A space-efficient quantum computer simulator suitable for high-speed FPGA implementation

Michael P. Frank\textsuperscript{a}, Liviu Oniciuc\textsuperscript{a}, Uwe H. Meyer-Baese\textsuperscript{*a}, Irinel Chiorescu\textsuperscript{b}

\textsuperscript{a}FAMU-FSU College of Engineering, 2525 Pottsdamer St., Tallahassee, FL, USA 32310-6046;
\textsuperscript{b}National High Magnetic Field Lab., 1800 E. Paul Dirac Dr., Tallahassee, FL, USA 32310-3706

ABSTRACT

Conventional vector-based simulators for quantum computers are quite limited in the size of the quantum circuits they can model, due to the worst-case exponential growth of even sparse representations of the full quantum state vector as a function of the number of quantum operations applied. However, this exponential-space requirement can be avoided by using general space-time tradeoffs long known to complexity theorists, which can be appropriately optimized for this particular problem in a way that also illustrates some interesting reformulations of quantum mechanics. In this paper, we describe the design and empirical measurements of a working software prototype of a quantum computer simulator that avoids excessive space requirements. Due to its space-efficiency, this design is well-suited to embedding in single-chip environments, permitting especially fast execution that avoids access latencies to main memory. We plan to prototype our design on a standard FPGA development board.

Keywords: Quantum computing, simulation, special-purpose architectures, FPGAs, embedded design

1. INTRODUCTION

Since a scalable, widely accessible quantum computer has not yet been built, it is important to be able to demonstrate the theoretical operation of quantum computers using simulators based on existing classical computing hardware. Such tools are useful for the validation and testing of new quantum algorithms in research settings, as well as for the education of students as well as more experienced scholars who may be new to the emerging field of quantum computing.

Unfortunately, most or all of the existing widely-available quantum computer simulators are severely limited in the size of the quantum circuits that they can simulate. This is due to the fact that traditional simulators operate by updating an explicit representation of the quantum state vector of the simulated quantum circuit. In the worst case (which is also the case that is typically encountered in practice, in most of the interesting quantum algorithms), the number of nonzero elements of the state vector increases exponentially with the number of operations (gates) that are dynamically applied in the quantum circuit, and therefore increases exponentially with the size of the problem to be solved. This means that, even when a sparse representation of the state vector is used, the finite available memory on any given platform imposes a rather strict limit on the size of the quantum circuits that can be feasibly simulated.

For example, if a given machine has 8 GB of main memory, then it might only be able to simulate general quantum circuits containing 30 or fewer nontrivial gates, since representing the resulting quantum state would typically require storing $\2^{30} = 1$G eight-byte floating-point complex numbers. Furthermore, since accessing main memory (as opposed to on-chip caches) is relatively slow, the large amount of memory required for simulating even circuits of sizes within this limit can still impair the simulator’s performance.

It would be desirable to have a simulator whose capabilities were not so strictly limited by the available memory, so that the simulator can be implemented on a fast single-chip hardware platform, and so it can also simulate circuits of sizes beyond the limits of traditional simulators, with more graceful performance degradation.

Fortunately, computational complexity theorists have long been aware that there is a general algorithmic transformation that can be applied to reduce the space requirements of algorithms. The basic concept is simply to recalculate data values dynamically when needed, rather than storing them explicitly. As long as an algorithm’s dataflow graph is not as deep as it is wide, using this approach can reduce the algorithm’s space complexity.

\textsuperscript{*}umb@eng.fsu.edu; phone 1 850 410-6220; fax 1 850 410-6479; www.eng.fsu.edu
This condition applies to the simulation of quantum computers. Each element of the quantum state vector at a given step of a quantum algorithm typically depends only on the values of 1 or 2 elements of the state vector at the preceding step. These values can be recomputed on demand from the values at the next preceding step, and so forth, in recursive fashion. The recursion back through the entire history of \( N \) previously applied quantum operations requires only \( O(N) \) space on a stack, to keep track of which path through the dataflow graph is currently being traversed.

