Entanglement improves classical control

Nguyen Truong Duy,[1,2] Matthew McKague,[3,4] and Stephanie Wehner[3,2]

[1] *Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543*
[2] *School of Computing, National University of Singapore, 13 Computing Drive, Singapore 117417*
[3] *Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2,Singapore 117543*
[4] *Department of Physics, University of Otago, Dunedin, New Zealand*

Electronic devices all around us contain classical control circuits. Such circuits consist of a network of controllers which can read and write signals to wires of the circuit with the goal to minimize the cost function of the circuit's output signal. Here, we propose the use of shared entanglement between controllers as a resource to improve the performance of otherwise purely classical control circuits. We study a well-known example from the classical control literature and demonstrate that allowing two controllers to share entanglement improves their ability to control. More precisely, we exhibit a family of circuits in which the the cost function using entanglement stays constant, but the minimal cost function without entanglement grows arbitrarily large. This demonstrates that entanglement can be a powerful resource in a classical control circuit.

Control theory has been used with great success to enable the control of *quantum* systems. Whether it comes to battling decoherence [1], or ensuring error-correction [2], it is certainly fair to say that quantum control forms a necessary tool to build quantum devices of any kind (see e.g. [3] for a recent survey). Indeed, quantum control has even been employed to study fundamental aspects of quantum mechanics itself [4]. In all of these works, the aim was to extend techniques and methods from classical control theory to gain a handle on *quantum* systems, which forms a daunting task due the possibility of coherent control in superposition and entanglement.

Here, however, we raise the opposite question, namely whether quantum effects can improve *classical* [5] control itself. More precisely, we consider a classical control circuit in which all signals are classical and the overall aim of the circuit is to solve a *classical* control problem. The only quantum ingredient will be that we allow the controllers in the circuit to share entanglement and to perform quantum measurements on the entangled state. To see whether such entanglement can at all be useful, we study the simple toy circuit depicted in Figure 1 which was proposed by Witsenhausen [6] to study aspects of classical control and is well-known under the name "Witsenhausen's counterexample" (see e.g. [7, 8] for surveys). In its original form, Witsenhausen's circuit [6] dealt with continuous signals but it was later discretized [9, 10], which is the approach we will follow here. The objective of this circuit is to take an input signal on a line and subsequently damp it down in as efficient a manner as possible. We can accomplish this using two controllers, c_1 and c_2, which can each add a signal to the line. The first controller has perfect information, but its use incurs a cost. The second controller has imperfect information, but there is no penalty for its use. The problem is made richer by giving the the controllers c_1 and c_2 access to additional resources summarized by ρ. Classically, ρ is a shared random variable, but quantumly ρ can also be a shared entangled state.

The performance of the controllers is evaluated using a cost function which grows with the remaining signal left on the line following the controllers. More precisely, in terms of the variables defined in Figure 1

$$C^{\mathcal{N},P_X,k}(c_1,c_2,\rho) =$$
$$\mathop{\mathbb{E}}_{P_X,\mathcal{N}} \left[k(c_1(x))^2 + (x + c_1(x) + c_2(s))^2 \right] \quad (1)$$

where $k \in \mathbb{R}^+$ is a parameter. A specific instance of the discrete Witsenhausen circuit is characterized by the choice of input distribution P_X, the channel $\mathcal{N}$ and the parameter k defining the cost function, i.e., by $(\mathcal{N}, P_X, k)$. Meanwhile, a specific control strategy is described by (c_1, c_2, ρ).

We consider three classes Ω of strategies. The first two are *classical* strategies. These can be deterministic strategies D, where c_1 and c_2 are functions $\mathbb{Z} \to \mathbb{Z}$ and ρ may be taken to be 0. Classically, we also allow strategies SR that employ some shared randomness ρ. In this case we have that for all $(c_1, c_2, \rho) \in SR$, c_1, c_2 are functions $\mathbb{Z} \times R \to \mathbb{Z}$ where $\rho \in R$ is the shared random variable. Finally, we allow for *quantum* strategies, where we use SE to denote this class of strategies with shared entanglement ρ. Note that the input and output of the controllers remain purely classical. For any $(c_1, c_2, \rho) \in SE$, $\rho \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is a quantum state shared between controller 1 holding system $\mathcal{H}_1$ and controller 2 holding system $\mathcal{H}_2$. Without loss of generality, c_1 performs a measurement on $\mathcal{H}_1$ indexed by its input x, and outputs the measurement outcome $m = c_1(x)$ on the wire. Formally, this means that the strategy of c_1 is defined by measurement operators $\{\{\Pi^1_{x,m} \geq 0\}_m\}_x$ where for each x we have
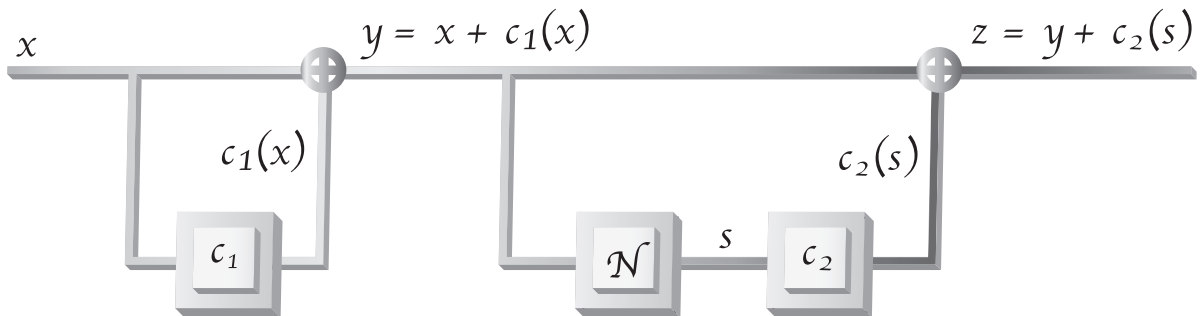
FIG. 1. Discrete Witsenhausen Circuit - The input to the circuit is a random integer $x \in \mathbb{Z}$ chosen according to a probability distribution P_X. The input is fed into a controller c_1 which outputs an integer. The output from c_1 is then added to the line, giving $y = x + c_1(x)$. At this point the second controller, c_2, receives the signal from the line. However, the signal is received via a classical noisy channel $\mathcal{N}$, characterized by the probability $\mathcal{N}(s|y)$ of outputting the symbol $s \in \mathbb{Z}$ given input y. The second controller c_2 then again outputs an integer which is added to the line, giving the final output of the circuit $z = y + c_2(s)$.

$\sum_m \Pi^1_{x,m} = I_{H_1}$. Similarly, the second controller c_2 performs a measurement $\{\Pi^2_{s,\ell}\}_\ell$ on $\mathcal{H}_2$ indexed by its input s and outputs the outcome $\ell = c_2(s)$ on the wire. Given some instance $(\mathcal{N}, P_X, k)$ we define the *minimum cost* for a class of strategies $\Omega \in \{D, SR, SE\}$ to be

$$C^{\mathcal{N}, P_X, k}_\Omega = \min_{(c_1, c_2, \rho) \in \Omega} \mathrm{C}^{\mathcal{N}, P_X, k}(c_1, c_2, \rho) . \qquad (2)$$

I. RESULT

Here, we prove that a strategy using entanglement outperforms any classical strategy for the circuit given in Figure 1 and hence entanglement improves classical control. More precisely, we show that for the generalized discrete Witsenhausen circuit, there exist a classical channel $\mathcal{N}_0$, a probability distribution P_X^0 and a constant k such that

$$C^{\mathcal{N}_0, P_X^0, k}_{SE} < C^{\mathcal{N}_0, P_X^0, k}_{SR} . \qquad (3)$$

We prove this theorem by exhibiting a particular strategy in SE and a one-parameter family of instances $(\mathcal{N}_t, P_X^t, k)$ such that $C^{\mathcal{N}_t, P_X^t, k}_{SE}$ remains constant for all t, but $C^{\mathcal{N}_t, P_X^t, k}_{SR}$ increases with t without bound. Hence for sufficiently large t our quantum strategy is better than the minimum cost for SR. What's more, it can be made arbitrarily large, showing a strong separation between classical and quantum strategies.

