On the Role of Coherence in Shor's Algorithm
Supplemental Material

In this Supplemental Material, we give the proofs of the results presented in the main text and some further information. This includes additional dynamic resource measures and their properties.

I. On the appearing measures

In this section, we present properties of the resource measures employed in our analysis of the performance of Shor's algorithm. To begin with, we discuss the functional

$$\mathscr{D}(\Lambda) = \max_{\rho} \left\| \Delta\Lambda(\mathbb{1} - \Delta)\rho \right\|_1 , \tag{1}$$

which is interesting from a resource theoretical perspective and we will see later that $\mathscr{D}(\Lambda)$ appears naturally when connecting the success probability of the investigated order-finding protocol with the ability to detect coherence. Moreover, it seems to be a natural candidate for a resource measure under detection incoherent operations. However, it is not monotonic under $\mathcal{DIS}$ as we show here. To relate the performance of Shor's algorithm with a rigorous dynamical measure in Thm. 14 and Thm. 16 we make use of the fact that the functional $\mathscr{D}$ shares sufficient similarities with the NSID measure $\tilde{M}_\diamond(\Lambda) = \min_{\Phi\in\mathcal{DI}} \max_\rho \left\| \Delta(\Lambda - \Phi)\rho \right\|_1$. In particular, that $\mathscr{D}$ provides an upper bound on $\tilde{M}_\diamond$ and that the two functionals coincide on qubit channels.

Proposition 1. *Let $\Phi \in \mathcal{DI}$. The functional $\mathscr{D}(\Lambda) = \max_\rho \left\| \Delta\Lambda(\mathbb{1} - \Delta)\rho \right\|_1$ has the following properties:*

1. *$\mathscr{D}(\Lambda) = 0 \Leftrightarrow \Lambda \in \mathcal{DI}$,*
2. *$\mathscr{D}(\Lambda^{C\leftarrow A} \otimes \mathbb{1}_B) = \mathscr{D}(\Lambda^{C\leftarrow A})$,*
3. *$\mathscr{D}(\Phi\Lambda) \leq \mathscr{D}(\Lambda)$,*
4. *$\mathscr{D}$ is convex.*

Proof. Let us begin by pointing out that convexity in the argument follows from the convexity of the trace norm itself. Notice that for any $\Lambda \in \mathcal{DI}$, i.e., $\Delta\Lambda = \Delta\Lambda\Delta$, it follows that the functional vanishes. Furthermore, for any detecting channel Λ there exists some state ρ such that $\Delta\Lambda(\mathbb{1} - \Delta)\rho \neq 0$, which proves faithfulness since $\|\cdot\|_1$ is a norm.
The functional behaves monotonically under post-processing with free channels even with an identity channel attached in parallel. Using [**?** , Lem. 14] for the inequality, we see that

$$\begin{aligned}
\mathscr{D}(\Lambda^{C\leftarrow A} \otimes \mathbb{1}_B) &= \max_{\rho_{AB}} \left\| \Delta_{CB}(\Lambda^{C\leftarrow A} \otimes \mathbb{1}_B)(\mathbb{1}_{AB} - \Delta_{AB})\rho_{AB} \right\|_1 \\
&= \max_{\rho_{AB}} \left\| \Delta_{CB}(\Lambda^{C\leftarrow A} \otimes \mathbb{1}_B)(\mathbb{1}_A \otimes \mathbb{1}_B - \Delta_A \otimes \Delta_B)\rho_{AB} \right\|_1 \\
&= \max_{\rho_{AB}} \left\| \left[(\Delta_C\Lambda^{C\leftarrow A} - \Delta_C\Lambda^{C\leftarrow A}\Delta_A) \otimes \mathbb{1}_B \right] (\mathbb{1}_A \otimes \Delta_B)\rho_{AB} \right\|_1 \\
&\leq \max_{|\psi\rangle,|i\rangle} \left\| (\Delta_C\Lambda^{C\leftarrow A} - \Delta_C\Lambda^{C\leftarrow A}\Delta_A)|\psi\rangle\langle\psi|_A \otimes |i\rangle\langle i| \right\|_1 \\
&= \max_{|\psi\rangle,|i\rangle} \left\| (\Delta_C\Lambda^{C\leftarrow A} - \Delta_C\Lambda^{C\leftarrow A}\Delta_A)|\psi\rangle\langle\psi|_A \right\|_1 \left\| |i\rangle\langle i| \right\|_1 \\
&= \max_{|\psi\rangle} \left\| (\Delta_C\Lambda^{C\leftarrow A} - \Delta_C\Lambda^{C\leftarrow A}\Delta_A)|\psi\rangle\langle\psi|_A \right\|_1 ,
\end{aligned} \tag{2}$$

which coincides with $\mathscr{D}(\Lambda^{C\leftarrow A})$ due to the convexity of the trace norm. The reverse inequality follows from restricting ρ_{AB} to product states in the first line.

Additionally, the properties of the trace norm allow us to write

$$\begin{aligned}
\mathscr{D}(\Phi\Lambda) &= \max_{\rho} \left\| \Delta\Phi\Lambda(\mathbb{1} - \Delta)\rho \right\|_1 \\
&= \max_{\rho} \left\| \Delta\Phi\Delta\Lambda(\mathbb{1} - \Delta)\rho \right\|_1 \\
&\leq \max_{\rho} \left\| \Delta\Lambda(\mathbb{1} - \Delta)\rho \right\|_1 .
\end{aligned} \tag{3}$$

□

To show that $\mathscr{D}$ is a resource measure, we would need in addition that $\mathscr{D}(\Lambda\Phi) \leq \mathscr{D}(\Lambda)$. However, this condition is violated in general for $\Phi \in \mathcal{DI}$. To prove this, we first describe how we can evaluate $\mathscr{D}$ numerically.

Proposition 2. *Consider a quantum channel $\Theta^{C\leftarrow B}$ and let $N = \dim(C)$. Let further $(s_{m,n})_{m,n}$ be the matrix of dimension $2^N \times N$ that contains as rows all N-dimensional vectors $\vec{s}_m$ whose entries are ± 1. The numerical value of $\mathscr{D}(\Theta^{C\leftarrow B})$ is then equivalent to the maximum of the solutions of the following 2^N semidefinite programs (each for a fixed m)*

$$\begin{aligned} &\textit{maximize:} \quad t_m = \sum_{n=0}^{N-1} s_{m,n} \left\langle n\right|_C \Theta^{C\leftarrow B} \left(\mathbb{1} - \Delta\right) \sigma \left|n\right\rangle_C , \\ &\textit{subject to:} \quad \sigma_B \geq 0, \\ &\qquad\qquad\ \ \mathrm{Tr}\left[\sigma_B\right] = 1. \end{aligned} \tag{4}$$

Proof. We use that for real f_n

$$\sum_n |f_n| = \max_{\vec{s}_m} (\vec{s}_m \cdot \vec{f}), \tag{5}$$

where the vectors $\vec{s}_m$ have been introduced in the statement of the Proposition. In addition

$$\mathscr{D}(\Theta^{C\leftarrow B}) = \max_\sigma \left\| \Delta\Theta^{C\leftarrow B} \left(\mathbb{1} - \Delta\right) \sigma \right\|_1 = \max_\sigma \sum_n \left| \left\langle n\right| \Theta^{C\leftarrow B} \left(\mathbb{1} - \Delta\right) \sigma \left|n\right\rangle \right|. \tag{6}$$

□

This method of evaluating $\mathscr{D}$ is certainly not the most efficient. However, with the help of the following Proposition, it allows us to disprove monotonicity.

Proposition 3. *Let $\Theta^{C\leftarrow B}$ be a quantum channel and A a third and fixed quantum system. Denote by $\mathcal{S}$ the set of all diagonal matrices of dimension $\dim(C)$ with diagonal elements ± 1, and by M_A the matrix on system A with all entries equal to one. Furthermore, we define*

$$\mathcal{X} = \{X = (M_A - \mathbb{1}_A) \otimes \Theta^\dagger S : S \in \mathcal{S}\}. \tag{7}$$

The solution of the maximization problem

$$\max_{\sigma, \Phi^{B\leftarrow A} \in \mathcal{DI}} \left\| \Delta\Theta^{C\leftarrow B} \Phi^{B\leftarrow A} \left(\mathbb{1} - \Delta\right) \sigma \right\|_1 \tag{8}$$

is then given by the maximum of the solutions of the following (finite number of) semidefinite programs to be evaluated for all fixed $X \in \mathcal{X}$

$$\begin{aligned} &\textit{maximize:} \quad \mathrm{Tr}\left[X Y_{AB}\right], \\ &\textit{subject to:} \quad Y \geq 0, \\ &\qquad\qquad\ \ \mathrm{Tr}_B\left[Y\right] = \Delta\sigma_A, \\ &\qquad\qquad\ \ \sigma_A \geq 0, \\ &\qquad\qquad\ \ \mathrm{Tr}\left[\sigma_A\right] = 1, \\ &\qquad\qquad\ \ \mathrm{diag}\left(\left\langle i\right|_A Y \left|j\right\rangle_A\right) = 0 \ \forall i \neq j. \end{aligned} \tag{9}$$

Proof. We write quantum states as $\rho = \sum_{i,j} \rho_{ij} |i\rangle\langle j|$ and the action of quantum channels as $\Phi(|i\rangle\langle j|) = \sum_{k,l} \Phi_{kl}^{ij} |k\rangle\langle l|$. With this notation, we find

$$\begin{aligned} \left\| \Delta\Theta\Phi \left(\mathbb{1} - \Delta\right) \rho \right\|_1 &= \left\| \Delta\Theta \left[\sum_{i\neq j} \sum_{kl} \Phi_{kl}^{ij} |k\rangle\langle l| \rho_{ij} \right] \right\|_1 \\ &= \left\| \Delta\Theta \left[\sum_{i\neq j} \left\langle i\right|_A \left(\sum_{op} \rho_{op} |o\rangle\langle p|_A \otimes \sum_{kl} \Phi_{kl}^{op} |k\rangle\langle l|_B \right) \left|j\right\rangle_A \right] \right\|_1 \\ &= \left\| \Delta\Theta \left[\sum_{i\neq j} \left\langle i\right|_A \left(\left(\mathbb{1}^A \otimes \Phi^{B\leftarrow \tilde{A}}\right) \sum_{op} \rho_{op} |oo\rangle\langle pp|_{A\tilde{A}} \right) \left|j\right\rangle_A \right] \right\|_1 . \end{aligned} \tag{10}$$

Using [**?** , Lem. 12] we thus have

$$\max_{\sigma,\Phi^{B\leftarrow A}\in\mathcal{DI}} \left\|\Delta\Theta^{C\leftarrow B}\Phi^{B\leftarrow A}\left(\mathbb{1}-\Delta\right)\sigma\right\|_1 = \max_{Y_{AB}\in\mathcal{Y}} \left\|\Delta\Theta^{C\leftarrow B}\left[\sum_{i\neq j}\langle i|_A Y_{AB}\left|j\right\rangle_A\right]\right\|_1, \tag{11}$$

where the set $\mathcal{Y}$ is defined as

$$\mathcal{Y} := \{Y_{AB} | Y\geq 0,\ \mathrm{Tr}_B(Y)=\Delta\sigma_A,\ \mathrm{diag}\left(\langle i|_A Y\left|j\right\rangle_A\right)=0\ \forall i\neq j,\ \sigma_A \text{ quantum state}\} \tag{12}$$

and therefore characterized by semidefinite constraints. Using the absolute value technique from Prop. 2, we can solve this optimization problem via a set of SDPs: With the definitions from the Proposition, we note that $\sum_{i\neq j}\langle i|_A Y_{AB}\left|j\right\rangle_A = \mathrm{Tr}_A\left[(M_A-\mathbb{1}_A)\otimes\mathbb{1}_B Y_{AB}\right]$ and therefore

$$\begin{aligned}
&\max_{Y_{AB}\in\mathcal{Y}}\left\|\Delta\Theta^{C\leftarrow B}\left[\sum_{i\neq j}\langle i|_A Y_{AB}\left|j\right\rangle_A\right]\right\|_1\\
&=\max_{S\in\mathcal{S}}\max_{Y_{AB}\in\mathcal{Y}}\mathrm{Tr}\left(S\Theta^{C\leftarrow B}\left[\sum_{i\neq j}\langle i|_A Y_{AB}\left|j\right\rangle_A\right]\right)\\
&=\max_{S\in\mathcal{S}}\max_{Y_{AB}\in\mathcal{Y}}\mathrm{Tr}\left(\left[\Theta^\dagger S\right]\left[\sum_{i\neq j}\langle i|_A Y_{AB}\left|j\right\rangle_A\right]\right)\\
&=\max_{S\in\mathcal{S}}\max_{Y_{AB}\in\mathcal{Y}}\mathrm{Tr}\left(\left[\Theta^\dagger S\right]\mathrm{Tr}_A\left[(M_A-\mathbb{1}_A)\otimes\mathbb{1}_B Y_{AB}\right]\right)\\
&=\max_{S\in\mathcal{S}}\max_{Y_{AB}\in\mathcal{Y}}\mathrm{Tr}\left(\left[\mathbb{1}_A\otimes\Theta^\dagger S\right](M_A-\mathbb{1}_A)\otimes\mathbb{1}_B Y_{AB}\right)\\
&=\max_{X\in\mathcal{X}}\max_{Y_{AB}\in\mathcal{Y}}\mathrm{Tr}\left(XY_{AB}\right).
\end{aligned} \tag{13}$$

□

Due to this Proposition, for every fixed system A, we can evaluate

$$\tilde{\mathscr{D}}_A(\Theta^{C\leftarrow B}) := \max_{\Phi^{B\leftarrow A}\in\mathcal{DI}}\mathscr{D}(\Theta^{C\leftarrow B}\Phi^{B\leftarrow A}) = \max_{\Phi^{B\leftarrow A}\in\mathcal{DI}}\max_{\sigma_A}\left\|\Delta\Theta^{C\leftarrow B}\Phi^{B\leftarrow A}(\mathbb{1}_A-\Delta_A)\sigma_A\right\|_1 \tag{14}$$

numerically by solving a collection of SDPs. We now move to the equivalence on qubit channels, which we will use to connect the performance of Shor's algorithm to the ability to detect coherence.

