2024 IEEE 6th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS-ISA) | 979-8-3503-8674-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/TPS-ISA62245.2024.00064

Study of Attacks on the HHL Quantum Algorithm

Yizhuo Tan
Yale University
New Haven, US
yizhuo.tan@yale.edu

Hrvoje Kukina
TU Wien
Vienna, Austria
hrvoje.kukina@student.tuwien.ac.at

Jakub Szefer
Yale University
New Haven, US
jakub.szefer@yale.edu

***Abstract*—As the quantum research community continues to grow and new algorithms are designed, developed, and implemented, it is crucial to start thinking about security aspects and potential threats that could result in misuse of the algorithms, or jeopardize the information processed with these quantum algorithms. This work focuses on exploration of two types of potential attacks that could be deployed on a cloud-based quantum computer by an attacker circuit trying to interfere with victim circuit. The two attacks, called Improper Initialization Attack (IIA) and Higher Energy Attack (HEA), are for the first time applied to a well-known and widely used quantum algorithm: HHL. The HHL algorithm is used in the field of machine learning and big data for solving systems of linear equations. This work evaluates the effect of the attacks on different qubits within the HHL algorithm: ancilla qubit, clock qubit, and b qubit. This work demonstrates that the two attacks are able to cause incorrect results, even when only one of the qubits in the victim algorithm is attacked. Having discovered the vulnerabilities, the work motivates the need for future work to develop defense strategies for each of these attack scenarios.**

***Index Terms*—Higher Energy Attacks, HEA, Improper Initialization Attacks, IIA, HHL Algorithm, Defense**

I. Introduction

In last few years, there has been a significant surge in the development of quantum computers and availability of of quantum processing units (QPUs) [16], which can be easily accessed online. Numerous technologies form basis of different types of QPUs, including: superconducting qubits [10], trapped ions [17], neutral atoms [18], silicon spin qubits [14], photons [26], and diamond NV centers [15]. Quantum computers based on these technologies are mostly so-called Noisy Intermediate-Scale Quantum (NISQ) computers, which do not have error correction yet. However, there are already algorithms that researchers are developing which can run on these quantum computers.

Among the promising algorithms for use with NISQ quantum computers is the HHL [9] quantum algorithm (named after its authors, Harrow, Hassidim, and Lloyd). It is designed for generating solutions to a set of linear equations. The HHL algorithm serves as a quantum counterpart to classical linear equation solvers such as Gaussian elimination. The HHL algorithm can have many applications, including in quantum machine learning [2], [5], [19].

With advent of cloud-based quantum computers, the algorithms such as HHL, will not be run on stand-alone machines. Rather, the Quantum Processing Unit (QPU) hardware in a cloud-based setting will be shared among different users and algorithms. Already today many cloud providers offer access to cloud-based QPUs, such as with IBM Quantum [10]. The QPU hardware can be time-shared where circuits, or shots of circuits, may be interleaved at different granularities. In future, multi-tenant quantum computers, which are already proposed in research [4], [6], [21], maybe deployed and further extend the amount of sharing among different, possibly distrusting, users. Consequently, security threats in the temporal and spatial sharing scenarios of QPUs need to be explored.

In the setting where adversary and victim may share the same quantum computer, this work explores two types of attacks and their impact on the important HHL algorithm. The first attack is what we call Improper Initialization Attack (IIA), involves malicious setting of the initial state of a qubit to $|1\rangle$, while it is expected that qubit should be set to $|0\rangle$. The second attack is the Higher Energy Attack (HEA), which involves setting the initial state of a qubit to a higher energy states such as $|2\rangle$ or $|3\rangle$, instead of the expected $|0\rangle$. In both cases, the incorrect setting of the qubit causes, as we demonstrate, incorrect results. We are first to show that even changing just one qubit in the victim algorithm is sufficient to launch the attack. The IIA attack could be launched when attacker is able to influence initial state of the qubit, for example through cross-talk; or the IIA can be part of software supply-chain attack where attacker is able to manipulate user's circuit without user's knowledge. The HEA is based on prior work that has demonstrated that setting a qubit to higher energy causes quantum gates applied to that qubit to not work correctly. An attacker may set a qubit to a higher energy state, and even if qubits are reset between attacker and victim circuit shots, the higher energy state propagates and causes victim's circuit to not work correctly. We perform in-depth evaluation of effect of the attacks on different qubits within HHL algorithm: ancilla qubit, clock qubit, and b qubit. In all cases, malicious output can be achieved even when only one qubit is attacked. This work motivates the need to think about how to secure algorithms such as HHL. The attacks show that HHL can be manipulated in a number of ways; therefore, defenses need to be developed.

This work was supported in part by NSF grant 2312754.

