
Quantum Reinforcement Learning für industrielle Anwendungen

Abschlussbericht



QLINDA
QUANTUM REINFORCEMENT
LEARNING FÜR INDUSTRIELLE
ANWENDUNGEN

31. März 2025

Vorhabensbezeichnung	Quantum Reinforcement Learning für industrielle Anwendungen (QLINDA)
Teilvorhaben	Bewertung und Entwicklung von QRL-Algorithmen
FKZ	13N15645
Zuwendungsempfänger	Fraunhofer Institut für Integrierte Schaltungen IIS Nordostpark, 84, 90411 Nürnberg
Kontakt	Dr. Daniel Scherer, daniel.scherer2@iis.fraunhofer.de
Laufzeit des Vorhabens	01.04.2021 bis 30.09.2024 (verlängert um 6 Monate)
Berichtszeitraum	01.01.2021 bis 30.09.2024

Inhaltsverzeichnis

1	Motivation und Ziele des Vorhabens	1
1.1	<i>Motivation des Vorhabens</i>	1
1.2	<i>Geplante Ziele und erzielte Ergebnisse.....</i>	2
2	Konsortium.....	2
3	Einzelergebnisse der Arbeitspakete	3
3.1	<i>AP1: Bewertung Stand der Forschung</i>	3
3.2	<i>AP2: Auswahl geeigneter Anwendungsklassen</i>	4
3.3	<i>AP3: Simulatorschnittstelle für industrielle Anwendungen</i>	5
3.4	<i>AP4: Auswahl, Anpassung und Entwicklung der Quantenalgorithmen.....</i>	5
3.5	<i>AP5: Einsatz auf Quantencomputer-Simulator und realem Quantencomputer</i>	7
3.6	<i>AP6: Praxisnahe Anwendungen.....</i>	7
3.7	<i>AP7: Bibliothek zur Nutzbarmachung für industrielle Anwendungen.....</i>	7
3.8	<i>AP8: Dissemination</i>	7
4	Beschreibung der Ergebnisse und Veröffentlichungen	9
4.1	<i>Klassifikationsmatrix.....</i>	9
4.2	<i>Aufdeckung von Instabilitäten in QRL.....</i>	10
4.3	<i>Inkrementelles Daten-Reloading.....</i>	12
4.4	<i>Batch-Constrained Quantum Q-Learning (BCQQ).....</i>	14
4.5	<i>SPSA-Optimierer.....</i>	15
4.6	<i>Guided-SPSA</i>	16
5	QRL-Demonstrator	17
5.1	<i>Anki Overdrive.....</i>	17
6	Externer Fortschritt.....	19
7	Relevanz und Nutzen der erzielten Ergebnisse	19
8	Zahlenmäßigen Nachweises	20

1 Motivation und Ziele des Vorhabens

1.1 Motivation des Vorhabens

Aus dem intellektuellen Umbruch der Physik des späten 19. Jahrhunderts hervorgegangen, entwickelte sich die Quantenmechanik als wissenschaftliche Theorie von unvergleichlicher Präzision und veränderte grundlegend unser Verständnis der Natur, während sie aufeinanderfolgende technologische Transformationen ermöglichte. Die erste Quantenrevolution bescherte der Menschheit grundlegende Technologien wie Laser und Halbleiter, die einen Großteil der modernen Technologie stützen. Nun, inmitten einer zweiten Quantenrevolution, erleben wir das aufstrebende Potenzial von Quantensensoren, Quantenkryptographie und Quantenkommunikation – Innovationen, die auf der gezielten und sorgfältigen Kontrolle quantenmechanischer Zustände für völlig neue technologische Anwendungen basieren. In diesem aufregenden Umfeld sticht das Quantencomputing als besonders eindrucksvolle Technologie hervor, das die Nutzung quantenmechanischer Zustände zur Revolutionierung der Informationsverarbeitung selbst erkundet. Dieser Fortschritt erfolgt parallel zu und überschneidet sich mit den bemerkenswerten Fortschritten in der Künstlichen Intelligenz und im maschinellen Lernen (ML), insbesondere im Bereich des Deep Learning (DL). Angetrieben von exponentiellen Steigerungen der Rechenleistung von Graphical Processing Units (GPU), der Integration von Central Processing Units (CPU), der Speicherkapazität und dem schier unendlichen Volumen verfügbarer Daten sind DL-Methoden zum Stand der Technik in zahlreichen Bereichen geworden. Ihre Fähigkeit, komplexe, nichtlineare Zusammenhänge innerhalb massiver Datensätze zu erkennen, ermöglicht es DL-Techniken, Probleme anzugehen, die einst nur durch komplexe, regelbasierte Systeme gelöst werden konnten, die auf umfangreichem Expertenwissen beruhten. Interessanterweise und entscheidend für die zukünftige Entwicklung wurde empirisch nachgewiesen, dass Quantencomputer bei bestimmten sorgfältig ausgewählten Berechnungsproblemen bereits selbst State-of-the-Art klassische Systeme wie CPUs und GPUs übertreffen können [1]. Dieser beobachtete Quantenvorteil bietet einen starken Anreiz für Projekte, die darauf abzielen, das aufstrebende Feld des Quantencomputings mit den bereits vorhandenen Fähigkeiten modernster ML- und DL-Ansätze synergistisch zu kombinieren. Die übergeordnete Vision besteht darin, hochinnovative und völlig neuartige Supertechnologien mit transformativem Potenzial über ein breites Spektrum industrieller Anwendungen hinweg zu entwickeln. Eine zentrale Motivation für Forschung im Bereich Quantencomputing ist die Allgegenwärtigkeit von Aufgaben aus der linearen Algebra, z.B. Matrixinversionen und das Suchen der Lösung linearer Gleichungssysteme, neben komplexen Optimierungsproblemen, die für DL genutzt werden. Quantencomputing bietet von Natur aus enormes Potenzial zur Beschleunigung und erheblichen Verbesserung der Effizienz dieser rechenintensiven maschinellen Lernprozesse. Ein kritisches Problem ergibt sich jedoch aus der grundlegend unterschiedlichen Natur der Informationsverarbeitung auf einem Quantencomputer im Vergleich zu konventionellen klassischen Computern. Dies erfordert einen komplexen Übersetzungsprozess, der die Übertragung etablierter klassischer ML- und DL-Algorithmen in ihre entsprechenden quantenmechanischen Gegenstücke verlangt. In einigen Fällen kann es sogar erforderlich sein, völlig neue algorithmische Ansätze zu entwickeln, die speziell darauf ausgelegt sind, das einzigartige Berechnungsparadigma von Quantencomputern auszunutzen. Diese Anpassungsnotwendigkeit erstreckt sich über verschiedene ML-Paradigmen, einschließlich überwachtem Lernen, unüberwachtem Lernen und Reinforcement Learning. Tatsächlich untersucht ein erheblicher Teil der theoretischen Arbeiten aktiv den komplexen Prozess der Übertragung von Lernverfahren und den damit verbundenen Algorithmen auf Quantencomputer [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14].

Auf der Grundlage dieser umfangreichen bestehenden wissenschaftlichen und theoretischen Beiträge ist daher das Hauptziel dedizierter Projekte, das Reinforcement Learning auf Quantencomputern effektiv zu ermöglichen – und damit das Feld des Quantum Reinforcement Learning (QRL) zu etablieren. Das ultimative Ziel besteht darin, diese mächtige Synergie zu nutzen und das QRL auf einen Reifegrad zu entwickeln, auf dem es angewendet werden kann, um eine breite und wirkungsvolle Palette von realen Problemen mit direkter Relevanz und signifikantem Wert in verschiedenen Industriesektoren zu lösen.

1.2 Geplante Ziele und erzielte Ergebnisse

Angetrieben von Fortschritten im Quantencomputing untersucht dieses Forschungsprojekt die Anwendung von QRL in industriellen Umgebungen. Das Kernziel besteht darin, spezifische industrielle Anwendungsbereiche zu identifizieren, in denen QRL signifikante Vorteile bietet, die Wirksamkeit bestehender QRL-Algorithmen für diese Bereiche zu bewerten und die Entwicklung verbesserter Algorithmen zu fördern. Die Methodik beginnt mit einer rigorosen Bewertung der aktuellen QRL-Algorithmen, um ihre inhärente Eignung für industrielle Herausforderungen zu beurteilen. Vielversprechende Algorithmen werden dann in realistischen, simulierten industriellen Umgebungen getestet, wobei ein maßgeschneidertes „Industrial Benchmark“ (IB)-Rahmenwerk [15] verwendet wird, um Relevanz und Strenge sicherzustellen. Ein wichtiger Aspekt des Projekts ist die Untersuchung der Übertragung dieser QRL-Lösungen von simulierten Umgebungen auf tatsächliche Quantencomputing-Plattformen – sowohl klassisch simuliert als auch auf echter Quantenhardware. In Anbetracht der praktischen Einschränkungen befasst sich die Forschung mit dem kritischen Thema der Datenkompatibilität und konzentriert sich auf die Entwicklung von Schnittstellentechnologien und Datenumwandlungstechniken, wie z.B. Vektorisierung, um die Kluft zwischen kontinuierlichen industriellen Daten und den diskreten Datenformaten zu überbrücken, die oft von quantenkompatiblen QRL-Algorithmen benötigt werden. Das Projekt untersucht und passt speziell vielversprechende QRL-Ansätze an, für die effiziente Implementierungen auf Quantencomputern durch variationelle Quantenschaltkreise (VQCs) [16], [17] bekannt sind. Ein anderer vielversprechender Algorithmus ist der Quantum Approximate Optimization Algorithm (QAOA) [18], der insbesondere für kombinatorische Optimierungsprobleme von Bedeutung ist, die in industriellen Anwendungen mit relationalen Daten häufig vorkommen. Letztendlich zielt das Projekt darauf ab, QRL-Methoden für die praktische Lösung industrieller Probleme zu validieren und zu verfeinern, indem es von ersten Simulationen hin zu einer zukünftigen Implementierung in der realen Welt auf Quantencomputern übergeht, wobei der Fokus darauf liegt, relevante industrielle Herausforderungen anzugehen und die Leistung zu optimieren. Eine detaillierte Darstellung der durchgeführten Arbeiten und der erzielten Ergebnisse im Rahmen dieses Projekts findet sich in Abschnitt 3 und Abschnitt 4 dieses Berichts. Die Ergebnisse dieses Projekts stellen einen bedeutenden Fortschritt dar, um das Potenzial von QRL und QML für praktische industrielle Anwendungen innerhalb des aktuellen NISQ-Paradigmas zu realisieren. Es ist jedoch wichtig zu beachten, dass das Erreichen eines definitiven Quantenvorteils durch Quantenalgorithmen, die variationelle Quantenschaltkreise nutzen, eine offene Frage bleibt und im Rahmen dieses Projekts nicht abschließend nachgewiesen werden konnte. Der Hauptgrund für dieses inkonklusive Ergebnis ist die derzeitige Unfähigkeit, die Größe der Algorithmen zu skalieren, um größere, komplexere Probleme anzugehen, aufgrund des überwältigenden Einflusses von Rauschen auf die verfügbare Quantenhardware. Darüber hinaus macht der erhebliche Aufwand, der durch Fehlerkorrektur- und Fehlerminderungs-techniken eingeführt wird, einfache iterative Lernansätze in der gegenwärtigen Landschaft des Quantencomputings unpraktisch.

2 Konsortium

Das Projektkonsortium besteht aus vier Partnern, die jeweils unterschiedliche und komplementäre Fachkenntnisse in das Vorhaben einbringen, QRL für industrielle Anwendungen auf NISQ-Hardware zu realisieren. Die Siemens AG (SAG), als industrieller Partner, liefert wertvolle Einblicke in reale industrielle Anwendungsfälle und bringt ein tiefes Verständnis für branchenspezifische Herausforderungen und Anforderungen mit, um die Relevanz und praktische Anwendbarkeit der Projektergebnisse sicherzustellen. Das Fraunhofer IIS (IIS) fungiert als Partner für angewandte Forschung und konzentriert sich auf die Kernaufgabe der Algorithmusentwicklung und Innovation im Bereich QRL, indem theoretische Konzepte in praktische Lösungen übersetzt werden. Die OTH Regensburg (OTH), der akademische Partner im Konsortium, trägt entscheidende Fachkenntnisse im Benchmarking und in der rigorosen Validierung der entwickelten Algorithmen bei, um sicherzustellen, dass ihre Leistung und Zuverlässigkeit gründlich bewertet werden. Schließlich bringt IQM Germany (IQM), der Hersteller von Quantenhardware, kritisches Wissen und Verständnis für die Eigenschaften und Einschränkungen aktueller Quantenhardware in das Projekt ein, um die Design- und Implementierungsanstrengungen der Algorithmen in Richtung praktikabler und effektiver Lösungen auf verfügbaren Plattformen zu lenken. Durch diese synergistische Zusammenarbeit arbeiten die Konsortialpartner auf das gemeinsame Ziel hin, QRL für praktische industrielle Anwendungen innerhalb der Grenzen verfügbarer

Quantenhardware voranzutreiben.

3 Einzelergebnisse der Arbeitspakete

3.1 AP1: Bewertung Stand der Forschung

Arbeitspaket 1 hatte das Ziel, eine robuste Grundlage für das Projekt zu schaffen und als Vorbereitung für die nachfolgenden Arbeitspakete zu fungieren. Hier führten wir eine umfassende Bewertung des aktuellen Stands der Technik im Bereich des QRLs durch. Das spezifische Ziel war es, die verfügbaren Algorithmen hinsichtlich ihrer Kompatibilität mit Noisy Intermediate-Scale Quantum (NISQ)-Geräten zu bewerten und zu bestimmen, ob sie sinnvoll auf der bestehenden Hardware ausgeführt werden können. Wesentliche Aktivitäten, die zur Erreichung dieses Ziels durchgeführt wurden, umfassten:

1. Umfassende QRL-Literaturübersicht: Eine tiefgehende und systematische Untersuchung der QRL-Algorithmen wurde durchgeführt, die die Überprüfung von über 140 Forschungsarbeiten, Preprints und wichtigen Veröffentlichungen umfasste. Dieser gründliche Prozess gewährte ein umfassendes Verständnis der verschiedenen QRL-Algorithmen und ihrer Eigenschaften.
2. NISQ-Kompatibilitätsevaluation: Jeder in der QRL-Literaturübersicht identifizierte Algorithmus wurde rigoros auf seine Kompatibilität mit den inhärenten Einschränkungen der NISQ-Hardware bewertet. Diese Bewertung berücksichtigte kritische Faktoren wie Schaltkreisstruktur, Komplexität des Algorithmus und Gesamtressourcenerfordernisse, die für die praktische NISQ-Implementierung entscheidend sind. Diese Bewertung führte zur Auswahl geeigneter Algorithmen für die weitere Erforschung in den kommenden Arbeitspaketen.
3. Periodische Aktualisierung der QRL-Literaturübersicht: Um die Relevanz und Genauigkeit der Litteraturübersicht im sich schnell entwickelnden Forschungsumfeld aufrechtzuerhalten, wurde ein System für laufende periodische Aktualisierungen eingerichtet. Dieser kontinuierliche Überwachungsprozess verfolgt neue Veröffentlichungen, aufkommende Forschungstrends und wichtige Fortschritte und stellt sicher, dass die Literaturübersicht eine aktuelle und wertvolle Ressource für das Projekt und die breitere Community bleibt.
4. arXiv-Veröffentlichung zur Steigerung der Sichtbarkeit in der Community: Der Abschluss der umfassenden Literaturübersicht und rigorosen Analyse wurde auf dem arXiv-Preprint-Server veröffentlicht. Diese öffentliche Verbreitung dient dazu, eine wertvolle, frei zugängliche Ressource für die breitere Quantencomputing-Community beizutragen, die Zusammenarbeit zu fördern, redundante Forschungsanstrengungen zu verhindern und den allgemeinen Fortschritt des Feldes durch die weite Verbreitung wichtiger Einblicke und Erkenntnisse zu beschleunigen.

