

In memoriam Katalin Marton

Classical simulations of communication channels

Péter E. Frenkel

Eötvös Loránd University, Pázmány Péter sétány 1/C, Budapest, 1117 Hungary
and Rényi Institute, Budapest, Reáltanoda u. 13-15, 1053 Hungary

We investigate whether certain non-classical communication channels can be simulated by a classical channel with a given number of states and a given ‘amount’ of noise. It is proved that any noisy quantum channel can be simulated by a corresponding classical channel with ‘the same amount’ of noise. Classical simulations of general probabilistic channels are also studied.

Introduction

A communication protocol with l possible inputs and k possible outputs can be described by a *transition matrix* $A = (a_{ij}) \in [0, 1]^{k \times l}$, where a_{ij} is the conditional probability of output i if the input is j . This is a *stochastic* matrix: for all j , we have $\sum_{i=1}^k a_{ij} = 1$. A *communication channel* can be described by the set of transition matrices that it affords. Channel Q can be simulated by channel C if all transition matrices afforded by Q are convex combinations of transition matrices afforded by C . Such convex combinations occur naturally in information theory; they correspond to the sender and receiver having access to (unlimited) shared randomness. The relation ‘can be simulated by’ is obviously reflexive and transitive. Two channels are *equivalent* if each can be simulated by the other.

The *classical channel with n states* affords stochastic 0-1 matrices with at most n nonzero rows. The *quantum channel of level n* affords channel matrices of the form $(\text{tr } E_i \rho_j)$, where $\rho_1, \dots, \rho_l \in M_n(\mathbb{C})$ are *density matrices*, and $E_1, \dots, E_k \in M_n(\mathbb{C})$ is a *positive operator valued measure (POVM)*. It is easy to see that the classical channel with n states can be simulated by the quantum channel of level n . By [4, Theorem 3] of Weiner and the present author, the converse also holds. The present paper is about variants of this theorem for general probabilistic channels (Section 1) and for noisy quantum channels (Section 2). In Section 3, we discuss

Péter E. Frenkel: frenkelp265@gmail.com, Research partially supported by MTA Rényi “Lendület” Groups and Graphs Research Group, by ERC Consolidator Grant 648017 and by NKFIH grants K 124152 and KKP 139502.

noiseless classical simulations of noisy channels. Section 4 contains an open problem tentatively linking classical simulations of quantum channels to the more traditional way of comparing efficiency of classical and quantum communication, involving von Neumann entropy, mutual information and Holevo's inequality. The reader who is interested in quantum information theory but not in general probabilistic theory can safely skip Section 1.

Notations and terminology. The set $\{1, \dots, k\}$ is denoted by $[k]$. For a real number a , we write $a_+ = \max(a, 0)$. The indicator of an event A is written $\mathbb{1}(A)$. A *convex body* is a convex compact set with nonempty interior.

A matrix is *stochastic* if all entries are nonnegative reals and each column sums to 1. The set of n -square matrices with complex entries is written $M_n(\mathbb{C})$. The identity matrix is $\mathbf{1}$. A complex matrix A is *psdh* if it is positive semi-definite Hermitian, written $A \geq 0$. A *positive operator valued measure (POVM)* is a sequence E_1, \dots, E_k of psdh matrices summing to $\mathbf{1}$. A *density matrix* is a psdh matrix with trace 1.

For $0 \leq \delta \leq 1$, the δ -noisy classical channel with n states affords transition matrices of the form $EX \in [0, 1]^{k \times l}$, where $E \in \{0, 1\}^{k \times n}$ is a stochastic 0-1 matrix and X is an $n \times l$ stochastic matrix with each column containing $n - 1$ entries equal to δ/n . Here E can be interpreted as a classical decoding map $[n] \rightarrow [k]$, and the columns of X can be interpreted as extremal δ -noisy classical states. The presence of noise impedes the exact transmission of pure states, the pure state chosen by the sender is transmitted unchanged with probability $1 - (n - 1)\delta/n$ but turns into each one of the other $n - 1$ pure states with probability δ/n . Note that the 0-noisy classical channel with n states is the same as the classical channel with n states defined previously.

Following the terminology of [2, 3], the *signalling dimension* $\text{sign.dim } Q$ of a channel Q is the smallest positive integer n such that Q can be simulated by the (noiseless) classical channel with n states.

1 General probabilistic theory

Let S be a convex body in a finite dimensional real affine space. Let E be the cone of *effects*, i.e., affine linear functions $e : S \rightarrow [0, \infty)$. A *partition of unity* is a sequence $e_1, \dots, e_k \in E$ of effects such that $e_1 + \dots + e_k = 1$ (the constant 1 function). The *channel with state space S* affords transition matrices of the form $(e_i(x_j)) \in [0, 1]^{k \times l}$, where $x_1, \dots, x_l \in S$, and e_1, \dots, e_k is a partition of unity.

1.1 Signalling dimension vs. information storability

Following terminology introduced in [2], the *signalling dimension* $\text{sign.dim } S$ of S is the signalling dimension of the channel with state space S , i.e., the smallest positive integer n such that the channel with state space S can be simulated by the classical channel with n states. By [4, Theorem 3] mentioned in the Introduction, the signalling dimension of the set of n -square density matrices is n .

Calculating, or even efficiently estimating the signalling dimension of a given convex body seems to be a difficult problem, and strong general theorems are yet

to be searched for. In this section, we start with weak general results and work our way towards deeper results for special cases.

The *affine dimension* $\text{aff.dim } S$ of S is the minimal dimension of an affine space containing S . Adding 1, we get the *linear dimension* $\text{lin.dim } S$ of S , i.e., the dimension of the vector space of affine linear functions on S . For example, the affine dimension of the set of n -square density matrices is $n^2 - 1$, while its linear dimension is n^2 .

