Fraunhofer
IAF

FRAUNHOFER-INSTITUT FÜR ANGEWANDTE FESTKÖRPERPHYSIK IAF

KOMPETENZZENTRUM QUANTENCOMPUTING BADEN-WÜRTTEMBERG

VERBUNDFORSCHUNGSPROJEKT

QUANTENOPTIMIERUNG MIT RESILIENTEN ALGORITHMEN II (QORA II)

Abschlussbericht
März 2024

Koordinator:
Dr. Thomas Wellens
Fraunhofer-Institut für Angewandte Festkörperphysik IAF
in Freiburg

Projektnummer: WM3-4332-149/44

Projektpartner:

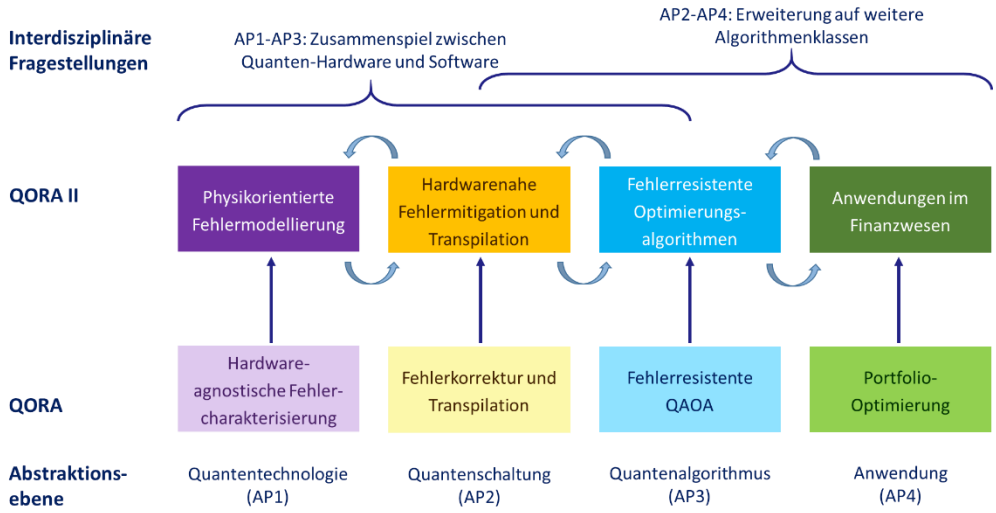
- Prof. Ilia Polian, Institut für Technische Informatik, Universität Stuttgart (USTUTT-ITI)
- Prof. Stefanie Barz, Institut für Funktionelle Materie und Quantentechnologien, Universität Stuttgart (USTUTT-FMQ)
- Prof. Guido Burkard, Fachbereich Physik, Universität Konstanz (UKON)
- Prof. Daniel Braun, Institut für Theoretische Physik, Eberhard-Karls-Universität Tübingen (EKUT)
- Prof. Gerhard Hellstern, Duale Hochschule Baden-Württemberg, Stuttgart (DHBWS)
- Prof. Martin Zaefferer, Duale Hochschule Baden-Württemberg, Ravensburg (DHBWR)
- Dr. Vamshi Katukuri, Fraunhofer-Institut für Arbeitswirtschaft und Organisation (IAO)

Inhalt

1 Zusammenfassung

1.1 Projektziele

Ziel von QORA II war es, die in QORA gewonnenen Erkenntnisse in mehrere Richtungen zu vertiefen und zu erweitern. Hierfür sollten auf die Hardware des Quantenrechners besser zugeschnittene Verfahren zur Charakterisierung und Korrektur der darauf auftretenden Fehler entwickelt und neue, über die bislang im Zentrum von QORA stehenden variationellen QAOA-Ansätze hinausgehende algorithmische Konzepte entwickelt werden. Unsere hinsichtlich ihrer Fehlerresistenz optimierten Quantenalgorithmen sollten neben der Portfoliooptimierung schließlich auch auf weitere für den Finanzbereich relevante Fälle (z.B. Kreditscoring, Clustering von Kundendaten und Risikomodellierung) angewendet werden. Das sich im Vorgängerprojekt QORA bewährt habende Konzept des schichtenübergreifenden Ansatzes mit 4 Arbeitspaketen wurde beibehalten:



Bei der Vertiefung der bereits gewonnenen Erkenntnisse sowie der Erschließung neuer Themen wurden zwei Hauptentwicklungsrichtungen verfolgt:

- **Stärkere Einbindung der Technologie-Aspekte:** In der ersten Förderperiode konnten wir einen umfangreichen Erfahrungsschatz über den Betrieb des IBM-QC und seine Ausfallmechanismen aufbauen. Diese Erkenntnisse wurden in der zweiten Förderperiode auf verschiedenen Abstraktionsebenen gewinnbringend eingesetzt, um technologieorientierte Fehlermodellierung, Fehlermitigation, Transpilation und Entwicklung von resilienzorientierten Quantenoptimierungs-verfahren zu ermöglichen.
- **Erweiterung der betrachteten Algorithmenklassen** von Portfolio-Optimierung aus der ersten Förderperiode auf eine Reihe von Anwendungen mit Fokus auf Finanzwesen, darunter Kunden-Scoring und Betrugserkennung basierend auf maschinellen Lernverfahren, Analyse von Kundendaten und Monte-Carlo-Analysen. Neben dem zuvor betrachteten QAOA fanden weitere Optimierungsverfahren, etwa basierend auf der dissipativen Relaxation, Berücksichtigung. Für ihre Resilienz sollten unterstützende Maßnahmen auf unteren Abstraktionsebenen, etwa speziell angepasste Transpilationsverfahren, entwickelt werden.

1.2 Projektverlauf

Um einen Kurzeinblick in den Verlauf des Projekts und die Erreichung der oben genannten Ziele zu geben, stellen wir die Ergebnisse zunächst anhand der im Antrag formulierten Meilensteine dar:

M1	Fehlermodell mit einfachen Dekohärenzprozessen liegt vor	AP1	Monat 7	erfüllt
M2	Erweitertes Fehlermodell liegt vor	AP1	Monat 12	nicht erfüllt
M3	Weiterentwickelte Transpilationsverfahren bereit und gebenchmarkt	AP2	Monat 9	erfüllt
M4	Robuste Erzeugung verschränkter Zustände mit Methoden aus AP2.2 demonstriert	AP2	Monat 12	teilweise erfüllt
M5	Wirksamkeit der Dynamischen Entkopplung auf IBM-QC analysiert	AP2	Monat 14	erfüllt
M6	Machbarkeitsstudie zur dissipativen Optimierung mit 2 Qubits erstellt	AP3	Monat 6	erfüllt
M7	Weiterentwicklung hybrider Algorithmen beendet	AP3	Monat 12	erfüllt
M8	Einfluss korrelierter Fehler für verschiedene NISQ-Algorithmen analysiert	AP3	Monat 13	nicht erfüllt
M9	Dissipative Optimierung auf größere Probleme erweitert und auf IBM-QC getestet	AP3	Monat 13	nicht erfüllt
M10	Resilienter Quantenalgorithmus für Merkmalsauswahl erstellt und getestet	AP4	Monat 6	erfüllt
M11	Demonstrator für Risikomodellierung mit Potenzialanalyse erstellt	AP4	Monat 12	erfüllt
M12	Übersetzung von Clustering-Algorithmen in QC-kompatible Form untersucht	AP4	Monat 13	erfüllt
M13	Studie zur Abschätzung des Quantenvorteils für alle Algorithmen von QORA II	AP4	Monat 15	erfüllt

Die meisten Meilensteine wurden wie geplant umgesetzt und somit das Projekt erfolgreich abgeschlossen werden. Die hohe wissenschaftliche Qualität der Ergebnisse wird auch durch die Anzahl der Publikationen (siehe Kapitel 3.1) belegt: aus QORA und QORA II entstanden insgesamt 22 Publikationen, von denen 11 in begutachteten Fachzeitschriften (darunter 5 Phys. Rev.) angenommen wurden (weitere befinden sich noch in Begutachtung). Auch die Arbeiten mit nicht erfüllten Meilensteinen haben Erkenntnisse geliefert, welche teilweise zu neu definierten Arbeitspaketen mit in der obigen Liste nicht aufgeführten Ergebnissen geführt haben, welche zum Zeitpunkt der Antragstellung noch nicht abzusehen waren.

Ziel von **AP 1** (Physikalische Fehlermodellierung) war eine genaue Modellierung der auf IBMQ Ehningen auftretenden Fehler auf der Basis eines physikalisch motivierten Hamiltonians. Insbesondere berücksichtigt dieser das höhere Energieniveau $|2\rangle$ der Transmon-Systeme, da dieses, wie wir in QORA nachgewiesen haben, in Verbindung mit der Kopplung zwischen Qubits teilweise zu erheblichen Crosstalk-Fehlern bei der Ausführung von CNOT-Gattern führt. Wie in Kapitel 2.1 beschrieben ist es uns gelungen, einen geeigneten Hamiltonian aufzustellen und dessen Parameter so zu bestimmen, dass eine gute Übereinstimmung zwischen Simulation und Experiment erreicht wird. Somit haben wir nun ein mikroskopisches Modell, welches die aufgrund der oben genannten Effekte (höhere Energieniveaus und Kopplung zwischen Qubits) auftretenden Fehler beschreibt (Meilenstein M1). Da dies jedoch aufwändiger war als erwartet, konnten wir

die ursprünglich geplante Erweiterung mit über einfache Dekohärenz- und Relaxationsprozesse hinausgehender Modellierung der Umgebung (Meilenstein M2) nicht innerhalb der Projektlaufzeit umsetzen. Als Schritt in diese Richtung haben wir unser schon in QORA begonnenes effizientes Tomographie-Protokoll weiterentwickelt (ein Artikel hierzu ist noch in Vorbereitung).

In **AP 2** wollten wir hardwarenahe Methoden entwickeln, die das Wissen über die Physik des Systems nutzen, um Fehler auf Quantenrechnern abzuschwächen. Dabei haben wir einerseits erfolgreiche Arbeiten zu Transpilationsverfahren aus QORA weitergeführt und mittels einer selektiven Pulsoptimierung auf bipotenten Architekturen (welche zwei verschiedene Implementierungen des CNOT-Gatters für verschiedene Qubit-Paare aufweisen) weiter verbessert (Meilenstein M3), siehe Kapitel 2.2.1. Daneben wurden weitere Fehlermitigationskonzepte, welche auf Zwischenmessungen und Postselektion basieren, implementiert und getestet. Hierbei wurden die theoretischen Vorarbeiten für Meilenstein M4 durchgeführt, wobei dessen Umsetzung auf dem IBM-Quantencomputer nicht gelang, da die hierfür nötigen dynamischen Schaltkreise nur mit zu geringer Wahrscheinlichkeit ausgeführt werden konnten (siehe Kap. 2.2.2). Ferner wurde die Wirkung von Dynamical Decoupling analysiert (Meilenstein M5), siehe Kap. 2.2.3. Im Zuge dieser Arbeiten ergaben sich wissenschaftlich spannende Fragestellungen mit neuen Ergebnissen zum Qubit-Routing-Problem für Optimierungsprobleme mit raum- und zeitperiodischer Struktur. Diese wurden in einem neuen, im Antrag nicht vorhergesehenen Arbeitspaket 2.5, zu dessen Gunsten das für das Gesamtprojekt nicht kritische AP 3.3 (mit Meilenstein M8, Kap. 2.3.3) zurückgestellt wurde, bearbeitet und erfolgreich gelöst (Kap. 2.2.4).

Während wir uns auf der algorithmischen Ebene des Vorgängerprojekts QORA auf die Standard-Version von QAOA mit verschiedenen Mixern konzentriert haben, wurden in **AP 3** weitere Ansätze der Quantenoptimierung untersucht, die eine höhere Resilienz gegenüber Fehlern erhoffen lassen. Zum innovativen Konzept der dissipativen Quantenoptimierung wurden erste theoretische Analysen durchgeführt und Ergebnisse für 2 Qubits erzielt (Meilenstein M6), wobei sich jedoch die Erweiterung auf größere Anzahl von Qubits (Meilenstein M9) als zu herausfordernd erwies (siehe Kap. 2.3.1). Als weitere Stoßrichtung innerhalb von AP 3 arbeiteten wir an einer besseren Verknüpfung von Quantenoptimierungsalgorithmen mit effizienten klassischen Algorithmen. Als Beispiel hierfür haben wir den Warmstart-QAOA untersucht, welcher von der klassisch effizient berechenbaren Lösung des „relaxierten" (d.h. kontinuierliches statt diskreten) Optimierungsproblem ausgeht. Insbesondere konnten wir zeigen, dass die Ergebnisse des Warmstart-QAOA sich durch ein klassisches Preprocessing des ursprünglichen Optimierungsproblems gefolgt von Standard-QAOA reproduzieren lassen. Außerdem haben wir eine vielversprechende, neue QAOA-Variante entwickelt, welche weniger Aufwand bei der Optimierung der Schaltkreisparameter erfordert (Kap. 2.3.2). Meilenstein M7 wurde somit erfüllt.

In **AP 4** haben wir uns neben der während QORA im Mittelpunkt stehenden Portfoliooptimierung anderen Anwendungen im Finanzbereich zugewandt. Die Arbeiten zur Auswahl von Merkmalen für Maschinelles Lernen wurden (Meilenstein M10, Kap. 2.4.1), zur Optionspreisbewertung mit Quantenalgorithmen (Meilenstein M11, Kap. 2.4.2) und zu Clustering-Algorithmen (Meilenstein M12, Kap. 2.4.3) wurden wie geplant abgeschlossen. Die Untersuchungen zur Abschätzung des Quantenvorteils (Meilenstein M13, Kap. 2.4.4) mittels Skalierungsanalyse ergaben Hinweise auf einen möglichen Quantenvorteil der in AP 3 entwickelten QAOA-Variante im Fall der Portfoliooptimierung.

1.3 Verwertung und Transfer

Das Projekt QORA II hat dazu beigetragen, Kompetenzen im Bereich Quantencomputing in Baden-Württemberg aufzubauen. Das Fraunhofer IAF wurde hierdurch in die Lage versetzt, neben den im Rahmen des Kompetenzzentrums Quantencomputing Baden-Württemberg geförderten Landesprojekten weitere Projektmittel einzuwerben (EU-Projekte AI4QT und SPINUS). Ein Vortrag über die Ergebnisse von QORA bei der KQCBW Developer Conference in Ehningen im November 2022 war Anlass zu Gesprächen mit der Bundesdruckerei GmbH, welche zum Industrieprojekt Quadnis führten.

Die Ergebnisse von QORA II flossen auch in die Vorbereitung der Schulungsmaterialen für das gemeinsam mit Fraunhofer IAO organisierte Schulungsprogramm ein, im Rahmen dessen in zwei Veranstaltungen (Juli und November 2023) Erkenntnisse zum aktuellen Stand des Quantencomputings und seiner Anwendungsmöglichkeiten an Teilnehmer interessierter Unternehmen vermittelt wurden.

Neben den bereits erwähnten Publikationen (siehe Kap. 3.1), waren Mitarbeitende von QORA II auch in zahlreichen Vorträgen und Veranstaltungen bei der Verbreitung der Projektergebnisse aktiv, siehe Kapitel 3.2, einige davon auch mit Publikum aus Industrie und Wirtschaft (z.B. die Messen World of Quantum in München, Quantum Effects in Stuttgart, QBN-Summit in Stuttgart, IT-Tage in Frankfurt a. M.).

Für das mit Hinblick auf zukünftige, praxisorientierte Anwendungen von Quantencomputing vielleicht interessanteste Ergebnis von QORA II halten wir die zum Projektende erhaltene Studie zur „Abschätzung des Quantenvorteils“ (AP 4.4), bei der für verschiedene anwendungsrelevante QUBO-Probleme die Laufzeiten eines leistungsfähigen klassischen Algorithmus mit der einer von uns entwickelten QAOA-Variante direkt verglichen und in einem der drei Fälle (Portfoliooptimierung) Hinweise auf einen möglichen Quantenvorteil gefunden wurden. Da hier längst noch nicht alle Fragen geklärt sind, wollen wir diese Studien auch im Fall der Bewilligung der nächsten Förderperiode fortsetzen und so dazu beitragen, die Potentiale des Quantencomputings für Unternehmen nutzbar zu machen.

Der Kooperationsvertrag wurde von allen Partnern unterzeichnet. Sitzungen des Industriebeirats mit Vertretern von Wüstenrot & Württembergische AG, Bausparkasse Schwäbisch Hall AG und JoS Quantum GmbH fanden am 7. Juli 2023 und am 15. März 2024 statt.

Abschließend seien die im Zuwendungsbescheid geforderten Kennzahlen genannt:

- Anzahl der projektbezogenen wissenschaftlichen Publikationen: 13 (Quellenangabe siehe Kap. 4.1)
- Projektbezogene Patentanmeldungen: keine
- Anzahl der Personen in Forschungseinrichtungen, die am Projekt mitgearbeitet haben: 19
- Davon die Anzahl der Personen, die Schulungen im Quantencomputing erhalten haben: 3 (im Rahmen des Programms von Fraunhofer IAO/IAF 2022)
- Anzahl der am Projekt beteiligten Unternehmen: 3
- Davon die Anzahl der am Projekt beteiligten kleinen und mittleren Unternehmen (KMU): 1

2 Überblick der wissenschaftlichen Arbeiten

Im Folgenden geben wir Überblick der in QORA II durchgeführten Arbeiten, wobei wir gegebenenfalls auf Abweichungen vom geplanten Projektverlauf eingehen.

2.1 Physikalische Fehlermodellierung (AP 1)

2.1.1 1-Qubit-Gatter

Als Erweiterung gegenüber dem QORA-Ansatz wurde in Arbeitspaket AP 1 des QORAII-Arbeitsprogramms von IAF eine physikalische Modellierung der Fehler des IBMQ-Systems durchgeführt. Hierfür wurde zunächst ein geeignetes Hamiltonsches Modell aufgestellt, welches das Netzwerk gekoppelter Transmon-Qubits möglichst gut beschreibt. Dieses beruht auf einem System gekoppelter Duffing-Oszillatoren, welche eine nichtlineare Erweiterung des gewöhnlichen harmonischen Oszillator-Modells darstellen. Sie sind neben einer Eigenfrequenz ω_q durch eine Anharmonizität Δ_q charakterisiert, welche die Abweichung der Energieniveaus von denen des harmonischen Oszillators beschreibt. Kopplungen zwischen den Qubits wurden hierbei als Flip-Flop-Wechselwirkungen angenommen. Zusätzlich nehmen wir an, dass die durch das Mikrowellenfeld induzierten Kopplungsstärken der individuellen Energieübergänge ($|0\rangle \leftrightarrow |1\rangle$, $|1\rangle \leftrightarrow |2\rangle$ usw.) der Transmon-Qubits im Prinzip beliebige Werte annehmen können, weswegen wir diese mit geeigneten Rabi-Experimenten für alle Qubits direkt gemessen haben. Generell beschränken wir uns im Folgenden auf die niedrigsten drei Energieniveaus der nichtlinearen Oszillatoren.

Im Allgemeinen führt hierbei die Anwesenheit des Niveaus $|2\rangle$ zu einer Störung der innerhalb des Qubit-Unterraums ($|0\rangle$ und $|1\rangle$) gewünschten Dynamik. Für 1-Qubit-Gatter kann diese Störung durch einen sogenannten „Drag-Puls“ kompensiert werden, wenn dessen Parameter β geeignet gewählt wird [F. Motzoi *et al.*, PRL 103, 110501 (2009)]. Letzterer kann gemessen werden, indem man den Puls anwendet, diesen anschließend durch einen inversen (d. h. um π phasenverschobenen) Puls negiert und letztlich (nach mehrmaliger Wiederholung dieser Sequenz) die Wahrscheinlichkeit $P_{|0\rangle}$, das Qubit im Zustand $|0\rangle$ zu finden, ausliest (siehe Abb. 1.1).

