# Energy use in quantum data centers: Scaling the impact of computer architecture, qubit performance, size, and thermal parameters

Michael James Martin, Caroline Hughes, Gilberto Moreno, Eric B. Jones, David Sickinger, Sreekant Narumanchi, and Ray Grout

**Abstract**—As quantum computers increase in size, the total energy used by a quantum data center, including the cooling, will become a greater concern. The cooling requirements of quantum computers, which operate at temperatures near absolute zero, are determined by computing system parameters, including the number and type of physical qubits, the packaging efficiency of the system, and the split between circuits operating at cryogenic temperatures and those operating at room temperature. When combined with thermal system parameters such as cooling efficiency and cryostat heat transfer, the total energy use can be determined using a first-principles energy model. These models show that cooling of quantum computers differs in two fundamental ways from conventional data centers: (1) the energy required for cooling is much greater than the energy required for computation, and (2) the cooling loads are sensitive to the computational architecture. The temperature requirements for different qubit types can change energy requirements by orders of magnitude. Power use and computational power, as quantified by quantum volume, are analytically correlated. Approaches are identified for minimizing energy use in integrated quantum systems relative to computational power. Designing a sustainable quantum computer will require both efficient cooling and system design that minimizes cooling requirements.

**Index Terms**—Cryogenics, data center integration, energy efficiency, quantum computing, sustainability

## 1 Introduction

QUANTUM computers capable of solving the scientific and cryptography problems that drive research interest in the technology will be much larger than current systems. For instance, a computer capable of using Shor's algorithm for current cryptography problems will require on the order of $10^6$ physical (as opposed to computational) qubits [1],while using quantum computing to simulate nitrogen fixation, a computational chemistry problem widely used as an example of the scientific potential of quantum computing, will require on the order of $10^5$ to $10^8$ physical qubits, depending on the error rate [2]. This compares to the 53-qubit machine recently used to sample the output of a pseudorandom quantum circuit [3]. As computers scale up, their power usage will also increase.

Because of the extremely low temperatures (mK to K) required by the majority of proposed quantum computing systems, quantum computers require very different themal management approaches from conventional data centers. The cooling system must move all power dissipated by electronics operating at low temperatures from a cryostat, in addition to heat entering the cryostat from the ambient environment. Because of this, the power consumption of the cooling system is likely to be much greater than the power consumption of the electronics. This has already been demonstrated with existing quantum annealing systems [4] and was identified as a major challenge in a recent National Academies of Sciences, Engineering, and Medicine overview of the state of quantum computing [5]. These challenges are not limited to quantum computing; they have been identified as a limiting factor for incorporating other cryogenic components into conventional computing architectures [6], [7]. These changes in the balance of power supplied to electronics compared to the power for cooling mark a significant change from conventional computing, where cooling requirements are typically 10%–30% of the power used by the electronics [8]. Identifying the combined cooling and electronics power requirements of these systems is therefore crucial to engineering quantum computers to decrease power consumption and in predicting the impact of quantum systems on national and global energy use [9], [10].

## 2 Energy Model Development

### 2.1 Model Derivation

The majority of quantum information systems are *heterogeneous* computer systems composed of (1) qubits and associated circuits and sensors operating at a temperature $T_c$ in a cryostat, and (2) external control and communications circuits operating at a temperature $T_H$ at or above the ambient temperature $T_o$, as shown in Fig. 1. [11] The external environment serves as a heat reservoir. There are two electronic power loads: $P_1$ is the power used by the qubits and other circuits operating at cryogenic temperatures, and $P_2$ is the power used by the external circuits. These terms can be combined to form a total electronic power load $P_c$. This model does not describe photonic quantum computers, which operate either entirely at room temperature [12] or have individual cryogenic components deployed throughout the system [13].

Scaling these power loads based on the size of the system is complicated by the range of electronics that will operate in the system. In addition to the qubits and their associated control and readout electronics, most quantum

• *M. J. Martin, C. Hughes, G. Moreno, E. B. Jones, D. Sickinger, S. Narumanchi, and R. Grout are with the National Renewable Energy Laboratory, Golde, CO, 80401. E-mail: Michael.martin@nrel.gov.*

xxxx-xxxx/0x/$xx.00 © 200x IEEE Published by the IEEE Computer Society

systems will contain a set of gates and other associated electronics. Their power use must also be taken into account.

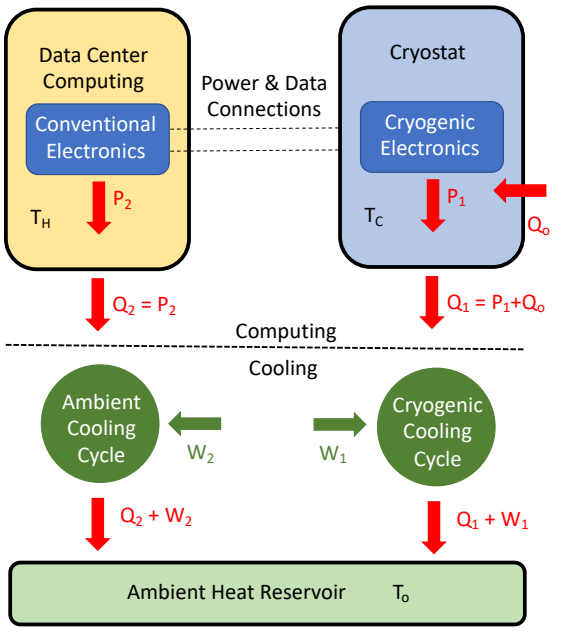


Fig. 1: Quantum Computing Data Center Model and Energy Movement

A reasonable first approximation is that the total power load will scale as some power $a$ of the number of physical qubits in the computer $n_p$:

$$P_c \propto (n_p)^a. \tag{1}$$

The most likely scaling is a linear scaling, where $a$ is equal to one. In this case $P_c$ will be equal to $n_p q$, where $q$ is the value of power used per qubit:

$$P_c = n_p q \tag{2}$$

The power use can then, at both the qubit and overall system level, be split between the power used in the cryostat and the ambient environment. Because all power used by these electronics will ultimately be dissipated as heat, either inside of or outside of the cryostat, it is appropriate to think of these values as heat loads. The power used per qubit $q$ will be split between the power used per qubit by the electronics inside the cryostat ($q_1$), and the power used per qubit by the electronics at ambient temperature ($q_2$):

$$q = q_1 + q_2. \tag{3}$$

The parameter $\phi$ can then be used define how the total energy is split between the circuits operating at cryogenic temperatures and the external circuits:

$$q_1 = \phi q, \tag{4}$$

$$q_2 = (1-\phi)q, \tag{5}$$

$$P_1 = \phi P_c, \tag{6}$$

$$P_2 = (1-\phi)P_c. \tag{7}$$

A value of $\phi$ of 1 then corresponds to a system where all electronics operate at cryogenic temperatures, while a value of 0 represents a system where there is no energy usage in quantum systems.

To maintain the cryogenic electronics at a temperature $T_c$ that is below the ambient temperature $T_o$, the cryogenic cooling system must remove both the power dissipated by the circuits in the cryostat, equal to $P_1$, and the heat entering the chamber $Q_o$. $Q_o$ will depend on the cryogenic chamber area $A$, the heat transfer coefficient $U$, and the temperatures $T_c$ and $T_o$:

$$Q_o = UA(T_o - T_c). \tag{8}$$

Because $T_c$ is much less than $T_o$, $Q_o$ can be approximated as being equal to $UAT_o$. Two additional assumptions allow this value to be rewritten in terms of the number of qubits. The first assumption is that $A$ is equal to $C_1 V^{2/3}$, where $V$ is the volume of the chamber and $C_1$ is a geometric constant. (For a cube, $C_1$ is 6; for a sphere, $C_1$ is $\sqrt[3]{36\pi}$.) The second assumption is that volume can be re-written as the product of the number of qubits and a volume per qubit $v_q$, so that $V$ is equal to $n_p v_q$. Note that $v_q$ is not the volume of the qubit, but the volume required to house a qubit when all interconnections and other geometric constraints are taken into account. The heat load $Q_o$ now depends on the number of qubits:

$$Q_o = UC_1 T_o v_q^{2/3} n_p^{2/3}. \tag{9}$$

The total heat $Q_1$ that must be removed from the cryogenic chamber $Q_c$ is the sum of $P_1$ and $Q_o$:

$$\begin{aligned} Q_1 &= P_1 + Q_o = n_p q_1 + UC_1 T_o v_q^{2/3} n_p^{2/3} \\ &= n_p q_1 \left(1 + \left(UC_1 T_o v_q^{2/3}/q_1\right) n_p^{-1/3}\right) \\ &= n_p q_1 \left(1 + \beta n_p^{-1/3}\right) \\ &= \phi n_p q \left(1 + \beta n_p^{-1/3}\right), \end{aligned} \tag{10}$$

where the parameter $\beta$ equal to $UC_1 T_o v_q^{2/3}/q_1$ captures the importance of heat transfer into the chamber relative to the electronic heat generation within the chamber. This scaling has one limitation: it assumes that power of the electronics in the cryostat is not negligible compared to the heat entering the cryostat, which would lead to a value of $\beta$ of infinity. This situation could occur if the packaging is not optimized to minimize heat entering the cryostat. Conversely, a perfectly insulated system would have a $\beta$ of zero.

