Layout Optimization in Quantum Circuit using Simulated Annealing

Somayyeh Shahbazi, and Naser Mohammadzadeh

Abstract- The placement is one of the main steps of the quantum physical design process. It has prominent effects on the total area and the delay of a circuit. However, this process is an NP-complete problem. Therefore, we need some heuristics to solve the placement problem. Focusing on this issue, in this paper a placement algorithm is proposed based on Simulated Annealing heuristic. Experimental results show that the proposed technique decreases the average area of quantum circuits up to 43% against maximum 3% latency penalty for the attempted benchmarks.

Keywords- Quantum Circuits, Placement, Simulated Annealing (SA), Ion Trap Technology.

I. INTRODUCTION

In layout generation, one of the main challenges of accurately modeling of large-scale quantum architecture is the ability to generate a layout for a quantum application by taking into account the area and the maximum parallelism that can be exploited [1]. Most of the important works [2]-[6] performed on quantum algorithms and its physical platform have focused on the constant layout and a few numbers of works [1][7]-[10] have been done in designing and producing layouts for quantum circuits and improving of them. In most of methods, the layout is generated and improved manually. Although the proposed methods in the primary steps and in an overall point of view seem efficient, but when the scale of the quantum circuit gets bigger, producing and improving of layouts manually seems to be impractical[11].

The circuit model for quantum computation allows us to represent any application as a sequence of logic gates [4],[12] to which we will refer to as operations or instructions interchangeably. Several major differences between quantum logic circuits and their classical counterparts make accurate placement and scheduling of a sequence of quantum operations a challenging task. First, the inability to copy quantum data[13] makes it impossible for the data to be transmitted on a wire or distributed to multiple destinations without error. In addition, the no-copy rule forces multi-qubit operations to bring the participating qubits physically together across some distance by traversing an empty channel as in the ion-trap technology [14][15] or by successive swapping with adjacent qubits as in the Kane silicon-based technology [16],[17]. The second important difference is that operations can occur simultaneously, and any or all physical qubits used in those operations may need to be physically transported. The choice of which qubits to transport affects the time of execution and the reliability of the application. The third major difference is that unlike classical data, quantum data is inherently very unstable. Experimental failure rates of multi-qubit operations for ion-traps are as high as 3% [18].

In this paper, we propose a heuristic for layout optimization of quantum circuits which takes a netlist and a physical layout [10], and optimizes the layout. We use a modified and improved version of dataflow-based layout generation approach proposed by Withney et al. [10].

By studying the physical limits of quantum computing, we can prevent from some problems such as obstruction and congestion between qubits which lead to inordinate increase of execute time and find the best location for placement of gates. Furthermore, qubit movement in channels and gate locations have direct effect on the execution time that is one of the main metrics in quantum circuit design.

The rest of this paper is organized as follows: We describe some basic concepts about quantum computation and ion trap technology in section II, followed by an overview of prior work in the field in Section III. Section IV contains our CAD flow and placement strategy. Experimental results are mentioned in Section V. Finally, Section VI concludes the paper.

II. BASIC CONCEPTS

A. Quantum Computation

Quantum computation and quantum information is the study of the information processing tasks that can be accomplished using quantum mechanical systems. Quantum mechanics is a mathematical framework or set of rules for the construction of physical theories. Quantum computation certainly offers challenges aplenty to physicists, but it is perhaps a little subtle what quantum computation and quantum information offers to physics in the long term. It is built upon an analogous concept, the quantum bit, or qubit for short. The beauty of treating qubits as abstract entities is that it gives us the freedom to construct a general theory of quantum computation and quantum information which does not depend upon a specific system for its realization [19].

Changes occurring to a quantum state can be described
using the language of quantum computation. Analogous to the way a classical computer is built from an electrical circuit containing wires and logic gates, a quantum computer is built from a quantum circuit containing wires and elementary quantum gates to carry around and manipulate the quantum information [19].

B. Ion Trap Technology Abstraction

Ion trap technology [20],[21] was chosen as the underlying technology to study our proposed flow. Trapped ions have shown good potential for scalability [7]. Ion trap technology has been physically realized using universal elements for quantum computation with a clear scalable model [22]. Fig. 1 shows the main elements of quantum computing with ion traps [23].

