

Anticipative measurements in hybrid quantum-classical computation

Teiko Heinosaari

*Quantum Algorithms and Software, VTT Technical Research Centre of Finland Ltd, Espoo, Finland and
Department of Physics and Astronomy, University of Turku, Turku 20014, Finland*

Daniel Reitzner

Quantum Algorithms and Software, VTT Technical Research Centre of Finland Ltd, Espoo, Finland

Alessandro Toigo

*Dipartimento di Matematica, Politecnico di Milano,
Piazza Leonardo da Vinci 32, 20133 Milano, Italy and
Istituto Nazionale di Fisica Nucleare, Sezione di Milano, Via Celoria 16, 20133 Milano, Italy*

Before the availability of large scale fault-tolerant quantum devices, one has to find ways to make the most of current noisy intermediate-scale quantum devices. One possibility is to seek smaller repetitive hybrid quantum-classical tasks with higher fidelity, rather than directly pursuing large complex tasks. We present an approach in this direction where the quantum computation is supplemented by a classical result. While the presence of the supplementary classical information helps alone, taking advantage of its anticipation also leads to a new type of quantum measurements, which we call anticipative. Anticipative quantum measurements lead to improved success rate over cases where we would use quantum measurements optimized without assuming the later arriving supplementing information. Importantly, in an anticipative quantum measurement the combination of the results from classical and quantum computations happens only in the end, without the need for feedback from the one to the other computation, a feature which hence allows for running both computations in parallel. We demonstrate the method using an IBMQ device and show that it leads to an improved success rate even in a real noisy setting.

I. INTRODUCTION

It is a trope of quantum computation to look for problems with a more advantageous complexity than their classical counterparts. With the advent of practical quantum computation, it became apparent that it is worthwhile to look also at problems where even a modest speedup can become useful. At this point, an additional aspect to be aware of is constituted by the limited resources of these devices. In particular, current noisy intermediate-scale quantum (NISQ) devices have too many imperfections to be used for general universal quantum computation [1, 2], hence approaches that are able to deal with these imperfections are sought both in circuits [3–5] and in measurements [6–9]. A possible application of limited NISQ devices is quantum-classical hybrid computation [11–14], where the NISQ device has only a partial role, for example as a subroutine that is able to speed up a repeated computation task. One of the questions arising in these kinds of schemes is how to optimally combine classical and quantum computations so that their overall functioning is as good as possible.

We are considering a class of tasks where one is required to compute the value of a function $g : X \rightarrow Y$ on any given input $x \in X$. Concrete examples of interest could be the following:

- $X = \{2, \dots, n\}$, $Y = \{1, 2, \dots, \lfloor n/2 \rfloor\}$, $g(x)$ = the largest proper divisor of x .
- X = the set of graphs with at most n vertices, $Y = \{1, 2, \dots, n\}$, $g(x)$ = the chromatic number of x .

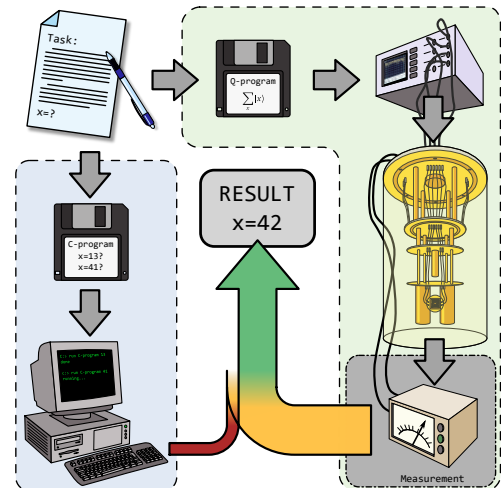


FIG. 1. We consider hybrid computation where classical and quantum computations are performed simultaneously. The outputs of the two computations are then combined to reduce the possible final answers and to provide a more precise output than from either of the two computations.

In the current work, we do not concentrate on any particular problem, but rather on a method that is applicable to a variety of such tasks. We focus on a scheme where the computation has two parallel parts: a classical part and a quantum part. The essential assumption in our investigation is that these computations are run in parallel, meaning that they take an input at the same time and

the computations are carried out simultaneously. The combination of the computations and the final inference happens only after both parts have been completed (see Fig. 1). The question is then if there is a way to optimize the two parts jointly that would lead to a better performance than individually optimizing each part.

The motivation for this parallel computing assumption is the following. Naturally, we want the total computing time to be minimal, which can be achieved by increasing the total success rate. Considering current NISQ devices, the motivation may be even more practical. The available quantum computer may be small and less efficient than the computational problem would require, leading to intrinsically imperfect computation. A classical computer may provide a useful aid even if it cannot solve the problem efficiently alone. For instance, the classical computation can be used to rule out some wrong answers, which is much more easier than to solve the question in full. This partial information can still help the quantum computation part, and that is exactly the main idea of the current study.

We note that the method we are introducing does not prevent from having also sequential steps in the total computation process. Indeed, the method can be thought to be applied separately in all phases. We can even find further motivation in the sequential arrangement, since this kind of hybrid tasks uses classical computation between separate quantum computations and so, during each run of the quantum computation, classical computers are idle. In fact, the advantage of our approach is that it allows to make use of this ineffective time.

II. PARALLEL QUANTUM-CLASSICAL COMPUTATION WITH ANTICIPATIVE MEASUREMENTS

A. Parallel quantum-classical computation

In our setting, the quantum computing part uses a d -level quantum system in the encoding of inputs and this size evidently limits the power of the quantum device. Each input x is encoded into a quantum state ϱ_x . Then, the quantum system is acted on by a quantum process Φ , hence the state is transformed into $\Phi(\varrho_x)$. The process can be a complicated array of quantum gates, it can use higher dimensional ancillary systems, and the produced states may possibly be not orthogonal. Further, the transformed state $\Phi(\varrho_x)$ can be a state of a different quantum system having a different dimension than d . The system in the transformed state is measured by a quantum measurement \mathbf{M} and an outcome z is recorded. The conditional probability to obtain z is given by the Born rule

$$p(z|x) = \text{tr}[\Phi(\varrho_x)\mathbf{M}(z)]. \quad (1)$$

In the most favorable case, one has $p(z|x) = \delta_{z,g(x)}$, i.e., the recorded outcome is the value $g(x)$ that we wanted to

compute. However, due to the limited size of the quantum computer and the noise involved in the process and in the final measurement, we expect $p(z|x)$ to be a probability distribution that has nonzero variance. The quantum computing part is therefore giving a probabilistic information about the correct answer for the value $g(x)$.

In addition to the quantum computation, we assume that there is a classical computation part where the aim is to rule out one or more wrong answers. We denote by $S \subset Y$ the set of values that the classical computation determines to be wrong. For example, in NP-problems checking answers is a fast process that can be used to generate the set S . Getting back to our earlier examples, if $g(x)$ is the largest proper divisor of an integer $x \geq 2$, the set S may be obtained by performing divisibility tests for some choice of potential divisors. Or, if x is a graph and $g(x)$ is the chromatic number of x , S could be the result of some random coloring heuristics carried out on x . The aim of the current investigation is not to identify specific problems where the classical computation can be explicitly described, and we actually let the set S be any non-trivial subset of Y .

The limits on the computation time and the efficiency of the classical computer may imply that S is a singleton set or has some other fixed size. In a typical application that we have in mind, S is a small set compared to the size of Y . The information in the form of S is provided only after the quantum part has concluded and we will therefore call it the *classical posterior information*, abbreviated to CPOST. We further assume that the classical computation is error-free, so that $g(x) \notin S$. However, we allow it to be non-deterministic, and we denote the conditional probability to obtain S given x as $p(S|x)$. For instance, if we would use a quantum computer to speed up some NP-problem, alongside we could run some number of classical computations in which we may randomly choose potential answers $y_1, y_2, \dots, y_m \in Y$, and either learn that one of our choices is the correct answer, or learn that $g(x) \notin S = \{y_1, y_2, \dots, y_m\}$. As the classical and quantum computations work in parallel and do not interact up to this point, we further require that the set S of the excluded answers is independent of the measurement outcome z when conditioned on the input x , that is, $p(S, z|x) = p(S|x)p(z|x)$.

The final step is then to combine the results from the two parts and make the final guess y , based on the outcome z of the quantum computation and the set S consisting of wrong values. In general, again, this guess is probabilistic and described by the conditional probability $p(y|S, z)$. Moreover, it depends on the input x solely through the classical and quantum computations, meaning that $p(y|S, z) = p(y|S, z, x)$. The efficiency of the full computation is characterized by its success probability, i.e., the probability of having $y = g(x)$ when the input x is sampled from X . For simplicity, we will consider uniform sampling, although we remark that this is not a necessary requirement in our approach.

Let us now discuss different ways of utilizing poste-

rior information. The setting provides a four-fold distinction between different scenarios. Firstly, we can distinguish between the cases in which CPOST is present or absent. A motivation for this distinction can be the case when CPOST might improve the quality of the result from the quantum computation, but the result might be useful even without CPOST. Secondly, and more importantly, we distinguish the cases where the quantum measurement is optimized without considering CPOST (the resulting measurements will be called *standard*) and where the quantum measurement is optimized by having CPOST in mind (measurements of this kind will be called *anticipative*).

