Quantum computing technology is currently developing at a very fast pace. The main obstacle to scaling up the number of qubits on the Quantum Processing Unit (QPU) is noise [1]: QPU are still subject to a number of noise sources that make them, at the current stage of development, still prone to large error. Noise may affect a quantum computation at each stage thereof, from initial qubit state preparation, to gate application, to readout and storage. Here we focus on the preparation. The starting point of any quantum algorithm, a so called quantum circuit, is a tensor product of the ground states of all qubits on the QPU that participate to the computation. From a thermodynamical perspective that is a zero temperature state. The third law of thermodynamics actually forbids its achievement: such a state can only be achieved to some degree of approximation [2]. That is, the unavoidable starting point of any quantum circuit is a state of some finite (no matter how small) temperature, rather than an ideal pure quantum state. Then, a question of crucial technological relevance is how to achieve smaller and smaller temperature of the initial preparation. The most direct way of addressing this problem is to control and reduce to a minimum all sources of noise that may affect the preparation.

Here we propose to adopt an alternative thermodynamic approach instead. As we learn from standard thermodynamic textbooks a thermodynamic refrigerator is a machine that takes heat away from a cold body to heat up a hotter one by consuming some power coming from an external energy source [3]. Thus, our idea is to do the same on a QPU, where one qubit would be cooled down at the cost of heating up another qubit (or more qubits as we shall see below), while some energy is spent to make that happen. That energy comes, as we shall see below, from application of a properly designed entangling gate on the set of involved qubits.

**Qubit Refrigeration Method**

Our quantum refrigeration scheme is a modification of the so called quantum SWAP engine [4–9]. A quantum SWAP engine is composed of two qubits, a hot qubit (labelled as qubit H from now on) being at temperature $T_H$ and a cold one (labelled as qubit C) being at temperature $T_C < T_H$. As reported previously [7, 10], application of the SWAP unitary to the two qubits results in the cold qubit getting to a colder temperature $T'_C < T_C$ and the hot one to a hotter temperature, provided the ratio of the two qubits resonant frequencies $\omega_C/\omega_H$, is smaller than the ratio of their initial temperatures $T_C/T_H$. Besides, among all the unitaries, the SWAP is the one that achieves the highest cooling coefficient of performance (COP), reading $\eta = (\omega_H/\omega_C - 1)^{-1}$.

The main difficulty that one encounters when trying to implement this simple scheme on current QPUs, is that they are engineered to have ideally identical qubits. If that is the case then the SWAP engine described above would not work: for $\omega_C = \omega_H$ the condition $\omega_C/\omega_H < T_C/T_H$, implies $T_C > T_H$, which contradicts that label C denotes the colder qubit. We have evidenced this unfortunate situation with a previous set of experiments performed on an IBM QPU [11]. In order for the cooling mechanism to work, a necessary condition is:

$$\omega_H > \omega_C.$$  \hspace{1cm} (1)

The larger is $\omega_H$ as compared to $\omega_C$, the more robust will be the cooling operation [11], while the coefficient of performance $\eta$, will decrease.

The question is then whether one can modify the SWAP engine design, in order to implement a working cooling mechanism on a QPU with identical qubits with non-tunable resonant frequency. One simple way to achieve that is to combine two or more qubits together to form a compound multi-level system having a larger resonant frequency.

Our solution is then to replace the hot qubit with a compound system made of two qubits, and focus only...
on its ground and most excited states, $|00\rangle_H$ and $|11\rangle_H$, respectively. Those two states will play the role of a qubit with a doubled resonant frequency. That would make the “hot” resonance $\omega_H$ be twice the “cold” resonance $\omega_C$, and opens for the possibility of refrigerating the cold qubit.