Possibly the first person to realize that this general kind of procedure could be applied to the calculation of quantum-mechanical amplitudes was the famous physicist Richard Feynman, who in his dissertation work\(^1\) showed how quantum mechanics could be reformulated in terms of a quantity he called the path integral, which essentially amounted to a continuous analogue of a sum over paths through a discrete dataflow graph.

When the complexity theory of quantum computing was being developed in the early 1990s, it was quickly realized\(^2\) that the same idea, back now in the discrete realm, applied to the simulation of quantum computers as well, leading to the important complexity-theoretic relation that \( \text{BQP} \subseteq \text{PSPACE} \), where \( \text{BQP} \) is the set of problems solvable by probabilistic quantum algorithms with a polynomial number of operations (as a function of problem size), and \( \text{PSPACE} \) is the set of problems solvable by classical computers using a polynomial amount of memory. More generally, we can show\(^3\) that a quantum algorithm with \( s \) qubits and \( t \) operations can be simulated using space \( O(s + t) \).

Although this essential insight has been known for at least 16 years now, to our knowledge it has not yet been applied to develop a flexible and widely-available tool for simulating quantum algorithms in such a way that the available memory is not a significant limiting factor on the size of the quantum computations that can be simulated. It is the goal of the SEQCsim (say “SEEK-sim”) project at Florida State University to remedy this situation by providing flexible, well-optimized freely-available software and hardware implementations of a Space-Efficient Quantum Computer Simulator.

To date, we have developed a working software prototype of our simulator in C++, and have empirically demonstrated its correctness and space-efficiency on a variety of simple test cases. We present some of these results in sections 2-4. Next steps include the development of a more powerful programming environment for the software version of the simulator, as well as a performance-optimized special-purpose hardware implementation of the simulator, to be prototyped using a standard FPGA (field-programmable gate array) platform, which we will describe in section 5. Section 6 concludes.

### 2. SEQCsim Algorithm

The presently-available software prototype of our simulator (version 0.8) operates according to an extremely simple procedure.

The simulator first obtains a definition of the quantum circuit to simulate from a set of four ASCII text input files, called \( \text{qconfig.txt} \), \( \text{qinput.txt} \), \( \text{qoperators.txt} \), and \( \text{qopseq.txt} \), examples of which are shown in listings 2-5. These files are structured in a simple throwaway file format, which will be replaced in later versions of our simulator by a general-purpose quantum programming language based on C++. The configuration file, given in \( \text{qconfig.txt} \), specifies the width of the quantum circuit, and assigns various named registers to bit-fields within it. The quantum algorithm, specified explicitly as a gate sequence in \( \text{qopseq.txt} \), may use any fixed-width quantum gates, whose matrix elements are given in \( \text{qoperators.txt} \). The initial input state (which must be a classical state in the computational basis) is given in \( \text{qinput.txt} \).

The goal of the simulator is to progress forwards through the quantum algorithm (circuit) one operation (gate) at a time, while keeping track of the amplitude of only a single basis state, in the classical computational basis, which is selected randomly at each step in accordance with the flow of probability mass in the quantum algorithm, such that the simulator’s probability of ending up at each final basis state precisely matches what would be obtained from a complete calculation of the final quantum state vector.

This approach evokes an old interpretation of quantum mechanics by Bohm\(^4\-5\), who showed that a quantum system can be conceived of as having a unique classical state at each time which evolves (either deterministically or nondeterministically) in accordance with the probability current through the system’s phase space that is induced by the Schrödinger time-evolution. In this model, a complete wavefunction still exists mathematically, but it is conceived of as just being a “pilot” wave that guides the evolution of the physical state, rather than being thought of as being the actual physical state itself. In accordance with the subordinate status of the wavefunction in Bohm’s philosophy, rather than storing the entire wavefunction, and conceiving of it as being the simulated state, we only calculate values of the wavefunction at points...
that are needed to compute the transition probabilities along the specific possible trajectory through the classical configuration space that is presently being explored.

The core algorithms for updating the stochastically-evolved basis state (procedure “run()”) and calculating wavefunction amplitudes for specific basis states (function “calcAmp()”) are outlined below in Listing 1.