II. METHODS

Let us now give an overview of our proof - full details can be found in the appendix. First, let us specify the parameters of the Witsenhausen instance. We will use a channel $\mathcal{N}_t = \mathcal{N} \circ \epsilon_t$, where $\mathcal{N} : [q] \times [d] \to ([q] \times [d])^{\times 2}$ with the notation $[q] = \{0, \ldots, q - 1\}$. Such a channel for e.g. $(q, d) = (6, 4)$ was studied in [11], where the task was to transmit messages $m \in [q]$ over $\mathcal{N}$ with *zero* error using an encoder $\mathcal{E}$ and decoder $\mathcal{D}$. Here, we will make use of two properties of $\mathcal{N}$ established in [11]:

1. For any classical encoder and decoder, the maximal number of symbols that can be transmitted over $\mathcal{N}$ with *zero* error is $q - 1$.

2. For a quantum encoder and decoder, there exists a strategy to send q symbols with zero error. More precisely, there exists a quantum state $|\Psi\rangle$ shared between the encoder and decoder, as well as measurements by the encoder depending on its input $m \in [q]$ giving outcome j, and a measurement by the decoder depending on its input $s = \mathcal{N}((m, j))$ such that the decoder recovers (m, j) with probability 1 for all $m \in [q]$.

The purpose of the channel ϵ_t is to map integers $y = mt + c_1(x)$ to inputs (m, j) accepted by the channel $\mathcal{N}$. This is accomplished by letting $\epsilon_t : \mathbb{Z} \to [q] \times [d]$ for $t \geq d$:

- If $x = at + b$ for some $a \in [q]$ and $b \in [d]$ then ϵ_t outputs (a, b).

- Otherwise ϵ_t outputs (a, b) chosen uniformly at random from $[q] \times [d]$.

It is not difficult to see that the maximum number of symbols that can be transmitted over $\mathcal{N}_t$ with *zero* error using a purely classical encoder and decoder remains $q - 1$. The parameter $k > 0$ can be chosen arbitrarily, independent of t and $\mathcal{N}$.

The input distribution we will use in our Witsenhausen instance is $P_X^t(x) = 1/q$ if $x = mt$ and $P_X^t(x) = 0$ otherwise. Intuitively, one can thus think of the input as being m, amplified by t. Our argument is identical for any other distribution on m.

For concreteness the reader may assume $(q, d) = (6, 4)$ and $k = 1$.

1. Quantum cost function

Let us first show that there exists a quantum strategy for which the cost function is independent of t. This is an easy consequence of the quantum strategy for zero error coding given in [11] adapted to our setting.

Let controllers one and two share the state $\rho = |\Psi\rangle\langle\Psi|$ from the zero error coding scheme above. Upon receiving $x = mt$, controller c_1 first applies $\epsilon_t(x) = (m, 0)$ himself to find m. He then uses the measurement for m performed by the encoder $\mathcal{E}$ in [11] on his part of $|\Psi\rangle$ to obtain j. He outputs $c_1^t(x) = j$ to be added to the signal which is then $y = x + j = mt + j$.

Note that controller c_2 receives output $s = \{(m, j), (m', j')\}$ from the channel $\mathcal{N}_t$. He performs the measurement by the decoder of [11] to obtain (m, j). He then outputs $c_2^t(s) = -mt - j$ to be added to the signal, which is gives $z = y + c_2^t(s) = 0$.

Since the final signal on the line is $z = mt + j - mt - j = 0$, the cost for this strategy is determined only by the term $k(c_1^t(x))^2$. Using the fact that $c_1^t(x) \le d - 1$ we can show that for our strategy

$$C^{\mathcal{N}_t, P_X^t, k}(c_1^t, c_2^t, |\Psi\rangle\langle\Psi|) = k \sum_{m \in [q]} P_X(mt)(c_1^t(mt))^2 \tag{4}$$

$$\le kd^2 . \tag{5}$$

2. Classical cost function

The challenging part is to show that the classical cost function can be made arbitrarily large with t. We accomplish this in a series of lemmas, which together yield the theorem below. The main intuition behind our argument is the observation that when t grows very large, but the cost function is still bounded, then controller two must have obtained a reasonable estimate

for any possible $m \in [q]$. We then argue by contradiction and show that if the cost function is bounded, then controller one and controller two can transform their control strategy into a strategy for sending more then $q - 1$ symbols over the channel $\mathcal{N}$ with *zero* error, which is a contradiction.

Theorem 1. *Let* $(\mathcal{N}_t, P_X^t, k)$ *be a Witsenhausen instance as outlined above. Then for any* $M \in \mathbb{R}$ *there exists a* $t_0 \ge d$ *such that for all* $t \ge t_0$

$$C_{SR}^{\mathcal{N}_t, P_X^t, k} > M. \tag{6}$$

Proof. We first show that we can restrict ourselves to deterministic strategies (Lemma 1). Hence it suffices to show that

$$C_D^{\mathcal{N}_t, P_X^t, k} > M \tag{7}$$

for large enough t. To this end, let $(c_1^t, c_2^t, 0)$ be a family of deterministic strategies which achieve a cost of $C_D^{\mathcal{N}_t, P_X^t, k}$. By assuming that

$$C_D^{\mathcal{N}_t, P_X^t, k} \le M \tag{8}$$

for all t, we will derive a contradiction. Using the definition of the cost function, this means that

$$C_D^{\mathcal{N}_t, P_X^t, k}(c_1^t, c_2^t, 0) = \mathop{\mathbb{E}}_{P_X, \mathcal{N}_t} \left[kc_1^t(x) + z^2 \right] \le M , \tag{9}$$

where $z = x + c_1^t(x) + c_2^t(s)$.

Since the expectation is taken over a finite set of events of non-zero probability we can show that $|c_1^t(x)|$ and $|z|$ are uniformly bounded for all possible inputs x (see Lemma 2). That is,

$$|c_1^t(x)| \le M_X , \tag{10}$$

$$|z| \le M_Z , \tag{11}$$

for some M_X and M_Z which are independent of t since they can be defined in terms of the parameters M, k, $P_X^{t\,min}$ and $P_Z^{t\,min}$ which are all independent of t. (The distributions P_X^t and P_Z^t depend on t, but the *minimal* probability in these distributions does not.)

Then, for large enough t, $|c_1^t(x)|$ is small compared with the inputs which are of the form $x = mt$. Dividing by t to make this more apparent, the signal $y = x + c_1^t(x)$ after the first controller satisfies

$$\frac{y}{t} = m + \frac{c_1^t(mt)}{t} \tag{12}$$

$$\le m + \frac{M_X}{t} \tag{13}$$

$$\approx m \tag{14}$$

for large t since $M_X/t \to 0$ as $t \to \infty$. Now the final output satisfies

$$\frac{z}{t} = m + \frac{c_1^t(mt)}{t} + \frac{c_2^t(s)}{t} \tag{15}$$

$$\approx m + \frac{c_2^t(s)}{t} \tag{16}$$

$$\approx 0 \tag{17}$$

for large t. The second line follows from the fact that $c_1^t(x)$ is uniformly bounded for all possible x (Lemma 2), and the final line follows from the fact that also $|z|$ is uniformly bounded. Hence, we have that $z/t \to 0$ as $t \to \infty$. Lemma 3 makes this argument formal. In particular, when $t_0 = 2(M_X + M_Y) + 1$, for any $t \geq t_0$ we have

$$\left| \frac{x}{t} + \frac{c_2^t(s)}{t} \right| < \frac{1}{2} \tag{18}$$

for all possible (x, s) such that $P_{X,S}(x, s) > 0$ (in particular, for all $x = mt$ with $m \in [q]$).