The work $W_1$ required to remove heat from the cryogenic chamber is calculated by dividing the heat by the coefficient of performance, or $COP(T_C)$ [14]. This value can be expressed as a product of the value of the Carnot efficiency $COP(T_C)|_C$ and a correction factor $\eta_c$:

$$COP(T_c) = \frac{Q_1}{W_1} = \eta_c COP(T_c)|_C = \eta_c \left(\frac{T_c}{T_o - T_c}\right). \tag{11}$$

Combining (10) and (11) gives a value for $W_1$:

$$\begin{aligned} W_1 &= n_p q_1 \left(\frac{1 + \beta n_p^{-1/3}}{\eta_c COP(T_c)|_C}\right) \\ &= \phi n_p q \left(\frac{1 + \beta n_p^{-1/3}}{\eta_c COP(T_c)|_C}\right). \end{aligned} \tag{12}$$

The heat that must be removed from the external electronics $Q_2$ is equal to $P_2$, which is equal to $n_p q_2$, or $(1-\phi) n_p q$. Because the cooling of room-temperature electronics is not performed using a direct refrigeration cycle, the cooling energy cost $W_2$ is characterized using a cooling figure of

merit $FOM(T_o)$, which is defined similarly to a coefficient of performance:

$$W_2 = \frac{P_2}{FOM(T_o)} = \frac{n_p q_2}{FOM(T_o)} = \frac{(1-\phi) n_p q}{FOM(T_o)}. \quad (13)$$

The total energy required to cool the system $W_S$ is then the sum of $W_1$ and $W_2$:

$$W_S = n_p q_1 \left( \frac{1+\beta n_p^{-1/3}}{\eta_c COP(T_c)|_C} \right) + \frac{n_p q_2}{FOM(T_o)} = n_p q \left[ \phi \left( \frac{1+\beta n_p^{-1/3}}{\eta_c COP(T_c)|_C} \right) + \left( \frac{1-\phi}{FOM(T_o)} \right) \right]. \quad (14)$$

The total power required by the system $P_T$ is then the sum of $P_C$ and $W_S$:

$$P_T = n_p q \left[ 1 + \phi \left( \frac{1+\beta n_p^{-1/3}}{\eta_c COP(T_c)|_C} \right) + \left( \frac{1-\phi}{FOM(T_o)} \right) \right]. \quad (15)$$

A list of all system parameters, with their definitions and units, is given as table 1.

Table 1. Quantum computing system parameters

| Variable | Units | Physical Significance |
|---|---|---|
| $C_1$ | n/a | Cryostat geometric constant |
| $COP(T_C)\|_C$ | n/a | Carnot Coefficient of Efficiency (COP) of cryogenic cooling system. |
| $FOM(T_o)$ | n/a | Figure of Merit for cooling of ambient electronics |
| $n_p$ | n/a | Number of qubits |
| $q$ | W | Heat/power load per qubit |
| $T_c$ | K | Operating temperature in the cryostat (qubit operating temperature) |
| $To$ | K | Ambient temperature |
| $U$ | W/m²-K | Cryostat heat transfer coefficient |
| $v_q$ | m³ | Cryostat volume per qubit |
| $\beta$ | n/a | Scaling parameter comparing cryostat external heat transfer to internal power dissipation, equal to $UC_1T_o v_q^{2/3}/\phi q$ |
| $\eta_c$ | n/a | Fraction of Carnot efficiency achieved by cryogenic cooling system. |
| $\phi$ | n/a | Split in power use between cryogenic and non-cryogenic electronics. |

### 2.2 Derivation of scaled energy usage

Because $q$ is the most difficult of these terms to estimate for future quantum information systems, it is useful to look at the ratio of the energy use $P_T$ to the total computational power usage $q$ required for a single qubit, which is defined as $P_T^*$:

$$P_T^* = n_p \left[ 1 + \phi \left( \frac{1+\beta n_p^{-1/3}}{\eta_c COP(T_c)|_C} \right) + \left( \frac{1-\phi}{FOM(T_o)} \right) \right]. \quad (16)$$

Note that $q$ is not completely eliminated from the equation: it still appears in the scaling of $\beta$. However, if $\beta$ is seen fundamentally as a scaling of the importance of heat dissipation within the cryostat to heat entering the cryostat, this simplification remains useful and appropriate. The identical scaling can be applied to $P_1$, $P_2$, $Q_c$, $W_2$, $W_3$, $P_S$, and $P_c$.

A range of metrics are available for quantifying the energy use and sustainability of data centers [15]. The efficiency of conventional data centers is frequently characterized by the power usage efficiency, or PUE [16] . This metric is the ratio of the total energy used by the data center divided by the energy used by computation. If PUE is calculated instantaneously, or if all loadings are constant, power may be used instead of energy, and PUE will be equal to the total power $P_T$ divided by the power going to computation $P_C$. If other possible loadings such as lighting are excluded, the PUE for a quantum system can be found simply by dividing (15) by a value of $P_C$ of $n_p q$:

$$PUE = 1 + \phi \left( \frac{1+\beta n_p^{-1/3}}{\eta_c COP(T_c)|_C} \right) + \left( \frac{1-\phi}{FOM(T_o)} \right). \quad (17)$$

## 3 Review of Representative Physical Parameters

Though there is considerable uncertainty in how quantum computers will evolve, enough is known about key parameters to allow initial assessments of energy use in quantum computing. As shown in Fig. 2, several key parameters can already be identified for quantum systems. Two types of qubits have already been integrated into quantum computers: superconducting qubits and ion-trap qubits; each has a range of reported operating temperatures [17], [18], [19], [20], [21], [22]. The temperature ranges for silicon and diamond qubits, which have been demonstrated as stand-alone components, are also shown [23], [24], [25], [26], [27]. The total operating range for existing qubit technologies ranges from 0.01 K to 10 K. The type of qubit used will determine $T_c$ and $COP_c(T_c)$ for the system.

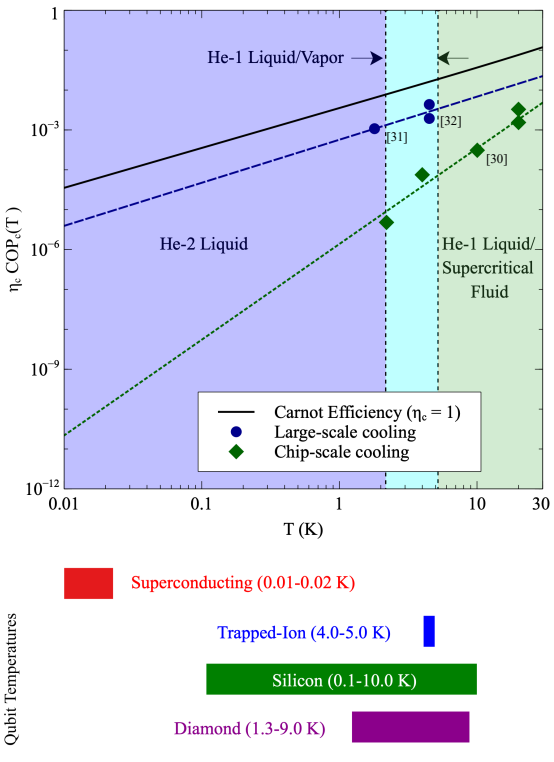

Fig. 2. Quantum data center operating temperatures and cryogenic cooling efficiencies with relevant helium temperature ranges.

