Backend compiler phases for trapped-ion quantum computers

Tobias Schmale*1, Bence Temesi*2, Alakesh Baishya*, Nicolas Pulido-Mateo†‡, Ludwig Krinner†‡, Timko Dubielzig†, Christian Ospelkaus†‡, Hendrik Weimer*, Daniel Borcherding*3

*Institut für Theoretische Physik, Leibniz Universität Hannover, Appelstraße 2, 30167 Hannover, Germany
†Institut für Quantenoptik, Leibniz Universität Hannover, Welfengarten 1, 30167 Hannover, Germany
‡Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany

Abstract—A promising architecture for scaling up quantum computers based on trapped ions are so called Quantum Charged-Coupled Devices (QCCD). These consist of multiple ion traps, each designed for solving specific tasks, that are connected by transport links. In this paper we present the backend compiler phases needed for running quantum circuits on a QCCD architecture, while providing strategies to solve the optimization problems that occur when generating assembly instructions. We implement and test these strategies for the QVLS-Q1 chip architecture.

Index Terms—Quantum Computing, Compiler, Trapped-Ions, Optimization

I. INTRODUCTION

Trapped ions are a promising candidate for realizing universal quantum processors. This is due to their high-fidelity state preparation, readout and universal gate operations as well as their long qubit coherence times [1].

Several architectures have been proposed for scaling up quantum devices based on trapped ions. In the simplest architecture, ions are arranged in a 1D string with a shared potential, where gates are implemented by addressing individual ion states and shared motional modes using lasers and microwaves [1]. However, it was shown that performing high-fidelity two-qubit gates becomes impractical for more than approximately 100 ions [1]. To overcome these challenges, an architecture, called Quantum Charge-Coupled Design (QCCD) was proposed, which consists of multiple ion traps optimized for their specific function as well as transportation links between them [2] [3]. The first implementation of a QCCD architecture can be found in [4]. See [1] for a review of trapped ion architectures.

Similar to classical computers, a compiler is needed to run algorithms on hardware. In this case, the compiler needs to turn quantum circuits into concrete sequences of ion shuttle operations and gate executions. The main challenge hereby is to simultaneously optimize the order of quantum gates and the number of ion shuttles while maintaining an appropriate compilation time.

In the following, we present backend compilation phases for a trapped-ion QCCD architecture and propose strategies to optimize gate executions and ion movements. While these strategies are applicable to general QCCD architectures we demonstrate our approach using a new QCCD architecture (see Fig. 1) that is currently being developed by the Quantum Valley Lower Saxony (QVLS) for the QVLS-Q1 quantum computer.

The code developed for this project is available under [5].

II. CONCEPT

A. Hardware

The QCCD architecture developed for the QVLS-Q1 quantum computer carries at most 50 ions and consists of the following registers connected by an X-junction:

- **SPAM**: State preparation and measurement register with a maximal trap capacity of one ion.
- **STORAGE / TEMPSTORAGE**: Storage registers with a maximal trap capacity of 50 ions each.

![Fig. 1. Simplistic illustration of the QVLS-Q1 chip architecture from a top view showing the STORAGE, TEMPSTORAGE, SPAM and COMPUTE register connected by an X-junction. In the depicted configuration four ions are stored in the STORAGE register and one is located in the COMPUTE register. Movements of ions are only allowed along the transportation links (black lines).](image-url)
**COMPUTE:** Computational register based on [6] with a maximal trap capacity of 2 ions and the possibility to apply the single-qubit gate
\[
R(\phi, \theta) = \cos \phi e^{i \sigma_x \theta / 2} + \sin \phi e^{i \sigma_y \theta / 2},
\]
where we additionally define RX(\theta) \equiv R(0, \theta) and RY(\theta) \equiv R(\pi / 2, \theta). We can also apply the two-qubit gate
\[
RXX(\theta) \equiv e^{i \sigma_x \sigma_z \theta / 2}.
\]
Here,
\[
\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}
\]
are the Pauli matrices and \( \theta \) is some real angle.

In this architecture, only the ions closest to the junction can be moved from one register to another following the transportation links. However, directly exchanging neighboring ions within one register is not allowed. See Fig. 1 for a simplistic illustration of the registers and transportation links of the QVLS-Q1 chip.