For a more detailed illustration of the functioning of this algorithm on a particularly simple example circuit, please refer to our previous paper\cite{frank2009}.

Listing 1. Outline of the core algorithm used in the present version of SEQCSim. The selection of a particular operation (gate and operand bits) on lines 5 and 18 determines a set of possible “neighbor” or “predecessor” basis states of the current one, differing from the present state on the operand bits.

```
1 procedure SEQCSim::run():
2   curState := inputState; // Current basis state, in the computational basis
3   curAmp := 1; // Amplitude of current basis state
4   for PC := 0 to #gates, // Index of current operation in the gate sequence
5      with respect to the operator gate[PC] and its operands,
6         for each neighbor nbr, of curState,
7            if nbr = curState, amp[nbr] := curAmp;
8            else amp[nbr] := calcAmp(nbr);
9            amp[] := opMatrix * amp[]; // Complex matrix product
10           prob[] := normSqr(amp[]); // Calc probs as normalized squares of amplitudes.
11           i := pickFromDist(prob[]); // Pick a random successor of the current state.
12           curState := nbr; // Go to that neighbor.
13           curAmp := amp[nbr]. // Remember its amplitude, calculated earlier.
14
15 function SEQCSim::calcAmp(State nbr): // Recursive amplitude-calculation procedure
16   curState := nbr;
17   if PC=0, return (curState = inputState) ? 1 : 0; // At t=0, input state has all the amplitude.
18   else, with respect to the operator gate[PC−1] and its operands,
19      for each predecessor pred. of curState,
20         PC := PC − 1;
21         amp[pred] = calcAmp(pred); // Recursive calculation of pred. amp.
22         PC := PC + 1;
23         amp[] := opMatrix * amp[];
24   return amp[curState];
```

3. EXAMPLE QUANTUM CIRCUIT USED IN TESTING

For purposes of testing the correctness and performance of our algorithm, we focused on a simple family of in-place binary adder circuits based on an algorithm by Draper\cite{draper2001}. These adders use a Quantum Fourier Transform (QFT) and its inverse to convert one of the addends into and out of a phase representation, and uses phase gates between addends to carry out the addition in the phase representation. These adders are not particularly efficient (since they require order \( n^2 \) gates for an \( n \)-bit add) but they require no ancilla bits and provide a good test case which includes both trivial and nontrivial gates. For our purposes, a “trivial” gate means a gate like the phase gate whose unitary matrix is diagonalizable in the computational basis; with such gates, each basis state has only one possible predecessor, so these gates do not have a significant impact on the time complexity of the simulation – in our approach, a sequence of trivial gates can be simulated in linear time.

An example of the adders used is shown in fig. 1 below. This illustration was prepared using the freely-available QCAD design/simulation tool, version 1.96, available from http://apollon.cc.u-tokyo.ac.jp/~watanabe/qcad/index.html. In this figure, \( H \) represents the Hadamard gate \( H = (\sigma_x + \sigma_z)2^{-1/2} = [1, 1; 1, -1]/2^{1/2} \), or in displayed form,
\[ H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \]  

(1)

A number \( k \) in a box represents a controlled-phase gate \( \varphi_k \) for a phase rotation of \( k \) degrees. In terms of the rank-2 identity operator \( \hat{I} = [1, 0; 0, 1] \), number operator \( \hat{n} = [0, 0; 0, 1] \), its complement \( \hat{n} = \hat{I} - \hat{n} \), and tensor product \( \otimes \), or more explicitly as a displayed matrix, this operator can be written

\[ \varphi_k = \hat{n} \otimes \hat{I} + \hat{n} \otimes \exp(i k \pi / 180) \]

(2)

The phase gate rotates the phase of a given basis state by the specified number of degrees if and only if both of the input bits are 1. It is symmetrical with respect to the control and target bits. It is a trivial gate – note its matrix is diagonal.