At 10 mK, distillation refrigeration using helium and laser cooling are currently the only viable cooling technologies. Laser cooling has been successfully used to maintain ion-trap qubits at 4 K [20]. However, in general, laser cooling systems have low thermodynamic efficiencies: though theoretically up to 20% of the Carnot efficiency is achievable, in practice only 3% has been achieved [28]. Distillation refrigerators have been used on a range of large- and small-scale scientific instrumentation to produce low-temperature helium II. Figure 2 shows the values of COP for both large-scale [29], [30] and chip-scale [31], [32] cooling systems. While chip-scale systems show a large drop in efficiency as the temperature decreases, the efficiency of large-scale systems appears to remain between 12% and 18% of the Carnot efficiency, even when the system temperature is below the boundary between He-1 and He-2 (also referred to as helium I and helium II).

For purposes of initial discussions of the impact of qubit type, heat transfer, and system size, $\eta_c$ is estimated to be 0.15 for the entire range of 10 mK–10.0 K based on the large-scale cooling results shown in Fig. 2. The impact of this assumption will be examined by varying the cooling system parameter $\eta_c$ for representative systems after the impact of system parameters is examined. The value for the $FOM(T_o)$ can be estimated based on experience with conventional data centers. An efficiently designed data center will typically have a PUE of 1.2, which corresponds to a value for $FOM(T_o)$ of 5.0, which is used in this study. The remaining thermal parameter, $U$, is estimated to be 0.3–0.8 mW/m$^2$-K, based on experience with existing cryostats used in the Large Hadron Collider [33]. This value should be modified based on interconnects. A coaxial cable will allow on the order of 1 mW of heat loss between a 3K cryostat and 300K ambient, corresponding to a UA on the order of 3 μW/K. An optical fiber will allow on the order of 0.01 mW over the same temperature range, corresponding to a UA on the order of 0.03 μW/K [34].

The only computer architecture parameter that can be identified that would be relevant for future systems is the number of physical qubits $n_p$ to perform useful computational tasks. The required power for electronics per qubit $q$ that will be seen in future quantum systems is challenging to estimate. The Landauer limit suggests the value of modifying information will be proportional to, or will at least increase with, $T_c$ [35] Though this limit has been extended to quantum systems [36], it does not reflect the power usage of related electronics either inside or outside the cryostat. Additional work suggests a lower bound for energy use per qubit [37], while other research suggests that proper manipulation of information in quantum systems can lower the total energy used [38]. However, the value of $q$ for future quantum systems cannot be reliably determined. Similarly, the required cryostat volume per qubit, $v_q$, cannot be characterized for a rapidly changing technology.

Though $\phi$ is bounded between 0 and 1, it may vary by orders of magnitude within that range in future systems. The value of $\phi$ reflects not only the qubit type and the power required by associated sensors, control systems, and electronics, but also a design decision about where to place conventional silicon electronics used to control the qubits. Placing these electronics inside the cryostat may simplify integration and decrease latency [39], but doing so leads to reliability challenges for CMOS electronics [40]. Doing so also creates a cooling penalty that has already been observed in laboratory experiments [41]. These reliability and power concerns have led to the development of specialized conventional CMOS circuits for use with quantum electronics [42], [43], [44], [45], [46], [47], [48].

The design decisions made to move electronics in or out of the cryostat, increasing or decreasing $\phi$, will impact other parameters. Increasing $\phi$ is likely to decrease $q$, as low-temperature electronics may consume less power than room-temperature electronics to perform the same function. Increasing $\phi$ will also increase $v_q$ and therefore $A$ and $\beta$ as more cryostat volume and surface heat transfer area will be required per qubit as these electronics are added. Finally, increasing $\phi$ may decrease the need for interconnects, reducing the effective value of $U$. These relationships can be evaluated in depth for an individual technology. For purposes of this paper, they will be considered as independent parameters.

## 4 Energy Use Results

In spite of the inability to assign values to many parameters for quantum systems, the power used for integrated quantum systems can still be studied using values of $P^*$ and $PUE$, in which $q$ and $n_p q$ respectively are factored out. The impact of different values of $\phi$ and of $\beta$, as well as system operating temperatures and size, can then be explored.

### 4.1 Ideal Results: No heat transfer

The limiting case for maximum system efficiency will be a perfectly insulated cryostat, which corresponds to a value of $\beta$ of zero. In this case, $P^*$ is proportional to $n_p$, and the PUE is independent of the size of the computer. Figure 3 shows the PUE as a function of $\phi$ for the two types of qubits that have been successfully integrated into quantum computers for the range of operating temperatures $T_c$ demonstrated for both qubit types. The results show that the power usage is extremely sensitive to values of $\phi$, which is consistent with previous findings that moving electronics inside the cryostat tended to exceed the cooling capacity of laboratory experiments. For a superconducting qubit, the energy used for cooling is much larger than the electronics energy usage at values of $\phi$ ranging from $10^{-7}$ to $10^{-6}$, depending on the operating temperature. The cooling requirements of a system operating at 10 mK are roughly twice those of a system operating at 20 mK. For an ion-trap system, the cooling power requirements dominate the electronics power requirements at values of $\phi$ of around $10^{-3}$. There is also much less variation over the range of possible operating temperatures.

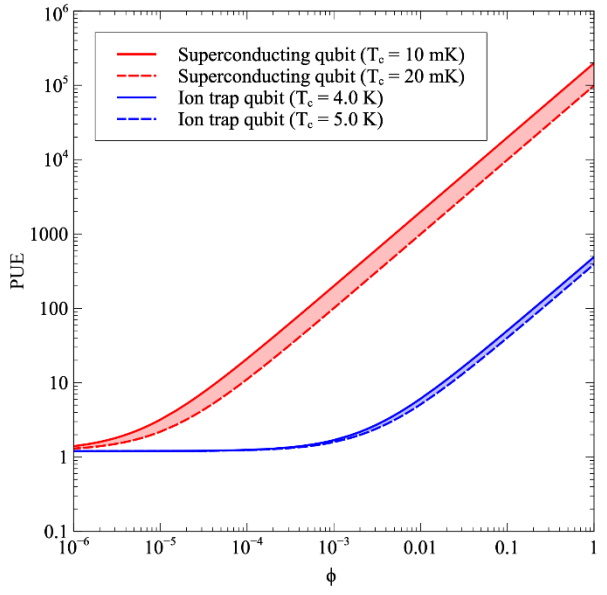


Fig. 3: Quantum Data Center Power Usage Efficiencies (PUEs) with no external heat transfer for qubit types currently integrated into quantum computers.

Figure 4 shows the range of possible PUEs for silicon and diamond qubits. Because silicon qubits have been demonstrated over temperatures from 0.1 to 10.0 K, the potential values for PUE range over three orders of magnitude. Depending on the operating temperature, the cooling power usage becomes much larger than the electronics power usage anywhere from a value of $\phi$ of $10^{-5}$ to $10^{-3}$. For diamond qubits, the range of potential power usage is much narrower. These results show that the type of qubit used, the operating temperature, and how the electronics are integrated all dramatically impact the power required to cool quantum systems.

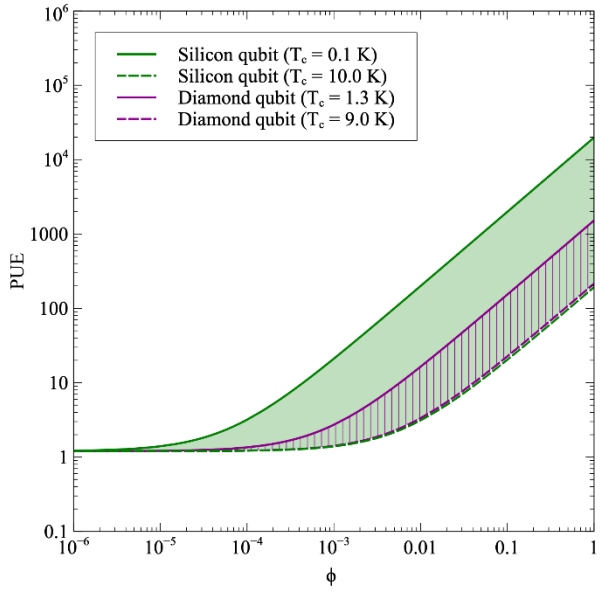


Fig. 4: Quantum Data Center Power Usage Efficiencies (PUEs) with no external heat transfer for qubit types currently demonstrated in laboratory environments.