The focus of this paper lies on the anticipative measurements with CPOST. Our method thus optimizes the quantum and classical steps together instead of treating them separately, even if they are run independently in parallel. In practice, this means that we adjust the measurement performed in the end of the quantum computation. The crux is to take into account that there will be CPOST, in our case in the form of wrong answers, before we have to make the final decision, although we cannot know the specific wrong answers before we have to perform the measurement.

It is not evident from this general description that the anticipative measurement method works and, in fact, it does not necessarily lead to a better success probability than the independently optimized parts. However, we demonstrate with examples that the anticipative measurement does give a benefit on a class of problems. The main goal of the current investigation is to show that the anticipative method can be better than the standard method and therefore might be a valuable tool in hybrid quantum-classical computation.

To summarize, for a given task, we have four scenarios to which we assign the respective success probabilities \mathbb{P} .

1. \mathbb{P}_0^{st} : standard quantum measurement without CPOST. In this scenario, we use quantum computation only.
2. $\mathbb{P}_{\text{CPOST}}^{\text{st}}$: standard quantum measurement with CPOST. This means that we run both quantum and classical computations, optimize them independently and combine in the end.
3. $\mathbb{P}_{\text{CPOST}}^{\text{an}}$: anticipative quantum measurement with CPOST. In this scenario, quantum measurement is optimized already having in mind the later arrival of classical information. This scenario is our main interest.
4. \mathbb{P}_0^{an} : anticipative quantum measurement without CPOST. This scenario is for comparison only. One can think of it as a scenario where classical computation breaks down and does not give any information even if we were waiting for it and hence chose anticipative measurement instead of the standard measurement.

In general, we have

$$\mathbb{P}_0^{\text{an}} \leq \mathbb{P}_0^{\text{st}} \leq \mathbb{P}_{\text{CPOST}}^{\text{st}} \leq \mathbb{P}_{\text{CPOST}}^{\text{an}}. \quad (2)$$

The first and last inequalities follow from the definitions. Indeed, the standard measurement means the optimal measurement for the problem without CPOST and the anticipative measurement means the optimal measurement for the task with CPOST. The middle inequality is true as posterior information cannot make the guessing probability worse if properly optimized (optimal solution must be at least as good as if one does not act based on CPOST). Depending on the task in question, these inequalities may be equalities, which would mean that for that task posterior information or the anticipative method do not help. Clearly, the anticipative method becomes interesting in tasks where $\mathbb{P}_{\text{CPOST}}^{\text{st}} < \mathbb{P}_{\text{CPOST}}^{\text{an}}$. In Section III, we demonstrate that this is actually the case already in a simple class of tasks.

B. Mathematical framework for anticipative measurements

We now present the mathematical framework for anticipative measurements that is needed in the applications to concrete problems.

We denote $X = \{1, \dots, m\}$, $Y = \{1, \dots, n\}$ and let $f : X \times Y \rightarrow \{0, 1\}$ be the function that defines the computational task by determining the wanted and unwanted input-output pairs $(x, y) \in X \times Y$. If each $x \in X$ is associated to only one correct answer $g(x) \in Y$, then the wanted pairs constitute the set $\{(x, g(x)) : x \in X\}$ and the task is defined by the Kronecker delta function $f(x, y) = \delta_{g(x), y}$. More generally, an input x may have several correct answers, which therefore constitute a subset $G_x \subset Y$. In this case, we choose $f(x, y) = 1_{G_x}(y)$, where 1_{G_x} is the indicator function of the set G_x .

In the current setting, we do not separate the initial quantum state encoding and the quantum process that transforms the states. This is due to the fact that the anticipative method alters only the final measurement and therefore only the form of the quantum states just before the measurement matters. We denote $\mathcal{E}(x) = (1/m)\Phi(\varrho_x)$, so that the mapping $x \mapsto \mathcal{E}(x)$ describes all what happens to the input x before the quantum measurement is performed. We call \mathcal{E} the *state ensemble* of our computational task. The uniform sampling of the inputs is included in the normalization constant $1/m$ of \mathcal{E} . A quantum measurement is mathematically described as a positive operator valued measure \mathbb{M} . Denoting by Z the outcome set of \mathbb{M} , the probability of choosing the input $x \in X$ and getting the outcome $z \in Z$ is hence $\text{tr}[\mathcal{E}(x)\mathbb{M}(z)]$.

On the classical side, we write $\alpha(S|x) = p(S|x)$ to denote the conditional probability of getting S from the classical computation when $x \in X$ is the input. We say that α is the *partial information map* of our quantum-classical hybrid computation. Note that the formalism

allows S can be any information about the task, not just the exclusion of outcomes, but for simplicity we just deal with the exclusion.

Finally, the combination of classical and quantum outcomes is a post-processing that gives the final answer $y \in Y$ with probability $\nu_S(y|z) = p(y|S, z)$. We refer to ν as the *post-processing map*. With these notations, the probability of guessing a correct answer is

$$\mathbb{P}_{\text{cpost}} = \sum_{x,y,z,S} f(x,y) \nu_S(y|z) \alpha(S|x) \text{tr}[\mathcal{E}(x)\mathbf{M}(z)]. \quad (3)$$

We typically assume that the state ensemble \mathcal{E} and the partial information map α are fixed, but we need to find a quantum measurement \mathbf{M} and a post-processing map ν performed after CPOST, so that the success probability of Eq. (3) is as high as possible.

Note, that even though Eq. (3) describes the success rate including the arrival of CPOST, its two extreme cases also describe situations when no classical information is provided, or when all wrong answers are excluded. The latter case means that $\alpha(S|x) = 1$ if $S = Y \setminus G_x$, where G_x is the set of the correct answers for the input x . The optimal post-processing map then necessarily satisfies $\nu_{Y \setminus G_x}(y|z) = 0$ for all $y \notin G_x$, and the final guess y can be chosen independently of the measurement outcome z . In this way, when all wrong answers are excluded, we can trivially achieve the equality $\mathbb{P}_{\text{cpost}} = 1$ for any measurement \mathbf{M} . The former case of a situation with no CPOST, instead, corresponds to setting $\alpha(S|x) = \alpha(S)$ independently of the input x . In this case, the summation over S can be carried out on the post-processings ν_S and the right-hand side of Eq. (3) becomes the probability

$$\mathbb{P}_0 = \sum_{x,y,z} f(x,y) \nu_0(y|z) \text{tr}[\mathcal{E}(x)\mathbf{M}(z)], \quad (4)$$

where

$$\nu_0(y|z) = \sum_S \nu_S(y|z) \alpha(S). \quad (5)$$

Different optimizations of the previous equations then lead to the cases described above:

$$\mathbb{P}_{\text{cpost}}^{\text{an}} = \max_{\mathbf{M}, \nu} \mathbb{P}_{\text{cpost}}(\mathbf{M}, \nu), \quad (6)$$

$$\mathbb{P}_0^{\text{st}} = \max_{\mathbf{M}, \nu_0} \mathbb{P}_0(\mathbf{M}, \nu_0), \quad (7)$$

where in the right-hand side we have explicitly indicated the dependence of the probabilities $\mathbb{P}_{\text{cpost}}$ and \mathbb{P}_0 on the chosen measurement \mathbf{M} and the post-processings ν and ν_0 . If we label the respective optimized measurements as \mathbf{M}^{an} and \mathbf{M}^{st} , we further have

$$\mathbb{P}_0^{\text{an}} = \max_{\nu_0} \mathbb{P}_0(\mathbf{M}^{\text{an}}, \nu_0), \quad (8)$$

$$\mathbb{P}_{\text{cpost}}^{\text{st}} = \max_{\nu} \mathbb{P}_{\text{cpost}}(\mathbf{M}^{\text{st}}, \nu). \quad (9)$$

C. Anticipative measurements with the exclusion of k wrong answers

For simplicity, from now on we assume that the number of wrong answers obtained by means of classical computation is fixed and we denote this number by k . Hence, the set S of the excluded answers is an element of the set

$$T = \{S \subset Y : |S| = k\}, \quad (10)$$

which in turn is a subset of the power set 2^Y . We further assume that S is drawn with uniform probability from the subset of all the elements of T which are disjoint from G_x .