Lemma 4. *Let Λ be any qubit channel defined in the index representation as $\Lambda(|n\rangle\langle m|)=\sum_{kl}\Lambda_{kl}^{nm}|k\rangle\langle l|$.*
Then the functional $\mathscr{D}$

1. *is given by $\mathscr{D}(\Lambda)=2|\Lambda_{00}^{01}|$,*
2. *coincides with the NSID measure [**?**] $\tilde{M}_\diamond(\Lambda)=\min_{\Phi\in\mathcal{DI}}\max_\rho\|\Delta(\Lambda-\Phi)\rho\|_1$.*

Proof. Due to convexity of the trace norm, the optimization in the definition of $\mathscr{D}(\Lambda)$ over all states can be reduced to pure states $\rho=|\psi\rangle\langle\psi|$. We expand a pure qubit state as $|\psi\rangle=p_0|0\rangle+p_1 e^{i\phi}|1\rangle$ and write $\Lambda_{kl}^{nm}=|\Lambda_{kl}^{nm}|e^{i\lambda_{kl}^{nm}}$. According to [**?** , Lem. 6], $|\lambda_{00}^{01}|=-|\lambda_{00}^{10}|=:\lambda$ and $|\Lambda_{00}^{01}|=|\Lambda_{00}^{10}|$. For the first part, from $\Delta\Lambda(\mathbb{1}-\Delta)$ being diagonal it follows then straightforward that

$$\begin{aligned}
\mathscr{D}(\Lambda)&=\max_\rho\|\Delta\Lambda(\mathbb{1}-\Delta)\rho\|_1=\max_\rho\ 2\left|\langle 0|\Delta\Lambda(\mathbb{1}-\Delta)\rho|0\rangle\right|\\
&=\max_{\{p_n,\phi\}}2\left|\sum_{n\neq m}\sqrt{p_n p_m}e^{i\phi(n-m)}\Lambda_{00}^{nm}\right|=\max_{\{p_n,\phi\}}2\left|\sum_{n\neq m}\sqrt{p_n p_m}e^{i\phi(n-m)}|\Lambda_{00}^{nm}|e^{i\lambda_{00}^{nm}}\right|\\
&=\max_{\{p_n,\phi\}}2\left|\sum_{n\neq m}\sqrt{p_n p_m}e^{i(\phi+\lambda)(n-m)}|\Lambda_{00}^{nm}|\right|\\
&=\max_{\{p_n,\phi\}}4\sqrt{p_0 p_1}\cos(\phi+\lambda)|\Lambda_{00}^{01}|,\\
&=2|\Lambda_{00}^{01}|.
\end{aligned} \tag{15}$$

Secondly, for a qubit map Λ, we find that

$$\begin{aligned}\tilde{M}_\diamond(\Lambda) &= \min_{\Phi\in\mathcal{DI}} \max_\rho \|\Delta(\Lambda-\Phi)\rho\|_1 \\ &= \min_{\Phi\in\mathcal{DI}} \max_\rho 2\,|\langle 0|\,\Delta(\Lambda-\Phi)\rho\,|0\rangle| \\ &= \min_{\Phi\in\mathcal{DI}} \max_{\{p_n,\phi\}} 2\left|\sum_n p_n(\Lambda_{00}^{nn}-\Phi_{00}^{nn}) + \sum_{n\neq m}\sqrt{p_n p_m}e^{i\phi(n-m)}\Lambda_{00}^{nm}\right| \\ &= \min_{\Phi\in\mathcal{DI}} \max_{\{p_n,\phi\}} 2\left|\sum_n p_n(\Lambda_{00}^{nn}-\Phi_{00}^{nn}) + \sum_{n\neq m}\sqrt{p_n p_m}e^{i(\phi+\lambda)(n-m)}|\Lambda_{00}^{nm}|\right|.\end{aligned} \tag{16}$$

Now we first consider the inner optimization problem, i.e., we fix Φ. The first sum always evaluates to a real number, and the phase ϕ only appears in the second sum. Let us assume that the first sum is positive. The optimum over ϕ is then obviously achieved for $\phi=-\lambda$. If the first sum is negative, the optimum is $\phi=\pi-\lambda$. In both cases, we have

$$\begin{aligned}\tilde{M}_\diamond(\Lambda) &= \min_{\Phi\in\mathcal{DI}} \max_{\{p_n\}} 2\left(\left|\sum_n p_n(\Lambda_{00}^{nn}-\Phi_{00}^{nn})\right| + 2\sqrt{p_0p_1}|\Lambda_{00}^{01}|\right) \\ &\geq \max_{\{p_n\}} 4\sqrt{p_0p_1}|\Lambda_{00}^{01}| = 2|\Lambda_{00}^{01}| \\ &= \mathscr{D}(\Lambda).\end{aligned} \tag{17}$$

Since $\Lambda\Delta\in\mathcal{DI}$, we also have

$$\begin{aligned}\tilde{M}_\diamond(\Lambda) &= \min_{\Phi\in\mathcal{DI}} \max_\rho \|\Delta(\Lambda-\Phi)\rho\|_1 \\ &\leq \max_\rho \|\Delta(\Lambda-\Lambda\Delta)\rho\|_1 \\ &= \mathscr{D}(\Lambda)\end{aligned} \tag{18}$$

and find

$$\tilde{M}_\diamond(\Lambda) = \mathscr{D}(\Lambda) = 2|\Lambda_{00}^{01}| \tag{19}$$

for qubit channels Λ. □

Furthermore, this allows us to prove that $\mathscr{D}(\Lambda)=\max_\rho\|\Delta\Lambda(\mathbb{1}-\Delta)\rho\|_1$ fails to form a measure as defined in the main text.

Proposition 5. *There exist $\Phi\in\mathcal{DI},\Theta$ such that $\mathscr{D}(\Theta)<\mathscr{D}(\Theta\Phi)$.*

Proof. Let $\Theta^{B\leftarrow B}$ be defined via the two Kraus operators

$$K_1=\frac{1}{\sqrt{2}}\begin{pmatrix}i & 1\\ 0 & 0\end{pmatrix},\quad K_2=\frac{1}{\sqrt{2}}\begin{pmatrix}0 & 0\\ 1 & i\end{pmatrix}, \tag{20}$$

i.e., $|B|=2$ (and it is straightforward to check that this defines indeed a CPTP map). With the help of Lem. 4, we find $\mathscr{D}(\Theta)=1$. Choosing $|A|=3$, we can use Prop. 3 to evaluate $\tilde{\mathcal{D}}_A(\Theta^{B\leftarrow B})$ numerically. Moreover, it is possible to extract optimal $\Phi^{B\leftarrow A}$ and σ_A from the solution of the semidefinite program. An optimal choice consists of

$$\sigma_A=\frac{1}{3}\begin{pmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{pmatrix}, \tag{21}$$

which is a maximally coherent state and the quantum operation $\Phi^{B\leftarrow A}$ given by its Choi state

$$J_\Phi=(\mathbb{1}_A\otimes\Phi)\sum_{n,m=1}^3|nn\rangle\langle mm| =: \sum_{n,m=1}^3|n\rangle\langle m|_A\otimes X_{nm} \tag{22}$$

where

$$X_{nn} = \begin{pmatrix} 1/2 & i/6 \\ -i/6 & 1/2 \end{pmatrix} \tag{23}$$

and

$$X_{nm} = \begin{pmatrix} 0 & -i/3 \\ i/3 & 0 \end{pmatrix} \tag{24}$$

for $n \neq m$. It is straightforward to check that J_Φ is hermitian, has eigenvalues $(0, 0, 0, 1, 1, 1)$, and

$$\mathrm{Tr}_B\left(J_\Phi\right) = \mathbb{1}_A, \tag{25}$$

i.e., $\Phi^{B\leftarrow A}$ is CPTP. Moreover, due to

$$J_{\Delta\Phi\Delta} = \Delta_{AB} J_\Phi = \frac{1}{2}\mathbb{1}_{AB} = \Delta_B J_\Phi = J_{\Delta\Phi}, \tag{26}$$

$\Phi \in \mathcal{DI}$. We are not going to prove optimality of Φ and σ, for example by deriving the dual program, but rather note that

$$\begin{aligned}
\tilde{\mathcal{D}}_A(\Theta^{B\leftarrow B}) &= \max_{\tilde{\Phi}^{B\leftarrow A}\in\mathcal{DI}} \max_{\tilde{\sigma}_A} \left\| \Delta\Theta^{B\leftarrow B}\tilde{\Phi}^{B\leftarrow A}\left(\mathbb{1}-\Delta\right)\tilde{\sigma}_A \right\|_1 \\
&\geq \left\| \Delta\Theta^{B\leftarrow B}\Phi^{B\leftarrow A}\left(\mathbb{1}-\Delta\right)\sigma_A \right\|_1 \\
&= \left\| \Delta\Theta^{B\leftarrow B}\,\mathrm{Tr}_A\left[\left(\left(\left(\mathbb{1}-\Delta\right)\sigma_A\right)^T \otimes \mathbb{1}_B\right) J_\Phi\right] \right\|_1 \\
&= \left\| \Delta\Theta^{B\leftarrow B}\,\mathrm{Tr}_A\left[\left(\left(\mathbb{1}-\Delta\right)\sigma_A \otimes \mathbb{1}_B\right) J_\Phi\right] \right\|_1 \\
&= \left\| \Delta\Theta^{B\leftarrow B}\frac{1}{3}\sum_{\substack{n,m=0 \\ n\neq m}}^{3} X_{ij} \right\|_1 \\
&= \frac{2}{3}\left\| \Delta\Theta^{B\leftarrow B}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \right\|_1 \\
&= \frac{2}{3}\left\| \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\|_1 \\
&= \frac{4}{3} > 1,
\end{aligned} \tag{27}$$

which finishes the proof by giving an explicit example. □

Whilst this is not the purpose of this Letter, we note that it is straightforward to show that the family of functionals

$$\tilde{\mathscr{D}}_A(\Theta^{C\leftarrow B}) = \max_{\Phi^{B\leftarrow A}\in\mathcal{DI}} \max_{\sigma_A} \left\| \Delta\Theta^{C\leftarrow B}\Phi^{B\leftarrow A}(\mathbb{1}_A - \Delta_A)\sigma_A \right\|_1 \tag{28}$$

defines resource measures in the detection-incoherent setting in itself. We notice similarities to the measures in Ref. [**?**], but leave further investigations on these measures, for example on an operational interpretation, to future work.

II. Shor's factorization algorithm

In this section we review Shor's algorithm, beginning with the basic prerequisites in number theory, moving on to Shor's protocol and a sequential version introduced in Ref. [**?**]. Additionally, the fine-tuned interplay of the quantum part and the classical post-processing via the continued fraction algorithm is discussed in detail, paving the way for a discussion of the protocols investigated in this work. Some notable examples of further reading on Shor's algorithm are the articles [**? ?**] and the textbook [**?**], on which the following brief review is based.

A. Reduction to order-finding

The first step in Shor's algorithm is the reduction of the integer factorization problem to the so-called order-finding problem [**?**]. Let N denote the integer to be factorized, which consists of m distinct prime factors and can be represented in an n bit string. Furthermore, let x be an integer $1 \leq x < N$ with x coprime to N, i.e., x and N share no common factor. The order-finding problem is then to find the *smallest* integer r such that $x^r = 1 \bmod N$. This integer r is called the order of x modulo N. The reduction of factoring to order-finding results from the following two statements. We omit the proofs at this point, for further reading see for example Ref. [**?**].

Lemma 6. *Given a composite (with more than one distinct prime factor), odd integer N and an integer solution a with $1 \leq a < N$ to the equation $a^2 = 1 \, mod \, N$, that is non-trivial, i.e., $a \neq \pm 1 \, mod \, N$, then at least one of $gcd(a \pm 1, N)$ is a non-trivial factor of N.*

Lemma 7. *For a uniformly chosen x in the range $1 \leq x < N$ and coprime to N, the probability that the order r of x modulo N is even and non-trivial is bounded by*

$$P(r \text{ even, and } x^{r/2} \neq -1 \, mod \, N) > 1 - \frac{1}{2^m}, \tag{29}$$

where m is the number of distinct prime factors of N.

With this at hand, a factorization algorithm is given by the following procedure: In a first step, catch exceptions like N having two as a (multiple) prime factor and check if N is a composite integer, i.e., has more than one distinct prime factor. This can be done efficiently on a classical device, see Ref. [**?**]. These two steps guarantee that the prerequisites of Lem. 6 and Lem. 7 are satisfied. In the next step, choose a random x, check if it is coprime to N, otherwise, repeat until it is. The bottleneck of the algorithm is the order-finding, but assuming we can solve this in polynomial time, determine the order r and subsequently check if it is even and non-trivial (which has sufficiently high probability due to Lem. 7). If so, compute $a = x^{r/2}$ (note that $x^{r/2}$ cannot be $1 \bmod N$ due to the definition of the order) and use Lem. 6 to find a factor of N, otherwise, repeat. The algorithm is run until all prime factors have been found. Since the greatest common divisor can be computed efficiently in polynomial time in the input length n (for example using Euclid's algorithm), having a polynomial time algorithm for order-finding results in a polynomial time algorithm for factorization.

B. Order-finding à la Shor

Shor's coup of an efficient order-finding protocol, depicted schematically in Fig. 2, is at the heart of the factorization algorithm. This *standard* protocol for order-finding utilizes two quantum systems A and B of dimension q and N respectively, where system A consists of L qubits such that $N^2 < q = 2^L < 2N^2$, with N being the number to factor. Along with the classical post-processing via the continued fraction algorithm, the quantum part of the protocol can be separated into three essential ingredients: preparation of an initial state, then the so-called modular exponentiation, and a measurement. The modular exponentiation is defined by the controlled-like unitary

$$U_c = \sum_{n=0}^{q-1} |n\rangle\langle n|_A \otimes U_B^n, \tag{30}$$

where $U_B \left|n\right\rangle_B = \left|xn \bmod N\right\rangle_B$. Note that the modular exponentiation can be implemented in polynomial time [**? ? ? ?**]. The modular exponentiation encodes information about the order r into the state of system A, only requiring knowledge about x and the number N to be factored. The encoding of this information depends on the initial state of the auxiliary system B, and a convenient choice is the state $\left|1\right\rangle_B$. Let us emphasize that other incoherent states can be used as well. For instance in Ref. [**?**], it is shown that choosing a normalized maximally mixed initial state $\mathbb{1}_B$ will increase the runtime of the algorithm at most polynomially. In fact, for factorization problems of the form $N = pq$, where p and q are primes, the increase is asymptotically negligible. After performing the modular exponentiation, the auxiliary system is discarded. For our purposes, the action of the modular exponentiation on system A will be fixed and labeled by $\mathcal{E}$. This channel $\mathcal{E}$ admits the following simple structure.

Lemma 8. *If system B is in the state $|1\rangle_B$, then the effect of the modular exponentiation on system A is given by*

$$\mathcal{E}(\rho_A) = \frac{1}{r}\sum_{j=0}^{r-1}\mathcal{E}_j(\rho_A) \quad \textit{with} \quad \mathcal{E}_j(\rho_A) = R_{j/r}\rho_A R_{j/r}^\dagger, \tag{31}$$

where the $R_{j/r}$ denote rotations around multiples of the fraction of r, i.e., $R_{j/r} = \sum_n e^{2\pi i \frac{j}{r}n}|n\rangle\langle n|$.