A partition of unity is *extremal* if it cannot be written as a convex combination of two partitions of unity in a nontrivial way. The nonzero effects appearing in an extremal partition of unity need not lie on extremal rays of the cone E of effects. When they do, a characterization of extremal partitions of unity is given in [2, Theorem 2]. We now give a necessary condition of extremality for a general partition of unity. Although this is implicitly contained in the paper cited above (see the proof given there), we include a proof.

Proposition 1.1. *The nonzero effects in an extremal partition of unity are linearly independent. Thus, their number is \leq the linear dimension of S .*

Proof. Let e_1, \dots, e_k be an extremal partition of unity. If $\lambda_1 e_1 + \dots + \lambda_k e_k = 0$ and $|\epsilon| \leq 1/\max\{|\lambda_i| : \lambda_i \neq 0\}$, then $(1 \pm \epsilon \lambda_1)e_1, \dots, (1 \pm \epsilon \lambda_k)e_k$ is also a partition of unity, which must coincide with e_1, \dots, e_k because of extremality. Thus $\lambda_i e_i = 0$ for all i . \square

Consider the transition matrix $A = (a_{ij}) \in [0, 1]^{k \times l}$ of some communication protocol, where a_{ij} is the conditional probability of output $i \in [k]$ if the input was $j \in [l]$. Let us try to guess the input based on the output, using a function $G : [k] \rightarrow [l]$. If input j occurs with probability q_j , then the probability of success will be

$$\sum_{j=1}^l q_j \sum_{i=1}^k a_{ij} \mathbb{1}(G(i) = j) = \sum_{i=1}^k q_{G(i)} a_{i, G(i)}.$$

Choosing the best possible guessing function G , the probability of success is

$$\sum_{i=1}^k \max_j q_j a_{ij}.$$

Without any communication, the probability of successfully guessing the input, with the optimal strategy, is $\max_j q_j$. The ratio $\sum_{i=1}^k \max_j q_j a_{ij} / \max_j q_j$ is maximized when $q_j = 1/l$ for all j , in which case it simplifies to $\sum_{i=1}^k \max_j a_{ij}$. Motivated by these considerations, and following [7] by Matsumoto and Kimura, the *information storability* $\text{inf.stor } S$ of S is defined to be the maximum of $\sum_{i=1}^k \max_j a_{ij}$ over all transition matrices (a_{ij}) afforded by S , or, equivalently, the maximum of $\sum_{i=1}^k \max_S e_i$ over all partitions of unity e_1, \dots, e_k . When taking these maxima, it suffices to consider extremal partitions of unity. Then Proposition 1.1 and a simple compactness argument shows that these maxima are attained.

As a simple example, let $S = [0, 1]$. Then we can choose the partition of unity $1 = x + (1 - x)$ to show that

$$\text{inf.stor } S \geq \max_{0 \leq x \leq 1} x + \max_{0 \leq x \leq 1} (1 - x) = 1 + 1 = 2.$$

On the other hand, as any affine linear function on S takes its maximum at 0 or 1, we have

$$\sum_{i=1}^k \max_S e_i = \sum_{e_i(0) \geq e_i(1)} e_i(0) + \sum_{e_i(0) < e_i(1)} e_i(1) \leq 1 + 1 = 2$$

for any partition of unity e_1, \dots, e_n on S , whence $\text{inf.stor } S = 2$. This is the easiest special case of [7, Theorems 1 and 4], cited below in relation to Theorem 1.2 and Proposition 1.3.

By [7, Theorem 4], $\text{inf.stor } S \leq \text{lin.dim } S$. We refine this inequality as follows.

Theorem 1.2. 1. $\text{inf.stor } S \leq \text{sign.dim } S \leq \text{lin.dim } S$.

2. If $\text{inf.stor } S \leq \text{aff.dim } S$, then $\text{sign.dim } S \leq \text{aff.dim } S$.

This theorem is closely related to [3, Theorem 1(i)].

Proof. (1) Let $n = \text{sign.dim } S$. Any transition matrix afforded by S is a convex combination of transition matrices afforded by the classical channel with n states. Such a matrix has $\leq n$ nonzero rows and therefore sum of row-maxima $\leq n$. This property is preserved when taking convex combinations. This proves the first inequality.

Any transition matrix afforded by S is a convex combination of transition matrices of the form $(e_i(x_j))$, where e_1, \dots, e_k is an *extremal* partition of unity, and $x_j \in S$. By Proposition 1.1, such a matrix has $\leq \text{lin.dim } S$ nonzero rows, and therefore is a convex combination of matrices afforded by the classical channel with $\text{lin.dim } S$ states. This proves the second inequality.

(2) Let $\text{inf.stor } S \leq \text{aff.dim } S = n$. Any transition matrix afforded by S is a convex combination of matrices of the form $A = (a_{ij}) \in [0, 1]^{k \times l}$, where $a_{ij} = e_i(x_j)$, e_1, \dots, e_k is an *extremal* partition of unity, and $x_j \in S$. We shall show that such an A is always a convex combination of transition matrices afforded by the classical channel with n states. Using Proposition 1.1, we may assume that $k = n + 1$. Set $m_i = \max_S e_i \in [0, 1]$ for each $i \in [k]$. Note that $\sum_{i=1}^k (1 - m_i) \geq n + 1 - \text{inf.stor } S \geq 1$. Choose a probability distribution p_1, \dots, p_k such that $p_i \leq 1 - m_i$ for all i . Then

$$p_i \leq 1 - a_{ij} = \sum_{i' \neq i} a_{i'j}$$

for all i and j , and

$$\sum_{i \in T} p_i \leq 1 = \sum_{i=1}^k a_{ij}$$

for all $T \subseteq [k]$.

