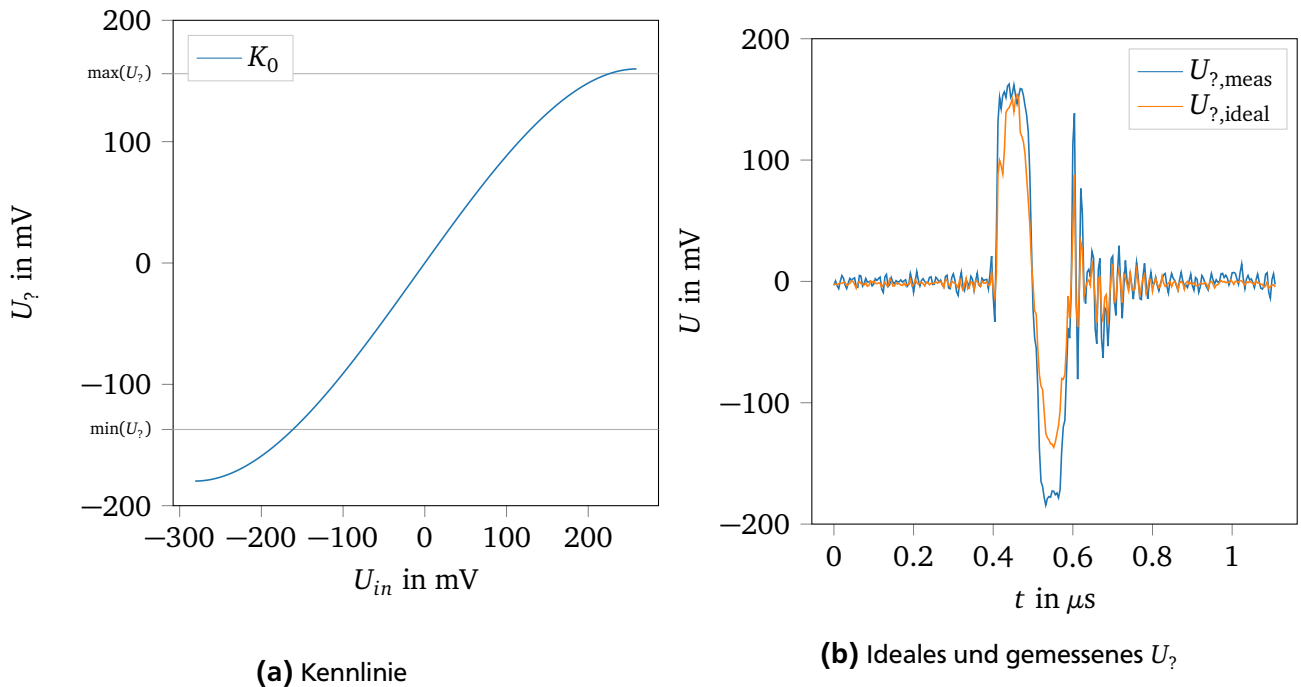


Abbildung 5.6: Nichtlineare Vorverzerrung

Für die Evaluierung der Grenzen wurden verschiedene Amplituden von $U_{?,ideal}$ über die Kennlinie nichtlinear vorverzerrt. Das Ausgangssignal wurde mit dem RF-Tool -Zitat fehlt noch- bewertet und ist in Abb. (5.7b) gezeigt. Dabei sind die Grenzen von $U_?$ in Abb. (5.7a) mit $\min(U_?)$ und $\max(U_?)$ beschriftet, dies entspricht mit $V_{pp} = 120$ mV etwa 34% des maximal möglichen V_{pp} , das von K im bijektiven Bereich zugelassen ist. Die anderen eingezeichneten Grenzen, die nicht beschriftet sind, stellen die Grenzen dar, bis zu denen die Güte des Ausgangssignals überprüft wurde.