Proof. Notice that by definition of the order-finding problem $x^r = 1 \bmod N$. It follows that $U_B^r = \mathbb{1}_B$, since $\forall n$ we find $U_B^r|n\rangle_B = |x^r n \bmod N\rangle_B = |(x^r \bmod N)(n \bmod N) \bmod N\rangle_B = |n\rangle_B$. Hence, orthonormal eigenstates $|\psi_j\rangle_B$ of U_B are simply given by

$$|\psi_j\rangle_B = \frac{1}{\sqrt{r}}\sum_{l=0}^{r-1} e^{-2\pi i l\frac{j}{r}}\,|x^l \bmod N\rangle_B\,, \tag{32}$$

with corresponding eigenvalues of $e^{2\pi i\frac{j}{r}}$. This allows us to expand the auxiliary state as $|1\rangle_B = \frac{1}{\sqrt{r}}\sum_{j=0}^{r-1}|\psi_j\rangle_B$. With this at hand, it is straightforward to calculate

$$\begin{aligned}
\mathcal{E}(\rho_A) &= \mathrm{Tr}_B\left[U_c(\rho_A\otimes|1\rangle\langle 1|_B)U_c^\dagger\right] \\
&= \mathrm{Tr}_B\left[\sum_{n,m}\rho_{nm}|n\rangle\langle m|_A\otimes\frac{1}{r}\sum_{j,j'=0}^{r-1}U_B^n|\psi_j\rangle\langle\psi_{j'}|_B(U_B^m)^\dagger\right] \\
&= \mathrm{Tr}_B\left[\sum_{n,m}\rho_{nm}|n\rangle\langle m|_A\otimes\frac{1}{r}\sum_{j,j'=0}^{r-1}e^{2\pi i(n\frac{j}{r}-m\frac{j'}{r})}|\psi_j\rangle\langle\psi_{j'}|_B\right] \\
&= \frac{1}{r}\sum_{j=0}^{r-1}\sum_{n,m}\rho_{nm}e^{2\pi i\frac{j}{r}(n-m)}|n\rangle\langle m|_A = \frac{1}{r}\sum_{j=0}^{r-1}\mathcal{E}_j(\rho_A).
\end{aligned}$$

□

Let us emphasize the resemblance of $\mathcal{E}$ to a symmetry operation that gives rise to the resource theory of asymmetry [**? ? ?**]. In this particular case, the symmetry group elements are simple rotations, being uniformly weighted to define the symmetry operation $\mathcal{E}$. This symmetry group gives rise to the resource theory of coherence as a special case [**? ? ?**]. Any incoherent state is left invariant under the action of $\mathcal{E}$, i.e., an incoherent state is symmetric with respect to the symmetry group, thereby naturally selecting a set of free states. On the contrary, any coherent state will encode information about r, thus being useful at least in principle for the task of order-finding. Analyzing the protocol in the framework of coherence theory is a natural consequence. Concretely, in this work the performance of the protocol will be quantitatively linked to the ability to create and detect coherence.

Furthermore, it has to be noted that not every single rotation $\mathcal{E}_j$ encodes the order r the way we wish. In fact, any rotation $\mathcal{E}_j$ where $\gcd(j,r) \neq 1$ is equivalent to a rotation around an angle depending on a factor of r rather than r itself. Fortunately, this is sufficiently rare to still allow for an efficient post-processing strategy that estimates r from the measurement statistics efficiently. After the modular exponentiation, a measurement of system A in the Fourier basis produces an outcome k that is forwarded to the continued fraction algorithm (CFA), which will then compute a continued fraction decomposition of k/q.

The continued fraction algorithm computes the decomposition of a number x in the following iterative form: the sum of its closest integer part and the reciprocal of another number, which is then written as the sum of its closets integer part and another reciprocal, and so on, see for example Ref. [**?**]. This decomposition is typically denoted as

$$x = [a_0, a_1, a_2, \ldots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}, \tag{33}$$

where the list is finite for rational x, i.e., $x = [a_0, a_1, \ldots, a_n]$, and infinite otherwise. The so-called convergents, or specifically the m-th convergent of x, is defined by $[a_0, a_1, \ldots, a_m]$. The post-processing of measurement results will be done by computing the convergents of k/q. Some measurement results give sufficiently good approximations to some j/r that allow recovering the latter fraction from k/q by using the CFA to compute the convergents, where one matches j/r.

To clarify which measurement outcomes do so, we continue with the following result from number theory involved in the study of Diophantine Approximation, i.e., approximations of irrational numbers by rational ones. The following statement can be found in various textbooks on number theory, see for example Ref. [**?**]. The first part is also treated in the textbook [**?**], and for completeness, we give a short proof of the statement based on Ref. [**?**].

Theorem 9. *Let x be a positive number and p/q a positive rational number. If*

$$\left| x - \frac{p}{q} \right| < \frac{1}{2q^2} \tag{34}$$

then p/q is a convergent of x. Conversely if p/q is a convergent of x, then

$$\left| x - \frac{p}{q} \right| \leq \frac{1}{q^2}. \tag{35}$$

Proof. Let $\frac{p_n}{q_n}$ denote the convergents to the continued fraction decomposition of x. The sequence $(q_n)_n$ is increasing [**?**], thus there exists some integer n such that $q_n \leq q < q_{n+1}$. Now assume that $\frac{p}{q}$ satisfies the inequality (34) but is not a convergent to the continued fraction algorithm, i.e., $\frac{p}{q} \neq \frac{p_n}{q_n} \, \forall n$. The convergents $\frac{p_n}{q_n}$ are precisely the best approximations to x in the second sense, thus, $|qx - p| < |q_n x - p_n|$ implies $q > q_{n+1}$ [**?**]. Therefore, if $\frac{p}{q}$ is not a convergent with $q < q_{n+1}$ (if $q = q_{n+1}$ there is nothing to show) we find $|q_n x - p_n| \leq |qx - p| < \frac{1}{2q}$, since $\frac{p}{q}$ satisfies Eq. (34) by assumption. This yields

$$\begin{aligned} \frac{1}{qq_n} \leq \frac{|pq_n - qp_n|}{qq_n} = \left| \frac{p}{q} - \frac{p_n}{q_n} \right| \leq \left| x - \frac{p_n}{q_n} \right| + \left| x - \frac{p}{q} \right| \\ < \frac{1}{2qq_n} + \frac{1}{2q^2}, \end{aligned} \tag{36}$$

and thus $q_n > q$, which is a contradiction to $q_n \leq q < q_{n+1}$. Therefore, we find that $q = q_n$ and consequentially $p = p_n$, which concludes the first part of the statement.
For the second part, we can make use of the so-called complete quotients a'_i, see for example Ref. [**?**], defined as $a'_i = [a_i, a_{i+1}, ...]$ which allows us to express arbitrary x as

$$x = \frac{a'_{i+1} p_i + p_{i-1}}{a'_{i+1} q_i + q_{i-1}}, \tag{37}$$

in terms of an arbitrary convergent $\frac{p_i}{q_i}$. Then it follows

$$\begin{aligned} \left| x - \frac{p_i}{q_i} \right| &= \left| \frac{a'_{i+1} p_i + p_{i-1}}{a'_{i+1} q_i + q_{i-1}} - \frac{p_i}{q_i} \right| = \left| \frac{\left(a'_{i+1} p_i + p_{i-1}\right) q_i - p_i \left(a'_{i+1} q_i + q_{i-1}\right)}{q_i \left(a'_{i+1} q_i + q_{i-1}\right)} \right| \\ &= \left| \frac{p_{i-1} q_i - p_i q_{i-1}}{q_i \left(a'_{i+1} q_i + q_{i-1}\right)} \right| = \left| \frac{(-1)^i}{q_i \left(a'_{i+1} q_i + q_{i-1}\right)} \right| \\ &\leq \frac{1}{q_i q_{i+1}}, \end{aligned} \tag{38}$$

where we used in the last line that $a'_{i+1} q_i + q_{i-1} \geq q_{i+1}$. Lastly, since $q_{i+1} \geq q_i$ every convergent and thus also the particular convergent p/q satisfies the inequality $\left| x - \frac{p}{q} \right| \leq \frac{1}{q^2}$. Recall that in the case of a rational x, i.e., a simple finite continued fraction expansion $x = [a_0, a_1, ..., a_n]$, we define the denominator of the $n+1$ convergent simply as q_n, such that the proof also holds for rational x. □

This Theorem provides a sufficient and necessary condition on the absolute difference of the number x and a rational approximation p/q such that said approximation is a convergent of x in the continued fraction decomposition. Coming back to the question of which measurement outcomes are useful, we employ the following Corollary.

Corollary 10. *Let k be an integer with $0 \leq k < q$ where $N^2 < q = 2^L < 2N^2$ that satisfies the inequality*

$$\left| \frac{j}{r} - \frac{k}{q} \right| \leq \frac{\beta}{2q} \tag{39}$$

for some coprime (j, r) with $0 < j < r$ and $\beta = \frac{q-1}{r^2}$. Then the continued fraction expansion of k/q will yield j/r and thereby r, as a convergent.

Proof. According to the first part of Thm. 9, any integer k that satisfies $\left| \frac{j}{r} - \frac{k}{q} \right| < \frac{1}{2r^2}$ will yield j/r as a convergent. Obviously $\frac{\beta}{2q} < \frac{1}{2r^2}$. In particular, since $\beta > 1$ all integers k that obey the inequality $\left| \frac{j}{r} - \frac{k}{q} \right| < \frac{1}{2q}$ will yield j/r as a convergent. □

This justifies the choice of the dimension of quantum system A with $\dim(A) = q$ at the beginning of the discussion. Let us emphasize that extending the margin of error like in Cor. 10 for a $\beta > 1$ has allowed to sharpen Shor's original bound (which basically utilizes a weaker bound with $\beta = 1$) on the coherent protocol, see for example Refs. [**?** **?**]. Looking at the following result, the reason why the post-processing via the CFA works well for a measurement result as in Cor. 10 can be better understood.

Lemma 11. *Consider fixed integers N and $q > N^2$.*

i) Assume you have a fixed integer $0 \leq k < q$. Then there exists at most one pair of integers (j, r) with $1 \leq r < N$, $0 \leq j < r$, and $gcd(j, r) = 1$ such that $\left|\frac{j}{r} - \frac{k}{q}\right| < \frac{1}{2q}$.

ii) Assume that you have a pair of integers (j, r) with $1 \leq r < N$ and $0 \leq j < r$. Then there exists an integer $0 \leq k < q$ such that $\left|\frac{j}{r} - \frac{k}{q}\right| < \frac{1}{2q}$ is satisfied.

Proof. We begin with i). Assume that there exist two distinct fractions $\frac{j'}{r'} \neq \frac{j}{r}$ that satisfy $\left|\frac{j}{r} - \frac{k}{q}\right| < \frac{1}{2q}$ and $\left|\frac{j'}{r'} - \frac{k}{q}\right| < \frac{1}{2q}$. It follows that

$$\left|\frac{j'}{r'} - \frac{j}{r}\right| = \left|\frac{j'}{r'} - \frac{k}{q} + \frac{k}{q} - \frac{j}{r}\right| < \frac{1}{q} < \frac{1}{N^2}. \tag{40}$$

On the other hand $\left|\frac{j'}{r'} - \frac{j}{r}\right| = \left|\frac{j'r - jr'}{rr'}\right| > \frac{1}{N^2}$, since $r, r' < N$ and there exists an integer i such that $|j'r - jr'| = |i| \geq 1$. By contradiction the two fractions are identical.

For ii), we note that the distance between neighboring fractions $\frac{k}{q}$ is given by $\frac{1}{q}$. Therefore, there always exists a k' such that $\left|\frac{j}{r} - \frac{k'}{q}\right| \leq \frac{1}{2q}$. However, equality can only hold if r is a power of 2, in which case there exists a k such that $\frac{k}{q}$ samples $\frac{j}{r}$ exactly. □

Combining the results of Cor. 10 and Lem. 11 tells us, that given a single rotation $\mathcal{E}_j$ as defined in Lem. 8 and with j coprime to r, there always exists a measurement outcome k that will yield r via the continued fraction algorithm. With this at hand, we define the following two sets for a fixed j coprime to r

$$\begin{aligned} \mathcal{K}_1^j &= \left\{k : 0 \leq k < q \wedge \left|\frac{j}{r} - \frac{k}{q}\right| < \frac{1}{2q}\right\}, \\ \mathcal{K}_2^j &= \left\{k : 0 \leq k < q \wedge \left|\frac{j}{r} - \frac{k}{q}\right| \leq \frac{1}{r^2}\right\}. \end{aligned} \tag{41}$$

Additionally we define the sets $\mathcal{K}_1, \mathcal{K}_2$ as $\mathcal{K}_i = \cup_j \mathcal{K}_i^j$, where the union is formed over all j smaller than and coprime to r. The set $\mathcal{K}_1$ contains all measurement outcomes that will yield the correct order r by putting the outcome in the continued fraction algorithm. The second set $\mathcal{K}_2$ consists of all outcomes that obey the necessary condition to be a convergent of the CFA according to Thm. 9, i.e., it contains all outcomes that will yield the correct r via the CFA but potentially also outcomes that do not. Let us conclude this preliminary discussion by noting what happens for an unknown and randomly chosen j (or equivalently a uniformly weighted $\mathcal{E}_j$, as we got here) during the post-processing. Suppose for the sampled $\mathcal{E}_j$, j and r share a common factor. The post-processing will then maximally yield a factor of r and thus fail. This case is however rare: the probability that a randomly chosen j is coprime to r is given by $\varphi(r)/r$, where $\varphi(r)$ denotes Euler's totient function. This ratio between Euler's totient function and its argument is bounded by $\frac{\varphi(r)}{r} > \frac{\delta}{\log\log r} > \frac{\delta}{\log\log N}$, for some positive constant δ, according to a well-known result by Hardy [**?** , Theorem 328]. In fact, the latter inequality is asymptotically tight for infinitely many values of r.

C. Sequential order-finding

Furthermore, we have to discuss a sequential version of Shor's original order-finding protocol that allows reducing the number of qubits drastically for large factorization problems. The protocol is based on a semi-classical implementation of the combination of an inverse quantum Fourier transform and a measurement in the computational basis following directly afterward (see Refs. [**?** **?**]): Assume the inverse Fourier transform is implemented via its *standard* decomposition into Hadamard gates and controlled rotations as depicted in Fig. 1, see also Ref. [**?**]. Fig. 2 therefore shows an implementation of Shor's algorithm in which the measurement outcome has to be reordered in reverse order. As explained in detail in Ref. [**?**], it is then possible to do the measurement on the first qubit directly after the first Hadamard gate belonging to the inverse Fourier transform was implemented (the second Hadamard gate in the figure) and use its outcome to classically control all the following rotations that

depend on this qubit. A similar argument holds for the other qubits as well: after the respective Hadamard gate in their line, one can directly measure them and control all following rotations classically depending on the outcome. Since all the controlled rotations in one line lead to an effective rotation, in this way, one can replace them with a single effective classically controlled rotation R_l' that depends on the previous measurement outcomes. This is shown in Fig. 3.

Thereby, the gates and measurements on the individual qubits are performed sequentially, which allows to split Shor's protocol into blocks, see Fig. 4, that each utilize only a single control qubit on which the Hadamard gates and the classically controlled rotations R_l' are performed. The single control qubit can be *recycled* after each block, such that the total amount of qubits required decreases to $\log N + 1$. Due to this decomposition, Shor's original protocol and the sequential version lead to identical measurement statistics if the auxiliary systems are initialized in the same state.

