# Grover search and the no-signaling principle
# Supplemental Material

Ning Bao
*Institute for Quantum Information and Matter and*
*Walter Burke Institute for Theoretical Physics,*
*California Institute of Technology 452-48, Pasadena, CA 91125, USA*

Adam Bouland
*Computer Science and Artificial Intelligence Laboratory,*
*Massachusetts Institute of Technology, Cambridge, MA*

Stephen P. Jordan
*National Institute of Standards and Technology, Gaithersburg, MD and*
*Joint Center for Quantum Information and Computer Science, University of Maryland, College Park, MD*

PACS numbers: 03.67.-a, 03.65.-w, 04.70.Dy

In this supplemental material, we provide more detailed proofs of the claims of our paper.

## SECTION A: FINAL-STATE PROJECTION

Recent developments, particularly the AMPS firewall argument [1], have generated renewed interest in models of black hole physics in which quantum mechanics is modified. Here, we explore some difficulties associated one such scheme, namely the Horowitz-Maldecena final state projection model [2]. In this model, black hole singularities are thought of as boundaries to spacetime with associated boundary conditions on the quantum wavefunction [2]. That is, at the singularity, the wavefunction becomes projected onto a specific quantum state. (This can be thought of as a projective measurement with postselection.)

If one prepares infalling matter in a chosen initial quantum state $|\psi\rangle \in V$, allows it to collapse into a black hole, and then collects all of the the Hawking radiation during the black hole evaporation, one is left with a new quantum state related to the original by some map $S : V \to V$. (We assume that black holes do not alter the dimension of the Hilbert space. Standard quantum mechanics and the Horowitz-Maldecena proposal share this feature.) Within standard quantum mechanics, all such $S$ correspond to multiplication by a unitary matrix, and hence the term $S$-matrix is used. If one instead drops matter into an existing black hole and collects part of the outgoing Hawking radiation, one is considering an open quantum system. We leave the analysis of this more general scenario to future work.

It is possible for the Horowitz-Maldecena final state projection model to induce a perfectly unitary process $S$ for the black hole. However, as pointed out in [3], interactions between the collapsing body and infalling Hawking radiation inside the event horizon generically induce deviations from unitarity. In this case, the action $S$ of the black hole is obtained by applying some linear but not unitary map $M$, and then readjusting the norm of the quantum state back to one[4]. Correspondingly, if a subsystem of an entangled state is collapsed into a black hole and the Hawking radiation is collected then the corresponding transformation is $M \otimes \mathbb{1}$ followed by an adjustment of the normalization back to 1. Thus, aside from its interest as a potential model for black holes, the Horowitz-Maldecena model provides an interesting example of nonlinear quantum mechanics in which subsystem structure remains well-defined (i.e. the issues described in Section A do not arise).

In sections A 1 and A 2 we show that if Alice has access to such a black hole and has foresightfully shared entangled states with Bob, then Alice can send instantaneous noisy signals to Bob and vice-versa independent of their spatial separation. We quantify the classical information-carrying capacity of the communication channels between Alice and Bob and find that they vanish only quadratically with the deviation from unitarity of the black hole dynamics, as measured by the deviation of the condition number of $M$ from one. Hence, unless the deviation from unitarity is negligibly small, detectable causality violations can infect the entirety of spacetime. Furthermore, the bidirectional nature of the communication makes it possible in principle for Alice to send signals into her own past lightcone, thereby generating grandfather paradoxes.

In section A 3 we consider the use of the black hole dynamical map $S$ to speed up Grover's search algorithm [5]. We find a lower bound on the condition number of $M$ as a function of the beyond-Grover speedup. By our results of sections A 1 and A 2 this in turn implies a lower bound on the superluminal signaling capacity induced by the black hole. In section A 4 we prove the other direction: assuming one can signal superluminally we derive a lower bound on the condition number of $M$, which in turn implies a super-Grover speedup[6]. We find that the black-box solution of NP-hard problems in polynomial time implies superluminal signaling with inverse polynomial capacity and vice versa.

### 1. Communication from Alice to Bob

**Theorem 1.** *Suppose Alice has access to a black hole described by the Horowitz-Maldecena final state projection model. Let $M$ be the linear but not necessarily unitary map describing the dynamics of the black hole. The non-unitarity of $M$ is quantified by $\delta = 1 - \kappa$, the deviation of its condition number from one. Alice can transmit instantaneous signals to Bob by choosing to drop her half of a shared entangled state into the black hole or not. The capacity of the resulting classical communication channel from Alice to Bob is at least*

$$C \geq \frac{3}{8 \ln 2} \delta^2.$$

*Proof.* We prove the lower bound on the channel capacity $C$ by exhibiting an explicit protocol realizing it. Suppose the black hole acts on a $d$-dimensional Hilbert space and correspondingly $M$ is a $d \times d$ matrix. Then, $M$ has a singular-value decomposition given by

$$M = \sum_{i=0}^{d-1} |\psi_i\rangle \lambda_i \langle \phi_i| \tag{A1}$$

with

$$\langle \psi_i | \psi_j \rangle = \langle \phi_i | \phi_j \rangle = \delta_{ij}. \tag{A2}$$

and $\lambda_0, \ldots, \lambda_{d-1}$ all real and nonnegative. We can choose our indexing so that $\lambda_0$ is the smallest singular value and $\lambda_1$ is largest singular value. Now, suppose Alice and Bob share the state

$$\frac{1}{\sqrt{2}} \left( |\phi_0\rangle |0\rangle + |\phi_1\rangle |1\rangle \right). \tag{A3}$$

Here $|0\rangle$ and $|1\rangle$ refer to Bob's half of the entangled state, which can be taken to be a qubit. If Alice wishes to transmit the message "0" to Bob she does nothing. If she wishes to transmit the message "1" to Bob she applies the black hole dynamical map $S$ to her half of the state. In other words, Alice drops her half of the state into the black hole, and waits for the black hole to evaporate. Correspondingly, one applies $M \otimes \mathbb{1}$ to the above state, yielding the unnormalized state

$$\frac{\lambda_0}{\sqrt{2}} |\psi_0\rangle |0\rangle + \frac{\lambda_1}{\sqrt{2}} |\psi_1\rangle |1\rangle. \tag{A4}$$

After normalization, this becomes:

$$\frac{\lambda_0}{\sqrt{\lambda_0^2 + \lambda_1^2}} |\psi_0\rangle |0\rangle + \frac{\lambda_1}{\sqrt{\lambda_0^2 + \lambda_1^2}} |\psi_1\rangle |1\rangle. \tag{A5}$$

Thus, recalling (A2), Bob's reduced density matrix in this case is

$$\rho_1 = \frac{\lambda_0^2}{\lambda_0^2 + \lambda_1^2} |0\rangle\langle 0| + \frac{\lambda_1^2}{\lambda_0^2 + \lambda_1^2} |1\rangle\langle 1|, \tag{A6}$$

whereas in the case that Alice's message was "0" his reduced density matrix is

$$\rho_0 = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|. \tag{A7}$$

If $M$ is non-unitary then $\lambda_1 \neq \lambda_0$ and thus the trace distance between these density matrices is nonzero. Consequently, $\rho_1$ is distinguishable from $\rho_0$ and some fraction of a bit of classical information has been transmitted to Bob.

More quantitatively, one sees that Bob's optimal measurement is in the computational basis, in which case Alice and Bob are communicating over a classical binary asymmetric channel. Specifically, if Alice transmits a 0, the probability of bit-flip error is $\epsilon_0 = 1/2$ whereas if Alice transmits a 1, the probability of bit-flip error is

$$\epsilon_1 = \frac{\lambda_0^2}{\lambda_0^2 + \lambda_1^2}. \tag{A8}$$

A standard calculation (see *e.g* [7]) shows that the classical capacity of this channel is

$$C = h\left(\frac{1}{1+z}\right) - \frac{\log_2(z)}{1+z} + \epsilon_0 \log_2(z) - h(\epsilon_0), \tag{A9}$$

where

$$z = 2^{\frac{h(\epsilon_1) - h(\epsilon_0)}{1-\epsilon_1-\epsilon_0}} \tag{A10}$$

and $h$ is the binary entropy

$$h(p) = -p\log_2(p) - (1-p)\log_2(1-p). \tag{A11}$$

Specializing to $\epsilon_0 = \frac{1}{2}$ simplifies the expression to

$$C = h\left(\frac{1}{1+y}\right) - \frac{\log_2(y)}{1+y} + \frac{1}{2}\log_2(y) - 1 \tag{A12}$$

where

$$y = 2^{\frac{h(\epsilon_1)-1}{1/2-\epsilon_1}}. \tag{A13}$$

Lastly, we consider the limiting case $\epsilon_1 = \frac{1}{2} - \Delta$ for $\Delta \ll 1$. In this limit, we get by Taylor expansion that

$$C = \frac{3}{2\ln 2}\Delta^2 + O(\delta^3). \tag{A14}$$

By (A8), $\Delta = \frac{1}{2}(1-\kappa) + O((1-\kappa)^2)$, which completes the proof. $\square$

### 2. Communication from Bob to Alice

**Theorem 2.** *Suppose Alice has access to a black hole described by the Horowitz-Maldecena final state projection model. Let $M$ be the linear but not necessarily unitary map describing the dynamics of the black hole. The non-unitarity of $M$ is quantified by $\delta = 1 - \kappa$, the deviation of its condition number from one. Bob can transmit instantaneous signals to Alice by choosing to measure his half of a shared entangled state or not. The capacity of the resulting classical communication channel from Bob to Alice is at least*

$$C \geq \frac{3}{8\ln 2}\delta^2.$$

*Proof.* Suppose again that Alice and Bob share the state $\frac{1}{\sqrt{2}}(|\phi_0\rangle|0\rangle + |\phi_1\rangle|1\rangle)$. If Bob wishes to transmit the message "0" he does nothing, whereas if he wishes to transmit the message "1" he measures his half of the entangled state in the $\{|0\rangle, |1\rangle\}$ basis. Then, Alice applies the black hole dynamical map $S$ to her half of the state[8], and then performs a projective measurement in the basis $\{|\psi_1\rangle, \ldots, |\psi_d\rangle\}$. We now show that this procedure transmits a nonzero amount of classical information from Bob to Alice unless $\lambda_0 = \lambda_1$, in which case $M$ is unitary.

In the case that Bob does nothing, the post-black hole state is again

$$\frac{\lambda_0}{\sqrt{\lambda_0^2 + \lambda_1^2}}|\psi_0\rangle|0\rangle + \frac{\lambda_1}{\sqrt{\lambda_0^2 + \lambda_1^2}}|\psi_1\rangle|1\rangle. \tag{A15}$$

Thus, Alice's post-black-hole reduced density matrix is

$$\frac{\lambda_0^2}{\lambda_0^2 + \lambda_1^2}|\psi_0\rangle\langle\psi_0| + \frac{\lambda_1^2}{\lambda_0^2 + \lambda_1^2}|\psi_1\rangle\langle\psi_1|. \tag{A16}$$

Alice's measurement will consequently yield the following probability distribution, given that Bob's message was "0":

$$p(0|0) = \frac{\lambda_0^2}{\lambda_0^2 + \lambda_1^2} \tag{A17}$$

$$p(1|0) = \frac{\lambda_1^2}{\lambda_0^2 + \lambda_1^2}. \tag{A18}$$

Now, suppose Bob's message is "1". Then, his measurement outcome will be either $|0\rangle$ or $|1\rangle$ with equal probability. We must analyze these cases separately, since the connection between ensembles of quantum states and density matrices is not preserved under nonlinear transformations[9]. If he gets outcome zero, then Alice holds the pure state $|\phi_0\rangle$, which gets transformed to $|\psi_0\rangle$ by the action of the black hole. If Bob gets outcome one, then Alice holds $|\phi_1\rangle$, which gets transformed to $|\psi_1\rangle$ by the action of the black hole. Hence, Alice's measurement samples from the following distribution given that Bob's message was "1":

$$p(0|1) = 1/2 \tag{A19}$$

$$p(1|1) = 1/2. \tag{A20}$$

Hence, the information transmission capacity from Bob to Alice using this protocol is the same as the Alice-to-Bob capacity calculated in section A 1. □

### 3. Super-Grover Speedup implies Superluminal Signaling

**Theorem 3.** *Suppose one has access to one or more black holes described by the Horowitz-Maldecena final state projection model. If the non-unitary dynamics induced by the black hole(s) allow the solution of a Grover search problem on a database of size $N$ using $q$ queries then the same non-unitary dynamics could be used transmit instantaneous signals by applying them to half of an entangled state. The capacity of the resulting classical communication channel (bits communicated per use of the nonlinear dynamics) is at least*

$$C = \Omega\left(\left(\frac{\eta}{2q^2} - \frac{2}{N}\right)^2\right)$$

*in the regime $0 < \frac{\eta}{2q^2} - \frac{2}{N} \ll 1$, where $\eta = (\sqrt{2} - \sqrt{2 - \sqrt{2}})^2 \simeq 0.42$.*

*Proof.* Let $V$ be the set of normalized vectors in the Hilbert space $\mathbb{C}^N$. We will let $S: V \to V$ denote the nonlinear map that a black hole produces by applying the matrix $M$ and then readjusting the norm of the state to one. We will not assume that all black holes are identical, and therefore, each time we interact with a black hole we may have a different map. We denote the transformation induced on the $k^{\text{th}}$ interaction by

$S_k : V \to V$. We treat $S_k$ as acting on the same state space for all $k$, but this is not actually a restriction because we can simply take this to be the span of all the Hilbert spaces upon which the different maps act.