979-8-3503-8674-5/24/$31.00 ©2024 IEEE
DOI 10.1109/TPS-ISA62245.2024.00064

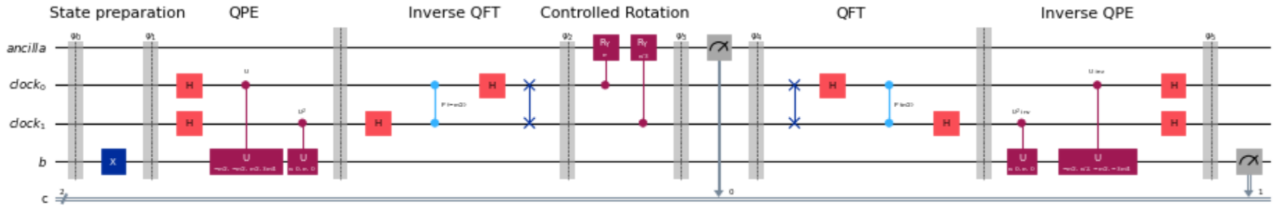

Fig. 1: Quantum circuit for the HHL algorithm, the vertical barriers are used to separate the different phases of the algorithm; the phases are labeled at the top of the circuit.

II. Background

This section presents background on the HHL algorithm. This section also presents background on supply-chain attacks, which are one means of launching our IIA attack, and this section also overviews higher energy states, which are key to the HEA attack.

A. HHL Algorithm

HHL is a quantum algorithm named after its authors, Harrow, Hassidim, and Lloyd, designed for quantum mechanical solutions to linear equations. It may also be referred to as a quantum matrix inversion algorithm. The HHL algorithm serves as a quantum counterpart to classical linear equation solvers such as Gaussian elimination and the conjugate gradient algorithm [3]. Its standout feature lies in its capacity to solve N linear equations with N variables (under certain conditions and limitations) in a time span proportional to $O(logN)$, in contrast to the classical $O(N)$, resulting in an exponential speedup. The HHL algorithm comprises three main components: quantum phase estimation, eigenvalue inversion rotation, and inverse quantum phase estimation. The initial stage of the HHL algorithm is quantum phase estimation (QPE), utilized for eigenvalue estimation. QPE itself consists of three subroutines: applying Hadamard gates on N-qubit clock register, applying controlled unitary transformations on the encoded b state vector achieved by the Hamiltonian simulation and applying the inverse Quantum Fourier transform on the N-qubit clock register at the end, before the second part of the algorithm – the eigenvalue inversion rotation.

NxN matrix A can be expressed in terms of its eigenvectors and eigenvalues as:

$$A = \sum_{i=0}^{N-1} \psi_i |u_i\rangle\langle u_i| \tag{1}$$

Vector b can also be constructed using the eigenbasis of A:

$$b = \sum_{j=0}^{N-1} b_j |u_j\rangle \tag{2}$$

The solution vector x can be written as follows:

$$|x\rangle = \sum_{i=0}^{N-1} \psi_i^{-1} b_j |u_i\rangle \tag{3}$$

In the context of the HHL algorithm, Hamiltonian simulation represents a subroutine within the quantum phase estimation. The Hamiltonian represents the energy operator and Hamiltonian simulation seeks algorithms capable of efficient quantum state time evolution implementation. Quantum Fourier Transform (QFT) represents an essential part of quantum phase estimation. Inverse QFT is needed to readout the eigenvalue. It consists of a sequence of Hadamard gates and controlled unitary rotation gates applied to N-qubits. In the quantum domain, QFT stands for basis change – from computational to Fourier.

After the measurement, the ancilla qubit has to be in the state $|1\rangle$ for the algorithm to produce the correct solution. If the state $|0\rangle$ is output after the measurement, the solution is discarded, and the process must be repeated, which makes the algorithm probabilistic in terms of obtaining the correct solution. An example of HHL circuit for a 2x2 matrix A and a two-dimensional vector b is presented in Fig. 1.

B. Software Supply Chain Attacks

The first attack, IIA, is based on attackers ability to modify the initial state of the victim's qubits. This can be achieved in different ways. Existing crosstalk attacks [1], [8], used in multi-tenant setting, have been used to manipulate victim's qubits by an attacker who executes operations on physically adjacent qubits. Another means of getting the qubits into incorrect initial state is to manipulate the user's circuit. In classical computing, well-known software supply-chain attacks [12] have been studied and demonstrated, where attackers are able to modify code packages, or otherwise manipulate the supply chain of the software so that user's code is modified unbeknown to them.

C. Higher Energy States

The second attack, HEA, is based on higher energy states. In practical quantum computing, qubits are generally manipulated between low-energy states such as $|0\rangle$ and $|1\rangle$. The ground state, the lowest energy state of a quantum system, is commonly represented by $|0\rangle$ for a qubit. The $|1\rangle$ is the elevated energy state in the typical two-level quantum mechanical system. Excited states, which have higher energy than the ground state, also exist. For example, excited states with even higher energy levels such as $|2\rangle$, $|3\rangle$, and so on, can exist. Most gate-level quantum computers only require $|0\rangle$ and $|1\rangle$, and the higher energy states are not used – although they can still be generated.

The higher energy states are not typically used directly in quantum computations, but there is nothing that prevents users with pulse-level access to quantum computers, which is common in machines such as IBM’s quantum computers, from generating such states. It is thus crucial for understanding the security of quantum systems to think about these higher energy states.