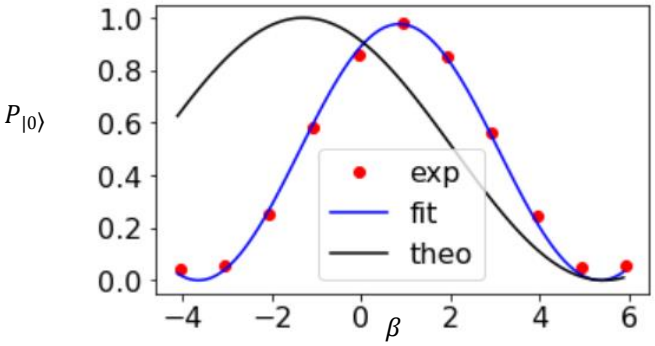


Abbildung 1.1: Resultate eines Experiments zur Kalibrierung des X-Gatters auf Qubit 2 von IBMQ Ehningen. Der durch Fit (blaue Kurve) an die experimentellen Daten (rote Punkte) erhaltene optimale Wert des Drag-Puls-Parameters $\beta_{opt} \approx 1$ (Maximum der blauen Kurve) stimmt nicht mit der Vorhersage aus der Simulation (schwarze Kurve) überein, was vermutlich auf eine Verzerrung des Mikrowellenpulses zurückzuführen ist.

Die gemessenen optimalen β-Werte der einzelnen Qubits (wie z.B. $\beta_{opt} \approx 1$ im Fall des X-Gatters von Qubit 2, siehe Abb. 1.1) variieren jedoch erheblich, was durch das bisher

aufgestellte Hamiltonsche Modell nicht erklärt werden kann und vermutlich auf eine im Experiment auftretende Verzerrung des Mikrowellenpulses zurückzuführen ist. Aus diesem Grund wählen wir im Folgenden die aus der Simulation erhaltene optimale Pulsform, wodurch sich im Fall isolierter Qubits eine mittlere Fehlerrate von ca. 10^{-6} ergibt. Durch die Kopplung an benachbarte Qubits im Zustand $|0\rangle_N$ steigt die Fehlerrate in einigen Fällen bis auf ca. 10^{-4} (z.B. für Qubit 1 und 11, siehe die orangenen Balken in Abbildung 1.2), selbst wenn der Puls in Gegenwart der Kopplung neu kalibriert wird. Wie erwartet liegen die Fehlerraten der simulierten Gatter jedoch alle unter den von IBM angegebenen, mittels Randomized Benchmarking (RB) mit Nachbarn im Zustand $|0\rangle_N$ gemessenen Werten (blaue Balken, wobei in den RB-Schaltkreisen X- und $\sqrt{X}$-Gatter im Verhältnis 1:4 auftreten). Hieraus folgt, dass bei korrekter Kalibrierung der Gatter die durch die Kopplung an die Umgebung verursachten Dekohärenz- bzw. thermischen Relaxationseffekte überwiegen. Letztere haben wir durch eine einfache Mastergleichung berücksichtigt (rote Balken, basierend auf den von IBM angegebenen T_1- und T_2-Werten) und dadurch grobe Übereinstimmung mit den IBM-Werten erreicht.

Werden nun in der Simulation die Nachbar-Qubits im Zustand $|1\rangle_N$ initialisiert, nehmen die kohärenten Fehler aufgrund der Anwesenheit des Niveaus $|2\rangle$ zum Teil stark zu (siehe grüne Balken in Abbildung 1.2). Im Extremfall von Qubit 3 werden dadurch Fehlerraten von bis zu 10^{-2} verursacht, was mit der in Ref. [14] berichtete Frequenzkollision der Qubits 2, 3 und 5 bei Anwesenheit des Niveaus $|2\rangle_3$ erklärt werden kann.

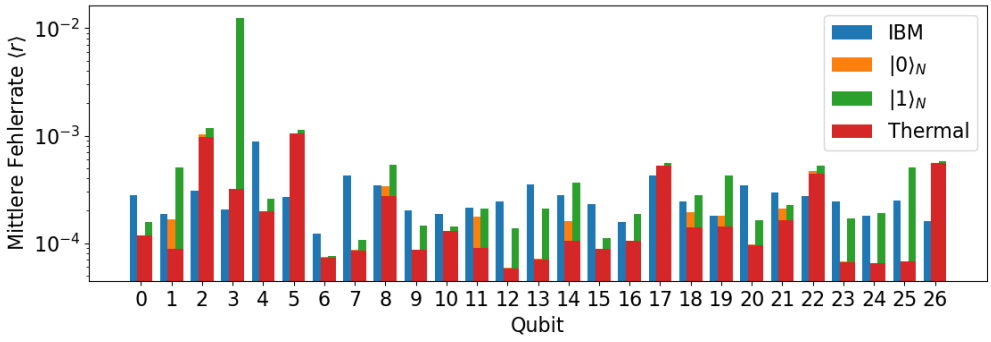


Abbildung 1.2: Mittlere Fehlerrate $\langle r\rangle = \frac{4\,r_{\sqrt{X}}+r_X}{5}$ der mit unserem Modell simulierten X- und $\sqrt{X}$-Gatter für Nachbarn im Zustand $|0\rangle_N$ (orange) und $|1\rangle_N$ (grün) im Vergleich zu den von IBM angegebenen Fehlerraten (blaue Balken) für all 27 Qubits von IBMQ Ehningen. Der simulierte Fehler setzt sich aus einem kohärenten, auf Wechselwirkungen zwischen einzelnen Qubits zurückgehenden Anteil (orange/ grüne Balken) sowie einem sich aus den von IBM angegebenen T_1- bzw. T_2- Relaxations bzw. Dekohärenzzeiten ergebenden thermischen Anteil (rote Balken) zusammen. Falls kein orangener Balken sichtbar ist, ist der kohärente Fehler kleiner als 10^{-5}.

2.1.2 2-Qubit-Gatter

Um obiges Hamiltonsche Fehlermodell auf beliebige Quanten-Schaltkreise zu erweitern, müssen nun zusätzlich auch Multi-Qubit-Gatter modelliert werden. Auf der IBM-Hardware werden diese durch sogenannte Cross-Resonance-Gatter zwischen zwei benachbarten Qubits realisiert. Benachbart bedeutet in diesem Fall, dass die Qubits durch einen Resonator miteinander gekoppelt sind. Diese Kopplung erzeugt ein hybrides System, in dem die Energieniveaus und damit Resonanzfrequenzen der Qubits leicht von dem Zustand des benachbarten Qubit abhängen. Das Cross-Resonance Gatter treibt eines der Qubits (das *Kontroll*-Qubit K) durch einen Mikrowellenpuls mit der Resonanzfrequenz des zweiten Qubits (das *Ziel*-Qubit Z). Das Kontroll-Qubit wird also nicht-resonant getrieben und ändert seinen Zustand nahezu nicht. Das Ziel-Qubit dagegen erfährt über die Kopplung einen resonanten Puls und durchläuft eine Dynamik, die durch die oben erwähnte Hybridisierung maßgeblich vom Zustand des Kontroll-Qubits abhängt. Startet das Kontroll-Qubit im Zustand $|0\rangle_K$, so ergibt sich für das Ziel-

Qubit eine andere Dynamik als im Fall des Anfangszustands $|1\rangle_K$. Folglich kann das Cross-Resonance-Gatter Verschränkung erzeugen, wen das Kontroll-Qubit beispielsweise im Zustand $|+\rangle_K = (|0\rangle_K + |1\rangle_K)/\sqrt{2}$ startet. Durch Hintereinanderschaltung von Cross-Resonance- und Ein-Qubit-Gattern kann das bekanntere CNOT-Gatter implementiert werden. Die Funktionsweise des Cross-Resonance Gatters ist in Abbildung 1.3. skizziert.

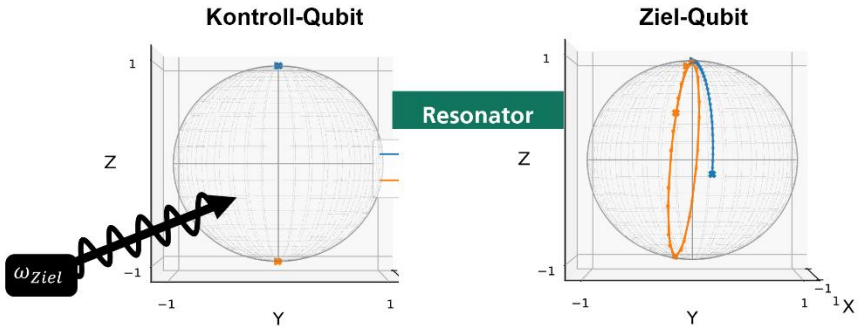


Abbildung 1.3: Skizze der Funktionsweise des Cross-Resonance-Gatters. Zwei anhand ihrer Bloch-Sphären dargestellte Qubits sind über einen Resonator gekoppelt. Das Kontroll-Qubit (links) wird von einem Puls mit der Frequenz des Ziel-Qubits (rechts) getrieben. Da der Puls für das Kontroll-Qubit nicht resonant ist, ändert dieses seinen Zustand kaum. Das Ziel-Qubit dagegen ist resonant und ändert seinen Zustand abhängig vom Zustand des Kontroll-Qubits ($|0\rangle_K$ in blau und $|1\rangle_K$ in orange). Das Ziel-Qubit startet im Zustand $|0\rangle_Z$ (Nordpol der Bloch-Sphäre), und die Linien zeigen die Dynamik in Abhängigkeit der Dauer des Cross-Resonance Pulses. In der Regel wird die Dauer des Pulses so gewählt, dass die finalen Zustände (große Kreuze) sich auf der Bloch-Sphäre genau gegenüber liegen. Für die Implementierung des sogenannten direkten CNOT-Gatters muss das Ziel-Qubit nur noch durch Ein-Qubit-Gatter in die Standardbasis (Nord- und Südpol) rotiert werden.

In Erweiterung des in Meilenstein M1 entwickelten Fehlermodells müssen also zusätzliche Qubits und deren Kopplungen im Hamiltonoperator berücksichtigt werden. Wie zuvor werden die Qubits als nichtlineare Oszillatoren mit drei Energieniveaus $|0\rangle, |1\rangle$ und $|2\rangle$ genähert. In Abbildung 1.2 ist erkenntlich, dass benachbarte Qubits einen relevanten Einfluss auf die Genauigkeit der Ein-Qubit-Gatter haben. Dieser Effekt kann für das Cross-Resonance-Gatter noch verstärkt auftreten [14]. In dem für IBM typischen Layout von *ibmq_ehningen* haben zwei benachbarte Qubits insgesamt zwei oder drei weitere Nachbarn. Das Modell muss somit die Dynamik von bis zu $3^{2+3} = 243$ gekoppelten Zuständen berücksichtigen. Neben den bereits ermittelten Kopplungsstärken der Mikrowellenpulse and die Energieniveaus $|2\rangle$ berücksichtigt dieses erweiterte Modell zusätzlich die Kopplungsstärken J zweier benachbarter Qubits und die Kopplungsstärken der Resonatoren an die Energieniveaus $|2\rangle$. Neben diesen erwartbaren Faktoren mussten zwei weitere Terme im Hamiltonoperator hinzugefügt werden, damit das Modell die experimentellen Daten von *ibmq_ehningen* mit hoher Genauigkeit reproduzieren kann:

(i) Der Cross-Resonance Puls scheint zu rund 1% in einen falschen Resonator zu streuen und damit das Target-Qubit direkt zu treffen. Die Amplitude und Phase dieser Streuung werden mit zwei zusätzlichen Parametern beschrieben.

(ii) Die Kopplung der Transmon-Qubits über einen Resonator führt zu der oben beschrieben Hybridisierung. Jedoch stimmt das Modell mit dem theoretisch erwarteten Wert für die Frequenzverschiebung nicht mit den experimentellen Daten überein. Diese Diskrepanz wird über einen weiteren Parameter korrigiert, der die Energie des Zustands $|1\rangle_K \otimes |1\rangle_Z$ anpasst.

Die experimentellen Daten für den Fit des Modells wurden mit *ibmq_ehningen* durchgeführt. Dafür wurde jedes Paar zweier benachbarten Qubits in den Zuständen $|0\rangle_K \otimes |0\rangle_Z$ beziehungswiese $|1\rangle_K \otimes |0\rangle_Z$ präpariert und der Zustand des Ziel-Qubits nach der Anwendung von Cross-Resonance Pulsen verschiedener Dauern gemessen. Danach wurde das Experiment unter Vertauschung des Kontroll- und Ziel-Qubits wiederholt. Die experimentellen Daten sind exemplarisch für Qubits 0 und 1 in Abbildung 1.4. in blau ($|0\rangle_K \otimes |0\rangle_Z$) und orange ($|0\rangle_K \otimes |1\rangle_Z$) dargestellt. Aufgrund des Rechenaufwands des Modells und der Menge an Parametern wurde der Fit des Modells an die Daten in mehreren Schritten ansteigender Komplexität vorgenommen. Dafür wurde zuerst eine einfache Rabi-Oszillation an das Ziel-Qubit gefittet und daraus Startwerte für die Parameter im Hamiltonoperator extrahiert. Daraufhin wurden mehrere Fits des Hamiltonschen Modells durchgeführt, indem das Hochfahren des Pulses und die unbeteiligten Nachbarn zuerst ignoriert und erst in späteren Durchgängen hinzugefügt wurden. Die Resultate des an die Daten angepassten Modells für Qubits 0 und 1 sind in Abbildung 1.4 dargestellt.

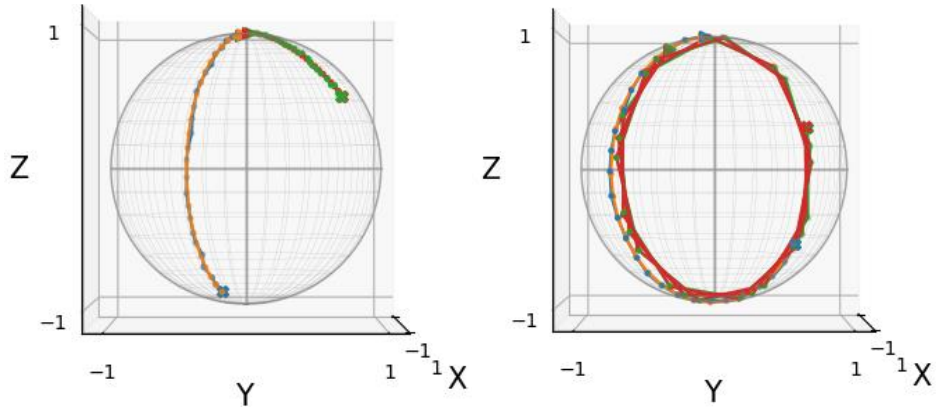


Abbildung 1.4: Experimentelle und simulierte Daten exemplarisch für die Qubits 0 und 1. Gezeigt wird die Dynamik des Ziel-Qubits unter Anwendung von Cross-Resonance-Pulsen unterschiedlicher Längen. In blau (grün) sind die experimentellen Daten für den Anfangszustand $|0\rangle_K \otimes |0\rangle_Z$ ($|1\rangle_K \otimes |0\rangle_Z$) dargestellt und in orange (rot) die zugehörigen simulierten Daten aus dem Hamiltonschen Modell. In der linken Figur ist $K = 0$ und $Z = 1$ und rechts $K = 1$ und $Z = 0$. Die Simulationen stimmen gut mit den experimentellen Daten überein.

Mit den nun angepassten Parametern ist das Hamiltonsche Modell für Zwei-Qubit-Gatter vollständig. Zusammen mit dem Modell für Ein-Qubit-Gatter können jetzt alle von IBM implementierten Pulse inklusive ihrer Fehler aufgrund der Energieniveaus $|2\rangle$ sowie durch unerwünschte Interaktion mit Nachbar-Qubits simuliert werden. Einzig eine virtuelle Z-Rotation des Kontroll-Qubits kann aus den experimentellen Daten nicht bestimmt werden und fehlt bisher im Modell. Für die Berechnung der Fehlerraten von CNOT-Gattern kann das Fehlermodell jedoch auch ohne diese Phase eine untere Schranke angeben. In Abbildung 1.5 werden die von IBM angegebenen CNOT-Fehlerraten aller Zwei-Qubit-Verbindungen auf *ibmq_ehningen* (blau) mit den simulierten Werten für zwei unterschiedliche Zustände der Nachbar-Qubits (N) verglichen. Alle Nachbarn werden entweder im Zustand $|0\rangle_N$ (orange) oder im Zustand $|1\rangle_N$ (grün) präpariert. Zusätzlich zu den so simulierten kohärenten Fehlern wurde aus den T_1- und T_2-Relaxationszeiten ein thermischer Fehler berechnet (rot) und auf den kohärenten Fehler addiert. Es ist erkenntlich, dass die simulierten Fehlerraten oft etwas kleiner sind als die von IBM angegebenen, generell aber in der gleichen Größenordnung liegen. Besonders interessant ist aber der Unterschied zwischen den Simulationen mit $|0\rangle_N$ und $|1\rangle_N$. Hier kommt insbesondere für die Qubit Paare 1_0, 3_2, 3_5 und 16_19 zum Teil zu signifikanten Diskrepanzen. Dies zeigt erneut, dass die Zustände der Nachbarn einen großen Effekt auf die Genauigkeit von CNOT-Gattern haben und zu Fehlerraten von bis zu 20% führen können. Dies ist in guter Übereinstimmung zu den Ergebnissen in Ref.

[14]. (Die in Ref. [14] berichtete Frequenzkollision zwischen Qubits 22 und 24 beim CNOT-Gatter $K = 25$ und $Z = 22$ ist in Abbildung 1.5 nur leicht erkenntlich, da sie das Nachbar-Qubit 24 in den Zustand $|2\rangle_{24}$ anregt, die Fehlerrate des CNOT-Gatters aber nur geringfügig verändert. Dies ist im vollständigen Zustandsvektor aber gut erkenntlich.)

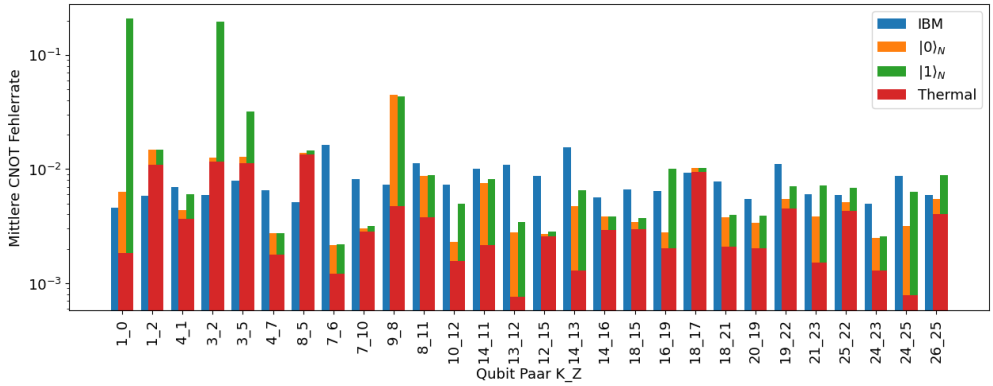


Abbildung 1.5: Mittlere Fehlerraten der CNOT-Gatter auf *ibmq_ehningen*. Die von IBM angegebene Raten (blau) werden mit den aus den T_1- und T_2-Relaxations-/ Dekohärenzzeiten berechneten thermischen Fehlerraten (rot), sowie den aus dem Fehlermodell stammenden und vom Zustand $|0\rangle_N$ (orange) oder $|1\rangle_N$ (grün) der Nachbarn abhängigen Raten verglichen. Auch wenn es zu keiner exakten Übereinstimmung von angegebenen und simulierten Fehlerraten kommt, liegen beide meist in einer ähnlichen Größenordnung. Auffällig sind die Qubit-Paare 1_0, 3_2 und 3_5, bei denen die Fehlerrate mit Nachbarn in $|1\rangle_N$ stark erhöht sind. Ähnliche Phänomene lassen sich auch für die Qubit-Paare 10_12, 16_19, 21_23 und 24_25 erkennen. Dies ist in guter Übereinstimmung mit Ref. [9] und kann durch Frequenzkollisionen erklärt werden. Die hohen simulierten Raten des Qubit-Paars 9_8 lässt sich vermutlich auf einen sub-optimalen Fit aufgrund eines temporär auftretenden „Two-Level-System"-Defektes zurückführen und können ignoriert werden.