Now suppose we wish to use the operations $S_1, S_2, \ldots$ to speed up Grover search. Let $x \in \{0, \ldots, N-1\}$ denote the solution to the search problem on $\{0, \ldots, N-1\}$. The corresponding unitary oracle on $\mathbb{C}^N$ is[10]

$$O_x = \mathbb{1} - 2|x\rangle\langle x|. \tag{A21}$$

The most general algorithm to search for $x$ is of the form

$$S_q O_x \ldots S_2 O_x S_1 O_x |\psi_0\rangle \tag{A22}$$

followed by a measurement. Here $|\psi_0\rangle$ is any $x$-independent quantum state on $\mathbb{C}^N$, and $S_k$ is any transformation that can be achieved on $\mathbb{C}^N$ by any sequence of unitary operations and interactions with black holes. Note that our formulation is quite general and includes the case that multiple non-unitary interactions are used after a given oracle query, as is done in [11]. Also, for some $k$, $S_k$ may be purely unitary. For example, one may have access to only a single black hole, and the rest of the iterations of Grover's algorithm must be done in the ordinary unitary manner. If the final measurement on the state described in (A22) succeeds in identifying $x$ with high probability for all $x \in \{0, \ldots, N-1\}$ then we say the query complexity of Grover search using the black hole is at most $q$.

We now adapt the proof of the $\Omega(\sqrt{N})$ quantum query lower bound for Grover search that was given in[12] [13] to show that any improvement in the query complexity for Grover search implies a corresponding lower bound on the ability of $S_k$ for some $k \in \{1, \ldots, q\}$ to "pry apart" quantum states. This then implies a corresponding lower bound on the rate of a superluminal classical information transmission channel implemented using $S_k$.

The sequence of quantum states obtained in the algorithm (A22) is

$$\begin{aligned} |\psi_0^x\rangle &= |\psi_0\rangle \\ |\phi_1^x\rangle &= O_x|\psi_0\rangle \\ |\psi_1^x\rangle &= S_1 O_x|\psi_0\rangle \\ |\phi_2^x\rangle &= O_x S_1 O_x|\psi_0\rangle \\ |\psi_2^x\rangle &= S_2 O_x S_1 O_x|\psi_0\rangle \\ &\vdots \\ |\psi_q^x\rangle &= S_q O_x \ldots S_1 O_x|\psi_0\rangle. \end{aligned} \tag{A23}$$

Let

$$|\psi_k\rangle = S_k S_{k-1} \ldots S_1 |\psi_0\rangle \tag{A24}$$

$$C_k = \sum_{x=0}^{N-1} \| |\phi_k^x\rangle - |\psi_{k-1}\rangle \|^2 \tag{A25}$$

$$D_k = \sum_{x=0}^{N-1} \| |\psi_k^x\rangle - |\psi_k\rangle \|^2. \tag{A26}$$

$|\psi_k\rangle$ can be interpreted as the state which would have been obtained after the $k^{\text{th}}$ step of the algorithm with no oracle queries (or of the Grover search problem lacked a solution).

Now, assume that for all $x \in \{0, \ldots, N-1\}$ the search algorithm succeeds after $q$ queries in finding $x$ with probability at least $\frac{1}{2}$. Then,

$$|\langle x|\psi_q^x\rangle|^2 \geq \frac{1}{2} \quad \forall x \in \{0, \ldots, N-1\} \tag{A27}$$

which implies

$$D_q \geq \eta N, \tag{A28}$$

with $\eta = (\sqrt{2} - \sqrt{2-\sqrt{2}})^2 \simeq 0.42$, as shown in [13] and discussed in [14]. By (A23), (A25), and (A26),

$$C_k = \sum_{x=0}^{N-1} \| O_x |\psi_{k-1}^x\rangle - |\psi_{k-1}\rangle \|^2 \tag{A29}$$

$$\leq D_{k-1} + 4\sqrt{D_{k-1}} + 4, \tag{A30}$$

where the above inequality is obtained straightforwardly using the triangle and Cauchy-Schwarz inequalities.

Next, let

$$R_k = D_k - C_k. \tag{A31}$$

Thus, by (A23), (A25), and (A26),

$$R_k = \sum_{x=0}^{N-1} \| S_k|\phi_k^x\rangle - S_k|\psi_{k-1}\rangle \|^2 - \sum_{x=0}^{N-1} \| |\phi_k^x\rangle - |\psi_{k-1}\rangle \|^2. \tag{A32}$$

Hence, one sees that $R_k$ is some measure of the ability of $S_k$ to "pry apart" quantum states. (In ordinary quantum mechanics $S_k$ would be unitary and hence $R_k$ would equal zero.)

Combining (A31) and (A30) yields

$$D_k \leq R_k + D_{k-1} + 4\sqrt{D_{k-1}} + 4. \tag{A33}$$

Let

$$B = \max_{1 \leq k \leq q} R_k. \tag{A34}$$

Then (A33) yields the simpler inequality

$$D_k \leq B + D_{k-1} + 4\sqrt{D_{k-1}} + 4. \tag{A35}$$

By (A23) and (A26),

$$D_0 = 0. \tag{A36}$$

By an inductive argument, one finds that (A35) and (A36) imply

$$D_k \leq (4+B)k^2. \tag{A37}$$

Combining (A37) and (A28) yields

$$(4+B)q^2 \geq \eta N, \tag{A38}$$

or in other words

$$B \geq \frac{\eta N}{q^2} - 4. \tag{A39}$$

Thus, by (A34) and (A32), there exists some $k \in \{1, \ldots, q\}$ such that

$$\sum_{x=0}^{N-1} \left( \| S_k|\phi_k^x\rangle - S_k|\psi_{k-1}\rangle \|^2 - \| |\phi_k^x\rangle - |\psi_{k-1}\rangle \|^2 \right) \geq \frac{\eta N}{q^2} - 4 \tag{A40}$$

Hence, there exists some $x \in \{0, \ldots, N-1\}$ such that

$$\| S_k|\phi_k^x\rangle - S_k|\psi_{k-1}\rangle \|^2 - \| |\phi_k^x\rangle - |\psi_{k-1}\rangle \|^2 \geq \frac{\eta}{q^2} - \frac{4}{N}. \tag{A41}$$

To simplify notation, define

$$|A\rangle = |\phi_k^x\rangle \tag{A42}$$
$$|B\rangle = |\psi_{k-1}\rangle \tag{A43}$$
$$|A'\rangle = S_k|\phi_k^x\rangle \tag{A44}$$
$$|B'\rangle = S_k|\psi_{k-1}\rangle. \tag{A45}$$

Then (A41) becomes

$$\||A'\rangle - |B'\rangle\|^2 - \||A\rangle - |B\rangle\|^2 \geq \frac{\eta}{q^2} - \frac{4}{N}. \tag{A46}$$

Recalling that $\||\psi\rangle\| = \sqrt{\langle\psi|\psi\rangle}$, (A46) is equivalent to

$$\mathrm{Re}\langle A|B\rangle - \mathrm{Re}\langle A'|B'\rangle \geq \epsilon \tag{A47}$$

with

$$\epsilon = \frac{\eta}{2q^2} - \frac{2}{N}. \tag{A48}$$

Next we will show that, within the framework of final-state projection models, (A47) implies that Alice can send a polynomial fraction of a bit to Bob or vice versa using preshared entanglement and a single application of black hole dynamics. Recall that, within the final state projection model,

$$|A'\rangle = \frac{M|A\rangle}{\sqrt{\langle A|M^\dagger M|A\rangle}} \tag{A49}$$
$$|B'\rangle = \frac{M|B\rangle}{\sqrt{\langle B|M^\dagger M|B\rangle}} \tag{A50}$$

Thus, (A47) is equivalent to

$$\mathrm{Re}\left[\langle A|\left(\mathbb{1} - \frac{M^\dagger M}{\sqrt{\langle A|M^\dagger M|A\rangle\langle B|M^\dagger M|B\rangle}}\right)|B\rangle\right] \geq \epsilon \tag{A51}$$

Hence,

$$\left\|\mathbb{1} - \frac{M^\dagger M}{\sqrt{\langle A|M^\dagger M|A\rangle\langle B|M^\dagger M|B\rangle}}\right\| \geq \epsilon. \tag{A52}$$

Again using $\lambda_0$ to denote the smallest singular value of $M$ and $\lambda_1$ to denote the largest, we see that, assuming $\epsilon$ is nonnegative, (A52) implies either

**Case 1:**

$$\frac{\lambda_1^2}{\sqrt{\langle A|M^\dagger M|A\rangle\langle B|M^\dagger M|B\rangle}} \geq 1 + \epsilon, \tag{A53}$$

which implies

$$\frac{\lambda_1^2}{\lambda_0^2} \geq 1 + \epsilon, \tag{A54}$$

or

**Case 2:**

$$\frac{\lambda_0^2}{\sqrt{\langle A|M^\dagger M|A\rangle\langle B|M^\dagger M|B\rangle}} \leq 1 - \epsilon, \tag{A55}$$

which implies

$$\frac{\lambda_0^2}{\lambda_1^2} \leq 1 - \epsilon. \tag{A56}$$

Examining (A47), one sees that $\epsilon$ can be at most 2. If $0 \leq \epsilon \leq 1$ then (A56) implies (A54). If $1 < \epsilon \leq 2$ then case 2 is impossible. Hence, for any nonnegative $\epsilon$ one obtains (A54). Hence, by the results of sections A 1 and A 2, Alice and Bob can communicate in either direction through a binary asymmetric channel whose bitflip probabilities $\epsilon_0$ for transmission of zero and $\epsilon_1$ for transmission of one are given by

$$\epsilon_0 = \frac{1}{2} \tag{A57}$$

$$\epsilon_1 = \frac{\lambda_0^2}{\lambda_0^2 + \lambda_1^2} \leq \frac{1}{2+\epsilon}. \tag{A58}$$

For $0 \leq \epsilon \leq 2$, $\frac{1}{2+\epsilon} \leq \frac{1}{2} - \frac{\epsilon}{8}$. Thus, (A58) implies the following more convenient inequality

$$\epsilon_1 \leq \frac{1}{2} - \frac{\epsilon}{8}. \tag{A59}$$

In section A 1 we calculated that the channel capacity in the case that $\epsilon_0 = \frac{1}{2}$ and $\epsilon_1 = \frac{1}{2} - \delta$ is $\Omega(\delta^2)$ for $\delta \ll 1$. Thus, (A57) and (A59) imply a channel capacity in either direction of

$$C = \Omega\left(\left(\frac{\eta}{2q^2} - \frac{2}{N}\right)^2\right) \tag{A60}$$

in the regime $0 < \frac{\eta}{2q^2} - \frac{2}{N} \ll 1$. □

The above scaling of the superluminal channel capacity with Grover speedup shows that polynomial speedup for small instances or exponential speedup for large instances imply 1/poly superluminal channel capacity. In particular, to solve $\mathsf{NP}$ in polynomial time without exploiting problem structure we would need $q \propto \log^c N$ for some constant $c$. In this setting $N = 2^n$ where $n$ is the size of the witness for the problem in NP. In this limit, (A60) implies instantaneous signaling channels in each direction with capacity at least

$$C = \Omega\left(\frac{1}{\log^{4c} N}\right) = \Omega\left(\frac{1}{n^{4c}}\right). \tag{A61}$$

If we assume that superluminal signaling capacity is limited to some negligibly small capacity $C \leq \epsilon$ then, by (A61), NP-hard problems cannot be solved by unstructured search in time scaling polynomially with witness size (specifically $n^c$ for some constant $c$) except possibly for unphysically large instances with $n = \Omega\left(\left(\frac{1}{\epsilon}\right)^{\frac{1}{4c}}\right)$.

