13

Quantum phase estimation

The authors are grateful to Patrick Rall and Ronald de Wolf for reviewing this chapter.

Rough overview (in words)

The quantum phase estimation (QPE) subroutine produces an estimate of an eigenvalue of a unitary operator. It is a cornerstone of quantum algorithms primitives and has numerous applications. For example, Shor's algorithm for factoring can be viewed as an application of QPE together with modular exponentiation. Similarly, when combined with Hamiltonian simulation, QPE can produce an estimate for an eigenvalue of a Hamiltonian (given an appropriate initial state), an important problem in areas such as quantum chemistry. Generally, since quantum computations enact unitary operators, quantum phase estimation is an essential algorithmic tool for accessing information about these operators, specifically, information about their periodicities, and the properties of their eigenstates.

As one of the oldest quantum primitives discovered [624, 299], QPE has played a significant historical role in the development of quantum algorithms. In a typical use case, QPE is used as a first step to compute an estimate of the eigenvalue of the unitary into an ancilla register. Then, the ancilla register is used as a control for subsequent operations. However, in some applications, such as Gibbs sampling and solving the quantum linear system problem, this procedure must be applied *coherently* to a superposition of eigenstates with different eigenvalues, and the estimate of the eigenvalue must be uncomputed at the end. As discussed below, coherent usage of the QPE primitive in this manner must be handled with care, due to several identified caveats. While QPE still provides essential intuition for how these applications work, in some

cases, modern techniques leveraging quantum signal processing and the quantum singular value transformation [431] lead to a cleaner and more direct analysis than QPE.

In the discussion below, we begin with the textbook presentation of QPE [299, 801], and expound on the aforementioned caveats. We also present example use cases, noting the instances where QPE was originally a key ingredient but no longer features directly in state-of-the-art solutions.

Rough overview (in math)

Let U be a unitary with eigendecomposition $U = \sum_j \mathrm{e}^{\mathrm{i}2\pi\phi_j}|\psi_j\rangle\langle\psi_j|$. Given as input the state $|\psi_j\rangle$, the QPE subroutine produces an estimate $\hat{\phi}_j$ for ϕ_j. The algorithm requires the ability to apply controlled U^{2^p} for non-negative integers p. If ϕ_j is an exact multiple of 2^{-P}, then an exact estimate of ϕ_j can be learned with certainty using only $p \in \{0, 1, \ldots, P-1\}$. In general, an estimate $\hat{\phi}_j$ of ϕ_j satisfying $|\phi_j - \hat{\phi}_j| \leq \epsilon$ can be learned with high probability by taking the maximum value of 2^p on the order of $1/\epsilon$. The algorithm also requires application of an inverse quantum Fourier transform to orchestrate the constructive interference near the estimate for ϕ_j. The quantum circuit for the standard approach to QPE is shown in Fig. 13.1.

Phase estimation can also be applied coherently onto a superposition of eigenstates. Suppose that the input state is $|\psi\rangle = \sum_j \alpha_j|\psi_j\rangle$. By linearity, if each phase ϕ_j is a multiple of 2^{-P} and phase estimation is run with sufficient resolution, then QPE enacts the following unitary

$$|\psi\rangle|0\rangle \mapsto \sum_j \alpha_j|\psi_j\rangle|\phi_j\rangle, \tag{13.1}$$

where $|\phi_j\rangle$ holds a P-bit binary representation of ϕ_j. If the auxiliary register is measured—here assuming for simplicity that the eigenvalues ϕ_j are nondegenerate—then with probability $|\alpha_j|^2$ (consistent with the Born rule) the estimate ϕ_j is obtained and the state collapses to the corresponding eigenstate $|\psi_j\rangle$.[1] If the phases ϕ_j are not multiples of 2^{-P}, an approximate version of this operation can still be accomplished as long as the precision is sufficiently small to resolve the eigenvalues, subject to some caveats (discussed below).

[1] Alternatively, if ϕ_j is known ahead of time (to sufficient precision), QPE can be wrapped inside of amplitude amplification and the state $|\psi_j\rangle$ can be prepared using $O(|\alpha_j|^{-1})$ applications of the QPE circuit, rather than $O(|\alpha_j|^{-2})$. Note that amplitude amplification can be understood through the QSVT [431] formalism, and in many applications, such as projecting onto the ground state of a Hamiltonian [688], one can achieve this sort of scaling directly without explicitly relying on QPE.