Having written the computational tasks in Eq. (3), we can interpret it as a quantum guessing game with posterior information and apply the mathematical results developed in [15, 16]. The same mathematical formalism can be used to construct and study incompatibility witnesses, but here our aim is different. One of the main facts contained in the aforementioned works is that any quantum guessing game with posterior information reduces to a usual state discrimination task for an auxiliary state ensemble. The reduction means that, even if the tasks are different, the optimal measurements are the same and the success probabilities are connected via a simple formula. To see the reduction in practice, let us first assume that $k = 1$, i.e., the classical computation part is giving one wrong answer. Moreover, let us suppose that each input x is associated to only one correct answer $g(x)$. In this case, we can identify the set T of Eq. (10) with Y and write the partial information map as $\alpha(t|x) = (1 - \delta_{g(x),t})/(n-1)$. The auxiliary state ensemble, denoted by $\bar{\mathcal{E}}$, is then defined on the Cartesian product Y^n and given as

$$\begin{aligned} \bar{\mathcal{E}}(y_1, \dots, y_n) &= \frac{1}{|Y|^{|T|-1} \Delta} \sum_{x,t} f(x, y_t) \alpha(t|x) \mathcal{E}(x) \\ &= \frac{1}{(n-1)n^{n-1}} \sum_x r(x) \mathcal{E}(x), \end{aligned} \quad (11)$$

where

$$\Delta = \sum_{x,y} f(x,y) \text{tr}[\mathcal{E}(x)] = 1, \quad (12)$$

$$r(x) = |\{t \in Y : y_t = g(x) \text{ and } t \neq g(x)\}|. \quad (13)$$

With the known techniques of minimum-error state discrimination [19, 20], one can then find the optimal measurement $\bar{\mathbf{M}}$ that yields the largest success probability in discriminating $\bar{\mathcal{E}}$. It turns out that the same measurement $\bar{\mathbf{M}}$ is optimal also for the original guessing game. Indeed, the interpretation of $\bar{\mathbf{M}}$ in the initial setting is that we obtain a measurement outcome that is a tuple, namely $(y_1, \dots, y_n) \in Y^n$, and the wrong answer given as the posterior information t from the classical computation refines this outcome to the final guess y_t . The success probabilities of the original task and of the auxiliary state discrimination task are not the same but have a

simple relation: if the success probability in the auxiliary state discrimination problem is \mathbb{P} , then

$$\mathbb{P}_{\text{CPOST}}^{\text{an}} = n^{n-1} \mathbb{P}. \quad (14)$$

The mathematical machinery works in the same way also in the cases with $k > 1$. Only the form of the auxiliary state ensemble $\tilde{\mathcal{E}}$ is different and it can be obtained from [16, Sec. 4.4].

III. QUBIT APPLICATION

While the intention of the anticipative method is to be used for hybrid computation tasks, in order to illustrate the method and test it in current quantum devices we consider a simple example having no direct computational interest, but exhibiting the same quantum features that may appear during quantum computations with a NISQ device. In this example, two bits of information are encoded into one qubit, hence the size of the quantum device is smaller than one would need for perfect encoding. After all quantum processing, we assume that the final states are from two bases. The task is to infer the values of the original bits with as high success probability as possible.

Qubit states can be identified with vectors on the Bloch sphere. The final states belong to two bases and the angle between these bases in the Bloch sphere description is parametrized by $\theta \in (0, \pi/2]$. There is a global unitary freedom to choose the directions, hence we can fix the states corresponding to the Bloch vectors $\pm \mathbf{a}$ and $\pm \mathbf{b}$ with

$$\begin{aligned} \mathbf{a} &= \cos\left(\frac{1}{2}\theta\right) \mathbf{i} + \sin\left(\frac{1}{2}\theta\right) \mathbf{j}, \\ \mathbf{b} &= \cos\left(\frac{1}{2}\theta\right) \mathbf{i} - \sin\left(\frac{1}{2}\theta\right) \mathbf{j}, \end{aligned} \quad (15)$$

where \mathbf{i} and \mathbf{j} are two orthogonal coordinate vectors. The setting is depicted in Fig. 2. With respect to the framework explained earlier, we have

$$X = \{+a, -a, +b, -b\} \quad (16)$$

and the state ensemble \mathcal{E} is

$$\begin{aligned} \mathcal{E}(\pm \mathbf{a}) &= \frac{1}{8} (\mathbb{1} \pm \mathbf{a} \cdot \boldsymbol{\sigma}), \\ \mathcal{E}(\pm \mathbf{b}) &= \frac{1}{8} (\mathbb{1} \pm \mathbf{b} \cdot \boldsymbol{\sigma}). \end{aligned} \quad (17)$$

Although the states within each of the two bases are orthogonal and hence perfectly distinguishable, the four states together are not. We will look now on the minimum-error discrimination schemes with or without CPOST both for the standard measurements and the anticipative measurements. In the following, we denote by \mathbb{P}_k^{st} and \mathbb{P}_k^{an} the success probabilities when CPOST consists of k wrong answers. This notation agrees also with the previous case of no CPOST which corresponds to $k = 0$. Furthermore, we take $Y = X$ and $f(x, y) = \delta_{x,y}$, which stems from the definition of the problem and our intention to identify the states with highest probability.

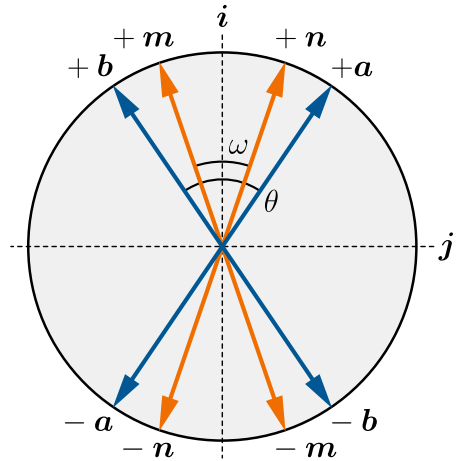


FIG. 2. Discrimination problem for four qubit states labelled by $\pm \mathbf{a}$ and $\pm \mathbf{b}$. The standard optimal measurement is a probabilistic projective measurement in the two bases given by the four states. The anticipative measurement is a probabilistic projective measurement in the two bases given by the vectors $\pm \mathbf{m}$ and $\pm \mathbf{n}$.

A. Standard measurement

One of the optimal quantum measurements to distinguish the four different inputs (without CPOST) is

$$\begin{aligned} \mathbf{M}^{\text{st}}(\pm \mathbf{a}) &= \frac{1}{4} (\mathbb{1} \pm \mathbf{a} \cdot \boldsymbol{\sigma}), \\ \mathbf{M}^{\text{st}}(\pm \mathbf{b}) &= \frac{1}{4} (\mathbb{1} \pm \mathbf{b} \cdot \boldsymbol{\sigma}). \end{aligned} \quad (18)$$

This measurement results from the maximization of Eq. (4) with the fixed postprocessing $\nu_0(y|z) = \delta_{y,z}$, as can be checked e.g. by testing the optimality conditions in [19, Eqs. (4)-(5)] (the latter choice of ν_0 is not restrictive, since we can always include the post-processing in the optimized measurement). For the forthcoming calculations we need all the input-output probabilities of the Born rule $\text{tr}[\mathcal{E}(x)\mathbf{M}^{\text{st}}(z)]$, which are given in the following table:

input (x)	measurement outcome (z)			
	$+a$	$-a$	$+b$	$-b$
$+a$	$\frac{1}{8}$	0	$\frac{1}{8} \cos^2 \frac{\theta}{2}$	$\frac{1}{8} \sin^2 \frac{\theta}{2}$
$-a$	0	$\frac{1}{8}$	$\frac{1}{8} \sin^2 \frac{\theta}{2}$	$\frac{1}{8} \cos^2 \frac{\theta}{2}$
$+b$	$\frac{1}{8} \cos^2 \frac{\theta}{2}$	$\frac{1}{8} \sin^2 \frac{\theta}{2}$	$\frac{1}{8}$	0
$-b$	$\frac{1}{8} \sin^2 \frac{\theta}{2}$	$\frac{1}{8} \cos^2 \frac{\theta}{2}$	0	$\frac{1}{8}$

The success probability for the standard measurement without CPOST given by Eq. (7) simply reduces to

$$\mathbb{P}_0^{\text{st}} = \sum_x \text{tr}[\mathcal{E}(x)\mathbf{M}^{\text{st}}(x)] = \frac{1}{2}. \quad (19)$$

This is the sum of the probabilities from the diagonal of the table.

By using Eq. (9), we can also infer the success probabilities \mathbb{P}_k^{st} when k outcomes are excluded by CPOST. The optimal strategy consists in guessing the non-excluded input having the highest probability conditioned to the outcome of the measurement \mathbf{M}^{st} . For $k = 1, 2$ one then obtains the best guess following the mapping rules of the next table:

measurement outcome (z)	relabeling priority (y)			
	1 st	2 nd	3 rd	4 th
$+a$	$+a$	$+b$	$-b$	$-a$
$-a$	$-a$	$-b$	$+b$	$+a$
$+b$	$+b$	$+a$	$-a$	$-b$
$-b$	$-b$	$-a$	$+a$	$+b$

This mapping is determined based on the mapping priority where only non-excluded outcomes from the lower right corner of the table are used. For example, having obtained the outcome $+a$ and excluded the outcomes $\{+a, -b\}$, the next available choice is $+b$. If $\{+a, +b\}$ are excluded, the next best choice is $-b$. This observation has also a visual interpretation. Looking at the setting from Fig. 2, we see that the mapping priority is ordered from the closest state to the orthogonal one. Performing the calculations we obtain

$$\mathbb{P}_1^{\text{st}} = \frac{1}{6} \left(3 + \cos^2 \frac{\theta}{2} \right), \quad (20)$$

$$\mathbb{P}_2^{\text{st}} = \frac{1}{6} \left(4 + \cos^2 \frac{\theta}{2} \right). \quad (21)$$

B. Anticipative measurement

As explained in Section II B, the anticipative measurement strategy for the given qubit state ensemble with the exclusion of one or two wrong answers can be calculated with the mathematical framework developed in [16]. If there is one excluded answer, then the problem is the same as the quantum guessing game studied in [16, Sec. 7.4] and one of the corresponding anticipative quantum measurement was found to be

$$\begin{aligned} \mathbf{M}^{\text{an}}(\pm \mathbf{m}) &= \frac{1}{4} (\mathbf{1} \pm \mathbf{m} \cdot \boldsymbol{\sigma}), \\ \mathbf{M}^{\text{an}}(\pm \mathbf{n}) &= \frac{1}{4} (\mathbf{1} \pm \mathbf{n} \cdot \boldsymbol{\sigma}), \end{aligned} \quad (22)$$

where

$$\begin{aligned} \mathbf{m} &= \frac{\mathbf{a} + 3\mathbf{b}}{\sqrt{10 + 6 \cos \theta}}, \\ \mathbf{n} &= \frac{3\mathbf{a} + \mathbf{b}}{\sqrt{10 + 6 \cos \theta}}. \end{aligned} \quad (23)$$