### 4. Signaling implies Super-Grover Speedup

In sections A 1 and A 2 we showed that if final-state projection can be used to speed up Grover search it can also be used for superluminal signaling. In this section we show the converse. Unlike in section A 3, we here make the assumption that we can make multiple uses of the same non-unitary map $S$ (just as other quantum gates can be used multiple times without variation). Since signaling cannot be achieved by performing unitary operations on entangled quantum degrees of freedom, superluminal signaling implies non-unitarity. Furthermore, as shown in Section F, iterated application of any nonlinear but differentiable map allows the Grover search problem to be solved with only a single oracle query. The nonlinear maps that arise in final-state projection models are differentiable (provided $M$ is invertible), and thus within the final-state projection framework signaling implies single-query Grover search. In the remainder of this section we quantitatively investigate how many iterations of the nonlinear map are needed to achieve single-query Grover search, as a function of the superluminal signaling capacity. We find that unless the signaling capacity is exponentially small, logarithmically many iterations suffice. Specifically, our main result of this section is the following theorem.

**Theorem 4.** *Suppose Alice has access to a linear but not necessarily unitary maps on quantum states, as can arise in the Horowitz-Maldecena final state projection model. Suppose she achieves instantaneous classical communication capacity of $C$ bits transmitted per use of nonunitary dynamics. Then she could solve the Grover search problem on a database of size $N$ using a single query and $O\left(\frac{\log(N)}{\log(1+C^2)}\right)$ applications of the available nonunitary maps.*

*Proof.* Suppose Alice has access to black hole(s) and Bob does not. Alice will use this to send signals to Bob using some shared entangled state $|\psi\rangle_{AB}$. Her most general protocol is to apply some map $M_0$ to her half of the state if she wishes to transmit a zero and some other map $M_1$ if she wishes to transmit a one. (As a special case, $M_0$ could be the identity.) Here, per the final state projection model, $M_0$ and $M_1$ are linear but not necessarily unitary maps, and normalization of quantum states is to be adjusted back to one after application of these maps. The possible states shared by Alice and Bob given Alice's two possible messages are

$$|\psi_0\rangle_{AB} \propto M_0|\psi\rangle_{AB} \tag{A62}$$
$$|\psi_1\rangle_{AB} \propto M_1|\psi\rangle_{AB}. \tag{A63}$$

The signaling capacity is determined by the distinguishability of the two corresponding reduced density matrices held by Bob

$$\rho_0 = \mathrm{Tr}_A\left[|\psi_0\rangle_{AB}\right] \tag{A64}$$
$$\rho_1 = \mathrm{Tr}_A\left[|\psi_1\rangle_{AB}\right]. \tag{A65}$$

We can define

$$|\psi'\rangle \propto M_0|\psi\rangle_{AB} \tag{A66}$$

in which case

$$|\psi_1\rangle_{AB} \propto M_1 M_0^{-1}|\psi'\rangle. \tag{A67}$$

(We normalize $|\psi'\rangle$ so that $\langle\psi'|\psi'\rangle = 1$.) Thus, the signaling capacity is determined by the distinguishability of

$$\rho_0 = \mathrm{Tr}_A\left[|\psi'\rangle\right] \tag{A68}$$
$$\rho_1 = \mathrm{Tr}_A\left[\frac{1}{\eta}M|\psi'\rangle\right] \tag{A69}$$

where

$$M = M_1 M_0^{-1} \tag{A70}$$
$$\eta = \sqrt{\langle\psi'|M^\dagger M|\psi'\rangle}. \tag{A71}$$

We have thus reduced our analysis to the case that Alice applies some non-unitary map $M$ to her state if she wants to transmit a one and does nothing if she wants to transmit a zero. We will next obtain a lower bound $\kappa_{\min}$ on the condition number of $M$ as a function of the signaling capacity from Alice to Bob. This then implies that one of $M_0, M_1$ has a condition number at least $\sqrt{\kappa_{\min}}$ for the general case.

Suppose that $M$ has the following singular value decomposition

$$M = \sum_i \lambda_i |\psi_i\rangle\langle\phi_i|. \tag{A72}$$

We can express $|\psi'\rangle$ as

$$|\psi'\rangle = \sum_{i,j} \alpha_{ij}|\phi_i\rangle|B_j\rangle \tag{A73}$$

where $|\phi_1\rangle, |\phi_2\rangle, \ldots$ is the basis determined by the singular value decomposition (A72) and $|B_1\rangle, |B_2\rangle, \ldots$ is the basis Bob will perform his measurement in when he tries to extract Alice's message. If Alice wishes to transmit one then she applies $M$ yielding

$$|\psi_1\rangle \propto \sum_{i,j} \lambda_i \alpha_{ij} |\psi_i\rangle |B_j\rangle. \tag{A74}$$

So

$$\rho_0 = \sum_{i,j,k} \alpha_{ij} \alpha_{ik}^* |B_j\rangle\langle B_k| \tag{A75}$$

$$\rho_1 = \sum_{i,j,k} \frac{\lambda_i^2}{\eta} \alpha_{ij} \alpha_{ik}^* |B_j\rangle\langle B_k|. \tag{A76}$$

Consequently, Bob's measurement will yield a sample from the following probability distributions conditioned on Alice's message.

$$p(j|0) = \sum_i |\alpha_{ij}|^2 \tag{A77}$$

$$p(j|1) = \sum_i \frac{\lambda_i^2}{\eta} |\alpha_{ij}|^2. \tag{A78}$$

The total variation distance between these distributions, which determines the capacity of the superluminal channel is

$$\Delta = \frac{1}{2} \sum_j |p(j|0) - p(j|1)| = \frac{1}{2} \sum_j \left| \sum_i |\alpha_{ij}|^2 \left(1 - \frac{\lambda_i^2}{\eta}\right) \right|. \tag{A79}$$

From a given value of this total variation distance we wish to derive a lower bound on the condition number of $M$, that is, the ratio of the largest singular value to the smallest. Applying the triangle inequality to (A79) yields

$$\Delta \leq \frac{1}{2} \sum_{ij} |\alpha_{ij}|^2 \left| 1 - \frac{\lambda_i^2}{\eta} \right|. \tag{A80}$$

Because $\alpha_{ij}$ are amplitudes in a normalized quantum state,

$$p(i) = \sum_j |\alpha_{ij}|^2 \tag{A81}$$

is a probability distribution. We can thus rewrite (A80) as

$$\Delta \leq \frac{1}{2} \sum_i p(i) \left| 1 - \frac{\lambda_i^2}{\eta} \right| \tag{A82}$$

$$\leq \frac{1}{2} \max_i \left| 1 - \frac{\lambda_i^2}{\eta} \right|. \tag{A83}$$

In keeping with the notation from previous sections, we let $\lambda_0$ denote the smallest singular value of $M$ and $\lambda_1$ the largest. Thus, (A83) yields

$$\Delta \leq \frac{1}{2} \max \left\{ 1 - \frac{\lambda_0^2}{\eta}, \frac{\lambda_1^2}{\eta} - 1 \right\}. \tag{A84}$$

Similarly,

$$\eta = \sum_{jk} |\alpha_{jk}|^2 \lambda_j^2 \tag{A85}$$

$$= p(j)\lambda_j^2 \tag{A86}$$

$$\in [\lambda_0^2, \lambda_1^2]. \tag{A87}$$

Applying (A87) to (A84) yields

$$\Delta \leq \frac{1}{2} \max \left\{ 1 - \frac{\lambda_0^2}{\lambda_1^2}, \frac{\lambda_1^2}{\lambda_0^2} - 1 \right\}. \tag{A88}$$

As shown in section A 5, the channel capacity $C$ is related to the total variation distance $\Delta$ according to

$$C \leq \Delta - \Delta \log_2 \Delta \tag{A89}$$

for $\Delta \leq 1/e$. For small $\Delta$, the $-\Delta \log_2 \Delta$ term dominates the $\Delta$ term. We can simplify further by noting that for all positive $\Delta$, $\sqrt{\Delta} > -\Delta \log_2(\Delta)$. Hence, $C = O(\sqrt{\Delta})$. Thus to achieve a given channel capacity $C$ we need

$$\Delta = \Omega(C^2). \tag{A90}$$

By (A88), this implies that achieving a channel capacity $C$ requires

$$|1 - \kappa_{\min}^2| = \Omega(C^2), \tag{A91}$$

where $\kappa_{\min}$ is the condition number of the nonlinear map $M = M_1 M_0^{-1}$. This implies that one of $M_0$ or $M_1$ must have condition number at least $\kappa = \sqrt{\kappa_{min}} = \Omega\left((1 - C^2)^{1/4}\right)$. This in turn implies Grover search with one query and $O(\log_\kappa(N))$ applications of the nonlinear map via the methods of [11]. □

### 5. Channel Capacity and Total Variation Distance

Alice wishes to transmit a message to Bob. If she sends zero Bob receives a sample from $p(B|0)$ and if she sends one Bob receives a sample from $p(B|1)$. Here, $B$ is a random variable on a finite state space $\Gamma = \{0, 1 \ldots, d-1\}$. The only thing we know about this channel is that

$$|p(B|0) - p(B|1)| = \delta, \tag{A92}$$

where $|\cdot|$ denotes the total variation distance (*i.e.* half the $l_1$ distance). In this section we derive an upper bound on the channel capacity as a function of $\delta$. Specifically, we show that the (asymptotic) capacity $C$ obeys

$$C \leq \delta - \delta \log_2 \delta. \tag{A93}$$

Any strategy that Bob could use for decoding Alice's message corresponds to a decomposition of $\Gamma$ as

$$\Gamma = \Gamma_0 \sqcup \Gamma_1 \tag{A94}$$

where $\Gamma_0$ are the outcomes that Bob interprets as zero and $\Gamma_1$ are the outcomes that Bob interprets as one.

From (A92) it follows that

$$|p(b \in \Gamma_0 | A = 0) - p(b \in \Gamma_0 | A = 1)| \leq \delta. \tag{A95}$$

(The defining property of total variation distance is that this holds for any set $\Gamma_0$.)

Let $F = 0$ whenever $B \in \Gamma_0$ and $F = 1$ whenever $B \in \Gamma_1$. That is, the random variable $F$ is Bob's guess as to Alice's message. By standard Shannon theory [15], the channel capacity is the mutual information $I(F; A)$ maximized over Alice's choice of $p(A)$.

From (A95) it follows that

$$|p(F|A = 0) - p(F|A = 1)| \leq \delta. \tag{A96}$$

Let $p_\alpha$ be the probability distribution

$$p_\alpha(F) = \alpha p(F|A = 0) + (1 - \alpha)p(F|A = 1) \tag{A97}$$

for some $\alpha \in [0, 1]$. From the elementary properties of total variation distance it follows that

$$|p_\alpha(F) - p(F|A = 0)| \leq \delta \tag{A98}$$

and

$$|p_\alpha(F) - p(F|A = 1)| \leq \delta \tag{A99}$$

for any choice of $\alpha$. In particular, we may set $\alpha = p(A = 0)$, in which case we have

$$|p(F) - p(F|A = 0)| \leq \delta \tag{A100}$$

$$|p(F) - p(F|A = 1)| \leq \delta. \tag{A101}$$

Next, we recall the Fannes inequality. This says that for any two density matrices $\rho, \sigma$ on a $d$-dimensional Hilbert space whose trace distance satisfies $T \leq \frac{1}{e}$

$$|S(\rho) - S(\sigma)| \leq T \log_2 d - T \log_2 T. \tag{A102}$$

Specializing to the special case that $\sigma$ and $\rho$ are simultaneously diagonalizable, one obtains the following statement about classical entropies.

**Corollary 1.** *Let $p$ and $q$ be two probability distributions on a state space of size $d$. Let $T$ be the total variation distance between $p$ and $q$. Suppose $T \leq \frac{1}{e}$. Then*

$$|H(p) - H(q)| \leq T \log_2 d - T \log_2 T. \tag{A103}$$

Applying corollary 1 to (A100) and (A101) yields[16]

$$|H [p(F)] - H [p(F|A = 0)]| \leq \delta - \delta \log_2 \delta \tag{A104}$$

$$|H [p(F)] - H [p(F|A = 1)]| \leq \delta - \delta \log_2 \delta. \tag{A105}$$

Thus,

$$\begin{aligned} I(F; A) &= H(F) - H(F|A) && \text{(A106)} \\ &= H [p(F)] - p(A = 0)H [p(F|A = 0)] - p(A = 1)H [p(F|A = 1)] && \text{(A107)} \\ &\leq \delta - \delta \log_2 \delta, && \text{(A108)} \end{aligned}$$

which completes the derivation.