A higher energy state attack could involve an attacker manipulating a qubit to force it from its predetermined state, usually $|0\rangle$, into a higher energy state above $|0\rangle$. For superconducting quantum systems, we assume that the user has pulse-level control over qubits, which is reasonable because IBM Quantum [10] already provides such tools such as Qiskit Pulse [11]. With malicious experiments to obtain information from the specific qubit, one can build a custom pulse to excite a qubit to higher energy state. Previous researches have shown that higher energy state such as $|2\rangle$ is harmful to the fidelity of quantum circuit outputs [22]. First, the higher energy state can disturb measurement and state discriminator, which leads to wrong measure output ’1’ for all higher energy states. Second, the frequency of the higher energy state is different from normal states $|0\rangle$ and $|1\rangle$. This property will disable all predefined and well-calibrated gates for $|0\rangle$ and $|1\rangle$ because superconducting systems employ microwave pulses, which match the corresponding frequency of qubits, to manipulate their qubits. Another property previously discovered is that the higher energy states cannot be properly reset without using specially designed CSR gate [22]. This means if the adversary excites his or her qubits to higher energy states, the following user who is allocated to the same set of qubits may encounter wrong initialization issue. These can, as a consequence, prevent the quantum system from producing the correct output and become a serious security issue for superconducting quantum systems.

D. IIA and HEA Attacks in NISQ Quantum Computer Setting

This work evaluates the IIA and HEA attacks on Noisy Intermediate-Scale Quantum (NISQ) quantum computers. The NISQ quantum computers do not have error correction nor error mitigation, this makes them susceptible to various types of noise. The noise can affect the fidelity of the HHL algorithm, as well as the attacks themselves. The experiments in this paper we performed on NISQ quantum computers in presence of noise, and the attacks work. Detailed study of noise impact on IIA and HEA, however, is left as future work.

III. Threat Model

To understand the IIA and HEA attacks, this work makes the following assumptions. We assume the attacker does not have any special privileges. It only has user-level, remote access to quantum computers. We assume they can use tools such as Qiskit Pulse to generate custom pulses, but this requires no special privileges.

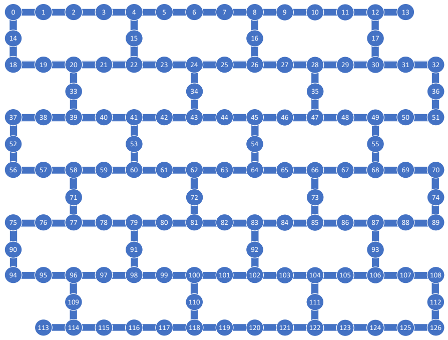

Fig. 2: Topology of a IBM Quantum QPU, the Eagle r3 processor, with 127 qubits. Circles represent qubits, thick lines represent fixed couplings between the qubits.

We assume the attacker and victim share the same quantum computer, either spatially or temporarily. We further assume the attacker is able to know which qubits the victim is using. As our experimentation shows, it is not necessary to know which victim qubit is used for ancilla, clock, or b register, since the attack on any one of them works.

For the IIA attack, we assume the attacker is able to set the victim’s qubit or qubits into $|1\rangle$ state at beginning of each shot of victim’s circuit. For the HEA attack, we assumed the attacker is able to set the victim’s qubit or qubits into $|2\rangle$ or other higher energy states at beginning of each shot of victim’s circuit. Existing work has already considered how to set qubit states into improper state, e.g., it has showed that higher energy states are not properly reset by `reset` gates [20]. This work does not explore how to launch such attacks, but instead focuses on their impact on the HHL algorithm.

IV. Experimental Setup

In this section we present the experimental setup. The experiments were performed both in simulation and on real superconducting quantum computers that were accessed remotely through the IBM Quantum service. In all castes, the number of shots of a quantum circuit is set to 1000 shots to obtain reasonable number of outputs used in computing the output probabilities, while limiting the experiment duration time.

A. Quantum Simulators Used

Two types of quantum simulators were used for the experiments: BasicSimulator and AerSimulator, both from IBM. BasicSimulator is part of Qiskit and is one of the “providers”. A provider is credited with supplying external services (such as objects) to Qiskit. It can be accessed using the following class: `qiskit.providers.basic_provider.BasicSimulator` in Qiskit.

On the other hand, AerSimulator is part of Qiskit Aer and can be accessed using the following class: `qiskit.providers.aer.backends.aerbackend.AerBackend`.

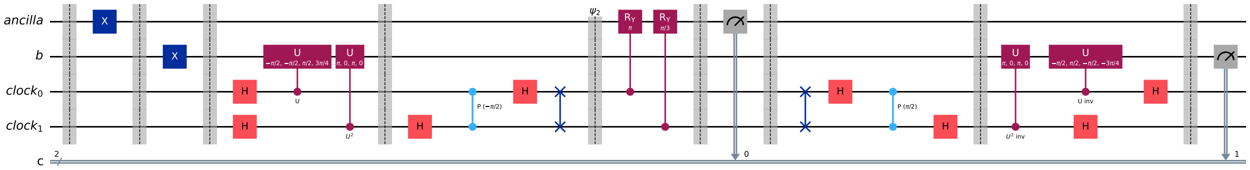

Fig. 3: Emulation of improper initialization attack on HHL ancilla qubit. To emulate the attack, an additional X gate is inserted at the beginning of the circuit in order to set the ancilla qubit to $|1\rangle$ state before the circuit executes.