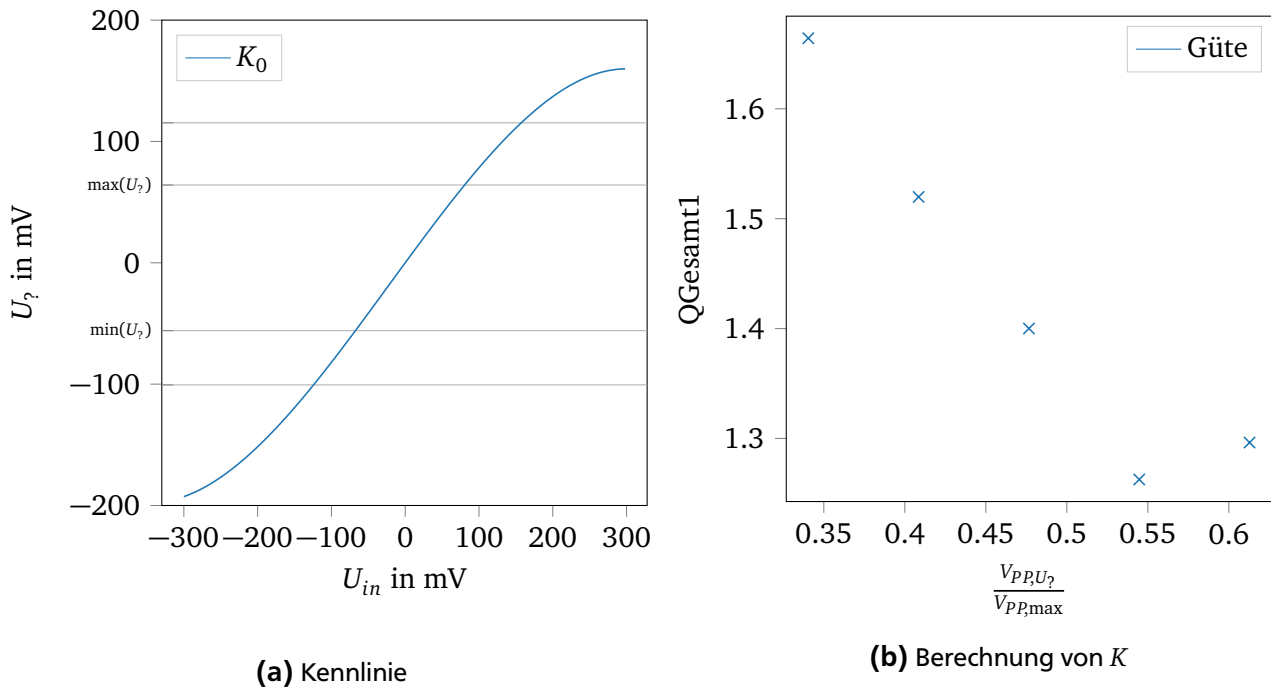


Abbildung 5.7: Bewertung des Ausgangssignals

Dabei ist der kleinste und damit der beste Wert der Güte mit 190 mV etwa 54% des maximal möglichen V_{pp} das von K aufgezeichnet worden. Diese Messung wurde nur für diese Kennlinie durchgeführt und lässt keine Schlüsse auf die allgemeine Gültigkeit für alle Kennlinien zu, deshalb sollte der optimale Wert für V_{pp} sicherheitshalbe für alle Kennlinien einzeln bestimmt werden.

5.2.2 Zweiter Ansatz

Aus den Ergebnissen aus 5.2.1 und der Evaluierung der Grenzen aus dem vorherigen Abschnitt wurde ein neuer Ansatz zur Optimierung aufgestellt, bei dem die Kennlinie K in einem kleineren Bereich optimiert wird als der, aus dem sie vorher zur Initialisierung berechnet wurde.

Als zweiter Ansatz wurde folgendes Setup verwendet:

- K_0 bestimmt über $U_{out,ideal}$ mit $V_{pp} = 6\text{ V}$
- K angepasst über $U_{out,ideal}$ mit $V_{pp} = 3\text{ V}$

Damit wurde K mit einem $U_{?,ideal}$ optimiert, bei dem V_{pp} auf 66% des maximal zulässigen Wertes gesetzt wurde. Der Bereich ist in Abb. (5.8a) eingezeichnet.

Ergebnisse und Erkenntnisse

Für diesen Ansatz wurden aufgrund der späten Erkenntnis keine ausführlichen Tests durchgeführt, deshalb lässt sich darüber nur bedingt eine Aussage über die Konvergenz treffen. Aber diese niedrigen Werte der Güte konnten bisher nur mit diesem Ansatz erreicht werden. Im Vergleich zu den Ergebnissen des vorigen Ansatzes in Abb. (5.5b) fällt jedoch auf, dass in den ersten Iterationen keine Verschlechterung der Qualität verglichen mit dem Initialwert auftritt.

Die Schrittweite wurde bei all diesen Messungen auf $\sigma_a = 0.5$ gelassen.