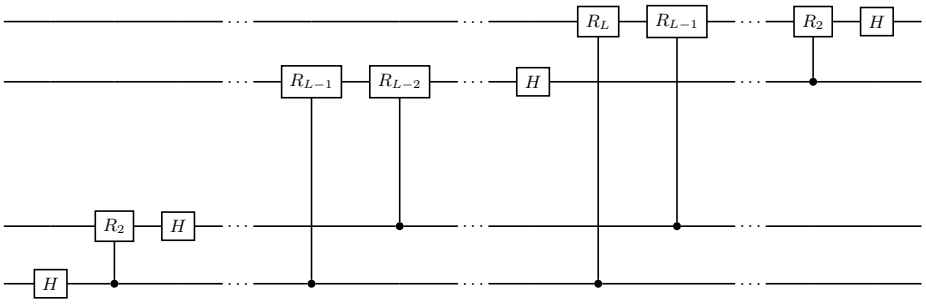

Figure 1: Standard decomposition of the inverse Fourier transform into Hadamard gates and controlled rotations R_l. The controlled rotation R_l adds a phase of $-2\pi i/2^l$ to $|1\rangle$ and leaves $|0\rangle$ unchanged. An additional initial reordering of the qubits in reverse order is not shown.

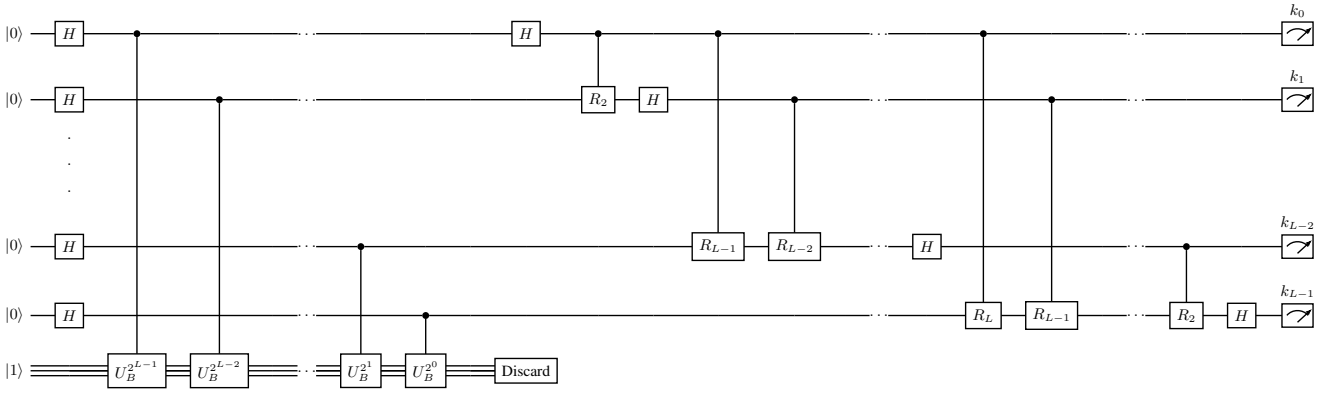

Figure 2: Decomposition of Shor's algorithm, with the inverse Fourier transform decomposed into Hadamard gates and controlled rotations R_l. A controlled rotation R_l adds a phase of $-2\pi i/2^l$ to $|1\rangle$ and leaves $|0\rangle$ unchanged. This leads to a total outcome $k = \sum_{i=0}^{L-1} 2^i k_i$.

III. Choosing the free operations

As discussed in the main text, we fix the overall protocol that we investigate and vary only parts of it. Here, we will explain our choices a bit more in detail. First, we assume that the post-processing of measurement results after a single round is achieved by the continued fraction algorithm. If this fails, then we restart the algorithm and perform the post-processing without accounting for the previous outcomes, thereby ignoring possible correlations between results of failed trials. In general, this is not the best possible post-processing strategy. An example of a more involved strategy can be found in Ref. [**?**]. Nevertheless, for simplicity, we assume this fixed post-processing involves only the outcome of individual trials since we are not optimizing over post-processing strategies anyway. The ability to create, then utilize, and finally detect coherence is a key feature in the protocol. Imposing constraints on these abilities can be done naturally within the framework of dynamical resource theories of coherence.

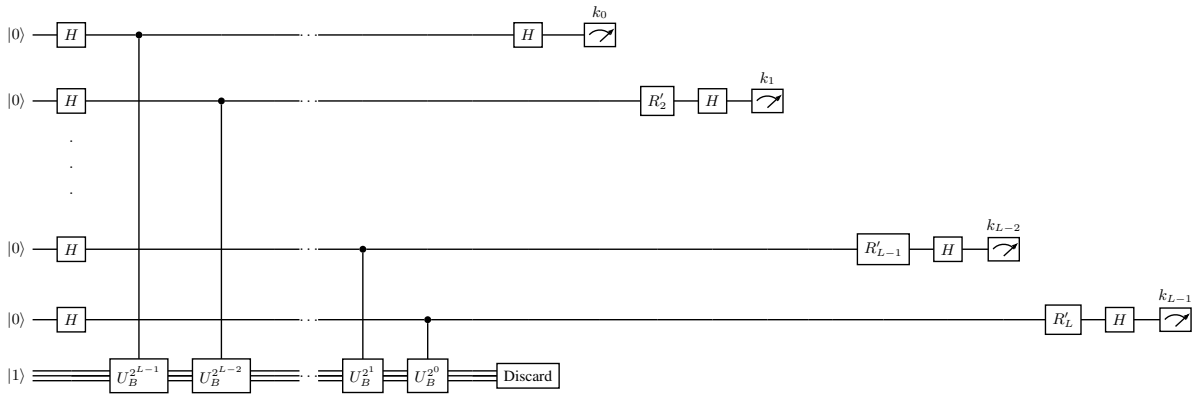

Figure 3: Rewriting Shor's algorithm using classically controlled effective rotations $R'_l = \sum_{n=0}^{1} e^{-2\pi i n \phi'_l} |n\rangle\langle n|$ that depend on the outcomes of previous measurements via $\phi'_l = \sum_{a=2}^{l} k_{l-a}/2^a$.

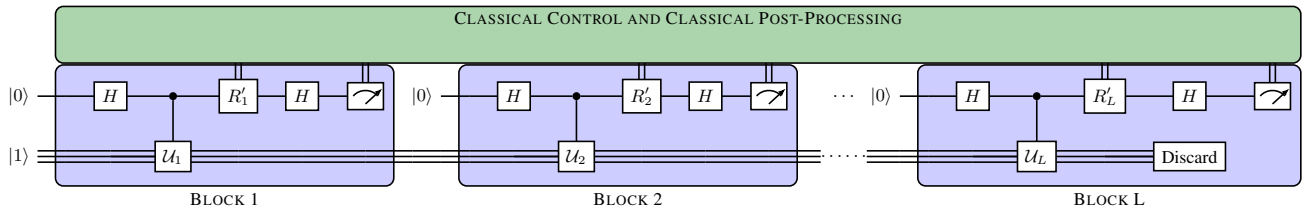

Figure 4: Sequential order-finding protocol using the semi-classical version of the Fourier transform. The modular exponentiation factors into single qubit controlled-operations given by $\mathcal{U}_l = U_B^{2^{L-l}}$ and the classically controlled rotations $R'_l = \sum_n e^{-2\pi i n \phi'_l} |n\rangle\langle n|$, where the phases ϕ'_l depend on the previous measurement outcomes k_l via $\phi'_l = \sum_{a=2}^{l} k_{l-a}/2^a$, see Refs. [**?** **?**].

Notice that expressing the protocol in the form of Fig. 4 makes it clear that except for the Hadamard gates, the protocol utilizes only incoherent input states, channels $\mathcal{U}_l$ and R'_l that can neither detect nor create coherence, and measurements in the incoherent basis. Replacing Hadamard gates by quantum channels $S_1^{(l)}[\Theta_l]$ and $S_2^{(l)}[\Lambda_l]$ respectively results in the protocol depicted in Fig. 5. If no particular block is considered, we omit the label l and refer to the channels for creation and detection simply as Θ and Λ.

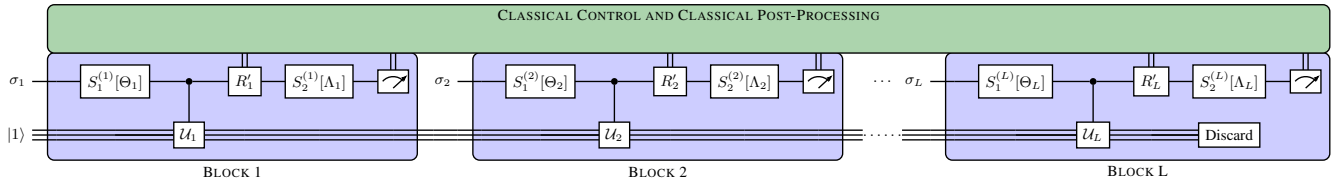

Figure 5: Circuit representation of the order-finding protocol using only channels Θ_l and Λ_l to create and detect coherence. The outcomes of an (incoherent) projective measurement in the computational basis are forwarded to the classical control and post-processing unit, which re-initializes the single control qubit, classically controls the rotations R'_l to implement the inverse Fourier transform, and lastly computes the continued fraction decomposition to yield an estimate of the order r.

Let us now explain why the symmetry of the fully *coherent* protocol that uses the same channel to create and detect coherence (i.e., the Hadamard gate) has to be broken in the more general case: The ability to create and detect coherence are two fundamentally different properties a quantum channel can possess, which in turn gives rise to two different resources that are generally not interconvertible (e.g., a channel $\Gamma(\sigma) = \rho\,\mathrm{Tr}(\sigma)$ can prepare coherence if ρ is chosen suitably, but not detect, whilst a destructive measurement in the Fourier basis can detect but not prepare coherence). The Hadamard gate can however create a maximally coherent state (by applying it to $|0\rangle$), but also maximizes the NSID-measure [**?**]. Therefore, it plays a dual role, i.e., it both creates and detects coherence.

As mentioned in the main text, the choice of free channels follows naturally. If Θ is incapable of creating coherence,

no information about the order can be encoded. If Λ cannot detect coherence, none of this information can influence the measurement statistics. Thus, the lack of either ingredient renders the protocol practically "useless" by reducing it to a random number generator independent of the order that it is supposed to estimate, and is moreover classically simulable. Therefore, the choices of free channels are maximally incoherent channels $\mathcal{MIO}$ [**?**] and detection-incoherent channels $\mathcal{DI}$ [**?**], also known as non-activating [**?**]. Let us mention that this random number generator gives rise to different probability distributions depending on the structure of the free channels Θ_{free} and Λ_{free}. Details will follow in the next section.
It is tempting to choose the set of creation-detection incoherent channels $\mathcal{CDI}$ as the set of free channels, i.e., the channels that can neither create nor detect coherence, also known as dephasing-covariant channels [**?** **?** **?**], classical [**?**], or commuting [**?**]. This would keep the symmetry of the protocol and seems to be an intuitive choice as it leads to a "fully classical" protocol. However, it does not lead to a consistent connection between operational advantages and deployed resources: Imagine we would use an channel $\Lambda \in \mathcal{DI}$ with $\Lambda \notin \mathcal{MIO}$ for detection. Although not granting any operational advantage, this channel has to be considered resourceful. In contrast, our choice of different sets of free channels naturally leads to an operationally meaningful use of resources.

Furthermore, the channel Λ utilized in the detection scheme is assumed to be a unital map. This assumption is physically motivated: The measurement statics of the incoherent measurement are uniquely determined by the pre-measurement populations. To be maximally sensitive to information about r, we want that the deviation of the measurement statistics from a flat distribution purely depends on the coherences that Λ mapped to populations, and not on a reshuffling of populations that does not include information about r. Without knowing r, we can choose both free super-channels S_1 and S_2 such that this is the case iff Λ is unital. Since the state before $\mathcal{U}_l$ is still independent of r, we can always choose S_1 such that its populations are equal to a maximally mixed state, without affecting the coherences (because the phases of the coherences are still independent of r and therefore known). After $\mathcal{U}_l$, the phase of the coherences depends however on r, and we can thus not alter the populations without varying the coherences (or knowing r). Thus, if Λ were not unital but could detect coherence, the following might happen: The measurement statistics depend stronger on the population reshuffling than on the detected coherences. In this case, we would perform worse than with a free channel that leads to equally distributed random numbers and therefore on average produces better guesses of r than random numbers that are weighted in a way that does not depend on r. To avoid this, we must choose Λ_l to be unital and similarly choose the super-channels $S_2^{(l)}$ to be unitality-preserving.

IV. Success probability

The success probability of the order-finding protocol, consisting of the quantum part combined with the continued fraction algorithm, can now be expressed. To ease up the notation, we make use of the equivalence between Shor's original version and the sequential version. That way, there is no need to laboriously track the back-action of the measurements in each block on the auxiliary system, which allows us to express the success probability compactly. Recall that the detection part, i.e., the *standard* implementation of the inverse Fourier transform (see Fig. 1), was altered only by replacing the Hadamard gates with channels $S_2^{(l)}[\Lambda_l]$. Let us denote the resulting channel by $F_{S_2^{(l)}[\Lambda_l]}$. Furthermore, we use $\tilde{\sigma} = \bigotimes_l \sigma_l$ and the POVM elements $M_k = \bigotimes_l M_{k_l}$. With this notation, the incoherent measurement $\mathbb{M} = \{M_k\}_k$ results in the measurement statistics

$$p_k(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l]; \tilde{\sigma}, \mathbb{M}) = \mathrm{Tr}\left[M_k \Delta F_{S_2^{(l)}[\Lambda_l]} \mathcal{E} \bigotimes_{l=1}^{L} S_1^{(l)}[\Theta_l] \tilde{\sigma} \right], \tag{42}$$

where $\mathcal{E}$ denotes the uniformly weighted rotations described in Lem. 8. After completing all blocks, the measurement outcome k is forwarded to the CFA, which will return the order r with a probability of $P(k \to r \,|\, \text{CFA})$. Therefore, the probability that the order-finding protocol in Fig. 5 succeeds, is given by

$$P^{\text{succ}}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l]; \tilde{\sigma}, \mathbb{M}) = \sum_k P(k \to r \,|\, \text{CFA})\; p_k(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l]; \tilde{\sigma}, \mathbb{M}). \tag{43}$$

Since all incoherent input states σ_l, incoherent measurements, and free super-channels $S_1^{(l)}$ and $S_2^{(l)}$ are available at no cost, we choose them optimally (but without knowledge of r and in a way that is implementable efficiently), which ensures that the available resources are used adequately. The resulting success probability is then given by

$$
\begin{aligned}
P^{\text{succ}}(\Theta_l,\Lambda_l) &= \max_{\substack{\tilde{\sigma}\in\mathcal{I}\\ \mathbb{M}\in\mathcal{IM}}} \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}} P^{\text{succ}}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{M}) \\
&= \max_{\substack{\tilde{\sigma}\in\mathcal{I}\\ \mathbb{M}\in\mathcal{IM}}} \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}} \sum_k P(k\to r\,|\,\text{CFA})\; p_k(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{M}).
\end{aligned} \tag{44}
$$

Since every incoherent POVM $\mathbb{M}$ is equivalent to a detection-incoherent channel followed by a projective measurement $\mathbb{P}$ in the incoherent basis [**?**], the optimization over the measurement can be absorbed into the optimization of the detection-incoherent super-channel, i.e.,

$$
P^{\text{succ}}(\Theta_l,\Lambda_l) = \max_{\tilde{\sigma}\in\mathcal{I}} \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}} \sum_k P(k\to r\,|\,\text{CFA})\, p_k(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P}). \tag{45}
$$

In general, this expression seems hard to evaluate exactly. However, in the following section, we will provide bounds allowing us to compare performance and resource content.