Arbeitspaket 1 hat effektiv eine dynamische und robuste Wissensbasis zu QRL-Algorithmen und deren NISQ-Eignung geschaffen und somit die Bühne für die nachfolgenden Phasen des Projekts bereitet, während es eine wertvolle, öffentlich zugängliche Ressource für die breitere Forschungscommunity bietet.

AP1.1: Bewertung nach Komplexität

Innerhalb von Arbeitspaket 1 wurde ein gezielter Aufwand (Arbeitspaket 1.1) unternommen, um die Komplexität der im ursprünglichen Literaturüberblick identifizierten QRL Algorithmen systematisch zu analysieren. Das Hauptziel war es, eine detaillierte Untersuchung der Rechenkomplexität jedes Algorithmus durchzuführen, mit einem speziellen Schwerpunkt auf der Quantifizierung der Gatterkomplexität der entsprechenden Quantenschaltkreise. Diese Untersuchung umfasste eine gründliche Analyse der operationalen Schritte innerhalb jedes Algorithmus, um deren Rechenanforderungen präzise zu bestimmen. Entscheidende Erkenntnisse dieser detaillierten Studie, einschließlich der abgeleiteten Gatterkomplexitäten, wurden direkt in die Literaturübersicht integriert. Diese Integration stellte sicher, dass die Literaturübersicht nicht nur einen breiten Überblick über verfügbare Algorithmen bietet, sondern auch wertvolle, Algorithmus-spezifische Einblicke in deren Ressourcenanforderungen liefert, wodurch die praktische Nützlichkeit der Literaturübersicht für das Projekt und die breitere Community verbessert wurde.

AP1.2: Berücksichtigung von Hardware-Eigenschaften

In Arbeitspaket 1.2 wurde der Bewertungsrahmen für QRL weiter verfeinert, um die beabsichtigte Hardwareplattform explizit zu berücksichtigen. Anstatt sich ausschließlich auf die allgemeine Komplexität zu konzentrieren, zielte dieses Arbeitspaket darauf ab, Algorithmen basierend auf spezifischen Hardwaretypen wie fehlertoleranter, NISQ und klassischer Hardware zu kategorisieren. Basierend auf den Ressourcenschätzungen von AP 1.1 war das Ziel, eine QRL-Klassifikationsmatrix zu erstellen, die die Hardware-Ressourcen darstellt, die erforderlich sind, um kleine RL-Probleme durch die Algorithmen zu lösen. Die Klassifikationsmatrix wurde basierend auf den folgenden Faktoren erstellt:

1. Hardware-fokussierte Algorithmenkategorisierung: Eine systematische Kategorisierung wurde durchgeführt, die QRL-Algorithmen gemäß ihren Ressourcenanforderungen klassifizierte: klassisch, NISQ oder fehlertoleranter Quantencomputer. Diese Klassifizierung bot eine klare ressourcenfokussierte Strukturierung der untersuchten Algorithmen.
2. "Grad der Quanten-Klassischen Hybridisierung" für die Hardwareklassifikation: Durch detaillierte Studien erwies sich der "Grad der Quanten-Klassischen Hybridisierung" als wertvolles Klassifikationsinstrument für diese hardwarezentrierte Kategorisierung. Dieser Indikator spiegelt effektiv das erforderliche Gleichgewicht zwischen quanten- und klassischen Ressourcen wider, das direkt mit den Stärken und Einschränkungen des Algorithmus in Verbindung steht.
3. Agent-Environment-Hybridisierung zur Ressourcenschätzung: Die Metrik "Grad der Quanten-Klassischen Hybridisierung" wurde weiter verfeinert, indem ihre Definition an den Ressourcenanforderungen sowohl des Agenten als auch des Environments, die innerhalb des Reinforcement Learnings interagieren, verankert wurde. Diese verfeinerte Definition bietet ein umfassenderes Verständnis der Hybridisierung in QRL-Algorithmen.

Arbeitspaket 1.2 lieferte eine strukturierte Analyse verschiedener QRL-Algorithmen und deren Komplexität. Die resultierende Klassifikationsmatrix der verschiedenen QRL-Methoden ist in Abschnitt 4.1 dargestellt. Darüber hinaus wurde das Potenzial von Fehlerresilienz in QRL-Algorithmen, die auf variationellen Quantenschaltkreisen beruhen, untersucht. Weitere Einzelheiten sind in Abschnitt 4.2 zu finden.

3.2 AP2: Auswahl geeigneter Anwendungsklassen

Arbeitspaket 2 war der wesentlichen Vorbereitungsaufgabe gewidmet, geeignete Anwendungsbereiche für die Forschungsarbeit zum Quanten Reinforcement Learning auszuwählen. Das übergeordnete Ziel war es, Anwendungsdomänen zu identifizieren, die nicht nur das inhärente Potenzial für Anwendung von Quantencomputern bieten, sondern auch die Möglichkeit bieten, verschiedene Kombinationen algorithmischer Ansätze und Hardwareplattformen sinnvoll zu erkunden. Um ein umfassendes und nuanciertes Verständnis potenzieller Anwendungsbereiche zu gewinnen, verfolgte Arbeitspaket 2 einen vielschichtigen Ansatz, der Literaturrecherche und direkte Interaktion mit Experten aus Industrie und Universitäten kombinierte. Eine Reihe strukturierter Diskussionen wurde mit den Projektpartnern von der OTH Regensburg, der SAG und IQM initiiert, um ihre unterschiedlichen Perspektiven und Fachkenntnisse zu nutzen, um vielversprechende Anwendungsbereiche zu identifizieren. Durch diesen kollaborativen Austausch wurde zunehmend deutlich, dass die aktuellen Einschränkungen der NISQ-Hardware die praktisch umsetzbaren algorithmischen Ansätze erheblich limitieren. Insbesondere wurde erkannt, dass die NISQ-Eignung effektiv die QRL-Forschung in der absehbaren Zukunft auf variationelle Quantenschaltkreise ausrichtet, wobei diese als Kernelemente in Modellen des maschinellen Lernens genutzt werden, sei es für überwachte oder unüberwachte Lernaufgaben oder in QRL-Algorithmen. Dieses Resultat fokussierte nicht nur die Forschungsausrichtung, sondern brachte auch wichtige Erkenntnisse hinsichtlich der Bedeutung und folglich der Bewertung von VQC-basierten Algorithmen mit sich. Bedenken wurden hinsichtlich der Fähigkeit geäußert, definitive Leistungsvergleiche mit klassischen Gegenstücken anzustellen und den endgültigen Einfluss und die Vorteile von VQC-basierten Ansätzen in realen Anwendungen sicher vorherzusagen. Darüber hinaus wurde ein bedeutender praktischer Engpass in Form der begrenzten Systemintegration der aktuellen NISQ-Geräte identifiziert. Dieser Mangel an nahtloser Integration wurde als wesentliche Einschränkung für das praktische

Training und die Implementierung erfolgreicher QRL-Algorithmen erkannt, was potenziell die Fähigkeit beeinträchtigen könnte, die theoretischen Vorteile von Quantencomputing Ansätzen in realistischen Umgebungen vollständig zu realisieren. Während die in Arbeitspaket 2 geleistete Arbeit erfolgreich die Forschungsfokussierung des Projekts auf eine realistischere und praktischere Erkundung von NISQ-kompatiblen QRL-Algorithmen eingrenzte und schärfte, hob sie auch hervor, dass viele grundlegende Forschungsfragen weiterhin unbeantwortet bleiben. Diese offenen Fragen verdeutlichen die Notwendigkeit fortlaufender, detaillierter und anwendungsorientierter Forschungsanstrengungen, um das Potenzial des QRL in der NISQ-Ära und darüber hinaus vollständig zu verstehen und effektiv zu nutzen.

3.3 AP3: Simulatorschnittstelle für industrielle Anwendungen

Diese Aufgaben wurden von der Siemens AG übernommen.

3.4 AP4: Auswahl, Anpassung und Entwicklung der Quantenalgorithmen

Arbeitspaket 4 wurde strategisch entworfen, um die Kluft zwischen theoretischem Verständnis und praktischer Anwendung im Bereich des QRL zu überbrücken. Das zentrale Ziel war es, die Erkenntnisse aus den vorherigen Arbeitspaketen in greifbare algorithmische Fortschritte zu übersetzen. Dies begann mit der sorgfältigen Auswahl der vielversprechendsten QRL-Algorithmen aus der umfassenden Literaturübersicht aus AP1, einem Auswahlprozess, der rigoros durch die Bewertungskriterien geleitet wurde, die in AP2 und AP3 festgelegt wurden und auf die Anforderungen spezifischer Zielanwendungen zugeschnitten war. Diese ausgewählten Algorithmen wurden nicht einfach übernommen, sondern aktiv angepasst und weiterentwickelt, um genau den Anforderungen dieser Anwendungen gerecht zu werden, indem auf das angesammelte Wissen aus AP2 und AP3 zurückgegriffen wurde, um ihre Leistung zu optimieren. Neben der Verfeinerung bestehender Methoden diente Arbeitspaket 4 auch als Innovationsmotor und förderte aktiv die Schaffung neuartiger QRL-Algorithmen. In Anerkennung der Tatsache, dass eine effektive Algorithmusbereitstellung mehr erfordert als nur den Algorithmus selbst, umfasste Arbeitspaket 4 auch die kritische Entwicklung wesentlicher unterstützender Werkzeuge. In direkter Verfolgung dieser Ziele erreichte Arbeitspaket 4 Folgendes: Angetrieben vom Ziel, vielversprechende Algorithmen für gezielte Anwendungen auszuwählen, identifizierte AP4.1 erfolgreich einen geeigneten Algorithmus aus dem Pool der zuvor untersuchten quanten- und klassischen RL Methoden, wobei Algorithmen priorisiert wurden, die ein starkes Potenzial für praktische Anwendungen in industriellen Anwendungsfällen zeigten. In Übereinstimmung mit dem Ziel der Algorithmusentwicklung und Innovation konzentrierte sich AP4.2 auf die Schaffung neuartiger QRL-Ansätze, was in der erfolgreichen Entwicklung eines neuen offline QRL-Algorithmus gipfelte, der das algorithmische Toolkit für QRL erweitert und neue Wege zur Bewältigung von RL-Herausforderungen eröffnet. In direkter Unterstützung des Ziels, wesentliche Werkzeuge für die Implementierung und Verbesserung von Algorithmen zu entwickeln, konzentrierte sich AP4.3 auf die Werkzeugentwicklung, was zur Entwicklung eines neuen Satzes von Werkzeugen führte, einschließlich effizienter Methoden zur Gradientenabschätzung, die speziell zur Verbesserung der praktischen Anwendbarkeit, Leistung und Benutzerfreundlichkeit von QRL-Algorithmen entwickelt wurden. Durch diese fokussierten Anstrengungen übersetzte Arbeitspaket 4 effektiv Forschung und Analyse in greifbare Ergebnisse, indem sowohl verfeinerte Algorithmen für spezifische Anwendungen als auch neuartige Werkzeuge zur Förderung der praktischen Anwendung von QRL bereitgestellt wurden.

AP4.1: Auswahl und Anpassung der Quantenalgorithmen für Reinforcement Learning

Arbeitspaket 4.1 konzentrierte sich auf die entscheidende Anpassung von QRL-Algorithmen und erkannte die Notwendigkeit, diese Algorithmen sowohl an die spezifischen Anforderungen industrieller Optimierungs- und Steuerungsaufgaben als auch an verschiedene Simulationsumgebungen und angestrebte Quantenhardware anzupassen. AP4.1 begann mit der sorgfältigen Auswahl der vielversprechendsten Quantenalgorithmen basierend auf den Anwendungskriterien, die in AP2 definiert und in AP3 weiter verfeinert wurden. Der ausgewählte Algorithmus zur Anpassung heißt Deep Q-Learning (DQN). DQN wurde als Zielalgorithmus ausgewählt wegen seiner Off-Policy-Struktur und seiner Fähigkeit, aus vorab gesammelten Daten zu lernen. Dieser Auswahlprozess ebnete dann den Weg für die Algorithmusanpassung, die zwei Hauptpfade verfolgte: Erstens durch die

Einbeziehung anwendungsspezifischer Anforderungen wie größere Dimensionen der Datenpunkte, und zweitens durch die Berücksichtigung der Einschränkungen und Möglichkeiten, die sich aus verschiedenen Simulationsumgebungen und Hardware ergeben, wie z.B. langsame sequenzielle Verarbeitung von Eingaben. Um die Anforderung der Kodierung von Datenpunkten mit großen Dimensionen zu erfüllen, untersuchte AP4.1 verschiedene Kodierungsstrategien für klassische Informationen in quantenmechanische Zustände. Diese Untersuchung führte zu einer neuartigen Kodierungsstrategie namens inkrementales Daten-Uploading, die in Abschnitt 4.3 vorgestellt wird. Im Verlauf von AP1, AP2 und AP3 wurde zunehmend deutlich, dass die Möglichkeit, Quantenvorteile mit NISQ-Hardware und variationellen Quantenalgorithmen oder Heuristiken zu erreichen, ungewiss bleibt. Selbst für Quantenalgorithmen für QRL mit nachweisbaren Vorteilen in asymptotischer Komplexität bleiben die praktischen Vorteile aufgrund des Aufwands für die erforderliche Fehlerkorrektur unklar. Folglich wurde der Prozess der Algorithmusanpassung hauptsächlich von den Anforderungen an den Datenzugriff der NISQ-Hardware und den spezifischen Bedürfnissen des in AP2 und AP3 betrachteten industriellen Anwendungsfalls getrieben. Um die Parallelverarbeitungsfähigkeiten zu verbessern und das Training auf echter Quantenhardware zu beschleunigen, wurde der Deep Q-Network (DQN)-Algorithmus angepasst, um sich nahtlos in das Qiskit Session-Framework zu integrieren, wodurch er für die Ausführung auf Quantencomputern optimiert wurde.

AP4.2: Entwicklung von neuen Quantenalgorithmen

Direkt aufbauend auf der Auswahl- und Anpassungsarbeit von AP4.1 konzentrierte sich Arbeitspaket 4.2 auf das zentrale Ziel der Entwicklung neuer QRL-Algorithmen. Dieses Arbeitspaket hatte zum Ziel, die Grenzen bestehender QRL-Methoden zu erweitern und neuartige Ansätze zu schaffen, die darauf zugeschnitten sind, die Herausforderungen und Möglichkeiten anzugehen, die in früheren Phasen des Projekts identifiziert wurden. Das Fraunhofer IIS entwickelte einen neuartigen QRL-Algorithmus namens Batch Constrained Quantum Q-Learning (BCQQ). Dieser Algorithmus zeigte eine überlegene Leistung beim Lernen einer optimalen Policy im Vergleich zu seinem klassischen Gegenstück mit einer vergleichbaren Anzahl von Parametern in Bezug auf die Dateneffizienz und hebt die potenziellen Vorteile des quanteninspirierten Ansatzes hervor. Ergänzend zur Entwicklung von BCQQ wurde auch eine neue Kodierungsstrategie namens zyklisches Daten-Reuploading innoviert. Diese neuartige Kodierungstechnik wurde entwickelt, um die Expressivität von Funktionsapproximatoren, die auf variationellen Quantenschaltkreisen beruhen, zu erhöhen und somit das Toolkit für das Quantum Machine Learning im Kontext des Reinforcement Learning und darüber hinaus zu bereichern. Eine detailliertere Erklärung dieser Ergebnisse ist in Abschnitt 4.4 enthalten. Durch diese fokussierten Anstrengungen lieferte Arbeitspaket 4.2 konkrete Fortschritte in der Entwicklung von QRL-Algorithmen und produzierte sowohl Anpassungen bestehender Methoden für Quantencomputer als auch völlig neue Algorithmen mit erweiterten Fähigkeiten.