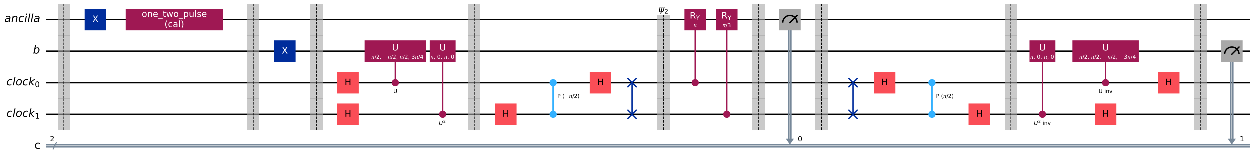

Fig. 4: Emulation of higher energy attack on HHL ancilla qubit. To emulate the attack, an additional X gate is inserted at the beginning of the circuit in order to set the ancilla qubit to $|1\rangle$ followed by a custom pulse used to excite the qubits from $|1\rangle$ to $|2\rangle$ state, i.e. the higher energy state, before the circuit executes.

The two simulators where used as-is without any modifications. Since the simulators are gate-level simulators that support only $|0\rangle$ and $|1\rangle$ states, only the IIA attacks were evaluated on them. For HEA, real quantum hardware was used.

B. Quantum Hardware Used

In addition to the simulation, we used real quantum hardware, namely the $IBM\_brisbane$ machine, which has 127 qubits with the heavy-hexagonal layout. The physical qubits used in experiments are qubit 0, 1, 2, and 3 or 14[1], since the tested HHL algorithm requires 4 qubits. Larger HHL algorithms were not tested due to very noisy outputs on the real hardware we have access to. All experiments are set to repeat for 10000 shots and optimization level was set to 0. Each kind of experiment is repeated 3 times due to limited time available on the real hardware. The topology of the $IBM\_brisbane$ used in testing on real hardware is shown in Fig. 2.

C. Generating Higher Energy States

As mentioned in II-C, IBM Quantum provides pulse-level control over superconducting qubits through Qiskit Pulse [11]. Although Qiskit Pulse allows the user to self-define their custom pulse, he or she still needs to specify the parameters for the custom pulse. We, and any attacker, can perform frequency sweep and Rabi experiments to obtain the parameters for the pulse needed to excite a qubit into higher energy state. The parameters are specific to each physical qubit, and the frequency sweep and Rabi experiments need to be performed on each qubit that is to be excited into higher energy states.

It should be noted that, if the transpiler optimizes an input circuit and changes the layout during the circuit execution, then the higher energy state may be wrongly injected. The reason is that logical qubits are assigned initially to some physical qubits. But during the execution, swap gates may be added by the transpiler to move the logical qubits to different physical qubits. However, swap gates are ineffective when higher energy states present, so the logical qubits will be moved, but the higher energy states remain on the physical qubits they were set to initially. As a result, we need to carefully arrange the initial layout between logical qubits and physical qubits for each experiment to fit $IBM\_brisbane$'s topology and check the transpiled circuit to make sure we put the attack on the target qubit correctly. That is also the reason why we use two sets of initial layouts $[0, 1, 2, 3]$ and $[0, 1, 2, 14]$ for different experiments.

V. Circuits Used for Testing Attacks

This work considers the two IIA and HEA attacks. The circuits used to evaluate the two attacks are shown in Fig. 3 and Fig. 4 respectively. The examples in the figures demonstrate the attacks on the ancilla qubits. But same approach is taken for attack on all the other qubits.

For the HHL part of the testing circuits, in our experiments, we use the following matrix A:

$$\begin{bmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{bmatrix} \tag{4}$$

and vector b:

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{5}$$

The result of the HHL circuit is interpreted from the output probabilities when the b register is measured. The way a solution encoded in a quantum state can be compared to

[1] Either qubit 3 or 14 is used in some experiments to deal with transpilation issues.

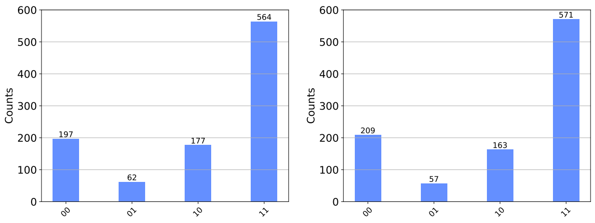

Fig. 5: Comparison between baseline results of running HHL on the BasicSimulator (left) and the AerSimulator (right).

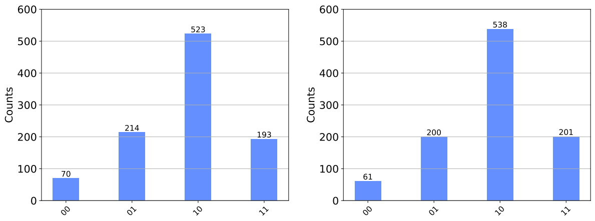

Fig. 6: Comparison of results of the IIA attack on the ancilla qubit tested on BasicSimulator (left) and the AerSimulator (right).

a classical solution vector for a particular system of linear equations is through the ratio of the squares of the magnitudes of its components. In our case, the correct output should be in a ratio of 1:9, meaning that among all the measurements when the ancilla qubit is 1, the number of measurements of the b register that give 0 versus the number that give 1 should be in the ratio 1:9.