![Fig. 1 Quantum Computing with Ion Traps](image)

There are two strictly quantum operation types that a low-level circuit is composed of: single- or multi-qubit logical gates and a single qubit measurement. All other operations are classical control operations, which includes communication instructions [1].

![Fig. 2 (a) Physical layout demonstrated for a T-junction (three-way intersection), (b) Abstraction of the circuit in (a), built using the straight channel and three way intersection macro blocks shown in (c) MEMS mirrors placed above the ion traps plane guide the laser beams to gate locations](image)

Fig. 2 shows a possible mapping of a demonstrated layout (Fig. 2a) to macroblock abstractions (Fig. 2b). As Fig. 2c shows, the laser pulses are guided to the gate locations by an array of MEMS mirrors located above the ion trap plane to apply quantum gates [24]. Fig. 3 shows the library defined in [10]. As this Fig. shows, each macroblock consists of a 3x3 structure of trap regions and electrodes with some ports to allow qubit movement between the macroblocks. The black squares are gate locations, which may not be performed at intersections or turns in the ion trap technology. Different orientations of each of these macroblocks can be used in a layout.

Some key characteristics of ion trap technology can be summarized as follows:

- Rectangular channels lined with electrodes make “wires” in ion traps. Atomic ions (qubits) can be suspended above the channel regions and moved ballistically by application of voltages on the channel electrodes [25]. Therefore, a movement control circuitry is required for each wire to handle any qubit communication.
- Any operation available in the ion trap technology can be performed at each gate location. This makes it possible to reuse gate locations for different operations within a quantum circuit.
- Fabrication and control of ion traps in the third dimension is difficult. Thus, scalable ion trap systems are two dimensional [17]. Therefore, routing channels should have T-junction(s) or cross-junction(s) to allow ions to move from one channel to another.
- Aside from Manhattan distance between the source and target locations for an ion movement, the geometry of the wire channel is also important in the calculation of movement latency. Experiments have shown that a right angle turn takes substantially longer than a straight channel over the same distance [25].

![Fig. 3 Library of basic macroblocks used in this paper. Ports (P0–P3) and electrodes of each macroblock make it possible for the ion to be moved and trapped. Some macroblocks contain a trap region where gates may be performed (black square)](image)

C. Simulated annealing

Simulated annealing is a method of local search algorithm. The process of annealing can be simulated with the Metropolis algorithm, which is based on Monte Carlo techniques. This technique can be used to generate a solution to combinatorial optimization problems assuming an analogy between them and physical many-particle systems with two equivalences: solutions in the problem are equivalent to states in a physical system, and the cost of a solution is equivalent to the “energy” of a state [26].

The search is started with a randomized state. In a polling loop we will move to neighboring states always accepting the moves that decrease the energy while only accepting bad moves accordingly to an acceptance probability can be given by $e^{-\frac{\Delta E}{T}}$ dependent on the “temperature” of the system (T).
Decrease the temperature slowly, accepting less bad moves at each temperature level until at very low temperatures. In this way, the algorithm converges to an optimal or near optimal configuration. The outline of the simulated annealing algorithm is shown in Fig. 4. The parameters and functions used in a simulated annealing algorithm determine the quality of the placement produced. These parameters and functions include the cooling schedule consisting of initial temperature (init-temp), final temperature (final-temp) and the function used for changing the temperature (SCHEDULE), inner-loop-criterion which is the number of trials at each temperature, the process used for shuffling a configuration (PERTURB), acceptance probability (F), and the cost function (COST). A good choice of these parameters and functions can result in a good placement in a relatively short time [27].

Algorithm SIMULATED-ANNEALING

Begin

Temp = INIT-TEMP;
Place = INIT-PLACEMENT;
While (temp > FINAL-TEMP) do
    While (inner-loop-criterion = FALSE) do
        New-place = PERTURB (place);
        ∆C = COST (new-place) – COST (place);
        If (∆C < 0) then
            Place = new-place;
        Else if (RANDOM (0, 1) < \frac{-∆C}{temp}) then
            Place = new-place;
        Temp = SCHEDULE (temp);
    End.
End.