In our concrete example with $(q, d) = (6, 4)$ and $k = 1$ we can bound M_X and M_Z from above by $\sqrt{6M}$ and $\sqrt{54M}$, respectively. In this case, t_0 is bounded above by $20\sqrt{M} + 1$.

From now on let $t \geq t_0$. Note that (18) means that $c_2^t(s)/t$ forms a good estimate for $m = x/t$. This allows us to construct a zero error encoding scheme for the channel $\mathcal{N}_t$. We use message set $\mathcal{M} = [q]$ and let

$$\eta = \frac{-c_2^t(s)}{t} \ . \tag{19}$$

We then define the encoding scheme $\mathcal{E}$ and decoding scheme $\mathcal{D}$ as

- $\mathcal{E}(m) = mt + c_1^t(mt)$.

- $\mathcal{D}(s)$ is given by rounding off η to the nearest integer.

Since $|m - \eta| < 1/2$ by (18), the nearest integer to η is always m and $\mathcal{D}$ always decodes correctly. Here the fact that we have used a particular distribution for the inputs does not matter. Since we achieve zero probability of error the encoding strategy must work with probability 1 for every input with positive probability. Put differently, our encoding strategy works with zero error for an alphabet of q symbols. This contradicts the fact that only $q - 1$ symbols can be sent over $\mathcal{N}_t$ with zero error. □

III. DISCUSSION

We demonstrated that entanglement improves classical control by studying a very simple circuit. It is an interesting open question to decide for which other control circuits entanglement is useful, and whether one could maybe even characterize the set of circuits for which entanglement helps. We note that a number of existing tasks in quantum information could be converted into a problem of classical control. First, Bell inequalities [12] expressed in the language of non-local games (see e.g. [13] for a survey) can be turned into a control problem in which two non-communicating controllers, Alice and Bob, receive inputs and have to produce outputs trying to maximize their winning probability which in this context can be seen as the negative of the cost function. The circuit obtained this way is somewhat simpler than Figure 1 because there is no feedforward, but nevertheless allows the construction of a primitive classical circuit in which known results on Bell inequalities tell us that entanglement does help. Indeed, one may see such non-local games as very restricted cases of the discrete team decision problem studied in the context of Witsenhausen's counterexample [10, 14].

Another task that allows us to come up with a classical circuit in which entanglement can help is the one of sending information over a noisy-channel, where the encoder and the decoder can share entanglement. In this context, it is known that entanglement helps to increase the zero-error capacity of the channel [11, 15] and one can turn this problem into a less trivial control circuit by defining a cost function that introduces a penalty for incorrect decoding. Looking at Figure 1 it becomes clear that this circuit is already somewhat related to Witsenhausen's counterexample, because transmitting information correctly from one controller to the other helps them to minimize the cost function, underscoring the information-theoretic flavor of this circuit [16, 17] Indeed, Witsenshausen's counterexample can be cast in the language of information transmission [18] where minimizing one part of the cost function can be seen as minimizing the mean-square error of correctly transmitting information from one controller to the other. As such, the techniques of [11, 15] are useful for us to establish one half of our proof, namely that the strategy using shared entanglement leads to a cost function that is independent of the parameter t. However, Witsenhausen's counterexample has the additional twist that the use of controller 1 introduces an additional penalty term into the cost function that is normally not considered in quantum information theory.

Taking an extremely broad perspective, one could even take computational problems and define a cost function based on whether a particular circuit correctly computes the solution of some function f on a given input x. Here, measurement based computing [19] could

be seen as a possible control strategy to minimize the cost, using controllers sharing a quantum state.

We emphasize, however, that our aim is not to take the many tasks of quantum information and turn them into specific instances of classical control circuits. Rather, our goal is to raise the general question whether entanglement can improve classical control circuits, and we have shown that this is indeed possible by introducing entanglement into a well-studied and established circuit from the control literature.

Witsenhausen's counterexample was originally conceived to demonstrate that linear control laws are not always optimal [6]. It has since been studied extensively to determine the computational difficulty of computing optimal classical control strategies [18, 20, 21], and various heuristic algorithms are known [22–24]. Intriguingly, the optimal strategy for Witsenhausen's original circuit remains unknown (see e.g. [18]), and discrete versions [9] have even been proven to be NP-complete [10]. As such, it would be very interesting to determine the complexity of determining the optimal control strategy in the presence of entanglement. Note that this can be seen as a relaxation of the original question because strategies involving entanglement include all classical strategies as a special case, and it is

hence conceivable that solving the quantum question could be easier than determining the optimal classical strategy. This would give a method for obtaining bounds for classical strategies. In the setting of Bell inequalities discussed above, it is known that for some non-local games, i.e. some specific control problems, known as XOR games, the optimal quantum strategy is easy to compute [25, 26] whereas finding the optimal classical strategy is hard. Yet, for other games the question is known to be hard even in the presence of entanglement [27, 28]. The question of how difficult it is to compute the optimal *quantum* strategy for Witsenhausen's circuit is thus wide open.

ACKNOWLEDGMENTS

[1] L. Viola and S. Lloyd, Phys. Rev. A , 2733 (1998).

[2] C. Ahn, A. C. Doherty, and A. J. Landahl, Phys. Rev. A **65**, 042301 (2002).

[3] D. Dong and I. R. Petersen, IET Control Theory and Applications **4**, 2651 (2010).

[4] R. Ionicioiu and D. R. Terno, Phys. Rev. Lett. **107**, 230406 (2011).

[5] Throughout, the term "classical" refers to non-quantum and not a particular form of control as in [17].

[6] H. S. Witsenhausen, Siam Journal on Control **6** (1968), 10.1137/0306011.

[7] Y. Ho, in *Proceedings of the 47th IEEE Conference on decision and control (CDC)* (2008) pp. 1611–1613.

[8] A. Sahai, IEEE Control Systems Magazine , 20 (2010).

[9] Y.-C. Ho and T. Chang, IEEE Transactions on Automat. Contr. **25**, 537 (1980).

[10] C. H. Papadimitriou and J. Tsitsiklis, SIAM J. Control Optim. **24**, 639 (1986).

[11] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, Physical Review Letters **104**, 230503 (2010).

[12] J. S. Bell, **1**, 195 (1965).

[13] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," (2013), arXiv:1303.2849.

[14] R. Rander, Annals of Mathematical Statistics (1962).

[15] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, IEEE Trans. Inf. Theor. **57**, 5509 (2011).

[16] S. Mitter and A. Sahai, *Information and control: Witsenhausen revisited*, Lecture notes in control and information sciences (Springer, 1999) pp. 281–293.

[17] H. S. Witsenhausen, in *Proceedings of the IEEE*, Vol. 59 (1971) pp. 1557–1566.

[18] P. Grover, A. Sahai, and S. Y. Park, in *Proceedings of the 7th international conference on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks* (IEEE, 2009) pp. 604–613.

[19] R. Raussendorf, D. E. Browne, and H. J. Briegel, Phys. Rev. A **68**, 022312 (2003).