In Appendix A, we summarize that proof and, in addition, we show that the measurement of Eq. (22) is optimal also in the case with $k = 2$, i.e., when classical information excludes two wrong answers. The Born rule probabilities $\text{tr}[\mathcal{E}(x)\mathbf{M}^{\text{an}}(z)]$ are the following:

input (x)	measurement outcome (z)			
	$+m$	$-m$	$+n$	$-n$
$+a$	Q_+	Q_-	P_+	P_-
$-a$	Q_-	Q_+	P_-	P_+
$+b$	P_+	P_-	Q_+	Q_-
$-b$	P_-	P_+	Q_-	Q_+

Here we have denoted

$$\begin{aligned} P_{\pm} &= \frac{1}{16} \left(1 \pm \frac{1 + 3 \cos \theta}{\sqrt{10 + 6 \cos \theta}} \right), \\ Q_{\pm} &= \frac{1}{16} \left(1 \pm \frac{\cos \theta + 3}{\sqrt{10 + 6 \cos \theta}} \right). \end{aligned} \quad (24)$$

It is easy to check that $Q_+ \geq P_+ \geq P_- \geq Q_-$.

Similarly as for the standard measurements, we can visualize the assignment of the measured results to the final guess as a table. Since the input-output probabilities are ordered in the same way as in the standard measurement case and differ only in their magnitude, the mapping follows the same ordering:

measurement outcome (z)	relabeling priority (y)			
	1 st	2 nd	3 rd	4 th
$+n$	$+a$	$+b$	$-b$	$-a$
$-n$	$-a$	$-b$	$+b$	$+a$
$+m$	$+b$	$+a$	$-a$	$-b$
$-m$	$-b$	$-a$	$+a$	$+b$

This yields the following expressions for the success probabilities defined in Eq. (6):

$$\mathbb{P}_1^{\text{an}} = \frac{1}{12} \left(4 + \sqrt{10 + 6 \cos \theta} \right), \quad (25)$$

$$\mathbb{P}_2^{\text{an}} = \frac{1}{12} \left(6 + \sqrt{10 + 6 \cos \theta} \right). \quad (26)$$

Finally, to calculate the success probability \mathbb{P}_0^{an} defined in Eq. (8), we observe that the optimal assignment without CPOST is $\pm \mathbf{m} \mapsto \pm \mathbf{b}$ and $\pm \mathbf{n} \mapsto \pm \mathbf{a}$, which again consists in guessing the non-excluded input having the highest probability conditioned on the outcome of the measurement \mathbf{M}^{an} . In this way, we get

$$\mathbb{P}_0^{\text{an}} = 4Q_+ = \frac{1}{4} \left(1 + \frac{\cos \theta + 3}{\sqrt{10 + 6 \cos \theta}} \right). \quad (27)$$

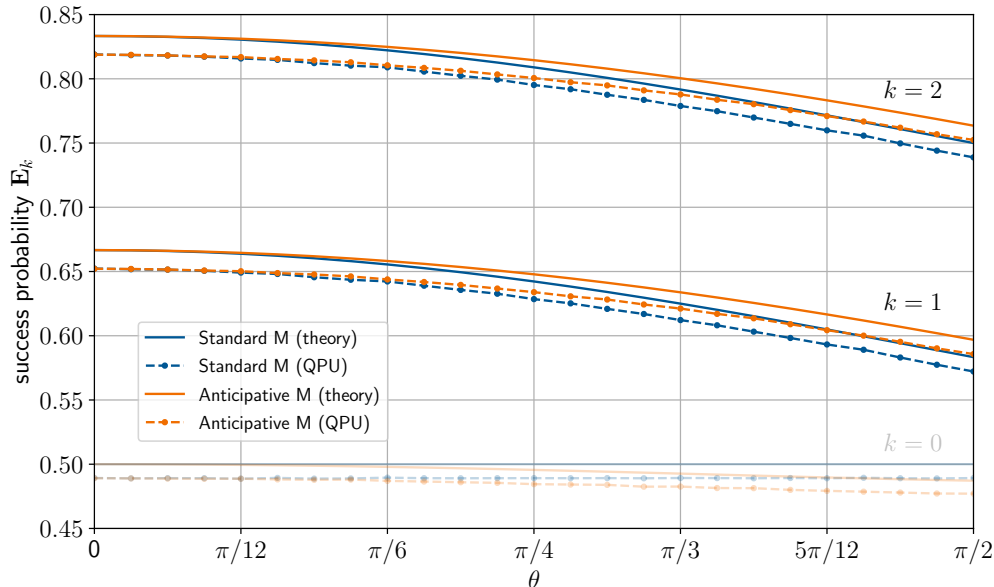


FIG. 3. Success probabilities of the discrimination task for four qubit states from two bases spanning the angle θ . The comparison is made between standard (blue) and anticipative (orange) measurements for different posterior information — $k = 0$ for measurements without CPOST, $k = 1$ for CPOST excluding one wrong answer and $k = 2$ for CPOST excluding two wrong answers. Solid lines correspond to the theoretical predictions and dashed lines are results from the *ibmq_manila*.

All the studied situations are depicted for different θ angles in Fig. 3 (solid lines). We can notice that while for the case without CPOST ($k = 0$) the anticipative measurement operates worse than the standard measurement, with CPOST ($k = 1, 2$) the anticipative measurement improves the success probability.

C. Demonstrations

The investigated state discrimination problem can be easily applied to current quantum devices. The possibility to do so comes from the fact that the measurements involved, whether standard or anticipative, are probabilistic projective measurements. These measurements, a subclass of projective simulable measurements [21, 22], are implementable as a random choice from the set of projective measurements [23]. In the qubit case, this relates to a random choice of a base change prior to the standard z -measurement. For example, the statistics of the measurement M^{an} can be obtained as a union of the statistics of the projective measurement in the direction $\pm \mathbf{m}$ and the projective measurement in the direction $\pm \mathbf{n}$ with equal number of runs. If the measurements would not belong to the subclass of projective simulable measurements, we could still implement them but that would require ancillary dimensions.

Using IBMQ device *ibmq_manila*, we implemented the

measurements in Qiskit, where for the rotations we used only RY gates (more information in Appendix B). Since the native gate set does not contain the RY gate, prior to the computation it needs to be decomposed using the SX and RZ native gates. This was automatically performed by the transpiler.

For the standard measurement in the basis $\pm \mathbf{a}$ we used the angle $+\theta/2$, while for the measurement in the basis $\pm \mathbf{b}$ we used the angle $-\theta/2$. For the anticipative measurement we needed to rotate the basis either to the $\pm \mathbf{n}$ direction by the angle $+\omega/2$ or to the $\pm \mathbf{m}$ direction by the angle $-\omega/2$. The angle ω is obtained from the dot product of the vectors \mathbf{m} and \mathbf{n} defined in Eq. (23):

$$\cos \omega = \mathbf{m} \cdot \mathbf{n} = \frac{3 + 5 \cos \theta}{5 + 3 \cos \theta}. \quad (28)$$

As explained earlier, this angle remains the same in both cases with $k = 1$ and $k = 2$.

The submissions to *ibmq_manila* contained circuits with 25 different choices for the parameter θ , 4 choices of the initial state ($\pm \mathbf{a}$ or $\pm \mathbf{b}$), 2 choices for the measurement (anticipative / standard), and 2 choices for the measurement basis (\mathbf{a} or \mathbf{b} for standard and \mathbf{m} or \mathbf{n} for anticipative), thus leading to 400 circuits, which were run for 20,000 repetitions each. The preparation scheme of these circuits is illustrated in Fig. 4.

All computations were performed on qubit 0 with applied optimization for the circuits, possibly combining

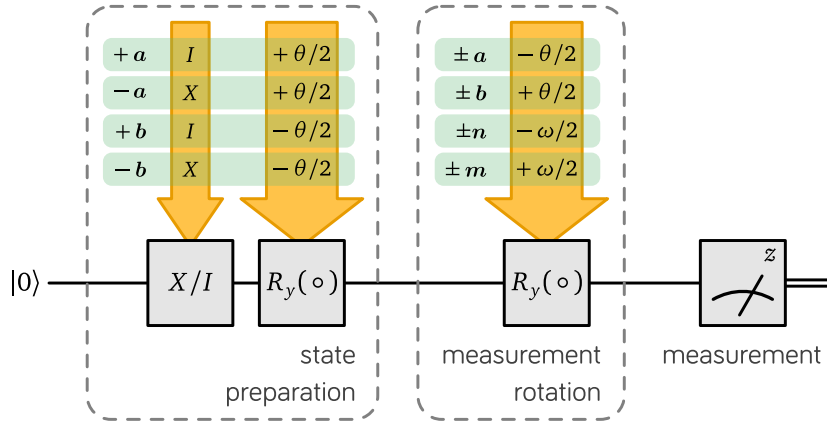


FIG. 4. Diagram of circuit preparation for the data collection phase. For each choice of the angle θ , the prepared state and the measurement direction, one circuit was constructed and executed 20,000 times.

preparation and measurement rotations into a more convenient form. Collected data were then post-processed to obtain the probabilities. The post-processings were not simulated any more but were computed from obtained data — each datapoint (executed circuit) was used to obtain the success probability for all allowed exclusions (those not containing the prepared state) and, therefore, it was used multiple times. The number of repetitions was high enough to make variances minimal and so there was no need to keep track of the precise number of shots for each measurement and post-processing, allowing us to extract probabilities this way.