The above example circuit can be easily prepared for input into the SEQCsim simulator by describing it in a simple text input format in the four files \{qconfig, qinput, qoperators, qopseq\}.txt, as illustrated in listings 2-5 below. The precise format of these files has some limited flexibility – keywords may be abbreviated, whitespace is ignored, and lines beginning with “comment:” are ignored. The format specifier on the first line allows for future extensions of the file format while retaining backwards-compatibility with older input files.

Listing 2. Contents of the ASCII text input file qconfig.txt, which is used to tell SEQCsim the size and registers of the input circuit, for the circuit shown in fig. 1. (Line numbers shown at the left are not included in the file.)

```
1 qconfig.txt format version 1
2 bits: 8
3 named bitarray: a[4] @ 0
4 named bitarray: b[4] @ 4
```

Listing 3. Contents of the ASCII text input file qinput.txt used to tell SEQCsim the decimal values of the input registers for the circuit shown in fig. 1. (Line numbers shown at the left are not included in the file.)

```
1 qinput.txt format version 1
2 a = 1
3 b = 1
```
Listing 4. Contents of the ASCII text input file `qoperators.txt` used to tell SEQCsim the definitions of the quantum operators (gates) used in the circuit shown in fig. 1. (Line numbers at the left are not included in the file.)

```
1 qoperators.txt format version 1
2 operators: 8
3 operator #: 0
4 name: H
5 size: 1 bits
6 matrix:
7 (0.7071067812 + i*0) (0.7071067812 + i*0)
8 (0.7071067812 + i*0) (-0.7071067812 + i*0)
9 operator #: 1
10 name: cPiOver2
11 size: 2 bits
12 matrix:
13 (1 + i*0) (0 + i*0) (0 + i*0) (0 + i*0)
14 (0 + i*0) (1 + i*0) (0 + i*0) (0 + i*0)
15 (0 + i*0) (0 + i*0) (1 + i*0) (0 + i*0)
16 (0 + i*0) (0 + i*0) (0 + i*0) (0 + i*1)
```

... (six additional operators elided for brevity)...

Listing 5. Contents of the ASCII text input file `qopseq.txt` used to tell SEQCsim the sequence of quantum operations (gate instances) used in the circuit shown in fig. 1. (Line numbers shown at the left are not included in the file.)

```
1 qopseq.txt format version 1
2 operations: 30
3 operation #0: apply unary operator H to bits a[3]
4 operation #1: apply binary operator cPiOver2 to bits a[3], a[2]
5 operation #2: apply binary operator cPiOver4 to bits a[3], a[1]
6 operation #3: apply binary operator cPiOver8 to bits a[3], a[0]
```

.. (22 additional gate operations elided for brevity)...

Listing 6. Text output from SEQCsim when run with the above text files (listings 2-5) as input. (Line numbers shown at the left are not included in the output.) Note that at the conclusion of the computation, the value of register `a` (least significant 4 bits) is 0010₂ = 2, verifying that the simulator has correctly determined that 1 + 1 = 2.

```
2 Welcome to SEQCsim, the Space-Efficient Quantum Computer SIMulator.
3 (C++ console version)
4 By Michael P. Frank, Uwe Meyer-Baese, Irinel Chiorescu, and Liviu Oniciuc.
5 Copyright (C) 2008-2009 Florida State University Board of Trustees.
6 All rights reserved.
7 ..(2 blank lines)...
```

```
9 SEQCsim::run(): Initial state is 7->00010001<-0 (8 bits) ==> (1 + i*0).
10 SEQCsim::Bohm_step_forwards(): (tPC=0)
11   The new current state is 7->00011001<-0 (8 bits) ==> (0.707107 + i*0).
12 SEQCsim::Bohm_step_forwards(): (tPC=1)
13   The new current state is 7->00011010<-0 (8 bits) ==> (0.707107 + i*0).
```