For testing the IIA attack, an additional X gate is inserted at the beginning of the circuit in order to set the target qubit to $|1\rangle$ state before the circuit executes. For testing the HEA attack, an additional X gate is inserted at the beginning of the circuit in order to set the target qubit to $|1\rangle$ followed by a custom pulse used to excite the qubits from $|1\rangle$ to $|2\rangle$ state, i.e. the higher energy state, before the circuit executes.

In the experiments, we have limited the testing to the HHL circuits with 4 qubits (ancilla, b, and two clock qubits). Larger HHL circuits do not work correctly on the current hardware we have access to, thus testing of larger HHL circuits is left as future work.

VI. Evaluation and Results for Simulators

In this section we present results from testing in simulation. We focus on the two simulators available in Qiskit as discussed before.

A. Baseline HHL Outputs

Fig. 5 shows the result from the BasicSimulator and the AerSimulator. The BasicSimulator gives results of 1:9.0968 and the result from the AerSimulator is 1:10.0175. Note that both ancilla and b qubits are measured, resulting in four possible states shown in figures: 00, 01, 10, and 11. The first bit is b and second bit is ancilla. The HHL output is only considered correct when ancilla is measured to be 1. Thus, when computing the ratios, we compute the ratio of the state 01 to state 11.

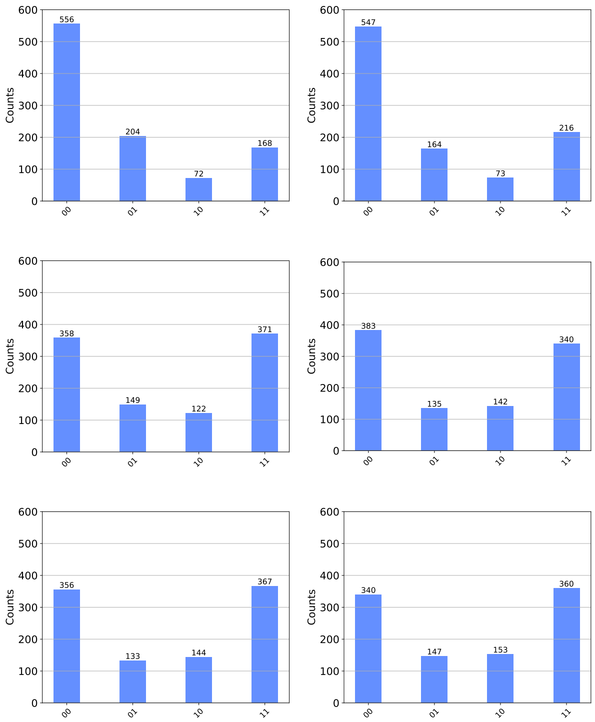

Fig. 7: Comparison of results of the IIA attack on the two clock qubits tested on BasicSimulator (left) and the AerSimulator (right). Top row is attack on the first clock qubit, middle row is attack on the second clock qubit, and last row is attack on both clock qubits.

B. HHL Outputs Under Attack

In the following text we report the results of the IIA attack on the ancilla qubit, clock qubit, b qubit.

1) IIA Attack on Ancilla Qubit: Fig. 6 shows that for the improper initialization attack on the ancilla qubit, the result from the BasicSimulator is 1:0.9019, while the result from the AerSimulator is 1:1.0050, which clearly shows that the attack worsened the result by 9 times.

2) IIA Attack on Clock Qubit: Fig. 7 shows that for the improper initialization attack on the clock0 qubit, the result from the BasicSimulator is 1:0.8235, while the result from the AerSimulator is 1:1.3171. For the improper initialization attack on the clock1 qubit, the result from the BasicSimulator is 1:2.4899, while the result from the AerSimulator is 1:2.5185. For the improper initialization attack on both clock qubits, the result from the BasicSimulator is 1:2.7594, while the result from the AerSimulator is 1:2.4490.

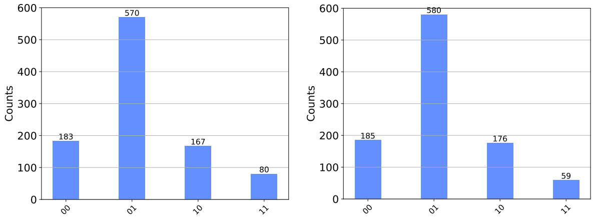

Fig. 8: Comparison between the BasicSimulator and the AerSimulator for the improper initialization attack on the b qubit.

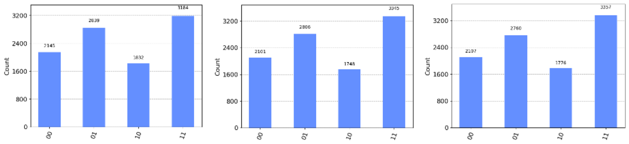

Fig. 9: Baseline HHL outputs testing HHL without attacks when using qubit 0,1,2,3 on *IBM_brisbane*. Three graphs represent the three tests, each with 1000 shots.