Assuming the qubit to be cooled is initially at some temperature $kT_C = 1/\beta_C$ ($k$ is Boltzmann’s constant), the first step of the method is to populate the states $|00\rangle_H$ and $|11\rangle_H$ according to a Gibbs distribution of temperature $kT_H = 1/\beta_H$, so that the total system is initially described by the density operator

$$\rho = \frac{e^{-\beta_H H_H - \beta_C H_C}}{Z_H Z_C}$$  \hspace{1cm} (2)$$

where

$$H_H = -\frac{\hbar \omega_H}{2} (|00\rangle\langle00|_H - |11\rangle\langle11|_H) \otimes 1_C$$  \hspace{1cm} (3)$$

with $\omega_H$ being the sum of the resonant frequencies of the two qubits composing system $H$; and

$$H_C = -\frac{\hbar \omega_C}{2} (|00\rangle\langle00|_C - |11\rangle\langle11|_C)$$  \hspace{1cm} (4)$$

being the cold qubit Hamiltonian ($\hbar$ is the reduced Planck constant). The symbol $Z_r$ stands for the according partition function $Z_r = Tr_r e^{-\beta H_r}$. $Tr_r$ and $1_r$ denote trace operation and the identity operator in the $r$ subsystem Hilbert space, respectively. In the following we adopt the notation $|ij,k\rangle$, with $i, j, k = 0, 1$ to denote the energy eigen-basis of the compound 3 qubit system, with the first two indices referring to system $H$, and the third index referring to system $C$.

The second step of the method consists in applying a unitary operation $U$ that maps $|ii,j\rangle$ onto $|jj,i\rangle$. One such gate generally reads

$$U = \frac{W}{\sqrt{\nu}}$$ \hspace{1cm} (5)$$

where $V$ is a generic unitary acting on the subspace spanned by $\{|01,0\rangle, |01,1\rangle, |10,0\rangle, |10,1\rangle\}$ and $W$ is the relevant swap operation occurring in the subspace spanned by $\{|00,0\rangle, |00,1\rangle, |11,0\rangle, |11,1\rangle\}$. Its matrix representation in that basis reads

$$W = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$ \hspace{1cm} (6)$$

The specific form of $V$ does not have any impact on the thermodynamics of the device. This is because in the preparation of Eq. (2), the states $\{|01,0\rangle, |01,1\rangle, |10,0\rangle, |10,1\rangle\}$ are not populated, hence any dynamics occurring in the space they span is immaterial from the energetic point of view. As we shall see below, however, the choice of $V$ may have a great impact from the practical point of view. The quantum circuit representation of the method is sketched in Fig. 1(b).

**RESULTS**

We have implemented the method on IBM *ibmq_jakarta* QPU with two different choices of $V$. Its topology is depicted in Fig 1(b). Qubit 1 is the cold qubit that needs to be refrigerated. Qubits 0, 2 form the $H$ system. Their resonant frequencies were $\omega_0 \approx 5.24$ GHz, $\omega_1 \approx 5.01$ GHz and $\omega_2 \approx 5.11$ GHz, hence $\omega_C \approx 5.01$ GHz, $\omega_H \approx 10.35$ GHz. The methods used to obtain the experimental data are described in detail below.

Figure 2(a) shows the theoretical “phase diagram” in the $T_H, T_C$ plane showing which thermodynamic operation mode is expected. We recall that, based on general quantum mechanical arguments, only 3 operations modes are possible besides Refrigeration [R] [12]. They are: Heat Engine [E], when heat is transferred from the hot to the cold subsystem while energy is output in the form of work; Thermal Accelerator [A], when heat is transferred from the hot to the cold subsystem while work is spent; Heater [H], when both subsystems receive energy from the work source. Note the extended connected blue region indicating that refrigeration can in principle be robustly implemented.