For any fixed j , put supply a_{ij} and demand p_i at each node i of the complete (but loopless) graph on k nodes. Then, for the total supply at the neighbors of any subset $T \subseteq [k]$, we have

$$\sum_{i \in N(T)} a_{ij} \geq \sum_{i \in T} p_i.$$

By the Supply–Demand Theorem [6, 2.1.5. Corollary], the demands can be met: there exist stochastic column vectors $b_j(1), \dots, b_j(k)$ such that the i -th entry of $b_j(i)$ is zero for all i , and $\sum_{i=1}^k p_i b_j(i)$ is the j -th column of A . Now let $B(i)$ be the

matrix with columns $b_1(i), \dots, b_l(i)$. Then the i -th row of $B(i)$ is zero, so $B(i)$ has $\leq k - 1 = n$ nonzero rows, so $B(i)$ is a convex combination of transition matrices afforded by the classical channel with n states. Then so is A , since

$$A = \sum_{i=1}^k p_i B(i).$$

□

For the remainder of this section, assume that S is not just a point. A *chord* of S is a segment AB whose endpoints A and B belong to the boundary of S . We write AOB for a chord AB with a distinguished point O on the chord. The convex body S is *centrally symmetric* if there exists a point $O \in S$ such that for any chord AOB of S , we have $|OA| = |OB|$ for the lengths of the segments OA and OB . The *Minkowski measure of asymmetry* $\text{asymm } S$ of S is the smallest real number $m \geq 1$ such that there exists a point $O \in S$ such that for any chord AOB of S , we have $|OB| \leq m|OA|$.

By [7, Theorem 1] of Matsumoto and Kimura, the information storability is related to the Minkowski measure of asymmetry as follows.

Proposition 1.3. $\text{inf.stor } S = \text{asymm } S + 1$

Although this is a known statement, we include the sketch of a geometric proof for the convenience of the reader.

Proof. \leq : There exists a point $O \in S$ such that for any chord AOB of S , we have $|OB| \leq (\text{asymm } S)|OA|$. Let $n = \text{asymm } S + 1$. Then $e(x) \leq ne(O)$ for all $e \in E$ and $x \in S$, whence

$$\sum_{i=1}^k \max_S e_i \leq n \sum_{i=1}^k e_i(O) = n$$

for all partitions of unity e_1, \dots, e_k .

\geq : Let $n = \text{inf.stor } S$. Then $\sum_{i=1}^k \max_S e_i \leq n$ for all partitions of unity e_1, \dots, e_k . When k is the linear dimension of S , this tells us that for any simplex Δ containing S , there exists a point each of whose barycentric coordinates with respect to Δ is at least $1/n$ times the maximum value of that barycentric coordinate on S . Using Helly's theorem, we see that there exists a point O that divides the distance between any two parallel supporting hyperplanes of S in a ratio at least as equitable as $1 : (n - 1)$. Then, for any chord AOB of S with $|AO| \leq |OB|$, considering the supporting hyperplane of S at A and the parallel supporting hyperplane, we get that $|OB| \leq (n - 1)|OA|$. □

Corollary 1.4. *For the regular octahedron, we have $\text{asymm} = 1$, $\text{inf.stor} = 2$, $\text{sign.dim} = \text{aff.dim} = 3$, and $\text{lin.dim} = 4$.*

Proof. The regular octahedron is centrally symmetric, which means that $\text{asymm} = 1$. By Proposition 1.3, we have $\text{inf.stor} = \text{asymm} + 1 = 2$. Obviously, $\text{aff.dim} = 3$ and $\text{lin.dim} = \text{aff.dim} + 1 = 4$.

By Theorem 1.2(2), we have $\text{sign.dim} \leq 3$. To prove the converse inequality, let

$$X = \begin{pmatrix} 1 & -1 & & & & \\ & & 1 & -1 & & \\ & & & & 1 & -1 \end{pmatrix}$$

be the matrix whose columns are the vertices of the octahedron (the entries not shown are zero). Let

$$V = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix},$$

then

$$VX = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}.$$

Adding 1 to each entry and dividing by 4, we get the stochastic matrix

$$A = \frac{1}{2} \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix},$$

which is therefore a transition matrix afforded by the octahedron. Since any two rows of A have an $1/2$ at the same position, we have

$$\sum_{1 \leq i < i' \leq 4} \max_{1 \leq j \leq 6} (a_{ij} + a_{i'j}) = \binom{4}{2} = 6.$$

On the other hand, any 4×6 transition matrix afforded by the classical channel with 2 states has at least $4 - 2 = 2$ zero rows, so the sum above would be $\leq \binom{4}{2} - \binom{4-2}{2} = 5$ — note that this is a special case of [4, inequality (3.6)]. This inequality is preserved under convex combinations. Therefore, the octahedron cannot be simulated by the classical 2-state channel, hence its signalling dimension is (at least) 3. \square

1.2 Noisy balls

If an origin is chosen in S , and $0 \leq \delta \leq 1$, then the δ -noisy channel with state space S affords the transition matrices $(e_i(x_j))$, where e_1, \dots, e_k is a partition of unity and $x_j \in (1 - \delta)S$ for all j . This is analogous to the partial depolarization channel in quantum information theory, cf. Subsection 3.1. Note that $e_i \geq 0$ is required on all of S .

It is easy to see that if $S' = f(S)$ is an affine image of S , then S' can be simulated by S . If, in addition, $O' = f(O)$, then δ -noisy S' can be simulated by δ -noisy S . In particular, a classical bit can be simulated by S unless S is just a point, and a δ -noisy classical bit can be simulated by any δ -noisy $S \neq \{O\}$ that is symmetric with respect to O .

Theorem 1.5. *Let n be an even positive integer. Put*

$$S = \{x \in \mathbb{R}^d : \|x\|_{n/(n-1)} \leq 1\},$$

the unit ball of the $n/(n-1)$ -norm. Let $0 \leq \delta \leq 1$.

1. *The δ -noisy channel with state space S can be simulated by the δ -noisy classical channel with n states.*
2. *The signalling dimension of S is $\leq n$.*
3. *For an ellipsoid of arbitrary affine dimension ≥ 1 , the signalling dimension is 2. A δ -noisy ellipsoid can be simulated by a δ -noisy classical bit.*

The proof below is similar to that of [4, Theorem 3]. However, the mixed discriminant used there (and used in Section 2 of the present paper) must be replaced by a different n -linear symmetric function $\{\cdot, \dots, \cdot\}$.