... (26 intermediate steps elided for brevity)...

```
66 SEQCsim::Bohm_step_forwards(): (tPC=28)
67   The new current state is 7->00010100<-0 (8 bits) ==> (0.707107 + i*0).
68 SEQCsim::Bohm_step_forwards(): (tPC=29)
69   The new current state is 7->00010010<-0 (8 bits) ==> (1 + i*0).
70 SEQCsim::done(): The PC value 30 is >= the number of operations 30.
71   We are done!
```
4. EMPIRICAL MEASUREMENTS OF SPACE/TIME COMPLEXITY

To show that the space and time complexity of SEQCSim indeed respond in the predicted manner to changes in the size of the simulated circuit, we used the following procedure. One of us (Oniciuc) wrote a simple tool to automatically generate the required `qconfig.txt` and `qopseq.txt` input files for SEQCSim, as well as corresponding `.qcd` circuit files for QCAD, for Draper adders of any desired number and size of operands. Using this tool, we generated adders of size 2×2 (2 addends, 2 bits each) up through size 2×14 (2 addends, 14 bits each), and ran QCAD and SEQCSim on each one, on a typical Dell desktop running Windows Vista, while measuring each application’s peak memory usage and CPU time using the shareware Kiwi application monitor. The results were tabulated and used to generate the charts below.

Fig. 1 shows how that overall memory usage of QCAD (which presumably internally uses a traditional state-vector based simulation technique) increases exponentially with the circuit size, whereas the memory usage of SEQCSim remains essentially flat – it is dominated by the space required for the standard C++ libraries which we use to provide functions such as text I/O and pseudo-random-number generation.

![QCAD vs. SEQCSim memory usage](chart)

**Fig. 2.** Overall peak memory usage, in kilobytes, of QCAD versus SEQCSim, for Draper adder circuits of width 4 (2×2) through 28 (2×14), as measured using the Kiwi application monitor. Figures include pages allocated for shared DLLs, but a comparison of private working set sizes, as measured by the Windows Task Manager in Vista, gives similar results. QCAD’s higher base memory usage is unsurprising since it requires more libraries to support its GUI. Note that the vertical scale is logarithmic. Beyond about 18 bits, QCAD’s memory usage increases exponentially, as the dynamic data set size exceeds the memory requirements of its base API libraries; this behavior would be expected for any simulator based on an explicit state-vector representation. Note that, in comparison, SEQCSim’s memory usage remains essentially flat, at about 2 MB, throughout this range, and most of this resides in shared libraries. There is no data point for QCAD for circuit width 28 because the required memory (about 4 GB) exceeded what was available on the PC that was used for testing.
The next chart, shown in fig. 3, shows in more detail how the memory usage of SEQCsims increases as a function of the number of gates in the quantum circuit (for the same set of test circuits). There are some irregularities in the graph, which we hypothesize result from the fact that the number of pages allocated by the dynamic memory allocator may vary slightly from run to run depending on unpredictable factors. These irregularities could be minimized by averaging the results over multiple runs, but we have not yet done this. Despite the irregularities, there is a clear almost-linear trend in the memory usage which is well accounted for by the increasing number of stack frames that must be allocated as the number of levels of recursion (which is proportional to the depth of the circuit) increases. The actual growth rate, if measured more precisely and over a larger range, may actually be slightly faster than linear, because our current implementation is somewhat inefficient in that it stores a representation of the entire current basis state at each level of the recursion. It would be possible to eliminate this inefficiency through a more direct implementation that replaced the recursion with a more specialized iterative routine for traversing the tree of predecessors in an Euler tour.

![Linear growth of SEQCsims memory usage with size of quantum circuit](image)