AP4.3: Entwicklung von Werkzeugen zur Korrektheitsprüfung

Arbeitspaket 4.3 konzentrierte sich auf die Entwicklung wesentlicher Werkzeuge, um die effiziente Entwicklung und Implementierung von QRL-Algorithmen zu erleichtern. Basierend auf den Erkenntnissen und praktischen Erfahrungen aus AP4.1 und AP4.2 hatte dieses Arbeitspaket das Ziel, das Toolkit für QRL-Anwender zu erweitern. Der Aufwand konzentrierte sich auf die Verbesserung der Methoden zur Gradientenschätzung und klassischer Optimierer, da diese die Hauptengpässe in Bezug auf die NISQ-Ressourcenerfordernisse darstellen und für robuste und effiziente QRL-Algorithmen von entscheidender Bedeutung sind. Im Rahmen dieses Ziels führte Arbeitspaket 4.3 eine detaillierte Studie durch, um die geeignetsten klassischen Optimierer für die Gradientenschätzung zu identifizieren. Hier verglichen wir die Leistung verschiedener Optimierer bei einer Lernaufgabe unter Verwendung unterschiedlicher Quantencomputer. Weitere Einzelheiten sind in Abschnitt 4.5 zu finden. Darüber hinaus wurde auf dieser Analyse aufbauend eine neuartige Methode zur Gradientenschätzung namens Guided-SPSA eingeführt, die eine schnellere Konvergenz demonstrierte und eine signifikante Reduzierung der Anzahl der nötigen Schaltkreisausführungen erreichte, und zwar konkret eine Reduzierung um 15% bis 25% im Vergleich zu bestehenden Methoden. Eine detaillierte Erklärung von Guided-SPSA wird in Abschnitt 4.6 bereitgestellt. Dieser Fortschritt trägt direkt zu einem effizienteren und weniger ressourcenintensiven Training von QRL-Algorithmen bei.

3.5 AP5: Einsatz auf Quantencomputer-Simulator und realem Quantencomputer

Arbeitspaket 5 befasste sich mit der praktischen Bereitstellung von QRL-Algorithmen sowohl für Simulationsumgebungen als auch für Quantenhardware. Das Kernziel bestand darin, bestehende Simulations- und Hardwarelösungen zu bewerten, geeignete Plattformen auszuwählen und anschließend die entwickelten Quantenalgorithmen zu implementieren und zu testen. In Anerkennung der Tatsache, dass die klassische Simulation für die Entwicklung und Machbarkeitstests von QRL-Algorithmen vor der Ausführung auf Quantenhardware von entscheidender Bedeutung ist, priorisierte das Arbeitspaket eine gründliche Evaluierungsphase. Diese Evaluierungs- und Implementierungsanstrengung wurde hauptsächlich von IQM und SAG durchgeführt, wobei das Fraunhofer IIS eine Bewertung der Simulationsumgebungen und der Quantenhardwareangebote von IBM beitrug. Die Untersuchung der verfügbaren Quantenhardware offenbarte erhebliche Einschränkungen. Die Fehlerquoten für Ein- und Zweiqubitgatter wurden als erheblich befunden und lagen zwischen 0,1% und 0,7%. Diese hohen Fehlerquoten machten die Ergebnisse von Quantenschaltkreisen mit mehr als 12 Qubits unzuverlässig. Darüber hinaus führten die zu diesem Zeitpunkt geltenden Zugriffsmodelle auf Quantencomputer zu erheblichen Wartezeiten für die Schaltkreisausführung, was das Training sequenzieller Algorithmen wie dem Reinforcement Learning auf echter Quantenhardware praktisch unmöglich machte.

3.6 AP6: Praxisnahe Anwendungen

Arbeitspaket 6 hatte zum Ziel, die praktische Anwendbarkeit der im Rahmen anderer Arbeitspakete bewerteten Methoden des QRL innerhalb spezifischer Anwendungsszenarien, die in seinen Unterarbeitspaketen behandelt wurden, zu validieren. Ein zentrales Ziel war es, Anforderungen aus diesen Anwendungsstudien zu sammeln und in die anderen Arbeitspakete zurückzuführen, um die Forschungs- und Entwicklungsanstrengungen zu verfeinern. Die erwarteten Ergebnisse dieses Arbeitspakets waren Berichte, die die Anwendbarkeit, Vorteile, Herausforderungen und offenen Forschungsthemen im Zusammenhang mit den entwickelten Methoden detailliert darlegten, betrachtet durch die Linse jeder Anwendung. Arbeitspaket 6 wurde überwiegend von den Projektpartnern (SAG, IQM) bearbeitet. Das Team des IIS nahm aktiv an mehreren Programmierworkshops teil, um die Eignung und Leistung der entwickelten Algorithmen und Werkzeuge in praktischen Anwendungen zu bewerten. Darüber hinaus wurden alle während des Projekts erstellten Werkzeuge und Algorithmen rigoros in verschiedenen Anwendungsfällen getestet, und entsprechende Machbarkeitsstudien wurden durchgeführt, um ihr Potenzial und ihre Einschränkungen in der realen Welt zu bewerten.

3.7 AP7: Bibliothek zur Nutzbarmachung für industrielle Anwendungen

Arbeitspaket 7 konzentrierte sich auf die Erstellung einer Softwarebibliothek, die die Nutzung der QRL-Algorithmen des Projekts für potenzielle industrielle Anwendungen erleichtern sollte. Das Hauptziel bestand darin, eine robuste Softwaregrundlage zu schaffen, die nicht nur unmittelbare Anwendungen unterstützt, sondern auch mittel- und langfristig eine breitere Bedeutung erzielt. Um dies zu erreichen, war ein Teil der Softwarebibliothek für die Veröffentlichung als Open Source vorgesehen. Diese Open-Source-Strategie zielte speziell darauf ab, akademischen Partnern zu ermöglichen, die Ergebnisse des Projekts in zukünftigen Kooperationen zu nutzen und eine Plattform für die weitere Forschung an den entwickelten Algorithmen zu bieten. Dieser Ansatz sollte sicherstellen, dass die Ergebnisse des Projekts auf Anwendungsbereiche übertragen und angepasst werden können, die über den unmittelbaren Fokus des Projekts hinausgehen. In Übereinstimmung mit diesem Ziel wurden wichtige Erkenntnisse von Fraunhofer IIS, wie der BCQQ-Algorithmus, in die Softwarebibliothek des Projekts integriert. Darüber hinaus wurde diese Softwarebibliothek der breiteren Forschungs- und Entwicklungsgemeinschaft öffentlich zugänglich gemacht [19].

3.8 AP8: Dissemination

Arbeitspaket 8 hatte zum Ziel, die Verbreitung und kritische Bewertung der Zwischenergebnisse des Projekts sicherzustellen. Das Hauptziel war es, die Ergebnisse frühzeitig im Projektzyklus Fachpublikum zu präsentieren und

zu diskutieren, während auch Fortschritte an die Industrie und die breite Öffentlichkeit kommuniziert wurden. Darüber hinaus strebte AP8 an, das gewonnene Wissen in akademische Lehrpläne zu integrieren, um den langfristigen Einfluss des Projekts zu maximieren. Diese proaktive Kommunikationsstrategie wurde entwickelt, um wertvolles Feedback von externen Parteien zu bestehenden Berichten und laufenden Arbeiten innerhalb von AP2, AP4 und AP5 zu erhalten. Folglich sollten die verschiedenen Unterarbeitspakete innerhalb von AP8 wichtige Erkenntnisse und Bewertungen generieren, die für den Gesamterfolg des Projekts als wesentlich erachtet wurden. Die spezifischen Aktivitäten, die unternommen wurden, um diese Ziele zu erreichen, sind in den einzelnen Unterarbeitspaketen von AP8 detailliert beschrieben.

AP8.1: Sichtbarmachung in der Fachöffentlichkeit

Arbeitspaket 8.1 konzentrierte sich auf die entscheidende Aufgabe, die Forschungs- und Entwicklungsergebnisse des Projekts sowohl an akademisches als auch industrielles Fachpublikum zu verbreiten. Das Hauptziel war es, Diskussionen zu erleichtern und Networking-Möglichkeiten innerhalb der QML-Gemeinschaft zu fördern. Dies wurde durch die Präsentation der Projektergebnisse auf Fachkonferenzen erreicht, gekoppelt mit der Erkundung potenzieller Entwicklungsrichtungen, wie der Organisation eines QML-Workshops auf einer bedeutenden wissenschaftlichen Konferenz. Diese Interaktionen sollten speziell wertvolles Feedback sammeln, um die methodische Richtung der Entwicklungsarbeit innerhalb des Projekts zu steuern. Zu diesem Zweck wurden zwei Iterationen eines QML-Workshops auf der IEEE Quantum Computing and Engineering Conference organisiert (<https://qml.lfdr.de/2024/#/home>). Darüber hinaus wurde auch ein spezieller QRL-Workshop auf derselben renommierten Konferenz durchgeführt. Die Forschungsarbeiten des Projekts wurden zudem aktiv durch die Präsentation und Veröffentlichung von 7 wissenschaftlichen Artikeln auf führenden Konferenzen im Fachgebiet kommuniziert. Diese Veranstaltungsorte boten wertvolle Gelegenheiten für Diskussionen und den Austausch von Wissen mit Experten aus Industrie und Wissenschaft.

AP8.2: Integration in die akademische Lehre

Arbeitspaket 8.2 hatte zum Ziel, das im Projekt entwickelte Wissen und die Konzepte in die akademische Lehre zu integrieren, mit einem starken Fokus auf den Beitrag zum deutschen Arbeitsmarkt. Das Hauptziel war die zeitnahe Einbindung der Projektergebnisse in Seminare und Vorlesungen, wobei der dringende Bedarf an qualifizierten Quantencomputing-Wissenschaftlern in Deutschland erkannt wurde. Diese unmittelbare Integration in die akademischen Lehrpläne, durchgeführt von der OTH Regensburg, dient direkt dem aktuellen Projekt und spielt zudem eine wichtige Rolle in der langfristigen Entwicklung der deutschen Arbeitskräfte, indem sie qualifizierte Ingenieure und Wissenschaftler in diesem entscheidenden aufstrebenden Feld ausbildet. Darüber hinaus förderte das Projekt aktiv die nächste Generation von Quantenexperten durch die Betreuung von studentischen Arbeiten. Das Fraunhofer IIS trug dazu bei, indem ein Praktikum und ein laufendes Promotionsvorhaben betreut wurden, wodurch der Wissenstransfer an die deutsche Industrie, einschließlich großer Unternehmen und international tätiger KMUs, erleichtert und direkt zur Schaffung eines starken nationalen Talentpools im Bereich der Quantentechnologien beigetragen wurde. Des Weiteren wurde im Rahmen des "Fut3ech Symposium" an der TH Deggendorf auf Einladung von IBM ein Vortrag zum Thema Quantum Reinforcement Learning und den Forschungsarbeiten am Fraunhofer IIS gehalten.

AP8.3: Vermittlung an die Industrie

Arbeitspaket 8.3 hatte zum Ziel, die Projektzwischenergebnisse einem breiten industriellen Publikum aus der Industrie zu vermitteln. Das Hauptziel war es, Ergebnisse auf relevanten Messen und Workshops zu präsentieren, um eine Vielzahl von Industrievertretern zu erreichen. Angesichts der erheblichen technologischen Hürden, die durch Quantentechnologien entstehen, konzentrierte sich AP8.3 darauf, ein Verständnis für deren Anwendungspotenzial aufzubauen. Zu diesem Zweck wurden die in verschiedenen Arbeitspaketen entwickelten Anwendungsklassen an mehreren Veranstaltungsorten präsentiert und diskutiert, um weitere strategische Partner für den Transfer der Projektergebnisse und des Wissens in vielfältige Anwendungen zu gewinnen. Das gesammelte Feedback wurde zur Optimierung der Zwischenergebnisse in das Projekt integriert. Um die Arbeiten des Projekts zu

präsentieren, wurde ein Demonstrator unter Verwendung eines Anki Overdrive-Rennsystems entwickelt. Eine detailliertere Erklärung des Demonstrators ist in Abschnitt 5.1 zu finden. Um weitere Anknüpfungspunkte an die Industrie zu finden, wurden Projektergebnisse im Rahmen mehrerer Veranstaltungen von Bayern Innovativ vorgestellt und diskutiert. Besonders hervorzuheben sind hierbei die Workshops zu den Themen "*Quantum DevOps - Neue Einsatzgebiete für Machine Learning und mathematische Optimierung*" und "*Quanten Computing: Optimierung und Quantum Machine Learning - wie implementieren?*".

AP8.4: Kommunikation an die Öffentlichkeit

Arbeitspaket 8.4 konzentrierte sich auf die aktive Kommunikation der Projektergebnisse an akademische und industrielle Interessengruppen sowie an die breite Öffentlichkeit. Das Hauptziel war es, zugängliche Kanäle zur Verbreitung von Informationen zu schaffen, wie eine dedizierte Projektwebsite mit einem Blog und einer E-Mail-Liste. Um die breite Öffentlichkeit effektiv zu erreichen, wurde Wert darauf gelegt, das komplexe Thema greifbar zu präsentieren. Soziale Medien wie LinkedIn wurden für eine weitere Verbreitung genutzt. Diese Kommunikationsbemühungen zielten auch darauf ab, Unternehmen mit peripherem Interesse zu engagieren, damit sie dem Konsortium leicht Feedback und Anregungen geben können. Um diese Ziele zu erreichen, wurde eine Projektwebsite, QLinda (<https://www.iis.fraunhofer.de/de/ff/lv/dataanalytics/anwproj/qlinda.html>), erstellt. Im Verlauf des Projekts wurden mehrere LinkedIn-Beiträge geteilt, um Zwischenergebnisse, Veranstaltungen und Erfolge zu kommunizieren. Darüber hinaus wurden mehrere öffentliche Vorträge bei Veranstaltungen wie der "Langen Nacht der Wissenschaft" gehalten, um mit Personen außerhalb des Bereichs Quantencomputing in Kontakt zu treten.

4 Beschreibung der Ergebnisse und Veröffentlichungen

Dieser Abschnitt skizziert die zentralen Ergebnisse und Publikationen, die aus der Beteiligung des Fraunhofer IIS am QLinda-Projekt hervorgegangen sind. Die Arbeiten "Aufdeckung von Instabilitäten in QRL" [20], "Batch-Constrained Quantum Q-Learning" [21], "SPSA-Optimierer" [22] und "Guided-SPSA" [23] repräsentieren die direkten Ergebnisse unserer Bemühungen, QRL für industrielle Anwendungen innerhalb des NISQ-Paradigmas im Rahmen des QLinda-Projekts zu realisieren. Es ist auch erwähnenswert, dass der Forschungsfortschritt innerhalb von QLinda durch Ergebnisse wie "Inkrementelles Daten-Reloading" [24] und andere Arbeiten [25], [26], [27], [28], [29] weiter unterstützt wurde, die als Ergebnis der synergistischen Interaktion zwischen QLinda und anderen zeitgleich laufenden Projekten entwickelt wurden.