Fig. 4 The Simulated annealing algorithm[27]

D. Maze Routing Algorithm

Maze routing algorithms are used to find a path between a pair of points, called the source and the target respectively, in a planar rectangular grid graph. The objective of a maze routing algorithm is to find a path between the source and the target vertex without using any blocked vertex. The process of finding a path begins with the exploration phase, in which several paths start at the source, and are expanded until one of them reaches the target. Once the target is reached, the vertices need to be retraced to the source to identify the path [28].

III. RELATED WORK

We are not the first to build a computer-aided design flow for quantum circuits. Svore et al. [2],[3] proposed a design flow that starts with a quantum program and generates its corresponding physical operations. Their work outlined various file formats and provided initial implementations of some of the necessary tools. Their design flow which has four phases, converts a high-level program specified in the mathematical abstractions of quantum mechanics and linear algebra into a low-level set of machine instructions scheduled on a fixed H-tree-based layout [2]. Balensiefer et al. [4],[5] proposed a design flow which takes a quantum description in QCL (Quantum Computation Language) [29] and generates a technology-dependent netlist. QCL that is defined by Omer [29] utilizes a syntax derived from C and provides a quantum simulator for code development and testing on a classical computing platform. In the physical design phase, the generated netlist is scheduled on a fixed layout by a list-scheduling algorithm [20]. No optimization is done on the laid out circuit. Metodi et al. [6] proposed a uniform Quantum Logic Array architecture, and extended and improved it later in [7]. The focus of the work was on the architectural research and the details of physical layout or scheduling were not explored. Whitney et al. [10] evaluate circuits by using the tools which allow us to lay out circuits. They proposed architectures for quantum computers have all consisted of computation regions connected by an interconnection network using quantum teleportation [18]. They also suggested [10] a quantum design flow that takes a description and generates its layout in ion trap technology. They proposed new heuristics for the layout generation and scheduling. Their physical design stage includes laying out and scheduling a fixed netlist.

Dousti et al. [30] focus on minimizing the total latency of the circuit to minimize the error in the circuit. A CAD tool, called Quantum mapper based on Scheduling, Placement, and Routing or QSPR, was developed to perform this task automatically. More precisely, the destination qubit is fixed in one trap while the source qubit is moved to reach the destination. Quantum physical operations scheduler (QPOS) distinguishes between the source and destination operands of a two-qubit instruction during the routing step [12]. QPOS extracts a routing path for each of the ready-to-issue instructions. If there are any overlaps among these paths, QPOS selects an instruction to execute based on the following criteria: 1) highest initial priority, 2) lowest among of congestion that is going to be introduced by using the path, and 3) shortest path length. Finally, QPOS maps these paths to the quantum circuit fabric and uses a deadlock prevention algorithm to prohibit qubits to locate in a position that further movement is impossible.

In earlier works, Mohammadzadeh et al. [31] introduced the physical synthesis concept for a quantum design flow and proposed a technique for it that exchanges the gates in the design after layout generation to improve the latency of quantum circuit execution. They also proposed[32] a new technique for physical synthesis using auxiliary qubit selection to improve the latency of quantum circuits. Recently, this group proposed an optimization technique using gate location changing to improve the latency of quantum circuits [33].

The main difference between our methodology and most another related works is using variable layout in quantum design. In most before works, a quantum design flow performs the synthesis and physical design processes separately. They take a fixed netlist and lay it out on a fix layout. But we consider a new placement that creates an optimum layout with suitable area in acceptable perform time for larger circuits.
with more qubits. The goal of the placement is to minimize the total area.

IV. OUR PLACEMENT STRATEGY

A vastly simplified version of typical classical circuit CAD flow is shown in Fig. 5. A user-specified application circuit specification into some sort of gate network is first synthesized then physical components are geometrically mapped to a substrate to make physical design. Verification steps ensure equivalence between stages. The purpose of the flow is to take in some sort of abstract circuit specification and produce a physical design that can be fabricated then [23].