To introduce $\{\cdot, \dots, \cdot\}$, we can think of an affine linear function $e : S \rightarrow \mathbb{R}$ as a formal sum of a number and a vector: $e = c + v \in \mathbb{R} \oplus \mathbb{R}^d = \mathbb{R}^{d+1}$, meaning that $e(x) = c + vx$ for $x \in S$, where vx is the usual inner product. For an effect $e \in E$, the condition $e \geq 0$ translates to $\|v\|_n \leq c$ because

$$(n/(n-1))^{-1} + n^{-1} = 1.$$

Given $e_1, \dots, e_n \in \mathbb{R}^{d+1}$, where $e_i = c_i + v_i$, we define

$$\{e_1, \dots, e_n\} = c_1 \cdots c_n - v_1 \cdots v_n,$$

where $v_1 \cdots v_n$ means that we take the coordinatewise product and then add up the coordinates (which is an n -linear generalization of the usual inner product). For $n = 2$, $\{\cdot, \cdot\}$ is the Lorentzian indefinite symmetric bilinear product well known from the special theory of relativity. For general n , $\{\cdot, \dots, \cdot\}$ is symmetric, multilinear and $\{1, \dots, 1\} = 1$. When $e_1, \dots, e_n \in E$, we have $\{e_1, \dots, e_n\} \geq 0$ by repeated application of Hölder's inequality. Further, if $0 \leq e \leq 1$ holds pointwise on S , then writing $e = c + v$ and $a = \|v\|_n$, we have $0 \leq a \leq \min(c, 1 - c)$ and therefore

$$\begin{aligned} \{e, \dots, e\} &= c^n - v^n \stackrel{*}{=} c^n - a^n = \\ &= (c - a)(c^{n-1} + c^{n-2}a + \dots + ca^{n-2} + a^{n-1}) \leq \\ &\leq (c - a)(c + (1 - c))^{n-1} = c - a = \min_{x \in S} e(x). \end{aligned}$$

Note that the equality marked by a $*$ holds because n is even.

We are now ready to start the proof of Theorem 1.5.

Proof. (1) Let $A \in [0, 1]^{k \times l}$ be a δ -noisy transition matrix afforded by S , i.e.,

$$a_{ij} = e_i((1 - \delta)x_j),$$

where $x_1, \dots, x_l \in S$, $e_i \in E$, and $e_1 + \dots + e_k = 1$. We shall prove that A is a convex combination of δ -noisy n -state classical transition matrices.

If $e_i = c_i + v_i$ as before, then $c_1 + \cdots + c_k = 1$, $v_1 + \cdots + v_k = 0$, and

$$a_{ij} = c_i + (1 - \delta)v_i x_j = \delta c_i + (1 - \delta)e_i(x_j),$$

so $A = \delta C + (1 - \delta)A'$, where C is the matrix with entries $c_{ij} = c_i$ not depending on j , and A' is the matrix with entries $a'_{ij} = e_i(x_j)$.

For $I = (i_1, \dots, i_n) \in [k]^n$, put

$$p_I = \{e_{i_1}, \dots, e_{i_n}\}.$$

We have $p_I \geq 0$ for all I . Thus, we get a measure P on $[k]^n$ defined by

$$P(T) = \sum_{I \in T} p_I.$$

Using the multilinearity of the bracket and the assumption that e_1, \dots, e_k is a partition of unity, we see that

$$P([k]^n) = \{1, \dots, 1\} = 1,$$

so P is a probability measure.

Let $D(I)$ be the matrix with entries $d(I)_{ij} = m(i, I)/n$ not depending on j , where $m(i, I)$ is the number of occurrences of i in the sequence I . Then $\int D dP = C$ because

$$\int d_{ij} dP = \sum_{I \in [k]^n} p_I m(i, I)/n = \{e_i, 1, \dots, 1\} = c_i = c_{ij}.$$

For any $R \subseteq [k]$, we may put $e_R = \sum_{i \in R} e_i$, and then we have

$$P(R^n) = \{e_R, \dots, e_R\} \leq \min_{x \in S} e_R(x) \leq e_R(x_j)$$

for all j since $0 \leq e_R \leq 1$. The right hand side here is $A'_j(R)$, where A'_j is the probability measure on $[k]$ given by the numbers $e_i(x_j)$. So we have

$$A'_j(R) \geq P(R^n) \quad \text{for all } R \subseteq [k].$$

Let us connect $I \in [k]^n$ to $i \in [k]$ by an edge if i occurs in I . This gives us a bipartite graph. The neighborhood of any set $T \subseteq [k]^n$ is the set $R \subseteq [k]$ of indices occurring in some element of T . We always have $T \subseteq R^n$, whence

$$A'_j(R) \geq P(R^n) \geq P(T).$$

Thus, by the Supply–Demand Theorem [6, 2.1.5. Corollary], and using the fact that both A'_j and P are probability measures, there exists a probability measure P_j on $[k]^n \times [k]$ which is supported on the edges of the graph and has marginals P and A'_j . Whenever $p_I \neq 0$, let $B'(I)$ be the $k \times k$ stochastic matrix whose j -th column is given by the conditional distribution $P_j|I$ on $[k]$. We have $A' = \int B' dP$.

Now $B(I) = \delta D(I) + (1 - \delta)B'(I)$ is a convex combination of δ -noisy n -state classical transition matrices, and, in turn, $A = \int B dP$ is a convex combination of the $B(I)$, as desired.

(2) Set $\delta = 0$ in (1).