Insgesamt wurde damit das Fehlermodell der 1-Qubit-Gatter auf Multi-Qubit-Prozesse und -Gatter erweitert und damit Meilenstein M1 (Fehlermodell mit einfachen Dekohärenzprozessen liegt vor) erfüllt. Für die Erstellung eines vollständigen Fehlermodells zur Simulation beliebiger Schaltkreise inklusive ihrer Fehler muss nun lediglich die Phase der virtuellen Z-Rotation der Ziel-Qubits bestimmt und eingesetzt werden.

2.1.3 Long Sequence Quantum Process Tomography

Ergänzend zur Entwicklung des Hamiltonschen Fehlermodells haben wir die im Vorgängerprojekt QORA begonnene Methode zur Tomographie aller 2-Qubit-CNOT Operationen weiter verfeinert. Hierdurch sollen insbesondere die aufgrund der Kopplung an die Umgebung auftretenden Dekohärenz- und Relaxationsprozesse, welche nicht in dem unserem in 2.1.2. erstellten Hamiltonschen Modell enthalten sind, genauer beschrieben werden. Diese Tomographiemethode stellt die Wirkung jedes Quantengatters durch eine Pauli-Transfermatrix (PTM) dar. Kern der neuen Tomographiemethode ist es, die Charakterisierung von 1-Qubit und 2-Qubit Basisgattern zu trennen, siehe Abbildung 1.6 (a). Dadurch können die 12 Fehlerparameter der PTM jedes 1-Qubit Basisgatters mit Langsequenz-Gattersatztomographie (LS-GST) [E. Nielsen *et. al.*, Quantum **5**, 557 (2021)] durch Ausführung weniger Quantenschaltkreise bestimmt werden. Anschließend werden die 240 Fehlerparameter der verbliebenen PTMs der 2-Qubit Basisgatter mit einer neuartigen Langsequenz-Quantenprozesstomographie (LS-QPT) charakterisiert. Ähnlich wie bei LS-GST wird dabei jedes Gatter mit sogenannten Germ-Schaltkreisen abgewechselt und p mal wiederholt, um die Empfindlichkeit der Methode gegenüber kleinen Gatterfehlern zu erhöhen. Um die numerische Stabilität des

anschließenden Fits der Messdaten zu verbessern, wird jeder Germ-Schaltkreis für verschiedene Tiefen p wiederholt, sodass insgesamt L Wiederholungen pro Germ-Schaltkreis nötig sind. Die Gesamtzahl der auszuführenden Quantenschaltkreise hängt dann von der gewünschten Genauigkeit (d.h. der maximalen Tiefe p) und der Anzahl an verschiedenen Germ-Schaltkreisen ab. Verglichen mit einer 2-Qubit LS-GST, skizziert in Abbildung 1.6(b), benötigt unsere neue Methode deutlich weniger verschiedene Germ-Schaltkreise, sodass – bei gleicher Genauigkeit – insgesamt ca. 20-mal weniger Schaltkreise auf dem Quantenprozessor ausgeführt werden müssen, siehe Abbildung 1.6(c).

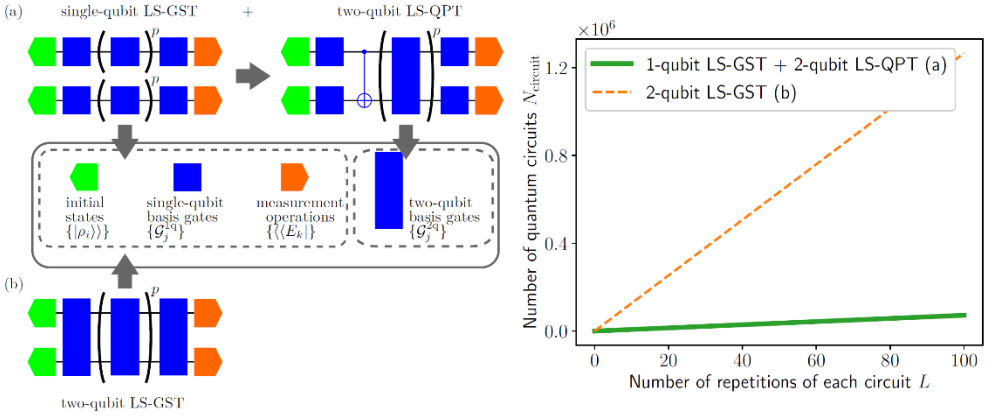


Abbildung 1.6: (a) Arbeitsprinzip der neuen effektiveren Charakterisierungsmethode für Quantenprozessoren: Alle 1-Qubit-Basisgatter werden mit einer Langsequenz-Gattersatztomographie (LS-GST) charakterisiert. Danach werden die restlichen 2-Qubit-Gatter mit der neuartigen Langsequenz-Quantenprozesstomographie (LS-QPT) vermessen. Zur Erhöhung der Genauigkeit werden die zu untersuchenden 2-Qubit-Gatter mit Germ-Schaltkreisen abgewechselt und p mal wiederholt. Jeder Germ-Schaltkreis wird für verschiedene Tiefen p ausgeführt, sodass insgesamt L Wiederholungen pro Germ-Schaltkreis nötig sind. (b) Konventionelle Charakterisierung mittels 2-Qubit LS-GST. (c) Vergleich der Anzahl an Schaltkreisen, die in den beiden Methoden ausgeführt werden müssen. Unsere Methode (grün durchgezogene Linie) benötigt bei gleicher Genauigkeit (gleiche Anzahl L von Wiederholungen jedes Germ-Schaltkreises) insgesamt ca. 20-mal weniger Schaltkreise.

Verglichen mit dem Vorgängerprojekt QORA wurde die Tomographiemethode folgendermaßen optimiert:

- Der 1-Qubit-LS-GST-Algorithmus arbeitet nun mit den Basisgattern $\sqrt{X}$, $R_z(\frac{\pi}{2})$ und dem Identitätsgatter, d.h. den tatsächlichen Basisgattern des *ibmq_ehningen* Quantenprozessors.
- Zusätzlich haben wir neue Germ-Schaltkreise für den LS-QPT-Algorithmus bestimmt, die nun auch auf den Basisgattern von *ibmq_ehningen* basieren. Durch diese Veränderungen wurde die Genauigkeit und Interpretierbarkeit der LS-GSTn und LS-QPTnMethoden verbessert.
- Schließlich haben wir den Fit-Algorithmus der LS-QPT Methode weiter verfeinert. Wie schon im Vorgängerprojekt QORA wird das CNOT-Gatter zunächst mit einem χ^2-Fit rekonstruiert. In einem zweiten Schritt wird ein zusätzlicher Term zur Kostenfunktion hinzugefügt, der Positivität des rekonstruierten CNOT-Gatters als zusätzliche Bedingung einführt. Schließlich wird (im Gegensatz zu QORA) als letzter Schritt noch ein Maximum-Likelihood-Fit durchgeführt, der im Gegensatz zum χ^2-Fit einen unverzerrter Schätzer für die Pauli-Transfermatrix des Quantengatters darstellt.

Die aus dem Fit gewonnene PTM beschreibt, wie das Quantengatter die 16 Basiselemente einer 4x4 Dichtematrix auf andere Basiselemente abbildet. Wir zerlegen die rekonstruierte PTM (G_{meas}) in das Produkt der PTM eines perfekten CNOT-Gatters (G_{ideal}) und einem Fehler (E), der nach dem idealen Gatter wirkt: $G_{meas} = E \cdot G_{ideal}$. Für kleine Abweichungen vom idealen CNOT-Gatter kann der Generator des Fehlers (L) durch einen Matrixlogarithmus bestimmt werden, $L = \log(E)$. Der Generator gibt an, welche kohärenten und dissipativen Prozesse den Gesamtfehler E erzeugt haben. Ein Beispiel für einen solchen Generator ist in Abbildung 1.7 gezeigt.

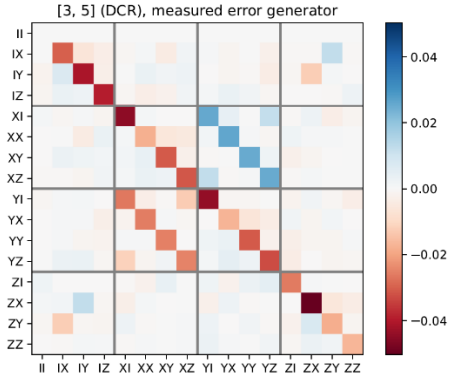


Abbildung 1.7: Pauli-Transfermatrix des Generators des CNOT-Gatterfehlers zwischen Kontroll-Qubit 3 und Ziel-Qubit 5 auf *ibmq_ehningen*, rekonstruiert mit der LS-QPT-Methode. Die grauen Hilfslinien trennen 4x4 Blöcke mit gleichem Basiselement auf dem Kontroll-Qubit.

Die rekonstruierten Generatoren zeigen eine charakteristische Blockstruktur, bei der im Rahmen der Messgenauigkeit lediglich 4x4 Blöcke auf der Haupt- und Nebendiagonalen von Null verschiedene Einträge enthalten (siehe Abbildung 1.7). Ferner zeigen die zentralen vier Blöcke eine charakteristische Struktur der Form

$$\begin{pmatrix} A & B \\ -B & A \end{pmatrix}$$

Diese Symmetrie lässt sich dadurch erklären, dass der CNOT-Gatterfehler nicht zeit-translationsinvariant ist, sondern (insbesondere der auf das Kontroll-Qubit wirkende Teil des Fehlerkanals) von dem spezifischen Zeitpunkt abhängt, zu dem ein CNOT-Gatter ausgeführt wird: Da Kontroll- und Ziel-Qubit verschiedene Resonanzfrequenzen besitzen, rotieren ihre Zustandsvektoren unterschiedlich schnell um die z-Achse, sodass sich eine zeitabhängige relative Phase zwischen Kontroll- und Ziel-Qubit ansammelt, je nachdem, wieviel Zeit seit der letzten Initialisierung des Registers vergangen ist. Diese Phase ist unerheblich für 1-Qubit-Gatter, aber sie wird relevant, sobald ein 2-Qubit-Gatter auf das Qubit-Paar angewendet wird. Für ein perfektes CNOT-Gatter wäre diese Phase irrelevant, da Rotationen um die z-Achse des Kontroll-Qubits mit dem CNOT-Gatter kommutieren. Für ein *fehlerbehaftetes* CNOT-Gatter muss der auf das Kontroll-Qubit wirkende Teil des (eigentlich zeittranslationsinvarianten) Gatterfehlers entsprechend der relativen Phase um die z-Achse rotiert werden. Je nach Ausführungszeitpunkt des CNOT-Gatters im Schaltkreis treten dadurch unterschiedliche effektive Gatterfehler auf. Die LS-QPT-Methode bestimmt den gemittelten Gatterfehler über viele verschiedene Anwendungen des CNOT-Gatters, der dadurch einem über zufällige z-Rotationen des Kontroll-Qubits gemittelten Fehlerkanal entspricht (sogenanntes „twirling“). Dieser mittlere Fehlerkanal muss eine Struktur der folgenden Form haben:

$$\begin{pmatrix} A & 0 & 0 & B \\ 0 & +C & +D & 0 \\ 0 & -D & +C & 0 \\ E & 0 & 0 & F \end{pmatrix}$$

Diese Struktur ist mit der experimentell gemessenen Form der Pauli-Transfermatrix kompatibel. Aktuell führen wir noch genaue Tests dieser Hypothese durch und bereiten eine Publikation vor, die innerhalb der nächsten Wochen auf dem arXiv erscheinen soll.

2.2 Robuste Quantenschaltungen (AP 2)

Die Arbeiten des Instituts für Technische Informatik (USTUTT-ITI) in AP 2 konzentrierten sich auf die Weiterentwicklung und das Benchmarking von Transpilationsverfahren (siehe 2.2.1) und die Untersuchung der Wirksamkeit von Dynamical Decoupling (siehe 2.2.3). Parallel dazu hat das Institut für Funktionelle Materie und Quantentechnologien (USTUTT-FMQ) in Zusammenarbeit mit Uni Konstanz (UKON) Untersuchungen verschiedener aus der Quantenoptik bekannter Methoden zur Erzeugung verschränkter Zustände sowie deren Übertragbarkeit auf gatterbasierte Quantencomputer (siehe 2.2.2) durchgeführt. Außerdem beschrieben sind die Ergebnisse von UKON zum neuen, im Antrag nicht vorhergesehenen Arbeitspaket zum Qubit-Routing-Problem für Optimierungsprobleme mit raum- und zeitperiodischer Struktur (siehe 2.2.4).

2.2.1 Weiterentwicklung und Benchmarking von Transpilationsverfahren (AP 2.1, AP 2.3)

Die Transpilation ist ein entscheidender Schritt bei der Ausführung von Quantenalgorithmen auf gegenwärtiger Quantenhardware, die durch Konnektivität und fehlerhafte Operationen begrenzt ist. In QORA haben wir einen Calibration-Aware Transpilationsansatz vorgeschlagen, der die zeitlich variablen Fehlerraten von Quantengeräten und die spezifischen Eigenschaften von variationellen Quantenalgorithmen berücksichtigt. Auf dieser Basis entwickelten und optimierten wir in QORA II eine Transpilationsmethodik, die die Schaltung zunächst auf Subtopologien des Zielgeräts anpasst und schließlich auf das Zielgerät abbildet.

Um Quantenalgorithmen optimal zu implementieren, fügen wir zuerst SWAP-Gatter hinzu, um sicherzustellen, dass der Quantenalgorithmus die Konnektivitäts-beschränkungen einer Subtopologie erfüllt. Die resultierende Schaltung für einen Fünf-Qubit-QAOA bei einer Tiefe von $p = 1$ ist in Abb. 2.1 dargestellt. Anschließend zerlegen wir die Schaltung in die nativen Basisgatter und optimieren sie, um ausführbare Schaltungen zu erzeugen, die auf das Zielgerät zugeschnitten sind. Schließlich verwenden wir eine Kostenfunktion zusammen mit aus den Kalibrationsdaten erhältlichen Informationen über die Fehlereigenschaften des Geräts, um die vorbereitete Schaltung optimal auf das Zielgerät abzubilden, wobei wir diejenigen Qubits auswählen, welche die Schaltungsfehler minimieren.

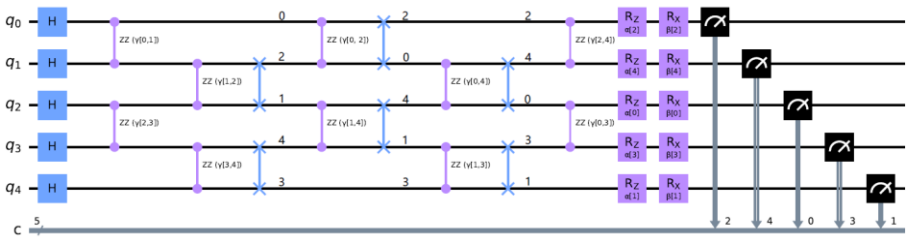


Abbildung 2.1: Qubit-Mapping-Lösung für ein Fünf-Qubit-QAOA bei einer Tiefe von $p = 1$ durch Einfügen von SWAP-Gattern, um die Konnektivitätsbeschränkungen einer linearen Subtopologie zu erfüllen.

Im Vergleich zu der im Vorgängerprojekt QORA entwickelten Calibration-Aware Transpilationsmethode verbesserten wir die Skalierbarkeit und Effizienz des Transpilationsprozesses erheblich. Wichtig ist, dass wir optimale Lösungen für

variationelle Quantenalgorithmen auf gemeinsamen Subtopologien bereitstellen. Diese optimalen Lösungen sind sowohl für jede Qubitzahl als auch für jede Tiefe von QAOA z.B. für die Portfoliooptimierung skalierbar. Unsere Methode weist im Vergleich zu anderen Ansätzen eine minimale Anzahl von CNOT-Gattern und eine minimale Schaltungstiefe auf. Darüber hinaus können diese Lösungen für andere variationelle Quantenalgorithmen verwendet werden. Abb. 2.2 zeigt eine Lösung für den Variational Quantum Eigensolver (VQE) mit einem sogenannten „vollständig verschränkten" Ansatz. Wir beobachten, dass das SWAP-Gatter vor dem Hadamard-Gatter angeordnet ist, um die Vorteile der CNOT-Gatteraufhebung bei der Kombination von CNOT und SWAP zu nutzen. Dies demonstriert die Flexibilität unserer Methode für die Anwendung auf andere Algorithmen. Darüber hinaus sind die Abbildungslösungen mühelos auf andere Quantengeräte anwendbar und können auf der Grundlage ihrer spezifischen Topologie und Fehlerinformationen auf diese Geräte abgebildet werden.

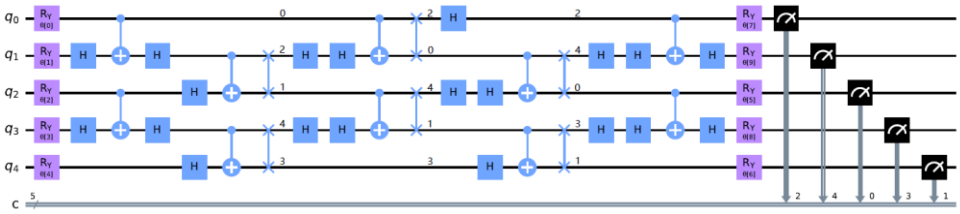


Abbildung 2.2: Qubit-Mapping-Lösung für fünf Qubits VQE mit einem vollständig verschränkten Ansatz bei einer Tiefe von $p = 1$ auf einer linearen Subtopologie. Experimente zeigen, dass unsere Methodik bessere Ergebnisse als herkömmliche Transpilationsstrategien erzielt und somit einen praktischen Kompilationsrahmen zur Verbesserung der Algorithmenleistung auf zukünftigen Quantengeräten bietet.

Um die Leistung des Algorithmus während des Transpilationsprozesses weiter zu verbessern, untersuchen wir die Eigenschaften des Ziel-Quantengeräts und schlagen einen Ansatz zur selektiven Pulsoptimierung vor, der nur die Gatter auf bestimmten Qubit-Paaren mit höheren Fehlerraten optimiert. Diese Arbeit wurde in Physical Review A unter dem Titel "Optimizing quantum algorithms on bipotent architectures" veröffentlicht [15]. Wir charakterisieren zunächst die bipotenten Quantenarchitekturen, bei denen zwei verschiedene Typen eines bestimmten Gatters existieren. Wie in Abb. 2.3 dargestellt weisen die beiden CNOT-Gattertypen, das direct CNOT- und das Echoed-Cross-Resonance (ECR)-CNOT-Gatter, unterschiedliche Eigenschaften auf.

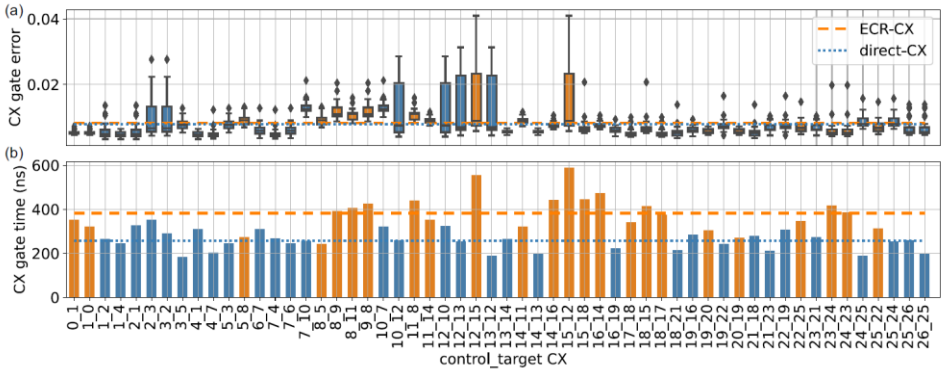


Abbildung 2.3: Vergleich der Eigenschaften von zwei Arten von CNOT-Gattern: direct CNOT und ECR-CNOT. (a) Gatter-Fehlerrate und (b) Gatter-Zeit.