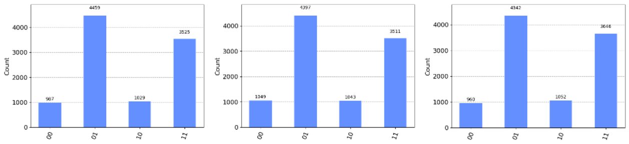

Fig. 10: Higher energy attack on HHL ancilla using qubit 0,1,2,3 on *IBM_brisbane*.

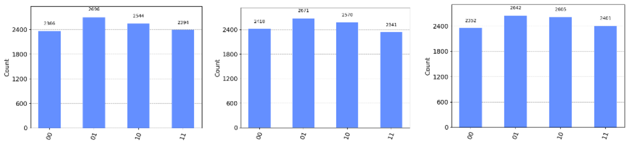

Fig. 11: Improper initialization attack on HHL ancilla using qubit 0,1,2,3 on *IBM_brisbane*.

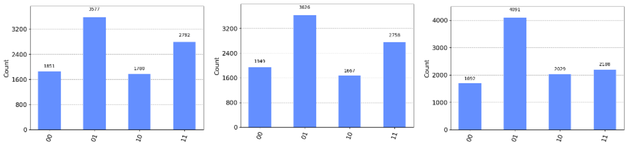

Fig. 12: Higher energy attack on HHL clock0 using qubit 0,1,2,3 on *IBM_brisbane*.

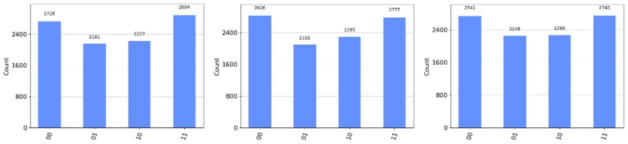

Fig. 13: Improper initialization attack on HHL clock0 using qubit 0,1,2,3 on *IBM_brisbane*.

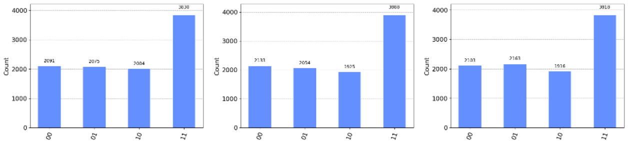

Fig. 14: Higher energy attack on HHL clock1 using qubit 0,1,2,3 on *IBM_brisbane*.

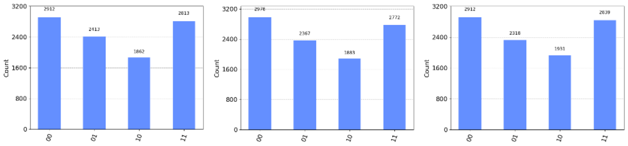

Fig. 15: Improper initialization attack on HHL clock1 using qubit 0,1,2,3 on *IBM_brisbane*.

3) IIA Attack on b Register: Fig. 8 shows that for the improper initialization attack on the b qubit, the result from the BasicSimulator is 1:0.1403, while the result from the AerSimulator is 1:0.1017.

C. Summary of the Effectiveness of the Attacks

To compare the results without and with attack, we used the variational distance metric. Table III shows the variational distance between original HHL probability distribution and improper initialization attack probability distributions under the different attacks on ancilla, clock, and b qubits. It can be seen that attacking any of the qubits results in significant variational distance. Interestingly, attacking clock1 or both clock0 and clock1 has less impact than attacking the other qubits. Nevertheless, we can surmise that attacking any one qubit is sufficient to generate incorrect results.

VII. Evaluation and Results for the Hardware

In this section we present results from testing the attacks on real hardware. We used the $IBM\_brisbane$ machine. We can test IIA and HEA attacks on this machine.

A. Baseline HHL Outputs

We again use the original 2x2 HHL algorithm mentioned in Sec. VI. The ratios of the three repeated HHL experiments in Fig. 9 on $IBM\_brisbane$ machine are 1:1.1215, 1:1.1921, and 1:1.2163. It can be immediately seen that even without the attack, due to the noisy nature of the real hardware, the output ratios are much worst than in simulation.

TABLE I: Variational distances of HHL outputs under improper initialization attacks on BasisSimulator and AerSimulator.

Victim qubit	BasicSimulator	AerSimulator
no attack	0	0
ancilla	0.4980	0.5180
clock0	0.5010	0.4450
clock1	0.2489	0.2520
clock0 and clock1	0.2300	0.2210
b	0.5080	0.5360

TABLE II: Variational distances of HHL outputs under higher energy attacks on $IBM\_brisbane$.

Victim qubit	Average ratio	Average variational distance
no attack	1:1.1766	0
ancilla	1:0.8096	0.1863
clock0	1:0.6920	0.1003
clock1	1:1.8077	0.0713
b	1:2.7822	0.2065

TABLE III: Variational distances of HHL outputs under improper initialization attacks on $IBM\_brisbane$.