Data obtained in the demonstration is depicted in Fig. 3 as dashed lines. We can observe the general behavior that was theoretically described above — the anticipative measurement is worse than the standard measurement in the absence of CPOST, but it shows improvement over the standard measurement with CPOST at hand. An interesting point is the region close to $\pi/2$, where we can not only see that the anticipative measurement is better than the standard measurement (for $k = 1$ and $k = 2$) but, even more, the results of the demonstration show an improvement over the theoretically best standard measurement.

IV. DISCUSSION

In this paper we have investigated a specific type of hybrid quantum-classical computation — parallel quantum-classical computation where one performs quantum and classical computations simultaneously and combines the results in the end. Quantum computation necessarily ends with a quantum measurement and we have focused on this particular part of the process. We have introduced a method of anticipative quantum measurements, where the quantum measurement is optimized not in isolation but taking into account that there will be classical supplementing information later. Crucially, the op-

timization does not depend on the actual outcome of the classical information but just on the general form of it. As an exemplary case we highlighted classical posterior information that rules out some wrong answers. We demonstrated that the method works even with a noisy real quantum device.

The success of the anticipative method raises some immediate thoughts and questions. Firstly, the scaling properties of the approach are not studied in this paper, but are important for further considerations. With the scaled dimensions and scaled number of outcomes, we generally observe scaling of the algorithmic time. This in turn provides space for additional (time) resources of the parallel classical computation that shall provide a more informative posterior information. It remains for further research to ascertain to what extent a measurable advantage is retained. It might well be that this question is even specific to the chosen task. Still, in the worst case, anticipative measurements might find their place in computations on current NISQ devices with limited resources and small tasks.

Secondly, one could possibly anticipate the classical posterior information not only in the last step of the quantum computation, namely in the quantum measurement, but also in the earlier steps of it. This would mean that one would modify the quantum process as a whole, allowing for more freedom in the optimization and thus potentially leading even to higher success rates. A drawback is that the method cannot then anymore be developed in the general form as we have done here, but it should take into account the actual computational task.

Finally, we have observed several peculiar properties in the specific exemplary qubit task. Namely, we saw that the anticipative measurement is the same for the cases of excluding one wrong answer and two wrong answers although in general every type of CPOST can lead to a different solution. Further, in these cases the anticipative measurement has only four outcomes that are nonzero. As explained in Section II C the anticipative measure-

ment is optimal also with respect to the auxiliary state ensemble discrimination, hence in both cases one would have expected more outcomes than four. A further investigation will show to what extent these observations are general and to what extent they are specific to the task.

V. ACKNOWLEDGMENT

We acknowledge the use of IBM Quantum services for this work. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team.

-
- [1] J. Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018.
- [2] K. Bharti, A. Cervera-Lierta, T.H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J.S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek and A. Aspuru-Guzik. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.*, 94: 015004, 2022.
- [3] J. Zhang, A.M. Souza, F.D. Brandao, and D. Suter. Protected Quantum Computing: Interleaving Gate Operations with Dynamical Decoupling Sequences. *Phys. Rev. Lett.*, 112:050502, 2014.
- [4] C.L. Edmunds, C. Hempel, R.J. Harris, V. Frey, T.M. Stace, and M.J. Biercuk. Dynamically corrected gates suppressing spatiotemporal error correlations as measured by randomized benchmarking. *Phys. Rev. Research*, 2:013156, 2020.
- [5] Y. Baum, M. Amico, S. Howell, M. Hush, M. Liuzzi, P. Mundada, T. Merkh, A.R.R. Carvalho, and M.J. Biercuk. Experimental Deep Reinforcement Learning for Error-Robust Gate-Set Design on a Superconducting Quantum Computer. *PRX Quantum*, 2:040324, 2021.
- [6] H. Kwon and J. Bae. A hybrid quantum-classical approach to mitigating measurement errors in quantum algorithms. *IEEE Trans. Comput.*, 70:1401–1411, 2020.
- [7] F.B. Maciejewski, Z. Zimborás, and M. Oszmaniec. Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography. *Quantum*, 4:257, 2020.
- [8] B. Nachman, M. Urbanek, W.A. de Jong, and C.W. Bauer. Unfolding quantum computer readout noise. *npj Quantum Inf.*, 6:84, 2020.
- [9] M.R. Geller. Rigorous measurement error correction. *Quantum Sci. Technol.*, 5:03LT01, 2020.
- [10] A. Eddins, M. Motta, T.P. Gujarati, S. Bravyi, A. Mezzacapo, C. Hadfield and S. Sheldon. Doubling the size of quantum simulators by entanglement forging. *PRX Quantum*, 3:010309, 2022.
- [11] S. Bravyi, G. Smith, and J.A. Smolin. Trading classical and quantum computational resources. *Phys. Rev. X*, 6:021043, 2016.
- [12] V. Dunjko, Y. Ge, and I.J. Cirac. Computational speedups using small quantum devices. *Phys. Rev. Lett.*, 121:250501, 2018.
- [13] P.P. Angara, U. Stege, H.A. Müller, and M. Bozzo-Rey. Hybrid quantum-classical problem solving in the NISQ era. In *Proceedings of the 30th Annual International Conference on Computer Science and Software Engineering*, pages 247–252, 2020.
- [14] A. Callison and N. Chancellor. Hybrid quantum-classical algorithms in the noisy intermediate-scale quantum era and beyond. *Phys. Rev. A*, 106:010101, 2022.
- [15] C. Carmeli, T. Heinosaari, and A. Toigo. State discrimination with postmeasurement information and incompatibility of quantum measurements. *Phys. Rev. A*, 98:012126, 2018.
- [16] C. Carmeli, T. Heinosaari, and A. Toigo. Quantum guessing games with posterior information. *Rep. Progr. Phys.*, 85:074001, 2022.
- [17] C. Carmeli, T. Heinosaari, and A. Toigo. Quantum incompatibility witnesses. *Phys. Rev. Lett.*, 122:130402, 2019.
- [18] C. Carmeli, T. Heinosaari, T. Miyadera and A. Toigo. Witnessing incompatibility of quantum channels. *J. Math. Phys.*, 60:122202, 2019.
- [19] J. Bae. Structure of minimum-error quantum state discrimination. *New J. Phys.*, 15:073037, 2013.
- [20] D. Ha and Y. Kwon. Discriminating N -qudit states using geometric structure. *Phys. Rev. A*, 90:022330, 2014.
- [21] M. Oszmaniec, L. Guerini, P. Wittek, and A. Acín. Simulating positive-operator-valued measures with projective measurements. *Phys. Rev. Lett.*, 119:190501, 2017.
- [22] L. Guerini, J. Bavaresco, M. T. Cunha, and A. Acín. Operational framework for quantum measurement simulability. *J. Math. Phys.*, 58:092102, 2017.
- [23] E. Haapasalo, T. Heinosaari, and J.P. Pellonpää. Quantum measurements on finite dimensional systems: relabeling and mixing. *Quantum Inf. Process.*, 11:1751, 2012.

Appendix A: Anticipative qubit measurements

We fix the set of inputs X and the state ensemble \mathcal{E} as in Eqs. (16) and (17), respectively. Our task is to discriminate any given input $\mathbf{x} \in X$, so we let the output and input sets coincide, $Y = X$, and we choose $f(\mathbf{x}, \mathbf{y}) = \delta_{\mathbf{x}, \mathbf{y}}$ following the notations of Section II B (we use the bold symbols \mathbf{x}, \mathbf{y} in place of x, y to stress that we deal with vectors in \mathbb{R}^3). We consider the parallel quantum-classical computation in which CPOST consists in excluding k wrong answers with either $k = 1$ or $k = 2$. The set S of the excluded answers is thus an element of the collection of sets T defined in Eq. (10). More precisely, for any fixed value of the input \mathbf{x} the set S is drawn from the subcollection

$$T_{\mathbf{x}} = \{S \subset Y : |S| = k \text{ and } \mathbf{x} \notin S\}. \quad (\text{A1})$$

We further assume that S is randomly picked with uniform probability within $T_{\mathbf{x}}$, which means that the partial information map α is

$$\alpha(S | \mathbf{x}) = \frac{1}{3} 1_{X \setminus S}(\mathbf{x}) \quad \forall S \in T. \quad (\text{A2})$$

In the particular case with $k = 1$, we can canonically identify $T \simeq X$ and $T_{\mathbf{x}} \simeq X \setminus \{\mathbf{x}\}$, so that the partial information map is $\alpha(\mathbf{t} | \mathbf{x}) = (1 - \delta_{\mathbf{x}, \mathbf{t}})/(m - 1)$ as described in Section II C.