Figure 2 shows that there is a range of values for cooling system efficiency, with large-scale helium cooling systems proving a value of $\eta_c$ of around 0.15, while chip-scale helium cooling systems will have values of around 0.01. Laser cooling systems will be below 0.03. Fig. 5 shows how different values of cooling system efficiency will impact the PUE of a superconducting qubit system operating at 15 mK. These results show that cooling system efficiency will have a large impact on the system PUE. This impact will increase as more electronics are moved into the cryostat. Fig. 6 shows similar results for a trapped ion system at operating at 4.5 K.

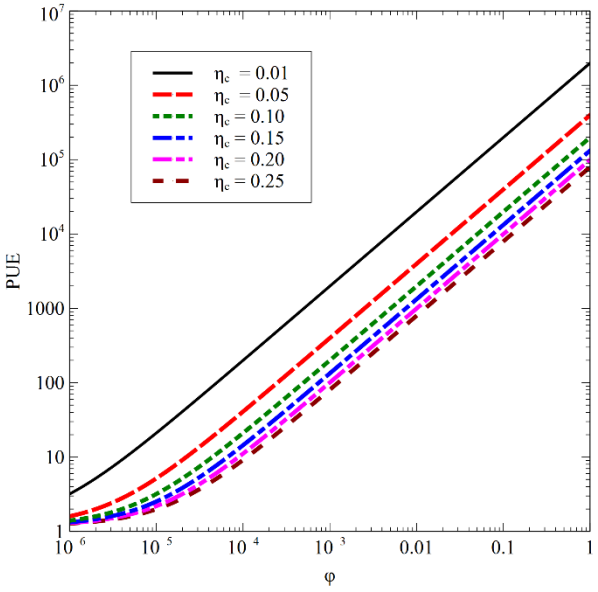


Fig. 5: Quantum Data Center Power Usage Efficiencies (PUEs) with no external heat transfer for superconducting qubits for different cooling system efficiencies.

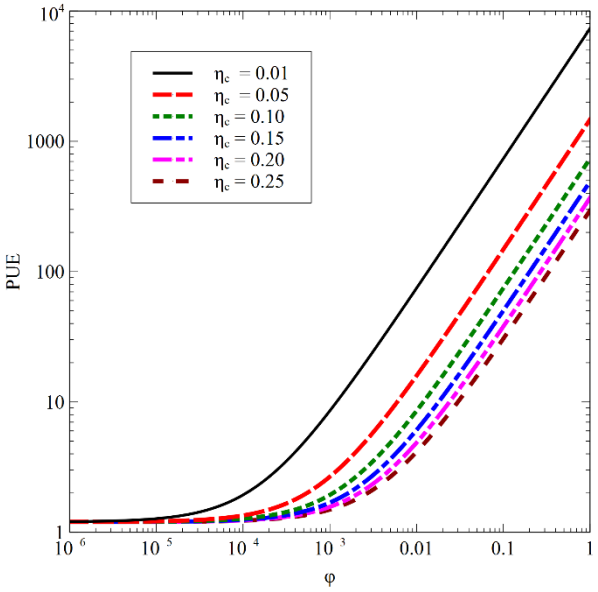


Fig. 6: Quantum Data Center Power Usage Efficiencies (PUEs) with no external heat transfer for trapped ion qubits for different cooling system efficiencies.

### 4.2 Results for systems with heat transfer

For real systems, the heat transfer term $\beta$ will not be zero, meaning the PUE will no longer be independent of the system size, and $P_T$ and $P_T^*$ will no longer be linear functions of the number of qubits. For any given value of $\phi$, $\beta$, or $n_p$, the shares of the power used (1) between the electronics $F_e$, (2) to cool the low-temperate electronics inside the cryostat $F_{lte}$, (3) to maintain the cryostat at temperature because of external heat transfer $F_o$, and (4) for external cooling $F_{ext}$ can be determined. Each of these can be written as a fraction of the overall power usage as found in Equation (16):

$$F_e = \frac{1}{\left[1+\phi\left(\frac{1+\beta n_p^{-1/3}}{\eta_c COP(T_c)|_C}\right)+\left(\frac{1-\phi}{FOM(T_o)}\right)\right]}, \quad (18.a)$$

$$F_{lte} = \frac{[\phi/\eta_c COP(T_c)|_C]}{\left[1+\phi\left(\frac{1+\beta n_p^{-1/3}}{\eta_c COP(T_c)|_C}\right)+\left(\frac{1-\phi}{FOM(T_o)}\right)\right]}, \quad (18.b)$$

$$F_o = \frac{\left[\phi\beta n_p^{-1/3}/\eta_c COP(T_c)|_C\right]}{\left[1+\phi\left(\frac{1+\beta n_p^{-1/3}}{\eta_c COP(T_c)|_C}\right)+\left(\frac{1-\phi}{FOM(T_o)}\right)\right]}, \quad (18.c)$$

$$F_o = \frac{[(1-\phi)/FOM(T_o)]}{\left[1+\phi\left(\frac{1+\beta n_p^{-1/3}}{\eta_c COP(T_c)|_C}\right)+\left(\frac{1-\phi}{FOM(T_o)}\right)\right]}. \quad (18.d)$$

Because of the complexity of the parameter space, the current results are limited to a value of $T_c$ of 15 mK, which represents a superconducting qubit system, and a value of $T_c$ of 4.5 K, which represents an ion-trap qubit system. For each case, values of $\phi$ of 0.1, 0.001, and 0.00001 with values of $\beta$ of 0, 0.01, 0.1, 1.0, 10.0, and 100.0 are considered for computers ranging in size from 10 to 1,000,000 qubits.

The results for a value of $T_c$ of 15 mK are shown in Fig. 7. The resulting $P_T^*$ and PUE values show that the external heat transfer only changes the power usage when $\beta$ is greater than 1. They also show the effect of external heat transfer diminishes as the number of physical qubits increases. The magnitude of these effects appears to also depend on how the electronic power usage is split between external electronics and the cryostat electronics.

As shown in Fig. 7(c), for a system with a $n_p$ of 1,000 qubits and a $\phi$ of 0.1 operating at 15 mK, the power for cooling the electronics $F_{lte}$ inside of the cryostat is more than 99% of the total power usage for values of $\beta$ of up to 0.01. At a value of $\beta$ of 1, the share of power going to cool electronics drops to around 90%, and the share of power going to cooling for external heat transfer $F_o$ increases to around 10%. At a value of $\beta$ of 10, the two cooling needs are equal, while at a $\beta$ of 100, cooling for external heat transfer dominates. The combined power needs for all electronics and external cooling $F_e$ + $F_{ext}$ never exceed 0.01% of the total power used. At a value of $\phi$ of 0.001, the power distributions are similar.

However, when $\phi$ is reduced to 0.0001, other forms of power usage become significant. When there is no external heat transfer, cooling the electronics in the cryostat requires approximately 50% of the total power usage, while the electronics themselves use 40% of the total power, and external cooling uses approximately 10% of the total power. However, as $\beta$ increases, the power required to remove heat from the cryostat because of external heat transfer dominates the power usage.

Fig. 8 shows the same information for an ion-trap system operating at a temperature of 4.5 K. Because $COP_c(T_c)$ is approximately 300 times higher at 4.5 K than at 15 mK, the values for $P_T^*$ and PUE decrease significantly. The distribution of energy usage within the system also changes. At a value of $\phi$ of 0.1, the combined cooling loads for the electronics in the cryostat and external heat transfer into the cryostat are still dominant. However, at a value of $\beta$ of 0, electronics account for 2.3% of total power use. This decreases to 1.1% of total power usage as $\beta$ increases to 10.0 and to less than 0.2% at a value of $\beta$ of 100.0, which reflects the decreased cost of cooling in the overall power budget.

As $\phi$ decreases to 0.001, the total cost of maintaining the low-temperature circuits at 4.5 K decreases dramatically. For values of $\beta$ of 10 or less, the electronics become the dominant consumer of power in the system. When $\phi$ decreases to 0.00001, the cost of maintaining low-temperature circuits at 4.5 K is only significant for values of $\beta$ greater than 10.0, and even then, it is less than 4% of total power usage. For these cases, $P_T^*$ and PUE do not change significantly as $\beta$ increases. The combination of reducing electronics in the cryostat to a minimal level and operating at higher temperatures lead to power usage that mimics a conventional computer. However, this configuration may compromise the ability of the conventional electronics to control the qubits effectively.

## 5 Quantum Volume Versus Power Scaling

In conventional computing, the power of a processor scales directly with the number of transistors, and the power of a parallel computer scales directly as the number of processors times some parallelization efficiency. For linear algebra problems, this can be quantified using the time required to solve standardized problems such as the commonly used LINPACK benchmark [49]. While there is no universally accepted equivalent for quantum computing, the concept of "quantum volume," which has been quantified for systems, has gained some acceptance [50], [51].