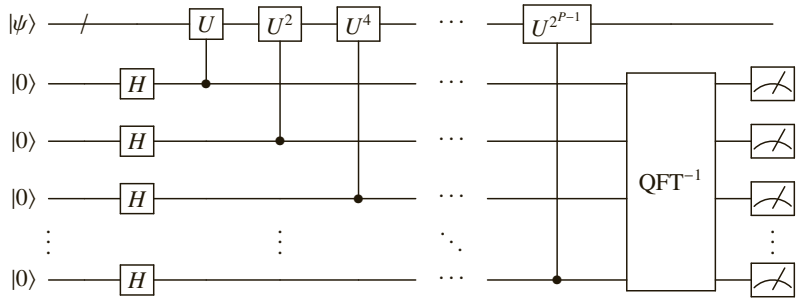

Figure 13.1 Quantum circuit implementation of QPE. The measurement outcomes on the P ancilla qubits give a P-bit estimate of the phase ϕ_j (correct up to error $O(2^{-P})$) with high probability.

Dominant resource cost (gates/qubits)

The QPE subroutine is typically dominated by calls to the controlled unitary U. If resolution ϵ is desired, one must perform controlled U^{2^p} operations for $p \in \{0, 1, \ldots, \lceil \log_2(1/\epsilon) \rceil + O(1)\}$; thus, the number of calls to a controlled U oracle will be $O(1/\epsilon)$. This dependence on ϵ is optimal; the $O(1/\epsilon)$ scaling is known as the *Heisenberg limit*.

In the context of estimating the eigenenergy of a Hamiltonian H, one can choose $U = e^{iH}$, and then implement controlled U^t, that is, controlled e^{iHt}, with Hamiltonian simulation. In this case, given the ability to prepare an eigenstate of H, an ϵ-approximation of the eigenvalue requires values of t up to $O(1/\epsilon)$.[2] However, one must also factor in the error in the Hamiltonian simulation. In a typical setting, access to the n-qubit Hamiltonian is given through a linear combination of L unitaries, for instance, Pauli matrices. Let $\|H\|_1$ denote the sum of the coefficients in the combination. Then, methods for Hamiltonian simulation based on quantum signal processing can approximate e^{iHt} to error $O(\epsilon)$ with $O(nL(\|H\|_1 t + \log(1/\epsilon)))$ gate complexity, whereas methods based on product formulas incur cost $O(nL(\|H\|_1 t)^{1+1/2k} \epsilon^{-1/2k})$ for $(2k)$th-order product formulas, although the actual cost can be lower after accounting for structure in the Hamiltonian terms. Balancing the error from phase estimation against the error from Hamiltonian simulation can cause sub-Heisenberg-limited performance, such as in the case of the product formula approach. The overhead associated with imperfect Hamiltonian simulation can be avoided by applying QPE to different functions of H; for example, a promising choice is the qubitization

[2] The fact that learning energies to greater precision requires a proportionally greater amount of time t is a manifestation of the energy-time Heisenberg uncertainty principle, and forms the origin of the term "Heisenberg limit."

operator, which acts in a similar way to $U = e^{i \arccos(H/\alpha)}$, where α is the normalization factor of the qubitization operator. The reason this is advantageous is that the qubitization operator can be implemented *exactly* given access to a block-encoding of H [717, 841, 139]. In general, we require the unitary U on which phase estimation is performed to be a known, classically invertible function of the Hamiltonian $U = f(H)$. The complexity of QPE depends on the desired uncertainty in the energy eigenvalue, which can be related to the uncertainty in the measured eigenphase via the magnitude of the derivative of the function, $\|f'(\cdot)\|$.

The number of qubits for QPE is simply the size of the register needed to hold the input state $|\psi_j\rangle$ plus the size of the register needed to hold the estimate $\hat{\phi}_j$ (i.e., roughly $\lceil \log_2(1/\epsilon) \rceil$ bits). Additionally, QPE requires an inverse quantum Fourier transform (QFT), which (using the textbook QFT implementation) adds only $O(\log^2(1/\epsilon))$ additional gates to the protocol.