Victim qubit	Average ratio	Average variational distance
no attack	1:1.1766	0
ancilla	1:0.8911	0.1049
clock0	1:1.2923	0.1125
clock1	1:1.1872	0.0906
b	1:1.0240	0.0616

B. HHL Outputs Under Attack

In the following text we report the results of the HEA attack on the ancilla qubit, clock qubit, b qubit.

1) IIA and HEA Attacks on Ancilla Qubit: Fig. 10 shows the higher energy attack on HHL ancilla qubit. This attack greatly harms the algorithm output and reduces the ratios to 1:0.7905, 1:0.7985, and 1:0.8397. Improper initialization attack is meanwhile presented in Fig. 11 and it also decreases the ratios to 1:0.8880, 1:0.8764, and 1:0.9088.

2) IIA and HEA Attacks on Clock Qubit: Fig. 12 and Fig. 13 show the two attacks on HHL clock0 qubit. The higher energy attack on clock0 decreases the ratios to 1:0.7895, 1:0.7606 and 1:0.5348, while improper initialization attack on clock0 does not affect HHL that much, with ratios of 1:1.3345, 1:1.3211 and 1:1.2211. Attacks on HHL clock1 qubit in Fig. 14 and Fig. 15 show interesting influence on the HHL output. The ratios of higher energy attack on clock1 are 1:1.7651, 1:1.8929 and 1:1.7651. The ratios of improper initialization attack on clock1 are 1:1.1657, 1:1.1711 and 1:1,2247, which are almost the same as original HHL output.

3) IIA and HEA Attacks on b Qubit: Fig. 16 present the higher energy attack on HHL b qubit. This attack is allocated to qubit 0, 1, 2, 14 on $IBM\_brisbane$ machine. The ratios for the higher energy attacks on b are 1:2.8028, 1:2.8626, and 1:2.6813. Even though it seems that the higher energy attack on b improves HHL performance closer to classical solution, this attack will actually force the ratio higher than classical solution if we have better quantum hardware than NISQ devices. The ratios for improper initialization attacks on HHL b qubit in Fig. 17 are 1:0.9157, 1:0.8959, and 1:1.2604.

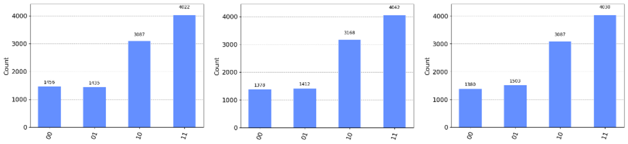

Fig. 16: Higher energy attack on HHL b using qubit 0,1,2,14 on *IBM_brisbane*.

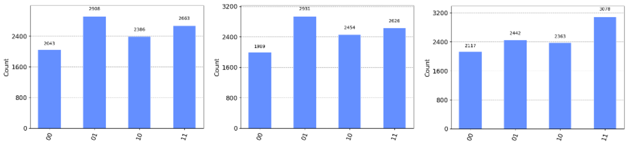

Fig. 17: Improper initialization attack on HHL b using qubit 0,1,2,14 on *IBM_brisbane*.

C. Summary of the Effectiveness of the Attacks

Table II and table III summarize these two attacks on HHL algorithm on $IBM\_brisbane$ machine. We use the average ratio and the variational distance to quantify the difference between probability distribution of original HHL hardware output and that of attacked output. On real hardware, higher energy attacks seem to have bigger impact in terms of the variational distance metric. On the other hand, the noisy nature of the NISQ computers means that effects of the attacks and the noise both affect the output, and further study of the attacks on real hardware is necessary.

VIII. Related Work

The security of quantum algorithms and quantum information, in general, has become a very interesting field lately. Given that quantum machines are developing at a rapid pace, researchers have started focusing on the security aspects. Therefore, [23] exhibits how to securely transmit information using the HHL algorithm to prevent information leakage, while [7] defines a new measure of information leakage for the quantum encoding of classical data. Error propagation in the HHL algorithm has been studied in [24], where the authors identified three major sources of errors: single-qubit flipping, gate infidelity, and error propagation. There is also a potential way to reduce the demands on physical qubits by evaluating the resource cost of quantum phase estimation, which is a crucial part of the HHL algorithm, before and after quantum error correction [25]. Similarly, quantum phase estimation, being one of the most computationally expensive components of the HHL algorithm, has been tested in terms of scaling properties and related noise resilience [13].

IX. Conclusion and Future Work

This work demonstrates two types of attacks that could be performed on the HHL algorithm, both on quantum simulators and on quantum hardware. To the best of our knowledge, this is the first paper that demonstrates and explains possible attacks on the HHL algorithm, both in theory and in practice.

A malicious user could perform improper initialization attacks and higher energy attacks. Higher energy attacks can only be studied on quantum hardware, while improper initialization attacks can be studied on both quantum simulators and quantum hardware. We demonstrated three attack scenarios: attacks on the ancilla qubit, the clock qubits, and the b qubit. We showed that the results obtained by running the HHL circuit under attack greatly differ from those obtained when the HHL circuit is not maliciously compromised. A potential follow-up is to design and implement possible defenses against each attack and each scenario on both quantum simulators and hardware. A potential defense strategy would incorporate adding new ancilla registers to the original HHL circuit and applying a combination of `CNOT`, `X`, and `Reset` gates, using measurement to detect the states of qubits before and after the execution of the circuit, and thereby concluding whether there is an attack and if it can be defined as IIA or HEA. Other possible defenses include leveraging the transpiler or scheduler. It's possible that the transpiler randomizes the assignment of victim's qubits, so that attacker cannot not easily predict which are the victim's qubits, which will make the attack more difficult.