Während ihre Fehlerraten vergleichbar sind, hat das direct CNOT-Gatter im Vergleich zum ECR-CNOT-Gatter eine deutlich geringere Dauer, was es insgesamt zu einem qualitativ hochwertigeren Gatter macht. Außerdem stellen wir fest, dass die Qubits, die

mit zwei Arten von CNOT-Gattern verbunden sind, deutlich längere Dekohärenzzeiten aufweisen als diejenigen, die nur mit einer Art von CNOT-Gatter verbunden sind [15]. Dies deutet auf einen Vorteil hin, wenn man eine Mischung aus direct CNOT- und ECR-CNOT-Gattern verwendet.

Die Leistung des QAOA wird erheblich von der Qualität der dominanten ZZ-Gatter beeinflusst. Wir nutzen die Eigenschaften bipotenter Architekturen und schlagen eine Methode vor, die die ZZ- und ZZ-SWAP-Gatter im Algorithmus, der durch ECR-CNOT-Gatter mit geringerer Qualität implementiert wird, individuell optimiert. Unser Ansatz zeigt eine beträchtliche Verbesserung auf zwei IBM-Quantengeräten, wie in Abb. 2.4 dargestellt. Die Ergebnisse zeigen, dass unsere Methode („Bipotent-ZZ-SWAP$_{OPT}$“) ein höheres approximation ratio und eine höhere Wahrscheinlichkeit der optimalen Lösung als ohne Pulsoptimierung („Bipotent-Default“) erzielt.

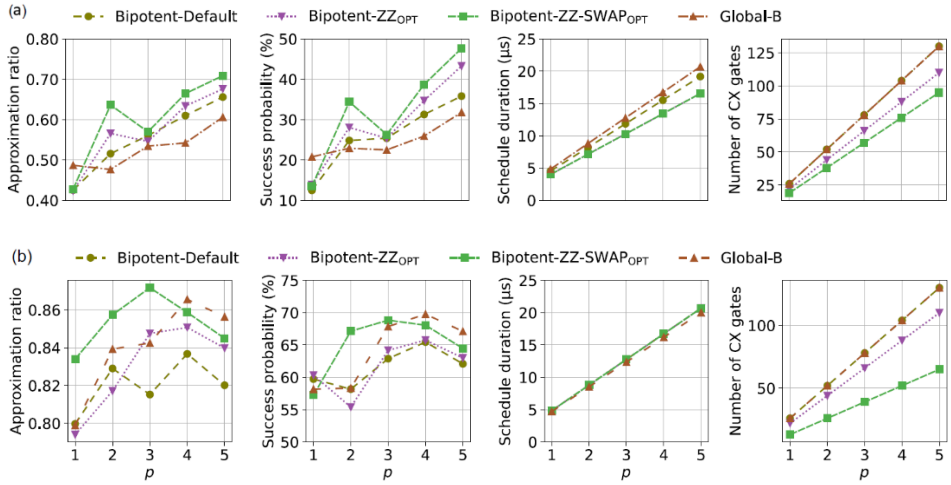


Abbildung 2.4: Selektive Pulsoptimierung von Fünf-Qubit-QAOA für (a) Portfolio-Optimierung auf ibmq_ehningen und (b) ein MaxCut-Problem auf ibm_auckland [10].

Die vorgeschlagene Methode kann auf andere Quantenalgorithmen erweitert werden und ist nicht auf QAOA beschränkt. Die Kalibrierung eines Quantengatters kann zwar ein zeitaufwändiger Prozess sein, die selektive Kalibrierung bietet aber einen effizienten Ansatz, indem nur Gatter auf Qubit-Paaren mit geringerer Qualität kalibriert werden. Dadurch wird der Aufwand für die Kalibrierung reduziert und gleichzeitig eine bessere Leistung erzielt. Darüber hinaus lässt sich die Methode an andere, in naher Zukunft entstehenden Quantengeräte anpassen, die andere native Basisgatter besitzen. Durch die Anpassung von Quantenalgorithmen an die spezifischen Eigenschaften des Ziel-Quantengeräts können wir ihre Leistung weiter verbessern.

Durch die oben beschriebenen Arbeiten wurde Meilenstein M3 („Weiterentwickelte Transpilationsverfahren bereit und gebenchmarkt“) erreicht. Im weiteren Projektverlauf wurden die Auswirkungen der Fehlerreduzierung durch dynamische Entkopplung (siehe 2.2.3) untersucht.

2.2.2 Zwischenmessungen u. Postselektion (AP 2.2)

Ein besonderes Augenmerk wurde in diesem in Zusammenarbeit von USTUTT-FMQ und UKON durchgeführten Arbeitspaket auf Methoden gelegt, welche Zwischenmessungen und Postselektion verwenden, um die Erzeugung verschränkter Zustände robuster zu machen.

GHZ-Zustände

Begonnen wurde mit der Untersuchung einer neuen Methode, die mithilfe von Zwischenmessungen einer Untermenge von Qubits die Erzeugung von verschränkten Zuständen ermöglicht [Piroli *et al.,* PRL **127**, 220503 (2021)]. Ausgehend von linearen Clusterzuständen können dadurch GHZ-Zustände mit konstanter Tiefe erzeugt und eventuelle Fehler korrigiert werden, indem jedes zweite Qubit zwischengemessen wird. Diese Methode kann einerseits durch gleichzeitige Messung aller Qubits und Postselektion, andererseits durch Ausnutzen der von IBM zur Verfügung gestellten Möglichkeit, dynamische Schaltkreise auszuführen, umgesetzt werden. Hierfür wurden verschiedene Techniken analysiert, welche sich für die Charakterisierung der dynamisch erzeugten GHZ-Zustände eignen, sprich besser skalieren als vollständige Quantenzustandstomographien und damit auch für mehr als fünf Qubits ausgeführt werden können. Besonders geeignet hierfür scheinen Parity Oscillations [Cruz *et al.*, Adv. Quantum Technol. **2**, 1900015 (2019] und die Auswertung von Entanglement Witnesses [Tóth, Gühne, PRL **94**, 060501 (2005)].

Darüber hinaus wurde die Methode erweitert, um Gruppen von GHZ-Zuständen mithilfe von Zwischenmessungen und ebenfalls mit konstanter Tiefe zu größeren GHZ-Zuständen zu verbinden. Hierfür wurde ein Algorithmus entworfen, der die Quantenschaltkreise zur Erzeugung von GHZ-Zuständen aus verschieden großen und vielen Gruppen beschreibt. Die Größe der einzelnen GHZ-Zustände hängt hierbei von der zur Verfügung gestellten Tiefe ab. Die Auswertung der Ergebnisse erfolgt identisch zur unveränderten Methode.

Im nächsten Schritt sollten die unveränderte und die erweiterte Methode auf IBMQ Ehningen für verschiedene Anzahlen von Qubits umgesetzt und mit der Standardmethode zur Erzeugung von GHZ-Zuständen verglichen werden. Aufgrund der von der Qubitanzahl unabhängigen konstanten Tiefe verspricht die neue Methode robustere Ergebnisse.

Um die Methode zu testen, wurden verschiedene Experimente gestartet. Durch Probleme mit den IBM-Quantencomputern, insbesondere den dynamischen Schaltkreisen, konnten hierbei jedoch keine verwertbaren oder reproduzierbaren Ergebnisse gewonnen werden. Zum momentanen Zeitpunkt lassen sich die dynamischen Schaltkreise nur mit zu geringer Wahrscheinlichkeit ausführen. Dass die unveränderte Methode Erfolg verspricht, konnte postselektiv in [de Jong et al., Phys. Rev. Research **6**, 013330, (2024)] gezeigt werden.

W-Zustände und andere verschränkte Zustände

Eine genaue Analyse der Methode zur Erzeugung des W-Zustands mittels Zwischenmessungen aus [Piroli et al.] hat ergeben, dass diese Methode - im Gegensatz zur Methode der Erzeugung von GHZ-Zuständen - im Bereich der schaltkreisbasierten Quantenrechner nicht als Schaltkreis konstanter Tiefe betrachtet werden kann.

Deshalb wurde eine Methode zur probabilistischen Fusion von W-Zuständen untersucht [Özdemir *et al.*, NJP **13**, 103003 (2011)]. Mithilfe dieser Methode können große W-Zustände mit konstanter Tiefe erzeugt werden, allerdings nur mit begrenzter Wahrscheinlichkeit. Da für dieses Projekt, sprich für die Ausführung von QAOA, vor allem allgemeine Dicke-Zustände wichtig sind, wurde die Methode erweitert und die Fusion von allgemeinen Dicke-Zuständen untersucht. Hierbei wurde festgestellt, dass die Fusion die charakteristischen Eigenschaften der Dicke-Zustände nicht erhält und die fusionierten Zustände nicht für QAOA verwendet werden können.

Darüber hinaus haben wir eine andere, vielversprechendere Methode zur Erzeugung eines wichtigen verschränkten Zustands untersucht: In [Smith *et al.*, PRX Quantum **4**,

020315 (2023)] wird eine schaltkreisbasierte Methode zur Erzeugung des AKLT-Zustands bei konstanter Tiefe mittels Zwischenmessungen vorgeschlagen. Dort wurde diese Methode zusätzlich auf IBM-Quantenrechnern getestet, was gezeigt hat, dass sie tatsächlich den herkömmlichen Methoden (ohne Zwischenmessungen) überlegen ist.

Wir haben die Methode gründlich analysiert, was zu einer erweiterten Menge von Zuständen geführt hat, die bei konstanter Tiefe erzeugt werden können. Eine ausführliche Suche nach Zuständen, die sich in dieser Menge befinden und gleichzeitig in der Literatur als bekannte Ressourcenzustände für Quantenaufgaben gelten, war jedoch bisher erfolglos. Es wurde festgestellt, dass die verallgemeinerte Methode zur Erzeugung verschränkter Zustände bei konstanter Tiefe eine große lokale Symmetriegruppe („on-site symmetry group") im Vergleich zur virtuellen Dimension der MPS-Darstellung (Matrix Product State representation) des Zustands erfordert. Diese Bedingung scheint nur für wenige Zustände erfüllt zu sein, für die Methoden zur Erzeugung bei konstanter Tiefe auf schaltkreisbasierten Quantenrechnern bereits bekannt sind.

2.2.3 Dynamische Entkopplung (AP 2.4)

Die dynamische Entkopplung (englisch: *dynamical decoupling*, DD) ist eine aufgrund ihrer Einfachheit und ihres minimalen Ressourcenbedarfs für *Noisy Intermediate-Scale Quantum* (NISQ)-Geräte geeignete Technik zur Minderung von Dekohärenzfehlern. Dieser Ansatz nutzt Sequenzen von Kontrollpulsen, die während der Leerlaufzeiten von Qubits angewendet werden. Diese Pulse unterdrücken effektiv unerwünschte Wechselwirkungen der Qubits mit ihrer Umgebung und schützen dadurch den gewünschten Quantenzustand. Im Rahmen des Projekts QORA II untersuchten wir die Wirksamkeit von DD bei der Fehlermitigation für den QAOA-Algorithmus, der auf das Problem der Portfoliooptimierung auf einer Reihe von acht IBM-Quantenrechnern angewendet wurde. Diese Untersuchung zielte darauf ab, die Verbesserung der QAOA-Leistung auf realen NISQ-Geräten durch die Anwendung von DD-Sequenzen zu quantifizieren.

Unsere Experimente beinhalten die Variation der Qubit-Anzahl von 3 bis 12 und eine QAOA-Tiefe von 1, sowie die Untersuchung verschiedener Kombinationen von Algorithmenimplementierungen, einschließlich der CX- und CZ-Versionen von QAOA, sowie Optimierungsstufen 1 und 3 innerhalb des Qiskit-Rahmenwerks. Wir führen diese Experimente mit acht QPUs durch, die aus vier 27-Qubit-Geräten bestehen: ibmq_mumbai, ibmq_kolkata, ibm_cairo und ibmq_ehningen, sowie vier 127-Qubit-Geräten: ibm_kyoto, ibm_cusco, ibm_brisbane und ibm_sherbrooke. Weitere Details finden sich in unserem Preprint [13].

Zur Quantifizierung des Einflusses von Rauschen und der Wirksamkeit von Fehlermitigationstechniken bei Quantenalgorithmen werden die Metriken NAR_B, NAR_D, NSP_B, NSP_D, Δ_{NAR} und Δ_{NSP} herangezogen. NAR_B bzw. NSP_B bezeichnen dabei das bezüglich des fehlerfreien Falles normalisierte Approximation Ratio bzw. die Wahrscheinlichkeit der optimalen Lösung, wenn Algorithmen auf einem realen verrauschten Quantencomputer ohne Fehlermitigation ausgeführt werden. NAR_D und NSP_D stellen dieselben Metriken dar, jedoch unter Anwendung einer bestimmten Fehlermitigationsstrategie, wie beispielsweise DD-Sequenzen. Δ_{NAR} und Δ_{NSP} quantifizieren die durch die Fehlermitigation erzielte Verbesserung, indem sie die Differenz zwischen den entsprechenden Werten mit und ohne Fehlermitigation berechnen. Weiterhin führen wir die Metriken $EMSR_{AR}$ und $EMSR_{SP}$ ein, um die Robustheit einer Fehlermitigationstechnik durch Messung ihrer Fehlermitigationserfolgsrate (EMSR) zu quantifizieren. Die EMSR repräsentiert den Prozentsatz der Durchläufe, bei denen die Anwendung der Mitigationstechnik zu besseren Ergebnissen als ohne Mitigationstechnik führte. Genauer gesagt verwenden $EMSR_{AR}$ und $EMSR_{SP}$ das Näherungsverhältnis bzw. die Wahrscheinlichkeit der optimalen

Lösung, um eine Verbesserung festzustellen. Eine hohe EMSR deutet auf eine konsistente Leistungsverbesserung über die Durchläufe hinweg hin und belegt damit die Robustheit der Mitigationstechnik.

Abbildung 2.5 zeigt einen allgemeinen Trend verbesserter Algorithmenleistung mit zunehmender Schaltkreisgüte. Die Schaltkreisgüte misst die Übereinstimmung zwischen der tatsächlichen Operation einer eines Schaltkreises und ihrer idealen Operation. Sie wird berechnet, indem die Güte aller einzelnen Gatter (Ein- oder Zwei-Qubit) innerhalb des Schaltkreises sowie die Güte der Messung multipliziert werden. DD-Sequenzen verbessern im Durchschnitt sowohl NAR als auch NSP. Allerdings weist NSP einen breiteren Variationsbereich für eine gegebene Schaltkreisgüte im Vergleich zum Approximation Ratio auf, insbesondere bei niedrigeren Güten. Darüber hinaus beträgt der mit dem Approximation Ratio berechnete $EMSR_{AR}$ 85,55%, während der auf der Wahrscheinlichkeit der optimalen Lösung basierende $EMSR_{SP}$ 66,8% beträgt. Im Durchschnitt zeigen DD-Sequenzen eine Erfolgsrate von 76,18% bei der Verbesserung der Algorithmenleistung auf, was ihre Robustheit bei der Minderung von Fehlern in praktischen Quantenalgorithmen hervorhebt.

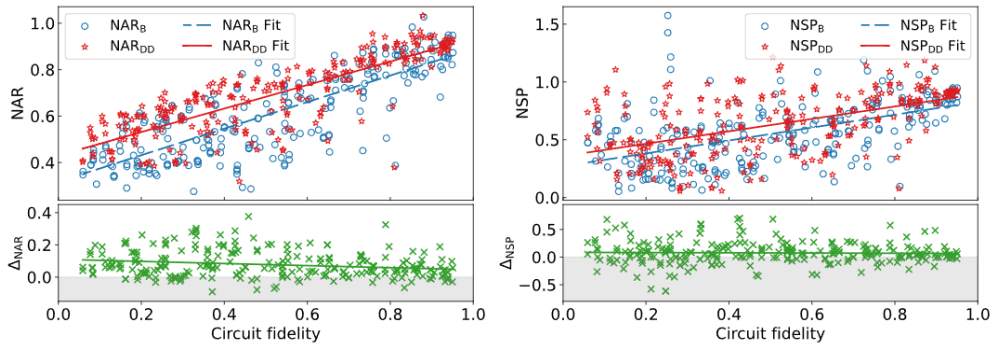


Abbildung 2.5: Einfluss der Circuit Fidelity auf die Algorithmusleistung und die DD-Wirksamkeit. Höhere Werte von NAR_B, NAR_D, NSP_B und NSP_D deuten auf eine bessere Leistung auf realen Quantengeräten hin. Werte über 1 zeigen an, dass die auf der IBM-Quantenhardware erzielte Leistung die Ergebnisse der rauschfreien Simulation übertrifft. Positive Werte von Δ_{NAR} und Δ_{NSP} zeigen Verbesserungen durch DD. Die CPMG-Sequenz wird für alle Datenpunkte verwendet. Jede Linie im Diagramm stellt einen linearen Fit an die Daten dar. Die aus diesen Daten berechneten Fehlermitigationserfolgsraten $EMSR_{AR}$ und $EMSR_{SP}$ betragen 85,55% bzw. 66,8%.

Wir untersuchen nun den Einfluss der Schaltkreisdauer (τ) und insbesondere deren Logarithmus $(\ln(\tau/\mathrm{dt}))$ auf die Algorithmenleistung und die Wirksamkeit von DD. Dabei steht dt für die Systemzykluszeit, die von den Hardware-Eigenschaften abhängig ist. Beispielsweise beträgt 1 dt in ibm_cairo etwa 0,22 ns (2/9 ns), während es in ibm_cusco 0,50 ns sind. Die Schaltkreisdauer spiegelt die Gesamtzeit wider, die zum Ausführen eines Quantenschaltkreises (ohne Messungen) erforderlich ist, und hängt von der Anzahl und Ausführungszeit der einzelnen Gatter ab. Kürzere Schaltkreisdauern verbessern potenziell die Schaltkreisgüte, da das System dann weniger lang Dekohärenzfehlern ausgesetzt ist, erfordern aber schnellere Gatter, deren Güten hardwarebedingt begrenzt sein können. Abbildung 2.6 zeigt die Auswirkung von $\ln(\tau/\mathrm{dt})$ auf die definierten Metriken unter Verwendung derselben Datensätze wie in Abbildung 2.5. Wir beobachten, dass sich die Algorithmenleistung mit zunehmender Schaltkreisdauer verschlechtert, während sich die Wirksamkeit von DD verbessert. Allerdings weist Δ_{NSP} bei längeren Schaltkreisdauern stärkere Fluktuationen auf, was darauf hindeutet, dass DD-Sequenzen zwar Dekohärenzfehler abmildern und damit die Leistung bei längeren Schaltkreisdauern potenziell verbessern können, sie aber auch andere Fehlermechanismen wie Operationsfehler einführen könnten, die diese Verbesserung relativieren.

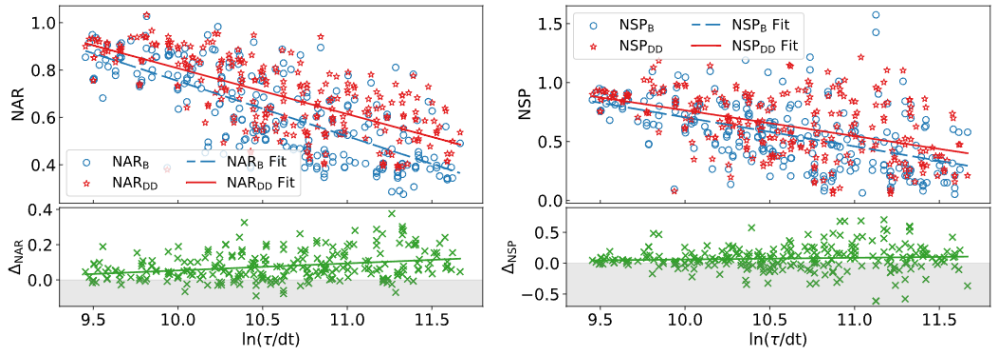


Abbildung 2.6: Einfluss der logarithmischen Schaltkreisdauer $(\ln(\tau/dt))$ auf die Algorithmenleistung und die Wirksamkeit von DD anhand derselben Datensätze wie in Abbildung 2.5.