**Fig. 3.** This graph shows how the peak memory usage of SEQCsims varied with the size in gates of the QFT-based adder circuits that were used for testing. The vertical scale here is linear, and we can see there is a slight, roughly linear increase in memory usage from 1900 KB to 1956 KB, in page-sized 4K increments, as we go from 9 operations for the 2×2 adder to 315 operations for the 2×14 adder. The results are somewhat noisy due to slight runtime variations in pages allocated by the memory manager because these results were not averaged over multiple runs. The extra 56K required for the larger circuit sizes is easily accounted for by the increased number of stack frames that must be allocated in order to get through all 315 levels of recursion in the calcAmp() function in the final step. If desired, the size of these stack frames could be further reduced from the current maximum of ~187 bytes to only a few bits, by replacing the recursive function with a more specialized iterative tree traverser that requires only enough bits at each level to keep track of which branch to visit next.
Finally, fig. 5 shows how the CPU time used by both QCAD and SEQCsim varies with the width of the simulated circuit. At present, SEQCsim is about a factor of 2 slower than QCAD for larger circuit sizes. This is unsurprising, given that presently SEQCsim is highly flexible (gates of any width can be defined) and uses a relatively elaborate recursive procedure to apply gates, as opposed to the more straightforward state-vector updating that must be done in QCAD. There is much room for further improvement in SEQCsim’s performance.

It would be fairly straightforward to modify SEQCsim to carry out an ordinary state-vector representation of the state until the limit of memory (or on-chip cache) is reached, and then revert to the recursive amplitude-calculation procedure only for further state evolution beyond that point. This would allow us to take full advantage of the available memory to boost performance, while retaining the ability to handle circuits of larger sizes without crashing.

Alternatively, we can keep SEQCsim’s memory usage minimal, while giving it a substantial constant-factor performance boost by reimplementing its kernel using a custom or semi-custom hardware architecture. That approach will be discussed in the next section.

[Diagram showing CPU time usage for QCAD vs. SEQCsim for 2×2 through 2×14 bit adders. QCAD’s greater CPU time usage for small circuit sizes can be accounted for by the fact that it has a GUI which is used to load the circuit and display results, whereas SEQCsim presently does not. We can see that the CPU time for both algorithms increases exponentially, as expected, as the circuit width increases. Each increase in circuit width by 2 bits results in about a factor of 4 increase in time complexity, as expected. In the case of QCAD this is because the state vector is 4× larger; whereas for SEQCsim it is because, for each additional bit in addend a, there are 2 additional non-trivial gates (Hadamard gates) in the circuit. The performance of SEQCsim could be significantly improved (beating that of QCAD) by either leveraging additional memory (recalculating fewer amplitudes), or by reimplementation in special-purpose, single-chip hardware, an approach which is made feasible because of SEQCsim’s low memory usage.]
5. FPGA-BASED EMBEDDED ARCHITECTURE

[Uwe to write this section.]

6. CONCLUSION AND FUTURE WORK

We have developed and demonstrated a working software prototype of an extremely memory-efficient quantum computer simulator, which will soon be released publicly through the site http://www.eng.fsu.edu/~mpf/SEQCSim.htm. Our prototype uses an amount of memory that increases only linearly in the size of the quantum circuit being simulated, at a proportion which was empirically found to be less than two hundred bytes of memory per gate in the present study.

A more carefully-optimized implementation should be able to do even better, and achieve an asymptotic memory usage of only 1 bit per qubit, plus 1 bit per typical nontrivial gate (such as a Hadamard gate or a general controlled-U gate).

Due to its miniscule memory requirements, a simulator of this class would be quite amenable for implementation in custom or semi-custom special-purpose hardware architecture, which can be easily prototyped using FPGAs loaded with embedded soft-core microprocessors such as Altera’s NIOS or Xilinx’s MicroBlaze, or a custom processor designed using the LISA tool set. Such a hardware-augmented implementation is expected to be able to outperform traditional software-only quantum computer simulators by a factor of 50-100×.

As for the software-only version of our simulator, its programming interface is presently rather cumbersome, requiring the user to define his or her quantum gates and gate sequences explicitly using an ad-hoc text input format. However, it would be straightforward to reimplement the simulator as, say, a set of classes in C++, or other object-oriented language, which would allow the programmer to describe quantum algorithms using the full expressive power of the host language, while observing a statistical behavior of his Qubit objects that matches what would be obtained from an ideal quantum computer, whose operation can be mimicked by a version of our simulator that is running “behind the scenes.”

This future version of our simulator would be a near-ideal tool for allowing students and researchers to venture into quantum programming and experiment with new quantum algorithms without having to either learn a new programming language first, or worry about running out of memory.

REFERENCES