Our first choice of $V$ was $V = 1$ with $1$ denoting the identity operator on the Hilbert space spanned by $\{|01,0\rangle, |01,1\rangle, |10,0\rangle, |10,1\rangle\}$. Figure 2(b) shows the according experimental “phase diagram” in the $T_H, T_C$...
plane. Note that in comparison with the theoretical expectation, presenting no region \([H]\) of heating for both subsystems, a large portion of the phase diagram is in fact taken by this region, especially at low temperatures. A blue \([R]\) region where refrigeration occurred exists, but it is very much shrunk as compared to theory. In particular no refrigeration was observed below temperature \(T_C = 172\, \text{mK}\) of the cold qubit (dashed vertical line). These effects were mostly due to the noise affecting the gate \(U\). We note, in fact, that in our experiments the gate \(U\) was decomposed and implemented by the IBM compiler as a sequence of more than 180 elementary gates \[13\]. Counting that each elementary gate comes with its load of noise, no matter how small, the high number thereof resulted in a good amount of noise, which greatly affected the functioning of the device.

In order to mitigate this problem (and to confirm that gate noise was indeed the source of the detrimental effects) we repeated the experiment with the choice of \(V \neq I\) being the unitary that maps \(|01,1\rangle\) onto \(|10,0\rangle\) (and vice-versa), while leaving the states \(|01,0\rangle, |10,1\rangle\) unaltered. In the following we shall refer to this choice as \(V^*\). Its matrix representation reads, in the basis \(|00,0\rangle, |00,1\rangle, |10,0\rangle, |11,1\rangle\), i.e., the matrix in Eq. \[6\].

At variance with the \(V = I\) case, with \(V = V^*\) the global unitary \(U\) was implemented with only 4 CNOTs as shown in Fig. \[3\].

Figure \[2\] shows the experimental “phase diagram” in the \(T_H, T_C\) plane obtained with the choice of \(V = V^*\). Note how this choice has resulted, in comparison with the choice \(V = I\), in a shrinking of the heating region \([H]\) (in red), while the refrigeration region \([R]\) (in blue) has enlarged, thus realising a more robust cooling operation. Most remarkably, the \([R]\) region now extends down to \(T_C = 52\, \text{mK}\) meaning that the improved implementation allows to cool a qubit to lower temperature, as compared to the more noisy case. This clearly indicates that decreasing the gate noise further will lead to even lower cooling temperature, and better performance.

Figure \[3\] shows the final temperature \(T_C'\) of the cold qubit, as a function of initial temperatures \(T_H, T_C\), in the refrigeration region \([R]\), for the \(V = V^*\) case.

\[\text{FIG. 2. “Phase diagrams” of the quantum heat engine. Panel a): Theory. Panel b): Experiment with the choice } V = I. \text{ Panel c): Experiment with the choice } V = V^*.\]

\[\text{FIG. 3. Final temperature } T_C' \text{ of the cold qubit as a function of initial temperatures } T_H, T_C, \text{ in the refrigeration region, for } V = V^*.\]

**METHODS**

We implemented the cooling protocol on the IBM Quantum processor \(ibmq\_jakarta\) which we remotely accessed through the Qiskit library \[14\]. The topology of the quantum processor is shown in Fig. \[4\]. Only qubits \(q_0, q_1\) and \(q_2\) were addressed in our experiments.

Four sets of experiments were performed each with the qubits being initialized in one of the states \(|00,0\rangle, |00,1\rangle, |11,0\rangle\) and \(|11,1\rangle\) forming the so-called-computational basis. After initialization, the gate \(U\), with either the choice \(V = I\) or \(V = V^*\), as described above, was applied. For the \(V = I\) case we let the IBM compiler find a decomposition in elementary gates, which was then applied to the hardware, whereas in the case \(V = V^*\) the decomposition in Fig. \[5\] was directly applied. Finally, projective
measurement in the computational basis were performed, and the outcome recorded. This procedure was repeated \(N = 8192\) times for each initial state and choice of \(V\) to obtain the statistics \(p_{ij',k'}|_{ij,k}\) that the compound system ends up in state \(|ij',k'\rangle\) given that it was prepared in state \(|ij,k\rangle\). These data were error mitigated following a calibration performed before the experiments, accordingly to the standard procedure described in [14, 15].