[20] Y. Wu and S. Verdú, in *Proceedings of the 50th IEEE Conference on Decision and Control (CDC)* (2011).

[21] P. Gover and A. Sahai, Int. J. Syst. Control Commun. (Special Issue on Information Processing and decision making in distributed control systems) **2**, 197 (2010).

[22] M. Baglietto, T. Parisini, and R. Zoppoli, IEEE Transactions on Automat. Contr. **46**, 1471 (2001).

[23] J. T. Lee, E. Lau, and Y. Ho, IEEE Transactions on Automat. Contr. **46**, 382 (2001).

[24] N. Li, J. Marden, and J. S. Shamma, in *Proceedings of the 48th IEEE Conference on Decision and Control (CDC)* (2009) pp. 157–162.

[25] R. Cleve, P. Høyer, B. Toner, and J. Watrous, in *Proceedings of 19th IEEE Conference on Computational Complexity* (2004) pp. 236–249.

[26] S. Wehner, Physical Review A **73**, 022110 (2006).

[27] J. Kempe, H. Kobayashi, K. Matsumoto, B. Toner, and T. Vidick, "Entangled games are hard to approximate," (2007), arXiv:0704.2903.

[28] T. Vidick, "Three-player entangled xor games are np-hard to approximate," (2013), arXiv:1302.1242.

[29] This channel is different from one used by Cubbit et al. [11] but it has the same confusability graph and deterministic capacity. As well, the same quantum encoding/decoding strategy also works.

Appendix A: Overview

In this appendix we provide the technical details of our claims. Our goal is thereby to produce a family of Witsenhausen instances, $(\mathcal{N}_t, P_X^t, k)$ parameterized by $t > d$ for some constant d such that

$$C_{SE}^{\mathcal{N}_t, P_X^t, k} \leq C , \tag{A1}$$

where $C = kd^2$ is independent of t, but

$$\lim_{t \to \infty} C_{SR}^{\mathcal{N}_t, P_X^t, k} = \infty \tag{A2}$$

That is to say, for large enough t the quantum strategy beats the classical bound, and that the difference can be made arbitrarily large as t increases.

The construction of our family of instances of the Witsenhausen problem is based on a classical channel $\mathcal{N}$ used by Cubitt et al. [11] and [15]. In Section B we recall their construction for completeness. In particular [11, 15] showed that the zero-error capacity of $\mathcal{N}$ without entanglement is strictly smaller than the entanglement-assisted capacity, and we will employ some of their techniques here to prove the quantum part of our task.

In Section C, we then give the construction for our family of instances of the Witsenhausen problem based on $\mathcal{N}$. This is done by choosing a suitable encoding from the inputs in our circuit to inputs acceptable by $\mathcal{N}$. This encoding depends on the parameter t. The combined encoding composed with $\mathcal{N}$ gives us our noisy channels $\mathcal{N}_t$ to be used in the Witsenhausen problem. We also define a probability distribution P_X^t, depending on t where the largest value with non-zero probability increases with t.

To establish our goal we have to prove two things: First, we have to show that using entanglement the cost function is bounded by some constant (see (A1)). For the quantum control strategy for the instance $(\mathcal{N}_t, P_X^t, k)$ we make some small modifications to the encoding and decoding strategies used by Cubitt et al. [11, 15] for $\mathcal{N}$. Roughly, the first controller will use the encoding strategy to modulate the signal on the line so as to encode the input x for transmission via $\mathcal{N}_t$. The second controller then uses the decoding strategy to determine x, which allows it to damp the signal down to 0.

Second, we have to show that by choosing t large enough we can make the classical cost function exceed the quantum one (see (A2)). We will establish this result by contradiction. In particular, if, instead of diverging to infinity, the classical cost was uniformly bounded for all t, we show that the first controller's signal $c_1(x)$ is also uniformly bounded. Based on this,

since the input values x grow without limit as t increases, the second controller c_2 must do the bulk of the damping for large enough t. In order to achieve this damping c_2 must have very good information about x, and for large enough t we find that c_2 must have perfect information on x. This implies that c_2 can perfectly decode the outputs of $\mathcal{N}$ for all inputs x just as the quantum controllers do. In turn, this allows us to construct a zero error coding scheme for $\mathcal{N}$ contradicting the classical zero-error capacity of $\mathcal{N}$.

Appendix B: Preliminaries

1. Zero-error classical communication

The scenario considered by Cubitt et al. involves a noisy classical channel $\mathcal{N}$ where the goal is to find an encoding scheme which allows the transmission of classical information over the noisy channel with *zero error*. $\mathcal{N}$ has input set $\mathcal{I}$ and output set $\mathcal{O}$ and the channel is given as $\mathcal{N}(o\,|\,i)$, the probability of obtaining output o when the input is i. In other words, $\mathcal{N}$ is a conditional probability distribution. We wish to use $\mathcal{N}$ to send messages using some encoding scheme $\mathcal{E}$ and decoding scheme $\mathcal{D}$, as in Figure 2. The encoding and decoding scheme may make use of some shared resource ρ.

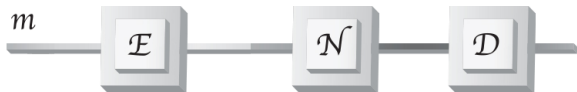$$m \quad \boxed{\mathcal{E}} \quad \boxed{\mathcal{N}} \quad \boxed{\mathcal{D}}$$

FIG. 2. Channel Communication with Correlation Assisted

When $\mathcal{M}$ is the input alphabet of $\mathcal{E}$ and $\mathcal{I}_\mathcal{M}$ is the identity channel on $\mathcal{M}$, then the *zero-error capacity* of $\mathcal{N}$ when using the encoding/decoding strategy $(\mathcal{E}, \mathcal{D}, \rho)$ is

$$Z^\mathcal{N}(\mathcal{E}, \mathcal{D}, \rho) = \begin{cases} |\mathcal{M}| & \text{if } \mathcal{D}(\mathcal{N}(\mathcal{E})) = \mathcal{I}_\mathcal{M} \\ 0 & \text{otherwise} \end{cases} \quad \text{(B1)}$$

Then the zero-error capacity of $\mathcal{N}$ using strategies from the set Ω is

$$Z^\mathcal{N}_\Omega := \max_{(\mathcal{E}, \mathcal{D}, \rho) \in \Omega} Z^\mathcal{N}(\mathcal{E}, \mathcal{D}, \rho) \ . \quad \text{(B2)}$$

As with Witsenhausen's problem, one can consider three classes of strategies: the deterministic strategies D, strategies using shared randomness SR, and strategies using shared entanglement SE. Strategies in D are $(\mathcal{E}, \mathcal{D}, 0)$, where $\mathcal{E} : \mathcal{M} \to \mathcal{I}$, $\mathcal{D} : \mathcal{O} \to \mathcal{M}$ and ρ can be taken to be 0. Strategies in SR are triples $(\mathcal{E}, \mathcal{D}, \rho)$

where ρ is a probability distribution on some set R, $\mathcal{E} : \mathcal{M} \times R \to \mathcal{I}$ and $\mathcal{D} : \mathcal{O} \times R \to \mathcal{M}$.