The quantum volume of a processor $V_Q$ quantifies the largest random quantum circuit of equal width and depth for which a heavy output distribution can be generated with at least probability 2/3. In the limit of independent stochastic errors, it depends on the number of qubits $n_p$ and the effective error rate $\varepsilon_{eff}$. The effective error will depend on the computing architecture, the noise of the qubits, and the total number of qubits, and is therefore not completely independent of $n_p$. In this limit quantum volume for a system of $n_p$ qubits with an effective error rate rate $\varepsilon_{eff}$ is approximated by:

$$V_Q = \max_{n<n_p}\left(min\left[n, \frac{1}{n\varepsilon_{eff}(n)}\right]^2\right). \quad (19)$$

This expression shows that the quantum volume is

This article has been accepted for publication in IEEE Transactions on Sustainable Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2022.3190242



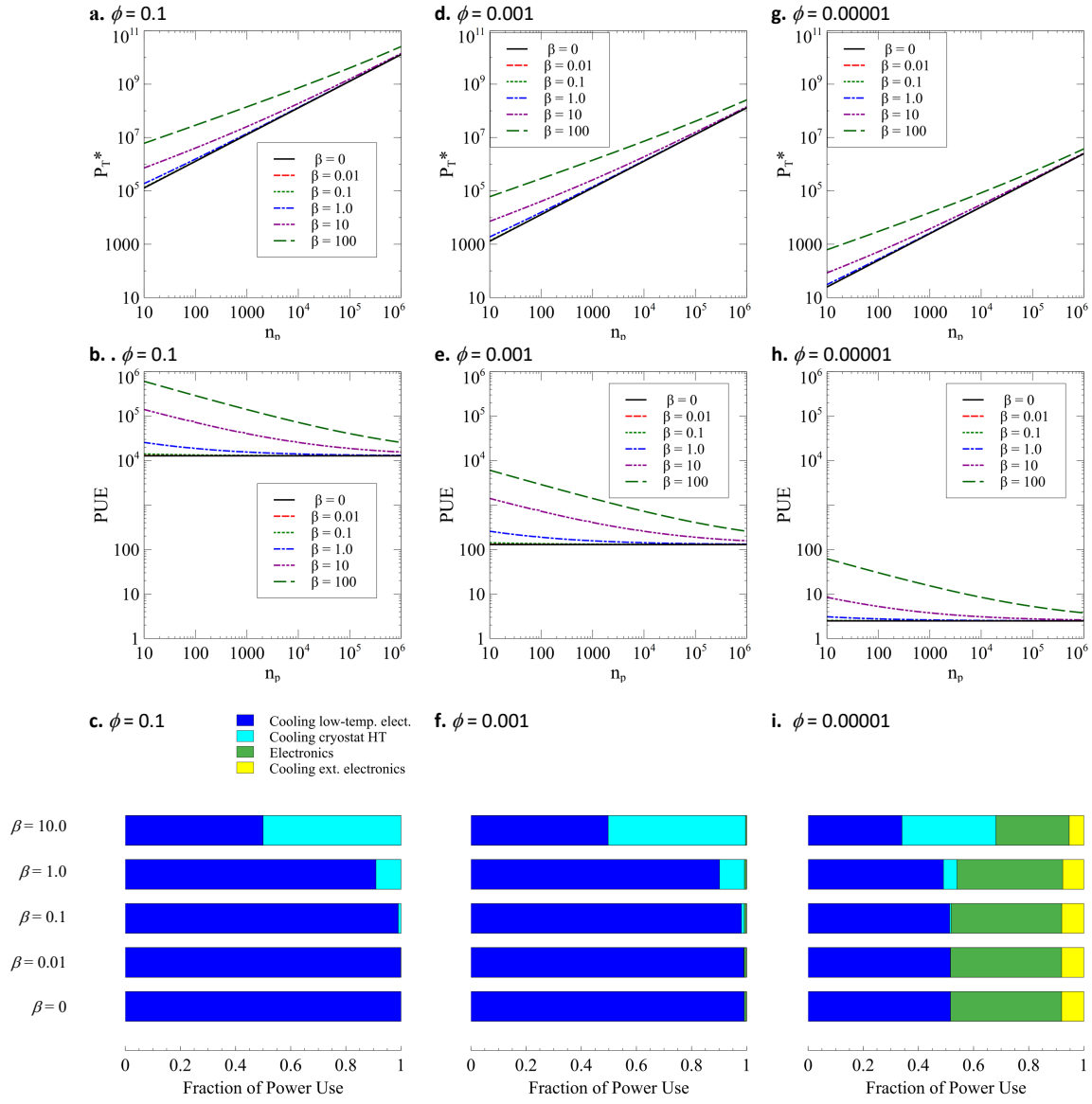


Fig. 7. Power usage, power usage efficiency (PUE), and power distribution for a quantum computer operating at a temperature of 15 mK. **a, b** $P_T^*$ and PUE for a system with $\phi$ of 0.1 and assorted values of $\beta$ as a function of $n_p$. **c** Power distribution for a system with $n_p$ of 1,000 and $\phi$ of 0.1. **d, e** $P_T^*$ and PUE for a system with $\phi$ of 0.001 and assorted values of $\beta$ as a function of $n_p$. **f** Power distribution for a system with $n_p$ of 1,000 and $\phi$ of 0.001. **g, h** $P_T^*$ and PUE for a system with $\phi$ of 0.00001 and assorted values of $\beta$ as a function of $n_p$. **i** Power distribution for a system with $n_p$ of 1,000 and $\phi$ of 0.00001.