References

[1] A. Ash-Saki, M. Alam, and S. Ghosh, "Analysis of crosstalk in nisq devices and security implications in multi-programming regime," in *Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design*, 2020, pp. 25–30.

[2] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.

[3] R. Chandra, "Conjugate gradient methods for partial differential equations." 1978. [Online]. Available: https://api.semanticscholar.org/CorpusID:118155422

[4] S. Deshpande, C. Xu, T. Trochatos, Y. Ding, and J. Szefer, "Towards an antivirus for quantum computers," in *2022 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, 2022, pp. 37–40.

[5] B. Duan, J. Yuan, C.-H. Yu, J. Huang, and C.-Y. Hsieh, "A survey on hhl algorithm: From theory to application in quantum machine learning," *Physics Letters A*, vol. 384, no. 24, p. 126595, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037596012030462X

[6] A. D'Onofrio, A. Hossain, L. Santana, N. Machlovi, S. Stein, J. Liu, A. Li, and Y. Mao, "Distributed quantum learning with co-management in a multi-tenant quantum system," in *2023 IEEE International Conference on Big Data (BigData)*. IEEE, 2023, pp. 221–228.

[7] F. Farokhi, "Maximal information leakage from quantum encoding of classical data," *Phys. Rev. A*, vol. 109, no. 2, p. 022608, 2024.

[8] B. Harper, B. Tonekaboni, B. Goldozian, M. Sevior, and M. Usman, "Crosstalk attacks and defence in a shared quantum computing environment," *arXiv preprint arXiv:2402.02753*, 2024.

[9] A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," *Physical review letters*, vol. 103, no. 15, p. 150502, 2009.

[10] IBM Quantum, "Ibm quantum platform," 2024, https://quantum.ibm.com/.

[11] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, "Quantum computing with Qiskit," 2024.

[12] P. Ladisa, H. Plate, M. Martinez, and O. Barais, "Sok: Taxonomy of attacks on open-source software supply chains," in *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2023, pp. 1509–1526.

[13] M. A. Marfany, A. Sakhnenko, and J. M. Lorenz, "Identifying bottlenecks of nisq-friendly hhl algorithms," 2024. [Online]. Available: https://arxiv.org/abs/2406.06288

[14] S. Neyens, O. K. Zietz, T. F. Watson, F. Luthi, A. Nethwewala, H. C. George, E. Henry, M. Islam, A. J. Wagner, F. Borjans *et al.*, "Probing single electrons across 300-mm spin qubit wafers," *Nature*, vol. 629, no. 8010, pp. 80–85, 2024.

[15] J. Pena, "Quantum diamond biomarker detection," *PhotonicsViews*, vol. 19, no. 1, pp. 48–50, 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/phvs.202270107

[16] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018. [Online]. Available: https://doi.org/10.22331/q-2018-08-06-79

[17] Quantinuum, "Quantinuum," 2024, https://www.quantinuum.com/.

[18] QuEra Computing Inc, "Quera," 2023, https://www.quera.com/.

[19] M. Schuld, I. Sinayskiy, and F. Petruccione, "An introduction to quantum machine learning," *Contemporary Physics*, vol. 56, no. 2, pp. 172–185, 2015.

[20] J. Tan, C. Xu, T. Trochatos, and J. Szefer, "Extending and defending attacks on reset operations in quantum computers," 2023. [Online]. Available: https://arxiv.org/abs/2309.06281

[21] S. Upadhyay and S. Ghosh, "Stealthy swaps: Adversarial swap injection in multi-tenant quantum computing," 2023. [Online]. Available: https://arxiv.org/abs/2310.17426

[22] C. Xu, J. Chen, A. Mi, and J. Szefer, "Securing nisq quantum computer reset operations against higher energy state attacks," in *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 2023, pp. 594–607.

[23] X. Yang, D. Li, J. Zhou, Y. Tan, Y. Zheng, and X. Liu, "Research on quantum dialogue protocol based on the HHL algorithm," *Quant. Inf. Proc.*, vol. 22, no. 9, p. 340, 2023.

[24] A. Zaman and H. Y. Wong, "Study of error propagation and generation in harrow-hassidim-lloyd (hhl) quantum algorithm," in *2022 IEEE Latin American Electron Devices Conference (LAEDC)*, 2022, pp. 1–4.

[25] M. Zheng, C. Liu, S. Stein, X. Li, J. Mülmenstädt, Y. Chen, and A. Li, "An Early Investigation of the HHL Quantum Linear Solver for Scientific Applications," 4 2024.

[26] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, "Quantum computational advantage using photons," *Science*, vol. 370, no. 6523, pp. 1460–1463, 2020.