Notice also, that in contrast to other QCCD architectures the gates can only be applied to at most two qubits simultaneously in the **COMPUTE** register, while all-to-all qubit connectivity is achieved by physically moving ions between the registers. The behavior of the registers can be best described in terms of multiple stack data structures: each register on the chip can be understood as a stack where the top element is the ion closest to the junction. Moving an ion from one register to another can then refer to popping and pushing an ion from the corresponding stacks.

**B. Assembly Language**

To describe operations on a QCCD architecture, assembly instructions for moving the ions and applying operations on them are needed. For the QVLS-Q1 chip this is realized by our trapped-ion assembly language (TIASM). The main components of the language needed for compiling a quantum circuit to TIASM are given by:

- **QUANTUM_REGISTER**(n): Loads \( n \) ions to the **STORAGE** register.
- **CLASSICAL_REGISTER**(n): Creates a classical register of \( n \) bits for storing the measurement results.
- **MOVE**(r1, r2): Moves the ion closest to the junction of register \( r_1 \) to register \( r_2 \).
- **MEASURE \rightarrow c:** Applies a measurement to the ion in the **SPAM** register and stores the measurement result (0 or 1) in a classical bit with label \( c \).
- **PREPARE**(r): If register \( r = **SPAM** \) a specific laser pulse is applied to the ion in the **SPAM** register to prepare it in a suitable ion state. If \( r = **COMPUTE** \) a microwave pulse is applied in the **COMPUTE** register to prepare the ion in the initial qubit state referred to as \( |0\rangle \).
- **RX(\theta) i_{\text{stack}} \rightarrow \text{RY}(\theta) i_{\text{stack}} \rightarrow R(\phi, \theta) i_{\text{stack}}:** Applies an RX(\theta), RY(\theta) or R(\phi, \theta) gate in the **COMPUTE** register to the ion at location \( i_{\text{stack}} \) within the stack.
- **RXX(\theta):** Applies the RXX(\theta) two-qubit gate to both ions in the **COMPUTE** register.

See [5] for the complete TIASM grammar.

![Fig. 2. Illustration showing the phases of the compiler. 1. A frontend transforms the source code to the intermediate representation (IR). 2. Optimizations dependent on the underlying hardware are applied to the IR. This includes the translation to native gates as well as circuit optimizations based on gate identities. 3. The assembly instructions are generated by orchestrating the gates and the ions using a circuit graph.](image)

**C. Backend compilation phases**

The task of the compiler is to translate a high-level quantum programming language [7] to our assembly language TIASM. As for classical compilers this task can be split into a frontend part, which consists of a lexer, parser and intermediate code generator and a backend part, which optimizes the intermediate code and generates the assembly instructions. See Fig. 2 for an overview of the different compilation phases.

In this work we assume that a frontend generating an intermediate representation (IR) exists and focus on the different phases of the backend part of the compiler. In the following we consider IRs based on OpenQASM [8]. The concepts for the backend compilation part are however also applicable when using different IRs. The different phases are now described in detail:

1) **Translate to native gates:** The native gates that can be applied on the QVLS-Q1 chip and on most trapped-ion chips are RX(\theta), RY(\theta), RXX(\theta) [9]. To convert a quantum circuit consisting of arbitrary quantum gates to the trapped-ion basis gates we are using the existing transpilation methods provided by the Python module called pytket [10].

2) **Circuit optimizations:** After efficiently unrolling an input circuit to our native set of gates, we continue to use the pytket module for gate-level circuit optimizations. Recent benchmarks comparing different transpilation methods [11] showed that pytket gives on average the fastest transpilation times due to its C++ core, while resulting in the lowest number of gates at the same time. pytket repeatedly merges single qubit rotations and uses commutative cancellation until the number of gates is optimized. On top of that, we extend its capabilities with methods unique to our ion-trap machine
A useful circuit identity in our case is given by:
\[
\text{RX}(\theta_1) - \text{RY}(\theta_2) + \text{RX}(\theta_1) - R(\theta_1, \theta_4 - \pi) = 0
\]
where for any \( \theta_1, \theta_2 \in \mathbb{R} \) we can compute the corresponding angles of \( \theta_3, \theta_4 \in \mathbb{R} \), more details in [12]. See a short example in Appendix A. Note however, that this method does not yet help with the reduction of two-qubit gates, which are generally more prone to errors than single qubit gates. While implementing the template above has no disadvantages, some of the rest of the strategies from [12] can introduce tradeoffs between circuit runtime and fidelity errors, which is why we omit these strategies here.