4.1 Klassifikationsmatrix

Abbildung 4.1 zeigt die Klassifikationsmatrix verschiedener existierender QRL-Algorithmusklassen. Die Matrix ist basierend auf den klassischen und quantenmechanischen Rechenressourcen gruppiert, die vom Reinforcement-Learning-Agenten und der Umgebung benötigt werden. Die Matrix bietet auch einen Einblick in die Kompatibilität der Algorithmen mit NISQ- und fehlertoleranten Quantencomputern.

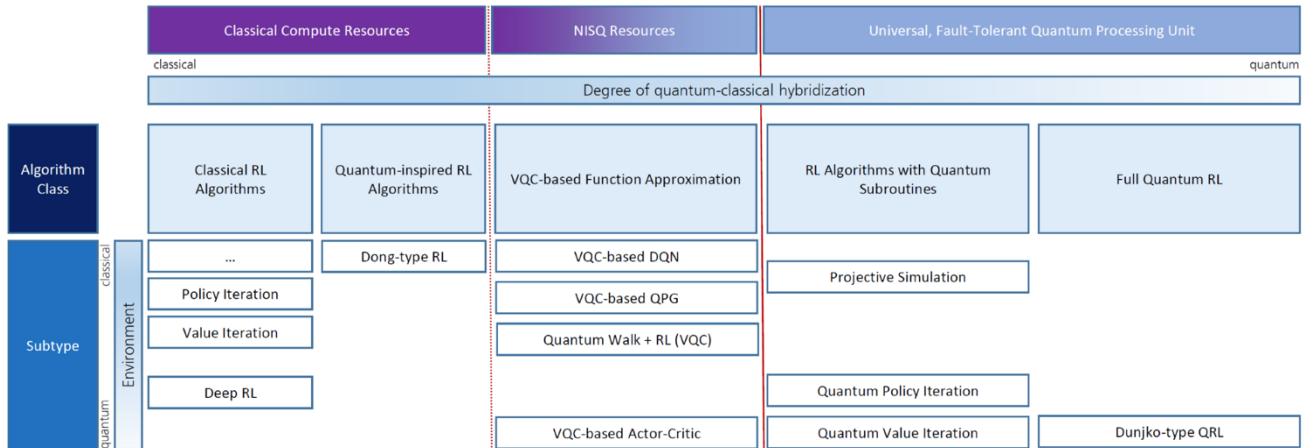


Abbildung 4.1: Klassifizierungsmatrix der verschiedenen QRL-Algorithmen

4.2 Aufdeckung von Instabilitäten in QRL

In dieser Arbeit [20] präsentieren wir eine umfassende Untersuchung von Variational Quantum Deep Q-Networks (VQ-DQNs), einer Klasse hybrider quanten-klassischer RL Algorithmen, die variationelle Quantenschaltkreise (VQCs) in den Deep Q-Learning-Rahmen integrieren. Unsere Arbeit ist motiviert durch die erheblichen Rechenanforderungen traditioneller Deep RL Algorithmen, die erhebliche Trainingszeit und Ressourcen erfordern, um zu einer optimalen Policy zu konvergieren. Wir untersuchen, ob Quantencomputing mit seinem Potenzial für die Steigerung der Rechengeschwindigkeit praktische Vorteile für RL durch die Verwendung von VQ-DQNs bieten kann. Wir beobachteten, dass VQ-DQNs trotz ihrer theoretischen Attraktivität oft Trainingsinstabilitäten aufweisen, die manchmal zu einer divergierenden Policy führen. Solche Instabilitäten beeinträchtigen nicht nur die praktische Nützlichkeit dieser Algorithmen, sondern werfen auch Bedenken hinsichtlich der Reproduzierbarkeit der in der Literatur berichteten Ergebnisse auf. Um diese Probleme zu untersuchen, führten wir eine gründliche Reproduktionsstudie bestehender VQ-DQN-Ansätze durch, insbesondere der von Chen et al. [30], Lockwood et al. [31], und Skolik et al. [32] veröffentlichten. Unter Verwendung ihres veröffentlichten Quellcodes replizierten wir Experimente in der CartPole-v1-Umgebung und identifizierten erhebliche methodische Probleme, die die Stabilität und Leistung der Algorithmen beeinflussen. Ein großes Problem, das wir fanden, hing mit der Frequenz der Parameterupdates zusammen. Einige Implementierungen führten Parameterupdates nur einmal pro Episode durch, was erheblich von den klassischen DQN-Praktiken abweicht, bei denen diese häufiger durchgeführt werden (z. B. nach jeder Aktion oder in regelmäßigen Abständen). Diese seltenen Updates können zu Instabilität führen, da der Replay-Puffer schneller wächst als der Agent seine Policy aktualisiert, was zu veralteten und nicht repräsentativen Erfahrungen führt, die das Lernen behindern. Darüber hinaus können die verwendeten Methoden zur Performanzbewertung, wie das Mitteln des Returns über Episoden mit variierenden Policy aufgrund von epsilon-greedy Exploration, zugrunde liegende Instabilitäten verschleiern, indem sie Schwankungen in der Performanz des Agenten glätten. Um diese Probleme anzugehen, implementierten wir den DQN mit VQC neu und hielten uns eng an etablierte RL Praktiken, um einen fairen Vergleich zu gewährleisten. Unsere Implementierung umfasst regelmäßige Updateintervalle, eine konsistente Performanzbewertung unter Verwendung der aktuellen greedy Policy und Strategien zur Parameterinitialisierung, die darauf abzielen, Probleme wie das Barren Plateau-Phänomen zu mildern, das zu verschwindenden Gradienten in VQC führt. Wir führten systematische Experimente in der CartPole-v0-Umgebung durch, um die Auswirkungen verschiedener Datenkodierungsstrategien und Methoden zur Q-Value-Extraktion auf die Leistung und Stabilität von VQ-DQNs zu bewerten. Konkret untersuchten wir verschiedene Kodierungsstrategien, einschließlich kontinuierlicher (C), skaliertes & kontinuierlicher (SC) und skaliertes & gerichteter (SD) Kodierung, sowie verschiedene Methoden zur Q-Value-Extraktion wie lokale Skalierung (LS) und globale Skalierung (GS). Unsere Ergebnisse zeigten, dass VQ-DQNs in allen Kombinationen von Kodierungs- und Extraktionsmethoden anfällig für Instabilitäten sind.

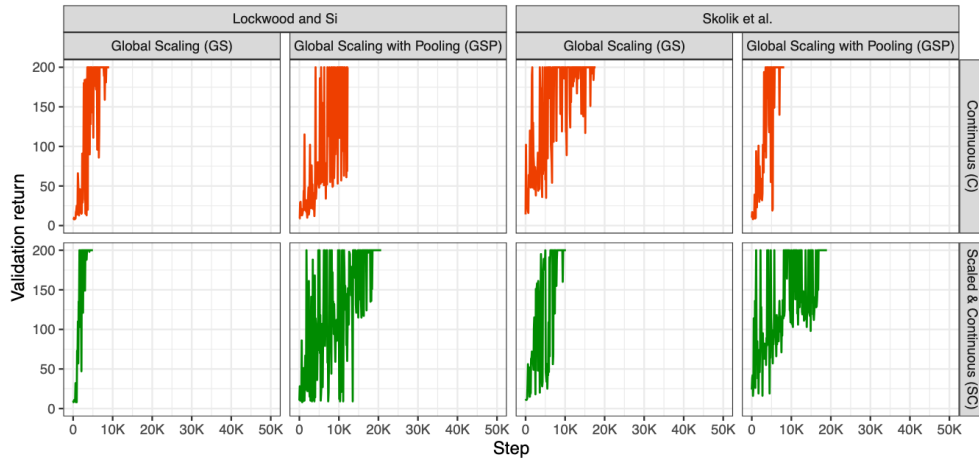


Abbildung 4.2: Validierungsreturn für unsere am besten abschneidenden Hyperparameter-Konstellationen in den Basis-Konfigurationen mit Daten-Reuploading

Durch umfangreiche Hyperparameteroptimierung entdeckten wir jedoch, dass bestimmte Einstellungen diese Probleme mildern konnten. Agenten, die die SC-Kodierung mit GS-Extraktion verwendeten, zeigten in einigen Fällen eine verbesserte Leistung und schnellere Konvergenz, wie in der Abbildung 4.2 dargestellt.

Wir untersuchten auch die Strategie des Daten-Reuploadings, die die Expressivität des VQC erhöht, indem die Datenkodierungslayer mehrfach innerhalb des Quantenschaltkreises eingefügt werden. Obwohl diese Technik in bestimmten Fällen die Leistung geringfügig verbesserte, führte sie nicht konsistent zu einer erhöhten Trainingsstabilität oder Konvergenzgeschwindigkeit. Darüber hinaus stellen die erhöhte Schaltkreistiefe, die für das Daten-Reuploading erforderlich ist, praktische Herausforderungen auf fehlerbehafteter Hardware dar. Um die Auswirkungen der Einschränkungen von Quantenhardware zu bewerten, validierten wir unsere trainierten Agenten auf einem IBM-Quantencomputer. Wir stellten fest, dass die Performanz des Agenten im Vergleich zu Simulationen erheblich abnahm. Diese Verschlechterung ist auf Faktoren wie Dekohärenz, Imperfektionen der Quantengatter, eingeschränkte Qubit-Verbindungen und praktische Einschränkungen wie lange Ausführungszeiten und lange Wartezeiten auf gemeinsam genutzter Hardware zurückzuführen. Das Training von RL-Agenten direkt auf Quantenhardware erwies sich als unpraktisch aufgrund des Zugriffsmodells der Geräte. Abbildung 4.3 zeigt die Validierungsergebnisse auf der IBMQ-Hardware.

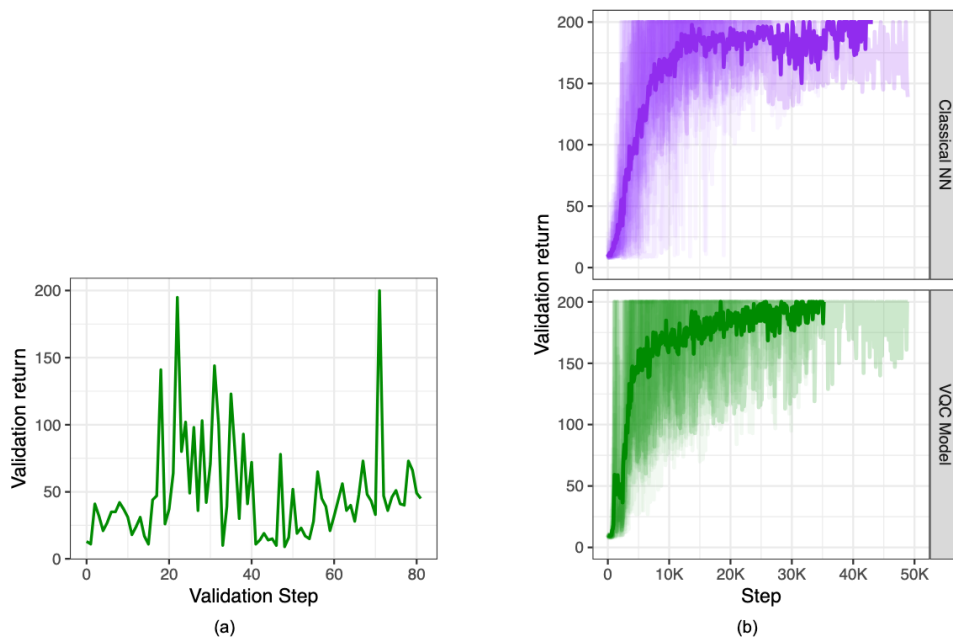


Abbildung 4.3: (a) Ergebnisse unseres Validierungslaufs auf ibmq_ehningen. (b) Vergleich zwischen VQ-DQN und klassischen NN, gemittelt über 30 verschiedene Agenten.

Beim Vergleich von VQ-DQNs mit klassischen neuronalen Netzwerken beobachteten wir, dass VQ-DQNs zwar in

bestimmten Settings zunächst schneller lernen konnten, klassische Netzwerke letztendlich jedoch vergleichbare oder bessere Leistungen mit größerer Stabilität und weniger Sensitivität gegenüber Hyperparametern erzielten. Dies lässt sich aus der Abbildung 4.3 ableiten. Dies deutet darauf hin, dass VQ-DQNs entgegen einiger Behauptungen in der Literatur derzeit möglicherweise keinen Vorteil gegenüber klassischen Ansätzen bieten, insbesondere angesichts der aktuellen Hardwareeinschränkungen. Unsere Untersuchung deckt erhebliche Instabilitäten in VQ-DQN-Ansätzen auf, die sowohl algorithmischen Faktoren als auch Hardwareeinschränkungen zuzuschreiben sind. Zu unseren Beiträgen gehört die Identifizierung methodologischer Probleme in früheren Studien, die Bereitstellung systematischer Bewertungen von Kodierungs- und Extraktionsmethoden, die Durchführung umfassender Hyperparameteranalysen und die Bereitstellung einer robusten, reproduzierbaren Implementierung von VQ-DQN. Wir heben auch die praktischen Herausforderungen bei der Bereitstellung von RL-Algorithmen auf aktueller Quantenhardware hervor. Unsere Ergebnisse unterstreichen die Bedeutung einer rigorosen Bewertung und einer vorsichtigen Interpretation der behaupteten Vorteile in QRL-Algorithmen. Behauptungen von Quantenvorteilen sollten mit Vorsicht behandelt werden, insbesondere wenn sie auf Simulationen basieren, die die realen Hardwareeinschränkungen nicht berücksichtigen. Zusammenfassend zeigt unsere Studie, dass Quantencomputing zwar erhebliches Potenzial zur Verbesserung des Reinforcement Learning auf lange Sicht birgt, die aktuellen VQ-DQN-Ansätze jedoch aufgrund von Trainingsinstabilitäten und Hardwareeinschränkungen erheblichen Herausforderungen gegenüberstehen. Die Bewältigung dieser Herausforderungen erfordert eine Kooperation zwischen Algorithmenspezialisten, Entwicklern von Quantenhardware und der breiteren Forschungscommunity.