(3) The signalling dimension of an ellipsoid is the same as that of the Euclidean unit ball. This is ≤ 2 by (2), and is ≥ 2 because the unit ball is not a point. The noisy claim follows from (1). \square

2 Noisy quantum channels

Let

$$K \subseteq \Delta_n = \{(\xi_1, \dots, \xi_n) : \xi_i \geq 0 \text{ for all } i, \xi_1 + \dots + \xi_n = 1\}$$

be a convex set of probability distributions that is invariant under all permutations of the n coordinates. The K -noisy classical channel affords transition matrices of the form $EX \in [0, 1]^{k \times l}$, where $X \in K^l$ is an $n \times l$ matrix with all columns in K , and E is a $k \times n$ stochastic 0-1 matrix. A density matrix is K -noisy if the sequence of its eigenvalues is in K . The K -noisy quantum channel affords transition matrices of the form $(\text{tr } E_i \rho_j)$, where E_1, \dots, E_k is a POVM and ρ_j is a K -noisy density matrix for $j = 1, \dots, l$.

It is easy to see that the K -noisy classical channel can be simulated by the K -noisy quantum channel. Our goal is to prove the converse, which is a far-reaching generalization of [4, Theorem 3] mentioned in the Introduction.

In fact, we may generalize further. Let $K_j \subseteq \Delta_n$ ($j = 1, \dots, l$) be convex sets, each of them invariant under all permutations of the n coordinates. The (K_1, \dots, K_l) -noisy classical channel affords transition matrices of the form $EX \in [0, 1]^{k \times l}$, where $X \in K_1 \times \dots \times K_l$ is an $n \times l$ matrix with j -th column in K_j , and E is a $k \times n$ stochastic 0-1 matrix. The (K_1, \dots, K_l) -noisy quantum channel affords transition matrices of the form $(\text{tr } E_i \rho_j)$, where E_1, \dots, E_k is a POVM and ρ_j is a K_j -noisy density matrix for $j = 1, \dots, l$.

It is easy to see that the (K_1, \dots, K_l) -noisy classical channel can be simulated by the (K_1, \dots, K_l) -noisy quantum channel. We shall prove the converse.

As in [4], our main tool is the *mixed discriminant*, the unique symmetric n -linear function D on $M_n(\mathbb{C})$ such that $D(E, \dots, E) = \det E$ for all $E \in M_n(\mathbb{C})$. Explicitly, if $E_i = [e_i^1, \dots, e_i^n]$ are the columns, then

$$D(E_1, \dots, E_n) = \frac{1}{n!} \sum_{\pi \in \mathfrak{S}_n} \det [e_{\pi(1)}^1, \dots, e_{\pi(n)}^n]. \quad (2.1)$$

We shall need the following inequalities.

Lemma 2.1. *For $\lambda_1, \dots, \lambda_n \in [0, 1]$ and $r = 1, 2, \dots, n$, we have*

$$\sum_{Q \subseteq [n]} (r - |Q|)_+ \prod_{m \notin Q} \lambda_m \prod_{m \in Q} (1 - \lambda_m) \leq \lambda_1 + \dots + \lambda_r, \quad (2.2)$$

where $a_+ = \max(a, 0)$.

Proof. We have

$$(r - |Q|)_+ \leq |[r] \setminus Q| = \sum_{s=1}^r \mathbb{1}(s \notin Q)$$

for all Q . Thus, the left hand side of (2.2) is

$$\leq \sum_{s=1}^r \sum_{Q \subseteq [n] \setminus \{s\}} \prod_{m \notin Q} \lambda_m \prod_{m \in Q} (1 - \lambda_m) = \sum_{s=1}^r \lambda_s \prod_{m \neq s} (\lambda_m + (1 - \lambda_m)) = \lambda_1 + \dots + \lambda_r.$$

□

Lemma 2.2. For an n -square Hermitian matrix $0 \leq E \leq \mathbf{1}$ with eigenvalues $\lambda_1, \dots, \lambda_n$, and $r = 1, 2, \dots, n$, we have

$$\sum_{q=0}^{r-1} (r-q) \binom{n}{q} D(\underbrace{E, \dots, E}_{n-q}, \underbrace{\mathbf{1} - E, \dots, \mathbf{1} - E}_q) \leq \lambda_1 + \dots + \lambda_r.$$

Proof. Since the spectrum and the mixed discriminant are both invariant under unitary conjugation, we may assume that E is a diagonal matrix. Then (2.1) reduces Lemma 2.2 to Lemma 2.1. \square

By Bapat's [1, Lemma 2(vi)], if E_1, \dots, E_n are all positive semidefinite Hermitian matrices, then

$$D(E_1, \dots, E_n) \geq 0. \quad (2.3)$$

Given a POVM $E_1, \dots, E_k \in M_n(\mathbb{C})$, we define

$$p_I = D(E_{i_1}, \dots, E_{i_n}) \quad (2.4)$$

for all $I = (i_1, \dots, i_n) \in [k]^n$. By multilinearity and (2.3), this defines a probability distribution on $[k]^n$.

Lemma 2.3. If $E_1, \dots, E_k \in M_n(\mathbb{C})$ is a POVM, u_1, \dots, u_k are real numbers, and $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $E = \sum_{i=1}^k u_i E_i$, then

$$\sum_{I \in [k]^n} p_I \min \left\{ \sum_{m \in S} u_{i_m} : S \subseteq [n], |S| = r \right\} \leq \lambda_1 + \dots + \lambda_r \quad (2.5)$$

for all $r = 1, 2, \dots, n$.

Proof. We may assume that all $u_i \geq 0$ because adding u to all u_i adds ru to both sides of (2.5). We may assume $u_1 \geq \dots \geq u_k$. Put $u_{k+1} = 0$. Write $E = \sum_{i=1}^k v_i F_i$, where $v_i = u_i - u_{i+1}$ and $F_i = E_1 + \dots + E_i$.

