The quantum cocktail party problem

Xiao Liang\(^1\), YaDong Wu\(^1\), and Hui Zhai\(^1\)*

\(^1\)Institute for Advanced Study, Tsinghua University, Beijing 100084, China;
\(^2\)Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, China

Received November 6, 2019; accepted December 10, 2019; published online February 21, 2020

Citation: X. Liang, Y. D. Wu, and H. Zhai, The quantum cocktail party problem, Sci. China-Phys. Mech. Astron. 63, 250362 (2020), https://doi.org/10.1007/s11433-019-1493-7

The cocktail party problem refers to the phenomenon that the brain of a listener can focus on a single voice while filtering out a range of other voices in a multi-talker situation, say, in a cocktail party [1]. This selective attention problem was firstly defined as the “cocktail party problem” (CPP) by Cherry [2] in 1953. In recent years, machine-learning based approaches to solve the CPP were essential for many industrial applications such as automated speech recognition. The independent component analysis (ICA) is such an algorithm particularly suitable for the CPP. The CPP also has applications in physical science such as astrophysics data analysis [3, 4]. Recently, it has also been proposed to use the spirit of CPP and ICA method to extract the eigen frequency of a quantum system from a dynamical probe [5].

Let us briefly review the classical CPP (c-CPP). Considering \(N\)-independent speakers in a room, they speak simultaneously and the voice of each speaker is a source denoted by a sequence \(s_i(t)\) \((i = 1, \ldots, N)\). There are also \(M\) detectors in the room. Each detector detects a signal \(x_j(t)\) \((j = 1, \ldots, M)\) that is considered to be a linear combination of all sources \(s_i(t)\), as schematically shown in Figure 1(a). That is to say, we have a \(M \times N\)-dimensional matrix \(A\) and

\[
x_j(t) = \sum_i A_{ji} s_i(t).
\]  

(1)

To concentrate on one of the speakers, it means that we should find out \(A^{-1}\) such that we can determine \(s_i(t)\) as

\[
s_i(t) = \sum_j (A^{-1})_{ij} x_j(t)
\]

from the signals of all detectors. To make the situation simpler, we usually consider that there are more detectors than sources, that is, \(M > N\). Unlike other blind signal separation problems [6], here both \(s_i\) and \(A\) are unknown, therefore, this is an ill-defined problem if no information of the source is known. In real cases, we utilize the information that being voice of an individual speaker, each source \(s_i(t)\) displays a certain feature and is more regular than a mix of several voices. By performing statistics over \(t\) for each sequence \(s_i(t)\), we can determine the entropy of the sequence and we use the criterion that the entropy of each sequence should be minimized to determine each \(s_i(t)\). This is how we solve the CPP with the ICA method [7].

In this letter we will propose a quantum analogy of the CPP, termed as the quantum CPP (q-CPP). We will discuss how to solve the q-CPP with an analogy of the ICA method. We should also present a mathematical statement that can help us to map the loss function to a Hamiltonian of Ising spins, therefore finding out each source can be mapped to

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure1.png}
\caption{(Color online) Schematic of the cocktail party problem. (a) Classical case. The voices emitted from different sources are mixed and detected by detectors. (b) Quantum case. The wave functions from different sources mix spatially and the density matrix at different places are detected by detectors.}
\end{figure}

\*Corresponding author (email: hzhai@tsinghua.edu.cn)
find the ground state of this spin Hamiltonian. Although classical Monte Carlo can solve the ground state’s spin configuration, we point out that this spin Hamiltonian can be solved more efficiently by quantum methods, for example, by quantum simulation and quantum annealing.

Quantum CPP  Here we propose the q-CPP. We consider \( N \) different sources, and each source \( s_i \) is a density matrix in a Hilbert space with dimension \( d \). There are \( M \) detectors, and the signal \( x_j \) detected by each detector is a density matrix of a mixed state denoted by \( \rho_j \):

\[
\rho_j = \sum_i A_{ji} |\phi_i\rangle \langle \phi_i|.
\]

We also normalize \( \rho_j \) to be trace unity, which requires \( \sum_i A_{ji} = 1 \) for all \( j \). The q-CPP is defined as follows: Suppose that we know a sufficient number of \( \rho_j \), can we find out \( A_{ji} \) to recover each \( |\phi_i\rangle \)?