4.3 Inkrementelles Daten-Reloading

Die Publikation mit dem Titel " Incremental Data-Uploading for Full-Quantum Classification" [24] befasst sich mit den Herausforderungen, hochdimensionale klassische Daten effektiv in Quantenschaltkreise zu kodieren, die für NISQ Hardware geeignet sind. Quantum Machine Learning bietet vielversprechende Möglichkeiten für Quantenvorteile, aber aktuelle Einschränkungen der Quantenhardware, wie eine begrenzte Anzahl von Qubits und niedrige Gattergenauigkeit, erschweren die Implementierung komplexer Quantenalgorithmen – insbesondere bei der Verarbeitung hochdimensionaler Daten. Im klassischen maschinellen Lernen spielt die Darstellung von Daten eine entscheidende Rolle für die Performanz von Modellen. Diese Bedeutung wird im QML verstärkt, in dem das Kodieren hochdimensionaler Daten in einen Quantenschaltkreis ohne Informationsverlust eine nicht triviale Aufgabe darstellt. Traditionelle Kodierungsschemata sehen häufig einen Kompromiss zwischen Einfachheit und Informationserhalt vor. Einfache Kodierungsmethoden, wie die Verwendung von Ein-Qubit-Rotationsgattern, können zu Informationsverlusten führen, während komplexe Kodierungsschemata, die Verschränkung und Daten-Reloading beinhalten, die Tiefe und die Anzahl der Gatter in Quantenschaltkreisen erhöhen. Solche Erhöhungen sind aufgrund der begrenzten Kohärenzzeiten und Fehlerraten von NISQ-Hardware unpraktisch. Um diese Herausforderungen anzugehen, schlagen wir eine neuartige Kodierungstechnik namens inkrementelles Daten-Reloading (IDU) vor. Die Kernidee hinter IDU besteht darin, die Kodierungsgatter im gesamten quantenmechanischen Schaltkreis zu verteilen und sie mit (trainierbaren) Layern zu mischen. Anstatt alle Kodierungsgatter zu Beginn des Schaltkreises zu platzieren – was zu Informationsverlust führen kann – speist die IDU-Methode die Daten schrittweise in den Schaltkreis, wie in Abbildung 4.4 dargestellt. Dieser Ansatz stellt sicher, dass die Daten während des gesamten Berechnungsprozesses für das Quantenmodell zugänglich bleiben, wodurch die Fähigkeit des Modells verbessert wird, komplexe Muster zu erlernen, ohne die Schaltkreistiefe oder die Gatteranzahl signifikant zu erhöhen.

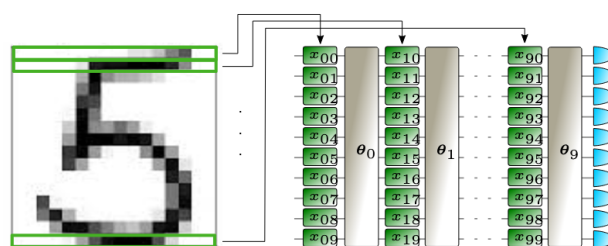


Abbildung 4.4: Vorgeschlagene Methode - Das Bild wird auf 10×10 heruntergerechnet und dann zeilenweise an den 10-Bit-Quantenschaltkreis übergeben, jedes Kodierungslayer (grün) wird von einem trainierbaren Layer (grau) gefolgt, dann werden die 10 Qubits gemessen (blau).

Wir bewerten die Performanz der IDU-Methode für Bildklassifikation unter Verwendung der MNIST- und Fashion-MNIST-Datensätze. Diese Datensätze werden standardmäßig als Benchmark im maschinellen Lernen verwendet und bestehen aus Graustufenbildern handschriftlicher Ziffern und Kleidungsstücken, die jeweils mit einer von zehn Klassen gekennzeichnet sind. Um die Experimente rechnerisch handhabbar zu machen, insbesondere da die Simulationen auf klassischer Hardware mit TensorFlow Quantum und Cirq durchgeführt wurden, wurden die Bilder von ihrer ursprünglichen Größe von 28×28 Pixeln auf 10×10 Pixel unter Verwendung eines bilinearen Filters heruntergerechnet. Diese Herabstufung reduzierte die Dimensionalität der Daten, während sie dennoch wesentliche Merkmale beibehielt, die für die Klassifikation notwendig sind. Wir verwenden einen 10-Qubit-Quantenschaltkreis, um ihn mit den 10 Klassen in den Datensätzen in Einklang zu bringen. Jeder Datenpunkt (Bild) bestand nach der Herabstufung aus 100 Merkmalen (Pixeln). Die Kodierungsschichten bestanden aus Ein-Qubit-Rotationsgattern Rx, eines für jedes Pixel, wobei die Pixelwerte entsprechend skaliert wurden, um als Rotationswinkel zu dienen. Die trainierbaren Layer bestanden aus Ein-Qubit-Ry- und Rz-Rotationsgattern sowie benachbarten kontrollierten R_y-Verschränkungs-Gattern. Diese Konfiguration balanciert den Bedarf an Expressivität mit den praktischen Einschränkungen von NISQ-Geräten. Durch die schrittweise Erhöhung der Anzahl der eingefügten variationellen Layer in den Kodierungsblock wollten wir untersuchen, wie die Datenzugänglichkeit innerhalb des Quantenschaltkreises die Performanz des Modells beeinflusst. Die Hypothese war, dass der Wechsel von Kodierungs- und variationellen Layern eine bessere Datenrepräsentation aufrechterhalten würde, was zu verbesserter Trainierbarkeit und Klassifikationsgenauigkeit führt. Die Experimente zeigten eine direkte Korrelation zwischen der Anzahl der eingefügten variationellen Layer in den Kodierungsblock und der Performanz des QML Modells. Konkret erzielten Modelle mit mehr Splits im Kodierungsblock eine höhere Klassifikationsgenauigkeit sowohl für die Trainings- als auch für die Testdaten. Beispielsweise erzielte die Architektur mit zehn Kodierungssplits (IDU 10), bei der Kodierungs- und variationelle Layer vollständig vermischt wurden, die höchste Genauigkeit von etwa 60% für den MNIST-Datensatz. Tabelle 4.1 zeigt die Genauigkeit für den Testdatensatz, die durch IDU im Vergleich zu naiver Kodierung und Daten-Reloading-Strategie erreicht wurde.

Datensatz	DRU	IDU				
		1	2	4	8	10
MNIST	33.2±0.01	34.0±0.01	47.3±0.03	50.9±0.01	51.5±0.00	56.7±0.02
gemischt	32.2±0.00	47.1±0.01	52.2±0.01	53.8±0.01	56.1±0.01	58.6±0.01
Fashion	43.5±0.17	43.8±0.01	48.3±0.01	52.5±0.01	53.6±0.03	56.9±0.03

Tabelle 4.1: Klassifikationsgenauigkeit des inkrementellen Daten-Reloadings für verschieden Testdatensätze.

Um sicherzustellen, dass die beobachteten Performanzverbesserungen nicht spezifisch für den MNIST-Datensatz gelten, führten wir zusätzliche Experimente mit dem Fashion-MNIST-Datensatz und einer gemischten Version des MNIST-Datensatzes durch, bei der die Pixel jedes Bildes zufällig permutiert wurden. Die auf diesen Datensätzen trainierten Modelle wiesen ähnliche Performanzmerkmale auf, was bestätigt, dass die Vorteile der IDU-Methode allgemein und nicht datenspezifisch sind. Wir analysierten auch die theoretischen Eigenschaften der IDU-Methode, indem wir die Trainierbarkeit und Expressivität der Modelle durch die Fisher Informations Matrix (FIM) und die Effektive Dimension untersuchten. Die FIM erfasst die Geometrie des Parameterraums und beeinflusst die Trainierbarkeit des Modells. Ein gleichmäßigeres Eigenwertspektrum der FIM deutet auf einen weniger verzerrten Parameterraum hin, was für Optimierungsmethoden wie Gradientdescent vorteilhaft ist. Die Effektive Dimension quantifiziert die Expressivität des Modells, indem sie die Größe der Funktionenmenge misst, deren Elemente das Modell approximieren kann. Die Analyse zeigte, dass die FIM für Modelle mit einer höheren Anzahl von Kodierungssplits in der IDU-Architektur ein gleichmäßigeres Eigenwertspektrum aufweist. Diese Gleichmäßigkeit deutet auf eine verbesserte Trainierbarkeit hin, da der Parameterraum weniger verzerrt ist, was eine effizientere Optimierung erleichtert. Darüber hinaus erhöhte sich die normierte effektive Dimension mit der Anzahl der Kodierungssplits, was auf eine größere Expressivität des Modells hinweist. Zusammenfassend bietet die Methode des inkrementellen Daten-Reloadings (IDU) einen neuartigen und effektiven Ansatz zur Kodierung hochdimensionaler Daten in Quantenschaltkreise, die für NISQ-Geräte geeignet sind. Durch den Wechsel von Kodierungs- und variationellen Layern verbessert die IDU-Methode die Datenzugänglichkeit und -Datenrepräsentation innerhalb des Quantenschaltkreises, ohne die Schaltkreistiefe oder die Gatteranzahl signifikant zu erhöhen. Dieser Ansatz verbessert sowohl die Trainierbarkeit als auch die Expressivität von QML Modellen, was zu einer besseren Klassifikationsgenauigkeit auf Benchmark-Datensätzen führt. Diese Arbeit trägt zur QML und QRL Forschung bei,

indem sie sowohl praktische als auch theoretische Einblicke in Datenkodierungsstrategien für hochdimensionale Datenpunkte, wie sie in industriellen Anwendungen vorkommen, bietet.

4.4 Batch-Constrained Quantum Q-Learning (BCQQ)

Diese Publikation [21] befasst sich mit den Herausforderungen der Anwendung von RL in realen, sicherheitskritischen Szenarien, in denen Umgebungsinteraktionen während des Trainings entweder unpraktisch oder unsicher sind. Traditionelle Methoden des Deep RL erfordern häufig umfangreiche Daten und Interaktionen mit der Umgebung, was in solchen Umgebungen nicht machbar ist. Batch Reinforcement Learning (Batch RL) bietet eine Lösung, indem Agenten mit vorab gesammelten Datensätzen trainiert werden, ohne weitere Interaktionen mit der Umgebung. Bestehende Batch RL Algorithmen benötigen jedoch typischerweise große Datensätze, um effektiv zu funktionieren, was Einschränkungen mit sich bringt, wenn die Datenverfügbarkeit begrenzt ist. Neueste Fortschritte im Quantencomputing deuten darauf hin, dass Quantenmodelle, insbesondere VQC möglicherweise weniger Daten für das Training benötigen als klassische Methoden. Motiviert durch diesen potenziellen Vorteil schlagen wir einen Batch RL Algorithmus vor, der VQC als Funktionsapproximatoren im Rahmen des diskreten Batch-Constrained Deep Q-Learning (BCQ) verwendet. Darüber hinaus führen wir ein neuartiges zyklisches Daten-Reloading-Schema ein, das die Ausdruckskraft der VQCs erhöht, ohne die Anzahl der Parameter zu erhöhen. Dies umfasst zwei Beiträge: (1) Batch-Constrained Quantum Q-Learning (BCQQ) (2) Zyklisches Daten-Reloading-Schema. Im vorgeschlagenen Verfahren integrieren wir VQC in den diskreten BCQ-Algorithmus, um den BCQQ-Algorithmus zu bilden. Im BCQ-Rahmen wird ein generatives Modell G_ω trainiert, um die Behavior-Policy aus den Batch-Daten zu schätzen. Die Policy ist darauf beschränkt, Aktionen auszuwählen, die unter der Behavior-Politik wahrscheinlich sind. Wir ersetzen sowohl das Q-Netzwerk als auch das generative Modell durch VQC, wodurch der Agent möglicherweise effektiver aus begrenzten Daten lernen kann. Das zyklische Daten-Reloading-Schema besteht darin, die Datenkodierungslayer vor jedem variationellen Layer nochmals einzuführen, während die Reihenfolge der Eingabedaten zyklisch unter den Qubits verschoben wird, wie in Abbildung 4.5 dargestellt. Dies bedeutet, dass jedes Qubit allen Eingabefeatures über verschiedene Layer hinweg ausgesetzt ist, wodurch die Fähigkeit des Schaltkreises, komplexe Muster in den Daten zu erfassen, verbessert wird.

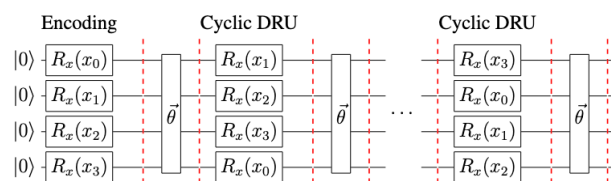


Abbildung 4.5: Quantenagent mit zyklischer Daten-Reloading-Strategie

Wir bewerten unseren BCQQ-Algorithmus mit zyklischem Daten-Reloading mit Hilfe der OpenAI CartPole-v1-Umgebung unter verschiedenen Bedingungen: (1) Training auf zufälligen Trajektorien: Wir trainieren den Agenten mit Datensätzen, die von einer zufälligen Policy mit Pufferspeichergrößen von 100, 10^4 und 10^6 Proben gesammelt wurden. Unsere Ergebnisse zeigen, dass der Quantenagent mit zyklischem Daten-Reloading selbst mit nur 100 zufälligen Proben nahezu optimale Leistungen erzielt und klassische Agenten, die mit denselben Daten trainiert wurden, übertrifft. (2) Training auf teilweise verrauschten Trajektorien: Um das Lernen von begrenzten Expertendaten zu simulieren, sammeln wir Datensätze von 25, 50 und 100 Environment-Übergängen von einer verrauschten Expertenpolicy. Der BCQQ-Agent lernt in allen Fällen erfolgreich eine optimale Policy und zeigt eine überlegene Dateneffizienz im Vergleich zu klassischen Agenten. Die Ergebnisse sind in Abbildung 4.6 dargestellt. (3) Modellauswahl und -analyse: Wir analysieren die effektive Dimension sowohl der QML als auch der klassischen Modelle mithilfe der Fisher Information Matrix (FIM). Die zyklische Daten-Reloading-Strategie erhöht die effektive Dimension des VQC und verbessert die Expressivität, ohne die Anzahl der Parameter zu erhöhen. (4) Validierung auf Quantenhardware: Wir testen den trainierten Quantenagenten auf einem verrauschten Simulator und Quantenhardware (IBM-Quantencomputer). Der Agent behält bei ausreichender Anzahl von Messungen eine gute Performanz und profitiert von Techniken zur Fehlerminderung wie der Zero-Noise Extrapolation, wie in Abbildung 4.6 dargestellt.

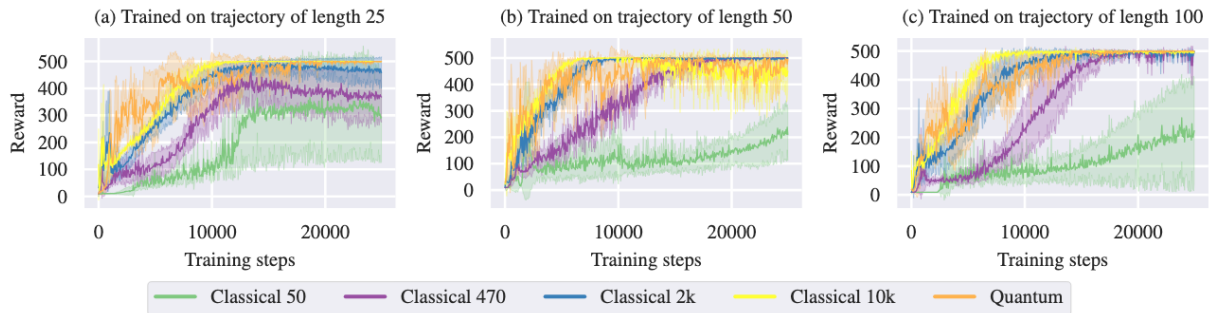


Abbildung 4.6: Abbildung (a), (b) und (c) zeigt die Lernkurven des Quantenagenten mit zyklischer Daten-Reloading-Strategie und verschiedenen klassischen Agenten, die auf teilweisen verrauschten Trajektorien der Länge 25, 50 und 100 trainiert wurden. Die gezeigten Ergebnisse sind über 3 Trainingsläufe gemittelt, wobei jede Bewertung aus den Returns besteht, die über 10 zufällige Seeds gemittelt wurden.