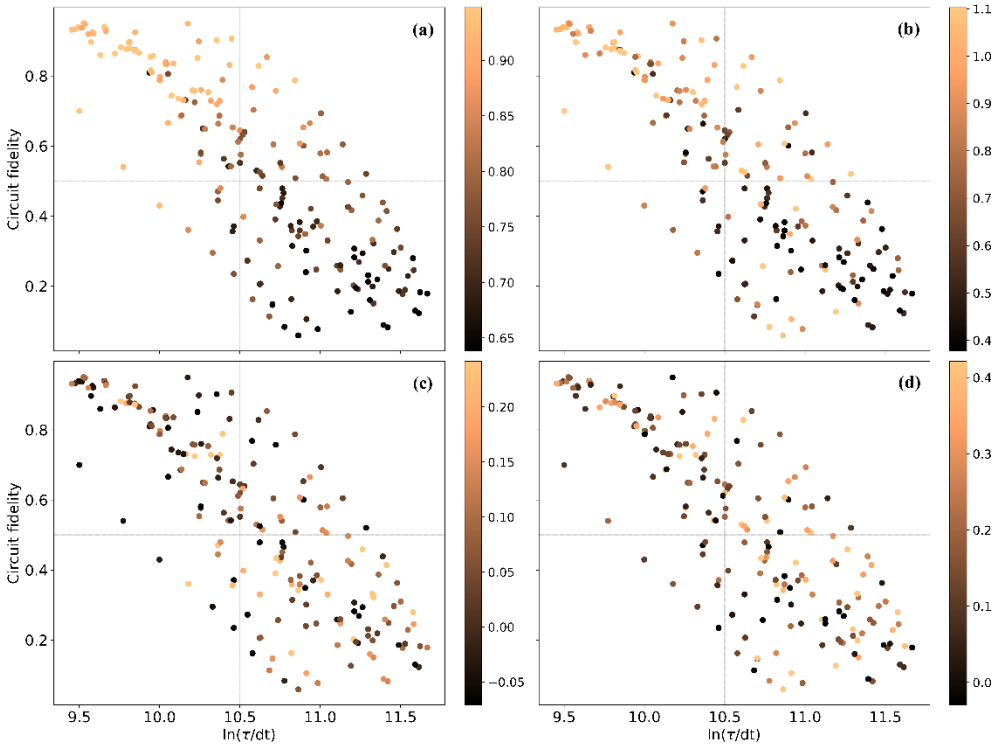


Abbildung 2.7: Einfluss der Schaltkreisgüte und von $(\ln(\tau/dt))$ auf die Algorithmenleistung und die Wirksamkeit von DD: (a) NAR_D, (b) NSP_D, (c) Δ_{NAR} und (d) Δ_{NSP}.

Die Abbildungen 2.7(a-d) veranschaulichen den kombinierten Einfluss der Schaltkreisgüte und -dauer auf NAR_D, NSP_D, Δ_{NAR} und Δ_{NSP}. Wie in Abbildung 2.7(a) zu beobachten ist, nimmt NAR_D mit zunehmendem logarithmischem Wert der Schaltkreisdauer $(\ln(\tau/dt))$ und abnehmender Schaltkreisgüte ab. Hohe Werte des Approximation Ratios konzentrieren sich auf den Bereich, in dem $\ln(\tau/dt)$ unter 2,1 liegt und die Schaltkreisgüte 0,5 übersteigt. Umgekehrt werden kleine Approximation Ratios hauptsächlich für $\ln(\tau/dt)$ Werte beobachtet, die 2,1 dt übersteigen, und für Güten unter 0,5. Ein ähnlicher Trend zeigt sich in Abbildung 2.7(b) für NSP_D. Im Gegensatz zu NAR_D bleibt jedoch auch bei längeren Schaltkreisdauern und niedrigeren Güten die Erreichung eines hohen NSP_D-Wertes möglich. Dies legt nahe, dass NSP_D weniger empfindlich auf diese Faktoren reagiert als NAR_D. Die Abbildungen 2.7(c) und 2.7(d) belegen weiterhin die Wirksamkeit von DD-Sequenzen, insbesondere bei längeren Schaltkreisdauern (höherem $\ln(\tau/dt)$). Dies könnte auf die Fähigkeit von DD-Sequenzen

zurückzuführen sein, Dekohärenzfehler abzuschwächen, die in diesen Zeitmaßstäben prominent werden.
Unsere Ergebnisse deuten darauf hin, dass DD-Sequenzen im Allgemeinen die Leistung von auf IBM-Quantengeräten ausgeführten Quantenalgorithmen verbessern können. Allerdings wird die Wirksamkeit von DD eindeutig durch die inhärente, ohne Fehlerminderung gemessene Performance des zugrundeliegenden Algorithmus beeinflusst. Algorithmen, die durch niedrigere Schaltkreisgüte und längere Ausführungszeiten (Schaltkreisdauern) gekennzeichnet sind, zeigen eine größere Verbesserung durch DD-Sequenzen. Umgekehrt erfahren Algorithmen mit von Natur aus höherer Schaltkreisgüte und kürzerer Ausführungszeit einen geringeren Nutzen von DD. Dieses Verhalten kann auf die intrinsisch niedrigere Fehlerrate hochleistungsfähiger Algorithmen zurückgeführt werden, einschließlich der von DD-Sequenzen angezielten Dekohärenzfehler. Darüber hinaus kann die Einbeziehung von DD-Pulssequenzen neue Gatterfehler einführen und so möglicherweise die Algorithmenleistung verringern und ihre Wirksamkeit für bestimmte Algorithmen begrenzen. Der Meilenstein M5 (Wirksamkeit der Dynamischen Entkopplung auf IBM-QC analysiert) wurde somit erfüllt.

2.2.4 Qubit-Routing-Problem aus Sicht der Vielteilchenphysik (neues AP 2.5)

Eine der Herausforderungen bei der Ausführung von Quantenschaltkreisen auf realen Quantenrechnern besteht darin, die Struktur des Schaltkreises an die Zielstruktur („Hardware Coupling Map“) des Quantenrechners anzupassen, ein Problem, das als „Qubit-Routing-Problem“ bezeichnet wird. Der Schaltkreis wird normalerweise durch die Einfügung von SWAP-Gattern an die Zielstruktur angepasst. Das Qubit-Routing-Problem besteht also darin, zu bestimmen, wo diese SWAP-Gatter eingefügt werden sollen. Die Einfügung von SWAP-Gattern erhöht jedoch die Tiefe des Schaltkreises, was im jetzigen Zeitalter von Noisy Intermediate-Scale Quantum (NISQ) Rechnern unerwünscht ist, da eine Erhöhung der Tiefe eine Erhöhung des Einflusses von Rauschen zur Folge hat. Daher ist eine optimale Lösung des Qubit-Routing-Problems, also die Lösung, die die niedrigste Erhöhung der Schaltkreistiefe zur Folge hat, hilfreich.

Darüber hinaus stellt bei Variational Quantum Eigensolvers (VQEs), wie zum Beispiel QAOA, die genaue Gatterfolge nicht fest, was zur Herausforderung führt, die Gatterfolge zu finden, die zu den niedrigsten Schaltkreistiefen führt – das „Gate-Ordering-Problem“. Es gilt sowohl für das Qubit-Routing-Problem als auch das Gate-Ordering-Problem, dass das Finden der optimalen Lösung NP-hard ist. In der Praxis bedeutet dies, dass es nicht möglich ist, diese Probleme optimal für relevante Problemgrößen und beliebige Probleminstanzen zu lösen.

Es wird jedoch möglich, in bestimmten Fällen das Gate-Ordering-Problem optimal zu lösen, indem wir eines der Grundprinzipien der Vielteilchenphysik verwenden, nämlich das der räumlichen Periodizität. Die Schaltkreise, denen man bei der Quantensimulation der magnetischen Strukturen von Kristallen sowie bei QAOA für Optimierungsprobleme mit einer gewissen periodischen Struktur begegnet, bestehen aus der räumlichen Wiederholung eines bestimmten, „Basis-Schaltkreis“ genannten Unterschaltkreises. In [9] haben wir eine Methode entwickelt, die das Gate-Ordering-Problem optimal und effizient löst, indem wir das Gate-Ordering-Problem nur für den Basis-Schaltkreis lösen. Aus dieser Lösung entsteht durch einfache räumliche Wiederholung die Lösung für Schaltkreise beliebiger Größe mit räumlicher periodischer Struktur, siehe Abbildung 2.8.

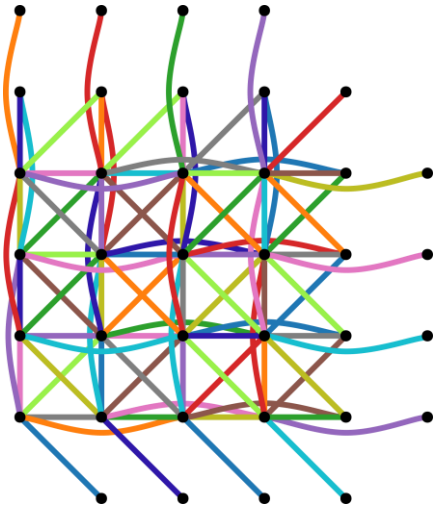


Abbildung 2.8: Optimale Lösung des Gate-Ordering-Problems für einen Quanten-schaltkreis für VQE/QAOA im Fall, dass die Problem-Hamiltonians Wechselwirkungen zwischen Nachbarn, übernächsten Nachbarn und über-übernächsten Nachbarn aufweisen. Gezeigt ist nur ein Basis-Schaltkreis, bei dem jede Farbe die Zwei-Qubit-Gatter einer Ebene des Schaltkreises darstellt. Die gesamte Struktur kann beliebig horizontal und vertikal wiederholt werden, ohne dass es dabei zu Gatterkollisionen kommt. Übernommen aus [9].

Indem wir räumliche Periodizität ausnutzen, wird es auch möglich, das Qubit-Routing-Problem optimal zu lösen. Nicht nur die Quantenschaltkreise besitzen manchmal eine periodische Struktur; auch die Zielhardware hat häufig eine periodische Struktur, zum Beispiel die Heavy-Hex-Coupling-Map von IBM. Hier besteht die Hardware-Coupling-Map aus einer räumlichen Wiederholung eines sogenannten Basis-Graphen. In Arbeit, die momentan in Bearbeitung ist, wird gezeigt, wie sich diese doppelte räumliche periodische Struktur ausnutzen lässt, um das Qubit-Routing-Problem sowie das Gate-Ordering-Problem *gleichzeitig* optimal und effizient zu lösen.

In [5] zeigen wir, dass es auch möglich ist, das Qubit-Routing-Problem effizient zu lösen, wenn wir keine periodischen Strukturen voraussetzen, sondern stattdessen fordern, dass die Struktur des Schaltkreises dem Kantengraphen der Hardware-Coupling-Map entspricht. Die formelle Optimalität unserer Methode wird gerade untersucht. Jedoch wurde gezeigt, dass in Fällen, wo unsere Methode angewendet werden kann, unsere Methode der heuristischen Methode von IBMs Qiskit deutlich überlegen ist [5].

2.3 Resiliente Optimierungsalgorithmen (AP 3)

2.3.1 Dissipative Quantenoptimierung (AP 3.1)

Ein zentrales Thema des Forschungsprogramms an der EKUT im Rahmen von QORA II war die Entwicklung dissipativer Quantenalgorithmen. Angesichts der relativ starken Dekohärenz und Dissipation in der aktuellen Hardware besteht die Idee darin, diese Prozesse zu nutzen, um effiziente Relaxationen zum Grundzustand eines gegebenen Problem-Hamiltons zu konstruieren.

Zu diesem Zweck untersuchten wir verkettete Markovsche Dekohärenzprozesse. Typische Problem-Hamiltonians sind Summen nicht-vertauschender „Wenigkörper-Hamiltonians“, die nur auf wenige Qubits gleichzeitig wirken. Dies wird normalerweise in der Zeitentwicklung durch Trotterisierung berücksichtigt. Für ausreichend kurze Zeitschritte addieren sich die in jedem Zeitschritt angewendeten Wenigkörper-Hamiltonians zum gesamten Hamiltonian. Die Lindbladians in der Markovschen Zeitentwicklung hängen vom Hamilton-Operator des Systems ab. Der Lindbladian, den man zur Relaxation auf den Grundzustand des Problems implementieren möchte, verbindet die Energieeigenräume dieses Hamilton-Operators mit Sprungoperatoren. Allerdings summieren sich die Lindbladians, die den Wenigkörper-Hamiltons

entsprechen, leider nicht zum Lindbladian des Problem-Hamiltons. Darüber hinaus würde eine direkte Konstruktion des Lindbladians des Problem-Hamiltons die Kenntnis der Energieeigenräume erfordern, also mehr als die Lösung des gesuchten Problems.

Angesichts dieser Situation haben wir verschiedene Ansätze untersucht: 1.) Wir haben versucht herauszufinden, welche dissipativen Prozesse aus theoretischer Sicht funktionieren könnten. 2.) Wir erforschten, welche dissipativen Prozesse prinzipiell auf IBM-Q implementiert werden können. 3.) Wir haben Techniken entwickelt, um Dissipations- und Relaxationsprozesse besser zu charakterisieren. 4.) Wir haben die Machbarkeit der Nutzung künstlich induzierter Dissipation untersucht.

Für 1.) begannen wir mit der Entwicklung eines hybriden Quantenalgorithmus, der QAOA ähnelt. Lokale dissipative Prozesse werden über variable Zeiträume angewendet, die als Optimierungsparameter in einer klassischen Schleife verwendet werden, um die gemessene Energie im Endzustand zu minimieren. Dies wurde zunächst mit dem Trotterisierungs-Ansatz durch Simulation zweier Qubits getestet. Ein Näherungsverhältnis der Grundzustandsenergie von nahezu 100 % konnte für ein einfaches Ising-Modell erzielt werden. Im Fall der XX-Wechselwirkung wich der stationäre Zustand jedoch erheblich vom korrekten Zustand für starke Wechselwirkungen ab. Abhilfe schaffte ein Einzel-Iterations-Ansatz, der $p = 1$ in QAOA entspricht, wobei Lindbladians für Einzelqubit-Hamiltonians lokale unitäre Transformationen ersetzen, und ein Lindbladian für die Wechselwirkung den Mischer, alle mit einer variablen Wirkungsdauer. Dies ergab gute Ergebnisse für 2 Qubits mit XX-Wechselwirkungen, wurde jedoch noch nicht auf größere Systeme ausgeweitet [6].

Was 2.) betrifft, haben wir umfangreiche Tests auf öffentlich verfügbarer IBM-Hardware mit wenigen Qubits durchgeführt, um die dissipativen Prozesse einzelner Qubits und die entsprechenden stationären Zustände besser zu verstehen. Das IBM-Fehlermodell erwies sich als unzureichend. Es wurden umfangreiche Literaturstudien und entsprechende Anpassungsversuche sowie ein enger Austausch insbesondere mit der Gruppe in Konstanz durchgeführt, um eine zuverlässige Modellierung der Zerfallsprozesse der Qubits zu finden. Wir untersuchten die Dissipation während zeitabhängigen Treibens und nicht-Markovsche Modelle im Detail. Bisher hat jedoch kein Modell zufriedenstellend funktioniert, wobei das Hauptproblem in der großen zeitlichen Variabilität der Hardware liegt. Ein wichtiges theoretisches Ergebnis ist jedoch, dass es relativ wenig Spielraum gibt (kleiner Parameter = Rabi-Frequenz/Resonanzfrequenz, max. ca 10%), die dissipativen Prozesse durch die Implementierung verschiedener Quantengatter zu modifizieren, da der native Qubit-Hamilton immer „an“ ist [6].

Bezüglich 3.) und um die oben aufgeführten Probleme anzugehen, haben wir die Standardtomographie von Quantengattern erweitert. Zusätzliche Variablen, die die Gedächtniseffekte repräsentieren, werden motiviert durch die allgemeine Form der sogenannten Post-Markovian-Master-Gleichung hinzugefügt. Daher ist eine größere Anzahl von Parametern erforderlich, um ein einzelnes Quantengatter zu charakterisieren. Mit dieser Änderung beobachteten wir eine Verbesserung der Gate-Set-Tomographie bei der Rekonstruktion der Prozessmatrizen von Quantengattern.

Um ein erweitertes Verständnis von Qubit-Kohärenz und -Relaxation auf der IBM-Hardware zu erhalten, haben wir einen muster-basierten Ansatz, der von klassischen Speichertestalgorithmen inspiriert ist, entwickelt [12]. Diese Tests ermöglichten es uns, wichtige Qubit-Eigenschaften wie T_1- und T_2-Zeiten effizient zu extrahieren, die für das Verständnis von Qubit-Kohärenz und -Relaxation wesentlich sind. Darüber hinaus konnten wir Interaktionen zwischen benachbarten Qubits identifizieren und analysieren, um Einblicke in das Verhalten von Qubits in nächster Nähe zueinander zu gewinnen. Durch die Nutzung dieses muster-basierten Ansatzes können wir den Bedarf an der

Bewertung der Leistung und Eigenschaften von zunehmend komplexen Quantensystemen adressieren.

Insgesamt wurde klar, dass es nicht ausreicht, Fehlermechanismen einzelner Qubits isoliert zu betrachten, da Dissipation und Dekohärenz eines Qubits stark von benachbarten Qubits abhängen, und damit eine Betrachtung des gesamten Chips notwendig wird.

2.3.2 Klassische und hybride Algorithmen (AP 3.2)

Ziel dieses Arbeitspakets war es, mögliche effizientere Zugänge zur Quantenoptimierung zu untersuchen, die sich durch die Verknüpfung von klassischen Algorithmen zur Lösung von kombinatorischen Optimierungsproblemen mit Quantenoptimierungsalgorithmen ergeben. Als Quantenalgorithmen standen nach wie vor der „Quantum Approximate Optimization Algorithm" (Standard-QAOA) und dessen Modifizierungen im Zentrum.

Warmstart-QAOA

Insbesondere haben wir den „Warmstart-QAOA", in welchem zunächst die kontinuierliche Relaxation des diskreten Optimierungsproblems klassisch gelöst wird, um damit den QAOA zu initialisieren, implementiert und für verschiedene Sets an Probleminstanzen für Portfoliooptimierung, mit je 20 Instanzen à 10 Qubits, untersucht.

Im ersten Schritt konnten wir zeigen, dass der Warmstart-QAOA, wie in [D. Egger *et al.*, Quantum **5**, 479 (2021)] gezeigt, bessere Ergebnisse als der Standard-QAOA erzielt, insbesondere für jene Probleminstanzen, deren klassische relaxierte Lösung näher an der optimalen Lösung liegt. Weiterhin haben wir ein Verfahren entwickelt, mit dem eine solche Verbesserung in Kombination mit dem Standard-QAOA ebenfalls erreicht werden kann. Dieses Verfahren besteht aus einem rein klassischen Preprocessing des ursprünglichen Optimierungsproblems, welches anschließend durch Standard-QAOA gelöst wird, und leitet sich von folgender Beobachtung ab: Obwohl die Lösungsvariablen des relaxierten Problems im Allgemeinen Werte zwischen 0 und 1 annehmen können, stellt sich heraus, dass einige von ihnen genau auf 0 oder 1 gesetzt werden und sich im Verlauf des Warmstart-QAOA nicht mehr ändern. Da sich diese Bits somit vom ursprünglichen Problem eliminieren lassen, erhalten wir ein reduziertes Problem. Wir haben nun dieses Verfahren Idee erweitert, indem auch jene Variablen, die Werte nahe bei 0 oder 1 aufweisen, bezüglich zweier von uns gesetzter Grenzparametern (δ_0, δ_1) zu 0 oder 1 gerundet und anschließend vom Problem eliminiert werden. Das resultierende reduzierte Problem wird nun für verschiedene Kombinationen der Grenzparameter mit dem Standard-QAOA gelöst.