Another version of QPE [627] achieves the same task with only a single ancilla qubit, but, as a result, learns only one bit of the output at a time. Additionally, it requires an exact eigenstate as input. The latter problem can be avoided using a statistical approach [690, 1013].

Caveats

The main caveats of QPE are related to the fact that eigenphases are not always exact integer multiples of 2^{-P}, resulting in noncertain outcomes of QPE, which can lead to complications in certain applications.

- Fat tails and boosting of success probability: Whenever the phases ϕ_j are not exact integer multiples of 2^{-P} for some integer P, phase estimation will not return the answer ϕ_j with certainty. Rather, there will be a distribution of possible estimates $\hat{\phi}_j$ that is peaked near ϕ_j. If one chooses $P = \lceil \log_2(1/\epsilon) \rceil + O(1)$, then most of the probability mass of this distribution lies within ϵ of ϕ_j. As P is increased further, the distribution becomes more sharply peaked near ϕ_j, and if an ϵ-accurate estimate with $1 - \delta$ probability is desired, one must take $P = \lceil \log_2(1/\epsilon) \rceil + O(\log(1/\delta))$, corresponding to a multiplicative $O(1/\delta)$ overhead in the query complexity to U and $O(\log(1/\delta))$ additional ancilla qubits. This poor δ dependence is due to "fat tails" on the distribution of estimates of $\hat{\phi}_j$. One way to avoid this overhead is to take the median of estimates obtained from $O(\log(1/\delta))$ repetitions of QPE [792, Lemma 1]. A downside of this approach is that it may be difficult to implement coherently on a superposition of eigenstates, in the sense of Eq. (13.1), since computing the median would require a coherent quantum sorting network. An alternative way to circumvent the fat tails problem

is to modify the QPE protocol to have a nonuniform superposition in the register that controls applications of U; a judicious choice of superposition leads the distribution over estimates $\hat{\phi}_j$ to be a Kaiser window (see [143, Appendix D] and [144]) or discrete prolate spheroidal sequence (DPSS) function [829], which minimizes the probability of deviating from ϕ_j by more than ϵ. See also [260], where a Gaussian profile is used to suppress the tails. Boosting the success probability to $1 - \delta$ in this fashion incurs multiplicative $O(\log(1/\delta))$ cost, rather than $O(1/\delta)$. The overall cost in queries to U by these methods matches a lower bound of $\Omega(\epsilon^{-1} \log(1/\delta))$ shown in [736].

- Performing coherent QPE: When ϕ_j are noninteger multiples of 2^{-P}, the coherent operation in Eq. (13.1) cannot be straightforwardly performed with exact fidelity. This is because for each value of j, the second register will be in a superposition of many values of $\hat{\phi}_j$ (most but not all of the amplitude will lie on estimates close to ϕ_j). To restore coherence, one might try coherently rounding the estimate $\hat{\phi}_j$ onto a coarser net of grid points (and then uncomputing the original estimate $\hat{\phi}_j$); however, there will always be edge cases where ϕ_j falls very near the midpoint between two grid points and rounding destroys some of the coherence in the input. This is true even as the precision of QPE is taken to zero ($\epsilon \to 0$). See [854] for a discussion. One possible way to mitigate this issue is presented in the "consistent phase estimation" protocol of [975, Section 5.2], where a random shift is applied to the grid points to avoid this situation for any particular eigenphase with high probability. However, this does not generically work simultaneously for all eigenphases. In [854], it is shown that performing the map of Eq. (13.1) is impossible without a "rounding promise" on the set of eigenphases $\{\phi_j\}$.
- Biased estimator: A further consequence of the noncertainty of the QPE output is that the estimate $\hat{\phi}_j$ is *biased*; that is, its expectation value is not exactly equal to ϕ_j. This issue can also be fixed with a random shift idea, yielding an unbiased (and symmetrically distributed) version of QPE [691, 49].

Example use cases

- In quantum chemistry and condensed matter physics, QPE can be used to measure the eigenvalues (and especially the ground state energy) of the Hamiltonian H, which gives knowledge about reaction mechanisms, stable configurations, and other equilibrium properties. For QPE to succeed, a trial state $|\psi\rangle$ with substantial overlap with the eigenstate of interest must be input to QPE, which is challenging in the general case. The problem of ground state preparation has garnered intense study, and state-of-the-art techniques do not always follow the textbook method that relies on the QFT, presented

above. For example, quantum signal processing can be leveraged directly to filter out unwanted eigenstates [689, 688], effecting a similar outcome as QPE.