3) Generate circuit graph: After these circuit optimizations, we can proceed with the actual code generation. Due to the sequential nature in which gates need to be executed on this hardware, it makes sense to group gates based on which qubits they act on. To preserve the non-trivial dependencies created by multi-qubit gates, we generate a directed, acyclic "gate dependency" graph (DAG) from the circuit [14]. A valid hardware-realization of the circuit, i.e. circuit path, consists of a traversal of this graph, that visits a node only once all parents have been visited.

4) Graph serialization: In general there are many possible circuit paths that satisfy the gate dependencies. By only following the edges in the DAG we generate a subset of those circuit paths that leads to fewer exchanges of the ions in the compute zone.

Following this strategy we generate up to \( n_p \) circuit paths of length \( n_g \), meaning that each path contains at most \( n_g \) RXX gates. These paths are then simulated using the strategies from the next section and the best one is chosen. The gates from this best path are then deleted from the graph and a next set of paths is generated until all gates have been executed and removed from the graph.

See Appendix B for an example showing the different graph paths that we generate for a given circuit.

5) Ion orchestration: For generating the actual ion movements, we assume that a circuit path, as described in the previous section, is given. Initially, we assign a qubit id to a physical ion once that qubit appears in a generated circuit path for the first time. That way the initial order in which the ions are placed in the storage register is optimal. The task then is to move ions on the chip, such that the correct ions end up in the COMPUTE register in the sequence determined by the path. To solve this problem, we apply the following heuristics:

- Smallest junction distance (JunctionDistance): When moving two ions to the compute zone we start with the ion that is closer to the junction.
- Compute ordering (ComputeOrder): When two ions have the same junction distance we first move the ion to the compute zone that will stay there longer as determined by the next gates in the current circuit path.
- Use SPAM as temporary storage (SpamStorage): Whenever there is a free location in the SPAM area we use it to keep upcoming ions in the graph path closer to the junction.
- Partner sorting (PartnerSorting): From the complete circuit graph we can conclude which ions will need to be paired up in the COMPUTE area in the future. We refer to these qubits as next "partner qubits". By using the SPAM and COMPUTE register we can exploit the information about the next partner qubits to improve the ordering of the ions. Whenever ions are moved from one storage register to another, we try to place these partners closer to each other.

An example of a graph path being executed using the JunctionDistance, SpamStorage and PartnerSorting strategy is shown in Fig. 4 in Appendix C, and Appendix D shows an example that demonstrates the necessity of the ComputeOrder heuristic.

This architecture as described above is motivated by our observation that the tasks of Graph Serialization and Ion Orchestration are intrinsically linked and cannot be optimized independently. Using this iterative architecture allows us to tackle this problem, by considering many small, local optimization problems, each dealing with a subset of the circuit (i.e. short graph paths), rather than trying to find the globally optimal serialized circuit. Furthermore, this design allows for gates to be reordered while at the same time ensuring that the next gate sequence is known during the ion orchestration step, which allows for more efficient move operations.

A simple example showing how an OpenQASM file is compiled using our backend compilation phases is shown in [5].

III. Evaluation

A. Circuit optimization

In order to see how the template performs, we looked at circuits with 5 qubits and a depth of 40, built up from a random sample of standard Qiskit gates consisting of single qubit, two-qubit and three-qubit gates [15]. Firstly, pytket optimizations already reduce the number of gates significantly (see also [11]). Then by implementing the circuit identity, we could see another 8-10% reduction, which contributes to higher overall circuit fidelities.

B. Graph serialization optimizations

The techniques for serializing the circuit graph are analyzed using a dataset of 500 random circuits on 1-50 qubits, with a mean RXX-gate count of 52 and a mean RXX-depth of 20. We compile each circuit and count the total number of movement operations generated. The results are shown in Fig. 5.