V. Proof of the results in the main text

In this section we give the proofs of the results presented in the main text, i.e., we derive bounds on the success probability given in Eq. (45).

A. Preliminaries

We start by presenting a bound on a product that we will later use to obtain a lower bound on the performance of the order-finding protocol.

Lemma 12. *For positive numbers $\{a_l\}_l$ with $0\le a_l\le 1\,\forall\, l$ the following inequalities hold:*

$$
\frac{4}{\pi^2}\prod_{l=1}^{L}\frac{1}{2}\left[1+a_l\right] \le \prod_{l=1}^{L}\frac{1}{2}\left[1+a_l\cos\left(\frac{\pi}{2^l}\right)\right] \le \prod_{l=1}^{L}\frac{1}{2}\left[1+a_l\right]. \tag{46}
$$

Proof. Since $a_l\ge 0$, the upper bound holds trivially. For the lower bound, notice that the term $0\le\cos\left(\frac{\pi}{2^l}\right)<1$ rapidly converges to one for increasing l. Thereby, it is reasonable that the deviation from the simple upper bound is small. First rewrite the product as

$$
\begin{aligned}
\prod_{l=1}^{L}\frac{1}{2}\left[1+a_l\cos\left(\frac{\pi}{2^l}\right)\right] &= \prod_{l=1}^{L}\frac{1}{2}\left[1+\cos\left(\frac{\pi}{2^l}\right)\right]\prod_{l=1}^{L}\left[\frac{1+a_l\cos\left(\frac{\pi}{2^l}\right)}{1+\cos\left(\frac{\pi}{2^l}\right)}\right] \\
&= \prod_{l=1}^{L}\frac{1}{2}\left[1+\cos\left(\frac{\pi}{2^l}\right)\right]\prod_{l=1}^{L}\left[\frac{1+a_l}{2}+\frac{1-a_l}{2}\frac{1-\cos\left(\frac{\pi}{2^l}\right)}{1+\cos\left(\frac{\pi}{2^l}\right)}\right] \\
&\ge \prod_{l=1}^{L}\frac{1}{2}\left[1+\cos\left(\frac{\pi}{2^l}\right)\right]\prod_{l=1}^{L}\left[\frac{1+a_l}{2}\right] \\
&= \prod_{l=1}^{L}\cos^2\left(\frac{\pi}{2^{l+1}}\right)\prod_{l=1}^{L}\left[\frac{1+a_l}{2}\right].
\end{aligned} \tag{47}
$$

Now utilize a special case of the Viète-Euler product formula, see for example Ref. [**?**], $\frac{\sin(x)}{x}=\prod_{l=1}^{\infty}\cos\left(\frac{x}{2^l}\right)$ with $x=\pi/2$ which results in

$$
\frac{4}{\pi^2}=\prod_{l=1}^{\infty}\cos^2\left(\frac{\pi}{2^{l+1}}\right)=\prod_{l=1}^{L}\cos^2\left(\frac{\pi}{2^{l+1}}\right)\prod_{l=L+1}^{\infty}\cos^2\left(\frac{\pi}{2^{l+1}}\right)\le\prod_{l=1}^{L}\cos^2\left(\frac{\pi}{2^{l+1}}\right), \tag{48}
$$

which concludes the proof. Notice that the last inequality is asymptotically tight for $L\to\infty$. □

Let us proceed by introducing a particular super-channel S_2 for the detection scheme. The channel $S_2[\Lambda]$ mimics a key property of the Hadamard gate that will allow us to mimic a key property of the inverse Fourier transform such that the protocol yields useful measurement outcomes with high probability.

Lemma 13. *Let Λ be a qubit quantum channel, defined in the index representation as*

$$\Lambda(|n\rangle\langle m|) = \sum_{kl} \Lambda_{kl}^{nm} |k\rangle\langle l|. \tag{49}$$

There exists an implementable super-channel $S_2 \in \mathcal{DIS}$, such that

$$\Delta S_2[\Lambda](|n\rangle\langle m|) = \sum_{k=0}^{1} |\Lambda_{kk}^{nm}| e^{\pi i k(n-m)} |k\rangle\langle k|. \tag{50}$$

It suffices to choose a super-channel of the form $S[\Lambda] = \Lambda\Phi_2$. We refer to the action of the channel $S_2[\Lambda]$ on any state as Hadamard-like, or shortly the channel is Hadamard-like.

Proof. Let us use the notation $\Lambda_{kl}^{nm} = |\Lambda_{kl}^{nm}| e^{i\lambda_{kl}^{nm}}$ and choose Φ_2 as the channel corresponding to the unitary $\sum_n e^{i\lambda_{00}^{01} n} |n\rangle\langle n|$. In the following, we will see that this choice satisfies our requirements. Note first that $\Lambda_{kk}^{nn} \geq 0 \, \forall k, n$, and therefore property (50) holds for populations. Moreover

$$\langle 0| \left(\Lambda\Phi_2 |0\rangle\langle 1|\right) |0\rangle = |\Lambda_{00}^{01}| \tag{51}$$

as required, and due to trace preservation

$$\langle 1| \left(\Lambda\Phi_2 |0\rangle\langle 1|\right) |1\rangle = |\Lambda_{00}^{10}| e^{i(\lambda_{11}^{01} - \lambda_{00}^{01})} = -|\Lambda_{00}^{01}|, \tag{52}$$

i.e., $e^{i(\lambda_{11}^{01} - \lambda_{00}^{01})} = -1 = e^{i\pi 1(0-1)}$, which finishes this case too. Finally,

$$\Lambda\Phi_2 |1\rangle\langle 0| = \left(\Lambda\Phi_2 |0\rangle\langle 1|\right)^\dagger, \tag{53}$$

from which the remainder of the proof follows. □

B. A lower bound

A lower bound on the success probability (45) is essential to bound the runtime of the algorithm. For the coherent protocol, which utilizes Hadamard gates, it has been shown that the success probability is lower bounded by a function that is slowly growing in the number N to factor [**?**]. In this section, we prove a similar bound for less resourceful channels, that will include the coherent bound derived by Shor as a limiting case.

For the lower bound on Eq. (45) discussed in the following, we can simply choose a specific set of free super-channels $S_1^{(l)}$ and $S_2^{(l)}$, which are depicted in Fig. 6.

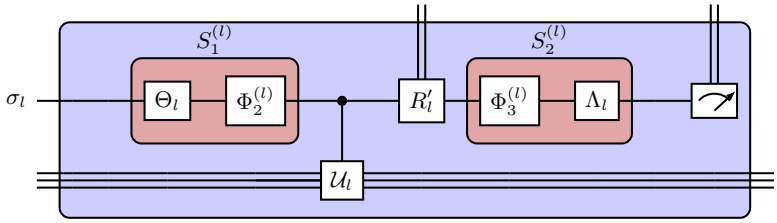

Figure 6: The particular super-channels that are employed for each individual block to derive the lower bound on the success probability in Thm. 14.

The super-channels $S_2^{(l)}$ employed in the detection part will be the ones that lead to Hadamard-like channels (see Lem. 13), whereas the $S_1^{(l)}$ will be introduced in the following Theorem.

Theorem 14. *The success probability of the order-finding protocol with qubit channels Θ_l and unital Λ_l is bounded by*

$$P^{succ}(\Theta_l,\Lambda_l) \geq \frac{4}{\pi^2}\frac{\varphi(r)}{r}\prod_{l=1}^{L}\left(\frac{1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)}{2}\right), \tag{54}$$

where $\mathscr{C}$ denotes the cohering power with respect to the robustness of coherence, $\tilde{M}_\diamond$ is the NSID-measure, both introduced in the main text, and $\varphi(r)$ denotes Euler's totient function.

Proof. Let us consider an idealized version of the order-finding protocol first. Assume that instead of a symmetry channel $\mathcal{E}$, derived in Lem. 8, only a single rotation $\mathcal{E}_j$, where j is coprime to r, is utilized. Let us denote the success probability of this order-finding protocol by $\tilde{P}_j^{\text{succ}}(\Theta_l,\Lambda_l)$, which is given by

$$\tilde{P}_j^{\text{succ}}(\Theta_l,\Lambda_l) = \max_{\tilde{\sigma}\in\mathcal{I}} \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}} \sum_k P(k\to r\,|\,\text{CFA})\, p_k^{(j)}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P}), \tag{55}$$

where $p_k^{(j)}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P}) = \text{Tr}\left[P_k \Delta F_{S_2^{(l)}[\Lambda_l]}\mathcal{E}_j \bigotimes_l S_1^{(l)}[\Theta_l]\tilde{\sigma}\right]$ (recall the notations introduced around Eq. (42)).

One way of obtaining a compact lower bound is the following: Instead of accounting for all possible measurement outcomes which may or may not yield the correct r via the classical post-processing, i.e., all outcomes contained in the set $\mathcal{K}_2^j$ in (41), we focus on the set $\mathcal{K}_1^j$. Since the set $\mathcal{K}_1^j$ contains exactly one outcome, we use this single measurement outcome k' obeying $|\frac{j}{r}-\frac{k'}{q}|<1/(2q)$ to provide a lower bound according to

$$\begin{aligned}
\tilde{P}_j^{\text{succ}}(\Theta_l,\Lambda_l) &= \max_{\tilde{\sigma}\in\mathcal{I}} \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}} \sum_k P(k\to r\,|\,\text{CFA})\, p_k^{(j)}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P}) \\
&\geq \max_{\tilde{\sigma}\in\mathcal{I}} \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}} P(k'\to r\,|\,\text{CFA})\, p_{k'}^{(j)}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P}) \\
&= \max_{\tilde{\sigma}\in\mathcal{I}} \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}} p_{k'}^{(j)}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P}),
\end{aligned} \tag{56}$$

where in the third line we used the results of Cor. 10 and Lem. 11. To further simplify this bound, we make use of particular super-channels $S_1^{(l)}, S_2^{(l)}$ depicted in Fig. 6. For the detection we choose $S_2^{(l)}[\Lambda_l] = \Lambda_l\Phi_2^{(l)}$ such that we obtain a Hadamard-like channel (see Lem. 13). For the adjustment of the channels Θ_l we do the following: after Θ_l was applied to σ_l, we perform a rotation removing the relative phase of the qubit state $\Theta_l(\sigma_l)$. Let us denote this rotation by $\mathcal{R}_1^{(l)}$. Then we apply the map $\tilde{\Phi}(\rho) = \frac{1}{2}(\rho+\sigma_x\rho\sigma_x)$. This post-processing of $\Theta_l(\sigma_l)$ results in a state of the form $S_1^{(l)}[\Theta_l](\sigma_l) = \frac{1}{2}\mathbb{1} + c_l\sigma_x$ where $c_l\geq 0$, which is then used to probe $\mathcal{E}_j$. Importantly, it does not carry any intrinsic phases that may interfere with the detection of the phases induced by $\mathcal{E}_j$. Enforcing uniformly distributed populations (which are preserved since Λ_l is unital by assumption) will ensure that the deviation in the measurement statistics caused by coherence can be maximized. Choosing super-channels defined in such a way, i.e., $S_1^{(l)}[\Theta_l] = \Phi_1^{(l)}\Theta_l = \tilde{\Phi}\mathcal{R}_1^{(l)}\Theta_l$ and $S_2^{(l)}[\Lambda_l] = \Lambda_l\Phi_2^{(l)}$, we obtain from Eq. (56) that

$$\begin{aligned}
\tilde{P}_j^{\text{succ}}(\Theta_l,\Lambda_l) &\geq \max_{\tilde{\sigma}\in\mathcal{I}} \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}} p_{k'}^{(j)}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P}) \\
&\geq \max_{\tilde{\sigma}\in\mathcal{I}}\; p_{k'}^{(j)}(\Phi_1^{(l)}\Theta_l, \Lambda_l\Phi_2^{(l)};\tilde{\sigma},\mathbb{P}) \\
&= \max_{\tilde{\sigma}\in\mathcal{I}}\; \text{Tr}\left[P_{k'}\Delta F_{\Lambda_l\Phi_2^{(l)}}\mathcal{E}_j \bigotimes_l \Phi_1^{(l)}\Theta_l\tilde{\sigma}\right].
\end{aligned} \tag{57}$$

At this point, we notice that we can express $\mathcal{E}_j$ (see Lem. 8) as a tensor product: expanding n into its binary representation, i.e.,