## SECTION B: VIOLATIONS OF THE BORN RULE

In this section we consider modification of quantum mechanics in which states evolve unitarily, but measurement statistics are not given by the Born rule. This is loosely inspired by the "state dependence" resolution of the firewalls paradox, put forth by Papadodimas and Raju [17]. In this theory, the measurement operators $O$ which correspond to observables are not fixed linear operators, but rather vary depending on the state they are operating on, i.e. $O = O(|\psi\rangle)$. (In general such dependencies lead to nonlinearities in quantum mechanics, but Papadodimas and Raju argue these are unobservable in physically reasonable experiments.) Recently Marolf and Polchinski [18] have claimed that such modifications of quantum mechanics lead to violations of the Born rule. We do not take a position either way on Marolf and Polchinski's claim, but use it as a starting point to investigate how violations of the Born rule are related to superluminal signaling and computational complexity.

Here we consider violations of the Born rule of the following form: given a state $|\psi\rangle = \sum_x \alpha_x |x\rangle$, the probability $p_x$ of seeing outcome $x$ is given by

$$p_x = \frac{f(\alpha_x)}{\sum_{x'} f(\alpha_{x'})} \tag{B1}$$

for some function $f(\alpha) : \mathbb{C} \to \mathbb{R}^+$. We assume that states in the theory evolve unitarily as in standard quantum mechanics. One could consider more general violations of the Born rule, in which the function $f$ depends not only on the amplitude $\alpha_x$ on $x$, but on the amplitudes on other basis states as well. However such a generalized theory seems impractical to work with, so we do not consider such a theory here.

We first show that, assuming a few reasonable conditions on $f$ (namely that $f$ has a reasonably behaved derivative and that measurement statistics do not depend on the normalization of the state), the only way to modify the Born rule is to set $f(\alpha) = |\alpha|^{2+\delta}$ for some $\delta \neq 0$. We then show that in theories where the Born rule is modified, superluminal signaling is equivalent to a speedup to Grover search. More precisely, we show that if one can send superluminal signals using states on $n$ qubits, then one can speed up Grover search on a system with $O(n)$ qubits, and vice versa. Hence one can observe superluminal signals on reasonably sized systems if and only if one can speed up Grover search using a reasonable number of qubits.

We are not the first authors to examine the complexity theoretic consequences of modifications to the Born rule. Aaronson [19] considered such modifications, and showed that if $\delta$ is any constant, then such modifications allow for the solution of #P-hard problems in polynomial time. Our contribution is to show the opposite direction, namely that a significant speedup over Grover search implies the deviation from the Born rule $\delta$ is large, and to connect this to superluminal signaling.

We prove our results in several steps. First, in Theorems 6 and 7, we show that deviations in the Born rule by $\delta$ allow the solution of NP-hard problems and superluminal signaling using $O(1/\delta)$ qubits. As noted previously, Theorem 6 follows from the work of Aaronson [19], but we include a proof for completeness.

In Theorem 8 we show that, assuming one has a superluminal signaling protocol using a shared state on $m$ qubits, the deviation from the Born rule $\delta$ must be $\geq \Omega(1/m)$. Likewise in Theorem 9 we show that if one can achieve a constant factor super-Grover speedup using $m$ qubits, that we must have $\delta \geq \Omega(1/m)$ as well. Combining these with Theorems 6 and 7 shows that a super-Grover speedup on $m$ qubits implies superluminal signaling protocols with $O(m)$ qubits and vice versa. Supplementary Figure 1 explains the relationship between these theorems below.

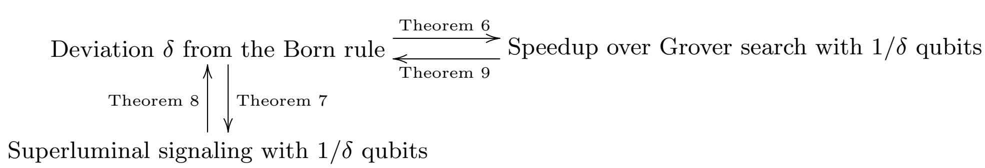



FIG. 1. Relationship between theorems connecting signaling and search.

In short, we find that a violation of the Born rule by $\delta$ is equivalent to allowing a super-Grover speedup and an instantaneous signaling protocol using $1/\delta$ qubits. Hence in theories in which $\delta$ is only polynomially suppressed (as a function e.g. of the number of fields $N$ in Super-Yang-Mills), then such theories allow for superluminal signaling and violations of the Grover lower bound with reasonable overheads. On the other hand, our results do not rule out violations of the Born rule in which $1/\delta$ is unphysically large.

### 1. Power law violations are unique

We now show that, given some reasonable assumptions about the function $f(\alpha)$, the only possible violation of the Born rule is given by $f(\alpha) = |\alpha|^p$. In particular we will demand the following properties of $f$:

1. Well-behaved derivative: $f(\alpha)$ is continuous and differentiable, and $f'(\alpha)$ changes sign at most a finite number of times on $[0, 1]$
2. Scale invariance: for any $k \in \mathbb{C}$, we have that $\frac{f(k\alpha)}{\sum_x f(k\alpha_x)} = \frac{f(\alpha)}{\sum_x f(\alpha_x)}$. In other words the calculation of the probability $p_x$ of seeing outcome $x$ is independent of the norm or phase of the input state; it only depends on the state of the projective Hilbert space.

There are a number of other reasonable constraints one could impose; for instance one could demand that the modified Born rule has to behave well under tensor products. Suppose you have a state $|\psi\rangle = \sum_x \alpha_x |y\rangle$ and a state $|\phi\rangle = \sum_y \beta_y |y\rangle$. A reasonable assumption would be to impose that in the state $|\psi\rangle \otimes |\phi\rangle$, the probability $p_{xy}$ of measuring outcome $xy$ should be equal to $p_x p_y$, i.e. a tensor product state is equivalent to independent copies of each state. More formally this would state that

$$\frac{f(\alpha_x \beta_y)}{\sum_{x'y'} f(\alpha_{x'} \beta_{y'})} = \frac{f(\alpha_x)}{\sum_{x'} f(\alpha_{x'})} \frac{f(\beta_y)}{\sum_{y'} f(\beta_{y'})}. \tag{B2}$$

Let us call this the Tensor product property. It will turn out that the Tensor product property is implied by the Scale invariance property, which we will show in our proof.

We now show that the Well-behaved derivative and Scale invariance properties imply $f(\alpha) = |\alpha|^p$ for some $p$.

**Theorem 5.** *Suppose that $f$ satisfies the Well-behaved derivative and Scale invariance properties. Then $f(\alpha) = |\alpha|^p$ for some $p \in \mathbb{R}$.*

*Proof.* First, note that the functions $f(\alpha)$ and $cf(\alpha)$ give the same measurement statistics for any scalar $c \in \mathbb{R}$. To eliminate this redundancy in our description of $f$, we'll choose $c$ such that $f(1) = 1$.

For any $\alpha \in \mathbb{C}$, consider the (non-normalized) state $\alpha|0\rangle + |1\rangle$. By scale invariance, for any $\beta \in \mathbb{C}$, we must have that

$$\frac{f(\alpha)}{f(\alpha) + f(1)} = \frac{f(\alpha\beta)}{f(\alpha\beta) + f(\beta)} \tag{B3}$$

which implies that $f(\alpha)f(\beta) = f(\alpha\beta)f(1) = f(\alpha\beta)$ for all $\alpha, \beta \in \mathbb{C}$. One can easily check that this implies the tensor product property.

In particular this holds for any phase, so if $\alpha = |\alpha|e^{i\theta}$, we must have that $f(\alpha) = \hat{f}(|\alpha|)g(\theta)$ for some functions $\hat{f} : \mathbb{R}^{\geq 0} \to \mathbb{R}^+$ and $g : [0, 2\pi) \to \mathbb{R}^+$. Note that taking $g \to cg$ and $\hat{f} \to \hat{f}/c$ leaves $f$ invariant for any scalar $c \in \mathbb{R}^+$. So without loss of generality, since $f(1) = 1$, we can set $\hat{f}(1) = g(0) = 1$ as well by an appropriate choice of scalar $c$. Now, for any phases $e^{i\theta}$ and $e^{i\phi}$, we have $f(e^{i\theta})f(e^{i\phi}) = f(e^{i(\theta+\phi)})$. Since $\hat{f}(1) = 1$ this implies $g(\theta)g(\phi) = g(\theta + \phi)$, i.e. $g$ must be a real one-dimensional representation of $U(1)$. The only such representation is $g = 1$, hence $f(\alpha) = f(|\alpha|)$.

Now we will show that $f(x) = x^p$ for some $p$. Consider any $0 < x < 1$ and $0 < x' < 1$ where $x \neq x'$. Since $f(\alpha)f(\beta) = f(\alpha\beta)$, we must have that $f(x^k) = f(x)^k$ and $f(x'^k) = f(x')^k$ for any $k \in \mathbb{N}$. Let $p = \log(f(x))/\log(x)$ and $p' = \log(f(x'))/\log(x')$. Then the above equations imply that $f(x^k) = x^{kp}$ and $f(x'^k) = x'^{kp'}$ for all $k \in \mathbb{N}$.

Now suppose by way of contradiction that there exist $x, x'$ such that $p \neq p'$. Since both $x < 1$ and $x' < 1$, as $k \to \infty$ we have that $f(x^{kp}) \to 0$ and $f(x'^{kp}) \to 0$. However, the sequence of points $f(x), f(x^2), f(x^3), \ldots$ approaches zero along the curve $h(x) = x^p$ while the sequence of points $f(x'), f(x'^2), f(x'^3), \ldots$ approaches zero along the curve $h'(x) = x^{p'}$. This implies $f$ must oscillate infinitely many times between the curves $h$ and $h'$, which implies $f'$ must change signs infinitely many times by the intermediate value theorem. This contradicts the Well-behaved derivative assumption.

Hence we have for all $0 < x < 1$, $f(x) = x^p$ for some $p$. Now if $x > 1$, we have $f(x)f(1/x) = f(1) = 1$. Since $1/x < 1$, then we have $f(1/x) = 1/x^p$, so $f(x) = x^p$ as well. Also $f(1) = 1^p = 1$, and by continuity $f(0) = 0$. Hence for all $x \geq 0$ we must have $f(x) = x^p$ for some $p$, as claimed. □

### 2. Born rule violations imply signaling and super-Grover speedup

We first show that large violations of the Born rule imply a large speedup to Grover search and allow for large amounts of superluminal signaling. This was previously shown by Aaronson [19], but for completeness we will summarize the proof here.

**Theorem 6** (Aaronson [19] Theorem 6). *Suppose that the Born rule is modified such that $f(\alpha) = |\alpha|^{2+\delta}$ where $\delta \neq 0$. Then one can solve* PP *problems on instances of size $n$ in time $O(\frac{n^2}{|\delta|})$. In particular one can search an unordered list of $2^n$ indices in $O(\frac{n^2}{|\delta|})$ time.*

*Proof.* We will use the modified Born rule to simulate postselection. Suppose one has a state $|\Psi\rangle = \sum_x (\alpha_x|0\rangle + \beta_x|1\rangle)|x\rangle$ and wishes to simulate postselection of the first qubit on the state $|0\rangle$. Suppose $\delta > 0$; the case $\delta < 0$ follows analogously. To simulate postselection on zero, simply append $k$ ancilla qubits in the $|0\rangle$ state. Then apply a Hadamard to each of the ancilla qubits controlled on the first qubit being a 1. The state now evolves to

$$\sum_x \left( \alpha_x|0\rangle|x\rangle|0\rangle^{n/\delta} + \beta_x|1\rangle|x\rangle \sum_y 2^{-k/2}|y\rangle \right) \tag{B4}$$

When measuring this state in the computational basis, the probability of measuring a 0 on the first qubit is proportional to $\sum_x |\alpha_x|^{2+\delta}$, while the probability of getting a 1 on the first qubit is proportional to $2^{-k\delta/2} \sum_x |\beta_x|^{2+\delta}$. Hence setting $k = n/\delta$, the probability of getting a 1 on the first qubit is exponentially suppressed by a factor of $2^{-n}$. This effectively postselects the first qubit to have value 0 as desired. The rest of the proof follows from the fact that Aaronson's PostBQP algorithm to solve PP-hard problems on instances of size $n$ runs in time $O(n)$ and involves $O(n)$ postselections; hence using this algorithm to solve PP-hard problems when the Born rule is violated takes time $O(\frac{n^2}{\delta})$ as claimed. □

Aaronson's result also implies that large violations of the Born rule imply one can send superluminal signals with small numbers of qubits.