For a strategy $(\mathcal{E}, \mathcal{D}, \rho) \in SE$, $\rho \in \mathcal{B}(\mathcal{H}_\mathcal{E} \otimes \mathcal{H}_\mathcal{D})$ is a quantum state shared between the encoder holding system $\mathcal{H}_\mathcal{E}$ and the decoder holding system $\mathcal{H}_\mathcal{D}$. Without loss of generality, $\mathcal{E}$ performs a measurement on $\mathcal{H}_\mathcal{E}$ indexed by its input m, and outputs the measurement outcome $i = \mathcal{E}(m)$ on the wire. Formally, this means that the encoding strategy $\mathcal{E}$ is defined by measurement operators $\{\{\Pi^1_{m,i} \geq 0\}_i\}_m$ where for each m we have $\sum_i \Pi^1_{m,i} = I_{\mathcal{H}_\mathcal{E}}$. Similarly, the decoder $\mathcal{D}$ performs measurements $\{\{\Pi^2_{o,m} \geq 0\}_m\}_o$ on $\mathcal{H}_\mathcal{D}$ indexed by its input o and outputs the outcome $m = \mathcal{D}(o)$.

2. Optimal deterministic encodings

For deterministic encodings the input and output encodings are simply relabellings of some subsets of $\mathcal{I}$ and $\mathcal{O}$. The critical consideration, then, is which subsets to use. To this end we consider the notion of *confusability*. For element $i \in \mathcal{I}$ let $T_i = \{o \in \mathcal{O}\,|\,\mathcal{N}(o|i) > 0\}$ be the set of outputs to which $\mathcal{N}$ maps i. Two symbols $i, j \in \mathcal{I}$ are said to be *confusable* if their output distributions have at least one common element, i.e. if $T_i \cap T_j \neq \emptyset$.

The *confusability graph* $G(\mathcal{N})$ of the classical channel, $\mathcal{N}$, is a graph whose vertex set is $\mathcal{I}$ where two vertices are adjacent whenever they are confusable.

$$V(G) = \mathcal{I} \quad \text{(B3)}$$
$$E(G) = \{(i,j)\,|\,T_i \cap T_j \neq \emptyset\} \quad \text{(B4)}$$

Note that, for any zero-error deterministic encoding scheme, two confusable elements of $\mathcal{I}$ cannot be used to encode two different messages since it is impossible to perfectly distinguish the two elements based on the output of $\mathcal{N}$. Hence we can put an upper bound on $Z^\mathcal{N}_D$ by determining the maximum size of a set of vertices in which no two are adjacent, i.e. the size of the largest independent set in $G(\mathcal{N})$, called the *independence number* of $G(\mathcal{N})$.

In fact, every independent set $S \in \mathcal{I}$ also gives a zero-error encoding strategy. By definition the T_i's are pairwise disjoint for $i, j \in S$ since all elements of S are pairwise non-confusable. We can then encode a message m by mapping it to an element $i \in S$, and decode by mapping all elements of T_i back to m. So $Z^\mathcal{N}_D$ equals the independence number of $G(\mathcal{N})$ [11].

3.   Channel Description and Classical Bound

In this subsection, we describe a channel $\mathcal{N}$ such that $Z_{SE}^{\mathcal{N}} > Z_D^{\mathcal{N}}$, due to Cubitt et al. [11], [15]). We will use $\mathcal{N}$ in the construction of our Witsenhausen problem.

We will specify the channel by first specifying the confusability graph for our desired channel $\mathcal{N}$, which is defined in terms of a Kochen - Specker basis set. We then construct the channel from this graph.

Definition 1. *Let $U = \{B_m | m \in [q]\}$ be a set of q orthonormal bases $B_m = \{|b_{mj}\rangle \, | \, j \in [d]\}$ for $\mathbb{C}^d$. If every set of vectors containing at least one vector from each B_m contains at least one pair of orthogonal vectors then we call U a* Kochen - Specker basis set *(or a KS basis set).*

Cubbit et al. use such a set of bases U with $(q, d) = (6, 4)$ in the construction of $\mathcal{N}$ [11].

For our channel[29] , set $\mathcal{I} = [q] \times [d]$ and $\mathcal{O}$ to the set of unordered pairs of $[q] \times [d]$, i.e., an element of $\mathcal{O}$ looks like $\{(m, j), (m', j')\}$ with $(m', j') \neq (m, j)$, corresponding to an edge. Set $\hat{T}_{(m,j)}$ to be the set of pairs (m', j') such that $\langle b_{mj} | b_{m'j'} \rangle = 0$. The channel is defined by

$$\mathcal{N}(\{(m, j), (m', j')\} | (m, j)) = \begin{cases} \frac{1}{|\hat{T}_{(m,j)}|} & o \in \hat{T}_{(m,j)} \\ 0 & o \notin \hat{T}_{(m,j)}. \end{cases}$$

$$(B5)$$

Note that if $i \notin o$, then $\mathcal{N}(o|i) = 0$.

For the bases U with $(q, d) = (6, 4)$ used by Cubbit et al. the degree of each vertex (m, j) is $9 = |\hat{T}_{(m,j)}|$. Hence $\mathcal{N}(o|i)$ is either $1/9$ or 0.

Consider the confusability graph of $\mathcal{N}$ which has vertices $\mathcal{I}$ and edges which are unordered pairs of vertices, i.e. elements of $\mathcal{O}$. Two elements (m, j), and (m', j') are confusabale whenever they can both be mapped to a common output. The only possible common output is $\{(m, j), (m', j')\}$. Hence the edge set of $G(\mathcal{N})$ is just the elements of $\mathcal{O}$ which occur with non-zero probability. These are the $\{(m, j), (m', j')\}$ for which $\langle b_{mj} | b_{m'j'} \rangle = 0$.

We now bound the classical zero-error capacity of $\mathcal{N}$ [11, 15]. The set of vertices in $G(\mathcal{N})$ can be partitioned into q cliques corresponding to the q bases in U. Any set of independent vertices can contain at most one vertex in each clique. Further, since U is a KS basis set, any set which contains a vertex in each of the q cliques contains two vertices which are adjacent, and hence the set is not independent. Thus the independence number of $G(\mathcal{N})$ is less than q and $Z_D^{\mathcal{N}} = Z_{SR}^{\mathcal{N}} < q$.

4.   Quantum zero-error encoding for $\mathcal{N}$

With an upper bound on the classical zero-error capacity of $\mathcal{N}$ in place, we recall the quantum strategy in SE that beats the classical bound for zero error coding [11].

To begin, we first specify the entangled state that the encoder and decoder will share, which is the maximally entangled state $|\psi\rangle \in \mathcal{H}_{\mathcal{E}} \otimes \mathcal{H}_{\mathcal{D}}$, where $\mathcal{H}_{\mathcal{E}}$ and $\mathcal{H}_{\mathcal{D}}$ have dimension d.

$$|\psi\rangle = \frac{1}{\sqrt{d}} \sum_{j \in [d]} |j\rangle \otimes |j\rangle \qquad (B6)$$

When the encoder $\mathcal{E}$ receives a message $m \in [q]$ it measures its half of the shared state $|\psi\rangle$ along the basis B_m^* (obtained by conjugating each state in B_m) to obtain outcome $j \in [d]$. The residual state on the decoder's subsystem is then $|b_{mj}\rangle$. The encoder then sends (m, j) through the channel $\mathcal{N}$.

The decoder $\mathcal{D}$ receives from $\mathcal{N}$ some s as its input, which is an edge in G incident with (m, j) and some other vertex (m', j'). The decoder thus cannot tell yet, whether the messages was m or m'. Since these vertices are adjacent, they correspond to orthogonal vectors $|b_{mj}\rangle$ (the residual state on $\mathcal{D}$'s subsystem) and $|b_{m'j'}\rangle$. Hence upon receiving s, $\mathcal{D}$ measures its subsystem in some basis which includes $|b_{mj}\rangle$ and $|b_{m'j'}\rangle$. The outcome is (m, j) with probability 1. $\mathcal{D}$ then outputs m. (Note that $\mathcal{D}$ also determines j, which will be important later in our Witsenhausen construction.)