Unsere Experimente legen nahe, dass der BCQQ-Algorithmus, der zyklisches Daten-Reloading integriert, es Quantenagenten ermöglicht, effektive Policies mit erheblich weniger Daten als klassische Agenten zu erlernen. Der Quantenagent bewältigt nicht nur erfolgreich das CartPole-v1-Environment, sondern zeigt auch eine verbesserte Generalisierung und Stabilität, insbesondere wenn er mit begrenzten oder verrauschten Daten trainiert wird. Die experimentellen Ergebnisse bieten mehrere wichtige Erkenntnisse über die Leistung und

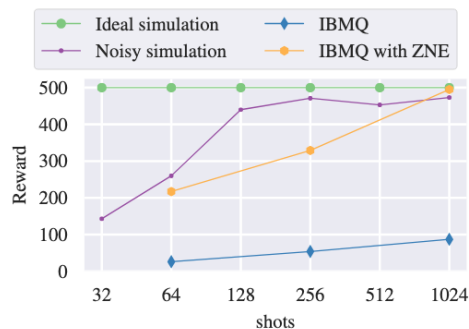


Abbildung 4.7: Validierungsergebnisse eines Quantenagenten, der zunächst auf einem idealen Simulator (ohne Simulation von Fehlern) trainiert wurde. Dieser wurde daraufhin auf einem Simulator, der Fehlereinfluss simuliert und auf Quantenhardware getestet. Die Experimente auf Hardware wurden mit und ohne Zero-Noise Extrapolation durchgeführt.

potenzielle Vorteile des BCQQ-Algorithmus: (1) Überlegene Dateneffizienz: Der Quantenagent übertraf konsistent klassische Agenten mit vergleichbarer Parameteranzahl, wenn er mit begrenzten Datenmengen oder verrauschten Daten trainiert wurde. Dies deutet darauf hin, dass VQC, insbesondere mit zyklischem Daten-Reloading, möglicherweise weniger Datenproben benötigen, um effektive Policies zu erlernen, was einen erheblichen Vorteil in realen Szenarien darstellt, in denen die Datensammlung kostspielig oder unpraktisch ist. (2) Verbesserte Modellexpressivität: Die zyklische Daten-Reloading-Strategie verbessert die Expressivität des VQC, ohne die Anzahl der Parameter zu erhöhen. Durch das zyklische Verschieben der Reihenfolge der Eingabefeatures, die in Qubits kodiert sind, wird jedes Qubit allen Eingabefeatures über verschiedene Layer hinweg ausgesetzt, was dem QRL Modell ermöglicht, komplexere Muster in den Daten zu erfassen. (3) Robustheit gegenüber limitierter Datenmengen und verrauschten Daten: Die Fähigkeit des Quantenagenten, aus nur 25 verrauschten Environment-Agent Interaktionen bzgl. einer Expertenpolicy zu lernen, zeigt eine Robustheit, die klassischen Agenten mit ähnlichen Parameteranzahlen fehlt. Diese Robustheit ist entscheidend in sicherheitskritischen Anwendungen, in denen die Daten unvollkommen oder rar sein können. (4) Praktikabilität auf NISQ-Geräten: Trotz der Herausforderungen durch fehlerbehaftete Hardware zeigt unsere Validierung auf Quantenhardware, dass der BCQQ-Algorithmus bei ausreichenden Messungen und Techniken zur Fehlermitigation wie Zero Noise Mitigation eine relativ stabile Leistung erbringen kann. Diese Erkenntnis ist ermutigend für die zukünftige Bereitstellung von QRL Algorithmen für Quantenhardware. Diese Ergebnisse deuten darauf hin, dass QRL Algorithmen vielversprechend für praktisch relevante Anwendungen sind, bei denen Daten begrenzt oder teuer zu beschaffen sind. Die Möglichkeit, über QRL mit VQC einen Quantenvorteil zu finden, bleibt jedoch aufgrund der fehlerbehafteten Natur der Quantenhardware, des extremen Aufwands, der durch Fehlerkorrektur- und Mitigationstechniken verursacht wird, und der inhärenten Einschränkungen, wie z.B. Barren Plateaus, ungewiss.

4.5 SPSA-Optimierer

Hier präsentieren wir einen umfassenden empirischen Vergleich verschiedener Optimierer für QML wobei wir uns

speziell auf die Verwendung von Simultaneous Perturbation Stochastic Approximation (SPSA)-basierten Gradientenschätzungsverfahren in Kombination mit gradientenbasierten klassischen Optimierern konzentrieren [22]. Unsere Motivation ergibt sich aus den Herausforderungen, die mit der exakten Berechnung von Gradienten in VQC unter Verwendung der Parameter Shift Rule verbunden sind, die einen erheblichen Sampling-Aufwand erzeugt, der proportional zur Anzahl der Parameter im Quantenschaltkreis ist. Die Parameter Shift Rule liefert exakte Gradienten, die Kosten für die Berechnung werden jedoch prohibitiv, da sie $2p$ Erwartungswertschätzungen für p Parameter erfordert. Im Gegensatz dazu erfordert SPSA nur zwei Schätzungen pro Aktualisierungsschritt, unabhängig von der Anzahl der Parameter, und skaliert somit auf große Quantenschaltkreise. Da die SPSA-Gradienten jedoch approximativ sind, können sie instabil bzgl. der Richtung maximaler Steigung sein. Da klassische Optimierer eine entscheidende Rolle in der Abstiegsrichtung der Optimierung, der Stabilität und der benötigten Anzahl von Schritten spielen, beabsichtigen wir, den am besten geeigneten klassischen Optimierer zu finden, der die SPSA-Gradienten stabilisiert und zu einer optimalen Konvergenz führt. Wir testeten die Kombination von SPSA mit modernen gradientenbasierten klassischen Optimierern, nämlich Adam, RMSProp, AMSGrad, Stochastic Gradient Descent (SGD) und SGD mit Momentum. Diese Analyse zielt darauf ab, die niedrigen Berechnungskosten von SPSA beizubehalten, während die Konvergenzeigenschaften fortschrittlicher gradientenbasierter Optimierer genutzt werden. Um die Effektivität der Kombination zu bewerten, führen wir umfangreiche numerische Experimente zu mehreren Regressionsaufgaben unter Verwendung von Datensätzen aus der scikit-learn-Bibliothek und realen Daten, wie dem Combined Cycle Power Plant-Datensatz, durch. Unsere Experimente beinhalten das Training von VQC mit vier bis fünf Qubits und mehreren variationellen Layers, wobei klassische Eingabedaten in Quantenzustände kodiert werden, und das Ziel darin besteht, eine Loss-Funktion durch Variation der Parameter zu minimieren. Wir führen Simulationen unter vier verschiedenen Bedingungen durch, um die Robustheit unserer Methode zu bewerten: (1) Ideale fehlerfreie Simulation: Analytische Berechnung der Erwartungswerte ohne Berücksichtigung von Fehler. (2) Shot-basierte Simulation: Einbeziehung von Shot-Noise durch Simulation mit einer endlichen Anzahl von Shots pro Erwartungswertschätzung. (3) Fehlersimulation mit realistischen Hardwarefehlern: Nachahmung der Fehlercharakteristika der IBM-Quantenhardware `ibmq_ehningen`. (4) Fehlersimulation mit Fehlermitigation: Anwendung von Techniken zur Zero Noise Extrapolation zur Minderung der Auswirkungen von Hardwarefehlern. Unsere Ergebnisse zeigen, dass die Kombination von SPSA-basierten Gradientenschätzungen mit fortschrittlichen Optimierern sowohl die Standard-SPSA (unter Verwendung von SGD) als auch die Parameter-Verschieberegeln hinsichtlich der Konvergenzgeschwindigkeit und des endgültigen absoluten Fehlers über alle Datensätze und Simulationsbedingungen hinweg erheblich übertrifft. Insbesondere erreicht der AMSGrad-Optimierer, wenn er mit SPSA-Gradienten verwendet wird, konstant die beste Leistung und reduziert sowohl die benötigte Anzahl an Trainingsschritten als auch den Berechnungsaufwand. Darüber hinaus zeigt unser Ansatz eine verbesserte Resilienz gegenüber Shot-Noise und Hardwarefehlern. Die Performanzverbesserung von SPSA mit AMSGrad gegenüber der Standard-SPSA wird unter Einbeziehung von Hardwarefehlern verstärkt, was die Praktikabilität der Kombination für aktuelle NISQ Geräte unterstreicht. Darüber hinaus stellen wir fest, dass die Verwendung von Fehlermitigation die relative Performanz der Optimierer nicht erheblich beeinflusst, obwohl sie die absoluten Werte der Zielfunktion aufgrund der erhöhten Schätzvarianz beeinflussen kann.

4.6 Guided-SPSA

In unserer Publikation mit dem Titel "Guided-SPSA: Simultaneous Perturbation Stochastic Approximation Assisted by the Parameter Shift Rule" [23] präsentieren wir eine neuartige Gradientenschätzungstechnik namens Guided-SPSA zur Optimierung von VQC in variationellen Quantenalgorithmien (VQA). Wie in Abschnitt 4.5 erklärt, liefern traditionelle gradientenbasierte Methoden wie die Parameter Shift Rule exakte Gradienten für VQA, leiden jedoch unter Berechnungskomplexitäten, die linear mit der Anzahl der Parameter im Quantenschaltkreis skalieren. Diese Skalierung wird für größere Quantenschaltkreise unpraktikabel, insbesondere angesichts der Einschränkungen aktueller NISQ-Hardware, wie begrenzte Qubitanzahlen, kurze Kohärenzzeiten und hohe Fehlerraten. Andererseits skalieren gradientenfreie Methoden wie SPSA nicht mit der Anzahl der Parameter, haben jedoch oft Probleme mit der Stabilität und können aufgrund der approximativen Natur der berechneten Gradienten zu suboptimalen Lösungen konvergieren. Unsere primäre Forschungsfrage befasst sich mit der Möglichkeit, eine Gradientenschätzungstechnik für VQA zu entwickeln, die eine stabile Konvergenz mit erheblich weniger Schaltkreisauswertungen im Vergleich zu bestehenden Methoden erreicht, auf aktueller Quantenhardware realisierbar ist und für NISQ-Hardware geeignet ist, ohne die Schaltkreiskomplexität in Bezug auf die Gatteranzahl oder die Anzahl der erforderlichen Messungen zu erhöhen. Um dies zu beantworten, führen wir Guided-SPSA ein, einen neuartigen Ansatz, der die Genauigkeit der Parameter Shift Rule sinnvoll mit der Rechen-effizienz der SPSA-basierten Gradientenschätzung kombiniert. Guided-SPSA nutzt die Vorteile beider Methoden, indem es die Eingabedaten zwischen den beiden Methoden zur Gradientenschätzung aufteilt und die Gradienten während des Trainingsprozesses sinnvoll kombiniert. Dieser Ansatz reduziert die insgesamt während des Trainings erforderlichen Schaltkreisauswertungen um etwa 15% bis 25% im Vergleich zur alleinigen Verwendung der Parameter Shift Rule, ohne die Stabilität der Konvergenz oder die Optimalität der Lösung zu beeinträchtigen. Im Guided-SPSA-Algorithmus

beginnen wir mit der Berechnung der maximalen und minimalen Anzahl von Perturbation-Samples, die während des Trainings basierend auf der Anzahl der Parameter im VQC und dem Parameter-Shift Sample-Ratio τ verwendet werden sollen. Die Perturbation-Samples für SPSA beginnen bei einem Mindestwert und steigen linear, während das Training fortschreitet. Diese Strategie nutzt den Vorteil von SPSA, indem sie in den frühen Phasen des Trainings kleinere Perturbation-Samples verwendet, wodurch die Berechnungskomplexität drastisch reduziert wird, die dadurch entstehenden Ungenauigkeiten aber akzeptierbar bleiben. Während jeder Trainingsepoche führen wir einen Forward-Pass durch, in dem wir Vorhersagen und Fehler auf Basis der aktuellen Parameter berechnen. Im Backward-Pass teilen wir die Daten in zwei Teile basierend auf τ . Für die Datensamples, die für die Parameter Shift Rule vorgesehen sind, berechnen wir die exakten Gradienten. Für die verbleibenden Datensamples, die für SPSA vorgesehen sind, berechnen wir approximative Gradienten und passen deren Größe unter Verwendung der durchschnittlichen Norm der durch die Parameter Shift Rule berechneten Gradienten an. Durch die Kombination dieser Gradienten stellen wir sicher, dass die Lösung nicht erheblich von den exakten Gradienten abweicht, was die Instabilität reduziert, insbesondere in den frühen Trainingsphasen. Wir führten umfangreiche numerische Experimente durch, um die Leistung von Guided-SPSA über verschiedene Paradigmen des QML zu bewerten, einschließlich Regression, Klassifikation und QRL. Für Regressionsaufgaben verwendeten wir die Friedman- und Boston-Housing-Datensätze. Der Funktionsapproximator war ein VQC, der aus vier oder fünf Qubits bestand, mit einem Kodierungslayer, mehreren variationellen Layer und einem Dekodierungslayer unter Verwendung von Observablen. Für Klassifikationsaufgaben wandten wir Guided-SPSA auf den Iris-Datensatz an, ein Multi-Klassen-Klassifikationsproblem. Zur Anwendung auf QRL trainierten wir einen Agenten mithilfe eines Quanten Policy Gradient Ansatzes, um das CartPole-v0-Environment zu lösen. In allen Szenarien zeigte Guided-SPSA eine vergleichbare Performanz zur Parameter Shift Rule, während 15% bis 25% weniger Schaltkreisbewertungen erforderlich waren. Abbildung 4.8 zeigt die Trainingsergebnisse des Quantenagenten für das CartPole-v0-Environment unter Verwendung der Parameter Shift Rule und Guided-SPSA. Zusammenfassend zeigt unsere Arbeit, dass Guided-SPSA eine effektive und praktikable Gradientenschätzungstechnik zur Optimierung von VQC in variationellen Quantenalgorithmen ist. Sie erreicht eine stabile Konvergenz mit weniger Schaltkreisbewertungen, was sie gut geeignet macht für die Implementierung auf aktueller NISQ Hardware, wo die Rechenressourcen begrenzt sind.

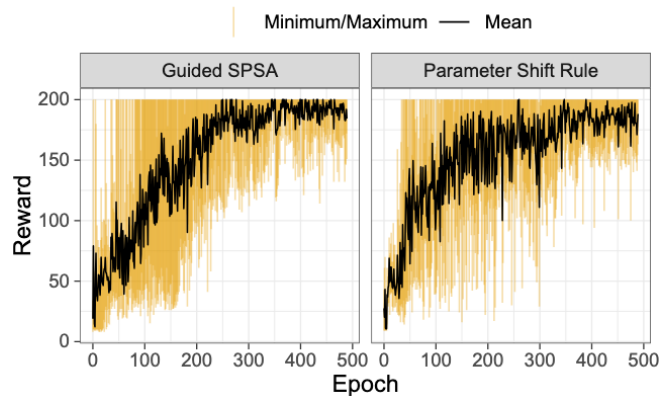


Abbildung 4.8: Trainingsverhalten eines QRL-Agenten für Guided SPSA und PS-Regel.