$n = n_1 n_2 ... n_L = \sum_{l=1}^{L} n_l 2^{L-l}$, we find

$$R_{j/r} = \sum_{n=0}^{2^L-1} e^{2\pi i \frac{j}{r} n} |n\rangle\langle n| \tag{58}$$

$$= \sum_{n_1=0}^{1} \cdots \sum_{n_L=0}^{1} e^{2\pi i \frac{j}{r} \sum_{l=1}^{L} n_l 2^{L-l}} |n_1 n_2 ... n_L\rangle\langle n_1 n_2 ... n_L| \tag{59}$$

$$= \bigotimes_{l=1}^{L} \sum_{n_l=0}^{1} e^{2\pi i \frac{j}{r} n_l 2^{L-l}} |n_l\rangle\langle n_l| \tag{60}$$

$$= \bigotimes_{l=1}^{L} R_{j/r}^{(l)}, \tag{61}$$

with $R_{j/r}^{(l)} = \sum_{n_l=0}^{1} e^{2\pi i \frac{j}{r} n_l 2^{L-l}} |n_l\rangle\langle n_l|$. We thus define $\mathcal{E}_j^{(l)}(\rho) := R_{j/r}^{(l)} \rho \left(R_{j/r}^{(l)}\right)^\dagger$ and notice that, with the equivalence of Figs. 2, 3, and 4 (and $\tilde{\sigma} = \bigotimes_l \sigma_l$),

$$\begin{aligned} \mathrm{Tr}\left[P_{k'} \Delta F_{\Lambda_l \Phi_2^{(l)}} \mathcal{E}_j \bigotimes_{l=1}^{L} \Phi_1^{(l)} \Theta_l \tilde{\sigma}\right] &= \mathrm{Tr}\left[P_{k'} \Delta F_{\Lambda_l \Phi_2^{(l)}} \bigotimes_{l=1}^{L} \left(\mathcal{E}_j^{(l)} \Phi_1^{(l)} \Theta_l \sigma_l\right)\right] \\ &= \prod_{l=1}^{L} \mathrm{Tr}\left[P_{k'}^{(l)} \Delta \Lambda_l \Phi_2^{(l)} R_l' \mathcal{E}_j^{(l)} \Phi_1^{(l)} \Theta_l \sigma_l\right], \end{aligned} \tag{62}$$

where $P_{k'}^{(l)} = |k'_{l-1}\rangle\langle k'_{l-1}|$ for a total $k' = \sum_{l=0}^{L-1} 2^l k'_l$ (see Figs. 2 and 3). Here it is important that we understand the product as ordered, since R_l' depends on all previous measurement outcomes. Recall that $c_l = |[\Theta_l(\sigma_l)]_{01}|$, with which

$$\begin{aligned} \Delta \Lambda_l \Phi_2^{(l)} R_l' \mathcal{E}_j^{(l)} \Phi_1^{(l)} \Theta_l \sigma_l =& \Delta \Lambda_l \Phi_2^{(l)} R_l' \mathcal{E}_j^{(l)} \left[\frac{1}{2}\mathbb{1} + c_l \sigma_x\right] \\ =& \Delta \Lambda_l \Phi_2^{(l)} R_l' \left[\frac{1}{2}\mathbb{1} + c_l \left(e^{-2\pi i \frac{j}{r} 2^{L-l}} |0\rangle\langle 1| + h.c.\right)\right] \\ =& \Delta \Lambda_l \Phi_2^{(l)} \left[\frac{1}{2}\mathbb{1} + c_l \left(e^{-2\pi i \left(\frac{j}{r} 2^{L-l} - \sum_{a=2}^{l} k'_{l-a}/2^a\right)} |0\rangle\langle 1| + h.c.\right)\right] \\ =& \frac{1}{2}\mathbb{1} + c_l \left(e^{-2\pi i \left(\frac{j}{r} 2^{L-l} - \sum_{a=2}^{l} k'_{l-a}/2^a\right)} \sum_{b_l=0}^{1} \left|(\Lambda_l)_{b_l b_l}^{01}\right| e^{i\pi b_l} |b_l\rangle\langle b_l| + h.c.\right). \end{aligned} \tag{63}$$

Since the robustness of coherence coincides with the l_1 measure of coherence for qubits, see Ref. [**?**], we find

$$\max_{\sigma_l \in \mathcal{I}} c_l = \max_{\sigma_l \in \mathcal{I}} |[\Theta_l(\sigma_l)]_{01}| = \max_{\sigma_l \in \mathcal{I}} C(\Theta_l \sigma_l)/2 = \mathscr{C}(\Theta_l)/2, \tag{64}$$

where $\mathscr{C}$ denotes the cohering power with respect to the robustness, and $\left|(\Lambda_l)_{00}^{01}\right| = \left|(\Lambda_l)_{11}^{01}\right|$ [**?** , Prop. 6] and thus

$$\begin{aligned} &\max_{\tilde{\sigma} \in \mathcal{I}} \mathrm{Tr}\left[P_{k'} \Delta F_{\Lambda_l \Phi_2^{(l)}} \bigotimes_{l} \mathcal{E}_j^{(l)} \Phi_1^{(l)} \Theta_l \tilde{\sigma}\right] \\ &= \prod_{l=1}^{L} \max_{\sigma_l \in \mathcal{I}} \left[\frac{1}{2} + c_l \left|(\Lambda_l)_{k'_0 k'_0}^{01}\right| \left(e^{-2\pi i \left(\frac{j}{r} 2^{L-l} - \sum_{a=1}^{l} k'_{l-a}/2^a\right)} + h.c.\right)\right] \\ &= \prod_{l=1}^{L} \frac{1}{2} \left[1 + \mathscr{C}(\Theta_l) \left|(\Lambda_l)_{00}^{01}\right| \left(e^{-2\pi i \left(\frac{j}{r} 2^{L-l} - \sum_{a=1}^{l} k'_{l-a}/2^a\right)} + h.c.\right)\right]. \end{aligned} \tag{65}$$

Following the usual procedure (see for example Ref. [**?**]), we note that $\sum_{a=1}^{l} k'_{l-a}/2^a = 2^{-l} \sum_{b=0}^{l-1} k'_b 2^b$ and $e^{2\pi i 2^{-l} \sum_{b=l}^{L-1} k'_b 2^b} = 1$. Therefore,

$$e^{-2\pi i \left(\frac{j}{r} 2^{L-l} - \sum_{a=1}^{l} k'_{l-a}/2^a\right)} = e^{-2\pi i 2^{L-l} \left(\frac{j}{r} - \frac{k'}{2^L}\right)} \tag{66}$$

and if we consider the worst-case scenario we find

$$\begin{aligned}\max_{\tilde{\sigma}\in\mathcal{I}} \operatorname{Tr}\left[P_{k'}\Delta F_{\Lambda_l\Phi_2^{(l)}}\mathcal{E}_j\bigotimes_l \Phi_1^{(l)}\Theta_l\tilde{\sigma}\right] &= \prod_{l=1}^{L}\frac{1}{2}\left[1+\mathscr{C}(\Theta_l)\left|(\Lambda_l)_{00}^{01}\right|2\cos\left(2\pi 2^{L-l}\left(\frac{j}{r}-\frac{k'}{2^L}\right)\right)\right] \\ &\geq \inf_{|\chi|<\frac{1}{2q}}\prod_{l=1}^{L}\frac{1}{2}\left[1+\mathscr{C}(\Theta_l)\left|(\Lambda_l)_{00}^{01}\right|2\cos\left(2\pi 2^{L-l}\chi\right)\right], \end{aligned} \tag{67}$$

where in the last line we used our assumption that $k'\in\mathcal{K}_1^j$. Since $2|\Lambda_{00}^{01}|=\tilde{M}_\diamond(\Lambda)$, as detailed in Lem. 4, it follows that

$$\begin{aligned}\tilde{P}_j^{\text{succ}}(\Theta_l,\Lambda_l) &\geq \inf_{|\chi|<\frac{1}{2q}}\prod_{l=1}^{L}\frac{1}{2}\left[1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)\cos\left(2\pi 2^{L-l}\chi\right)\right] \\ &= \prod_{l=1}^{L}\frac{1}{2}\left[1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)\cos\left(\pi 2^{-l}\right)\right] \\ &\overset{\text{Lem. 12}}{\geq} \frac{4}{\pi^2}\prod_{l=1}^{L}\frac{1}{2}\left[1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)\right]. \end{aligned} \tag{68}$$

Now remember that up to here, we assumed that we replaced $\mathcal{E}$ with $\mathcal{E}_j$. This is of course not possible since it would require knowledge of r. To get back to the original protocol, we note that applying $\mathcal{E}$ corresponds to applying $\mathcal{E}_j$ with $j\in 0,...,r-1$ chosen uniformly at random. The number of such j with $\gcd(j,r)=1$ is given by $\varphi(r)$, where $\varphi(r)$ denotes Euler's totient function. The overall success probability is therefore bounded by

$$\begin{aligned}P^{\text{succ}}(\Theta_l,\Lambda_l) &= \frac{1}{r}\sum_{j=0}^{r-1}\tilde{P}_j^{\text{succ}}(\Theta_l,\Lambda_l) \\ &\geq \left(\frac{\varphi(r)}{r}\right)\frac{4}{\pi^2}\prod_{l=1}^{L}\frac{1}{2}\left[1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)\right]. \end{aligned} \tag{69}$$

Particularly, if the same channels are utilized in each block the bound simplifies to

$$P^{\text{succ}}(\Theta,\Lambda)\geq\frac{4}{\pi^2}\left(\frac{\varphi(r)}{r}\right)\left[\frac{1+\mathscr{C}(\Theta)\tilde{M}_\diamond(\Lambda)}{2}\right]^L. \tag{70}$$

□

Euler's totient function grows almost linearly in its argument and is strictly bounded by $\varphi(r)>\frac{\delta r}{\log\log r}>\frac{\delta r}{\log\log N}$ for some $\delta>0$, where $\delta\approx e^{-\gamma}$ with γ being the Euler-Mascheroni constant, see for instance Ref. [**?** , Theorem 328], which connects this bound to the original bound derived by Shor [**?** **?**]. For a perfectly coherent protocol, this bound would take the form $P^{\text{succ}}>\frac{4}{\pi^2}\frac{\delta}{\log\log r}$, which equals the bound originally obtained by Shor [**?**]. In the following works, see for example Refs. [**?** **?**], it has been shown that for the fully coherent protocol, the factor $\frac{4}{\pi^2}\approx 0.4$ can be pushed to about 0.9 (at least in an average case) by a more careful, yet tedious, analysis. The basic idea behind these proofs is to consider not only the set $\mathcal{K}_1$ as useful outcomes but to stretch the definition of said set as it has been outlined in Cor. 10. Since continuity in the dynamical measures $\mathscr{C}(\Theta)$ and $\tilde{M}_\diamond(\Lambda)$ is to be expected, it would not be surprising if the bound in Eq. (54) can be sharpened analogously. For now, we leave this to future work.

C. Classical limit

As already pointed out, the classical limit of the protocol uses only free channels $\Theta_{\text{free}}^{(l)}$ and $\Lambda_{\text{free}}^{(l)}$ and corresponds to a random number generator. It returns a number in the range $0\leq k\leq 2^L-1$ with a probability distribution of

$\{p_k(S_1^{(l)}[\Theta_{\text{free}}^{(l)}], S_2^{(l)}[\Lambda_{\text{free}}^{(l)}]; \tilde{\sigma}, \mathbb{P})\}_k$ independent of the order r since

$$\begin{aligned} p_k(S_1^{(l)}[\Theta_{\text{free}}^{(l)}], S_2^{(l)}[\Lambda_{\text{free}}^{(l)}]; \tilde{\sigma}, \mathbb{P}) &= \operatorname{Tr}\left[P_k \Delta F_{S_2^{(l)}[\Lambda_{\text{free}}^{(l)}]} \mathcal{E} \bigotimes_l S_1^{(l)}[\Theta_{\text{free}}^{(l)}](\tilde{\sigma})\right] \\ &= \operatorname{Tr}\left[P_k \Delta F_{S_2^{(l)}[\Lambda_{\text{free}}^{(l)}]} \Delta\mathcal{E} \bigotimes_l S_1^{(l)}[\Theta_{\text{free}}^{(l)}](\tilde{\sigma})\right] \\ &= \operatorname{Tr}\left[P_k \Delta F_{S_2^{(l)}[\Lambda_{\text{free}}^{(l)}]} \bigotimes_l S_1^{(l)}[\Theta_{\text{free}}^{(l)}](\tilde{\sigma})\right]. \end{aligned} \tag{71}$$

Without prior knowledge about the order r (including factors of r itself which may be obtained by considering the outcomes of multiple rounds combined; not considered here though), the on average most beneficial probability distribution p_k is the uniform distribution. For all free channels $\Theta_{\text{free}}^{(l)}, \Lambda_{\text{free}}^{(l)}$ we can always choose a pair $S_1^{(l)}, S_2^{(l)}$ that achieves this uniform distribution. In fact, such super-channels can even be chosen independently of $\Theta_{\text{free}}^{(l)}$ and $\Lambda_{\text{free}}^{(l)}$ in the classical limit, even in the case of non-unital $\Lambda_{\text{free}}^{(l)}$. A simple example would be to choose a suitable replacement channel as the post-processing of $\Lambda_{\text{free}}^{(l)}$. The resulting uniformly random measurement outcome is forwarded to the continued fraction algorithm producing an estimate on r. Thus, the overall success probability (assuming no prior knowledge of r) in the classical limit is given by

$$P^{\text{succ}}(\Theta_{\text{free}}^{(l)}, \Lambda_{\text{free}}^{(l)}) = \frac{f(N,r)}{2^L}, \tag{72}$$

where we define $f(N,r) = \sum_k P(k \to r \,|\, \text{CFA})$. The function $f(N,r)$ mitigates the exponential term in the success probability to some extend and to quantify this notion we proceed to derive bounds on this function.

Proposition 15. *The function $f(N,r)$ is bounded by*

$$2\varphi(r)\left\lfloor \frac{q-1}{2r^2} \right\rfloor \leq f(N,r) \leq \varphi(r)\left(1 + 2\left\lfloor \frac{q}{r^2} \right\rfloor\right), \tag{73}$$

where $\varphi(r)$ denotes Euler's totient function and q is uniquely given by $N^2 < q = 2^L < 2N^2$.

Proof. Consider a single coprime pair (j,r). Let $\tilde{f}_j(N,r)$ denote the function that counts the number of measurement outcomes that lead to this particular convergent j/r, i.e.,

$$\tilde{f}_j(N,r) = \sum_k P(k \to (j,r) \,|\, \text{CFA}). \tag{74}$$

For a lower bound, recall Thm. 9 and Cor. 10. Let us define the set $\mathcal{K}_1^j(\beta)$ which contains all integers that surely allow for a successful post-processing, i.e.,

$$\mathcal{K}_1^j(\beta) = \left\{k : 0 \leq k < q \wedge \left|\frac{j}{r} - \frac{k}{q}\right| \leq \frac{\beta}{2q}\right\}, \tag{75}$$

where $\beta = \frac{q-1}{r^2}$. A lower bound on $\tilde{f}_j(N,r)$ is then given by

$$\tilde{f}_j(N,r) = \sum_k P(k \to (j,r) \,|\, \text{CFA}) \geq \sum_{k \in \mathcal{K}_1^j(\beta)} P(k \to (j,r) \,|\, \text{CFA}) = \sum_{k \in \mathcal{K}_1^j(\beta)} 1 = |\mathcal{K}_1^j(\beta)|. \tag{76}$$

Furthermore, according to the second part of Thm. 9 all measurement outcomes that yield the pair (j,r) as a convergent of k/q are contained in the set $\mathcal{K}_2^j$, as introduced in Eq. (41). Thus the function can be upper bounded as

$$\tilde{f}_j(N,r) = \sum_k P(k \to (j,r) \,|\, \text{CFA}) = \sum_{k \in \mathcal{K}_2^j} P(k \to (j,r) \,|\, \text{CFA}) \leq \sum_{k \in \mathcal{K}_2^j} 1 = |\mathcal{K}_2^j|. \tag{77}$$

To further simplify these bounds, we proceed to bound the cardinalities of $\mathcal{K}_2^j$ and $\mathcal{K}_1^j(\beta)$. We start with a lower bound on $|\mathcal{K}_1^j(\beta)|$. Consider the closest fraction $\frac{k'}{q}$ defined by the smallest distance to the fraction $\frac{j}{r}$. This integer k' is roughly in the

center of the set defined by $\mathcal{K}_1^j(\beta)$, and also the set $\mathcal{K}_2^j$. Now consider the adjacent integers $k = k' \pm n$ to the closets integer k'. First assume $\frac{j}{r} - \frac{k'}{q} > 0$. Then for the elements to the left of k', i.e., $k = k' - n$ contained in $\mathcal{K}_1^j(\beta)$ we have

$$\frac{\beta}{2q} \geq \frac{j}{r} - \frac{k}{q} = \left(\frac{j}{r} - \frac{k'}{q}\right) + \frac{n}{q} > \frac{n}{q}, \tag{78}$$

and therefore

$$n < \tfrac{q-1}{2r^2}. \tag{79}$$

If $\frac{q-1}{2r^2}$ is an integer, all natural numbers $n \leq n_{\max} = \frac{q-1}{2r^2} - 1 =$ satisfy this equation and therefore lead to a k in $\mathcal{K}_1^j(\beta)$. Moreover, due to our closeness assumption of k', larger n cannot be in $\mathcal{K}_1^j(\beta)$. In this case, also all $k = k' + m$ with $m \leq m_{\max} = \frac{q-1}{2r^2}$ are contained in $\mathcal{K}_1^j(\beta)$, because we assumed $\frac{j}{r} - \frac{k'}{q} > 0$, i.e., there cannot be less integers $k > k'$ in $\mathcal{K}_1^j(\beta)$ than integers $k < k'$ and we cannot hit the boundary twice exactly. In sum, we find $|\mathcal{K}_1^j(\beta)| = 1 + n_{\max} + m_{\max} = 2\frac{q-1}{2r^2}$. If $\frac{q-1}{2r^2}$ is not an integer, we take $n_{\max} = \lfloor \frac{q-1}{2r^2} \rfloor$ instead, and $|\mathcal{K}_1^j(\beta)| = 1 + 2n_{\max} = 1 + 2\lfloor \frac{q-1}{2r^2} \rfloor$. Combining both cases, we have $|\mathcal{K}_1^j(\beta)| \geq 2\lfloor \frac{q-1}{2r^2} \rfloor$.