Here we will briefly discuss the uniqueness of the solution. Firstly, we should emphasize that in order to ensure the solution is unique, it is important to require that different \( |\phi_i\rangle \) are not orthogonal to each other. Secondly, when the number of detectors increases by one, the constraints increase by \( d^2 \) and the free parameters increase by \( N - 1 \), so we will consider the situation that \( d^2 > N \). Lastly, it is always good to have a sufficient number of detectors, normally we consider \( M > N \). We do not rigorously prove the uniqueness of the solution, but we find that in practice, usually we can find unique solutions when these conditions are satisfied.

A physical realization of the q-CPP can be proposed as follows. Let us consider a particle whose internal Hilbert space is a product of two degrees of freedom as \( \mathcal{H} = \mathcal{H}_A \times \mathcal{H}_B \). The dimensionality of \( \mathcal{H}_A \) is \( d \) and the dimensionality of \( \mathcal{H}_B \) is \( N \), and \( \{|x_i\rangle\} \) \( (i = 1, \ldots, M) \) forms a complete set of orthogonal bases in \( \mathcal{H}_B \). A wave function \( |\Phi\rangle \) can be generally expanded in these bases as:

\[
|\Phi\rangle = \sum_{i=1}^N \varphi_i(r) |\phi_i\rangle |x_i\rangle.
\]

For a more physical picture, one can consider eq. (3) as a particle that can be emitted from \( N \) different sources, and the wave function in \( \mathcal{H}_B \) is \( |x_i\rangle \) for the particle emitted from the source-\( i \), as schematically shown in Figure 1(b). If we place \( M \) detectors in different places \( r \) and the measurement does not act on Hilbert space \( B \), the quantum measurement traces out the Hilbert space \( \mathcal{H}_B \) and results in \( M \) different density matrices as eq. (2) with \( A_{ji} = |\varphi_i(r_j)|^2 \). Thus, it is also natural to require \( A_{ji} \) to be positive numbers. In practice, these density matrices can be constructed by the quantum state tomography. We consider the situation that both \( \varphi(r) \) and \( |\phi_i\rangle \) are unknown. The q-CPP is to determine them from \( \rho_j \) \( (j = 1, \ldots, M) \).

It is worth making a comparison with the quantum interference problem. Quantum interference takes place between identical particles and one can read out information of each source from the interference pattern. It is also well known that interference gets lost when particles are not identical. In this case, because the wave functions in the Hilbert space \( \mathcal{H}_B \) are orthogonal, it plays the role of labelling particles from different sources. Therefore, our q-CPP problem can be regarded as how to recover information of each source from the mixing of a number of distinguishable particles. Note that because the coherence is essentially lost, one can at most recover the amplitude (here \( A_{ji} = |\varphi_i(r_j)|^2 \)) but not the phase.

To find out the pure state, the most important information we use here is that the density matrix of a pure state has the property that \( \rho^2 = \rho \). Thus, the scheme is to find out a proper combination of \( \rho_j \) as \( \rho = \sum_{j=1}^M w_j |\phi_j\rangle \langle \phi_j| \) with the normalization condition \( \sum_j w_j = 1 \) that can minimize \( |\rho^2 - \rho| \). This is equivalent to say, we define the loss function as:

\[
\mathcal{F} = \sum_{mn} (|\rho_j^2 - \rho|_{mn})^2 = \sum_{m} \left| \sum_{j=1}^M (\rho_j^2 w_j^2 - \rho_j w_j)_{mn} + \sum_{ij} (\rho_i \rho_j)_{mn} w_i w_j \right|^2.
\]

A comparison between c-CPP and q-CPP is summarized in Table 1.

| Source | c-CPP | q-CPP |
|---|---|---|
| Detector | voice of each speaker | each pure state |
| Loss function | minimizing entropy | minimizing \( |\rho^2 - \rho| \) |
normalize it under the constraint \( \sum_j A_{ij} = 1 \). For the example shown in Figure 2, we choose three pure states \( \rho_i = \rho(\phi_i) \) and the fidelities between the three pure states are \( F(\rho_1, \rho_2) \approx 0.56 \), \( F(\rho_1, \rho_3) \approx 0.54 \) and \( F(\rho_2, \rho_3) \approx 0.88 \), where the fidelity is defined as \( F(\rho, \rho_0) = \text{Tr} \sqrt{\sqrt{\rho_0} \rho \sqrt{\rho_0}} \).