5 QRL-Demonstrator

5.1 Anki Overdrive

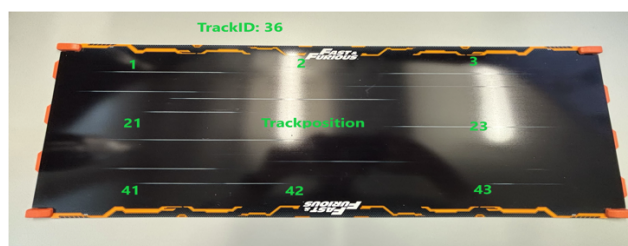


Abbildung 5.1: Streckenstück von Anki Overdrive mit Beispielpositionswerten.

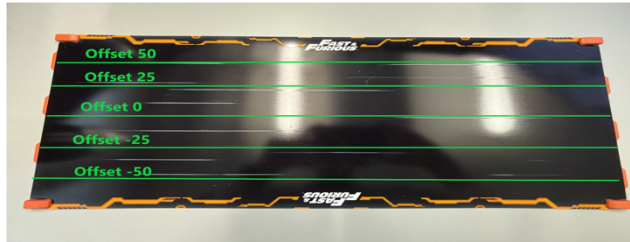


Abbildung 5.2: Streckenstück von Anki Overdrive mit markierten Versatzbereichen.

Die Kontrolle eines Fahrzeugs auf einer Spielzeugrennbahn des Herstellers Anki Overdrive, einem System für Roboter-Spielzeugauto-Rennen, stellt eine Herausforderung dar, die durch die Implementierung eines QRL-Agenten angegangen werden soll. Anki Overdrive ist ein Spiel, bei dem die Spieler mit App-gesteuerte Roboterautos auf modularen Strecken gegeneinander antreten und die Rennen durch Hochgeschwindigkeitsmanöver und Spurwechsel gekennzeichnet sind. Der Begriff „Overdrive“ beschreibt treffend einen Zustand unkontrollierten, oft hektischen Spielens, bei dem die präzise Kontrolle zugunsten aggressiver, jedoch ineffizienter Handlungen verloren geht. Dieses Overdrive-Phänomen äußert sich darin, dass Spieler, oder in unserem Fall autonome Agenten, die Autos über ihre optimalen Betriebsgrenzen hinausdrängen, was zu erratischen Trajektorien, verpassten Streckenabschnitten und suboptimalen Rundenzeiten führt. Um dem entgegenzuwirken, kann ein QRL-Agent so gestaltet werden, dass er optimale Steuerungsrichtlinien für die Navigation auf der Anki Overdrive-Strecke erlernt, indem er die Prinzipien der Quantenmechanik nutzt, um seine Lern- und Entscheidungsprozesse zu verbessern. Ein solcher Agent wurde durch iterative Interaktionen mit der Anki Overdrive-Rennstrecke trainiert und erhielt Feedback in Form von Belohnungen für wünschenswerte Ergebnisse wie das Beibehalten der Streckenposition und das Erreichen schneller Rundenzeiten, während er für unerwünschte Ergebnisse wie das Verlassen der Strecke oder Kollisionen bestraft wurde. Die Anki Overdrive-Strecke besteht aus zwei verschiedenen Arten von Streckenteilen, nämlich geraden und gebogenen Teilen. Wenn das Auto von einem Teil zum nächsten wechselt, erhält der Controller die Koordinaten des Autos auf der Strecke, die aktuelle Geschwindigkeit und die Beschleunigung des Autos als Live-Feedback. Jedes Streckenstück enthält eine eindeutige Kennung, die hilft, die Koordinaten des Autos auf der Strecke zuzuordnen. Das Streckenstück ist in neun Segmente unterteilt, wie in Abbildung 5.1 gezeigt. Der Versatz des Autos von der Mitte der Strecke, wie in Abbildung 5.2 gezeigt, zusammen mit der eindeutigen Kennung jedes Streckenteils und dem Segment innerhalb dieses Teils, wird verwendet, um den genauen Standort des Autos zu bestimmen.

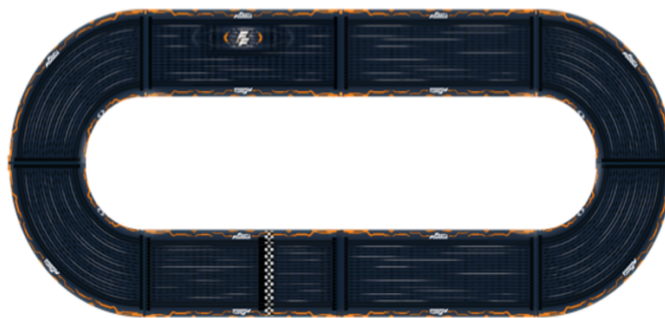


Abbildung 5.3: Anki Overdrive-Streckendesign, das für Training und Tests verwendet wird.

DQN-basiertes QRL wurde als Trainingsalgorithmus ausgewählt, um die Fähigkeiten des QRL-Agenten zu überprüfen, der aus Live-Feedback lernt. Zu einem gegebenen Zeitpunkt besteht das Feedback, das dem Agenten zugeführt wird, aus der Übergangszeit von einem Streckenstück zum anderen, dem aktueller Versatz, der aktuelle Strecken-ID, dem folgende Strecken-ID, der Streckenart als Boolescher Wert und die letzten beiden vom Agenten getroffenen Entscheidungen. Der Agent darf eine der folgenden Änderungen vornehmen, sobald er den aktuellen Zustand beobachtet: Geschwindigkeit um 100 erhöhen, Geschwindigkeit um 100 verringern, Versatz +50 (äußere linke Seite), Versatz +25, Versatz 0 (Mitte), Versatz -25, Versatz -50 (äußere rechte Seite). Die Zeit, die das Auto benötigt, um ein Streckenstück zu überqueren, wird als Belohnung an den Agenten übergeben. Um die Einfachheit zu gewährleisten, wird ein kreisförmiges Layout, wie in Abbildung 5.3 gezeigt, als Testumgebung verwendet, um die Performanz des QRL-Agenten zu trainieren und zu bewerten. Der Agent erkundet hier die Strecke, indem er zu Beginn zufällige Aktionen ausführt und eine Belohnung für diese Aktionen erhält. Während des Trainings lernt der Agent, immer bessere Aktionen zu wählen, sodass die Zeit zum Überqueren eines Streckenstücks minimiert und die erreichte Belohnung maximiert wird. Ein klassischer DQN-Agent wurde ebenfalls im selben Setup trainiert, um die Leistung des QRL-Agenten mit einem klassischen RL-Agenten zu vergleichen. Die von QRL- und klassischen RL-Agenten während des Trainings benötigte Rundenzeit ist in Abbildung 5.4 dargestellt. Die Ergebnisse zeigen, dass der

Quantenagent zu Beginn schneller konvergiert, jedoch dazu neigt, für mehrere Epochen in einem lokalen Optimum stecken zu bleiben, bevor er schließlich eine optimale Policy erlernt. Diese Beobachtung stimmt mit den Ergebnissen dieses Projekts überein, die zeigen, dass VQC-basiertes QRL das Potenzial hat, eine optimale Policy zu erlernen, das Training schneller konvergiert und dateneffizienter ist. Allerdings kann das Training eines QRL-Agenten auch sehr instabil sein.

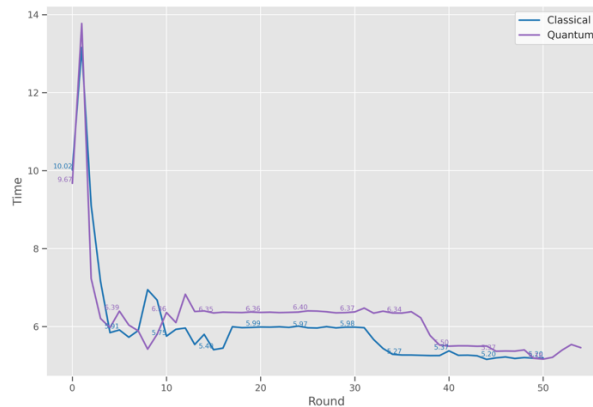


Abbildung 5.4: Rundenzeit, die von QRL- und klassischen RL-Agenten erreicht wird.

6 Externer Fortschritt

Im Laufe des Projekts behielten wir die bedeutenden Fortschritte in den breiteren Bereichen des QRL und des Quantum Machine Learning im Auge. Dazu gehörte das Aufkommen neuartiger Algorithmen und die Entwicklung innovativer Werkzeuge, die darauf abzielen, Herausforderungen wie Barren Plateaus in variationellen Quantenschaltkreisen zu bewältigen. Unser wesentliches Ergebnis verstärkt jedoch das anhaltende und signifikante Engpassproblem, das durch Rauschen auf realer Quantenhardware entsteht. Der erhebliche Aufwand, der mit Techniken zur Fehlerunterdrückung und Fehlerkorrektur verbunden ist, ist zwar notwendig, um dieses Rauschen zu bekämpfen, bleibt jedoch ein großes Hindernis für das Erreichen praktischer quantenmechanischer Vorteile über verschiedene Quantenalgorithmen hinweg, einschließlich derjenigen, die in diesem Projekt untersucht wurden. Diese fundamentale Einschränkung bleibt eine zentrale Herausforderung für das gesamte Feld.

7 Relevanz und Nutzen der erzielten Ergebnisse

Das Projekt brachte bedeutende Fortschritte in mehreren Schlüsselbereichen. Ein zentrales Ergebnis war die Entwicklung eines neuartigen QRL-Algorithmus, genannt BCQQ, der speziell für industrielle Anwendungen entwickelt wurde, bei denen Daten begrenzt und kostspielig zu erwerben sind. Dieser Offline-Trainingsansatz macht BCQQ besonders gut geeignet für reale Szenarien und stimmt mit den Einschränkungen der aktuellen NISQ-Hardware überein, einschließlich eingeschränktem Gerätezugang und der Notwendigkeit eines effizienten parallelen Trainings. Über den BCQQ-Algorithmus hinaus führte das Projekt mehrere innovative Methoden und Werkzeuge ein und veröffentlichte diese, die über QRL hinausgehen und eine breitere Anwendbarkeit im Bereich des Quantum Machine Learning haben. Dazu gehören verbesserte Kodierungsstrategien und verbesserte Gradientenabschätzungsmethoden, die zu effizienteren und effektiveren Trainingsverfahren für eine breitere Palette von Quantenalgorithmen beitragen, was der breiteren Quantenforschungs-Community zugutekommt. Darüber hinaus lieferte das Projekt eine umfassende Zusammenfassung der bestehenden QRL-Methoden, verbunden mit der Erstellung einer Open-Source-Softwarebibliothek. Diese wertvolle Ressource stellt einen bedeutenden Beitrag zur Quantencomputing-Community dar und ermöglicht Forschern und Praktikern den direkten Zugang zu den wegweisenden Ergebnissen des Projekts und fördert so den weiteren Fortschritt auf diesem Gebiet. Während das Projekt diese erheblichen Fortschritte erzielt hat, ist es wichtig, die anhaltende Herausforderung zu erkennen, einen definitiven quantenmechanischen Vorteil mit variationalen quantenbasierten Methoden wie QRL auf aktueller NISQ-Hardware zu erreichen. Der überwältigende Einfluss von Hardware-Rauschen und der Aufwand, der mit Fehlerkorrektur- und Minderungstechniken verbunden ist, stellt weiterhin erhebliche Hürden dar. Daher bleibt, während das Projekt erfolgreich wertvolle algorithmische Fortschritte, neuartige Methoden und wesentliche Ressourcen für die QRL- und QML-Communities generiert hat, ungewiss, ob ein klarer quantenmechanischer Vorteil

mit diesen variationalen Ansätzen auf aktuellen NISQ-Geräten realisiert werden kann. Darüber hinaus hat das Projekt auch einen wichtigen Beitrag zur Entwicklung der Arbeitskräfte in Deutschland geleistet, indem es zahlreiche Möglichkeiten für studentische Abschlussarbeiten, Praktika und Doktorarbeiten geschaffen hat. Diese Talentförderung im sich schnell entwickelnden Bereich der Quantencomputing-Technologien stellt einen entscheidenden langfristigen Einfluss des Projekts dar und schafft eine qualifizierte Arbeitskräfte für die Zukunft der Quantentechnologien in Deutschland. Dieses Projekt lieferte unschätzbare Einblicke in die Stärken und Schwächen von VQC-basierten Algorithmen, die Einschränkungen der aktuellen NISQ-Hardware und die allgemeine Machbarkeit der Anwendung quantenmechanischer Lösungen auf reale Probleme. Aufbauend auf diesem erworbenen Wissen beabsichtigt das Team, die Forschungsanstrengungen in den vielversprechenden Bereichen des Post-NISQ-Quantum-Machine-Learning und der Anwendung quantenmechanischer Techniken zur Verbesserung klassischer Machine-Learning-Algorithmen fortzusetzen.

8 Zahlenmäßigen Nachweises

	Budget	2021	2022	2023	2024	Total	Balance
Personalkosten	712.240,00 €	116.664,78 €	247.126,48 €	260.99,33 €	94.602,04 €	719.390,63 €	-7.150,63 €
Reisekosten	13.750,00 €	0,00 €	1.7963,39 €	8.487,43 €	6.038,44 €	16.322,26 €	-2.572,26 €
QC-Zugang	29.310,00 €	9.770,00 €	0,00 €	19.540,00 €	0,00 €	29.310,00 €	0,00 €
Summe	755.300,00 €	126.434,78 €	248.922,87 €	289.024,76 €	100.640,48 €	765.022,89 €	-9.722,89 €

Tabelle 8.1: Budget und Ausgaben

Die Projektmittel wurden wie vorgesehen verwendet, wie in Tabelle 8.1 gezeigt. Die Personal- und Reisekosten überschritten leicht das zugewiesene Budget. Alle anderen projektbezogenen Ausgaben wurden im Rahmen des festgelegten Budgets verwaltet.