If $\frac{j}{r} - \frac{k'}{q} < 0$, the same bound holds true, which can be seen by switching the role of $n_{\max}$ and $m_{\max}$, i.e., switching the intervals to the left and right. Lastly, consider the case of $\frac{j}{r} - \frac{k'}{q} = 0$ and $k = k' \pm n$. From

$$\frac{\beta}{2q} \geq \left|\frac{j}{r} - \frac{k}{q}\right| = \left|\left(\frac{j}{r} - \frac{k'}{q}\right) + \frac{n}{q}\right| = \frac{|n|}{q}, \tag{80}$$

follows that all integers $k = k' + n$ with $|n| \leq \lfloor \frac{q-1}{2r^2} \rfloor$ are contained in $\mathcal{K}_1^j(\beta)$, i.e., $|\mathcal{K}_1^j(\beta)| \geq 1 + 2\lfloor \frac{q-1}{2r^2} \rfloor$. Combining all cases, we we find

$$|\mathcal{K}_1^j(\beta)| \geq \min\left\{2\lfloor \tfrac{q-1}{2r^2} \rfloor, 1 + 2\lfloor \tfrac{q-1}{2r^2} \rfloor\right\} = 2\lfloor \tfrac{q-1}{2r^2} \rfloor. \tag{81}$$

Now we continue to obtain an upper bound on the cardinality of $|\mathcal{K}_2^j|$. All integers $k \in \mathcal{K}_2^j$ obey $|\frac{j}{r} - \frac{k}{q}| \leq \frac{1}{r^2}$ by definition. Again consider the closest fraction $\frac{k'}{q}$ defined as the one with the smallest difference to the fraction $\frac{j}{r}$. As in the discussion for the lower bound, assume $\frac{j}{r} - \frac{k'}{q} > 0$. From the analogue of Eq. (78) follows that $k = k' - n$ is an element of $\mathcal{K}_2^j$ if $n < \frac{q}{r^2}$. If q/r^2 is an integer then $n_{\max} = \frac{q}{r^2} - 1$, and for the same arguments as before, $m_{\max} = \frac{q}{r^2}$. If q/r^2 is not an integer, $n_{\max} = \lfloor \frac{q}{r^2} \rfloor$. In addition, $m_{\max} = \lfloor \frac{q}{r^2} \rfloor$. Depending on $\frac{q}{r^2}$ being an integer or not, the cardinality is given by $|\mathcal{K}_2^j| = 1 + 2\lfloor \frac{q}{r^2} \rfloor$ or $|\mathcal{K}_2^j| = 2\lfloor \frac{q}{r^2} \rfloor$, thereby the cardinality is bounded by

$$|\mathcal{K}_2^j| \leq 1 + 2\lfloor \tfrac{q}{r^2} \rfloor. \tag{82}$$

In the remaining case, i.e., if $\frac{j}{r} - \frac{k}{q} = 0$, we find $|\mathcal{K}_2^j| = 1 + 2\lfloor \frac{q}{r^2} \rfloor$, whether or not $\frac{q}{r^2}$ is an integer.

To conclude the proof, take into account all possible integers j that are smaller then and coprime to r. There are exactly $\varphi(r)$ such values for j, and correspondingly for each such j, there is a range of possible outcomes k that lead to the respective pair (j, r). Inserting the expressions Eqs. (81) and (82) into Eqs. (75) and (76) respectively, we see that the function $\tilde{f}_j(n, r)$ can be bounded by

$$f(N,r) = \sum_k P(k \to r \,|\, \text{CFA}) = \sum_{j \text{ coprime to } r} \tilde{f}_j(N,r) \geq \sum_{j \text{ coprime to } r} 2\left\lfloor \frac{q-1}{2r^2} \right\rfloor = 2\varphi(r)\left\lfloor \frac{q-1}{2r^2} \right\rfloor,$$

and

$$f(N,r) = \sum_k P(k \to r \,|\, \text{CFA}) = \sum_{j \text{ coprime to } r} \tilde{f}_j(N,r) \leq \sum_{j \text{ coprime to } r} \left(1 + 2\left\lfloor \frac{q}{r^2} \right\rfloor\right) = \varphi(r)\left(1 + 2\left\lfloor \frac{q}{r^2} \right\rfloor\right).$$

Note that for all (j, r) with j not coprime to r, the continued fraction algorithm will yield a factor of r. Whilst this information can be used in principle, it is not relevant for the fixed post-processing strategy that we chose. □

Let us conclude this section by noting that with the result of Prop. 15 and Eq. (72), we can provide bounds on the classical limit of the success probability, i.e.,

$$2\frac{\varphi(r)}{2^L}\left\lfloor \frac{2^L - 1}{2r^2} \right\rfloor \leq P^{\text{succ}}(\Theta_{\text{free}}^{(l)}, \Lambda_{\text{free}}^{(l)}) \leq \frac{\varphi(r)}{2^L}\left(1 + 2\left\lfloor \frac{2^L}{r^2} \right\rfloor\right), \tag{83}$$

where we used $q = 2^L$. The classical limit of the success probability is thus sensibly dependent on the ratio between 2^L and r^2. Since $N^2 < 2^L < 2N^2$, this is a purely problem specific expression, in the sense that it only depends on the number N to factor and a corresponding order r.

VI. An upper bound

From a complexity theoretic perspective, providing an upper bound on the success probability is rather uninteresting, since it corresponds to a best-case scenario. On the other hand, an upper bound is an interesting question, if we want to attribute a potential speed-up to a resource, i.e., in our case coherence. For this reason we use a similar technique as in the classical limit to provide a sufficiently general upper bound on the performance of the protocol. Nevertheless, the bound is general enough to provide quantitative insights on the role of coherence in the algorithm.

Theorem 16. *The success probability of the order-finding protocol with qubit channels Θ_l and unital Λ_l is bounded by*

$$P^{succ}(\Theta_l,\Lambda_l) \leq \min\left\{\frac{\varphi(r)}{2^L}\left(1+2\lfloor\tfrac{2^L}{r^2}\rfloor\right)\prod_{l=1}^{L}\left(1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)\right),\,1\right\}, \tag{84}$$

where $\mathscr{C}$ denotes the cohering power with respect to the robustness of coherence, $\tilde{M}_\diamond$ is the NSID-measure, and $\varphi(r)$ is Euler's totient function.

Proof. Again, consider an idealized protocol with only a single rotation $\mathcal{E}_j$ first. Recall the notations introduced around Eq. (42). We now need to be more careful than in the lower bound and use a similar technique as in the classical limit. From the quantum part of the protocol we obtain a measurement outcome, i.e., an integer k, with a probability depending on the rotation $\mathcal{E}_j$. The classical post-processing succeeds by definition if it returns a coprime pair (j',r), where r is the order we are looking for. Even if a measurement outcome k does not lead to j/r, it could still be close enough to another coprime fraction j'/r such that the post-processing succeeds. This means we have to account for all possible coprime integers j' and hence their corresponding integers k that allow to estimate j'/r. The success probability is given by

$$\begin{aligned}\tilde{P}_j^{\text{succ}}(\Theta_l,\Lambda_l) &= \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}}\max_{\tilde{\sigma}\in\mathcal{I}}\sum_k P(k\to r\,|\,\text{CFA})p_k^{(j)}(S_1^{(l)}[\Theta_l],S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P})\\ &= \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}}\max_{\tilde{\sigma}\in\mathcal{I}}\sum_{j'\perp r}\sum_k P(k\to(j',r)\,|\,\text{CFA})p_k^{(j)}(S_1^{(l)}[\Theta_l],S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P}),\end{aligned} \tag{85}$$

where in the second line we used that we only consider direct estimates of r but not factors of r, thus we sum only over all coprime j'. Recall the necessary condition that integers k leading to j'/r are contained in $\mathcal{K}_2^{j'}$, hence

$$\begin{aligned}\tilde{P}_j^{\text{succ}}(\Theta_l,\Lambda_l) &= \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}}\max_{\tilde{\sigma}\in\mathcal{I}}\sum_{j'\perp r}\sum_k P(k\to(j',r)\,|\,\text{CFA})p_k^{(j)}(S_1^{(l)}[\Theta_l],S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P})\\ &\leq \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}}\max_{\tilde{\sigma}\in\mathcal{I}}\sum_{j'\perp r}\sum_{k\in\mathcal{K}_2^{j'}} p_k^{(j)}(S_1^{(l)}[\Theta_l],S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P})\\ &\leq \sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}}\max_{\tilde{\sigma}\in\mathcal{I}}\sum_{j'\perp r}\max_{j'\perp r}\sum_{k\in\mathcal{K}_2^{j'}} p_k^{(j)}(S_1^{(l)}[\Theta_l],S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P})\\ &= \varphi(r)\sup_{\substack{S_1^{(l)}\in\mathcal{MIOS}\\ S_2^{(l)}\in\mathcal{DIS}}}\max_{\tilde{\sigma}\in\mathcal{I}}\max_{j'\perp r}\sum_{k\in\mathcal{K}_2^{j'}} p_k^{(j)}(S_1^{(l)}[\Theta_l],S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P}).\end{aligned} \tag{86}$$

Recall from the proof of Thm. 14 that

$$p_k^{(j)}(S_1^{(l)}[\Theta_l],S_2^{(l)}[\Lambda_l];\tilde{\sigma},\mathbb{P}) = \prod_{l=1}^{L}\text{Tr}\left[P_k^{(l)}\Delta S_2^{(l)}[\Lambda_l]R_l'\mathcal{E}_j^{(l)}S_1^{(l)}[\Theta_l]\sigma_l\right]. \tag{87}$$

As in the proof for the lower bound, we derive an expression for the measurement statistics for arbitrary super-channels. We can write

$$\begin{aligned}\Delta S_2^{(l)}[\Lambda_l]R_l'\mathcal{E}_j^{(l)}S_1^{(l)}[\Theta_l]\sigma_l &= \Delta S_2^{(l)}[\Lambda_l]\Delta R_l'\mathcal{E}_j^{(l)}S_1^{(l)}[\Theta_l]\sigma_l + \Delta S_2^{(l)}[\Lambda_l](\mathbb{1}-\Delta)R_l'\mathcal{E}_j^{(l)}S_1^{(l)}[\Theta_l]\sigma_l\\ &= \Delta S_2^{(l)}[\Lambda_l]\Delta S_1^{(l)}[\Theta_l]\sigma_l + \Delta S_2^{(l)}[\Lambda_l](\mathbb{1}-\Delta)R_l'\mathcal{E}_j^{(l)}S_1^{(l)}[\Theta_l]\sigma_l\\ &= \Delta S_2^{(l)}[\Lambda_l]\Delta S_1^{(l)}[\Theta_l]\sigma_l + \Delta S_2^{(l)}[\Lambda_l]R_l'\mathcal{E}_j^{(l)}(\mathbb{1}-\Delta)S_1^{(l)}[\Theta_l]\sigma_l,\end{aligned} \tag{88}$$

where in the last line we used that $\mathbb{1}-\Delta$ commutes with the rotations. Let us focus on the second term first. We define

$$\begin{aligned}\left[S_1^{(l)}[\Theta_l](\sigma_l)\right]_{01} &= \left|\left[S_1^{(l)}[\Theta_l](\sigma_l)\right]_{01}\right| e^{i\phi_1} = c_l e^{i\phi_1},\\ (S_2^{(l)}[\Lambda_l])^{01}_{b_l,b_l} &= |(S_2^{(l)}[\Lambda_l])^{01}_{b_l,b_l}| e^{i\lambda^{01}_{b_l,b_l}}.\end{aligned} \tag{89}$$