If we perform a measurement \mathbb{M} with the outcome set Z and we post-process its outcome by means of a post-processing map ν , we guess a correct answer with the probability $\mathbb{P}_{\text{CPOST}}$ of Eq. (3). As we did in Eqs. (6) and (9), we rewrite such probability as $\mathbb{P}_{\text{CPOST}}(\mathbb{M}, \nu)$ in order to stress its dependence on \mathbb{M} and ν . Further, following Eqs. (6), we denote by $\mathbb{P}_{\text{CPOST}}^{\text{an}}$ the value of $\mathbb{P}_{\text{CPOST}}(\mathbb{M}, \nu)$ optimized over all measurements \mathbb{M} and all post-processing maps ν . The aim of this appendix is then to determine the maximum $\mathbb{P}_{\text{CPOST}}^{\text{an}}$, find the measurements \mathbb{M}^{an} at which it is attained and describe the corresponding optimal post-processings. The essential tool is the following relation proved in [16, Prop. 2]:

$$\mathbb{P}_{\text{CPOST}}^{\text{an}} = \max_{\mathbb{M}} \mathbb{P}_{\text{CPOST}}(\bar{\mathbb{M}}, \pi). \quad (\text{A3})$$

In this expression, the maximum on the right-hand side ranges over all measurements $\bar{\mathbb{M}}$ with the outcome set X^T , where X^T is the set of all functions $\phi : T \rightarrow X$. Moreover, π denotes the post-processing map

$$\pi_S(\mathbf{x} | \phi) = \delta_{\mathbf{x}, \phi(S)}. \quad (\text{A4})$$

Note that the outcome set X^T of $\bar{\mathbb{M}}$ and the post-processing map π are fixed in the right-hand side of Eq. (A3). This compares with the definition of $\mathbb{P}_{\text{CPOST}}^{\text{an}}$, which in principle would require to optimize $\mathbb{P}_{\text{CPOST}}$ over both \mathbb{M} and ν , also allowing the outcome set of \mathbb{M} to vary among all finite sets.

Theorem 1. For $j \in \{+, -\}$, let $\phi_j^{\mathbf{a}, \mathbf{b}}$ be element of X^T defined as follows:

- if $k = 1$,

$$\phi_j^{\mathbf{a}, \mathbf{b}}(S) = \begin{cases} j\mathbf{a} & \text{if } j\bar{\mathbf{a}} \notin S \\ j\mathbf{b} & \text{if } j\bar{\mathbf{a}} \in S \end{cases}, \quad (\text{A5a})$$

- if $k = 2$,

$$\phi_j^{\mathbf{a}, \mathbf{b}}(S) = \begin{cases} j\mathbf{a} & \text{if } j\bar{\mathbf{a}} \notin S \\ j\mathbf{b} & \text{if } j\bar{\mathbf{a}} \in S \text{ and } j\bar{\mathbf{b}} \notin S \\ -j\mathbf{b} & \text{if } j\bar{\mathbf{a}} \in S \text{ and } j\bar{\mathbf{b}} \in S \end{cases}. \quad (\text{A5b})$$

Then, a measurement attaining the maximum in Eq. (A3) is the measurement $\bar{\mathbb{M}} = \bar{\mathbb{M}}^{\mathbf{a}, \mathbf{b}}$ given by

$$\begin{aligned} \bar{\mathbb{M}}^{\mathbf{a}, \mathbf{b}}(\phi) &= 0 \quad \text{for } \phi \notin \{\phi_+^{\mathbf{a}, \mathbf{b}}, \phi_-^{\mathbf{a}, \mathbf{b}}\}, \\ \bar{\mathbb{M}}^{\mathbf{a}, \mathbf{b}}(\phi_j^{\mathbf{a}, \mathbf{b}}) &= \frac{1}{2} \left(\mathbb{1} + j \frac{3\mathbf{a} + \mathbf{b}}{\|3\mathbf{a} + \mathbf{b}\|} \cdot \boldsymbol{\sigma} \right). \end{aligned} \quad (\text{A6})$$

Moreover, we have

$$\mathbb{P}_{\text{CPOST}}^{\text{an}} = \frac{1}{12} \left(2 + 2r + \sqrt{10 + 6\mathbf{a} \cdot \mathbf{b}} \right).$$

By exchanging the vectors \mathbf{a} and \mathbf{b} in Eqs. (A5) and (A6), we also obtain the optimal measurement

$$\begin{aligned} \bar{\mathbb{M}}^{\mathbf{b}, \mathbf{a}}(\phi) &= 0 \quad \text{for } \phi \notin \{\phi_+^{\mathbf{b}, \mathbf{a}}, \phi_-^{\mathbf{b}, \mathbf{a}}\}, \\ \bar{\mathbb{M}}^{\mathbf{b}, \mathbf{a}}(\phi_j^{\mathbf{b}, \mathbf{a}}) &= \frac{1}{2} \left(\mathbb{1} + j \frac{\mathbf{a} + 3\mathbf{b}}{\|\mathbf{a} + 3\mathbf{b}\|} \cdot \boldsymbol{\sigma} \right), \end{aligned} \quad (\text{A7})$$

and any convex combination of $\bar{\mathbb{M}}^{\mathbf{a}, \mathbf{b}}$ and $\bar{\mathbb{M}}^{\mathbf{b}, \mathbf{a}}$ is still a measurement attaining the maximum in Eq. (A3). As an example, the optimal measurement for the case with $k = 1$ that was derived in [16, Equations (65) and (66)] coincides with the sum $(1/2)\bar{\mathbb{M}}^{\mathbf{a}, \mathbf{b}} + (1/2)\bar{\mathbb{M}}^{\mathbf{b}, \mathbf{a}}$. In the particular case with $\mathbf{a} \cdot \mathbf{b} = 0$, Theorem 1 yields two additional optimal measurements besides $\bar{\mathbb{M}}^{\mathbf{a}, \mathbf{b}}$ and $\bar{\mathbb{M}}^{\mathbf{b}, \mathbf{a}}$, namely the measurements $\bar{\mathbb{M}}^{-\mathbf{a}, \mathbf{b}}$ and $\bar{\mathbb{M}}^{\mathbf{b}, -\mathbf{a}}$ that are obtained by replacing \mathbf{a} with $-\mathbf{a}$ in (A6) and (A7), respectively. The supports of the four measurements $\bar{\mathbb{M}}^{\mathbf{a}, \mathbf{b}}$, $\bar{\mathbb{M}}^{\mathbf{b}, \mathbf{a}}$, $\bar{\mathbb{M}}^{-\mathbf{a}, \mathbf{b}}$ and $\bar{\mathbb{M}}^{\mathbf{b}, -\mathbf{a}}$ are mutually disjoint subsets of X^T , and each of them contains two elements. By restricting the convex sum $(1/2)\bar{\mathbb{M}}^{\mathbf{a}, \mathbf{b}} + (1/2)\bar{\mathbb{M}}^{\mathbf{b}, \mathbf{a}}$ to its support $\{\phi_+^{\mathbf{a}, \mathbf{b}}, \phi_-^{\mathbf{a}, \mathbf{b}}, \phi_+^{\mathbf{b}, \mathbf{a}}, \phi_-^{\mathbf{b}, \mathbf{a}}\}$, evaluating the corresponding restriction of the post-processing map π and relabeling $\phi_j^{\mathbf{a}, \mathbf{b}} \rightarrow j\mathbf{n}$ and $\phi_j^{\mathbf{b}, \mathbf{a}} \rightarrow j\mathbf{m}$, we reduce the sum $(1/2)\bar{\mathbb{M}}^{\mathbf{a}, \mathbf{b}} + (1/2)\bar{\mathbb{M}}^{\mathbf{b}, \mathbf{a}}$ to the anticipative measurement \mathbb{M}^{an} of Eq. (22), while π becomes the post-processing map ν with

- if $k = 1$,

$$\begin{aligned} \nu_S(\mathbf{x} | j\mathbf{n}) &= \delta_{\mathbf{x}, j\mathbf{a}} 1_{X \setminus S}(j\mathbf{a}) + \delta_{\mathbf{x}, j\mathbf{b}} 1_S(j\mathbf{a}), \\ \nu_S(\mathbf{x} | j\mathbf{m}) &= \delta_{\mathbf{x}, j\mathbf{b}} 1_{X \setminus S}(j\mathbf{b}) + \delta_{\mathbf{x}, j\mathbf{a}} 1_S(j\mathbf{b}), \end{aligned}$$

- if $k = 2$,

$$\begin{aligned} \nu_S(\mathbf{x} | j\mathbf{n}) &= \delta_{\mathbf{x}, j\mathbf{a}} 1_{X \setminus S}(j\mathbf{a}) + \delta_{\mathbf{x}, j\mathbf{b}} 1_S(j\mathbf{a}) 1_{X \setminus S}(j\mathbf{b}) \\ &\quad + \delta_{\mathbf{x}, -j\mathbf{b}} 1_S(j\mathbf{a}) 1_S(j\mathbf{b}), \\ \nu_S(\mathbf{x} | j\mathbf{m}) &= \delta_{\mathbf{x}, j\mathbf{b}} 1_{X \setminus S}(j\mathbf{b}) + \delta_{\mathbf{x}, j\mathbf{a}} 1_{X \setminus S}(j\mathbf{a}) 1_S(j\mathbf{b}) \\ &\quad + \delta_{\mathbf{x}, -j\mathbf{a}} 1_S(j\mathbf{a}) 1_S(j\mathbf{b}). \end{aligned}$$

We observe that ν coincides with the relabeling priority described in the second table of Section IIIB for both cases with $k = 1$ and $k = 2$.