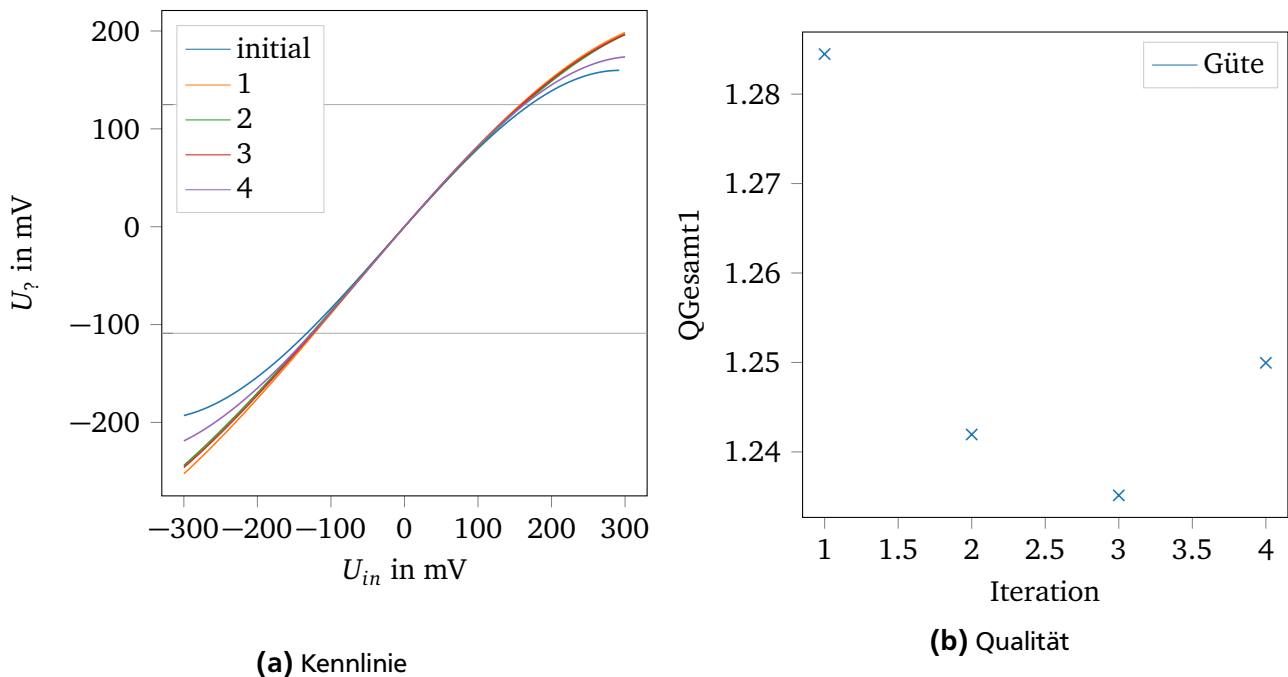


Abbildung 5.8: Zweite Ansatz zur Anpassung von K

6 Fazit

Das erweiterte Programm bietet die Möglichkeit, in einem Aufruf Kennlinie und Übertragungsfunktion des Hammerstein-Modells zu bestimmen, zu speichern und für eine weitere Auswertung vorzuhalten. Der Funktionsumfang wurde um Schleifen für die getrennte iterative Anpassung der beiden Blöcke ergänzt und ein erster Ansatz für diese Optimierung ausprobiert.

Dabei lässt sich feststellen, dass die Anpassung der Übertragungsfunktion theoretisch zielführend ist, jedoch für eine merkbliche Verbesserung am realen System noch der Einfluss von Rauschen betrachtet werden muss. Auch muss ein Umgang mit den Einflüssen von Interpolation und Diskretisierung auf die Spektren der genutzten Signale gefunden werden. Die hierzu vorgenommenen Anpassungen in der Optimierung konnten zwar die Problematiken aufgreifen, stellen allerdings noch keine Lösung dar.

Die Optimierung der Kennlinie führt zur Feststellung konzeptioneller Probleme. Der Definitionsbereich des zur initialen Berechnung genutzten Signals und die Amplituden der für weitere Iterationen genutzten Signale haben entscheidenden Einfluss auf Erfolg und Qualität der Optimierung. Hier konnte die Problematik erkannt und beschrieben werden, ohne jedoch eine verlässliche Lösung anzubieten.

Es lässt sich anhand der vorgenommenen Implementierung feststellen, dass die verfolgten Ansätze das Modell nicht verschlechtern. Eine Aussage über eine merkbliche Verbesserung ist mit den genutzten Algorithmen jedoch nicht abschließend möglich.