The energy variations of the two subsystems and the total work were computed as

\[
\langle \Delta E_H \rangle = \sum_{i,i'} \tilde{E}^H_{ii'} p_{i'i} \left( \sum_i \tilde{E}^H_{ii} p_i - \sum_i \tilde{E}^H_{ii'} p_{i'} \right),
\]

\[
\langle \Delta E_C \rangle = \sum_{i,i'} \tilde{E}^C_{ii'} p_{i'i} \left( \sum_i \tilde{E}^C_{ii} p_i - \sum_i \tilde{E}^C_{ii'} p_{i'} \right),
\]

\[
\langle W \rangle = \langle \Delta E_H \rangle + \langle \Delta E_C \rangle,
\]

where \(i(i')\) is a short notation for the multi index set \(i, j, k(i', j', k')\). The symbols \(\tilde{E}^H, \tilde{E}^C\) denote, respectively the hot and cold subsystem eigenenergies, reading, for the cold subsystem \(E^C_{ij0} = -\hbar \omega_C / 2, E^C_{ij1} = \hbar \omega_C / 2\) and, for the hot subsystem \(\tilde{E}^H_{00k} = -\hbar \omega_H / 2, \tilde{E}^H_{11k} = \hbar \omega_H / 2, \tilde{E}^H_{10k} = \Delta / 2, \tilde{E}^H_{01k} = -\Delta / 2\), where \(\Delta = \hbar \omega_0 - \hbar \omega_2 \approx 0.13\) GHz is the detuning between between qubit 0 and qubit 2. Note that the actual hot subsystem Hamiltonian \(\tilde{H}_H = \sum_i \tilde{E}^H_i |i\rangle \langle i|\), differs from the ideal Hamiltonian \(\hat{H}_H\), Eq. (3), because of the non-null detuning \(\Delta\). For the initial distribution \(p_i\) we used the expression

\[
P_i = e^{-\beta \mu E^H_i - \beta C E^C_i} / (Z_H Z_C)
\]

with \(E^H_i, E^C_i\) the eigenvalues of the ideal Hamiltonian, Eq. [3]. We remark that this procedure amounts to create the initial bi-thermal preparation artificially, rather than physically, a method that is often used in quantum thermodynamics experiments, see e.g., [16].

The plots in Fig. 2 were obtained by looking at the signs of \(\langle \Delta E_H \rangle, \langle \Delta E_C \rangle, \langle W \rangle\), for both theory and experiment. The region \([H]\) is the region \(\langle \Delta E_H \rangle > 0, \langle \Delta E_C \rangle > 0\); the region \([E]\) is the region \(\langle W \rangle < 0\); the region \([A]\) is the region \(\langle \Delta E_H \rangle < 0, \langle \Delta E_C \rangle > 0, \langle W \rangle > 0\); the region \([R]\) is the region \(\langle \Delta E_C \rangle < 0\).

The final temperature of the cold qubit, reported in Fig. 3 was calculated according to the formula

\[
kT'_C = -\hbar \omega_C [\ln(Q/(1 - Q))]^{-1}
\]

with \(Q\) being the final population of state \(|1\rangle_C\) of the cold qubit, namely:

\[
Q = \sum_{ij} p_{ij1}, \quad p_{ij1} = \sum_i p_{i'i} p_{i'}.
\]

CONCLUSIONS

We have practically demonstrated a thermodynamic method to cool a qubit on a QPU at the expense of heating up other qubits. The method allowed, in our implementation, to cool a qubit on a QPU down to 52 mK.

We have shown that the method is limited by gate noise, accordingly, its performance can be greatly increased by decreasing the gate noise level. In principle the method may cool down qubits to arbitrarily low (but finite) temperature. The method requires that a quantum system with a larger resonant frequency than that of the qubit that needs to be cooled is available on the QPU, and be linked to said qubit. Such a system can be either a single qubit of larger level spacing, or, as demonstrated here, a compound of several qubits of same level spacings. Based on the reported results, it is desirable that QPU’s be equipped with “sacrificial” qubits, purposely designed to be mere heat dumpsters, that do not participate to actual computations, while instead being used to take the noise away from the “chosen” computational qubits, in order to improve their purity.