We then use Newton’s method to solve the q-CPF. \( w \) is initialized in such a way that \( w_j (j = 1, \ldots, M - 1) \) are uniformly sampled in the range of \([-2, 2]\) and \( w_M \) is determined by the constraint \( \sum_j w_j = 1 \). Then, we can reach a convergent solution following eq. (5). In the example of Figure 2, three different \( \rho_f \) can be found by Newton’s method depending on different initializations, and their fidelities with \( \rho_i (i = 1, 2, 3) \) are shown in Figure 2(a)-(c). One can see that there is always one fidelity equalling unity. For instance, for the case \( \{ \rho_f, \rho_2, \rho_3 \} \) are consistent with \( F(\rho_f, \rho_2) \) and \( F(\rho_f, \rho_3) \), respectively. This means that the resulting \( \rho_f \) recovers \( \rho_1 \). Similarly, in the cases of Figure 2(b) and (c), the resulting \( \rho_f \) recovers \( \rho_2 \) and \( \rho_3 \), respectively. We have also tried different numbers of detectors. For the case with three sources, we find that the performance is good as long as the number of detectors is equal to or greater than three.

Mapping to a Hamiltonian problem: Since eq. (4) is a function of \( \{ w_j \} \), and if we restrict the value of all \( w_j \) to be \( \pm 1 \), minimizing eq. (4) can be regarded as finding the ground state of a Hamiltonian of the Ising spins. If we replace \( w_j \) with \( \sigma^z_j \), eq. (4) can be written into a Hamiltonian form as:

\[
\hat{H} = \sum_{m} \sum_{j=1}^{M} \left( (\rho^z_j)^{mm} - (\rho_j)^{mm} \sigma^z_j \right) + \sum_{ij} (\rho(i)\rho_j) \sigma^z_i \sigma^z_j \]

where

\[
A_{ij} = \text{Tr} (\rho(\rho(i)\rho(j)), (8)
B_{ij} = -\text{Tr} (\rho(i)\rho(j)\rho(k) - \text{Tr}(\rho(i)\rho(k)), (9)
C_{ij} = \text{Tr}(\rho(j)), (10)
\]

Here we set the energy unit of the Hamiltonian as unity. In order to satisfy the constraint \( \sum_j w_j = 1 \), we require the number of spins to be odd and the total magnetization to be unity. Note that this Hamiltonian contains four, three and two-body interactions. In this Hamiltonian, the number of sites is equal to the number of detectors. Here it is worth emphasizing that only computing the coefficients listed in eqs. (8)-(10) depends on the pure state Hilbert dimension \( d \) of the original quantum problem, and the complexity of the Hamiltonian eq. (7) itself will not increase as \( d \) increases. Given that in the previous discussion of uniqueness of the solution, we prefer to have a large \( d \), this is a great advantage of this approach.

Now the question is whether we can restrict all \( w_j \) to be \( \pm 1 \). Here we make the following statement:

Statement: We consider each source \( s_i \) as a vector, and \( M \)-number of signal \( x_j (j = 1, \ldots, M) \) as a mixing of \( N \)-number of sources \( s_i \), written as \( x_j = \sum_{j} A_{ij} s_i \), where all \( A_{ij} \) are positive numbers ranging between zero and unity without any other restrictions. We construct \( y = \sum_{j=1}^{M} w_j s_j \) where \( w_j \) can only take \( \pm 1 \). For each given \( M \) and for a specified target \( s_k \), we optimize \( w_j (j = 1, \ldots, M) \) to minimize \( |y - s_k| \) and the minimized value is denoted by \( |y - s_k|_{\text{min}} \). We state that

Figure 2: (Color online) (a)-(c) The fidelities versus number of detectors for three different solutions found by Newton’s method. (d)-(f) The fidelities versus number of spins found by looking for the ground state of the Ising Hamiltonian. Here the number of spins means the number of detectors. In all cases, the solid dots denote the fidelities between the output density matrix \( \rho_f \) and three input pure state density matrices, and the dashed lines denote the fidelities between the three input pure state density matrices.
shown in Figure 3, we plot $|y - s_k^M|_{\min} \to 0$, the meaning of this statement is that, as long as the number of the detectors is sufficient, we can always restrict $w_j$ to be $\pm 1$. We have verified this statement with numerical simulations. As an example, we consider five sources and each of them is an eight-dimensional vector. As shown in Figure 3, we plot $|y - s_k^M|_{\min}$ as a function of $1/M$ and find that it does converge to zero as $M$ increases.