To set these results into context, we try to give an estimate for the number of movements generated by a very naive compiler, that randomly picks the next available two-qubit
gate: The above circuits contain on average 25 qubits. A reasonable equilibrium distribution of ions on the chip for a large random circuit would consist of the STORAGE and TEMP STORAGE areas each containing half of the ions, i.e. around 12 ions per storage register. Therefore, the average ion requires just over 6 movements to be transported to the COMPUTE area. An execution of a single RXX gate requires 2 ions to be present in the COMPUTE area, i.e. around 12 movements need to be executed per RXX gate.

If we turn off all heuristics, set \( n_g = 1 \) and randomly choose a circuit path, our compiler best matches the above criteria. With these settings we get \( \approx 10.4 \) movements/RXX gate as seen in Fig. 3 (right, "Random gate"). We do not expect these two values to match exactly as the theoretical expectation is a very crude estimate and we have a limited dataset of finite circuits, however, the above argument gives a rough baseline for the empirical value and can explain the order of magnitude of typical movement counts generated by the compiler.

Next, we point out the fact that merely introducing the circuit path already gives a large reduction in the number of moves. This can be seen by comparing the datapoint "Random gate" with "Random path" (Fig. 3 (right)) where a random path of length \( n_g = 7 \) is selected. This vast improvement can be explained by the better reuse of ions located in the COMPUTE area that purely arises from walking along the graph edges instead of picking random gates from the front of the graph. If one then simulates all graph paths and picks the best one, as for data point "Simulate paths" in Fig. 3, one achieves a further reduction.

The length of the generated paths \( n_g \), also has a large influence on the generated number of movements, as evident from Fig. 3 (left). However, increasing \( n_g \) has the expected consequence of an exponentially increasing compile time, so for practical applications, \( n_g \) should be limited to a fixed, small value. This highlights the tuning capability of the compiler, allowing it to be adjusted to the required task: for real-time quantum computations the circuit needs to be compiled as fast as possible. However, when a circuit can be compiled beforehand, the quality of the assembly instructions might be of more importance than the compile time.

C. Ion orchestration optimizations

The heuristics given above for ion orchestration are also analyzed in Fig. 5 (right). These heuristics allow us to go from 5.34 movements/RXX gate, to 4.81 movements/RXX gate for \( n_g = 7 \), which is significant reduction for circuits with thousands to tens to thousands of RXX gates.

Overall we are thus able to reduce the 10.4 moves/RXX of the naive compiler (Fig. 3 (right, "Random gate")) all the way down to less than 4.7 moves/RXX (Fig. 3 (left, \( n_g = 10 \)), which is a reduction by more than 50%.

IV. RELATED WORK

An extensive study of compilers for different QCCD designs was performed in [16]. This work, however, assumes the ability to swap neighbouring ions within the same trap as well as gate applications in all traps. In [14], optimization techniques based on heuristics were presented for a QCCD with multiple traps arranged in a 1D array. Both works focus on QCCD architectures that are distinctly different from the relevant QVLS-Q1 chip.

The optimization techniques in [14] also first included the usage of a gate dependency graph to optimize the execution order of gates. Building upon this idea we are generating in this work multiple paths through the gate dependency graph to simultaneously optimize the gate execution order and the ion movements.

V. CONCLUSION

In this paper we presented backend compilation phases for trapped-ion QCCD architectures and provided strategies for simultaneously optimizing gate orderings and ion movements during the generation of assembly instructions.

As for classical compilers the backend was split into a phase optimizing the IR and a phase generating the assembly instructions. For the circuit optimization phase we demonstrated that pytket transpilation methods can be extended to reduce the amount of gates by up to 10%.

The assembly instructions are collected by simulating the ion movements corresponding to paths through the circuit graph. To obtain a useful tradeoff between compilation time
and quality of assembly instructions, we split the tasks of serializing the circuit graph and orchestrating the ions into multiple local optimization problems that we solved using simple heuristics. This novel strategy is applicable to trapped-ion QCCD architectures in general. The heuristics, however, are dependent on the specific hardware.