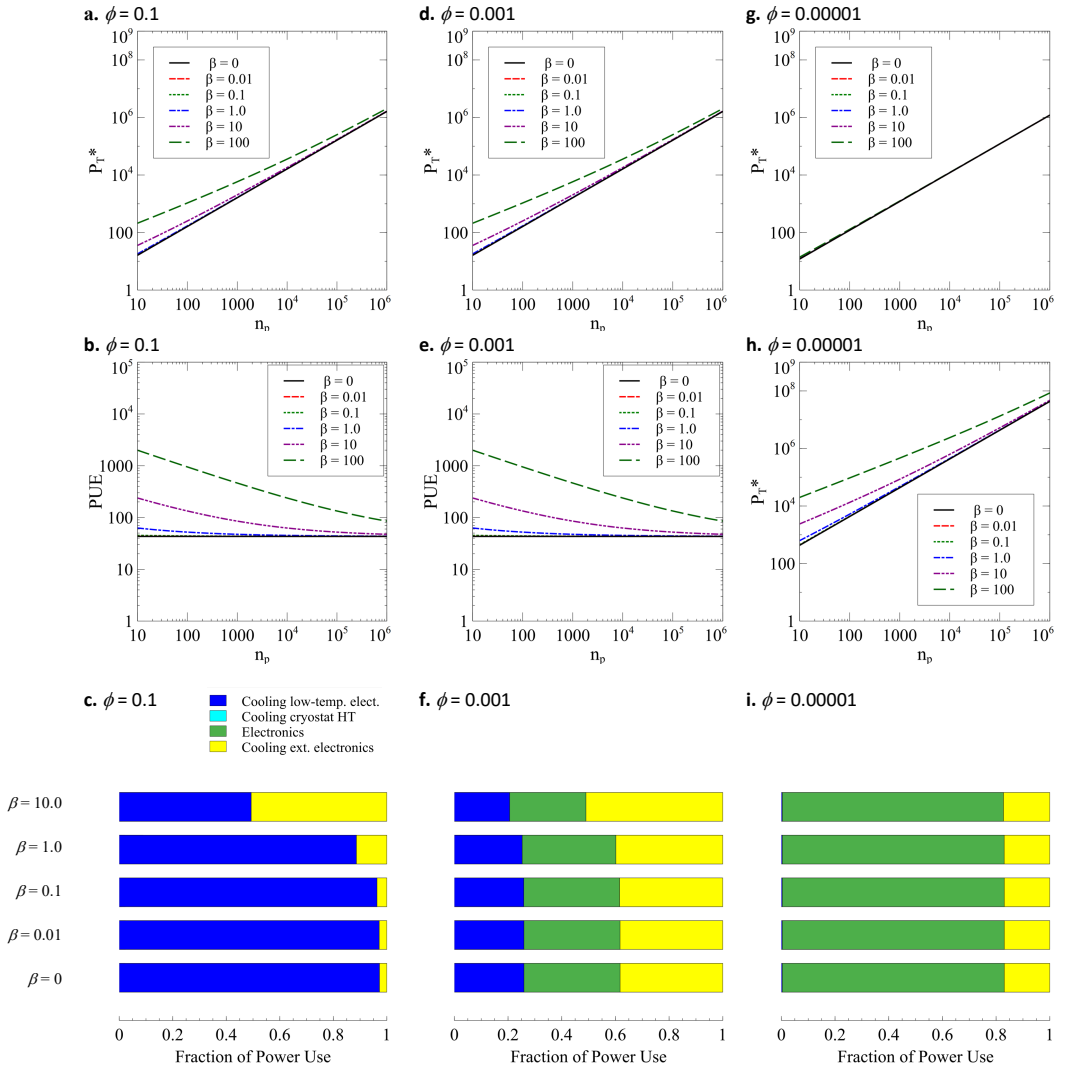


Fig 8. Power usage, power usage efficiency (PUE), and power distribution for a quantum computer operating at a temperature of 4 K.**a, b** $P_T^*$ and PUE for a system with $\phi$ of 0.1 and assorted values of $\beta$ as a function of $n_p$. **c** Power distribution for a system with $n_p$ of 1,000 and $\phi$ of 0.1. **d, e** $P_T^*$ and PUE for a system with $\phi$ of 0.001 and assorted values of $\beta$ as a function of $n_p$. **f** Power distribution for a system with $n_p$ of 1,000 and $\phi$ of 0.001. **g, h** $P_T^*$ and PUE for a system with $\phi$ of 0.00001 and assorted values of $\beta$ as a function of $n_p$. **i** Power distribution for a system with $n_p$ of 1,000 and $\phi$ of 0.00001.

This expression shows that the quantum volume is proportional to $n_p^2$ until the threshold value of $n_{peak}$ equal to $\sqrt{1/\varepsilon_{eff}(n)}$ is reached. At this point, adding additional qubits to the system will not increase the computational volume. Scaling quantum volume for proposed calculations is still a subject of considerable uncertainty. A processor and circuit with quantum volume of approximately 2000 ($V_o = 43^2$) would require roughly 140 Terabytes of memory to store the wavefunction for Schrodinger evolution, roughly the largest instance that would fit on the Jülich supercomputer [12] . The number of qubits required for Shor's algorithm, or chemical calculations, corresponds to a quantum volume on the order of $10^6$ or greater. Figure 9 shows $V_Q$ as a function of the number of qubits for various error rates. These effective error rates are generally far below those seen in current machines: current two-qubit error rates for state-of-the-art superconducting processors with tens of physical qubits are in the range of 0.36%–1.1% [3], [42], while two-qubit error rates for trapped ion processors with a few to roughly a dozen physical qubits are in the range of 0.79%–2.5% [52], [53].

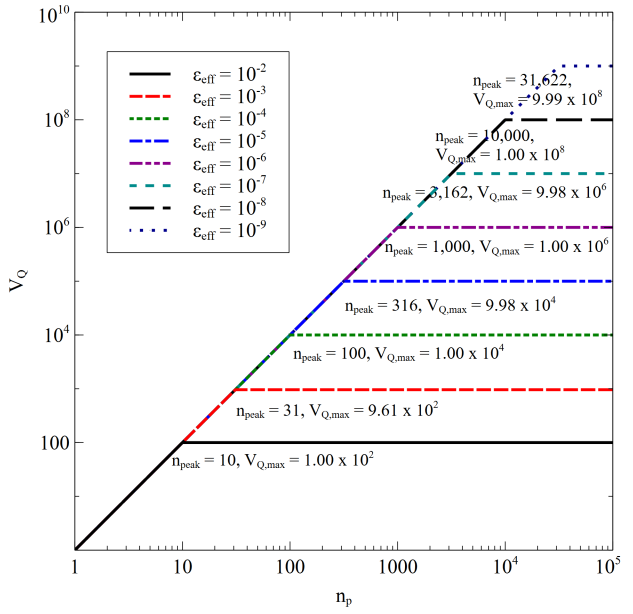


Fig. 9. Quantum volume as a function of qubits and effective noise

As long as the number of qubits is below $n_{peak}$, the quantum volume will be proportional to $n_p^2$, and equation (16) for the scaled power use can be re-written as:

$$P_T^* = V_Q^{1/2}\left[1 + \left(\frac{\phi}{\eta_c COP(T_c)|_C}\right) + \left(\frac{1-\phi}{FOM(T_o)}\right)\right] + Q_V^{1/6}\left[\frac{\phi\beta}{\eta_c COP(T_c)|_C}\right]. \quad (20)$$

As the limit of a large system, or negligible heat transfer is reached, the power required to operate a quantum computer scales with $V_Q^{1/2}$.

The scaled power use as a function of quantum volume for a system with a $\phi$ of 0.001 operating at 0.015 K is shown as Fig. 10. The maximum values of $Q_V$ for different error rates are shown as vertical lines. The large system limit, where cryostat heat transfer can be ignored, is generally above the limit imposed by quantum volume considerations scalings.

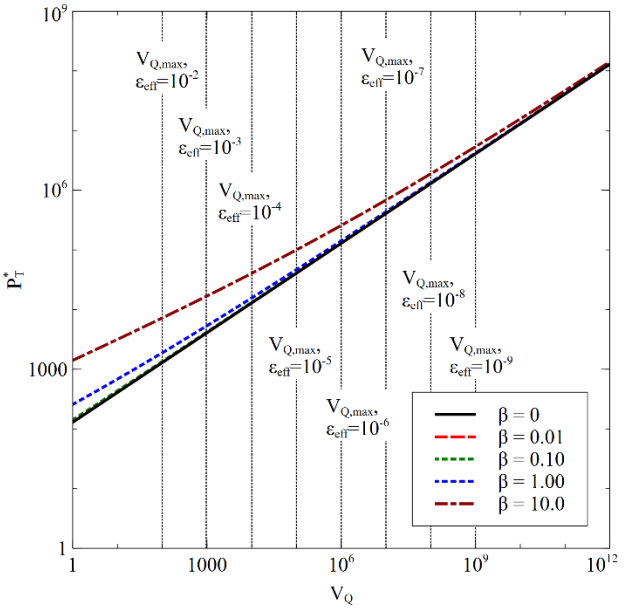


Fig. 10. Scaling of Power Use $P_T^*$ versus Quantum Volume $Q_V$ for a system withof 0.001 operating at 0.015 K. Vertical lines indicate the maximum achievable quantum volume for an effective error rate $\varepsilon_{eff}$

As shown in fig. 11, a similar pattern, with lower overall power usage, is obtained for a system operating at 4 K.

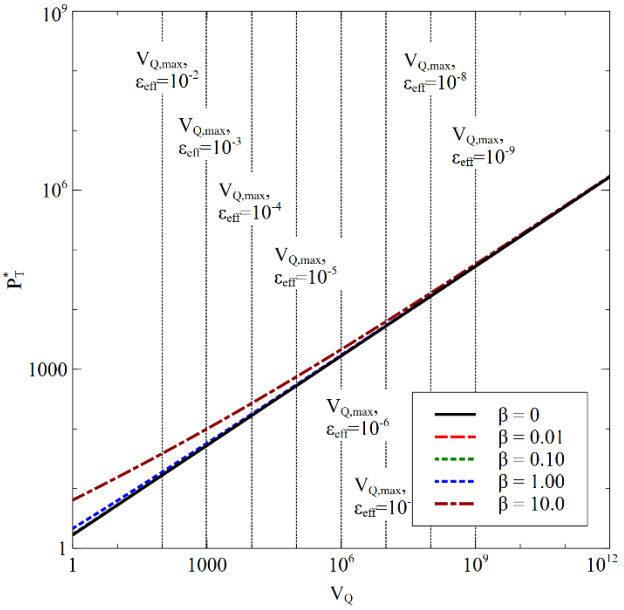


Fig. 11. Scaling of Power Use $P_T^*$ versus Quantum Volume $Q_V$. for a system withof 0.001 operating at 4.0 K. Vertical lines indicate the maximum achievable quantum volume for an effective error rate $\varepsilon_{eff}$

## 6 Conclusions

Quantum information systems use energy in very

different ways than conventional computers. Instead of the energy use being dominated by energy used within the electronic circuitry, energy use is dominated by cooling requirements. While this is well understood empirically, the current work quantifies the requirements across computational architectures, computer sizes, and thermal design parameters. A direct scale-up of power use in future quantum systems using this model based on published data is challenging, since key technologies, including both quantum electronics and conventional electronics that are integrated into quantum systems, are changing rapidly. Because the quantum computing has not yet considered sustainability, parameters relevant to energy use are generally not reported in descriptions of new technologies. However, even with this limitation, it is possible to use this framework to consider paths to energy optimization. This work points to two strategies for improving the energy efficiency of quantum systems, which will both lessen the energy impact of deploying quantum systems and reduce operational costs.