Proof of Theorem 1. By [16, Eq. (25)], we have

$$\mathbb{P}(\bar{\mathbf{M}}, \pi) = C \mathbb{P}_{\bar{\mathcal{E}}}(\bar{\mathbf{M}}),$$

where $\bar{\mathcal{E}} : X^T \rightarrow \mathcal{L}(\mathbb{C}^2)$ is an auxiliary state ensemble associated with \mathcal{E} and defined as

$$\bar{\mathcal{E}}(\phi) = \frac{1}{24C} \sum_{\mathbf{x}} |\phi^{-1}(\mathbf{x}) \cap T_{\mathbf{x}}| (\mathbf{1} + \mathbf{x} \cdot \boldsymbol{\sigma}),$$

the scalar $C > 0$ is a suitable normalization constant, and

$$\mathbb{P}_{\bar{\mathcal{E}}}(\bar{\mathbf{M}}) = \sum_{\phi} \text{tr} [\bar{\mathcal{E}}(\phi) \bar{\mathbf{M}}(\phi)]$$

is the success probability of the usual state discrimination task for the auxiliary state ensemble $\bar{\mathcal{E}}$ and the measurement $\bar{\mathbf{M}}$. In particular, the measurements $\bar{\mathbf{M}}$ which maximize the probabilities $\mathbb{P}_{\text{cpost}}(\bar{\mathbf{M}}, \pi)$ and $\mathbb{P}_{\bar{\mathcal{E}}}(\bar{\mathbf{M}})$ coincide. Thus, optimizing the state discrimination task with posterior information for the original ensemble \mathcal{E} reduces to a standard state discrimination problem for the auxiliary ensemble $\bar{\mathcal{E}}$. To solve the latter problem, we denote by $\Lambda(\bar{\mathcal{E}})$ the largest eigenvalue of all the operators $\bar{\mathcal{E}}(\phi)$, $\phi \in X^T$, and we observe that any measurement $\bar{\mathbf{M}}$ satisfies the inequality

$$\mathbb{P}_{\bar{\mathcal{E}}}(\bar{\mathbf{M}}) \leq 2 \Lambda(\bar{\mathcal{E}}).$$

As a consequence of [16, Proposition 1], the above bound is attained if and only if $\bar{\mathbf{M}}$ satisfies the relation

$$\bar{\mathcal{E}}(\phi) \bar{\mathbf{M}}(\phi) = \Lambda(\bar{\mathcal{E}}) \bar{\mathbf{M}}(\phi)$$

for all $\phi \in X^T$. We will show that this is actually the case for the measurement $\bar{\mathbf{M}}^{\mathbf{a}, \mathbf{b}}$ of (A6), and therefore

$$\mathbb{P}_{\text{cpost}}(\bar{\mathbf{M}}^{\mathbf{a}, \mathbf{b}}, \pi) = \max_{\bar{\mathbf{M}}} \mathbb{P}_{\text{cpost}}(\bar{\mathbf{M}}, \pi) = 2C \Lambda(\bar{\mathcal{E}})$$

for the value of $\Lambda(\bar{\mathcal{E}})$ that we will explicitly determine below.

For notational convenience, for all $\phi \in X^T$ and $j \in \{+, -\}$, we introduce the nonnegative integer numbers

$$\alpha_j^\phi = |\phi^{-1}(j\mathbf{a}) \cap T_{j\mathbf{a}}|, \quad \beta_j^\phi = |\phi^{-1}(j\mathbf{b}) \cap T_{j\mathbf{b}}|,$$

so that the auxiliary state ensemble rewrites

$$\bar{\mathcal{E}}(\phi) = \frac{1}{24C} \left\{ (\alpha_+^\phi + \alpha_-^\phi + \beta_+^\phi + \beta_-^\phi) \mathbf{1} + [(\alpha_+^\phi - \alpha_-^\phi) \mathbf{a} + (\beta_+^\phi - \beta_-^\phi) \mathbf{b}] \cdot \boldsymbol{\sigma} \right\}.$$

The largest eigenvalue of the operator $\bar{\mathcal{E}}(\phi)$ is

$$\begin{aligned} \lambda(\phi) &= \frac{1}{24C} \left\{ \alpha_+^\phi + \alpha_-^\phi + \beta_+^\phi + \beta_-^\phi \right. \\ &\quad \left. + \left\| (\alpha_+^\phi - \alpha_-^\phi) \mathbf{a} + (\beta_+^\phi - \beta_-^\phi) \mathbf{b} \right\| \right\} \\ &= \frac{1}{24C} \gamma(\alpha_+^\phi, \alpha_-^\phi, \beta_+^\phi, \beta_-^\phi), \end{aligned}$$

where γ is the function

$$\begin{aligned} \gamma(\alpha_+, \alpha_-, \beta_+, \beta_-) &= \alpha_+ + \alpha_- + \beta_+ + \beta_- \\ &\quad + [(\alpha_+ - \alpha_-)^2 + (\beta_+ - \beta_-)^2 \\ &\quad + 2(\alpha_+ - \alpha_-)(\beta_+ - \beta_-) \mathbf{a} \cdot \mathbf{b}]^{\frac{1}{2}}. \end{aligned}$$

If $\alpha_+^\phi \neq \alpha_-^\phi$ or $\beta_+^\phi \neq \beta_-^\phi$, the projection onto the $\lambda(\phi)$ -eigenspace of $\bar{\mathcal{E}}(\phi)$ is the rank-1 operator

$$\Pi(\phi) = \frac{1}{2} \left(\mathbf{1} + \frac{(\alpha_+^\phi - \alpha_-^\phi) \mathbf{a} + (\beta_+^\phi - \beta_-^\phi) \mathbf{b}}{\left\| (\alpha_+^\phi - \alpha_-^\phi) \mathbf{a} + (\beta_+^\phi - \beta_-^\phi) \mathbf{b} \right\|} \cdot \boldsymbol{\sigma} \right).$$

We now proceed to separately evaluate

$$\Lambda(\bar{\mathcal{E}}) = \max_{\phi} \lambda(\phi) = \frac{1}{24C} \max_{\phi} \gamma(\alpha_+^\phi, \alpha_-^\phi, \beta_+^\phi, \beta_-^\phi)$$

in the two cases with $k = 1$ and $k = 2$.

- *Case $k = 1$.* Since $|T_{\mathbf{x}}| = 3$ for all $\mathbf{x} \in X$ and $\sum_{\mathbf{x}} |\phi^{-1}(\mathbf{x})| = |T| = 4$, the numbers $\alpha_+^\phi, \alpha_-^\phi, \beta_+^\phi$ and β_-^ϕ satisfy the constraints

$$\begin{aligned} \alpha_+^\phi, \alpha_-^\phi, \beta_+^\phi, \beta_-^\phi &\in \{0, 1, 2, 3\}, \\ \alpha_+^\phi + \alpha_-^\phi + \beta_+^\phi + \beta_-^\phi &\leq 4. \end{aligned}$$

For $\mathbf{a} \cdot \mathbf{b} > 0$, the constrained maximum

$$\begin{aligned} \max \gamma(\alpha_+, \alpha_-, \beta_+, \beta_-) \quad \text{subject to} \\ \alpha_+, \alpha_-, \beta_+, \beta_- \in \{0, 1, 2, 3\}, \quad (*) \\ \alpha_+ + \alpha_- + \beta_+ + \beta_- \leq 4 \end{aligned}$$

was evaluated in [16, Appendix A] and found to be equal to $4 + \sqrt{10 + 6 \mathbf{a} \cdot \mathbf{b}}$. By an easy continuity argument, this result extends also to the case with $\mathbf{a} \cdot \mathbf{b} = 0$. We have

$$(\alpha_+^\phi, \alpha_-^\phi, \beta_+^\phi, \beta_-^\phi) = \begin{cases} (3, 0, 1, 0) & \text{if } \phi = \phi_+^{\mathbf{a}, \mathbf{b}} \\ (0, 3, 0, 1) & \text{if } \phi = \phi_-^{\mathbf{a}, \mathbf{b}} \end{cases},$$

which are two feasible points attaining the maximum (*). Therefore,

$$\Lambda(\bar{\mathcal{E}}) = \frac{1}{24C} \left(4 + \sqrt{10 + 6 \mathbf{a} \cdot \mathbf{b}} \right).$$

- *Case $k = 2$.* We still have $|T_{\mathbf{x}}| = 3$ for all $\mathbf{x} \in X$, but now $\sum_{\mathbf{x}} |\phi^{-1}(\mathbf{x})| = |T| = 6$. Moreover, if $\mathbf{x} \neq \mathbf{y}$, we have $\{\mathbf{x}, \mathbf{y}\} \notin T_{\mathbf{x}} \cup T_{\mathbf{y}}$ and $\phi^{-1}(\mathbf{x}) \cap \phi^{-1}(\mathbf{y}) = \emptyset$, hence it must be $|\phi^{-1}(\mathbf{x}) \cap T_{\mathbf{x}}| + |\phi^{-1}(\mathbf{y}) \cap T_{\mathbf{y}}| \leq |T| - 1 = 5$. It follows that the numbers $\alpha_+^\phi, \alpha_-^\phi, \beta_+^\phi$ and β_-^ϕ now satisfy the constraints

$$\begin{aligned} \alpha_+^\phi, \alpha_-^\phi, \beta_+^\phi, \beta_-^\phi &\in \{0, 1, 2, 3\}, \\ \alpha_+^\phi + \alpha_-^\phi + \beta_+^\phi + \beta_-^\phi &\leq 6, \\ \alpha_+^\phi + \beta_+^\phi &\leq 5, \\ \alpha_-^\phi + \beta_-^\phi &\leq 5. \end{aligned}$$