To demonstrate all aspects of the backend compiler phases we developed a trapped-ion assembly language (TIASM) for the QVLS-Q1 chip and implemented our backend compiler phases and specific heuristics for this hardware. That way we were able to reduce the amount of move operations by more than 50% compared to randomly executing gates when using our best heuristics for orchestrating the ions. This result highlights how our backend compiler phases together with well suited heuristics for a specific hardware can contribute to reducing the error that accumulates when applying a quantum circuit on a trapped-ion QCCD architecture.

VI. OUTLOOK

During the graph serialization step we generate circuit paths by starting from the qubits that are accessible with respect to the gate dependency graph. Depending on the current placement of the ions on the chip there might be some obvious paths that are not worth simulating. In future works we will investigate the effect of heuristics during the generation of graph paths in order to reduce the number of paths generated, and reducing the exponential complexity of generating the graph paths.

On top of that, we are planning to improve the rating of the graph paths after their simulation. In this work multiple graph paths are simulated and the graph path that resulted in the least amount of movements is chosen. Instead one could investigate different cost functions that not only take into account the amount of movements, but also other criteria like the expected number of future movements based on the placements of the ions after a graph path was applied.

The QVLS-Q1 chip architecture that was used to demonstrate our backend compiler phases can be seen as a fundamental building block of a larger trapped-ion quantum computer. To cope with multiple connected junctions we will extend the compiler in the future to also optimize the ion movements and gate executions in larger architectures. Here the additional challenge will be to decide which gates should be applied sequentially in one building block and which gates could be applied in parallel using the \text{COMPUTE} zones of all building blocks.

ACKNOWLEDGMENT

We gratefully thank Ina Schaefer for helpful discussions. This work was funded by the Quantum Valley Lower Saxony (QVLS) through the Volkswagen foundation and the ministry for science and culture of Lower Saxony and by Germany’s Excellence Strategy – EXC-2123 QuantumFrontiers – 390837967.

REFERENCES

[1] Colin D. Bruzewicz et al. “Trapped-Ion Quantum Computing: Progress and Challenges” In: Applied Physics Reviews 6.2 (June 2019), p. 021314.
[2] David J Wineland et al. “Experimental issues in coherent quantum-state manipulation of trapped atomic ions”. In: Journal of research of the National Institute of Standards and Technology 103.3 (1998), p. 259.
[3] D. Kielpinski, C. Monroe, and D. J. Wineland. “Architecture for a large-scale ion-trap quantum computer” In: Nature 417.6890 (June 2002), pp. 709–711.
[4] Juan M Pino et al. “Demonstration of the QCCD trapped-ion quantum computer architecture”. In: arXiv preprint arXiv:2003.01293 (2020).
[5] Daniel Borcherding et al. Tiasm Compiler. gitlab.uni-hannover.de/QVLS/TiasmCompiler 2022.
[6] G Zarantonello et al. “Robust and resource-efficient microwave near-field entangling Be+ 9 gate”. In: Physical review letters 123.26 (2019), p. 260503.
[7] Bettina Heim et al. “Quantum programming languages”. In: Nature Reviews Physics 2.12 (2020), pp. 709–722.
[8] Andrew W Cross et al. “Open quantum assembly language”. In: arXiv preprint arXiv:1707.03429 (2017).
[9] Robert McConnell Colin D. Bruzewicz John Chiaverini and Jeremy M. Sage. “Trapped-Ion Quantum Computing: Progress and Challenges” In: Applied Physics Reviews 6.2 (2019).
[10] Seyon Sivarajah et al. “\text{\textbf{loket}}: a retargetable compiler for NISQ devices” In: Quantum Science and Technology 6.1 (2020), p. 014003.
[11] Y Kharkov et al. “Arline Benchmarks: Automated Benchmarking Platform for Quantum Compilers”. In: arXiv preprint arXiv:2202.14025 (2022).
[12] Dmitri Maslov. “Basic circuit compilation techniques for an ion-trap quantum machine” In: New Journal of Physics 19.2 (2017), p. 023035.
[13] Dmitri Maslov et al. “Quantum circuit simplification and level compaction” In: IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 27.3 (2008), pp. 436–444.
[14] Abdullah Ash Saki, Rasit Onur Topaloglu, and Swaroop Ghosh. “Muzzle the Shuttle: Efficient Compilation for Multi-Trap Trapped-Ion Quantum Computers” In: arXiv:2111.07961 [quant-ph] (Nov. 15, 2021).
[15] Md Sajid Anis et al. Qiskit: An Open-source Framework for Quantum Computing 2021.
[16] Prakash Murali et al. “Architecting Noisy Intermediate-Scale Trapped Ion Quantum Computers” In: 2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA). 2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA). Valencia, Spain: IEEE, May 2020, pp. 529–542.
For \( n_g = 3 \) we generate the following 4 paths through the circuit by following the nodes of the dependency graph:

1) A, C, D  2) A, D, C
3) B, A, C  4) B, A, D

Notice that there are four more paths that satisfy the gate dependencies (e.g., A, B, C). However, we only consider paths that follow the nodes of the dependency graph. Executing gate B after gate A would lead to a complete exchange of the ions in the compute zone, and therefore probably to more move operations.

Following the same strategy we can generate 4 paths for \( n_g = 6 \):

1) A, C, D, E, B, F  2) A, D, C, E, B, F
3) B, A, C, D, E, F  4) B, A, D, C, E, F

C. Example ion movements

Fig. 4 shows an example of a graph path being applied by moving the ions on the chip, such that they end up in the compute area in the correct sequence. For moving the ions, some heuristics from Section II-C5 are applied in order to reduce the number of movements.

D. ComputeOrder heuristic example

The following circuit is one of the simplest examples, that highlights a possible optimization regarding ion orchestration.

Assuming ions \( q_0 \) and \( q_1 \) are not in the compute register and are equally close to the junction (as would for example happen at the beginning of a circuit when no qubits’ id’s have yet been assigned to the ions), a decision has to be made regarding the order in which ions \( q_0 \) and \( q_1 \) are put into the compute register. Here it is more efficient, to start with ion 1, as it can remain in compute for longer, and does not have to be moved again for the execution of the two subsequent RXX gates.

**APPENDIX**

A. Circuit optimization example

Let us also show an example for the circuit template that we introduced above. The main advantage of it is that it allows us to replace a particular sequence of the single qubit gates \( \text{RX}(\theta_1), \text{RY}(\theta_2), \text{RX}(\theta_1) \), with just one gate \( \text{R}() \). For instance, if we find the following sequence of gates in a circuit:

\[
\text{RX}(\theta_1) \quad \text{RY}(\theta_2) \quad \text{RXX}(\alpha) \quad \text{RX}(\theta_1)
\]

then to be able to use the identity, we need to move the last RX gate in between the RY and the RXX gates. This is possible since RX and RXX gates commute with each other. Consequently, the circuit can be reduced to:

\[
\text{R}(\theta_3, \theta_4) \quad \text{RXX}(\alpha)
\]

Note, that any general \( \text{R} \) gate can be directly implemented with the ion-trap quantum hardware at hand, thus the number of gates has been truly reduced. An example where we cannot find the right sequence of gates via commutation is when there is an additional RY gate in the circuit after the RXX gate. In this case, the RX gate cannot commute through the RY gate, preventing us from using the identity.

B. Graph path generation example

In the following circuit all gates are two-qubit gates (\( F \) acts on qubits 2 and 4).

For the computation of the graph path (RXX(2, 3), RXX(5, 8), RXX(4, 5)) being applied, starting from a possible mid-circuit configuration just after the first RXX gate of the path has been applied. The arrows indicate the movement operation that needs to be applied to obtain the next diagram. Additionally, ions 3 and 9 are next partners due to an upcoming RXX(3, 9) gate in the circuit graph. Between steps 1) and 2) the SpamStorage heuristic from Section II-C5 is made use of, because ion 4 will be needed for an upcoming gate in the graph path. This prevents the ion to be moved unnecessarily far from the junction. Between steps 3) and 5), the PartnerSort heuristic is made use of, as ions 3 and 9 are next partners and they can be placed next to each other, simply by delaying the removal of ion 3 from the compute area. Note: between 6) and 7), as well as between 7) and 8) two movement operations are applied.