**Theorem 7** (Aaronson [19]). *Suppose that the Born rule is modified such that $f(\alpha) = |\alpha|^{2+\delta}$ where $\delta \neq 0$. Then one can transmit a bit superluminally in a protocol involving a state on $O(n/|\delta|)$ qubits which succeeds with probability $1 - 2^{-n}$. Note one can use this protocol to send either classical bits or quantum bits.*

*Proof.* The proof follows almost immediately from the proof of Theorem 6. Suppose that Alice wishes to send a bit 0 or 1 to Bob. Alice and Bob can perform the standard teleportation protocol [20], but instead of Alice sending her classical measurement outcomes to Bob, Alice simply postselects her measurement outcome to be 00 (i.e. no corrections are necessary to Bob's state) using the trick in Theorem 6. If Alice uses $O(n/|\delta|)$ qubits to simulate the postselection, and then measures her qubits, she will obtain outcome 00 with probability $1 - 2^{-n}$ and the bit will be correctly transmitted as desired. □

### 3. Signaling implies large power law violation

We now show that if one can send a superluminal signal with bias $\epsilon$ using a shared state on $n$ qubits, then the violation of the Born rule $\delta$ must satisfy $|\delta| \geq O(\epsilon/n)$. Hence $\delta$ and $\epsilon$ must be polynomially related. Put less precisely, if a physically reasonable experiment can send a superluminal signal with a nontrivial probability, then there must be a nontrivial (and hence observable) violation of the Born rule. This in turn, implies by Theorem 6 that one can solve NP-hard problems with a reasonable multiplicative overhead.

**Theorem 8.** *Suppose that the Born rule is modified such that $f(\alpha) = |\alpha|^{2+\delta}$, and suppose there is a signaling protocol using an entangled state on $n$ qubits signaling with probability $\epsilon$. Then $|\delta| \geq O(\frac{\epsilon}{n})$.*

*Proof.* Consider the most general signaling protocol to send a bit of information. Suppose that Alice and Bob share an entangled state $|\Phi\rangle$ on $n$ qubits, $m$ of which are held by Bob and $n-m$ of which are held by Alice. To send a zero, Alice performs some unitary $U_0$ on her half of the state, and to send a one, Alice performs some unitary $U_1$ on her half of the state. Bob then measures in some fixed basis $B$. This is equivalent to the following protocol: Alice and Bob share the state $|\Psi\rangle = U_0|\Phi\rangle$ ahead of time, and Alice does nothing to send a 0, and applies $U = U_1 U_0^\dagger$ to obtain $|\Psi'\rangle = U|\Psi\rangle$ send a 1. Then Bob measures in basis $B$. We say the protocol succeeds with probability $\epsilon$ if the distributions seen by Bob in the case Alice is sending a 0 vs. a 1 differ by $\epsilon$ in total variation distance. As shown in section A 5 of Section B, the total variation distance is polynomially related to the capacity of the resulting classical communication channel.

Let $\alpha_{xy}$ be the amplitude of the state $|x\rangle|y\rangle$ in $|\Psi\rangle$, where the $|x\rangle$ is an arbitrary basis for Alice's qubits and $|y\rangle$ are given by the basis $B$ in which Bob measures his qubits. Let $\alpha'_{xy}$ be the amplitude of $|x\rangle|y\rangle$ in the state $|\Psi'\rangle$, so we have $\alpha'_{xy} = \sum_{x'} U_{xx'}\alpha_{x'y}$. In short

$$|\Psi\rangle = \sum_{xy} \alpha_{xy}|x\rangle|y\rangle \qquad\qquad |\Psi'\rangle = \sum_{xy} \alpha'_{xy}|x\rangle|y\rangle \tag{B5}$$

Assume that $\sum_{x,y} |\alpha_{xy}|^2 = 1$, i.e. the state is normalized in the $\ell_2$ norm. Since $U$ is unitary this implies the state $U|\psi'\rangle$ is normalized in the $\ell_2$ norm as well.

Now suppose that the protocol has an $\epsilon$ probability of success. Let $D_0$ be the distribution on outcomes $y \in \{0,1\}^m$ when Alice is sending a zero, and $D_1$ be the distribution when Alice is sending a 1. Let $D_b(y)$ denote the probability of obtaining outcome $y$ under $D_b$. Then the total variation distance between $D_0$ and $D_1$, given by $\frac{1}{2}\sum_y |D_0(y) - D_1(y)|$, must be at least $\epsilon$. Equivalently, there must be some event $S \subset \{0,1\}^m$ for which

$$\sum_{y\in S} D_0(y) - D_1(y) \geq \epsilon \tag{B6}$$

and for which, for all $y \in S$, we have $D_0(y) > D_1(y)$.

Assume for the moment that $\delta > 0$; an analogous proof will hold in the case $\delta < 0$. Let $N = 2^n$ be the dimension of the Hilbert space of $|\Psi\rangle$. Plugging in the probabilities $D_0(y)$ and $D_1(y)$ given by the modified Born rule, we obtain

$$\begin{aligned}
\epsilon &\leq \sum_{x\in\{0,1\}^{n-m}, y\in S} \frac{|\alpha_{xy}|^{2+\delta}}{\sum_{x'y'} |\alpha_{x'y'}|^{2+\delta}} - \frac{|\alpha'_{xy}|^{2+\delta}}{\sum_{x'y'} |\alpha'_{x'y'}|^{2+\delta}} && \text{(B7)}\\
&\leq \sum_{x\in\{0,1\}^{n-m}, y\in S} N^{\delta/2}|\alpha_{xy}|^{2+\delta} - |\alpha'_{xy}|^{2+\delta} && \text{(B8)}\\
&= \sum_{x\in\{0,1\}^{n-m}, y\in S} \left(1 + \frac{\delta}{2}\log(N)\right) |\alpha_{xy}|^2 \left(1 + \delta \log|\alpha_{xy}|\right) - |\alpha'_{xy}|^2 (1 + \delta\log|\alpha'_{xy}|) + O(\delta^2) && \text{(B9)}\\
&= \sum_{x\in\{0,1\}^{n-m}, y\in S} \left(|\alpha_{xy}|^2 - |\alpha'_{xy}|^2\right) + \frac{\delta}{2}\log(N)|\alpha_{xy}|^2 + \delta\left(|\alpha_{xy}|^2 \log|\alpha_{xy}| - |\alpha'_{xy}|^2\log|\alpha'_{xy}|\right) + O(\delta^2) && \text{(B10)}\\
&\leq \frac{\delta}{2}\log(N) + \frac{\delta}{2}\sum_{x\in\{0,1\}^{n-m}, y\in S} \left(|\alpha_{xy}|^2\log|\alpha_{xy}|^2 - |\alpha'_{xy}|^2\log|\alpha'_{xy}|^2\right) + O(\delta^2) && \text{(B11)}\\
&\leq \frac{\delta}{2}\log(N) + \frac{\delta}{2}\log(N) + O(\delta^2) = \delta n + O(\delta^2) && \text{(B12)}
\end{aligned}$$

On line (B8) we used the fact that for any vector $|\phi\rangle = \sum_y \beta_y |y\rangle$ of $\ell_2$ norm 1 over a Hilbert space of dimension $N$, we have $N^{-\delta/2} \leq \sum_y |\beta_y|^{2+\delta} \leq 1$ when $\delta > 0$. On line (B9) we expanded to first order in

$\delta$. On line (B11) we used the fact that the first term is zero because applying a unitary to one half of a system does not affect measurement outcomes on the other half of the system and the second sum is upper bounded by 1. On line (B12) we used the fact that the sum is given by a difference of entropies of (possibly subnormalized) probability distributions, each of which is between zero and $\log(N)$.

Hence we have that $\delta n + O(\delta^2) \geq \epsilon$, so to first order in $\delta$ we must have $\delta \geq \epsilon/n$ as claimed. □

The following corollary follows from Theorem 6, and hence we've shown that superluminal signaling implies a super-Grover speedup.

**Corollary 2.** *Suppose that the Born rule is modified such that $f(\alpha) = |\alpha|^{2+\delta}$, and that there is a signaling protocol using an entangled state on $n$ qubits which signals with probability $\epsilon$. Then there is an algorithm to solve #P-hard and NP-hard instances of size $m$ (e.g. #SAT on $m$ variables) in time $O(m^2 n/\epsilon)$.*

### 4. Super-Grover speedup implies signaling

We now show that even a mild super-Grover speedup implies that $\delta$ is large, and hence one can send superluminal signals. Our proof uses the hybrid argument of Bennett, Bernstein, Brassard and Vazirani [13] combined with the proof techniques of Theorem 8.

**Theorem 9.** *Suppose that the Born rule is modified such that $f(\alpha) = |\alpha|^{2+\delta}$, and there is an algorithm to search an unordered list of $N$ items with $Q$ queries using an algorithm over a Hilbert space of dimension $M$. Then*

$$\frac{1}{6} \leq \frac{2Q}{\sqrt{N}} + |\delta| \log(M) + O(\delta^2). \tag{B13}$$

*Proof.* Suppose that such an algorithm exists. It must consist of a series of unitaries and oracle calls followed by a measurement in the computational basis.

Let $|\psi^0\rangle = \sum_y \alpha_y^0 |y\rangle$ be the state of the algorithm just before the final measurement when there is no marked item, and let $|\psi^x\rangle = \sum_y \alpha_y^x |y\rangle$ be the state if there is a marked item. Let $D_0$ be the distribution on $y$ obtained by measuring $|\psi^0\rangle$ in the computational basis, and $D_x$ be the distribution obtained by measuring $|\psi^x\rangle$. We know that $|\psi^0\rangle$ and $|\psi^x\rangle$ must be distinguishable with 2/3 probability for every $x$. Hence we must have that the total variation distance between $D_0$ and $D_x$ must be at least 1/6 for every $x$ (otherwise one could not decide the problem with bias 1/6). This implies there must exist some event $S_x$ for which

$$\frac{1}{6} \leq \sum_{y \in S_x} D_0(y) - D_1(y) \tag{B14}$$

Assume $\delta > 0$; an analogous proof holds for $\delta < 0$. Plugging in the expressions for $D_0$ and $D_1$ and averaging over $x$ we obtain

$$\frac{1}{6} \leq \frac{1}{N} \sum_x \sum_{y \in S_x} \frac{|\alpha_y^0|^{2+\delta}}{\sum_y |\alpha_{y'}^0|^{2+\delta}} - \frac{|\alpha_y^x|^{2+\delta}}{\sum_y |\alpha_{y'}^x|^{2+\delta}} \tag{B15}$$

$$\leq \frac{1}{N} \sum_x \sum_{y \in S_x} M^{\delta/2} |\alpha_y^0|^{2+\delta} - |\alpha_y^x|^{2+\delta} \tag{B16}$$

$$= \frac{1}{N} \sum_x \sum_{y \in S_x} \left(1 + \frac{\delta}{2}\log(M)\right) |\alpha_y^0|^2 \left(1 + \delta \log|\alpha_y^0|\right) - |\alpha_y^x|^2 \left(1 + \delta \log|\alpha_y^x|\right) + O\left(\delta^2\right) \tag{B17}$$

$$= \frac{1}{N} \sum_x \sum_{y \in S_x} \left(|\alpha_y^0|^2 - |\alpha_y^x|^2\right) + \frac{\delta}{2}\log(M)|\alpha_y^0|^2 + \frac{\delta}{2}\left(|\alpha_y^0|^2 \log|\alpha_y^0|^2 - |\alpha_y^x|^2 \log|\alpha_y^x|^2\right) + O\left(\delta^2\right) \tag{B18}$$

$$\leq \delta \log(M) + O(\delta^2) + \frac{1}{N} \sum_x \sum_{y \in S_x} \left(|\alpha_y^0|^2 - |\alpha_y^x|^2\right). \tag{B19}$$

In line (B16) we used the fact that $M^{-\delta/2} \leq \sum_y |\alpha_y|^{2+\delta} \leq 1$ for any state $\alpha_y$ normalized in the $\ell_2$ norm, in line (B17) we Taylor expanded to first order in $\delta$, and in line (B19) we used the fact that the sum in the second term is upper bounded by one and the sum on third term is a difference of entropies of (subnormalized) probability distributions which is at most $\log(M)$.