Acknowledgments

We acknowledge the use of IBM Quantum services for this work [17]. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team. In this paper we used ibmq.jakarta, which is one of the IBM Quantum Falcon r5.11H Processors. Andrea Solfanelli and Alessandro Santini acknowledge that their research has been conducted within the framework of the Trieste Institute for Theoretical Quantum Technologies (TQT).

---

[1] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum 2, 79 (2018).
[2] H. B. Callen, Thermodynamics: an introduction to the physical theories of equilibrium thermostatics and irreversible thermodynamics (Wiley, New York, 1960).
[3] E. Fermi, Thermodynamics (Dover, New York, 1956).
[4] S. Lloyd, Quantum-mechanical maxwell’s demon, Phys. Rev. A 56, 3374 (1997).
[5] H. Quan, Y. X. Liu, C. Sun, and F. Nori, Quantum thermodynamic cycles and quantum heat engines, Phys. Rev. E 76, 031105 (2007).
[6] A. E. Allahverdyan, R. S. Johal, and G. Mahler, Work extremum principle: Structure and function of quantum heat engines, Phys. Rev. E 77, 041118 (2008).
[7] M. Campisi, J. Pekola, and R. Fazio, Nonequilibrium fluctuations in quantum heat engines: theory, example, and possible solid state experiments, New J. Phys. 17, 035012 (2015).
[8] A. M. Timpanaro, G. Guarnieri, J. Goold, and G. T. Landi, Thermodynamic uncertainty relations from exchange fluctuation theorems, Phys. Rev. Lett. 123, 090604 (2019).
[9] R. Uzdin and R. Kosloff, The multilevel four-stroke swap engine and its environment, New J. Phys. **16**, 095003 (2014).

[10] L. Buffoni, A. Solfanelli, P. Verrucchi, A. Cuccoli, and M. Campisi, Quantum measurement cooling, Phys. Rev. Lett. **122**, 070603 (2019).

[11] A. Solfanelli, A. Santini, and M. Campisi, Experimental verification of fluctuation relations with a quantum computer, PRX Quantum **2**, 030353 (2021).

[12] A. Solfanelli, M. Falsetti, and M. Campisi, Nonadiabatic single-qubit quantum otto engine, Phys. Rev. B **101**, 054513 (2020).

[13] Elementary gates are $R_z(\theta)$, $\sigma_x$, $\sqrt{\sigma_x}$, CNOT, namely single qubit rotation of arbitrary angle $\theta$ around the $z$ axis, the single qubit $\sigma_x$ operator, its square root, and the entangling controlled-NOT operator among two connected qubits.

[14] A. Abbas, S. Andersson, A. Asfaw, A. Corcoles, L. Bello, Y. Ben-Haim, M. Bozzo-Rey, S. Bravyi, N. Bronn, L. Capelluto, A. C. Vazquez, J. Ceroni, R. Chen, A. Frisch, J. Gambetta, S. Garion, L. Gil, S. D. L. P. Gonzalez, F. Harkins, T. Imamichi, P. Jayasinha, H. Kang, A. h. Karamlou, R. Loredo, D. McKay, A. Maldonado, A. Macaluso, A. Mezzacapo, Z. Minev, R. Movassagh, G. Nannicini, P. Nation, A. Phan, M. Pistoia, A. Rattew, J. Schaefer, J. Shabani, J. Smolin, J. Stenger, K. Temme, M. Tod, E. Wanzambi, S. Wood, and J. Wootton., Learn quantum computation using qiskit (2020).

[15] A. Santini and V. Vitale, Experimental violations of leggett-garg’s inequalities on a quantum computer (2021), arXiv:2109.02507 [quant-ph].

[16] S. Hernández-Gómez, N. Staudenmaier, M. Campisi, and N. Fabbri, Experimental test of fluctuation relations for driven open quantum systems with an NV center, New J. Phys. **23**, 065004 (2021).

[17] Ibm quantum, https://quantum-computing.ibm.com/ (2021).