With $\mathcal{E}$ and $\mathcal{D}$ as above, and $\rho = |\psi\rangle\langle\psi|$, we find that

$$Z_{SE}^{\mathcal{N}} \geq Z^{\mathcal{N}}(\mathcal{E}, \mathcal{D}, \rho) = q \qquad (B7)$$

which beats the classical bound of $q - 1$.

Appendix C: Construction of the Witsenhausen problem

For an instance of the Witsenhausen problem we need to give a channel, a probability distribution on $\mathbb{Z}$ for the inputs, and a constant k. We now construct a family of channels $\mathcal{N}_t$ which we will use for our instances of the Witsenhausen problem. We wish to use $\mathcal{N} : [q] \times [d] \to ([q] \times [d])^{\times 2}$ constructed in the previous section. It is clear that we cannot immediately use this channel since our wire yields inputs in $\mathbb{Z}$. We hence define a family of suitable encoding channels $\epsilon_t : \mathbb{Z} \to [q] \times [d]$ for $t \geq d$:

- If $x = at + b$ for some $a \in [q]$ and $b \in [d]$ then ϵ_t outputs (a, b).

- Otherwise ϵ_t outputs (a, b) chosen uniformly at random from $[q] \times [d]$.

The channels we use in constructing our Witsenhausen problems are

$$\mathcal{N}_t := \mathcal{N} \circ \epsilon_t \tag{C1}$$

We will also need to know the deterministic zero-error capacity of $\mathcal{N}_t$, given by

$$Z_D^{\mathcal{N}_t} = q - 1. \tag{C2}$$

To see this, note that the confusability graph G' of $\mathcal{N}_t$ has an induced subgraph isomorphic to G (the confusability graph of $\mathcal{N}$) and all other vertices are all adjacent to every other vertex. The latter vertices cannot be included in any independent set of size greater than 1, so the independence number cannot increase beyond that of G. On the other hand, an independent set in G maps to an independent set in G', so the independence number of G' is the same as that of G.

For the input probability distributions for our Witsenhausen problems, we define a distribution P_X^t of the input x by

$$P_X^t(x) := \begin{cases} \frac{1}{q} & \text{if } x = mt \text{ where } m \in [q] \\ 0 & \text{otherwise} \end{cases} \tag{C3}$$

Note that the value of q goes hand in hand with the channel $\mathcal{N}$. We have chosen this distribution since it ensures that no x occurs for which ϵ_t generates a random output. In other words, with probability 1 the input to the Witsenhausen problem is mapped directly to a particular input to $\mathcal{N}$. Note that for a fixed t, the distribution (C3) actually only depends on m, and hence we identify

$$P_M^t(m) := P_X^t(mt) . \tag{C4}$$

Finally, we need to specify k. Our results will work with any strictly positive k, but for concreteness one may take

$$k = 1. \tag{C5}$$

Appendix D: Quantum strategy for the Witsenhausen problem

Let us now show that there exists a quantum strategy using shared entanglement such that the cost function is a constant. We will make use of the quantum encoding strategy for $\mathcal{N}$ proposed in [11, 15], which achieves a zero-error capacity of q. The intuition is that the first controller, c_1^t, will perform the quantum

encoding $\mathcal{E}$ for $\mathcal{N}_t$, suitably modified to modulate the signal so as to encode the message $m = x/t$ into $\mathcal{N}_t$. Then c_2^t will later decode according to the quantum decoding strategy $\mathcal{D}$ for $\mathcal{N}_t$ in order to determine (m, j) and damp the signal on the line down to 0.

- Controller c_1^t does the following:
  - Receive $x = mt$ and then apply $\epsilon_t(x) = (m, 0)$ to find m.
  - Measure $|\psi\rangle$ in basis B_m^* to obtain (m, j).
  - Output $c_1^t(x) = j$ to be added to the signal which is then $y = x + j = mt + j$

- Controller c_2^t does the following:
  - Receive output $s = \{(m, j), (m', j')$ from the channel $\mathcal{N}_t$.
  - Perform projective measurement on $|\psi\rangle$ on a basis containing $|b_{m,j}\rangle$ and $|b_{m',j'}\rangle$ to obtain (m, j).
  - Output $c_2^t(s) = -mt - j$ to be added to the signal, which is then $z = y + c_2^t(s) = 0$.

Since final signal on the line is $z = mt + j - mt - j = 0$, the cost for this strategy is determined by the $k(c_1^t(x))^2$ term only:

$$C^{\mathcal{N}_t, P_X^t, k}(c_1^t, c_2^t, \rho) = k \mathop{\mathbb{E}}_{P_X^t, \mathcal{N}_t} \left[(c_1^t(mt))^2 \right] \tag{D1}$$

$$= k \sum_x P_X^t(x)(c_1^t(x))^2 . \tag{D2}$$

Recall that $P_X^t(x) = P_M(m)$ for $x = mt$ with $m \in [q]$ and 0 otherwise. Hence

$$C^{\mathcal{N}_t, P_X^t, k}(c_1^t, c_2^t, \rho) = k \sum_{m \in [q]} P_M(m)(c_1^t(mt))^2 \tag{D3}$$

Now since $0 \le c_1^t(mt) \le d - 1$

$$C^{\mathcal{N}_t, P_X^t, k}(c_1^t, c_2^t, \rho) < k \sum_{m \in [q]} P_M(m)d^2 \tag{D4}$$

$$= kd^2 . \tag{D5}$$

We thus have that $C_{SE}^{\mathcal{N}_t, P_X^t, k} < kd^2$ for all $t \ge d$ as promised in (A1).

Appendix E: Classical bound for the Witsenhausen problem

The challenging part of our result is to show that the classical cost function can be made to grow arbitrarily large by increasing t. Let us first give an outline of our

9

proof, which can be divided into several natural steps. First we reduce to the case of deterministic strategies. Next we assume for contradiction that there is some uniform upper bound M on the minimal deterministic cost for all t. Taking some family of strategies $(c_1^t, c_2^t, 0)$ which achieve the minimal deterministic cost, we first show that $|c_1^t|$ and $|z|$ are also uniformly bounded. Because of this, the signal after c_1^t will be dominated by the mt term for large t, i.e. $y \approx mt$. Hence c_2^t must have a very good estimate of m in order to achieve a bounded final signal. Put differently, for large enough t, $z \approx mt + c_2^t \approx 0$ so $-c_2^t/t \approx m$. We use this fact to construct an encoding scheme for $\mathcal{N}_t$. Roughly, the encoding scheme simulates the Witsenhausen problem to create y which is sent through the channel. The decoding scheme is to calculate $-c_2^t/t$ which will be equal to the input m. For large enough t this scheme has no error and beats the deterministic zero-error bound $Z_D^{\mathcal{N}_t}$, achieving our contradiction.