Let σ_i be the sum of the r smallest eigenvalues of F_i . Then

$$\sum_{i=1}^k v_i \sigma_i \leq \lambda_1 + \dots + \lambda_r. \quad (2.6)$$

As $0 \leq F_i \leq \mathbf{1}$, we have

$$\sum_{q=0}^{r-1} (r-q) \binom{n}{q} D(\underbrace{F_i, \dots, F_i}_{n-q}, \underbrace{\mathbf{1} - F_i, \dots, \mathbf{1} - F_i}_q) \leq \sigma_i \quad (2.7)$$

for all i , by Lemma 2.2.

On the other hand, since $u_i = v_i + \dots + v_k$, we have

$$\min \left\{ \sum_{m \in S} u_{i_m} : S \subseteq [n], |S| = r \right\} = \sum_{i=1}^k v_i (r - |\{m \in [n] : i_m > i\}|)_+.$$

It remains to check that

$$\begin{aligned} & \sum_{I \in [k]^n} p_I (r - |\{m \in [n] : i_m > i\}|)_+ = \\ &= \sum_{q=0}^{r-1} (r-q) \binom{n}{q} D(\underbrace{F_i, \dots, F_i}_{n-q}, \underbrace{\mathbf{1} - F_i, \dots, \mathbf{1} - F_i}_q) \end{aligned}$$

for all $i \in [k]$. This follows from

$$\begin{aligned} & \sum (p_I : I \in [k]^n, |\{m \in [n] : i_m > i\}| = q) = \\ &= \binom{n}{q} D(\underbrace{F_i, \dots, F_i}_{n-q}, \underbrace{\mathbf{1} - F_i, \dots, \mathbf{1} - F_i}_q), \end{aligned}$$

which is clear from the definitions of p_I and F_i , and from the symmetry and multilinearity of D . \square

We are ready for the main result of this paper.

Theorem 2.4. *The (K_1, \dots, K_l) -noisy quantum channel can be simulated by the (K_1, \dots, K_l) -noisy classical channel. In particular, the K -noisy quantum channel can be simulated by the K -noisy classical channel.*

Proof. It suffices to prove that for any POVM E_1, \dots, E_k , and any K -noisy density matrix ρ , there exist points $x_I = (x_{I,1}, \dots, x_{I,n}) \in K$ for each $I = (i_1, \dots, i_n) \in [k]^n$ such that

$$\text{tr } E_i \rho = \sum_{I \in [k]^n} p_I \sum (x_{I,m} : m \in [n], i_m = i) \quad (2.8)$$

for each $i \in [k]$. Here the p_I are defined as in (2.4).

Let the eigenvalues of ρ be $0 \leq \mu_1 \leq \dots \leq \mu_n$; we have $\mu_1 + \dots + \mu_n = 1$. Since ρ is K -noisy, we have $\mu = (\mu_1, \dots, \mu_n) \in K$. Since K is convex and invariant with respect to permutations, any convex combination of permutations of μ is in K . Thus, if $x \in [0, 1]^n$ is a stochastic vector, and any r distinct coordinates of x sum to $\geq \mu_1 + \dots + \mu_r$ for each $r = 1, 2, \dots, n$, then $x \in K$. If we require

- these 2^n inequalities for each x_I , together with
- $x_{I,m} \geq 0$ for all I and m , and
- (2.8) for all i ,

then each x_I will be a stochastic vector since setting $r = n$ yields

$$x_{I,1} + \dots + x_{I,n} \geq \mu_1 + \dots + \mu_n = 1,$$

while summing (2.8) for $i = 1, 2, \dots, k$ yields

$$1 = \sum_{I \in [k]^n} p_I (x_{I,1} + \dots + x_{I,n}).$$

Therefore, it suffices to prove that the system of $(2^n + n)k^n$ inequalities and k equations above has a solution. By the well-known Farkas Lemma, this is equivalent to saying that a linear combination of the inequalities and equations in the system cannot lead to the contradictory inequality $0 \geq 1$. That is, it suffices to prove that if nonnegative numbers $w_{I,H}$ ($I \in [k]^n, H \subseteq [n]$) and real numbers u_1, \dots, u_k satisfy

$$\sum (w_{I,H} : H \subseteq [n], H \ni m) \leq p_I u_{i_m} \quad (2.9)$$

for all $I \in [k]^n$ and all $m \in [n]$, then

$$\sum_{I \in [k]^n} \sum_{H \subseteq [n]} w_{I,H} (\mu_1 + \dots + \mu_{|H|}) \leq \sum_{i=1}^k u_i \operatorname{tr} E_i \rho. \quad (2.10)$$

Let $\lambda_1 \leq \dots \leq \lambda_n$ be the eigenvalues of $u_1 E_1 + \dots + u_k E_k$. By von Neumann's inequality, the right hand side of (2.10) is

$$\geq \lambda_1 \mu_n + \dots + \lambda_n \mu_1.$$

The coefficient of any μ_t on the left hand side of (2.10) is

$$\sum_{I \in [k]^n} \sum_{|H| \geq t} w_{I,H},$$

so it suffices to prove that

$$\sum_{t=n-r+1}^n \sum_{I \in [k]^n} \sum_{|H| \geq t} w_{I,H} \leq \lambda_1 + \dots + \lambda_r$$

for $r = 1, \dots, n$. In view of Lemma 2.3, this follows if

$$\sum_{t=n-r+1}^n \sum_{|H| \geq t} w_{I,H} \leq p_I \sum_{m \in S} u_{i_m}$$

for all $I \in [k]^n$ and all $S \subseteq [n]$ with $|S| = r$. This follows from (2.9) and the fact that

$$\sum_{n-r < t \leq |H|} 1 = (|H| + r - n)_+ \leq |S \cap H| = \sum_{m \in S \cap H} 1$$

for all $H, S \subseteq [n]$ with $|S| = r$. □

3 Simulation of a noisy channel by a noiseless one

Given the K -noisy channel, we might try to determine its signalling dimension, i.e., simulate it by a noiseless classical channel with as few states as possible. In view of Theorem 2.4, it makes no difference whether the given channel is classical or quantum.