![Fig. 5 Simplified view of a classical computer-aided design flow](http://dx.doi.org/10.15242/IIE.E0214516)

We proposed our placement flow that presented in Fig. 6. This flow has two main steps. During the first stage, an initial placement is generated based on a modified version of dataflow graph based algorithm [10]. The second stage optimizes the initial layout to reach the final improved layout. In the first stage, the algorithm provides an initial layout for the second stage. At the top, it begins with a netlist of the desired quantum circuit. Then, a dataflow graph is created that is equivalent to the instruction sequence. Each node represents an instruction, as labeled and each arc represents a qubit dependency. To increase the parallelism, the algorithm prioritized gate location. It means the instructions that have not any common qubits can have the same priority. Then, it prepares an initial placement of qubits and gate locations. Layout is constructed by going through dataflow graph. After that, long columns are folded into short columns to get a more compact layout, where every column is nearly the same height. Also, two vertical global routing channels are inserted between columns. The columns are then sorted to position gate locations that need to be connected roughly horizontal to one another. Thereafter, the algorithm inserts channel rout based on data flow graph between gate locations. It eliminates a weight to each edge of data flow. It is based on latency of qubit movement on that edge. The algorithm selects the longest edge on the longest critical path and merges these two node groups to eliminate this latency, in effect specifying that these two instructions should occur at the same gate location. Finally, it reroutes the paths and channels and the schedule again. Until create a triple gate in a location or the scheduler give us a time more than the previous layout, it does this work[10].

The generated layout is initial layout for our proposed methodology and is able to be fed into the SA algorithm stage. The SA algorithm includes four stages: 1) creating a new solution to find a close feasible solution (Perturb), 2) cost calculation for new solution, 3) analysis acceptance for new solution, 4) repeats management and periods changing (cooling rate).

![Fig. 6 The general idea](http://dx.doi.org/10.15242/IIE.E0214516)

A. Perturb Procedure

For perturb, we propose merge gate locations method which is described below. Merge method is the best way to shrink circuit, but during it the latency of implementations should be considered. During each perturb call, merge process is done to determine the pair of gate locations for eliminating row or column that source gate locations are in it. Merge is based on gates dependency and consist of three parts: merge maximum three gates in a location, calculation the latency distance and arrival time of qubits of merging gates to location. If in merging process, we face with the gate location that include three gates in a row or column, that row or column will be set aside for elimination. Calculating the distance is done with the assumption that the channel is empty. For example, according to Fig. 7, if gate number 18 is source gate and gates number 22, 19, 17 and 21 are input or output gates for it, the gate number 18 can be merged with one of them. According to
Table 1, we determine the distance of gate number 18 to its depended gates and measure total distances. The lowest total is the highest priority for merging (Sum1).

![Diagram of gate connections](image)