The first strategy is to design quantum systems in an energy-efficient manner. The qubit type will have a huge impact on overall energy use, since the Carnot efficiency of a system operating at the 4 K temperature of trapped ion qubits is more than 400 times the Carnot efficiency of a system operating at the 10 mK temperature required by some superconducting qubits. Even relatively small increases in operating temperatures for superconducting qubits are likely to significant increases in energy efficiency. While the effect of moving electronics out of the cryostat is somewhat harder to quantify due to potential changes in electronics power use, cryostat size, and heat lost through interconnects, the high cost of cooling at low temperatures means that this will be a critical decision in minimizing energy use. Finally, minimzing the physical size of electronics, even if the power use of the electronics does not change, will minimize the cooling power required.

The second strategy is to reduce the energy costs of cooling to quantum temperatures. Within conventional data center operation, experience shows that careful engineering design and integration to reduce non-computing data center energy costs can reduce PUE from 1.60 to 1.02. This reduction corresponds to a thirtyfold reduction in noncomputational energy costs. The current work shows that the largest opportunity for reducing energy use and obtaining similar improvements in quantum systems is through improving $\eta_c$. This approach will require both careful consideration of cycle efficiency and efficient detailed design of individual components.

The approach most likely to yield energy-efficient quantum systems is to combine these two strategies through *codesign*, where the impact of decisions made in deciding the computational architecture on the energy and cooling requirements of the system are considered during design and are matched with existing cooling capabilities. This may require designers to choose the optimal approaches to achieving a computational output within a set energy budget by selecting system architecture parameters.

## ACKNOWLEDGMENT

This work was authored by the National Renewable Energy Laboratory (NREL), operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. This work was supported by the Laboratory Directed Research and Development (LDRD) Program at NREL. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

## REFERENCES

[1] T. Häner, M. Roetteler, and K. M. Svore, "Factoring Using 2n+2 Qubits with Toffoli Based Modular Multiplication," *Quantum Information and Computation*, vol. 18, no. 7-8, pp. 673-684, June 2017.

[2] M. Reiher, *et al.*, "Elucidating Reaction Mechanisms on Quantum Computers," *Proceedings of the National Academy of Sciences of the U.S.A.* vol. 114, no. 29, pp. 7555-7560, July 2017.

[3] F. Arute, *et al*., "Quantum Supremacy Using a Programmable Superconducting Processor," *Nature*, vol. 574, pp. 505-510, October 2019.

[4] D-Wave Systems. *"The D-Wave 2000Q Quantum Computer,"* www.dwavesys.com/sites/default/files/D-Wave%202000Q%20Tech%20Collateral_1029F.pdf, 2018,

[5] E. Grumbling and M. Horowitz, *Quantum Computing: Progress and Prospects*, Washington, DC, National Academies Press, pp. 207-214, 2018.

[6] D. Gupta, *et al.,* "Digital Output Data Links from Superconductor Integrated Data Circuits," *IEEE Transactions on Applied Superconductivity,* vol 29, no. 5, Article 1303208, August 2019.

[7] D. S. Holmes, A. L. Ripple, and M. A. Manheimer, "Energy-Efficient Superconducting Computing- Power Budgets and Requirements," *IEEE Transactions on Applied Superconductivity,* vol. 23, no. 3, Article 1701610, June 2013.

[8] A. Shehabi, et al*, United States Data Center Energy Usage Report*, Berkeley, CA, Lawrence Berkeley National Laboratory, pp. 23, 2016.

[9] International Energy Agency, *Digitalization & Energy* IEA, Paris, France, pp. 116, 2017.

[10] E. Masanet, *et al.*, "Recalibrating Global Data Center Energy Usage Estimates," *Science,* vol. 367, no. 6481, pp.984-986, February 2020.

[11] D. J. Reilly, "Engineering the quantum-classical interface of solid-state qubits," *npj Quantum Information,* vol. 1, Article 15011, October 2015.

[12] H.-S. Zhong, *et al,* "Quantum computational advantage using photons," *Science,* vol 370, no. 6523, pp. 1460-1463, December 2020.

[13] J. M. Arrazola, *et al, "* Quantum circuits with many photons on a programmable nanophotonic chip," *Nature* vol. 591, no. 7848, pp. 54–60, March 2021.

[14] R. E. Sonntag and C. Borgnakke, *Fundamentals of Thermodynamics 10th Edition*, John Wiley & Sons, Hoboken, NJ, pp 162, 2019.

[15] V. Dinsesh Reddy, *et al.,* "Metrrics for Sustainable Data Centers," *IEEE Transactions on Sustainable Computing,* vol. 2, no. 3, pp. 290-303, July-September 2017.

[16] C. Belady, A. Rawson, J. Pfleuger, and T. Cader, "The green grid data center power efficiency metrics: PUE and DCiE," Green Grid, White Paper-06, 2007.

[17] Intel Delivers 17-Qubit Superconducting Chip with Advanced Packaging to QuTech. *Intel Newsroom* https://newsroom.intel.com/news/intel-delivers-17-qubit-superconducting-chip-advanced-packaging-qutech/, 2017.

[18] Baleegh Adbo *et al.* IBM Q 5 Tenerife V1.4.0 Version Log. *GitHub* https://github.com/Qiskit/ibmq-device-information/blob/master/backends/tenerife/V1/version_log.md, 2019.

[19] Rigetti. The Quantum Processing Unit (QPU). http://docs.rigetti.com/en/1.9/qpu.html, 2018.

[20] G. Pagano, *et al.,* "Cryogenic Trapped-ion System for Large Scale Quantum Simulation," *Quantum Science and Technogy,* vol. 4, no. 1, Article 014004, October 2018.

[21] K. R. Brown, J. Kim, and C. Monroe, "Co-Designing a Scalable Quantum Computer with Trapped Atomic Ions," *Npj Quantum Information* vol. 2, Article 16034, November 2016.

[22] M. F. Brandl, *et al.*, "Cryogenic Setup for Trapped Ion Quantum Computing," *Review of Scientific Instruments*, vol. 87, no. 11, Article 113103, November 2016.

[23] L. Hutin et al., "Si CMOS platform for quantum information processing," in *Proc. IEEE Symp. VLSI Technol.*, Honolulu, HI, USA, Jun. 2016, pp. 1–2, doi: 10.1109/VLSIT.2016.7573380.

[24] K. Ono, T. Mori, and S. Moriyama, "High-Temperature Operation of a Silicon Qubit," *Scientific Reports* vol. 9, Article 469, January 2019.

[25] L. Childress and R. Hanson, "Diamond NV Centers for Quantum Computing and Quantum Networks," *MRS Bulletin*, vol. 38, no. 2, pp. 134–138, February 2013.

[26] L. Robledo, *et al.*, "High-Fidelity Projective Read-Out of a Solid-State Spin Quantum Register," *Nature*, vol. 477, pp. 574–578, September 2011.

[27] S. Takahashi, *et al*, "Quenching Spin Decoherence in Diamond through Spin Bath Polarization," *Physical Review Letters*, vol. 101, no. 4, Article 047691, July 2008.

[28] X. L. Ruan, S. C. Rand, and M. Kaviany, "Entropy and Efficiency in Laser Cooling of Solids," *Physical Review B*, vol. 75, no. 21, Article 214304, June 2007.

[29] M. A. Green, "Helium Refrigeration During the 50 Years Since the 1968 Brookhaven Summer Study," *IEEE Transactions on Applied Superconductivity*, vol. 29, no. 5, pp. 1-55, August 2019.

[30] E. Jensen, "Energy Efficiency." CERN, https://doi.org/10.23730/CYRSP-2018-001.421, 2016.

[31] V. R. Kotsubo, *et al*, "Compact 2.2 K Cooling System for Superconducting Nanowire Single Photon Detectors." *IEEE Transactions on Applied Superconductivity*, vol. 27, no. 4, June 2017.