Theorem 3.1. *The K -noisy classical (or, equivalently, quantum) channel can be simulated by the noiseless d -state classical (or, equivalently, level d quantum) channel if and only if we have*

$$\mu_1 + \dots + \mu_r \geq \binom{r}{d} / \binom{n}{d} \quad (3.1)$$

for all $\mu = (\mu_1 \leq \dots \leq \mu_n) \in K$ and all integers $d \leq r \leq n$.

Proof. ‘Only if’: Let $\mu = (\mu_1 \leq \dots \leq \mu_n) \in K$. Let $A = (a_{ij})$ be an $n \times n!$ stochastic matrix whose columns are the $n!$ permutations of μ . Then A is a transition matrix afforded by the K -noisy channel, thus also by the noiseless d -state channel. By [4, Section 3], we then have

$$\binom{n}{r}(\mu_1 + \dots + \mu_r) = \sum_{|S|=r} \min_{j \in [l]} \sum_{i \in S} a_{ij} \geq \binom{n-d}{n-r}$$

for all $d \leq r \leq n$, which is equivalent to (3.1).

‘If’: Let S be a uniform random d -element subset of $[n]$. It suffices to prove that, for any $\mu \in K$, there is a random element m of S whose distribution is given by $\mathbb{P}(m = r) = \mu_r$ for all $r = 1, \dots, n$. We may assume $\mu_1 \leq \dots \leq \mu_n$. Let $\nu_r = \mathbb{P}(\max S = r)$, then

$$\nu_1 + \dots + \nu_r = \mathbb{P}(\max S \leq r) = \binom{r}{d} / \binom{n}{d} \leq \mu_1 + \dots + \mu_r,$$

so μ is a convex combination of the permutations of ν . But ν is the distribution of the greatest element of S , so each permutation of ν is the distribution of an element of S , thus μ is the distribution of a random element of S , as claimed. \square

For $0 \leq \delta \leq 1$, the δ -noisy quantum channel of level n affords transition matrices of the form $(\text{tr } E_i \rho_j)$, where $E_1, \dots, E_k \in M_n(\mathbb{C})$ is a POVM and $\rho_1, \dots, \rho_l \in M_n(\mathbb{C})$ are density matrices with all eigenvalues $\geq \delta/n$. This channel is equivalent to the δ -noisy classical channel with n states. This is a special case of Theorem 2.4. Alternatively, it can be shown by combining ideas from the proofs of Theorem 1.5(1) and [4, Theorem 3].

Corollary 3.2. *Let $0 \leq \delta \leq 1$. The signalling dimension of the δ -noisy n -state classical (or, equivalently, n -level quantum) channel is $\lceil (1 - \delta)n + \delta \rceil$.*

Proof. The δ -noisy n -state classical channel can be simulated by the noiseless d -state classical channel if and only if we have

$$r\delta/n \geq \binom{r}{d} / \binom{n}{d} \tag{3.2}$$

for all integers $d \leq r \leq n - 1$ — note that both sides of (3.1) are 1 for $r = n$. In inequality (3.2), the left hand side is linear in r , while the right hand side is convex for $r = 0, 1, \dots$. Also, the inequality holds for $r = 0, 1, \dots, d - 1$. Therefore, it holds for all integers $d \leq r \leq n - 1$ if and only if it holds for $r = n - 1$, i.e., $(n - 1)\delta/n \geq (n - d)/n$, or, equivalently, $d \geq (1 - \delta)n + \delta$. \square

3.1 Partial replacer quantum channels

The usual mathematical model for a noisy quantum channel is given in terms of a completely positive trace-preserving map $\mathcal{N} : M_m(\mathbb{C}) \rightarrow M_n(\mathbb{C})$. Let $\text{ran } \mathcal{N}$ stand for the set of density matrices $\mathcal{N}(\sigma) \in M_n(\mathbb{C})$, where $\sigma \in M_m(\mathbb{C})$ is a density matrix. The channel affords transition matrices of the form $(\text{tr } E_i \rho_j) \in [0, 1]^{k \times l}$,

where E_1, \dots, E_k is a POVM in $M_n(\mathbb{C})$ and each ρ_j is contained in $\text{ran } \mathcal{N}$. The *signalling dimension* $\text{sign.dim } \mathcal{N}$ of \mathcal{N} is the signalling dimension of this channel, i.e., the smallest d such that the channel can be simulated by the noiseless classical channel with d states. If the spectrum of every $\rho \in \text{ran } \mathcal{N}$ is contained in a given permutation-invariant set $K \subseteq \Delta_n$, then every transition matrix afforded by \mathcal{N} is also afforded by the K -noisy quantum channel, so we can use Theorem 2.4 to show that \mathcal{N} can be simulated by the K -noisy classical channel. Then Theorem 3.1 can be used to give an upper bound on the signalling dimension of \mathcal{N} .

An important special case is given by *partial replacer channels*. Let $m \leq n$. We embed $M_m(\mathbb{C})$ into $M_n(\mathbb{C})$ as the set of matrices that are zero outside of the upper left m -square block. We fix a density matrix $\rho \in M_n(\mathbb{C})$. The *replacer channel* $\mathcal{N}_\rho : M_m(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ is given by $\mathcal{N}_\rho(X) = (\text{tr } X)\rho$. Given $0 \leq \delta \leq 1$, the *partial replacer channel* $\mathcal{N}_\rho(\delta) : M_m(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ is given by $\mathcal{N}_\rho(\delta)(X) = (1 - \delta)X + \delta(\text{tr } X)\rho$. In [3, Theorem 3] by Doolittle and Chitambar, it is shown that

$$\lceil (1 - \delta)m + \delta \rceil \leq \text{sign.dim } \mathcal{N}_\rho(\delta) \leq \min\{m, \lceil (1 - \delta)m + 1 \rceil\}, \quad (3.3)$$

and the upper bound is tight for the *partial erasure channel* given by the *erasure flag* ρ which has entry 1 at position $(m + 1, m + 1)$ and zero elsewhere. Note that the difference between the upper and the lower bound in (3.3) is at most 1.