Now we show the ground state spin configuration of this Hamiltonian can determine the solution of the q-CPP. Here, we use the classical Monte-Carlo method to find the ground state approximately. In our simulations, the initial temperature is unity which is the same as the energy unit of the Hamiltonian. During the annealing process, the temperature is reduced epoch by epoch, and in each epoch the temperature is reduced by $\frac{1}{n}$, where $n$ is the epoch number. In each epoch, we randomly flip the spin for 12000 times. In a simulation, it suffers from the problem of being trapped into a local minimum. If so, when we regard $\sigma^2_j$ as $w_j$ and reconstruct density matrix $\rho_f = \sum_j w_j \rho_j$, $\rho_f$ may not be a positive definite matrix. Hence, during the annealing process, we simultaneously run two criteria. We require the von Neumann entropy of the reconstructed density matrix, defined as $-Tr\rho_f \log \rho_f$, to be close to zero, and we also require the reconstructed density matrix to be positive definite. We stop the annealing process at a temperature when these two criteria are well satisfied.

The results of the Monte Carlo calculation are presented in Figure 2(d)-(f). Here we also choose three different input density matrices with fidelity mutually as $F(\rho_1, \rho_2) = 0.13$, $F(\rho_1, \rho_3) = 0.70$ and $F(\rho_2, \rho_3) = 0.45$. Similar to the results by Newton’s method, depending on different initializations, we can find different spin configurations that can construct three different density matrices $\rho_f$. For instance, for the case shown in Figure 2(d), one can find the fidelity $F(\rho_f, \rho_1)$ is very close to unity, and $F(\rho_f, \rho_2)$ and $F(\rho_f, \rho_3)$ are consistent with $F(\rho_1, \rho_2)$ and $F(\rho_1, \rho_3)$, respectively. We can also see that the results get improved as the number of detectors increases. Finally, we should remark that the most efficient way to find out its ground state is through the quantum annealing approach [8].

**Summary and remarks** We have proposed a quantum analogy of the cocktail party problem. The essential point is to replace the classical signal from each source with quantum data. In the classical problem, it is the independent classical signals that are emitted from different sources, which are mixed and detected by the detectors. In the quantum problem, instead, it is the quantum wave functions that are emitted from different sources, and here the “independent” means that the wave functions are orthogonal in part of the Hilbert space and particles are essentially distinguishable. This leads to a reduced mixed state density matrix observed by detectors. In both cases, the goal is to recover the individual sources from the information of detectors only. We also show that solving this problem can be mapped to finding the ground state of an Ising type spin Hamiltonian, and we propose to solve this Hamiltonian by quantum simulation or quantum annealing. We note that, to ensure the uniqueness of the solution, it is preferred to keep both the Hilbert space dimension $d$ of the pure state wave function and the number of detector $M$ large enough. Under this situation, this quantum approach has the advantage that on one hand, the complexity of the Hamiltonian does not increase with the increasing of $d$; and on the other hand, the number of Ising spins equals $M$ and the quantum approach can exhibit its advantage when $M$ is large.

Finally, we should note that the blind signal separation problem is an important subject in quantum information processing. Various versions of blind signal separation problems have been investigated [6, 9-12], but the format of our problem is very different from any of them. As many of these problems discussed under this content, the problem proposed here adds an important twist to the quantum information processing.

This work was supported by the Beijing Outstanding Young Scientist Program, Ministry of Science and Technology (Grant No. 2016YFA0301600), and the National Natural Science Foundation of China (Grant No. 11734010).

---

1. M. A. Bee, and C. Micheyl, *J. Comp. Psychol.* 122, 235 (2008).
2. E. C. Cherry, *J. Acoust. Soc. Am.* 25, 975 (1953).
3. I. P. Waldmann, *Astrophys. J.* 747, 12 (2012).
4. J. Crowder, and N. J. Cornish, *Phys. Rev. D* 75, 043008 (2007).
5. Y. Wu, and H. Zhai, arXiv: 1904.05067.
6. Y. C. Eldar, and A. V. Oppenheim, *IEEE Signal Process. Mag.* 19, 12 (2002).
7. A. Hyvärinen, and E. Oja, *Neural Netw.* 13, 411 (2000).
8. S. E. Venegas-Andraca, W. Cruz-Santos, C. McGeoch, and M. Lanzagorta, *Contemp. Phys.* 59, 174 (2018).
9. A. Deville, and Y. Deville, arXiv: 1611.04002.
10. Y. Deville, and A. Deville, *Quantum Inf. Process.* 11, 1311 (2012).
11. G. M. D’Ariano, R. Demkowicz-Dobrzański, P. Perinotti, and M. F. Sacchi, *Phys. Rev. Lett.* 99, 070501 (2007).
12. G. M. D’Ariano, R. Demkowicz-Dobrzański, P. Perinotti, and M. F. Sacchi, *Phys. Rev. A* 77, 032344 (2008).