We next consider the final term

$$R \equiv \frac{1}{N} \sum_x \sum_{y \in S_x} \left( |\alpha_y^0|^2 - |\alpha_y^x|^2 \right). \tag{B20}$$

Let $\hat{S}_x$ be the observable

$$\hat{S}_x = \sum_{y \in S_x} |y\rangle\langle y|. \tag{B21}$$

Then

$$\begin{aligned} R &= \frac{1}{N} \sum_x \left[ \langle \psi^0 | \hat{S}_x | \psi^0 \rangle - \langle \psi^x | \hat{S}_x | \psi^x \rangle \right] && \text{(B22)} \\ &= \frac{1}{N} \sum_x \left[ \left( \langle \psi^0 | - \langle \psi^x | \right) \hat{S}_x | \psi^0 \rangle + \langle \psi^x | \hat{S}_x \left( |\psi^0\rangle - |\psi^x\rangle \right) \right] && \text{(B23)} \\ &\leq \frac{2}{N} \sum_x \left\| |\psi^0\rangle - |\psi^x\rangle \right\|, && \text{(B24)} \end{aligned}$$

where the last inequality uses the fact that $\|\hat{S}_x\| = 1$. Next we note that $\sum_x \left\| |\psi^0\rangle - |\psi^x\rangle \right\|$ is the $\ell_1$ norm of the $N$-dimensional vector whose $x^{\text{th}}$ component is $\left\| |\psi^0\rangle - |\psi^x\rangle \right\|$. For any $N$-dimensional vector $\vec{v}$, $\|v\|_1 \leq \sqrt{N} \|\vec{v}\|_2$. Thus,

$$R \leq \frac{2}{\sqrt{N}} \sqrt{\sum_x \left\| |\psi^0\rangle - |\psi^x\rangle \right\|^2}. \tag{B25}$$

As shown in [13], a unitary search algorithm using $Q$ oracle queries yields

$$\sum_x \left\| |\psi^0\rangle - |\psi^x\rangle \right\|^2 \leq 4Q^2. \tag{B26}$$

Together, (B25) and (B26) imply

$$R \leq \frac{2Q}{\sqrt{N}}. \tag{B27}$$

Now, (B27) bounds the last term in (B19) yielding our final result.

$$\frac{1}{6} \leq \delta \log(M) + O(\delta^2) + \frac{2Q}{\sqrt{N}}. \tag{B28}$$

□

The following Corollary follows immediately from Theorem 9 and Theorem 7.

**Corollary 3.** *Suppose that the Born rule is modified such that $f(\alpha) = |\alpha|^{2+\delta}$, and one can search a list of $N = 2^n$ items using $m$ qubits and $Q$ queries. Then to first order in $\delta$, we have*

$$|\delta| \geq \frac{1}{m} \left( \frac{1}{6} - \frac{2Q}{\sqrt{N}} \right).$$

*In particular, if one can search an $N$ element list with $Q \leq \sqrt{N}/24$ queries on a state of $m$ qubits, then $|\delta| \geq \frac{1}{12m}$, and hence by Theorem 7 one can send superluminal signals with probability 2/3 using $O(m)$ qubits.*

In contrast, Grover's algorithm uses $\frac{\pi}{4}\sqrt{N}$ queries to solve search, which is optimal [21]. So Corollary 3 shows that if one can achieve even a modest factor of ($6\pi \approx 19$) speedup over Grover search using $m$ qubits, then one can send superluminal signals using $O(m)$ qubits.

## SECTION C: CLONING OF QUANTUM STATES

One way to modify quantum mechanics is to allow perfect copying of quantum information, or "cloning". As a minimal example, we will here introduce the ability to do perfect single-qubit cloning. As with nonlinear dynamics, care must be taken to formulate a version of quantum cloning that is actually well defined. It is clear that perfect single qubit cloning should take $|\psi\rangle \mapsto |\psi\rangle \otimes |\psi\rangle$ for any single-qubit pure state. The nontrivial task is to define the behavior of the cloner on qubits that are entangled. It is tempting to simply define cloning in terms of the Schmidt decomposition of the entangled state. That is, applying the cloner to qubit $B$ induces the map $\sum_i \lambda_i |i_A\rangle|i_B\rangle \mapsto \sum_i \lambda_i |i_A\rangle|i_B\rangle|i_B\rangle$. However, this prescription is ill-defined due to the non-uniqueness of Schmidt decompositions. The two decompositions of the EPR pair given in (F2) and (F3) provide an example of the inconsistency of the above definition.

Instead, we define our single-qubit cloner as follows.

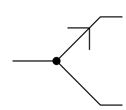
**Definition 1.** *Let $\rho_{AB}$ be a state on a bipartite system $AB$. Let $\rho_B$ be the reduced density matrix of $B$. Then applying the cloner to $B$ yields*

$$\rho_{AB} \mapsto \rho_{AB} \otimes \rho_B.$$

In particular, for pure input, we have $|\psi_{AB}\rangle\langle\psi_{AB}| \mapsto |\psi_{AB}\rangle\langle\psi_{AB}| \otimes \rho_B$. Thus, this version of cloning maps pure states to mixed states in general. Furthermore, the clones are asymmetric. The cloner takes one qubit as input and produces two qubits as output. The two output qubits have identical reduced density matrices. However, one of the output qubits retains all the entanglement that the input qubit had with other systems, whereas the other qubit is unentangled with anything else. By monogamy of entanglement it is impossible for both outputs to retain the entanglement that the input qubit had.

It is worth noting that the addition of nonlinear dynamics, and cloning in particular, breaks the equivalence between density matrices and probabilistic ensembles of pure states. Here, we take density matrices as the fundamental objects in terms of which our generalized quantum mechanics is defined.

In analyzing a model of computation involving cloning, we will treat the cloning operation as an additional gate, with the same "cost" as any other. In circuit diagrams, we denote the cloning gate as follows.



This notation reflects the asymmetric nature of our cloning gate; the arrow indicates the output qubit that retains the entanglement of the input qubit.

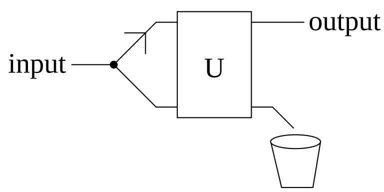
### 1. Grover Search using Quantum Cloning

Cloning is a nonlinear map on quantum states. As argued by Abrams and Lloyd [11], one can solve Grover search on a database of size $N$ using $O(1)$ oracle queries and $O(\log N)$ applications of $S$, for any nonlinear map $S$ from pure states to pure states, except perhaps some pathological cases. Here, with theorem 11, we have formalized this further, showing that this holds as long as $S$ is differentiable. However, the cloning gate considered here maps pure states to mixed states. Therefore, this gate requires a separate analysis. We cannot simply invoke theorem 11. Instead we specifically analyze the cloning gate given above and arrive at the following result.

**Theorem 10.** *Suppose we have access to a standard Grover bit-flip oracle, which acts as $U_f|y\rangle|z\rangle = |y\rangle|z \oplus f(y)\rangle$ where $f : \{0,1\}^n \to \{0,1\}$. Using one query to this oracle, followed by a circuit using* poly($n$)

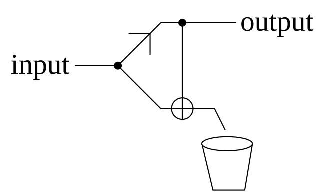
*conventional quantum gates and $O(n)$ of the single-qubit cloning gates described in definition 1, one can distinguish between the cases that $|f^{-1}(1)| = 0$ and $|f^{-1}(1)| = 1$ with high probability.*

*Proof.* For the design of nonlinear Grover search algorithms it is helpful to have a nonlinear map from a fixed state space to itself. To this end, we consider circuits of the following form, which implement nonlinear maps from the space of possible density matrices of a qubit to itself.




Here, one clones the input qubit, performs some unitary $U$ between the two resulting copies, and lastly discards one of the qubits.

With a small amount of trial and error one can find a choice of $U$ which enables single-query Grover search using an analogue of the Abrams-Lloyd algorithm. Specifically, we choose $U$ to be the controlled-not gate. That is, let

$\mathcal{M} =$ 


.

$\mathcal{M}$ is a quadratic map on density matrices. By direct calculation

$$\mathcal{M}\left(\begin{bmatrix} r_{00} & r_{01} \\ r_{10} & r_{11} \end{bmatrix}\right) = \begin{bmatrix} r_{00}^2 + r_{00}r_{11} & r_{01}^2 + r_{01}r_{10} \\ r_{10}^2 + r_{10}r_{01} & r_{11}^2 + r_{11}r_{00} \end{bmatrix}. \tag{C1}$$

One can find the fixed points of $\mathcal{M}$ by solving the system of four quadratic equations implied by $\mathcal{M}(\rho) = \rho$. The solutions are as follows.

$$r_{10} = 1 - r_{01}, \quad r_{11} = 1 - r_{00} \tag{C2}$$

$$r_{00} = 0, \quad r_{10} = 1 - r_{01}, \quad r_{11} = 0 \tag{C3}$$

$$r_{01} = 1, \quad r_{10} = 0, \quad r_{11} = 1 - r_{00} \tag{C4}$$

$$r_{00} = r_{01} = r_{10} = r_{11} = 0 \tag{C5}$$

The solutions (C3) and (C5) are traceless and therefore unphysical. Solution (C4) is an arbitrary mixture of $|0\rangle$ and $|1\rangle$. That is,

$$\rho_r = \begin{bmatrix} r & 0 \\ 0 & 1-r \end{bmatrix} \quad \text{is a fixed point for all } r \in [0,1]. \tag{C6}$$

As a matrix, the solution (C2) is

$$\rho_{a,b} = \begin{bmatrix} a & b \\ 1-b & 1-a \end{bmatrix}. \tag{C7}$$

This is only Hermitian if $b = (1-b)^*$, which implies that $b = \frac{1}{2} + \alpha i$ for some $\alpha \in \mathbb{R}$. However, if $\alpha \neq 0$ then $\rho_{a,b}$ fails to be positive semidefinite, which is unphysical. Thus, $b = \frac{1}{2}$. The eigenvalues of $\rho_{a,1/2}$ are

$$\frac{1}{2} \pm \sqrt{\frac{1 + 2a(a-1)}{2}}. \tag{C8}$$

Thus, unless $a = \frac{1}{2}$, the largest eigenvalue of $\rho_{a,1/2}$ exceeds one, which is unphysical. So, the only physical fixed point other than $\rho_r$ is

$$\rho_+ = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = |+\rangle\langle+|. \tag{C9}$$

Numerically, one finds that $\rho_r$ is an attractive fixed point and $\rho_+$ is a repulsive fixed point. Let

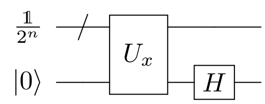
$$\rho_\epsilon = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} - \epsilon \\ \frac{1}{2} - \epsilon & \frac{1}{2} \end{bmatrix} = (1-\epsilon)|+\rangle\langle+| + \epsilon|-\rangle\langle-|. \tag{C10}$$

Then,

$$\mathcal{M}(\rho_\epsilon) = \rho_{2\epsilon + O(\epsilon^2)}. \tag{C11}$$

Consequently, $\mathcal{M}^r(\rho_\epsilon)$ is easily distinguishable from $\mathcal{M}^r(\rho_+) = \rho_+$ after $r = O(\log(1/\epsilon))$ iterations of $\mathcal{M}$.

Let $U_f$ be the standard Grover bit-flip oracle, which acts as $U_f|y\rangle|z\rangle = |y\rangle|z \oplus f(y)\rangle$ where $f : \{0,1\}^n \to \{0,1\}$. Now, consider the following circuit.




One sees that the bottom qubit emerges in the state $\rho_+$ if $f$ has no solution and emerges in the state $\rho_\epsilon$ with $\epsilon = \frac{1}{2^n}$ if $f$ has one solution. By making one such query and then applying the map $\mathcal{M}$ a total of $O(n)$ times to the resulting state, one obtains single-qubit states in the no-solution and one-solution cases that are easily distinguished with high confidence using conventional quantum measurements. □

For simplicity, in theorem 10, we have restricted our attention to search problems which are promised to have exactly one solution or no solutions and our task is to determine which of these is the case. Note that 3SAT can be reduced to UNIQUESAT in randomized polynomial time [22]. Hence solving the Grover problem in poly(n) time when there is either exactly one solution or no solutions suffices to solve NP-hard problems in randomized polynomial time.