We shall now prove that the lower bound is tight if $m = n$ and ρ is sufficiently mixed, in particular, if $\rho = \mathbf{1}/n$ is the *maximally mixed state*, yielding the *partial depolarization channel*

$$\mathcal{N}(\delta)(X) = (1 - \delta)X + (\delta/n)(\text{tr } X)\mathbf{1}.$$

From now on, we let $m = n$. Let $d = \lceil (1 - \delta)n + \delta \rceil$ stand for the lower bound in (3.3). Let $\mu_1 \leq \dots \leq \mu_n$ stand for the eigenvalues of a fixed density matrix ρ .

Proposition 3.3. 1. If $\delta(\mu_1 + \dots + \mu_r) \geq \binom{r}{d} / \binom{n}{d}$ holds for $r = d, \dots, n - 1$, then $\text{sign.dim } \mathcal{N}_\rho(\delta) = d$.

2. The partial depolarization channel is equivalent to the δ -noisy classical channel with n states.

3. The signalling dimension of the partial depolarization channel is d .

Proof. (1) The eigenvalues $\mu'_1 \leq \dots \leq \mu'_n$ of $\mathcal{N}_\rho(\delta)(\sigma) = (1 - \delta)\sigma + \delta\rho \geq \delta\rho$ satisfy $\mu'_1 + \dots + \mu'_r \geq \delta(\mu_1 + \dots + \mu_r)$ for any density matrix $\sigma \in M_m(\mathbb{C})$ and any $r = 1, \dots, n$. Thus, $\mu'_1 + \dots + \mu'_r \geq \binom{r}{d} / \binom{n}{d}$ for $r = 1, \dots, n - 1$, but also, trivially, for $r = n$. The claim now follows from Theorem 3.1 together with the first inequality in (3.3).

(2) The range $\text{ran } \mathcal{N}(\delta)$ is the set of density matrices with all eigenvalues $\geq \delta/n$, so the claim follows from Theorem 2.4.

(3) follows from (2) together with Corollary 3.2. \square

4 Future research

It is well known that quantum communication can outperform classical communication if entanglement is used cleverly. On the other hand, in certain scenarios

not involving entanglement, it can be proved that passing from classical to quantum cannot increase efficiency.

A fundamental result in this direction is the Holevo bound [5] which we now recall. For any stochastic matrix $A = (a_{ij}) \in [0, 1]^{k \times l}$ and input probabilities $q_j \geq 0$ ($j = 1, \dots, l$) summing to 1, we define the *mutual information*

$$\text{Info}(A, q) = H(j) + H(i) - H(i, j).$$

Here H stands for the Shannon entropy of a random variable, and the joint distribution of the random pair (i, j) is given by the probabilities $q_j a_{ij}$. Now, for any density matrices $\rho_j \in M_n(\mathbb{C})$ and any POVM $E_1, \dots, E_k \in M_n(\mathbb{C})$, the Holevo inequality reads

$$\text{Info}(A, q) \leq \chi, \tag{4.1}$$

where $a_{ij} = \text{tr } E_i \rho_j$ and the Holevo quantity χ is defined by

$$\chi = S \left(\sum_{j=1}^l q_j \rho_j \right) - \sum_{j=1}^l q_j S(\rho_j),$$

where S is von Neumann entropy, i.e., the Shannon entropy of the spectrum. If all ρ_j with $q_j > 0$ commute, then a POVM E_1, \dots, E_k can be found so that equality holds in (4.1). Otherwise, the inequality is strict for any POVM.

Another result in the above mentioned direction is [4, Theorem 3]: the n -level quantum channel can be simulated by the n -state classical channel.

It would be nice to unify these two results. Let a probability distribution q_1, \dots, q_l be given. Can every quantum transition matrix $A = (a_{ij}) = (\text{tr } E_i \rho_j) \in [0, 1]^{k \times l}$, where $E_1, \dots, E_k \in M_n(\mathbb{C})$ is a POVM, and $\rho_1, \dots, \rho_l \in M_n(\mathbb{C})$ are density matrices, be written as a convex combination $A = \sum p_I A_I$ of stochastic matrices A_I , each with $\leq n$ nonzero rows, and each satisfying $\text{Info}(A_I, q) \leq \chi$? Can the proof of Theorem 2.4 be modified to yield this result and thus, maybe, a new proof of Holevo's inequality?

Acknowledgement. I am grateful to Mihály Weiner for useful conversations.

References

- [1] R.B.Bapat: Mixed discriminants of positive semidefinite matrices. *Linear Algebra Appl.* **126** (1989), 107–124. [https://doi.org/10.1016/0024-3795\(89\)90009-8](https://doi.org/10.1016/0024-3795(89)90009-8)
- [2] Michele Dall'Arno, Sarah Brandsen, Alessandro Tosini, Francesco Buscemi, and Vlatko Vedral: No-Hypersignaling Principle, *Phys. Rev. Lett.* 119 (2017), 020401. <https://doi.org/10.1103/PhysRevLett.119.020401>
- [3] Brian Doolittle, Eric Chitambar: Certifying the Classical Simulation Cost of a Quantum Channel, *Phys. Rev. Research* 3, 043073. <https://doi.org/10.1103/PhysRevResearch.3.043073>
- [4] P. E. Frenkel and M. Weiner: Classical information storage in an n -level quantum system, *Communications in Mathematical Physics* 340 (2015), 563–574. <https://doi.org/10.1007/s00220-015-2463-0>

- [5] A. S. Holevo: Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel, *Probl. Peredachi Inf.*, 9:3 (1973), 3–11; *Problems Inform. Transmission*, 9:3 (1973), 177–183.
- [6] L. Lovász and M. D. Plummer: *Matching Theory*. North-Holland, 1986.
- [7] Keiji Matsumoto, Gen Kimura: Information-induced asymmetry of state space in view of general probabilistic theories, <https://doi.org/10.48550/arXiv.1802.01162>