Zusammenfassend konnten wir zeigen, dass dieser Ansatz für kleine Problemgrößen zu sichtbaren Verbesserungen im Vergleich zum Standard-QAOA und sogar zum Warmstart (für gewisse Probleminstanzen und genügend groß gewählte Grenzparameter) führt, siehe Abbildung 1.3. Allgemeiner haben wir gezeigt, dass sich die verbesserte Performance des Warmstart-QAOA durch klassische Methoden reproduzieren lässt und somit nicht allein durch Quanteneffekte induziert wird. Diese Ergebnisse [1] können mit weiterer Forschung dazu beitragen, ein besseres Verständnis darüber zu erlangen, welche Schritte in der Optimierungsroutine durch effiziente klassische Methoden ersetzt werden können, um Quantencomputing gezielter einzusetzen

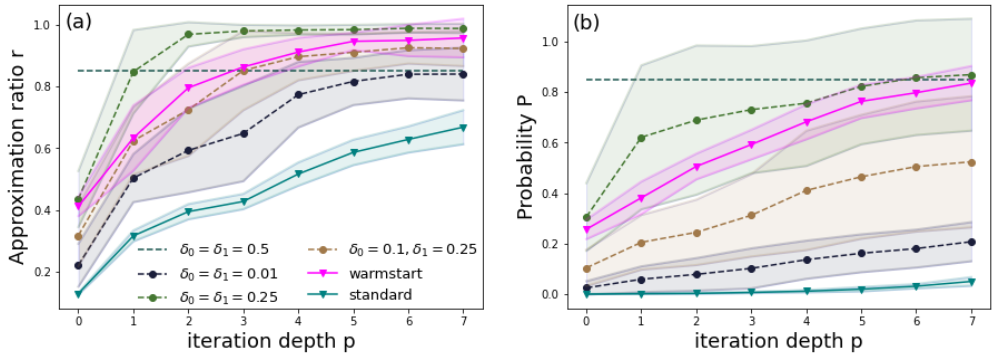


Abbildung 1.3: Vergleich verschiedener QAOA-Varianten für Portfoliooptimierung mit 10 Qubits (Auswahl von 5 aus 10 Aktien). Gezeigt sind jeweils für 20 verschiedene Probleminstanzen der Mittelwert und die Standardabweichung (Fehlerbalken) des (a) approximation ratio r und (b) der Wahrscheinlichkeit P der optimalen Lösung in Abhängigkeit der Iterationstiefe p des QAOA, berechnet mit Standard-QAOA (blaugrün), Warmstart-QAOA (pink) und Standard-QAOA mit klassischem Preprocessing (gestrichelt) für verschiedene Kombinationen der Grenzparameter (δ_0, δ_1). Die Variante mit $\delta_0 = \delta_1 = 0.25$ liefert die besten Ergebnisse und übertrifft insbesondere den Warmstart-QAOA bereits bei kleiner Tiefe p.

QAOA-Schaltkreise großer Tiefe mit wenigen Parametern: diskretes adiabatisches Theorem

In der zweiten Projekthälfte war im Antrag ursprünglich die Untersuchung von klassischen Algorithmen inspirierter, rekursiver QAOA-Varianten geplant. In der Zwischenzeit sind wir jedoch auf eine andere Idee gestoßen, die uns mit Hinblick auf einen möglichen „Quantenvorteil" (siehe Kapitel 2.4.4) erfolgversprechender erschien:

Schon im Vorgängerprojekt QORA hat sich nämlich abgezeichnet, dass mit dem üblicherweise verwendeten variationellen QAOA-Zugang voraussichtlich kein Quantenvorteil zu erzielen ist (siehe auch Abbildung 4.11 weiter unten, welche dem Abschlussbericht von QORA entnommen ist). Grund dafür ist die hohe Anzahl $2p$ von Schaltkreisparametern $(\gamma_1, \beta_1, \gamma_2, \beta_2, \dots, \gamma_p, \beta_p)$, die bei genügend großer Tiefe p (welche nötig ist, um klassische Algorithmen zu übertreffen) zu optimieren sind, und die entsprechend große Zahl dazu erforderlicher Schaltkreisauswertungen, von denen jede einzelne wiederum eine große Anzahl von „Shots" (Schaltkreisdurchläufen mit anschließender Messung) beinhaltet, um den zu minimierenden Erwartungswert der Zielfunktion präzise zu bestimmen. Wünschenswert wäre es daher, gute Werte für diese Parameter auch ohne aufwändiges Optimierungsverfahren bestimmen zu können.

Eine Möglichkeit hierfür besteht in der Analogie zum adiabatischen Quantencomputing, welche auch ursprünglich die Wahl des QAOA-Ansatzes motiviert hat [Farhi 2016]. Hiernach sollten die Parameter γ_i ($i = 1,2, \dots, p$) von Null ausgehend langsam ansteigen, während die Parameter β_i parallel dazu auf Null abfallen, z. B. gemäß folgendem einfachen linearen Ansatz:

$$\gamma_i = m_1 \cdot \frac{i}{p}$$
$$\beta_j = m_2 \cdot \frac{p-j}{p}$$

Wählt man $m_1 = m_2 = \Delta t$ und $p = T/\Delta t$, wobei T der Transformationsdauer sowie Δt der Diskretisierung der kontinuierlichen Zeitentwicklung entspricht, reproduziert man auf diese Weise die dem adiabatischen Quantencomputing zu Grunde liegende

kontinuierliche Transformation des Hamiltonians. Laut Theorie wird dann der gesuchte Grundzustand im Grenzfall $T \to \infty$, $\Delta t \to 0$ erreicht. Dies führt jedoch zu einer extrem hohen, für praktische Anwendungen ungeeigneten Iterationstiefe $p = T/\Delta t$.

Eine gewisse Abhilfe verspricht hier das bislang noch wenig bekannte *diskrete adiabatische Theorem* [Dranov, Kellendonk, Seiler, J. Math. Phys. (1997)]. Die diskrete adiabatische Evolution findet hierbei zwischen dem Eigenvektor $|\psi_0\rangle$ eines unitären Operators U_0 und einem Eigenvektor $|\psi\rangle$ eines anderen unitären Operators U statt und ist durch ein Produkt von sich allmählich ändernden unitären Operatoren beschrieben: $|\psi_p\rangle = U_p U_{p-1} \dots U_1 |\psi_0\rangle$, wobei p die Iterationstiefe angibt. Gilt nun $\|U_j - U_{j-1}\| \propto \frac{1}{p}$ (mit $j = 1, \dots, p$ und $U_p = U$) und weisen die Spektren der U_j's keine Kreuzungen auf, dann approximiert der Zustand $|\psi_p\rangle$ den Zustand $|\psi\rangle$ im Limes $p \to \infty$. Die Anwendung auf QAOA ergibt sich für $U_j = e^{-i\beta_j M} e^{-i\gamma_j F}$ mit dem obigen linearen Ansatz. Im Gegensatz zum kontinuierlichen Fall ist der Limes $\Delta t \to 0$ nun überflüssig, sodass die Parameter m_1 und m_2 nicht gegen Null gehen müssen und der Grenzfall $p \to \infty$ früher erreicht werden kann (siehe auch [V. Kremenetski *et al.*, arXiv:2305.04455v2]).

Mit dem obigen linearen Ansatz sind nun nur noch 2 Parameter (m_1 und m_2) zu optimieren, während gleichzeitig hohe Tiefen p erreicht werden können. In QORA II haben wir die Idee verfolgt, auf die variationelle Optimierung ganz zu verzichten, und stattdessen geeignet Werte von m_1, m_2 und p durch Extrapolation von kleineren hin zu größeren Zahlen von Qubits durchzuführen. Die Ergebnisse sind in Kapitel 2.4.4 gezeigt.

2.3.3 Analyse des Einflusses korrelierter Fehler auf NISQ-Algorithmen (AP 3.3)

Wie oben beschrieben haben die Arbeiten von AP 2.2 zu neuen Fragestellungen bezüglich des Qubit-Routing-Problems geführt, die wir zum Zeitpunkt der Antragstellung von QORA II noch nicht vorhergesehen haben, und somit in ein neues Arbeitspaket 2.5 mündeten. Im Antrag war für UKON stattdessen eine Fortführung von Arbeiten zu korrelierten Fehlern aus QORA geplant (AP 3.3). Da diese mit dem inzwischen in Phys. Rev. A (Editor's Suggestion) veröffentlichten Artikel [17] zu einem befriedigenden Abschluss gelangt sind, wurde die wissenschaftlich weniger interessante Übertragung auf andere Algorithmenklassen zu Gunsten der neuen Arbeiten zurückgestellt.

2.4 Anwendungsfälle im Finanzbereich (AP 4)

Die im Vorgängerprojekt QORA entwickelte verbesserte Methodik zur Lösung von QUBO-Problemen lässt sich neben der Portfoliooptimierung auch auf andere Anwendungsprobleme im Finanzsektor anwenden. Im speziellen hat sich DHBW dem Problem der Merkmalsauswahl beim maschinellen Lernen (AP 4.1, siehe Kap. 2.4.1) und dem Problem der Identifikation von Clustern in nicht-gelabelten Daten (AP 4.3, Kap. 2.4.3) zugewandt, während IAO mit der Optionsbewertung eine andere Klasse von Problemen untersucht hat (AP 4.2, Kap. 2.4.2). Die abschließende Abschätzung des Quantenvorteils für die Anwendungsfälle Portfoliooptimierung, Feature Selection und Clustering (AP 4.4, Kap. 2.4.4) wurde in Zusammenarbeit von IAF und DHBW durchgeführt.

2.4.1 Auswahl von Merkmalen für Maschinelles Lernen (AP 4.1)

Im Fokus von AP 4.1 steht die Frage, welche Merkmale z.B. im Rahmen eines Kundenscoringverfahrens auszuwählen sind, so dass die Trennschärfe des Verfahrens maximiert wird. Die auszuwählenden Merkmale sollten einerseits stark mit der

Zielvariablen korreliert, andererseits untereinander nur schwach korreliert sein. Diese beiden Bedingungen lassen sich als QUBO formalisieren, siehe Abbildung 4.1.

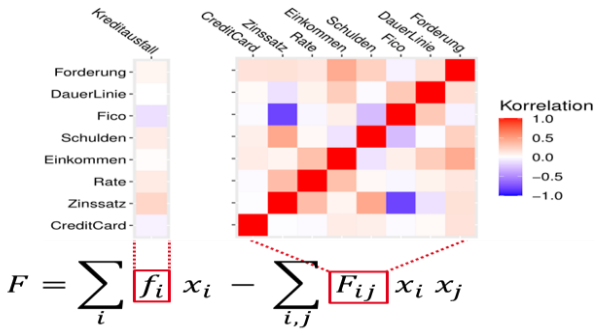


Abbildung 4.1: Formulierung der Merkmalsauswahl als QUBO.

Im nächsten Schritt haben wir mit verschiedenen klassischen Optimierungsverfahren, stochastischen Metaheuristiken, aber auch dem QAOA-Verfahren (mit Simulatoren sowie mit dem Quantensystem in Ehningen) die Lösungen des QUBO-Problems untersucht und verglichen. Grundlage für verschiedene Instanzen des QUBO_Problems sind dabei die Korrelationen zwischen Features von mehreren Benchmark-Datensätzen: „Statlog (German Credit Data)“[1], „Lending Club“[2] und „Breast Cancer Wisconsin“[3].

Dabei zeigt sich folgendes:

- Die Formulierung der Merkmalsauswahl als Optimierungsproblem ist ein gangbarer Weg, um die Lösungen dieses NP-schweren Problems besser einzuschränken und kann durchaus mit etablierten Verfahren wie z.B. LASSO oder RFE konkurrieren.
- Obwohl die Ergebnisse auf der physikalischen Hardware noch von Fehlern auf der Qubit-Ebene überlagert werden, ergeben sich insgesamt plausible Ergebnisse.
- Für die untersuchten Quantenalgorithmen zeigen unsere Ergebnisse, dass ein sorgfältig modifizierter und implementierter QAOA-Algorithmus der Standardlösung von IBM überlegen ist.

Die Arbeiten in AP 4.1 konnten insgesamt abgeschlossen werden und wurden zur Veröffentlichung bei IET Quantum Communication eingereicht [7]. Meilenstein M10 wurde damit erreicht.

2.4.2 Quantenalgorithmen zur Optionsbewertung (AP 4.2)

In diesem Arbeitspaket untersuchte Fraunhofer IAO Quantenalgorithmen zur Optionsbewertung, welche eine quadratische Beschleunigung im Vergleich zu den hierfür auf klassischen Computern durchgeführten Monte-Carlo-Simulationen verspricht. Zunächst wurde das Skalierungsverhalten eines ursprünglich für fehlertolerante Quantencomputer vorgeschlagenen Algorithmus untersucht. Danach wurden hybride Algorithmen bewertet, welche weniger Gatter erfordern und daher für fehlerbehaftete NISQ-Hardware geeigneter sein sollten.

[1] https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data

[2] https://www.kaggle.com/datasets/wordsforthewise/lending-club

[3] https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

Quantenamplitudenschätzung

Im ersten Fall handelt es sich um die auf Quantenphasenschätzung beruhende Quantenamplitudenschätzung (QAE). Das Problem ist hier in Form eines Operators $\mathcal{A}$ gegeben, der wie folgt auf $n+1$ Qubits wirkt:

$$\mathcal{A}\,|0\rangle_n\,|0\rangle = \sqrt{1-a}\,|\psi_0\rangle_n\,|0\rangle + \sqrt{a}\,|\psi_1\rangle_n\,|1\rangle$$

wobei die Größe a zu bestimmen ist. QAE erlaubt die Abschätzung von a mit einer Ungenauigkeit, die wie $1/M$ skaliert, wobei M die Anzahl der Anwendungen von $\mathcal{A}$ bezeichnet. Die auf Quantenphasenschätzung beruhende QAE benötigt jedoch eine große Zahl kontrollierter Operationen mit entsprechend vielen CNOT-Gattern. Aus diesem Grund haben wir als Alternative einen hybriden Ansatz verfolgt, welcher ein auf einer Maximum-Likelihood-Methode basierendes klassisches Postprocessing beinhaltet [Suzuki *et al.*, Quantum Inf. Process. **19**, 75 (2020)].

Als konkretes Szenario haben wir einen einfachen Fall einer „European Call Option" untersucht, bei dem der Wert des der Option zugrunde liegenden Wertpapiers in zwei Bits kodiert ist, siehe Abbildung 4.2.

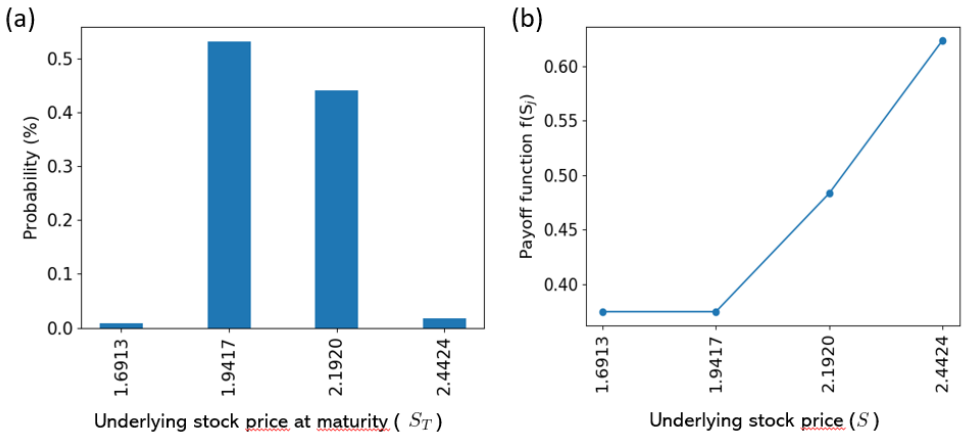


Abbildung 4.2: (a) Diskretisierte Wahrscheinlichkeitsverteilung $p(S)$ des Wertes des der Option zugrunde liegenden Wertpapiers zum Fälligkeitsdatum. (b) Dazugehörige Payoff-Funktion $f(S)$, welche den Wert der Call-Option bestimmt.

Der diesem Problem entsprechende Operator $\mathcal{A}$ wird dann wie folgt konstruiert [Rebentrost *et al.*, PRA 98, 022321 (2018)]:

$$\mathcal{A}|0\rangle_n|0\rangle = \sum_{j=0}^{N-1}\sqrt{1-f(S_j)}\sqrt{p_j}|S_j\rangle|0\rangle + \sqrt{f(S_j)}\sqrt{p_j}|S_j\rangle|1\rangle\,.$$

Der entsprechende Quantenschaltkreis für $\mathcal{A}$ und den daraus konstruierten Operator $\mathcal{Q} = \mathcal{A}S_0\mathcal{A}^\dagger S_{\psi_0}$, welcher zur Amplitudenschätzung benötigt wird, ist in Fig. 4.3(a) skizziert.

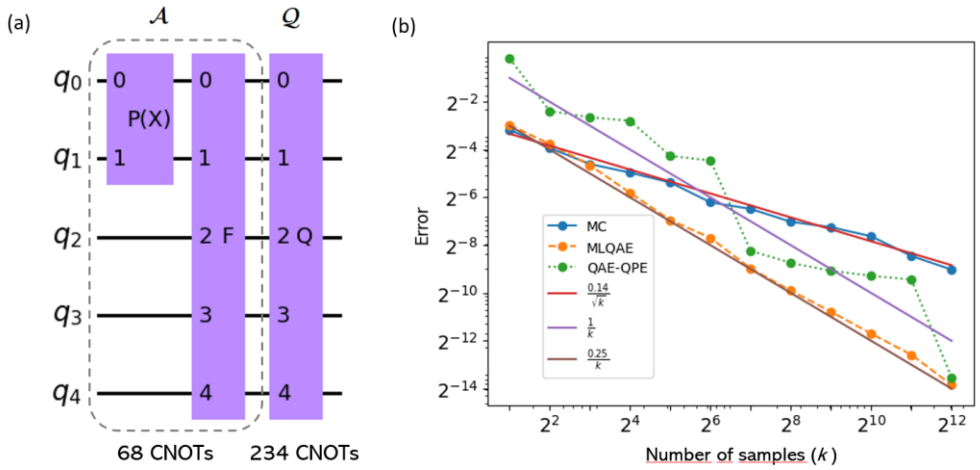


Abbildung 4.3: (a) Schema des Quantenschaltkreises für den Operator $\mathcal{A}$ (Präparation der Wahrscheinlichkeitsverteilung p sowie Anwendung der Payoff-Funktion f) gefolgt von einer Instanz des Operators $\mathcal{Q}$. Darunter ist jeweils die Anzahl der CNOT-Gatter aufgeführt, die zur Ausführung des Schaltkreises auf IBMQ Ehningen benötigt wird. (b) Skalierung des Fehlers der Quantenamplitudenschätzung (grün bzw. orange mit Quantenphasenschätzung bzw. Maximum Likelihood-Methode) verglichen mit dem klassischen Monte-Carlo-Algorithmus (blau).

Fig. 4.3(b) zeigt die Skalierung des Abschätzungsfehlers mit Anzahl k der Wiederholungen (d.h. Monte-Carlo-Samples im klassischen bzw. Anwendungen von $\mathcal{Q}$ im quantenmechanischen Fall). Während der Fehler des klassischen Monte-Carlo-Algorithmus wie $1/\sqrt{k}$ abfällt, weisen die Quantenalgorithmen die günstigere Skalierung $1/k$ auf.