Bei erfolgreicher Durchführung einer iterativen Optimierung der Bausteine des Hammerstein-Modells ausgehend von initialen Werten könnte es möglich sein für ein beliebiges Ausgangssignal hinreichend gute Werte erhalten zu können. Wird die Optimierung weiterhin auf ein ideales Hammerstein-Modell angewendet, bietet sich die Möglichkeit zum Vergleich zwischen Messaufbau und mathematischem Modell. Auf dieser Basis kann der Algorithmus zur Beurteilung der Grenzen der vorgenommenen Modellierung für die Kavität genutzt werden. Dies ist auf Basis der im Rahmen dieser Arbeit vorliegenden Werte noch nicht möglich.

Im Entstehungsprozess der Implementierung konnte festgestellt werden, dass eine umfassende Anpassung der Code-Struktur notwendig war. Dadurch konnte verwirklicht werden, dass für weitere Arbeiten an der Implementierung eine aussagekräftige Dokumentation und ansprechende Struktur vorliegen, die einen leichten Einstieg und die einfache Erweiterbarkeit ermöglichen sollen.

Das vorliegende Design ist in modulare Blöcke aufgeteilt, die jeweils einzelne in sich geschlossene Funktionalitäten darstellen. Auch wurde ein Programmier-Konzept genutzt, bei dem bereits lange vor den Messungen an der Kavität die Funktionalität des Codes überprüft werden konnte. Die dafür elementaren Tests der modularen Blöcke können sicherstellen, dass neue Änderungen keine bereits erfolgreich implementierte Funktionalität beeinträchtigen. Das Konzept umfasst ebenso die Simulation eines Hammerstein-Modells, sodass die Optimierungsalgorithmen nicht erst am Messaufbau ausprobiert und weiterentwickelt werden mussten. Dabei ist festzustellen, dass der Prozess der Umstrukturierung und die Einbettung dieser neuen Umgebung sich ausnehmend aufwendig gestaltet hat. Fehlersuche, Übersicht und Erweiterbarkeit konnten unverkennbar verbessert werden. Dadurch konnte die Einbindung von neuen Funktionen um ein Vielfaches vereinfacht werden, wenn dies auch mit einem nennenswerten Maß an Disziplin beim Coden verbunden ist. Es lässt sich also konstatieren, dass die vorliegende Form der Optimierung ohne dieses Design nicht zu verwirklichen gewesen wäre.

Zusammenfassend stellt die vorliegende Projektarbeit eine Basis für nachfolgende inhaltliche Erweiterungen von Optimierungsansätzen am Barrier-Bucket-System dar.

7 Ausblick

— In diesem Kapitel wird auf offene Fragen / neue Probleme / Anstöße für weitere Arbeiten eingegangen. Dabei sollte es um eher inhaltliche Aspekte gehen (u. U. wenig to do for Code-Design) gegebenenfalls darf hier bei vielem auf die Erfahrungen aus den vorigen Kapiteln verwiesen werden und damit einen Übersichts-Charakter haben (erleichtert nachfolgenden Projekten die Arbeit) —

7.1 Impressionen zum Code

In Bezug auf das Design des Codes verbleiben eine Reihe Anregungen und Gedanken, die nicht realisiert wurden, je nach Ausführung aber interessant zu bedenken sein könnten.

- Die Geräte AWG und DSO als Klassen einzubinden könnte den Vorteil bieten, alle SCPI Commands an einer Stelle zu bündeln und den Zugriff darauf allgemeingültig zu halten.
- Eine Einbindung des momentanen Programms in die RF-Tools des GSI-Standards könnte in zwei Teilen erfolgen. Die bestehende Funktionalität zur Berechnung von K und H kann unabhängig der Optimierungsalgorithmen genutzt werden. Dabei ist insbesondere zu prüfen, ob oder wie die Aspekte im Code-Design beibehalten werden.
- Die momentane Version ist in erster Linie über eine Python-IDE ausführbar. Die Ausführung über die Kommandozeile wurde nicht fokussiert.

7.2 Ausblick: Optimierung

- (offene Punkte K-Optimierung?)