It is interesting to note that probability distributions also cannot be cloned. The map $\vec{p} \mapsto \vec{p} \otimes \vec{p}$ on vectors of probabilities is nonlinear and hence does not correspond to any realizable stochastic process. Furthermore, one finds by a construction similar to the above that cloning of classical probability distributions also formally implies polynomial-time solution to NP-hard problems via logarithmic-complexity single-query Grover search. However, nonlinear maps on probabilities do not appear to be genuinely well-defined. Suppose we have probability $p_1$ of drawing from distribution $\vec{p_1}$ and probability $p_2$ of drawing from distribution $\vec{p_2}$. Normally this is equivalent to drawing from $p_1\vec{p_1} + p_2\vec{p_2}$. However, if we apply a nonlinear map $\mathcal{M}$ then $\mathcal{M}(p_1\vec{p_1} + p_2\vec{p_2})$ is in general not equal to $p_1\mathcal{M}(p_1) + p_2\mathcal{M}(\vec{p_2})$. It is not clear that a well-defined self-consistent principle can be devised for resolving such ambiguities.

### 2. Superluminal Signaling using Quantum Cloning

Suppose Alice and Bob share an EPR pair $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. If Alice wishes to transmit a zero she does nothing. If she wishes to transmit a one she measures her qubit in the computational basis. If Alice doesn't measure then Bob's reduced density matrix is maximally mixed. Hence if he makes several clones and measures them all in the computational basis he will obtain a uniformly random string of ones and zeros. If Alice does measure then Bob's reduced density matrix is either $|0\rangle\langle 0|$ or $|1\rangle\langle 1|$, with equal probability. If he makes several clones and measures them all in the computational basis he will get $000\ldots$ or $111\ldots$, with equal probability. Thus, by making logarithmically many clones, Bob can achieve polynomial certainty about the bit that Alice wished to transmit.

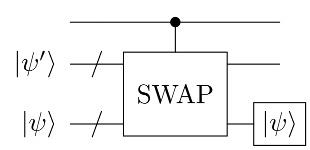
## SECTION D: POSTSELECTION

In [19] it was shown that adding the ability to postselect a single qubit onto the state $|0\rangle$ to the quantum circuit model yields a model of computation whose power is equal to the classical complexity class PP. Furthermore, postselection onto $|0\rangle$ allows perfect superluminal signaling by postselected quantum teleportation. Here we consider a more general question: suppose we have the ability to postselect on some arbitrary but fixed $n$-qubit state $|\psi\rangle$. Does this still yield efficient means of solving problems in PP and sending superluminal signals? It is clear that one can use postselection onto $|\psi\rangle$ to simulate postselection onto $|0\rangle$ given a quantum circuit for a unitary $U$ such that $U|00\ldots 0\rangle = |\psi\rangle$. However, for a generic $n$-qubit state $|\psi\rangle$, no polynomial-size quantum circuit for this tasks exists. Nevertheless, in this section we show that, for Haar random (but fixed) $|\psi\rangle$, postselection onto $|\psi\rangle$ can with high probability be used to simulate postselection onto $|0\rangle$ with exponential precision.

We first note that the maximally entangled state of $2n$ qubits:

$$|\Phi_{2n}\rangle = \sum_{x \in \{0,1\}^n} |x\rangle \otimes |x\rangle \tag{D1}$$

can be prepared using $n$ Hadamard gates followed by $n$ CNOT gates. Postselecting the second tensor factor of $|\Phi_{2n}\rangle$ onto $|\psi\rangle$ yields $|\psi\rangle$ on the first tensor factor. In this manner, one may extract a copy of $|\psi\rangle$. We assume that $|\psi\rangle$ is Haar random but fixed. That is, each time one uses the postselection "gate," one postselects onto the same state $|\psi\rangle$. Hence, using the above procedure twice yields two copies of $|\psi\rangle$. Applying $\sigma_x$ to one qubit of one of the copies of $|\psi\rangle$ yields a state $|\psi'\rangle = \sigma_x|\psi\rangle$. As shown below, the root-mean-square inner product between $|\psi\rangle$ and $|\psi'\rangle$ is of order $1/\sqrt{2^n}$. That is, they are nearly orthogonal. Thus, one can simulate postselection onto $|0\rangle$ with the following circuit.




Here, the top qubit gets postselected onto $|0\rangle$ with fidelity $1 - O(1/\sqrt{2^n})$, the middle register is discarded, and the bottom register is postselected onto $|\psi\rangle$, an operation we denote by —$\boxed{|\psi\rangle}$—.

Lastly, we prove the claim that the root-mean-square inner product between a Haar random $n$-qubit state $|\psi\rangle$ and $|\psi'\rangle = \sigma_x|\psi\rangle$ is of order $1/\sqrt{2^n}$. This mean-square inner product can be written as

$$\bar{I} = \int dU |\langle 0\ldots 0|U^\dagger \sigma_x U|0\ldots 0\rangle|^2 \tag{D2}$$

$$= \sum_{a,b \in \{0,1\}^n} \int dU U^\dagger_{0a} U_{\bar{a}0} U^\dagger_{0b} U_{\bar{b}0} \tag{D3}$$

where $\bar{a}$ indicates the result of flipping the first bit of $a$, $\bar{b}$ indicates the result of flipping the first bit of $b$, and 0 in the subscripts is shorthand for the bit string $0\ldots 0$. (We arbitrarily choose the $\sigma_x$ to act on the first qubit.)

Next we recall the following identity regarding integrals on the Haar measure over $U(N)$. (See [23] or appendix D of [24].)

$$\begin{aligned} \int dU\ U_{ij} U_{kl} U^\dagger_{mn} U^\dagger_{op} \ = \ & \tfrac{1}{N^2-1} \left(\delta_{in}\delta_{kp}\delta_{jm}\delta_{lo} + \delta_{ip}\delta_{kn}\delta_{jo}\delta_{lm}\right) \\ & - \tfrac{1}{N(N^2-1)} \left(\delta_{ij}\delta_{kp}\delta_{jo}\delta_{lm} + \delta_{ip}\delta_{kn}\delta_{jm}\delta_{lo}\right) \end{aligned} \tag{D4}$$

Applying (D4) to (D3) shows that the only nonzero terms come from $a = \bar{b}$ and consequently

$$\bar{I} = \sum_{a \in \{0,1\}^n} \int dU \; U_{\bar{a}0} U_{a0} U^\dagger_{0a} U^\dagger_{0\bar{a}} \tag{D5}$$

$$= \frac{N}{N^2 - 1} - \frac{1}{N^2 - 1}. \tag{D6}$$

Consequently, the RMS inner product for large $N$ is

$$\sqrt{\bar{I}} \simeq \frac{1}{\sqrt{N}}. \tag{D7}$$

Recalling that $N = 2^n$ completes the argument.

## SECTION E: GENERAL NONLINEARITIES

Our discussion of final-state projection models can be thought of as falling within a larger tradition of studying the information-theoretic and computational complexity implications of nonlinear quantum mechanics, as exemplified by [11, 25–27]. A question within this subject that has been raised multiple times [11, 25] is whether all nonlinearities necessarily imply that Grover search can be solved with a single query. In this note we shed some light on this question. However, note that the setting differs from that of section A 3 of Section B in that (following [11, 25]) we assume the nonlinear map is the same each time, and we can apply it polynomially many times. In section A 3 of Section B we have included the possibility that black holes (and the nonlinear maps that they generate) are scarce and that they may differ from one another.

We first note that, for dynamics that map normalized pure states to normalized pure states, the terms nonunitary and nonlinear are essentially interchangeable. Let $V$ be the manifold of normalized vectors on a complex Hilbert space $\mathcal{H}$, which could be finite-dimensional or infinite-dimensional. Let $S : V \to V$ be a general map, not necessarily linear or even continuous. We'll call $S$ a *unitary map* if it preserves the magnitude of inner products. That is, $|\langle S\psi | S\phi \rangle| = |\langle \psi | \phi \rangle|$ for all $|\phi\rangle, |\psi\rangle \in \mathcal{H}$. Wigner's theorem [28] states that all unitary maps are either unitary linear transformations, or antiunitary antilinear transformations. (Antiunitary transformations are equivalent to unitary transformations followed by complex conjugation of all amplitudes in some basis.) Extending quantum mechanics by allowing antiunitary dynamics does not affect computational complexity, as can be deduced from [29]. Thus, without loss of generality, we may ignore antiunitary maps. Hence, within the present context, if a map is unitary then it is linear. Conversely, by linear algebra, if map $S$ is linear, and maps $V \to V$, *i.e.* is norm-preserving, then it is also inner-product preserving, *i.e.* unitary.

A standard version of the Grover problem is, for some function $f : \{0,1\}^n \to \{0,1\}$, to decide whether the number of solutions to $f(y) = 1$ is zero or one, given that one of these is the case. The search problem of finding a solution is reducible to this decision problem with logarithmic overhead via binary search. In [11] Abrams and Lloyd show how to solve the decision version of Grover search using a single quantum query to $f$ and $O(n)$ applications of a single-qubit nonlinear map. This suffices to solve $\mathsf{NP}$ in polynomial time. We now briefly describe their algorithm. In contrast to section A 3 of Section B, it is more convenient here to assume a bit-flip oracle rather than a phase-flip oracle. That is, for $y \in \{0,1\}^n$ and $z \in \{0,1\}$ the oracle $O_f$ acts as

$$O_f |y\rangle|z\rangle = |y\rangle|z \oplus f(y)\rangle. \tag{E1}$$

Querying the oracle with the state $\frac{1}{\sqrt{2^n}} \sum_{y \in \{0,1\}^n} |y\rangle|0\rangle$ yields $\frac{1}{\sqrt{2^n}} \sum_{y \in \{0,1\}^n} |y\rangle|f(y)\rangle$. Applying a Hadamard gate to each qubit of the first register and measuring the first register in the computational basis yields the outcome $00\ldots0$ with probability at least $\frac{1}{4}$. Given that this occurs, the post-measurement state of the second register is

$$|\psi_s\rangle = \frac{(2^n - s)|0\rangle + s|1\rangle}{\sqrt{(2^n - s)^2 + s^2}}, \tag{E2}$$

where $s$ is the number of solutions, *i.e.* $s = |f^{-1}(1)|$. Thus, we can solve the Grover search problem by distinguishing two exponentially-close states, namely $|\psi_0\rangle$ and $|\psi_1\rangle$. For the particular nonlinear map on the manifold of normalized pure single-qubit states considered in [11], a pair of states $\epsilon$-close together can be separated to constant distance by iterating the map $O(\log(1/\epsilon))$ times.

We now show that any differentiable nonlinear map from pure states to pure states on any finite-dimensional Hilbert space can achieve this. (See theorem 11.) Let $V^{(n)}$ be the manifold of normalized pure states on $\mathbb{C}^n$. Thus, $V^{(n)}$ is a $2n-1$ dimensional real closed compact manifold. For points $a, b$ on $V^{(n)}$ let $|a-b|$ denote their distance. (Our choice of distance metric is not important to the argument, but for concreteness, we could choose the angle between quantum states, that is, $|a-b| = \cos^{-1} |\langle a|b\rangle|$. That this is a metric is proven in section 9.2.2 of [14].)

**Theorem 11.** *Let $S : V^{(n)} \to V^{(n)}$ be a differentiable map, that is, a self-diffeomorphism of $V^{(n)}$. Let $r = \max_{a,b\in V^{(n)}} \frac{|S(a)-S(b)|}{|a-b|}$. Then there exists some sufficiently short geodesic $l$ in $V^{(n)}$ such that for all $x, y \in l$, $\frac{|S(x)-S(y)|}{|x-y|} \geq r$.*

*Proof.* Choose two points $x, y$ on $V^{(n)}$ that maximize the ratio $r = \frac{|S(x)-S(y)|}{|x-y|}$. By assumption, $S$ is not unitary, so not all distances are preserved. Because $S$ is a map from $V^{(n)}$ to another manifold of equal volume (namely $V^{(n)}$ itself) it cannot be that all distances are decreased. Thus, this maximum ratio must be larger than one. The extent that this ratio exceeds one quantifies the deviation from unitarity.