Then we can write the second term as

$$\begin{aligned}&\Delta S_2^{(l)}[\Lambda_l] R_l' \mathcal{E}_j^{(l)}(\mathbb{1}-\Delta) S_1^{(l)}[\Theta_l]\sigma_l\\ &= \Delta S_2^{(l)}[\Lambda_l] R_l' \left(c_l e^{-2\pi i \frac{j}{r} 2^{L-l}} e^{i\phi_1}|0\rangle\langle 1| + h.c.\right)\\ &= \Delta S_2^{(l)}[\Lambda_l] \left(c_l e^{-2\pi i\left(\frac{j}{r}2^{L-l} - \sum_{a=2}^{l} k'_{l-a}/2^a\right)} e^{i\phi_1}|0\rangle\langle 1| + h.c.\right)\\ &= c_l \left(e^{-2\pi i\left(\frac{j}{r}2^{L-l} - \sum_{a=2}^{l} k'_{l-a}/2^a\right)} e^{i\phi_1} \sum_{b_l=0}^{1} \left|(S_2^{(l)}[\Lambda_l])^{01}_{b_l b_l}\right| e^{i\lambda^{01}_{b_l,b_l}} |b_l\rangle\langle b_l| + h.c.\right)\\ &= \left|\left[S_1^{(l)}[\Theta_l](\sigma_l)\right]_{01}\right| \left|(S_2^{(l)}[\Lambda_l])^{01}_{00}\right| \left(e^{-2\pi i\left(\frac{j}{r}2^{L-l} - \sum_{a=2}^{l} k'_{l-a}/2^a\right)} e^{i\phi_1} \sum_{b_l=0}^{1} e^{i\lambda^{01}_{b_l,b_l}} |b_l\rangle\langle b_l| + h.c.\right).\end{aligned} \tag{90}$$

where we used in the last line that $\left|(S_2^{(l)}[\Lambda_l])^{01}_{00}\right| = \left|(S_2^{(l)}[\Lambda_l])^{01}_{11}\right|$ [**?** , Prop. 6]. Let us introduce the abbreviation $q_{k_l}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l]; \sigma_l, \mathbb{P}) = \mathrm{Tr}\left[P_k^{(l)} \Delta S_2^{(l)}[\Lambda_l] \Delta S_1^{(l)}[\Theta_l]\sigma_l\right]$. We evaluate the projective measurement and bound the phase dependent terms by two to find

$$\begin{aligned}&\tilde{p}_k^{(j)}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l]; \tilde{\sigma}, \mathbb{P})\\ &\leq \prod_{l=1}^{L} \left(q_{k_l}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l]; \sigma_l, \mathbb{P}) + 2\left|\left[S_1^{(l)}[\Theta_l](\sigma_l)\right]_{01}\right| \left|(S_2^{(l)}[\Lambda_l])^{01}_{00}\right|\right)\\ &= \prod_{l=1}^{L} \left(q_{k_l}(S_1[\Theta], S_2[\Lambda]; \sigma_l, \mathbb{P}) + \tfrac{1}{2} C(S_1^{(l)}[\Theta_l](\sigma_l)) \tilde{M}_\diamond(S_2^{(l)}[\Lambda])\right),\end{aligned} \tag{91}$$

where we used the expressions for the measures from Eq. (64) and Lem. 4. Using this bound on the probability to measure an outcome k results in a bound on the success probability given by

$$\begin{aligned}\tilde{P}_j^{\text{succ}}(\Theta_l, \Lambda_l) \leq \varphi(r) \sup_{\substack{S_1 \in \mathcal{MIOS}\\ S_2 \in \mathcal{DIS}}} \max_{\tilde{\sigma}\in\mathcal{I}} \max_{j' \perp r}\\ \sum_{k\in\mathcal{K}_2^{j'}} \prod_{l=1}^{L} \left(q_{k_l}(S_1[\Theta], S_2[\Lambda]; \sigma_l, \mathbb{P}) + \tfrac{1}{2} C(S_1^{(l)}[\Theta_l](\sigma_l)) \tilde{M}_\diamond(S_2^{(l)}[\Lambda])\right).\end{aligned} \tag{92}$$

Only the contribution $q_{k_l}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l]; \sigma_l, \mathbb{P})$ emerging from the distribution of the initial populations depends implicitly on the integer j', in the sense that the integer j' determines which measurement outcomes k are contained in the respective set $\mathcal{K}_2^{j'}$. We assume no prior knowledge about r and therefore no knowledge about the fraction $\frac{j'}{r}$. Therefore, the integers $k \in \mathcal{K}_2^{j'}$, cannot be known prior to the experiment and we have to assume that the interval of integers defined by $\mathcal{K}_2^{j'}$ is distributed uniformly across the range $0 \leq k' < q$. Hence, the optimal initial population distribution is uniform, i.e., $q_{k_l}(S_1^{(l)}[\Theta_l], S_2^{(l)}[\Lambda_l]; \sigma_l, \mathbb{P}) = \frac{1}{2}\ \forall k_l$. Recall that we assume that all Λ_l are unital and all $S_2^{(l)}$ are unitality-preserving. Thus, we can construct super-operations $S_1^{(l)}, S_2^{(l)}$ that can achieves this uniform distribution without influencing the second term involving the measures, see for example

the super-operations utilized in the proof of the lower bound. Therefore, we find the upper bound

$$\begin{aligned}
\tilde{P}_j^{\text{succ}}(\Theta_l,\Lambda_l) &\leq \varphi(r) \sup_{\substack{S_1\in\mathcal{MIOS}\\ S_2\in\mathcal{DIS}}} \max_{\tilde{\sigma}\in\mathcal{I}} \max_{j'\perp r} \sum_{k\in\mathcal{K}_2^{j'}} \frac{1}{2^L}\prod_{l=1}^{L}\left(1+C(S_1^{(l)}[\Theta_l](\sigma_l))\tilde{M}_\diamond(S_2^{(l)}[\Lambda_l])\right)\\
&= \frac{\varphi(r)}{2^L}\prod_{l=1}^{L}\left(1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)\right)\max_{j'\perp r}\sum_{k\in\mathcal{K}_2^{j'}} 1\\
&= \frac{\varphi(r)}{2^L}\prod_{l=1}^{L}\left(1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)\right)\max_{j'\perp r}|\mathcal{K}_2^{j'}|\\
&\leq \frac{\varphi(r)}{2^L}\prod_{l=1}^{L}\left(1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)\right)\left(1+2\lfloor\tfrac{2^L}{r^2}\rfloor\right),
\end{aligned} \tag{93}$$

where we used the results from the proof concerning the classical limit in the last line. So far we only used a single rotation. Note that since this bound holds for an arbitrary rotation $\mathcal{E}_j$ and since it is independent of the j labeling a rotation $\mathcal{E}_j$, which is applied probabilistically, we find

$$\begin{aligned}
P^{\text{succ}}(\Theta_l,\Lambda_l) &= \frac{1}{r}\sum_{j=0}^{r-1}\tilde{P}_j^{\text{succ}}(\Theta_l,\Lambda_l)\\
&\leq \frac{1}{r}\sum_{j=0}^{r-1}\frac{\varphi(r)}{2^L}\prod_{l=1}^{L}\left(1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)\right)\left(1+2\lfloor\tfrac{2^L}{r^2}\rfloor\right)\\
&= \frac{\varphi(r)}{2^L}\prod_{l=1}^{L}\left(1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)\right)\left(1+2\lfloor\tfrac{2^L}{r^2}\rfloor\right).
\end{aligned} \tag{94}$$

Lastly, note that this bound can exceed unit probability and for this reason, we decide to formulate a bound of the form

$$P^{\text{succ}}(\Theta_l,\Lambda_l) \leq \min\left\{\frac{\varphi(r)}{2^L}\left(1+2\lfloor\tfrac{2^L}{r^2}\rfloor\right)\prod_{l=1}^{L}\left(1+\mathscr{C}(\Theta_l)\tilde{M}_\diamond(\Lambda_l)\right),\,1\right\}, \tag{95}$$

which concludes the proof. □

Again, note that for identical operations in each block we obtain the important special case of

$$P^{\text{succ}}(\Theta_l,\Lambda_l) \leq \min\left\{\frac{\varphi(r)}{2^L}\left(1+2\lfloor\tfrac{2^L}{r^2}\rfloor\right)\left(1+\mathscr{C}(\Theta)\tilde{M}_\diamond(\Lambda)\right)^L,\,1\right\}. \tag{96}$$

If the bound involving the resources measures exceeds unity, the upper bound reduces to a trivial bound. However, we emphasize that this is not only a trivial bound on the success probability. Most importantly, the expression exceeds unity if the prefactor becomes large. Comparing it with the classical success probability, we see that the prefactors are the same. This leads us to the conclusion that our bound is relevant whenever the order-finding problem is hard in the classical limit. Then the bound involving the dynamical resource measures is indeed useful. In that sense, we can argue that coherence is the resource that provides an advantage whenever there is an actual advantage to grant.

VII. Visualization

To conclude the Supplementary Material, we provide two plots that visualize the results derived in the previous sections. First, we note that solving the factorization problem given by the pair (L, r) using the classical limit of the quantum order-finding protocol becomes less efficient if the order r increases. To visualize this, see Fig. 7, where the upper bound on the classical performance is depicted (the lower bound behaves similarly). Note that for small orders r of order $\mathcal{O}(1)$ the classical protocol can perform reasonably well, which may be intuitively understood by the fact that an efficient classical search for r is feasible in this regime. For increasing orders r, the classical performance diminishes. The factorization problems (L, r) with large order r defy the classical limit of the order-finding protocol, rendering it inefficient. It is this regime where coherence will allow for a super-polynomial speed-up in L over the classical limit.

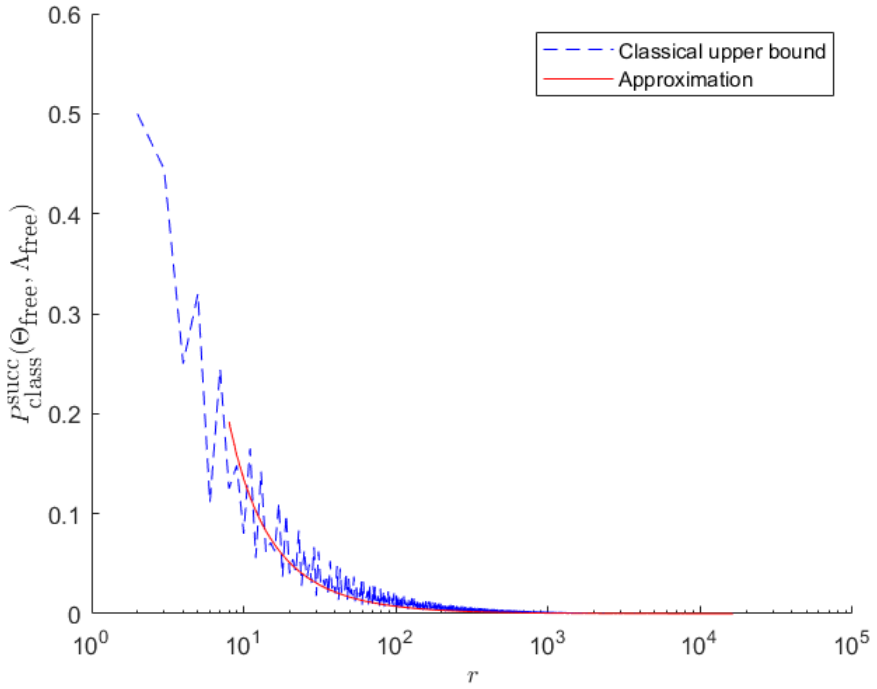

Figure 7: Behavior of the classical upper bound $P^{\text{succ}}_{\text{class}}(\Theta_{\text{free}}, \Lambda_{\text{free}}) \leq \frac{\varphi(r)}{2^L}\left(1 + 2\left\lfloor \frac{2^L}{r^2} \right\rfloor\right)$ in terms of the order r for a given $L = 30$. Furthermore, an approximation of the upper bound based on $\varphi(r) \approx \frac{r}{e^\gamma \log\log r}$.

To better visualize the bounds on the performance of the order-finding algorithm, let us consider specific resourceful channels Θ and Λ (used for the creation and detection of coherence). First, take the ideal Hadamard channels mixed with dephasing noise during the creation and detection of coherence, which are given by

$$\Theta_p(\rho) = pH\rho H + (1-p)\Delta(\rho), \tag{97}$$
$$\Lambda_q(\rho) = qH\rho H + (1-q)\Delta(\rho). \tag{98}$$

The parameters p and q interpolate between an optimal channel for creation and detection and a completely free channel, which simply erases all coherence. In Fig. 8 we compare our lower and upper bound on the performance of the classical limit (see Prop. 15) with the bounds of Thms. 14 and 16 given identical operations Θ_p and Λ_q on each qubit and $p = q$. It is straightforward to see that the measures evaluate to $\mathscr{C}(\Theta_p) = \tilde{M}_\diamond(\Lambda_p) = p$. Note that whilst the upper bound on the performance using resourceful operations (dashed blue line) also provides an upper bound on the best classical performance given any $p > 0$, the lower bound (solid blue line) can drop below the classical limit of the success probability. This is due to the fact that in the proof of Thm. 14, we neglected terms that give a non-vanishing contribution which we accounted for in the classical limit. Here, it is important to note that for values of probabilities to apply the correct quantum operations p that exceed approximately $p = 0.65$ an exponential speed-up (note the logarithmic scale) over the classical limit is guaranteed to be achieved, which demonstrates the quantitative role of coherence utilized by the specific operations Θ_p and Λ_p.

Moreover, let us introduce the partial dephasing map Δ_p defined by $\Delta_p(\rho) = p\,\mathrm{id} + (1-p)\Delta$. Then consider the channel

$$\Phi_{p,q}(\rho) = \Delta_p \mathcal{H} \Delta_q(\rho), \tag{99}$$

where $\mathcal{H}$ denotes the channel associated to the Hadamard gate, i.e., $\mathcal{H}(\rho) = H\rho H$. This channel corresponds to an optimal implementation of the Hadamard gate with partial dephasing noise prior and after its application. For the channel $\Phi_{p,q}$, the measures reduce to $\mathscr{C}(\Phi_{p,q}) = p$ and $\tilde{M}_\diamond(\Phi_{p,q}) = q$, which is intuitive since one of the partial dephasings does not affect creation respectively detection. In particular, for a symmetric channel $\Phi_{p,q}$, i.e., $p = q$, the measures, and thus the bounds, reduce to the mixing with a total dephasing map. The performance of the order-finding protocol with $\Phi_{p,p}$ is therefore also depicted in Fig. 8.

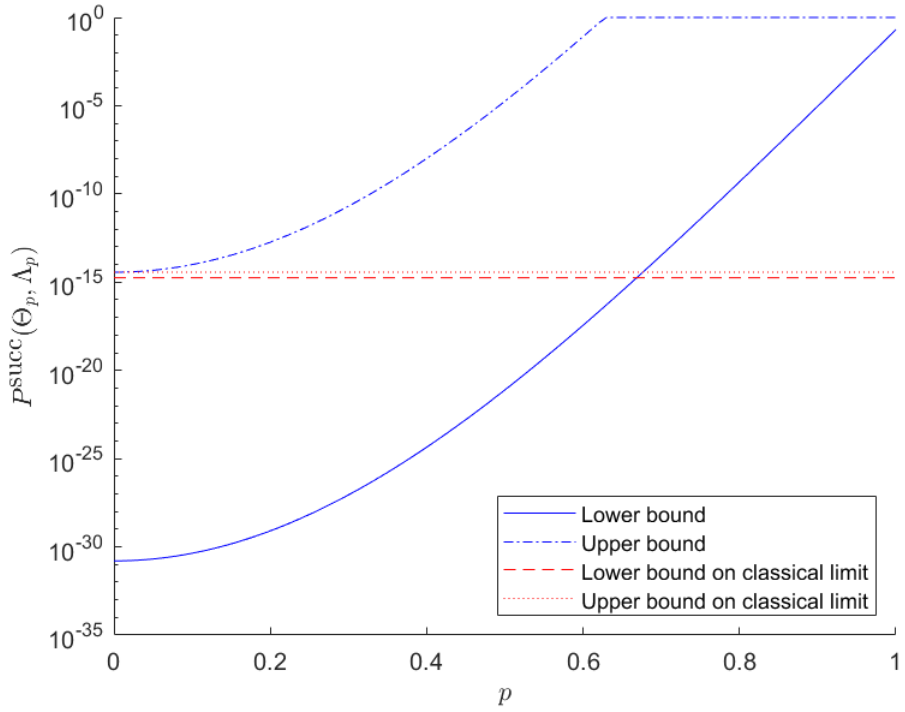

Figure 8: Comparison between the success probability with ideal operations subject to dephasing noise and the classical limit, where p denotes the probability to apply the ideal Hadamard gate for creation and detection. Here we picked a typical factorization problem (L, r) with $L = 100$ and $r = \frac{\sqrt{2^L}}{4} \sim N/4$.