- In Shor's algorithm, given a composite integer N and a (randomly chosen) base $x < N$, QPE is used to determine the order of x, that is, the minimum integer r for which $x^r \equiv 1 \mod N$, which is in turn used to infer the prime factors of N. Here, the unitary U is the modular multiplication unitary that sends $|y\rangle \mapsto |xy \mod N\rangle$.
- In amplitude estimation [186], given a unitary U that prepares a state $U|\psi_0\rangle = a|\psi_g\rangle + b|\psi_b\rangle$, QPE is used to estimate $|a|$ or $|a|^2$. More advanced approaches to amplitude estimation not relying on QPE have since been developed. These leverage Grover's algorithm, or more generally quantum signal processing, without using the QFT. While these do not surpass the QPE-based method in asymptotic complexity, they potentially offer other benefits, such as improved practical performance and versatility. See [855] and references therein.
- In the Monte Carlo–style quantum algorithms for Gibbs sampling of quantum (i.e., nondiagonal in the computational basis) Hamiltonians, roughly speaking, the quantum state undergoes a random walk on the eigenbasis of the Hamiltonian. Steps of this random walk are accepted or rejected according to how much the energy changes at each step. The QPE subroutine is used to simultaneously (approximately) project onto the eigenbasis of the Hamiltonian and to produce an estimate of the energy, used to determine whether the step should be accepted or rejected. Early studies [984, 1076, 1048] of this approach were hampered by the caveats related to rejecting quantum states and imperfect energy estimates, but recent works [856, 260, 259] circumvent these problems (by randomizing the grid points or completely abandoning phase estimation).
- To follow the ground state $|\psi_0(s)\rangle$ of a Hamiltonian $H(s)$ as some parameter s is varied from 0 to 1, one can run the adiabatic algorithm. Alternatively, one can consider a discretization of steps $s_t \in \{s_0, \ldots, s_T\}$, where $0 = s_0 < s_1 < s_2 < \cdots < s_{T-1} < s_T = 1$, and run QPE on $H(s_t)$ in succession, each time causing a measurement in the eigenbasis of $H(s_t)$. Due to the quantum Zeno effect, as long as sufficiently small steps are taken, each projection will be onto the ground space with high probability (see, e.g., [944]). Larger steps can be tolerated if one boosts the probability that each step succeeds with amplitude amplification [162]. This approach is similar to the idea in Hastings' short-path algorithm [505, 329], which solves combinatorial optimization problems. However, note that modern implementations along these lines would likely elect to perform the ground state projection

via eigenstate filtering [689] or related QSVT-based methods, rather than QPE.

- While state-of-the-art quantum linear system solvers (QLSSs) do not explicitly use QPE, the original QLSS by Harrow, Hassidim, and Lloyd [500] uses QPE to coherently measure the eigenvalues of a matrix A into an auxiliary register. These eigenvalue estimates are subsequently inverted with coherent arithmetic in order to produce the state $A^{-1}|b\rangle$ corresponding to the solution to the system $Ax = b$. Achieving optimal asymptotic performance requires additional ingredients beyond QPE, and is best understood through the language of block-encodings and quantum linear algebra. This framework allows for manipulation of eigenvalues without explicitly reading them into an ancilla register with QPE.
- In certain machine learning tasks related to linear algebra, such as principal component analysis [708] and recommendation systems [608], quantum algorithms have been proposed that leverage QPE to access the information about the eigenvectors and eigenvalues. As explained in [854], these have not always fully accounted for the caveat related to coherent QPE, although typically these caveats can be circumvented using the framework of quantum linear algebra [248, 431].

Further reading

- The standard circuit and analysis of QPE appears in Nielsen and Chuang [801]. See also [299].
- Many variants of the QPE algorithm have been explored, which can be superior to the standard version in certain settings. See, for example, [854, 690] for additional references and informative overviews of various methods, along with their advantages and drawbacks.
- Reference [687] contains a pedagogical overview of QPE including some of its variants and applications.