Die Zusammensetzung des Quantenschaltkreises nach Transpilation auf IBMQ Ehningen (mit Basigattern CX, RZ, SX und X und unter Verwendung des höchsten optimization levels 3) ist in Tabelle 4.1 gezeigt. Die Tiefe des Schaltkreises für $\mathcal{A}$ beträgt 144, wobei 68 CNOT-Gatter verwendet werden. Eine Instanz von $\mathcal{Q}$ hat Tiefe 360 mit 234 CNOT-Gattern. 16 Wiederholungen von $\mathcal{Q}$, welche laut Abb. 4.3(b) nötig sind, um die bessere Fehlerskalierung im Vergleich zu Monte-Carlo zu etablieren, würden demnach Schaltkreise mit bis zu 4000 CNOT-Gattern benötigt, welche in Anbetracht der begrenzten Kohärenzzeiten und Fehlerraten die gegenwärtigen IBMQ-Hardware deutlich überfordern. Wir kommen zu dem Schluss, dass das als Quantenamplitudenschätzung formulierte Problem der Optionspreisbewertung für gegenwärtige NISQ-Geräte noch nicht geeignet ist.

Operator	Schaltkreistiefe	# CNOTs
$\mathcal{A}$	144	68
$\mathcal{Q}$	360	234

Tabelle 4.1: Kennzahlen der auf IBMQ Ehningen transpilierten 5-Qubit-Schaltkreise für European Call Options.

Variationelle imaginäre Zeitentwicklung (VarQITE)

Ein alternativer Zugang zur Optionsbewertung, welcher möglicherweise für NISQ-Geräte besser geeignet ist, besteht darin, das klassische Problem in ein quantenmechanisches Problem umzuformulieren, nämlich der Lösung der Schrödingergleichung mit imaginärer Zeit.

Wie in [Fontanela *et al.*, arXiv:1912.02753 (2021)] gezeigt, kann mit Hilfe des Feynman-Kac-Formalismus [Alghassi *et al.*, Quantum **6**, 730 (2022)] der Optionspreis zur Zeit $t \in [0, T]$ für eine European Call Option mit Payoff-Funktion wie in Abb. 4.2 als Diffusionsgleichung der Form

$$\partial_t \, u(\tau, x) = \frac{1}{2} \, \partial_{xx} \, u(\tau, x)$$

formuliert werden, wobei $u(\tau, x)$ mit dem Optionspreis über $\tau = \sigma^2 (T - t)$ und dem Preis x des zugrunde liegenden Wertpapiers verknüpft ist. Diese Diffusionsgleichung wird nun durch Wicks Rotation $\xi = -i \, \tau$ in eine lineare Schrödingergleichung übersetzt:

$$-i \, \frac{\partial}{\partial \xi} \, |\psi\rangle = \hat{H} \, |\psi\rangle$$

wobei die Wellenfunktion $|\psi\rangle$ die Rolle des modifizierten Optionspreises $u(\xi, x)$ spielt, während der Hamiltonoperator durch $\hat{H} = \frac{1}{2} \partial_{xx}$ gegeben ist. Die Lösung der obigen Schrödingergleichung mit imaginärer Zeit liefert so die gesuchte Lösung der Diffusionsgleichung. Zur Lösung der Schrödingergleichung nutzen wir den „variational quantum imaginary time evolution" (VarQITE)-Algorithmus [McArdle *et al.*, npj Quantum Inf. **5**, 75 (2019)]. In Abbildung 4.4(a) ist ein variationeller Schaltkreis mit 2 Qubits gezeigt, welcher den Hamiltonian $\hat{H}$ darstellt. Der mit dem VarQITE-Algorithmus auf einem qasm-Simulator berechnete Erwartungswert $\langle \psi | \hat{H} | \psi \rangle$ als Funktion der imaginären Zeit ξ stimmt sehr gut mit dem exakten Ergebnis überein, siehe Abb. 4.4(b), während die auf der Quantenhardware ibm_torino erhaltenen Ergebnisse, siehe Abb. 4.4(c), zwar dem allgemeinen Trend folgen, aber teilweise erhebliche Schwankungen aufweisen.

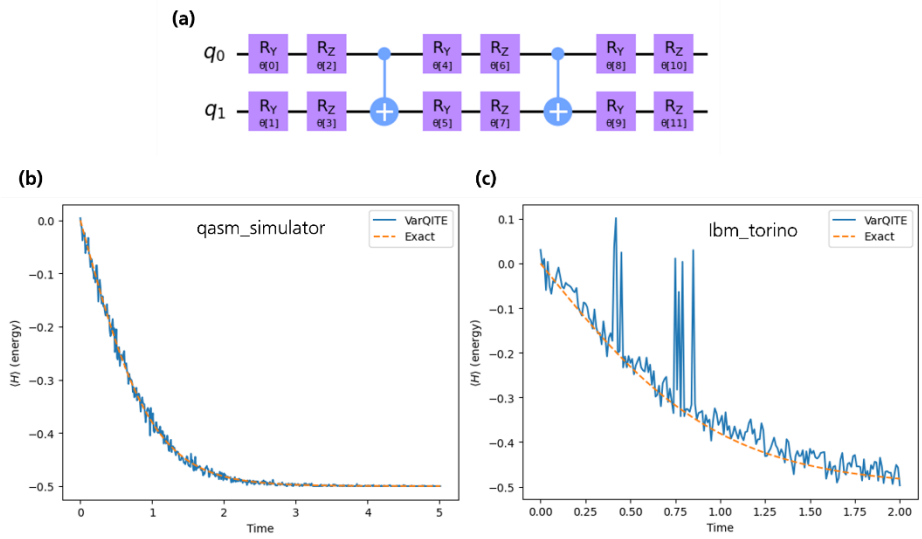
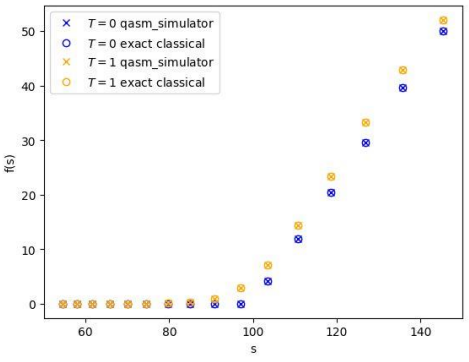


Abbildung 4.4: (a) Variationeller Schaltkreis mit zwei Qubits, der für den VarQITE-Algorithmus genutzt wird. (b) Vergleich der auf dem Quantensimulator und mit exakten klassischen Methoden erhaltenen Erwartungswerte als Funktion der imaginären Zeit. (c) Wie (b), aber mit Quantenhardware ibm_torino anstelle des Quantensimulators.

Abbildung 4.5 zeigt schließlich den durch einen variationellen Schaltkreis mit 4 Qubits auf dem qasm-Simulator erhaltenen Preis der European Call Option. Hierbei wählen wir die Volatilität $\sigma = 0.20$ und Strike Price $S_0 = K = 100$. Der Zustandsraum wird auf logarithmischer Skala auf dem Intervall $[S_{min}, S_{max}] = [50, 100]$ in $2^4 = 16$ Punkte diskretisiert.



Simulator erhaltene Preise der European Call Option zum Fälligkeitsdatum ($T = 1$) und am Anfang ($T = 0$).

Zusammenfassend haben wir den VarQITE-Algorithmus zur Preisbewertung von European Call Options auf Quanten-hardware mit 2 Qubits und nur auf dem Simulator mit 4 Qubits ausgeführt, wobei die erhaltenen Ergebnisse gut mit [Fontanela *et al.*] übereinstimmen.

VarQITE ist also ein vielversprechender NISQ-Algorithmus für Optionspreisbewertung, es jedoch muss noch geklärt werden, wie schnell dessen Komplexität und Konvergenzrate mit steigender Anzahl von Qubits anwächst [arXiv:2303.12839].

2.4.3 Unsupervised Learning und Clustering-Algorithmen (AP 4.3)

Im Rahmen des Arbeitspakets AP4.3 wurden als Ausgangspunkt geeignete Datensätze identifiziert und klassische Algorithmen für das Clustering (hier: K-Means) umgesetzt. Als nächstes wurde das binäre Clustering als QUBO formuliert und mit den in AP 4.1 beschriebenen Methoden (klassische Optimierungsverfahren sowie QAOA) gelöst. Die gefundenen Lösungen korrespondieren für kleine Problemgrößen mit den Erwartungen aus K-Means, welcher als „greedy-Algorithmus“ keine global optimale Lösung garantieren kann. Daraus ergibt sich die Fragestellung, ob für große Probleminstanzen der Optimierungsansatz ggf. vorteilhafter ist, sich ggf. auch ein Quantenvorteil ergibt und wie dieser abgeschätzt werden kann (siehe Kapitel 2.4.4).

Grundlage für die Experimente sind zwei übliche Benchmarkdatengeneratoren, „Blobs“ und „Moons“ (Scikit-Learn, kein Datum). Eine Beispielausgabe dieser Generatoren ist in Abbildung 4.6 zu sehen. Ziel ist die Trennung der dargestellten Datenpunkte in zwei Gruppen (Cluster).

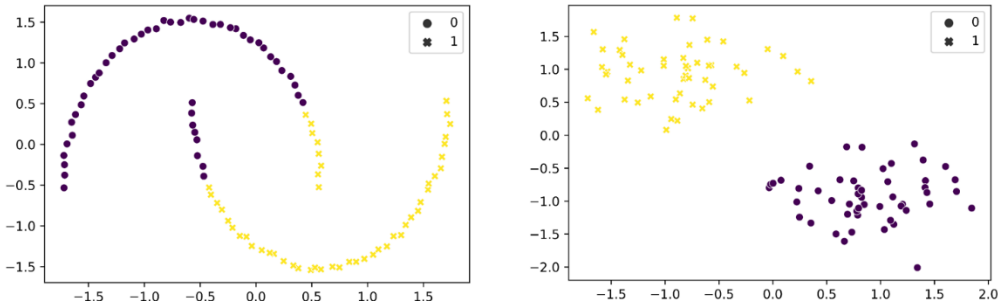


Abbildung 4.6: Beispieldatensätze, die mit den Generatoren „Moons“ (links) und „Blobs“ (rechts) erzeugt wurden. Die Cluster-Einteilung im dargestellten Beispiel erfolgt durch den k-Means Algorithmus.

Auf Basis dieser beiden Benchmarksdatensätze wurde zunächst die Laufzeit von klassischen Optimierungsalgorithmen untersucht.

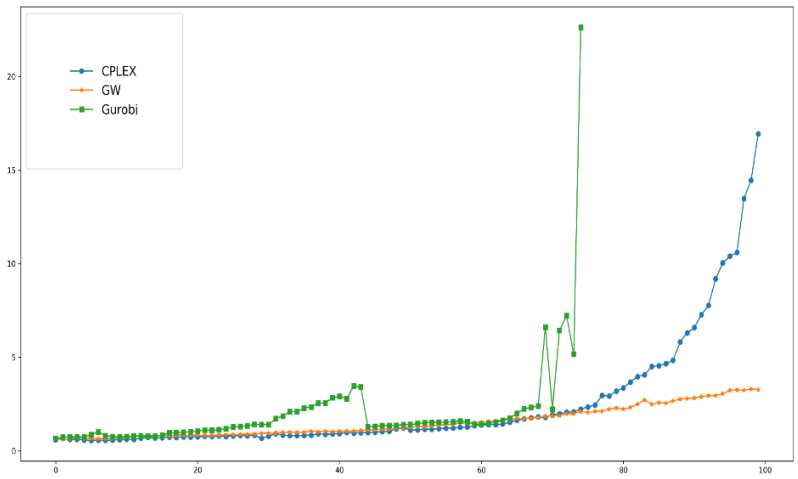


Abbildung 4.7: Laufzeiten unterschiedlicher klassischer Optimierungsalgorithmen (CPLEX, GW = Goemans-Williamson, Gurobi) für den Moons-Datensatz.

Wie in Abbildung 4.7 und 4.8 zu sehen ist, unterscheiden sich die Laufzeiten nicht nur erheblich, sondern sie zeigen auch ein anderes qualitatives Verhalten. Ein Problem bei der empirischen Abschätzung des Quantenvorteils besteht darin, dass die Angaben zu den Laufzeiten der Algorithmen oft unvollständig sind. Dies betrifft einerseits klassische Algorithmen, die eine Probleminstanz mit einer gewissen Wahrscheinlichkeit in einer vorgegebenen maximalen Zeit nicht lösen. Zum anderen kann bei Experimenten auf Quantenhardware nicht nur das oben genannte Problem auftreten, sondern ein Optimierungsprozess auch aus technischen Gründen vorzeitig abgebrochen werden. In beiden Fällen entstehen Daten, bei denen einzelne Läufe „zu kurz" beobachtet werden und ein exakter Zeitpunkt für die Lösung einer Probleminstanz nicht angegeben werden kann. Werden diese Läufe einfach weggelassen, reduziert sich nicht nur die Anzahl der Beobachtungen, sondern es kommt auch zu einer deutlichen Unterschätzung der benötigten Laufzeit.

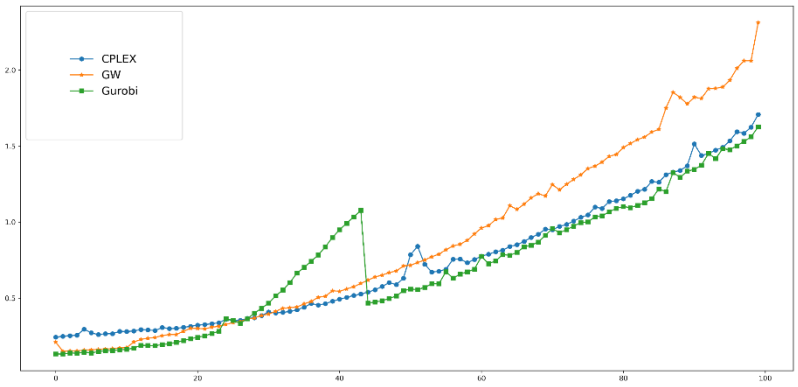


Abbildung 4.8: Laufzeiten unterschiedlicher klassischer Optimierungsalgorithmen (derselben wie in Abbildung 4.7) für den Blobs-Datensatz.

Ein Standardverfahren in der klassischen Optimierung ist die Annahme, dass abgebrochene Läufe neu gestartet werden (simulierte Neustarts). Unter dieser Annahme können Laufzeiten abgeschätzt werden, wenn mindestens ein Optimierungslauf „erfolgreich" war (die Instanz gelöst hat). Weniger verbreitet ist der Ansatz, die

aufgezeichneten Laufzeitdaten als „rechtszensierte“ Daten zu betrachten. Solche Daten entstehen typischerweise in der Medizin und werden mit Methoden der „survival analysis“ modelliert. Wir haben untersucht, inwieweit diese Methoden geeignet sind, die Laufzeiten klassischer und quantenbasierter Algorithmen zu analysieren und damit empirisch fundierte Abschätzungen eines Quantenvorteils zu erhalten. Abbildung 4.9 zeigt beispielhaft, dass eine Modellierung mit survival splines erlaubt, einer Unterschätzung von Laufzeiten durch rechts-zensierte Daten entgegen zu wirken. Zudem konnte festgestellt werden, dass die oben erwähnten simulierten Neustarts in fast allen Fällen zu einer deutlichen Überschätzung führen.

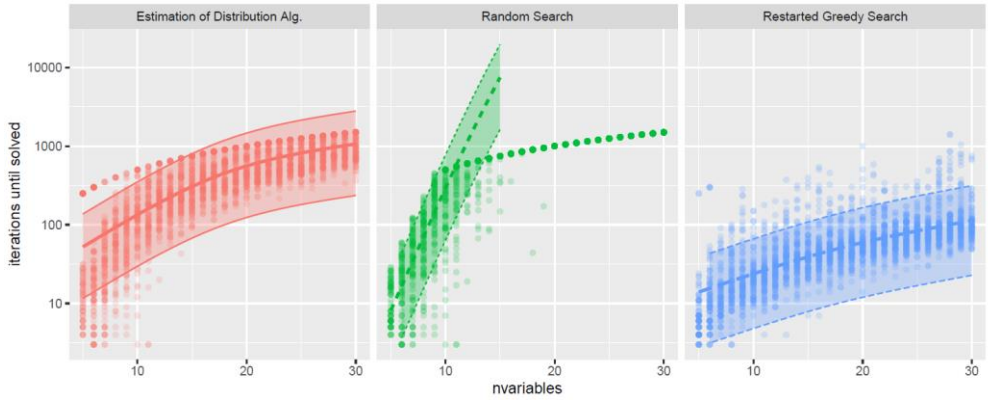


Abbildung 4.9: Modellierung von Laufzeitdaten mit Survival Spline Modellen für drei Algorithmen. Farbige Punkte sind aufgenommene Daten (rechts-zensiert durch maximale Laufzeit). Linien mit farbigem Band zeigen den Median sowie 5% und 95% Quantil der Modelle.

Für den Abgleich mit QAOA wurden zudem Benchmarks auf allen drei Anwendungsfällen (Portfolio Optimization, Feature Selection und Clustering) durchgeführt, um zumindest auf niedrig-dimensionalen Problemen einen direkten Vergleich auch mit vollständigen (unzensierten) Daten vornehmen zu können. Geprüft wurden dabei CPLEX (in verschiedenen Konfigurationen), Greedy Search mit Restarts (GRS), ein einfacher Estimation of Distribution Algorithmus (EDA), Goemans-Williamson (GW), sowie MQLib als Hyper-Heuristic (MQH) und die einzelne Heuristik „BURER2002“ aus MQLib. Hierbei ist insbesondere festzustellen, dass GRS und GW viele Instanzen nicht lösen (Ausnahme: Clustering Instanzen), während der EDA nur einzelne nicht löst. MQH löst alle Instanzen, weist aber stark variierende, teils übermäßig hohe Laufzeiten auf. Am schnellsten und konsistentesten löst BURER2002 diese Probleminstanzen, gefolgt von CPLEX. Ein Auszug dieser Ergebnisse für Clustering Instanzen mit 12 bis 28 Bits ist in Abbildung 4.10 zu finden.

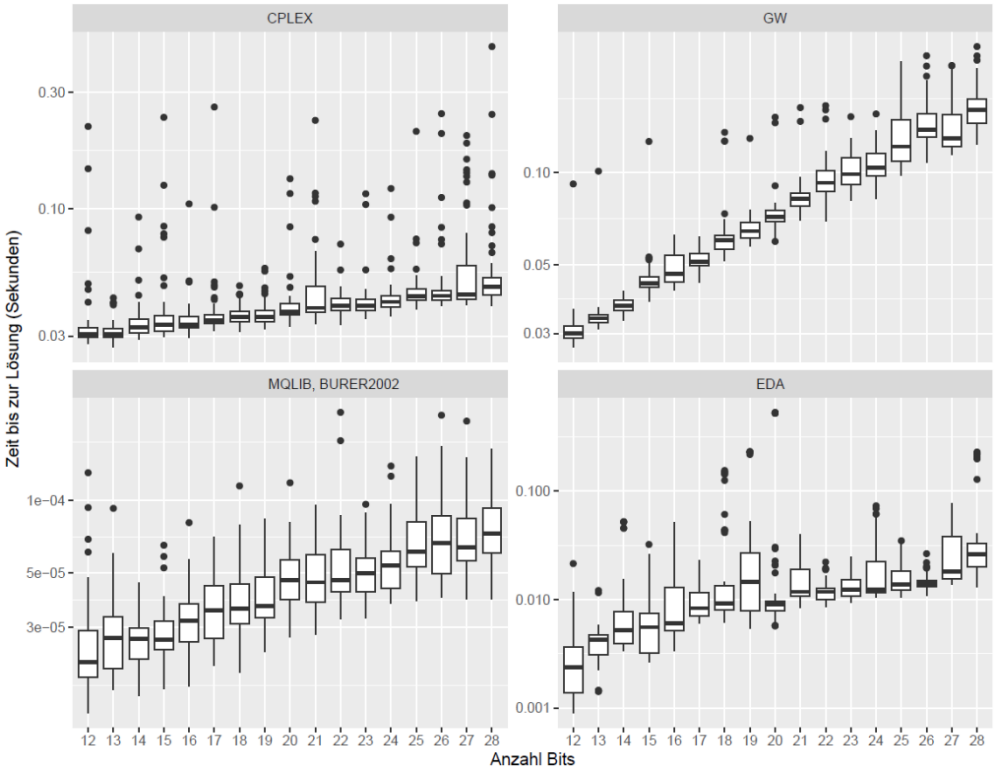


Abbildung 4.10: Benötigte Zeit bis zur optimalen Lösung von Clustering-Instanzen mit 121 bis 28 Bits für verschiedene klassische Optimierer.