Table I. Calculating the distance of gate number 18 to its depended gates

<table>
<thead>
<tr>
<th>Circuit Number</th>
<th>Area (Microblocks)</th>
<th>Area after SA Optimization</th>
<th>Area Improvement (%)</th>
<th>Latency (μs)</th>
<th>Latency after SA Optimization (μs)</th>
<th>Latency Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[5.1.3]</td>
<td>70</td>
<td>35</td>
<td>50</td>
<td>216</td>
<td>218</td>
<td>-0.93</td>
</tr>
<tr>
<td>[7.1.3]</td>
<td>70</td>
<td>35</td>
<td>50</td>
<td>206</td>
<td>192</td>
<td>6.8</td>
</tr>
<tr>
<td>[5.1.3]</td>
<td>143</td>
<td>91</td>
<td>36.36</td>
<td>298</td>
<td>285</td>
<td>4.36</td>
</tr>
<tr>
<td>[11.1.3]</td>
<td>208</td>
<td>117</td>
<td>43.75</td>
<td>515</td>
<td>481</td>
<td>6.6</td>
</tr>
<tr>
<td>[19.1.7]</td>
<td>651</td>
<td>420</td>
<td>35.48</td>
<td>1170</td>
<td>1044</td>
<td>10.77</td>
</tr>
<tr>
<td>[20.1.6]</td>
<td>782</td>
<td>646</td>
<td>17.39</td>
<td>1132</td>
<td>1120</td>
<td>1.06</td>
</tr>
<tr>
<td>[28.2.8]</td>
<td>1426</td>
<td>1150</td>
<td>19.35</td>
<td>1754</td>
<td>1697</td>
<td>3.25</td>
</tr>
<tr>
<td>[29.1.11]</td>
<td>1617</td>
<td>1334</td>
<td>17.8</td>
<td>1735</td>
<td>1800</td>
<td>-3.75</td>
</tr>
<tr>
<td>[32.2.8]</td>
<td>1518</td>
<td>1242</td>
<td>18.18</td>
<td>1750</td>
<td>1724</td>
<td>1.49</td>
</tr>
<tr>
<td>[35.1.10]</td>
<td>2262</td>
<td>1595</td>
<td>29.49</td>
<td>4612</td>
<td>4082</td>
<td>11.49</td>
</tr>
<tr>
<td>[40.3.10]</td>
<td>2752</td>
<td>1716</td>
<td>37.65</td>
<td>6039</td>
<td>5314</td>
<td>12.01</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C. Acceptance Analysis

After calculating ΔC, if the result is a negative number, we accept the new placement and replaces with the current layout. If ΔC is positive and Random (0,1) < e^{-\frac{\Delta C}{\lambda}} , then new layout is accepted and replaced. Otherwise, new placement doesn’t accept.

D. Cooling Rate

Final step in SA is to manage repetition and change periods flow. SA runs appropriate number of performed iterations during each period. These iterations are related to number of rows and columns in extant layout. Cooling rate consists of initial temperature, temperature steps and number of iteration at each temperature. The strategy of temperature reduction is calculated below:

\[ T_{new} = T_{old} - 1 \]  

V. Experimental Results

A. Experimental Setup

The proposed placement strategy is implemented in C#. The parameters used for the simulation are as follow: INIT-TEMP= initial layout size; INIT-PLACEMENT= upshot layout based on assumptions in [10]; FINAL-TEMP= 0; inner-loop= number of rows and columns in current layout; PERTURB (place) = described merge procedure; SCHEDULE (T) = T – 1. An Apple MacBook Pro laptop with Intel core 2 Duo CPU and 4 GB RAM was used to run the simulation.

B. Simulation Results

Table II and Fig. 8 and 9 show results of simulation for our proposed optimization algorithm with simulated annealing.
This paper includes a new technique for placement in quantum circuits. This style of placement is due to merge of gates location and repeat. In this way, we confront some problems such as, manner of leaving qubits from gate location, movement qubits in channel, blockage and congestion. We provided a solution for each of these problems.

When the goal is circuit run time, a skillful quantum physical designer can apply the special techniques to control entering and leaving qubits, prevent the achievement qubits that generate traffic in channel and calculate distances for shrinking the circuit to the best way and the least delay. In addition to mentioned problems, there are some physical limitations such as being impossible to merge more than three gates in a gate location or increase latency penalty in complex circuits toward reduction area. In general, simulated annealing placement outperforms the other placements regarding final results.

VI. CONCLUSION

This paper includes a new technique for placement in quantum circuits. This style of placement is due to merge of gates location and repeat. In this way, we confront some problems such as, manner of leaving qubits from gate location, movement qubits in channel, blockage and congestion. We provided a solution for each of these problems.

When the goal is circuit run time, a skillful quantum physical designer can apply the special techniques to control entering and leaving qubits, prevent the achievement qubits that generate traffic in channel and calculate distances for shrinking the circuit to the best way and the least delay. In addition to mentioned problems, there are some physical limitations such as being impossible to merge more than three gates in a gate location or increase latency penalty in complex circuits toward reduction area. In general, simulated annealing placement outperforms the other placements regarding final results.

REFERENCES


[34] http://iaks-www.ira.uka.de/home/grassl/QECC/circuits/index.html, Copyright © Markus Grassl