In the following, we will need to talk about distributions of z and s. From their definitions, we see that these distributions do of course depend on the strategies of the controllers along with the distribution on x and the channel $\mathcal{N}$. Note, however, that c_1^t and c_2^t are deterministic functions of x and s and hence the probabilities of z and s are completely determined by P_X^t and $\mathcal{N}_t$ once we fix c_1^t and c_2^t. More precisely, we have

$$P_{X,S}^t(x,s) = P_X^t(x)\mathcal{N}_t(s|x + c_1^t(x)) \qquad \text{(E1)}$$

$$P_Z^t(z) = \sum_{\substack{x \\ z=x+c_1^t(x)+c_2^t(s)}} P_{X,S}^t(x,s) \ . \qquad \text{(E2)}$$

Furthermore, since all of our probability distributions have finite support (i.e. there is a finite output set for $\mathcal{N}$ and a finite set of possible inputs x) we can define a minimum probability:

$$P_X^{t\,min} = \min_{P_X^t(x)>0} P_X^t(x) \qquad \text{(E3)}$$

$$P_Z^{t\,min} = \min_{P_Z^t(z)>0} P_Z^t(z) \qquad \text{(E4)}$$

For a given (q,d) we may calculate explicit values. In particular, for $(q,d) = (6,4)$ and $\mathcal{N}^t$ and P_X^t as given above, we find

$$P_X^{t\,min} = \frac{1}{6}. \qquad \text{(E5)}$$

Although $P_Z^{t\,min}$ depends explicitly on c_1^t and c_2^t we can still find lower bounds on the minimal probability since these are deterministic functions. Supposing that each (x,s) gives a different value of z so that there is no summing of probabilities, the minimal non-zero probability for a given z is the same as the minimal

probability for a given (x,s). In other cases the probability for particular values of z can only go up. Now, s can depend implicitly on x and c_1^t via the input to $\mathcal{N}^t$, but for any input there 9 possible outputs from $\mathcal{N}^t$, each occurring with equal probability. Thus we can lower bound the probability of any (x,s) by $1/6 \cdot 1/9$. So

$$P_Z^{t\,min} \geq \frac{1}{6} \cdot \frac{1}{9} = \frac{1}{54}. \qquad \text{(E6)}$$

1. Main theorem

Here we state and prove the main theorem for the bound on $C_{SR}^{\mathcal{N}_t, P_X^t, k}$. For convenience sake, we first re-state the theorem and provide some more details in the proof outline. We will use various lemmas in the proof, which appear with proofs in later sections.

Theorem 2. *Let $(\mathcal{N}_t, P_X^t, k)$ be as given in Appendix C. Then for any $M \in \mathbb{R}$ there exists a $t_0 \geq d$ such that for all $t \geq t_0$*

$$C_{SR}^{\mathcal{N}_t, P_X^t, k} > M. \qquad \text{(E7)}$$

Proof. We first show that we can restrict ourselves to deterministic strategies (Lemma 1). Hence it suffices to show that

$$C_D^{\mathcal{N}_t, P_X^t, k} > M \qquad \text{(E8)}$$

for large enough t. To this end, let $(c_1^t, c_2^t, 0)$ be a family of deterministic strategies which achieve a cost of $C_D^{\mathcal{N}_t, P_X^t, k}$. By assuming that

$$C_D^{\mathcal{N}_t, P_X^t, k} \leq M \qquad \text{(E9)}$$

for all t, we will derive a contradiction. Using the definition of the cost function, this means that

$$C_D^{\mathcal{N}_t, P_X^t, k}(c_1^t, c_2^t, 0) = \mathbb{E}_{P_X, \mathcal{N}_t}\left[kc_1^t(x) + z^2\right] \leq M \ , \qquad \text{(E10)}$$

where $z = x + c_1^t(x) + c_2^t(s)$.

Since the expectation is taken over a finite set of events of non-zero probability, we can show that $|c_1^t(x)|$ and $|z|$ are uniformly bounded for all possible inputs x (see Lemma 2). That is,

$$|c_1^t(x)| \leq M_X \ , \qquad \text{(E11)}$$

$$|z| \leq M_Z \ , \qquad \text{(E12)}$$

for some M_X and M_Z which are independent of t since they can be defined in terms of the parameters M, k, $P_X^{t\,min}$ and $P_Z^{t\,min}$ which are all independent of t. (The distributions P_X^t and P_Z^t depend on t, but the

minimal probability in these distributions does not.) Then, for large enough t, $|c_1^t(x)|$ is small compared with the inputs which are of the form $x = mt$. Dividing by t to make this more apparent, the signal $y = x + c_1^t(x)$ after the first controller satisfies

$$\frac{y}{t} = m + \frac{c_1^t(mt)}{t} \tag{E13}$$

$$\leq m + \frac{M_X}{t} \tag{E14}$$

$$\approx m \tag{E15}$$

for large t since $M_X/t \to 0$ as $t \to \infty$. Now the final output satisfies

$$\frac{z}{t} = m + \frac{c_1^t(mt)}{t} + \frac{c_2^t(s)}{t} \tag{E16}$$

$$\approx m + \frac{c_2^t(s)}{t} \tag{E17}$$

$$\approx 0 \tag{E18}$$

for large t. The second line follows from the fact that $c_1^t(x)$ is uniformly bounded for all possible x (Lemma 2), and the final line follows from the fact that also $|z|$ is uniformly bounded. Hence, we have that $z/t \to 0$ as $t \to \infty$. Lemma 3 makes this approximation formal. In particular, when $t_0 = 2(M_X + M_Y) + 1$, for any $t \geq t_0$ we have

$$\left| \frac{x}{t} + \frac{c_2^t(s)}{t} \right| < \frac{1}{2} \tag{E19}$$

for all (x, s) such that $P_{X,S}(x, s) > 0$ (in particular, for all $x = mt$ with $m \in [q]$).

In our concrete example with $(q, d) = (6, 4)$ and $k = 1$ we can bound M_X and M_Z from above by $\sqrt{6M}$ and $\sqrt{54M}$, respectively. In this case, t_0 is bounded above by $20\sqrt{M} + 1$.

From now on let $t \geq t_0$. Note that (E19) means that $c_2^t(s)/t$ forms a good estimate for $m = x/t$. This allows us to construct a zero error encoding scheme for the channel $\mathcal{N}_t$ as in Figure 2. We use messages set $\mathcal{M} = [q]$ and let

$$\eta = \frac{-c_2^t(s)}{t} . \tag{E20}$$

We then define the encoding scheme $\mathcal{E}$ and decoding scheme $\mathcal{D}$ as

- $\mathcal{E}(m) = mt + c_1^t(mt)$.

- $\mathcal{D}(s)$ is given by rounding off η to the nearest integer.

Since $|m - \eta| < 1/2$ by (18), the nearest integer to η is always m and $\mathcal{D}$ always decodes correctly. Here the fact that we have used a particular distribution for the inputs does not matter. Since we achieve zero probability of error the encoding strategy must work with probability 1 for every input with positive probability. Put differently, our encoding strategy works with zero error for an alphabet of q symbols. This contradicts the fact that only $q - 1$ symbols can be sent over $\mathcal{N}_t$ with zero error. $\square$

2. Step 1: Deterministic strategies are optimal

Here we reduce to the case of deterministic strategies. Roughly, any strategy using shared randomness is a convex combination of deterministic strategies. All the strategies in the convex combination have their costs bounded below by the deterministic cost, and the convex combination cannot take the cost any lower.