Now, consider the geodesic $g$ on $V^{(n)}$ from $x$ to $y$. Because it is a geodesic, $g$ has length $|x-y|$. Now consider the image of $g$ under the map $S$. Because $S$ is a continuous map, $S(g)$ will also be a line segment. By the construction, the endpoints of $S(g)$ are distance $r|x-y|$ apart. Therefore, the length of $S(g)$, which we denote $|S(g)|$, satisfies $|S(g)| \geq r|x-y|$, with equality if $S(g)$ happens to also be a geodesic. Thus, $S$ induces a diffeomorphism $S_g$ from the line segment $g$ to the line segment $S(g)$, where $|S(g)|/|g| \geq r$. Because $S_g$ is a diffeomorphism it follows that on any sufficiently small subsegment of $g$ it acts by linearly magnifying or shrinking the subsegment and translating to some location on $S(g)$. Because $|S(g)|/|g| \geq r$ it follows that there exists some subsegment $l$ such that this linear magnification is by a factor of at least $r$. (There could be some subsegments that grow less than this or even shrink, but if so, others have to make up for it by growing by a factor of more $|S(g)|/|g|$.) □

We now argue that the existence of $l$ suffices to ensure success for the Abrams-Lloyd algorithm. Let $f$ denote the "magnification factor" that $S$ induces on $l$. According to theorem 11, $f \geq r$. We are interested in asymptotic complexity, so the distance $\epsilon$ between $|\psi_0\rangle$ and $|\psi_1\rangle$ is asymptotically small. Therefore, we assume $\epsilon$ is smaller than the length of $l$. So, we can append ancilla qubits and apply a unitary transformation such that the resulting isometry maps $|\psi_0\rangle$ and $|\psi_1\rangle$ to two points $\left|\phi_0^{(0)}\right\rangle$ and $\left|\phi_1^{(0)}\right\rangle$ that lie on $l$. We then apply $S$, resulting in the states $\left|\phi_0^{(1)}\right\rangle$ or $\left|\phi_1^{(1)}\right\rangle$, which have distance $f\epsilon$. If $f\epsilon$ is larger than the length $l$ then we terminate. Because we have a fixed nonunitary map, the distance between our states is now a constant (independent of $\epsilon$ and hence of the size of the search space). If $f\epsilon$ is smaller than the length of $l$, then we apply a unitary map that takes $\left|\phi_0^{(1)}\right\rangle$ and $\left|\phi_1^{(1)}\right\rangle$ back onto $l$ and apply $S$ again. We then have states $\left|\phi_0^{(2)}\right\rangle$ and $\left|\phi_1^{(2)}\right\rangle$ separated by distance $f^2\epsilon$. We then iterate this process until we exceed the size $l$, which separates the states to a constant distance and uses $\log_f(1/\epsilon)$ of the nonunitary operations. States with constant separation can be distinguished within standard quantum mechanics by preparing a constant number of copies and collecting statistics on the outcomes of ordinary projective measurements.

## SECTION F: A CAUTIONARY NOTE ON NONLINEAR QUANTUM MECHANICS

The Horowitz-Maldecena final-state projection model, cloning of quantum states, and the Gross-Pitaevsky equation (if interpreted as a quantum wave equation) all involve nonlinear dynamics of the wavefunction. In such cases, one must be very careful to ensure that subsystem structure, which is captured by tensor product structure in conventional quantum mechanics, is well-defined. Indeed, subsystem structure is lost by introducing generic nonlinearities, and in particular by the nonlinearity of the Gross-Pitaevsky equation. This makes the question about superluminal signaling in the Gross-Pitaevsky model ill-posed. The Horowitz-Maldecena model does have a natural notion of subsystem structure, which is one of the features that makes it appealing. Furthermore, the model of cloning that we formulate in Section D preserves subsystem structure by virtue of being phrased in terms of reduced density matrices.

More formally, let $V$ be the manifold of normalized vectors in the Hilbert space $\mathbb{C}^d$. We will model nonlinear quantum dynamics by some map $S: V \to V$ which may not be a linear map on $\mathbb{C}^d$. In general, specifying a map $S$ on $V$ does not uniquely determine the action of $S$ when applied to a subsystem of a larger Hilbert space. For example, consider the map $S_0$ on the normalized pure states of one qubit given by

$$S_0|\psi\rangle = |0\rangle \quad \forall|\psi\rangle \tag{F1}$$

Now, consider what happens if we apply $S_0$ to half of an EPR pair $|\Psi_{\text{EPR}}\rangle$. We can write the EPR state in two equivalent ways

$$|\Psi_{\text{EPR}}\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle|0\rangle + |1\rangle|1\rangle\right) \tag{F2}$$

$$= \frac{1}{\sqrt{2}}\left(|+\rangle|+\rangle + |-\rangle|-\rangle\right) \tag{F3}$$

where

$$|\pm\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle \pm |1\rangle\right) \tag{F4}$$

Symbolically applying the rule $S_0|\psi\rangle \mapsto |0\rangle$ to the first tensor factor of (F2) yields $|0\rangle|+\rangle$, whereas applying this rule to the first tensor factor of $(F3)$ yields $|0\rangle|0\rangle$.

This example illustrates that one must specify additional information beyond the action of a nonlinear map on a fixed Hilbert space in order to obtain a well-defined extension to quantum theory incorporating the notion of subsystems.

## SECTION G: OPEN PROBLEMS

We have shown that in several domains of modifications of quantum mechanics, the resources required to observe superluminal signaling or a speedup over Grover's algorithm are polynomially related. We extrapolate that this relationship holds more generally, that is, in any quantum-like theory, the Grover lower bound is derivable from the no-signaling principle and vice-versa. A further hint in this direction is that, as shown in [30], the limit on distinguishing non-orthogonal states in quantum mechanics is dictated by the no-signaling principle. Thus, any improvement over the Grover lower bound based on beyond-quantum state discrimination can be expected to imply some nonzero capacity for superluminal signaling. There is a substantial literature on generalizations of quantum mechanics which could be drawn upon to address this question. In particular, one could consider the generalized probabilistic theories framework of Barrett [31], the category-theoretic framework of Abramsky and Coecke [32], the Newton-Schrödinger equation [33], quaternionic quantum mechanics [34], or the Papadodimas-Raju state-dependence model of black hole dynamics [17, 18, 35]. In these cases the investigation of computational and communication properties is inseparably tied with the fundamental questions about the physical interpretations of these models. Possibly, such investigation could help shed light on these fundamental questions.

Our finding can be regarded as evidence against the possibility of using black hole dynamics to efficiently solve NP-complete problems, at least for problem instances of reasonable size. Note however that there are other independent questions regarding the feasibility of computational advantage through final-state projection and other forms of non-unitary quantum mechanics. In particular, the issue of fault-tolerance in modified quantum mechanics remains largely open, although some discussion of this issue appears in [11, 25, 36]. Also, while our results focus on the query complexity of search, in practice one also is interested in the time complexity. Harlow and Hayden [37] have argued that decoding the Hawking radiation emitted by a black hole may require exponential time on a quantum computer. If the Harlow-Hayden argument is correct, then exponential improvement in query complexity for search does not imply exponential improvement in time-complexity. We emphasize however that query complexity sets a lower bound on time complexity, and therefore the reverse implication still holds, namely exponential improvement in time complexity implies exponential improvement in query complexity, which in the models we considered implies superluminal signaling. Hence an operational version of the Grover lower bound can be derived from an operation version of the no-signaling principle.

---


[1] Almheiri, A., Marolf, D., Polchinski, J. & Sully, J. Black holes: Complementarity or firewalls? *Journal of High Energy Physics* **1302**, 062 (2013). ArXiv:1207.3123.

[2] Horowitz, G. T. & Maldecena, J. The black hole final state. *Journal of High Energy Physics* **0402**, 008 (2004). ArXiv:hep-th/0310281.

[3] Gottesman, D. & Preskill, J. Comment on "The black hole final state". *Journal of High Energy Physics* **0403**, 026 (2004). ArXiv:hep-th/0311269.

[4] Some interpret the final state projection model as inducing a unitary map for observers who stay outside the event horizon, while inducing a non-unitary map for infalling observers [38]. Under this interpretation, our arguments still apply to an infalling observer in the context of a large black hole.

[5] Grover, L. K. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing (STOC)*, 212–219 (1996). ArXiv:quant-ph/9605043.

[6] By a "super-Grover speedup", we mean an algorithm which searches an unstructured $N$-element list using fewer queries than Grover's algorithm.

[7] Leslie, M. `http://mathoverflow.net/questions/96493` (2012).

[8] That is, Alice drops her half of the shared state into the black hole, and waits for the black hole to evaporate.

[9] Elsewhere we have used density matrices, but only after the application of the nonlinear operation.

[10] An alternate definition is to use a bit-flip oracle $U_x|y\rangle|z\rangle = |y\rangle|z \oplus f(y)\rangle$ where $f(y) = 1$ if $y = x$ and $f(y) = 0$ otherwise. This choice is irrelevant since the phase flip oracle can be simulated using a bit-flip oracle if the output register is initialized to $(|0\rangle - |1\rangle)/\sqrt{2}$, and a bit flip oracle can be simulated using a controlled-phase-flip oracle.

[11] Abrams, D. S. & Lloyd, S. Nonlinear quantum mechanics implies polynomial-time solution for NP-complete and #P problems. *Physical Review Letters* **81**, 3992 (1998). ArXiv:quant-ph/9801041.

[12] Our notation is based on the exposition of this proof given in [14].

[13] Bennett, C. H., Bernstein, E., Brassard, G. & Vazirani, U. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing* **26**, 1510–1523 (1997).

[14] Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).

[15] Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (Wiley, Hoboken, New Jersey, 1991).

[16] We have used $H[p]$ to denote the entropy of a probability distribution $p$ and $H(R)$ to denote the entropy of a random variable $R$.

[17] Papadodimas, K. & Raju, S. State-dependent bulk-boundary maps and black hole complementarity. *Physical Review D* **89**, 086010 (2014). ArXiv:1310.6335.

[18] Marolf, D. & Polchinski, J. Violations of the Born rule in cool state-dependent horizons (2015). ArXiv:1506.01337.

[19] Aaronson, S. Quantum computing, postselection, and probabilistic polynomial-time. *Proceedings of the Royal Society A* **461**, 3473 (2005). ArXiv:quant-ph/0412187.

[20] Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters* **70**, 1895 (1993).

[21] Zalka, C. Grover's quantum searching algorithm is optimal. *Physical Review A* **60**, 2746–2751 (1999). ArXiv:quant-ph/9711070.

[22] Valiant, L. & Vazirani, U. NP is as easy as detecting unique solutions. *Theoretical Computer Science* **47**, 85–93 (1986).

[23] Collins, B. & Sniady, P. Integration with respect to the Haar measure on unitary, orthogonal, and symplectic group. *Communications in Mathematical Physics* **264**, 773–795 (2006).

[24] Harlow, D. Jerusalem lectures on black holes and quantum information (2014). ArXiv:1409.1231.

[25] Aaronson, S. NP-complete problems and physical reality. *ACM SIGACT News* **36**, 30–52 (2005). ArXiv:quant-ph/0502072.

[26] Meyer, D. A. & Wong, T. G. Nonlinear quantum search using the Gross-Pitaevskii equation. *New Journal of Physics* **15**, 063014 (2013). ArXiv:1303.0371.

[27] Childs, A. M. & Young, J. Optimal state discrimination and unstructured search in nonlinear quantum mechanics (2015). ArXiv:1507.06334.

[28] Wigner, E. P. *Gruppentheorie und ihre Anwendung auf die Quanten mechanik der Atomspektren*, 251–254 (Friedrich Vieweg und Sohn, 1931).

[29] Aharonov, D. A simple proof that Toffoli and Hadamard are quantum universal (2003). ArXiv:quant-ph/0301040.

[30] Bae, J., Hwang, W.-Y. & Han, Y.-D. No-signaling principle can determine optimal quantum state discrimination. *Physical Review Letters* **107**, 170403 (2011). ArXiv:1102.0361.

[31] Barrett, J. Information processing in generalized probabilistic theories. *Physical Review A* **75**, 032304 (2005). ArXiv:quant-ph/0508211.

[32] Abramsky, S. & Coecke, B. Categorical quantum mechanics. In Engesser, K., Gabbay, D. & Lehmann, D. (eds.) *Handbook of Quantum Logic and Quantum Structures*, vol. 2, 261–325 (Elsevier, 2008). ArXiv:0808.1023.

[33] Ruffini, R. & Bonazzola, S. Systems of self-gravitating particles in general relativity and the concept of an equation of state. *Physical Review* **187**, 1767 (1969).

[34] Adler, S. L. *Quaternionic Quantum Mechanics and Quantum Fields* (Oxford University Press, Oxford, 1995).

[35] Harlow, D. Aspects of the Papadodimas-Raju proposal for the black hole interior. *Journal of High Energy Physics* 1411 (2014). ArXiv:1405.1995.

[36] Brun, T. Computers with closed timelike curves can solve hard problems. *Foundations of Physics Letters* **16**, 245–253 (2003). ArXiv:gr-qc/0209061.

[37] Harlow, D. & Hayden, P. Quantum computation vs. firewalls. *Journal of High Energy Physics* **1013:85** (2013). ArXiv:1301.4504.

[38] Harlow, D. Personal communication .