As in the previous case, we first evaluate the constrained maximum

$$\begin{aligned} \max \gamma(\alpha_+, \alpha_-, \beta_+, \beta_-) \quad \text{subject to} \\ \alpha_+, \alpha_-, \beta_+, \beta_- \in \{0, 1, 2, 3\}, \\ \alpha_+ + \alpha_- + \beta_+ + \beta_- \leq 6, \quad (**) \\ \alpha_+ + \beta_+ \leq 5, \\ \alpha_- + \beta_- \leq 5, \end{aligned}$$

and next we prove that $(\alpha_+^\phi, \alpha_-^\phi, \beta_+^\phi, \beta_-^\phi)$ is an optimal point for all $\phi \in \{\phi_+^{\mathbf{a}, \mathbf{b}}, \phi_-^{\mathbf{a}, \mathbf{b}}\}$. To this aim, we fix a feasible point $(\alpha_+, \alpha_-, \beta_+, \beta_-)$ and we start by assuming that $\alpha_j - \alpha_h = 3$ for some $j, h \in \{+, -\}$. It follows that $\alpha_j = 3, \alpha_h = 0$ and

$$\begin{aligned} \gamma(\alpha_+, \alpha_-, \beta_+, \beta_-) &= 3 + \beta_+ + \beta_- \\ &+ [9 + (\beta_+ - \beta_-)^2 + 6(\beta_j - \beta_h) \mathbf{a} \cdot \mathbf{b}]^{\frac{1}{2}} \end{aligned}$$

because $\alpha_+, \alpha_- \in \{0, 1, 2, 3\}$. If $(\alpha_+, \alpha_-, \beta_+, \beta_-)$ is a feasible point with $\beta_j \leq \beta_k$, then $(\alpha'_+, \alpha'_-, \beta'_+, \beta'_-) = (\alpha_+, \alpha_-, \beta_-, \beta_+)$ is a feasible point such that $\gamma(\alpha'_+, \alpha'_-, \beta'_+, \beta'_-) \geq \gamma(\alpha_+, \alpha_-, \beta_+, \beta_-)$ and $\beta'_j \geq \beta'_k$. Therefore, in order to find the constrained maximum (**), we can restrict to the feasible points which satisfy $\beta_j \geq \beta_k$. For any such point, the relation $\alpha_j + \beta_j \leq 5$ requires that $\beta_j \leq 2$. The possibility $\beta_+ = \beta_- = 2$ is excluded by the inequality $\alpha_+ + \alpha_- + \beta_+ + \beta_- \leq 6$. Therefore, the only remaining possibilities are

- $(\beta_j, \beta_k) = (2, 1)$, and then

$$\gamma(\alpha_+, \alpha_-, \beta_+, \beta_-) = 6 + \sqrt{10 + 6 \mathbf{a} \cdot \mathbf{b}};$$

- $(\beta_j, \beta_k) = (2, 0)$, and then

$$\begin{aligned} \gamma(\alpha_+, \alpha_-, \beta_+, \beta_-) &= 5 + \sqrt{13 + 12 \mathbf{a} \cdot \mathbf{b}} \\ &\leq 6 + \sqrt{10 + 6 \mathbf{a} \cdot \mathbf{b}}; \end{aligned}$$

- $(\beta_j, \beta_k) = (1, 0)$, and then

$$\begin{aligned} \gamma(\alpha_+, \alpha_-, \beta_+, \beta_-) &= 4 + \sqrt{10 + 6 \mathbf{a} \cdot \mathbf{b}} \\ &< 6 + \sqrt{10 + 6 \mathbf{a} \cdot \mathbf{b}}; \end{aligned}$$

- $(\beta_j, \beta_k) = (1, 1)$ or $(\beta_j, \beta_k) = (0, 0)$, and then

$$\begin{aligned} \gamma(\alpha_+, \alpha_-, \beta_+, \beta_-) &= 6 + \beta_+ + \beta_- \leq 8 \\ &< 6 + \sqrt{10 + 6 \mathbf{a} \cdot \mathbf{b}}. \end{aligned}$$

Next, we consider the case with $\beta_j - \beta_h = 3$, and we observe that it is similar to the previous case with $\alpha_j - \alpha_h = 3$. In particular,

$$\gamma(\alpha_+, \alpha_-, \beta_+, \beta_-) \leq 6 + \sqrt{10 + 6 \mathbf{a} \cdot \mathbf{b}};$$

also in this case. Finally, since the constraint $\alpha_+, \alpha_-, \beta_+, \beta_- \in \{0, 1, 2, 3\}$ requires that $\max\{|\alpha_+ -$

$\alpha_-|, |\beta_+ - \beta_-|\} \leq 3$, the only remaining possibility is $\max\{|\alpha_+ - \alpha_-|, |\beta_+ - \beta_-|\} \leq 2$, which implies

$$\begin{aligned} \gamma(\alpha_+, \alpha_-, \beta_+, \beta_-) &\leq 6 + \sqrt{8 + 8 \mathbf{a} \cdot \mathbf{b}} \\ &\leq 6 + \sqrt{10 + 6 \mathbf{a} \cdot \mathbf{b}}. \end{aligned}$$

In summary, the constrained maximum (**) is $6 + \sqrt{10 + 6 \mathbf{a} \cdot \mathbf{b}}$, and it is attained e.g. at the feasible points

$$(\alpha_+^\phi, \alpha_-^\phi, \beta_+^\phi, \beta_-^\phi) = \begin{cases} (3, 0, 2, 1) & \text{if } \phi = \phi_+^{\mathbf{a}, \mathbf{b}} \\ (0, 3, 1, 2) & \text{if } \phi = \phi_-^{\mathbf{a}, \mathbf{b}} \end{cases}.$$

Similarly to the case with $k = 1$, it then follows that

$$\Lambda(\bar{\mathcal{E}}) = \frac{1}{24C} \left(6 + \sqrt{10 + 6 \mathbf{a} \cdot \mathbf{b}} \right).$$

In both cases with $k = 1$ and $k = 2$, the measurement (A6) satisfies the equality

$$\bar{\mathbf{M}}^{\mathbf{a}, \mathbf{b}}(\phi) = \begin{cases} \Pi(\phi) & \text{if } \phi \in \{\phi_+^{\mathbf{a}, \mathbf{b}}, \phi_-^{\mathbf{a}, \mathbf{b}}\} \\ 0 & \text{otherwise} \end{cases},$$

hence $\bar{\mathcal{E}}(\phi) \bar{\mathbf{M}}^{\mathbf{a}, \mathbf{b}}(\phi) = \Lambda(\bar{\mathcal{E}}) \bar{\mathbf{M}}^{\mathbf{a}, \mathbf{b}}(\phi)$ for all $\phi \in X^T$. The relation (A3) and the optimality of $\bar{\mathbf{M}}^{\mathbf{a}, \mathbf{b}}$ then follow from the previous discussion. \square

Appendix B: IBMQ Manila specifications

Our demonstrations were performed on *ibmq-manila*, which is one of the 5-qubit IBM Quantum Falcon Processors, r5.11L (linearly coupled qubits), at that time having backend version 1.0.17 (dated 9. November 2021). Even though we performed computations on qubit 0 only, we present data for all qubits in the table, where we list qubit frequencies f , T_1 and T_2 times, single qubit errors ϵ_1 and z -measurement errors ϵ_M .

The native gate set is CX (controlled X), ID (identity), IF ELSE (classical dynamical conditioning), RZ (parametric z -rotation), SX (square root of X), X (bit flip / NOT operation in the computational z -basis).

For completeness we describe also some of the operations. The gate X can be represented by the Pauli σ_x matrix and as a result SX performs a unitary rotation described by the matrix

$$\sqrt{\sigma_x} = \frac{1}{2} \begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix}.$$

The gate RZ performs a parametric unitary rotation given by the matrix

$$R_z(\theta) = \begin{pmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{+i\frac{\theta}{2}} \end{pmatrix}.$$

Finally, we have used the gate RY which performs the parametric rotation

$$R_y(\theta) = \begin{pmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix}.$$

Since this gate does not belong to the native gate set of *ibmq_manila*, it needs to be further decomposed. One can use the identity

$$R_y(\theta) = i\sqrt{\sigma_x}R_z(\pi - \theta)\sqrt{\sigma_x}R_z(\pi).$$

As we used automatic regime of Qiskit transpiler, the

actual decomposition might vary and is possibly further optimized by the transpiler together with the rest of the optimized circuit.

qubit	f [GHz]	T_1 [μ s]	T_2 [μ s]	ϵ_1	ϵ_M
0	4.96(3)	208	116	2.06×10^{-4}	2.30×10^{-2}
1	4.83(8)	227	86	2.46×10^{-4}	2.80×10^{-2}
2	5.03(7)	179	25	2.41×10^{-4}	2.27×10^{-2}
3	4.95(1)	134	62	1.69×10^{-4}	2.14×10^{-2}
4	5.06(6)	147	41	3.13×10^{-4}	2.29×10^{-2}