Lemma 1. *Let $(\mathcal{N}, P_X, k)$ be an instance of the discrete Witsenhausen problem. Then*

$$C_D^{\mathcal{N}, P_X, k} = C_{SR}^{\mathcal{N}, P_X, k} . \tag{E21}$$

Proof. Let $(c_1, c_2, \rho) \in SR$ be a strategy that achieves $C_{SR}^{\mathcal{N}, P_X, k}$. Let us suppose that ρ is drawn from a set R and distributed according to some measure $d\rho$ with $\int_R d\rho = 1$. The cost function (1) can then be written as

$$C^{\mathcal{N}, P_X, k}(c_1, c_2, \rho) \tag{E22}$$

$$= \underset{P_X, \mathcal{N}, d\rho}{\mathbb{E}} \left[kc_1^2(x, \rho) + (c_2(s, \rho) + x + c_1(x, \rho))^2 \right] \tag{E23}$$

$$= \int_R \underset{P_X, \mathcal{N}}{\mathbb{E}} \left[kc_1^2(x, \rho) + (c_2(s, \rho) + x + c_1(x, \rho))^2 \right] d\rho \tag{E24}$$

For a fixed ρ, $(c_1(x, \rho), c_2(s, \rho), 0)$ is a deterministic strategy, and for each fixed ρ the integrand in the above is the cost of that deterministic strategy, which is bounded below by the deterministic cost. Hence the above is bounded below by

$$C_{SR}^{\mathcal{N}, P_X, k} = C^{\mathcal{N}, P_X, k}(c_1, c_2, \rho) \tag{E25}$$

$$\geq \int_R C_D^{\mathcal{N}, P_X, k} d\rho \tag{E26}$$

$$= C_D^{\mathcal{N}, P_X, k}. \tag{E27}$$

On the other hand, any deterministic strategy is also a shared randomness strategy in which ρ is ignored, so

$$C_D^{\mathcal{N}, P_X, k} \geq C_{SR}^{\mathcal{N}, P_X, k} . \tag{E28}$$

$\square$

3. Step 2: Bounding controller 1 and the output

Once we assume that there is some bound on the deterministic cost, then we can derive a bound on the output of c_1^t. This follows straightforwardly from the fact that the cost is an expectation over some finite support, so each possible output from c_1^t occurs with at least some minimal probability. As well, z's contribution to the cost is always positive, so we can omit it and the bound remains valid. Then, multiplying c_1^t by this minimal probability we must obtain some value which, when squared, is less than the bound. A similar argument applies to z.

Lemma 2. *Suppose* $C_D^{\mathcal{N}_t, P_X^t, k} \leq M$. *Let* $(c_1^t, c_2^t, 0) \in D$ *be a deterministic strategy such that* $C_D^{\mathcal{N}_t, P_X^t, k} = C^{\mathcal{N}_t, P_X^t, k}(c_1^t, c_2^t, 0)$. *Then for all* x *such that* $P_X^t(x) > 0$ *and* z *such that* $P_Z(z) > 0$

$$|c_1^t(x)| \leq M_X = \sqrt{\frac{M}{k P_X^{t\,min}}} \tag{E29}$$

$$|z| \leq M_Z = \sqrt{\frac{M}{P_Z^{t\,min}}} . \tag{E30}$$

Proof. By assumption, we have

$$M \geq C^{\mathcal{N}_t, P_X^t, k}(c_1^t, c_2^t, 0) . \tag{E31}$$

By definition and linearity of expectation,

$$C^{\mathcal{N}_t, P_X^t, k}(c_1^t, c_2^t, 0) \tag{E32}$$

$$= \mathop{\mathbb{E}}_{P_X^t, \mathcal{N}_t} \left[k(c_1^t(x))^2 + (c_2^t(s) + x + c_1^t(x))^2 \right] \tag{E33}$$

$$= \mathop{\mathbb{E}}_{P_X^t, \mathcal{N}_t} \left[k(c_1^t(x))^2 \right] + \mathop{\mathbb{E}}_{P_X^t, \mathcal{N}_t} \left[(c_2^t(s) + x + c_1^t(x))^2 \right] \tag{E34}$$

Since $(c_2^t(s) + x + c_1^t(x))^2 \geq 0$, we also have that $\mathop{\mathbb{E}}_{P_X^t, \mathcal{N}_t} \left[(c_2^t(s) + x + c_1^t(x))^2 \right] \geq 0$. Therefore, we have from (E34) that

$$C^{\mathcal{N}_t, P_X^t, k}(c_1^t, c_2^t, 0) \geq \mathop{\mathbb{E}}_{P_X^t, \mathcal{N}_t} \left[k(c_1^t(x))^2 \right] . \tag{E35}$$

Because c_1^t does not depend on $\mathcal{N}_t$, we have

$$\mathop{\mathbb{E}}_{P_X^t, \mathcal{N}_t} \left[k(c_1^t(x))^2 \right] = \mathop{\mathbb{E}}_{P_X^t} \left[k(c_1^t(x))^2 \right] \tag{E36}$$

Combining (E31), (E36) and (E35) we thus have

$$M \geq \mathop{\mathbb{E}}_{P_X^t} \left[k(c_1^t(x))^2 \right] \tag{E37}$$

$$= k \sum_{m \in [q]} P_X^t(x)(c_1^t(x))^2 \tag{E38}$$

$$\geq k \sum_{m \in [q]} P_X^{t\,min}(c_1^t(x))^2 \tag{E39}$$

where $P_X^{t\,min}$ is defined in (E3). All the terms in the final sum are positive, hence for all m

$$k P_X^{t\,min}((c_1^t(x))^2 \leq M , \tag{E40}$$

and thus

$$|c_1^t(x)| \leq \sqrt{\frac{M}{k P_X^{t\,min}}} . \tag{E41}$$

By an analogous argument

$$|z| \leq \sqrt{\frac{M}{P_Z^{t\,min}}} \tag{E42}$$

where $P_Z^{t\,min}$ is defined in (E4). $\qquad \square$

For the $\mathcal{N}^t$ and P_X^t defined in appendix C with $(q, d) = (6, 4)$ we can give explicit bounds on M_X and M_Z. Using $k = 1$ we find

$$M_X \leq \sqrt{6M} \tag{E43}$$

$$M_Z \leq \sqrt{54M}. \tag{E44}$$

4. Step 3: Approximating m

With $c_1^t(x)$ and z bounded, we find that c_2^t must have a very good estimate on $x = mt$. We divide out by t to make this more apparent.

Lemma 3. *Suppose* $C_D^{\mathcal{N}_t, P_X^t, k} \leq M$ *for all* $t \geq d$. *For each* t *let* $(c_1^t, c_2^t, 0) \in D$ *be a deterministic strategy such that* $C_D^{\mathcal{N}_t, P_X^t, k} = C^{\mathcal{N}_t, P_X^t, k}(c_1^t, c_2^t, 0)$. *Then there exists a* $t_0 \geq d$ *such that for all* $t \geq t_0$

$$\left| \frac{x}{t} + \frac{c_2^t(s)}{t} \right| < \frac{1}{2} \tag{E45}$$

for all (x, s) *such that* $P_{X,S}(x, s) > 0$.

Proof. As shown in Lemma 2, $|c_1^t(x)|$ and $|z| = |x + c_1^t(x) + c_2^t(s)|$ are uniformly bounded,

$$|c_1^t(x)| \leq M_X \tag{E46}$$

$$|x + c_1^t(x) + c_2^t(s)| \leq M_Z \tag{E47}$$

with M_X and M_Z independent of t. Combining these, we find

$$|x + c_2^t(s)| \leq M_X + M_Y \tag{E48}$$

Recall that $x = mt$ for some $m \in [q]$ for every x such that $P_X^t(x) > 0$. Dividing the above inequality by $t > 0$ to obtain

$$\left| m + \frac{c_2^t(s)}{t} \right| \leq \frac{M_X + M_Y}{t} \tag{E49}$$

for every (m, s) with $m \in [q]$. Then for $t \geq t_0 :=$ $2(M_X + M_Y) + 1$, we have our desired result. $\qquad \square$

We can make an estimate of t_0 for our explicit con-

struction with $(q, d) = (6, 4)$ and $k = 1$. Indeed

$$t_0 \leq 2(\sqrt{54} + \sqrt{6})\sqrt{M} + 1 \leq 20\sqrt{M} + 1. \qquad \text{(E50)}$$