[32] M. Linder, *et al.* "Cryogenics in Space- A Review of the Missions and Technologies," *ESA Bulletin*, vol. 107, pp. 92-105, August 2001.

[33] V. Parma, "Cryostat Design," in *Proceedings, CAS-CERN Accelerator School: Superconductivity for Accelerators*, pp 353-399, 2014.

[34] A. Youssefi, *et al.* "A cryogenic electro-optic interconnect for superconducting devices," *Nature Electronics vol.* 4, no. 5, pp. 326–332, 2021.

[35] R. Landauer, "Irreversibility and Heat Generation in the Computing Process," *IBM Journal of Research and Development* vol. 5, no. 3, pp.183–191, July 1961.

[36] R. Gaudenzi, *et. al*, "Quantum Landauer Erasure with a Molecular Nanomagnet," *Nature* Physics vol. 14, pp. 565-568, March 2018.

[37] J. Gea-Banacloche, "Minimum Energy Requirements for Quantum Computation," *Physical Review Letters*, vol. 89 no. 21, Article 217901, November 2002.

[38] J. Ikonen, J. Salmilehto, and M. Möttönen, "Energy-efficient quantum computing," *Npj Quantum Information*, vol. 3, Article 17, April 2017.

[39] J. M. Hornobrook, *et al.*, "Cryogenic Control Architecture for Large-Scale Quantum Computing," *Physical Review Applied*, vol. 3, no. 2, Article 0240101, February 2015.

[40] F. Jazaeri, A. Beckers, A. Tajalli, and J.-M. Sallese, "A review on quantum computing: From qubits to front-end electronics and cryogenic MOSFET physics," in Proc. 26th Int. Conf. Mixed Des. Integr. Circuits Syst. (MIXDES), Jun. 2019, pp. 15–25.

[41] T. Lehmann, "Cryogenic Support Circuits and Systems for Silicon Quantum Computers," in 2019 IEEE International Symposium on Circuits and Systems (ISCAS), May 2019, pp. 1–5.

[42] J. C. Bardin, *et al.* "Design and Characterization of a 28-nm Bulk-CMOS Cryogenic Quantum Controller Dissipating Less than 2 mW at 3 K," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 11, pp. 3043-3060, November 2019.

[43] Z. Xue, *et al.* "CMOS-based cryogenic control of silicon quantum circuits," *Nature*, vol. 593, no. 7858, pp. 205–210, May 2021.

[44] S. J. Pauka, *et al.* "A cryogenic CMOS chip for generating control signals for multiple qubits," *Nature Electronics* vol. 4, no. 1, pp. 64–70, February 2021.

[45] M. Mehrpoo, F. Sebastiano, E. Charbon and M. Babaie, "A Cryogenic CMOS Parametric Amplifier," *IEEE Solid-State Circuits Letters*, vol. 3, pp. 5-8, 2020.

[46] A. Esmailiyan *et al.*, "A Fully Integrated DAC for CMOS Position-Based Charge Qubits with Single-Electron Detector Loopback Testing," *IEEE Solid-State Circuits Letters*, vol. 3, pp. 354-357, 2020.

[47] B. C. Paz *et al.*, "Integrated Variability Measurements of 28 nm FDSOI MOSFETs down to 4.2 K for Cryogenic CMOS Applications," *2020 IEEE 33rd International Conference on Microelectronic Test Structures (ICMTS)*, 2020, pp. 1-5.

[48] E. Cha, N. Wadefalk, G. Moschetti, A. Pourkabirian, J. Stenarson and J. Grahn, "A 300-$\mu$W Cryogenic HEMT LNA for Quantum Computing," *2020 IEEE/MTT-S International Microwave Symposium (IMS)*, 2020, pp. 1299-1302.

[49] J. Dongarra, P. Luszczek, and A. Petitet, "The LINPACK Benchmark: Past, Present and Future," *Concurrency and Computation Practice and Experience*, vol. 15, no. 9, pp. 803-820, August 2003.

[50] N. Moll, *et al.*, "Quantum optimization using variational algorithms on near-term quantum devices," *Quantum Science and Technology*, vol. 3, Article 030503, 2018.

[51] A. W. Cross, *et al.*, "Validating Quantum Computers Using Randomized Model Circuits," *Physical Review A*, vol. 100, no. 3, Article 032328, September 2019.

[52] J. M. Pino, *et al.*, "Demonstration of the QCCD trapped-ion quantum computer architecture," *arXiv preprint arXiv:2003.01293*, 2020.

[53] K. Wright, *et al.*, "Benchmarking an 11-qubit quantum computer," *Nature Communications*, vol. 10, Article 5464, November 2019.

**Michael James Martin** received his PhD in Aerospace Engineering from the University of Michigan in 2007. He received his M.S. in Science and Technology Studies from Virginia Tech in 2018. He is currently a staff scientist at the National Renewable Energy Laboratory in the Computational Sciences Center. Prior to joining NREL he held research appointments at NASA-JPL, Louisiana State University, and the Naval Research Laboratory, as well as policy appointments in the Department of Energy and the United States Senate. His professional honors include the Department of Energy Secretary's Appreciation Award, the American Society of Mechanical Engineers (ASME) Congressional Fellowship, Associate Fellow of the American Institute of Aeronautics and Astronautics (AIAA) as well as multiple teaching awards. Dr. Martin is currently a member of the ASME's Energy and Public Policy committee. His current research interests include the physics and sustainability of new energy technologies.

**Caroline L. Hughes** earned a B.S. in Engineering Physics from the University of Colorado Boulder in 2015 and an M.S. in Nuclear Engineering from the University of California, Berkeley, in 2018, with emphases in applied mathematics and numerical analysis. She is a researcher in the Strategic Energy Analysis Center at the National Renewable Energy Laboratory in Golden, CO. Her research ranges from software development to strategic planning in energy research and quantum computing.

**Gilbert Moreno** received his B.S. in Mechanical Engineering from the University of Texas at El Paso (2001) and his M.S. (2005) and Ph.D. (2009) in Mechanical Engineering from the University of Texas at Arlington. He is a senior research engineer at the National Renewable Energy Laboratory (NREL) where he conducts research into the thermal management of automotive power electronics. He has worked at NREL for over 10 years developing advanced cooling technologies for various DOE and industry funded projects.

**Eric B. Jones** received a dual B.A. in mathematics and physics from Wesleyan University in 2013, followed by an M.Sc. in quantum fields and fundamental forces from Imperial College London in 2015, and a Ph.D. in applied physics from the Colorado School of Mines in 2020. He is currently a postdoctoral researcher in the Computational Sciences Center at the National Renewable Energy Laboratory in Golden, CO. His research revolves around using noisy intermediate-scale quantum computers to solve problems in condensed matter physics, optimization, chemistry, materials science, and artificial intelligence relevant to the development and deployment of renewable energy and sustainability technologies.

**David Sickinger** received his M.S. in Mechanical Engineering from the University of Arizona in 2001 and a PostGradDip in Computer Science from the University of Auckland in 2004. He is a researcher in the Computational Science Center at the National Renewable Energy Laboratory in Golden, CO. His research revolves around facility

cooling and power infrastructure, with a focus on energy and water efficient data center facilities (design, construction, operation).

**Sreekant Narumanchi** received his PhD in Mechanical Engineering from Carnegie Mellon University. He is currently a group manager for the Advanced Power Electronics and Electric Machines (APEEM) Group at NREL. Dr. Naramuchani's professional recognition includes Fellow of the American Society of Mechanical Engineers (ASME), the 2022 THERMI Award, the 2018 ASME EPPD-K16 Clock Award, a 2016 R&D 100 Award, the 2003 Best Paper Award from the ASME Journal of Electronic Packaging, and the ASME 2013 InterPACK Conference Outstanding Paper Award. He serves as Associate Editor for the ASME Journal of Electronic Packaging, as Guest Editor of the IEEE Components, Packaging, and Manufacturing Technologies Journal, as a member of the Executive Committee of the ASME Electronic and Photonic Packaging Division, and as a member of Thermal Working Group of the IEEE Heterogeneous Integration Roadmap Committee. His current research interests include thermal-fluids, electro-thermal, mechanical and reliability aspects of power electronics and electric machines.

**Ray Grout** received his PhD in engineering from the University of Cambridge. He is currently the director of NREL's Computational Science Center. Prior to joining NREL he was a postdoctoral researcher at Sandia National Laboratory. Dr. Grout's research interests include reacting flows, simulation using exa-scale computing and other new architectures, turbulence, and machine learning for physical systems.