Auf den Ergebnissen aus Iterative Optimierung des Hammerstein-Modells aufbauend, verbleiben eine Reihe von offenen Fragestellungen:

1. Wie wirkt sich die Optimierung von K aufgrund ihrer Nichtlinearität auf die Übertragungsfunktion H aus und muss dies in der Optimierung berücksichtigt werden, etwa in der Reihenfolge der Iterationsschritte?
2. Wie wird mit der Tatsache umgegangen, dass im Frequenzbereich unabhängig der Qualität der Messung nur etwa halb so viele Daten für die Anpassung zur Verfügung stehen, wie in H selbst vorliegen? Eine erste Verbesserung könnte das zero-padding genannte Hinzufügen von zusätzlichen 0-Werten an die Zeitsignale sein, womit sich für die Berechnung der FFT der Frequenzabstand zweier Werte verringern lässt.
3. In welcher Reihenfolge wird die Iteration durchgeführt? Wird zuerst H in mehreren Durchgängen angepasst und danach K ? Oder im Wechsel je eine Iteration?
4. Wie wird die Auswahl der Schrittweiten σ_H und σ_a vorgenommen? Werden diese pauschal einmal festgesetzt zu Beginn des Algorithmus oder ist eine dynamische Anpassung, etwa durch die Qualität des letzten gemessenen Signals oder in Abhängigkeit des Iterationsschritts vorzuziehen?
5. Wie lässt sich der Einfluss von zufälligem Rauschen auf die Optimierung reduzieren? Insbesondere die Auflösung im höherfrequenten Betragsspektrum ist hier problematisch. Und im Falle des Ignorieren von Korrekturen an als fehlerhaft befundenen Frequenzen: Auf welchen Wert wird der Korrekturterm bei diesen Frequenzen gesetzt?

-
6. Wie lassen sich die im idealen Spektrum des Einzelsinus enthaltenen Nulldurchgänge in der Optimierung von \underline{H} berücksichtigen, um interpolationsbedingt große Fehlerterme zu vermeiden? Ist das einfache Ignorieren dieser Frequenzen für die Anpassung eine Möglichkeit? Und in welchem kleinen Frequenzbereich um einen Nulldurchgang müssten dann die Werte ignoriert werden?
 7. Wie - insofern überhaupt - ist eine Optimierung der Phase von \underline{H} zu gestalten?
 8. Nach welchem Qualitätsmerkmal wird das Signal bewertet und wie wirkt sich dies auf den Algorithmus aus?
 9. Gibt es eine sinnvolle Abbruchbedingung, mit der die Iteration versehen werden sollte? Etwa, dass sich die Qualität im Vergleich zu den vorherigen Iterationen nicht mehr mit ähnlicher Rate verbessert hat? Der Trade-Off liegt zwischen Laufzeit und Signal-Qualität.
 10. Wie bestimmt man die Grenzen, in denen K genutzt werden kann? Wie kann man garantieren, dass K in den Bereichen, aus denen die Daten zur Berechnung von a_n benutzt werden, bijektiv ist?
 11. Gibt es eine Möglichkeit die Grenzen von K bei der Optimierung zu erweitern? Ein möglicher Indikator: Kann man die Rückgabe von `numpy.linalg.lstsq(LGS)` aus der Berechnung von a_n nutzen, um eine Aussage über die Abweichung der Werte zu treffen?

7.3 — Gerätekomm—

— in dieser Section werden weitere Punkte der Geräte-Komm aufgegriffen, etwa - die (geringe) Auflösung des AWG im Kontext der Optimierung (ggf. in 7.2 besser?) , die Einbindung des neuen Oszis oder die Idee der Klassen-Implementierung

Literaturverzeichnis

- [1] Leslie Lamport, *L^AT_EX: a document preparation system*, Addison Wesley, Massachusetts, 2nd edition, 1994.
- [2] Denys Bast, Armin Galetzka, *Projektseminar Beschleunigertechnik*, 2017.
- [3] Jens Harzheim *et al.*, *Input Signal Generation For Barrier Bucket RF Systems At GSI*, In Proceedings of IPAC2017 in Copenhagen, Dänemark. pp. 3948 - 3950 2017.
- [4] Kerstin Gross *et al.*, *Test Setup For Automated Barrier Bucket Signal Generation*, In Proceedings of IPAC2017 in Copenhagen, Dänemark. pp. 3948 - 3950 2017.
- [5] Keysight Technologies, *Keysight Trueform Series Operating and Service Guide*, 2015.
- [6] Keysight Technologies, *33600A Series Trueform Waveform Generators - Data Sheet*, 2014.
- [7] C. Tröser, *Application Note: Top Ten SCPI Programming Tips for Signal Generators*, Rohde & Schwarz, 2013.
- [8] I. Sommerville, *Software Engineering*, 9th edition
- [9] https://en.wikipedia.org/wiki/Helper_class
- [10] Code Complete 2 by Steve McConnell
- [11] <https://www.testingexcellence.com/pros-cons-test-driven-development/>
- [12] https://en.wikipedia.org/wiki/Mock_object
- [13] https://en.wikipedia.org/wiki/System_testing

8 Anhang

hier kommt eine Liste der Funktionen nochmals hin