2.4.4 Abschätzung des Quantenvorteils (AP 4.4)

Aufgrund der oben beschriebenen Ergebnisse verwenden wir (wie auch schon im Vorgängerprojekt QORA) die Heuristik BURER2002 aus MQlib, um die Skalierung der für das Auffinden der optimalen Lösung benötigten klassischen Rechenzeit mit steigender Problemgröße zu ermitteln. Auf der Quantenseite verwenden wir die in Kapitel 2.3.2 beschriebene Idee der QAOA-Schaltkreise großer Tiefe mit wenigen Parametern und ermitteln die gesamte Tiefe aller Quantenschaltkreise, welche auszuführen sind, bis die optimale Lösung gefunden wird.[4]

Als Ausgangspunkt wiederholen wir kurz das in QORA erhaltene Ergebnis für Portfoliooptimierung, siehe Abbildung 4.11. Hier wurde kein Hinweis auf Quantenvorteil gefunden, da die klassische Lösungsdauer mit steigender Problemgröße weniger schnell ansteigt als im Fall der damals verwendeten variationellen QAOA-Variante.

[4] Genauer gesagt ermitteln wir den Median dieser Größe, also die gesamte Tiefe aller Schaltkreise, die auszuführen sind, bis die optimale Lösung mit Wahrscheinlichkeit ½ gefunden worden ist. Für einen einzelnen Schaltkreis der Tiefe d, welcher mit Wahrscheinlichkeit P die optimale Lösung liefert, beträgt diese Größe $D = d\frac{\log(1/2)}{\log(1-P)}$.

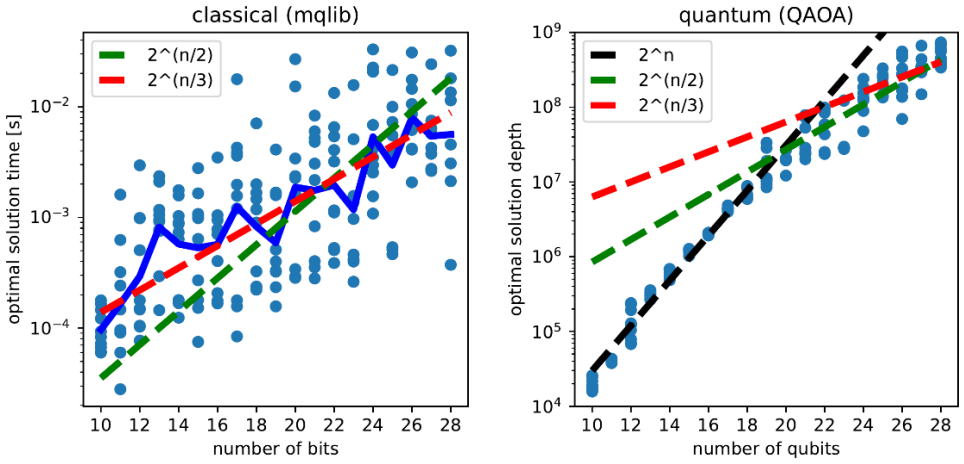


Abbildung 4.11: Vergleich eines klassischen Verfahrens zur Lösung von QUBO-Problemen mit variationellem QAOA (aus dem Abschlussbericht von QORA). Links: vom klassischen Algorithmus benötigte Zeit zum Auffinden der optimalen Lösung als Funktion der Problemgröße n (Anzahl binärer Variablen) für jeweils 10 Instanzen des Portfoliooptimierungsproblems (blaue Punkte, mit blauer Linie als geometrischem Mittel). Die benötigte Zeit skaliert wie $2^{n/3}$ (rote gestrichelte Linie). Rechts: Aufaddierte Gesamttiefe aller Quantenschaltkreise, welche auf dem fehlerfreien Simulator im Verlauf der QAOA-Optimierung (mit Standard-Mixer und 1000 Shots pro Schaltkreis) insgesamt ausgeführt werden, bis die optimale Lösung zum ersten Mal gemessen wird, als Funktion der Problemgröße n (Anzahl der Qubits). Für kleine Probleme beobachten wir eine Skalierung wie 2^n, entsprechend einer „Brute Force"-Suche. Ab ca. $n = 20$ führt die Quantenoptimierung zu einer gemäß $2^{n/2}$ (grüne gestrichelte Linie) beschleunigten Lösung. Die noch bessere Skalierung des klassischen Algorithmus (rote gestrichelte Linie) wird jedoch nicht erreicht. Insgesamt liefert die hier verwendete QAOA-Variante also keinen Hinweis auf einen Quantenvorteil.

Mit der in Kapitel 2.3.2 beschriebenen Idee konnten wir in QORA II jedoch das QAOA-Ergebnis erheblich verbessern. In Abbildung 4.12 sind die Ergebnisse für die drei Anwendungsfälle Portfoliooptimierung, Feature Selection und Clustering gezeigt. Tatsächlich erhalten wir nun im Fall der Portfoliooptimierung einen Hinweis auf Quantenvorteil, da die Rechendauer (bzw. Schaltkreistiefe) hier langsamer ansteigt als für den klassischen Algorithmus. Für Feature Selection und Clustering ist hingegen der klassische Algorithmus schneller.

Zu beachten ist jedoch, dass diese in der Endphase des Projekts entstandenen Ergebnisse auf fehlerfreien Quantensimulationen beruhen und noch nicht auf echter Quantenhardware getestet worden sind, um ihre Robustheit gegenüber Fehlern zu überprüfen. Außerdem ist der Aufwand zur Ermittlung der Schaltkreisparameter m_1, m_2 und p (siehe Kap. 2.3.2), welche die Auswertung von Quantenschaltkreisen mit kleinerer Zahl von Qubits (bis höchstens $n = 11$) erfordern, noch nicht auf der Quantenseite enthalten. Trotzdem halten wir diese Ergebnisse mit Hinblick auf die nächste Phase des Kompetenzzentrums Quantencomputing Baden-Württemberg für vielversprechend. Ähnliche Ergebnisse wurden vor kurzem auch von J. A. Montañez-Barrera und K. Michielsen aus Jülich bzw. Aachen veröffentlicht.[5]

[5] J. A. Montañez-Barrera und K. Michielsen, *Towards a universal QAOA protocol: Evidence of a scaling advantage in solving some combinatorial optimization problems*, arXiv:2405.09169v2 (Juni 2024).

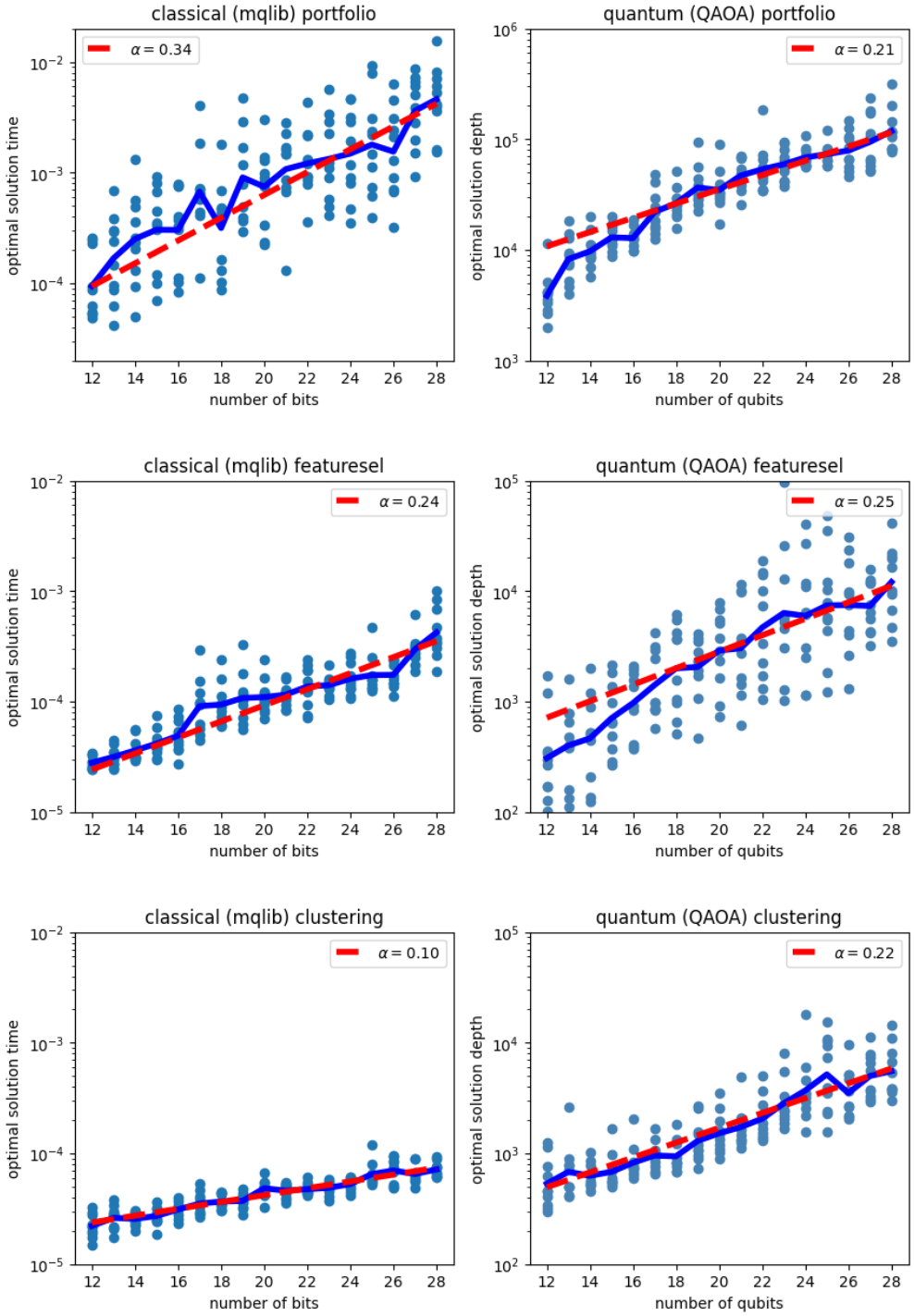


Abbildung 4.12: Skalierung der klassischen Laufzeit (links) zur Lösung von QUBO-Problemen im Vergleich zu unserer in QORA II entwickelten „linear ramp" - QAOA-Variante (rechts), siehe Kap. 2.3.2, für drei Anwendungsfälle: Portfoliooptimierung (oben), Feature Selection (Mitte) und Clustering (unten). Siehe auch Abbildung 4.11 für weitere Einzelheiten. Die rote gestrichelte Linie zeigt jeweils einen exponentiellen Fit mit Steigung α (d. h. $\propto 2^{\alpha n}$) an den aus jeweils 10 Instanzen ermittelten geometrischen Mittelwert (blau). Im Fall der Portfoliooptimierung (oben) sehen wir für QAOA einen langsameren Anstieg als für den klassischen Algorithmus und somit einen Hinweis auf Quantenvorteil. Die Feature Selection (Mitte) und vor allem das Clustering (unten) sind hingegen klassisch schneller zu lösen, so dass sich hier kein Quantenvorteil ergibt.

3 Anhang

3.1 Publikationen

[1] V. Dehn and T. Wellens, *A hybrid quantum-classical approach to warm-starting optimization*, Proc. SPIE 12911, Quantum Computing, Communication, and Simulation IV, 129110N (13 March 2024).
[2] D. Bhatti and S. Barz, *Generating Greenberger-Horne-Zeilinger states using multiport splitters*, Phys. Rev. A **107**, 033714 (2023).
[3] S. Kumar, D. Bhatti, A. E. Jones, and S. Barz, *Experimental entanglement generation using multiport beam splitters*, New J. Phys. **25**, 063027 (2023).
[4] J. Mackeprang, D. Bhatti, and S. Barz, *Non-adaptive measurement-based quantum computation on IBM Q*, Sci. Rep. **13**, 15428 (2023).
[5] J. Kattemölle and S. Hariharan, *Line-graph qubit routing: from kagome to heavy-hex and more*, arXiv:2306.05385 (2023)
[6] M. Hüls, *Using a Quantum Computer's Errors for Good*, Master Thesis, EKUT (June 2023).
[7] G. Hellstern, V. Dehn, M. Zaefferer, *Quantum computer based feature selection in machine learning*, IET Quant. Comm. 1-24 (2024).
[8] A. Naik, E. Yeniaras, G. Hellstern, G. Prasad, S. Vishwakarma, *From Portfolio Optimization to Quantum Blockchain and Security: A Systematic Review of Quantum Computing in Finance*, arXiv:2307.01155 (2023).
[9] J. Kattemölle, *Edge coloring lattice graphs*, arXiv:2402.08752 (2023)
[10] Y. Ji, X. Chen, I. Polian, and Y. Ban. *Algorithm-oriented qubit mapping for variational quantum algorithms.* arXiv:2310.09826 (2023).
[11] Y. Ji, K. F. Koenig, and I. Polian. *Improving the performance of digitized counterdiabatic quantum optimization via algorithm-oriented qubit mapping.* arXiv:2311.14624 (2023).
[12] E. Weiss, M. Cech, S. Soltan, M. A. Koppenhöfer, M. Krebsbach, T. Wellens and D. Braun, *Pattern-based quantum functional testing*, arXiv:2405.20828 (2024).
[13] Y. Ji and I. Polian, *Synergistic Dynamical Decoupling and Circuit Design for Enhanced Algorithm Performance on Near-Term Quantum Devices*, arXiv:2405.17230 (2024).

Bereits im Abschlussbericht von QORA (März 2023) aufgeführte Preprints, die inzwischen sämtlich in Fachzeitschriften mit Begutachtung veröffentlicht wurden:

[13] A. Ketterer and T. Wellens, *Characterizing Crosstalk of Superconducting Transmon Processors*, Phys. Rev. Applied **20**, 034065 (2023).
[15] Y. Ji, K. F. Koenig, and I. Polian, *Optimizing quantum algorithms on bipotent architectures*, Phys. Rev. A **108**, 022610 (2023).
[16] J. Mackeprang, D. Bhatti, M. Hoban, and S. Barz, *The power of qutrits for non-adaptive measurement-based quantum computing*, New J. Phys. **25**, 073007 (2023).
[17] J. Kattemölle and G. Burkard*, Ability of error correlations to improve the performance of variational quantum algorithms*, Phys. Rev. A **107**, 042426 (2023), Editor's Suggestion.
[18] N. Milazzo, O. Giraud, G. Gramegna, and D. Braun, *Principles of quantum functional testing*, Phys. Rev. A **108**, 022602 (2023).

Der Vollständigkeit halber seien schließlich auch die restlichen Veröffentlichungen aus dem Vorgängerprojekt QORA genannt:

[19] S. Brandhofer, S. Devitt, T. Wellens, and I. Polian: *Special Session: Noisy Intermediate-Scale Quantum (NISQ) Computers – How They Work, How They Fail, How to Test Them?* Proc. 39th IEEE VLSI Test Symposium (VTS'21), pp. 1-6 (2021).

[20] Y. Ji, S. Brandhofer, and I. Polian, *Calibration-Aware Transpilation for Variational Quantum Optimization*, 2022 IEEE Int. Conf. on Quantum Computing and Engineering (QCE), pp. 204-214 (2022).
[21] S. Brandhofer, D. Braun, G. Hellstern, M. Hüls, Y. Ji, I. Polian, A. Bhatia and T. Wellens, *Benchmarking the performance of portfolio optimization with QAOA*, Quantum Inf. Process. **22**, 25 (2023).
[22] Y. Ji and I. Polian, *Algorithm-Aware Qubit Mapping*, TechRxiv.21786146 (2022).

3.2 Verbreitung der Projektergebnisse (Vorträge/Poster)

- J. Kattemölle, *Ability of error correlations to improve the performance of variational quantum algorithms*, Posterpräsentation, 778. WE-Heraeus-Seminar *Coping with Errors in Scalable Quantum Computing Systems*, Physikzentrum Bad Honnef, 8.-11. Januar 2023
- F. Knäble, Vortrag im DigitalDialog Quantum Computing for Finance, Fraunhofer IAO, 1. Februar 2023
- G. Burkard, *High-fidelity quantum gates for spin qubits and spin-photon interfaces,* eingeladener Vortrag, IQST Day, Universität Ulm, 16. Februar 2023
- V. Dehn, *Portfolio Optimization with QAOA*, Vortrag auf der DPG-Frühjahrstagung in Hannover, 5.-10. März 2023
- A. Ketterer, *Characterizing crosstalk of superconducting transmon processors,* Vortrag auf der DPG-Frühjahrstagung in Hannover, 5.-10. März 2023
- G. Gramegna, *Quantum Functional Testing*, Vortrag auf der DPG-Frühjahrstagung in Hannover, 5.-10. März 2023
- M. Hüls, *Portfolio Optimization using a Quantum Computer*, Posterpräsentation, DPG-Frühjahrstagung in Hannover, 5.-10. März 2023
- T. Wellens, *Characterizing crosstalk of superconducting transmon processors*, Vortrag beim IBM Quantum Partner Forum, Paris, Mai 2023
- V. Dehn, *Performance of QAOA for Portfolio Optimization*, Posterpräsentation auf der Quantum Matter International Conference – QUANTUMatter 2023, Madrid, Mai 2023
- J. Kattemölle, *Ability of error correlations to improve the performance of variational quantum algorithms*, eingeladener Vortrag, Center for Quantum Technologies, Singapur, 16. Mai 2023
- J. Kattemölle, *Line-graph qubit routing: from kagome to heavy-hex and more*, Vortrag im Seminar *Quantum Algorithms for Many Body Systems*, Amsterdam & Konstanz (online), 26. Mai 2023
- G. Hellstern, *Finance meets Quantum*, Vortrag auf der Messe „World of Quantum, 28. Juni 2023, München
- G. Burkard, *Semiconductor spin qubits and how to connect them to photons*, eingeladener Vortrag, IQST mini workshop, ZAQuant, Universität Stuttgart, 4. Juli 2023
- V. Katakuri and V. Mohan, Vortrag im QuantumBW-Fokustreffen: *Quantensoftware-Engineering*, 13./14. Juli 2023
- D. Braun, *Principles of quantum functional testing*, invited key-note presentation, IQIS conference, Trieste, September 2023
- G. Hellstern, Quantum meets Finance, Vortrag auf dem QBN Summit, 10.10.2023, Stuttgart
- T. Wellens and V. Katakuri*, QORA II: Quantum optimization with resilient algorithms,* Vortrag auf der Messe Quantum Effects, Stuttgart, 11. Oktober 2023
- G. Gramegna, *Principles of Quantum Functional Testing*, Posterpräsentation, IQFA 13th Colloquium, Palaiseau, France, 16.-18. November 2022

- G. Hellstern, *Quantencomputing*, Vortrag auf den IT-Tagen, 13.12.20203, Frankfurt
- J. Kattemölle, *Optimal-depth Quantum Circuits by the Edge Coloring of Lattice Graphs*, Vortrag, KQCBW Developer Conference, IAF Freiburg, 7. März 2024
- J. Kattemölle, *Generalized brickwork quantum circuits*, eingeladener Vortrag, Exactly Solved Models and Quantum Computing, Leiden, 21. März 2024
- J. Kattemölle, *Efficient, deterministic, and explicit quantum circuits for the (approximate) preparation of Bethe ansatz states*, Posterpräsentation, Exactly Solved Models and Quantum Computing, Leiden, 21. März 2024
- J. Kattemölle, *Optimal quantum circuits for the quantum simulation of quantum matter*, Vortrag, QUANTUMatter, San Sebastián, 8. Mai 2024
- J. Kattemölle, *Optimal qubit routing of tilable circuits*, Vortrag im Seminar Quantum Algorithms for Many Body Systems, Amsterdam & Konstanz (online), 14. Mai 2024