Literaturverzeichnis

- [1] F. Arute *et al.*, “Quantum supremacy using a programmable superconducting processor,” *Nature*, vol. 574, no. 7779, pp. 505–510, Oct. 2019,
- [2] C. L. CHEN, D. Y. DONG, and Z. H. CHEN, “QUANTUM COMPUTATION FOR ACTION SELECTION USING REINFORCEMENT LEARNING,” *Int. J. Quantum Inform.*, vol. 04, no. 06, pp. 1071–1083, Dec. 2006,
- [3] D. Dong, C. Chen, H. Li, and T.-J. Tarn, “Quantum Reinforcement Learning,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 5, pp. 1207–1220, Oct. 2008,
- [4] M. Ganger and W. Hu, “Quantum Multiple Q-Learning,” *International Journal of Intelligence Science*, vol. 09, pp. 1–22, Jan. 2019,
- [5] V. Dunjko, J. M. Taylor, and H. J. Briegel, “Framework for learning agents in quantum environments.” 2015.
- [6] V. Dunjko, J. M. Taylor, and H. J. Briegel, “Advances in quantum reinforcement learning,” in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, Oct. 2017, pp. 282–287.
- [7] F. A. Cárdenas-López, L. Lamata, J. C. Retamal, and E. Solano, “Multiqubit and multilevel quantum reinforcement learning with quantum technologies,” *PLOS ONE*, vol. 13, no. 7, p. e0200455, Jul. 2018,
- [8] H. J. Briegel and G. D. las Cuevas, “Projective simulation for artificial intelligence.” 2013.
- [9] A. A. Melnikov, A. Makmal, V. Dunjko, and H. J. Briegel, “Projective simulation with generalization,” *Scientific Reports*, vol. 7, no. 1, Oct. 2017,
- [10] G. D. Paparo, V. Dunjko, A. Makmal, M. A. Martin-Delgado, and H. J. Briegel, “Quantum Speedup for Active Learning Agents,” *Physical Review X*, vol. 4, no. 3, Jul. 2014,
- [11] V. Dunjko, J. M. Taylor, and H. J. Briegel, “Quantum-Enhanced Machine Learning,” *Physical Review Letters*, vol. 117, no. 13, Sep. 2016,
- [12] D. Crawford, A. Levit, N. Ghadermarzy, J. S. Oberoi, and P. Ronagh, “Reinforcement Learning Using Quantum Boltzmann Machines.” 2019.
- [13] T. Sriarunothai *et al.*, “Speeding-up the decision making of a learning agent using an ion trap quantum processor,” *Quantum Science and Technology*, vol. 4, no. 1, p. 015014, Dec. 2018,
- [14] S. Y.-C. Chen, C.-H. H. Yang, J. Qi, P.-Y. Chen, X. Ma, and H.-S. Goan, “Variational Quantum Circuits for Deep Reinforcement Learning.” 2020.
- [15] D. Hein *et al.*, “A benchmark environment motivated by industrial control problems,” in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, Nov. 2017, pp. 1–8.
- [16] S. Jerbi, H. P. Nautrup, L. M. Trenkwalder, H. J. Briegel, and V. Dunjko, “A framework for deep energy-based reinforcement learning with quantum speed-up,” *ArXiv*, vol. abs/1910.12760, 2019,
- [17] Y. Ma and V. Tresp, “Quantum Machine Learning Algorithm for Knowledge Graphs.” 2020.
- [18] S. Khairy, R. Shaydulin, L. Cincio, Y. Alexeev, and P. Balaprakash, “Learning to Optimize Variational

- Quantum Circuits to Solve Combinatorial Problems,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 03, pp. 2367–2375, Apr. 2020,
- [19] “QLinda Software Library — qlinda.vercel.app.”
- [20] M. Franz *et al.*, “Uncovering instabilities in variational-quantum deep Q-networks,” *Journal of the Franklin Institute*, vol. 360, no. 17, pp. 13822–13844, Nov. 2023,
- [21] M. Periyasamy, M. Hölle, M. Wiedmann, D. D. Scherer, A. Plinge, and C. Mutschler, “BCQQ: Batch-Constraint Quantum Q-Learning with Cyclic Data Re-uploading,” in *2024 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jun. 2024, pp. 1–9.
- [22] M. Wiedmann *et al.*, “An Empirical Comparison of Optimizers for Quantum Machine Learning with SPSA-Based Gradients,” in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, IEEE, Sep. 2023, pp. 450–456.
- [23] M. Periyasamy, A. Plinge, C. Mutschler, D. D. Scherer, and W. Mauerer, “Guided-SPSA: Simultaneous Perturbation Stochastic Approximation assisted by the Parameter Shift Rule.” 2024.
- [24] M. Periyasamy, N. Meyer, C. Ufrecht, D. D. Scherer, A. Plinge, and C. Mutschler, “Incremental Data-Uploading for Full-Quantum Classification,” in *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*, IEEE, Sep. 2022, pp. 31–37.
- [25] N. Meyer *et al.*, “Qiskit-Torch-Module: Fast Prototyping of Quantum Neural Networks,” in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, IEEE, Sep. 2024, pp. 817–823.
- [26] N. Meyer *et al.*, “Warm-Start Variational Quantum Policy Iteration,” in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, IEEE, Sep. 2024, pp. 1458–1466.
- [27] N. Meyer, M. Röhn, J. Murauer, A. Plinge, C. Mutschler, and D. D. Scherer, “Comprehensive Library of Variational LSE Solvers,” in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, IEEE, Sep. 2024, pp. 1–4.
- [28] N. Meyer, D. D. Scherer, A. Plinge, C. Mutschler, and M. J. Hartmann, “Quantum Policy Gradient Algorithm with Optimized Action Decoding.” 2023.
- [29] N. Meyer, D. D. Scherer, A. Plinge, C. Mutschler, and M. J. Hartmann, “Quantum Natural Policy Gradients: Towards Sample-Efficient Reinforcement Learning,” in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, IEEE, Sep. 2023, pp. 36–41.
- [30] S. Y.-C. Chen, C.-H. H. Yang, J. Qi, P.-Y. Chen, X. Ma, and H.-S. Goan, “Variational Quantum Circuits for Deep Reinforcement Learning.” 2020.
- [31] O. Lockwood and M. Si, “Reinforcement Learning with Quantum Variational Circuits.” 2020.
- [32] A. Skolik, S. Jerbi, and V. Dunjko, “Quantum agents in the Gym: a variational quantum algorithm for deep Q-learning,” *Quantum*, vol. 6, p. 720, May 2022,

Quantum Reinforcement Learning für industrielle Anwendungen

Kurzbericht



QLiNDA
QUANTUM REINFORCEMENT
LEARNING FÜR INDUSTRIELLE
ANWENDUNGEN

31. März 2025

Vorhabensbezeichnung	Quantum Reinforcement Learning für industrielle Anwendungen (QLiNDA)
Teilvorhaben	Bewertung und Entwicklung von QRL-Algorithmen
FKZ	13N15645
Zuwendungsempfänger	Fraunhofer Institut für Integrierte Schaltungen IIS Nordostpark, 84, 90411 Nürnberg
Kontakt	Dr. Daniel Scherer, daniel.scherer2@iis.fraunhofer.de
Laufzeit des Vorhabens	01.04.2021 bis 30.09.2024 (verlängert um 6 Monate)
Berichtszeitraum	01.01.2021 bis 30.09.2024

Kurzbericht

Ursprüngliche Aufgabenstellung

In dem sich schnell entwickelnden Bereich der industriellen Automatisierung und Steuerung hat sich Reinforcement Learning (RL) als transformative Technologie erwiesen. Industrielle Sektoren suchen kontinuierlich nach fortschrittlichen Steuerungsmethoden, um komplexe Prozesse zu optimieren, die Effizienz zu steigern und sich an dynamische Betriebsumgebungen anzupassen. RL ermöglicht es Systemen, optimale Verhaltensweisen durch Interaktionen mit ihrer Umgebung zu erlernen, was zu Lösungen führt, die traditionelle Steuerungsmethoden möglicherweise nicht aufdecken. Durch kontinuierliches Lernen aus Feedback können sich RL-Algorithmen an komplexe und dynamische industrielle Prozesse anpassen und die Effizienz und Leistung verbessern. In industriellen Kontexten bedeutet dies das Potenzial zur Optimierung vielfältiger Abläufe, von dem Energiemanagement in Kraftwerksanlagen bis hin zur Rationalisierung des Materialflusses in Fertigungspipelines. Die Anwendung von RL verspricht, beispiellose Kontroll- und Betriebsverfeinerungsniveaus in komplexen industriellen Systemen zu erreichen. Die Implementierung von RL in industriellen Umgebungen stellt jedoch erhebliche Herausforderungen dar. RL-Algorithmen sind bekanntlich datenintensiv, d.h. sie benötigen große Datenmengen, um effektiv trainiert zu werden. In industriellen Kontexten kann die Generierung dieser Daten kostspielig, zeitaufwändig und manchmal aufgrund von Sicherheits- und Betriebsbeschränkungen unpraktisch sein. Das Quantum Reinforcement Learning (QRL) hat sich als vielversprechendes Feld zur potentiellen Überwindung dieser Einschränkung ergeben. QRL untersucht die Integration von Methoden des Quantencomputings, um die Fähigkeiten von RL-Algorithmen zu erweitern. Quantencomputer bieten das Potenzial für beschleunigte Berechnungen und verbesserte Informationsverarbeitung bei spezifischen Berechnungsaufgaben. Durch die synergetische Kombination der adaptiven Lernfähigkeiten von RL mit den rechnerischen Vorteilen der Quantenberechnung zielt QRL darauf ab, die Entwicklung effizienterer und datensensitiverer industrieller Steuerungssysteme zu erleichtern, insbesondere in Szenarien, in denen Datenknappheit eine Einschränkung darstellt. Diese Integration eröffnet neue Horizonte für die effektivere Lösung komplexer industrieller Probleme.

Stand der Wissenschaft und Technik

Das Projekt "Quantum Reinforcement Learning für industrielle Anwendungen (QLindA)" hat sich in der Projektlaufzeit darauf konzentriert, Quantencomputing zu nutzen, um RL-Methoden für industrielle Anwendungsfälle zu verbessern. Ziel war es, eine Plattform zu entwickeln, die Quantencomputing mit RL integriert, um die Einschränkungen der Datenintensität in industriellen Anwendungen anzugehen. Trotz des Potenzials des Quantencomputings ergaben sich mehrere bedeutende Herausforderungen, insbesondere mit aktuellen Noisy Intermediate-Scale Quantum (NISQ)-Geräten:

1. **Empfindlichkeit gegenüber Rauschen:** Quantencomputer sind anfällig für Fehler aufgrund von Umgebungsrauschen, was die naive Kodierung großer Datensätze in Quantenschaltkreise ineffizient macht. Ausgeklügelte Datenkodierungsstrategien sind erforderlich, um eine robuste und effektive Datennutzung innerhalb fehleranfälliger Quantencomputer zu gewährleisten.
2. **Unpraktikabilität langer Trainingszeiten:** Die fragile Natur quantenmechanischer Zustände macht lange Trainingszeiten auf NISQ-Hardware unmöglich. Verfahren des Reinforcement Learning erfordern häufig iterative Trainingszyklen. Methodologien, die effizientes Lernen innerhalb der zeitlichen Einschränkungen des NISQ-Hardwarebetriebs ermöglichen, sind für die praktische Anwendung unerlässlich.
3. **Schwierigkeit beim Training expressiver Quantenschaltkreise:** Während Quantenschaltkreise komplexe Funktionen approximieren können, ist es herausfordernd, sie effektiv zu trainieren.

Ablauf des Vorhabens und wesentliche Ergebnisse

Um diese Herausforderungen systematisch anzugehen und umfassende Einblicke in QRL zu gewinnen, wurde ein rigoroses Forschungs- und Entwicklungsprogramm durchgeführt. Diese intensive Untersuchung führte zur erfolgreichen Entwicklung mehrerer neuartiger Methoden, die darauf abzielen, die identifizierten Einschränkungen von QRL innerhalb des NISQ-Paradigmas zu verringern:

1. **Verbessertes Kodierungsverfahren:** Entwicklung einer fortschrittlichen Kodierungstechnik zur effizienten Übersetzung großer klassischer Datensätze in Quantenschaltkreise. Diese Methode minimiert Informationsverluste und maximiert die Effizienz quantenmechanischer Berechnungen, wobei sichergestellt wird, dass kritische Datenmerkmale während der Verarbeitung erhalten bleiben.

2. **Optimierung der Trainingsverfahren:** Identifizierung der effektivsten klassischen Optimierungsmethoden, die die Trainingszeiten auf Quantencomputern verkürzen. Durch die Auswahl von Optimierungsmethoden, die für Quantencomputer geeignet sind, wurde der Rechenaufwand reduziert und die Konvergenzraten verbessert.
3. **Neuartige Gradientenschätzungsmethode:** Entwicklung eines neuen Ansatzes zur Gradientenschätzung, der die Anzahl der auszuführenden Quantenschaltkreise um 15% bis 25% reduziert. Diese Methode strafft den Trainingsprozess und macht ihn praktikabler für NISQ-Geräte, indem die erforderlichen Rechenressourcen verringert werden.

Während diese Methoden einige der mit NISQ-Geräten verbundenen Probleme erheblich verringern, beseitigen sie sie nicht vollständig. Wir erkennen an, dass weitere Fortschritte in der Forschung zu Fehlerkorrekturmethode unerlässlich sind, um diese Herausforderungen vollständig zu überwinden und das volle Potenzial des QRL zu entfalten. Durch die Nutzung dieser methodologischen Fortschritte und basierend auf einer detaillierten Untersuchung sowohl klassischer als auch quantenmechanischer Reinforcement Learning Methoden wurde ein **offline Quanten Reinforcement Learning Algorithmus** erfolgreich entwickelt. Dieser Algorithmus ist darauf ausgelegt, aus vorhandenen Datensätzen zu lernen, was in industriellen Kontexten oft praktikabler ist, da Online-Training kostspielig oder störend für die Abläufe sein kann. Der entwickelte offline QRL-Algorithmus zeigt **erhöhte Expressivität** und **verbesserte Trainierbarkeit** im Vergleich zu klassischen neuronalen Netzen mit einer vergleichbaren Anzahl von trainierbaren Parametern. Dies deutet darauf hin, dass der quantenmechanische Ansatz das Potenzial hat, komplexere Funktionen zu lernen und effiziente Trainingsergebnisse zu erzielen. Simulationsexperimente, die an repräsentativen kleineren Problemen durchgeführt wurden, zeigten, dass variationelle Quantenschaltkreise, die als Funktionsapproximatoren im entwickelten QRL-Algorithmus eingesetzt werden, in der Lage sind, offline RL-Probleme effektiv zu lösen, während die Datenanforderungen im Vergleich zu klassischen neuronalen Netzwerken erheblich reduziert werden. Diese Erkenntnis ist von entscheidender Bedeutung und unterstreicht die dateneffizienten Vorteile, die im QRL inhärent sind, ein Schlüsselattribut für industrielle Anwendungen mit begrenzter Datenverfügbarkeit.

Um die praktische Nützlichkeit des entwickelten Algorithmus empirisch zu validieren, wurde ein QRL-Agent trainiert, um ein **Spielzeugfahrzeug basierend auf Echtzeit-Streckenfeedback** zu steuern. Der Agent lernte, die Strecke effizient zu navigieren, sich an Veränderungen anzupassen und die Leistung zu optimieren. Dieser Demonstrator dient als greifbare Illustration der Anwendbarkeit von QRL auf dynamische Steuerungssysteme, die reale industrielle Prozesskontrollszenerarien widerspiegeln. Obwohl dies einen vereinfachten Anwendungsfall darstellt, liefert diese Demonstration wertvolle empirische Validierung des entwickelten Ansatzes und hebt das Potenzial hervor, auf komplexere industrielle Steuerungsprobleme zu skalieren. Trotz dieser Fortschritte erkennen wir an, dass die Herausforderungen, die durch NISQ-Geräte entstehen, eine Barriere für die breite Anwendung von QRL darstellen. Fortlaufende Forschung zu Fehlerkorrekturmethode und die Entwicklung robusterer Quantenhardware sind notwendig, um diese Hürden vollständig zu überwinden. Unser Projekt trägt zur Grundlage des QRL in industriellen Kontexten bei und bietet einen Einblick in eine Zukunft, in der QRL-Algorithmen Effizienz und Innovation in industriellen Prozessen vorantreiben.

Über praktische Implementierungen hinaus haben wir zur breiteren wissenschaftlichen Gemeinschaft beigetragen, indem wir eine umfassende Literaturübersicht zu bestehenden Methoden des Quantum Reinforcement Learnings erstellt haben. Die entstandene Übersicht („A Survey on Quantum Reinforcement Learning“, arXiv:2211.03464 [quant-ph]) wurde auf dem arXiv Preprint Server veröffentlicht, bietet wertvolle Einblicke und dient als Ressource für Forscher und Praktiker, die das Feld des QRL vorantreiben möchten.