

The coherent measurement cost of coherence distillation

Varun Narasimhachar

Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Republic of Singapore 138632

10 April 2025

Abstract

Quantum coherence—an indispensable resource for quantum technologies—is known to be distillable from a noisy form using operations that cannot create it. However, distillation exacts a hidden coherent *measurement* cost, which has not previously been examined. We devise the *target effect* construction to characterize this cost through detailed conditions on the coherence-measuring structure necessary in any process realizing exact (maximal or non-maximal) or approximate distillation. As a corollary, we lower-bound the requisite measurement coherence, as quantified by operationally-relevant measures. We then consider the asymptotic limit of distilling from many copies of a given noisy coherent state, where we offer rigorous arguments to support the conjecture that the (necessary and sufficient) coherent measurement cost scales extensively in the number of copies. We also show that this cost is no smaller than the coherence of measurements saturating the scaling law in the generalized quantum Stein’s lemma. Our results and conjectures apply to any task whereof coherence distillation is an incidental outcome (e.g., incoherent randomness extraction). But if pure coherence is the only desired outcome, our conjectures would have the cautionary implication that the measurement cost is often higher than the distilled yield, in which case coherence should rather be prepared afresh than distilled from a noisy input.

1 Introduction

Coherence is a cornerstone of quantum mechanics, playing a central role in the wavelike interference effects that epitomize quantum phenomena. It is also a valuable resource, powering transformative quantum technologies such as quantum computing, quantum communication, and quantum metrology [1].

A central concept in quantum information science is that of *resource distillation*: the conversion of a resource (like coherence) from an impure form into a standard, pure form. The distillation protocol must not itself consume or generate the resource under question—it must just convert the resource to the desired form. Under the formalism of quantum processes, this condition is not unambiguous. Depending on how it is interpreted, it begets formal conditions of varying strength. In the case of coherence, the least-constrained distillation protocols are *coherence–non-creating channels* [2, 3], established in the literature as the *maximal incoherent operations* (MIO). When acting on incoherent input states, MIO produce outputs that are also incoherent.

But this condition is so weak that it may yet admit operationally-unreasonable possibilities, such as a process that amplifies certain aspects of coherence in an already-coherent input. An operational subclass of MIO, called the *incoherent operations* (IO) [4], mitigate this incongruity by placing the additional constraint that the process must be implementable through elementary sub-processes (formally, Kraus operators) that are *each* coherence–non-creating. Winter and Yang [5] found that the IO subclass is already powerful enough to distill coherence maximally, i.e. as efficiently as MIO.

Moving further in the operationalist direction, Yadin *et al.* [6] defined a subclass of IO called the *strictly incoherent operations* (SIO). These processes are implementable through sub-processes that individually *neither create nor detect* (i.e., measure) coherence. Lami *et al.* [7] found that this constraint is too strong to admit full distillation; subsequently, Lami [8] exactly quantified how efficiently SIO can distill.

It is evident from these developments that the coherence-measuring capability of the components implementing IO is essential to enabling maximal distillation. But there has been no in-

vestigation yet on exactly *how much* coherence-measuring power is required. This is the question that motivates our work.

We approach the problem by constructing a measurement that we call the *target effect* of a process. A process’ target effect captures the probability with which it succeeds in converting an arbitrary input to a desired standard form of the resource—in other words, the process’ efficacy for the purpose of distillation. The target effect also contains significant information about the structure of coherent measurements required to implement the process. Our foundational lemmas put stringent conditions on the structure of the target effect for the associated process to succeed in various instances of distillation—exact (maximal and non-maximal) and approximate. Based on these lemmas, our main results provide lower bounds on the requisite coherent measurement cost (formalized in the relevant technical section below) of the respective distillation instances. Our results apply not only to IO, but to the most general operational class MIO.

We then consider the so-called *asymptotic limit* of distillation, where the input is a large number of copies of a given resource state. Based on our results on approximate distillation, we conjecture that the necessary and sufficient coherent measurement cost of maximal asymptotic distillation is what we call the input’s *irretrievable coherence*—a quantity related to the irreversibility of resource conversion under IO. We make substantial progress towards proving this conjecture and discuss the technical difficulties that hinder the proof. In the process, we establish a relationship between our problem and an important quantum-statistical result called the generalized quantum Stein’s lemma [9, 10]: the requisite measurement coherence of asymptotically maximal distillation is bounded below by the coherence of hypothesis-testing measurement effects that saturate the error exponent scaling bound set by the lemma. Finally, we discuss the implications of our conjectures: (1) the asymptotic coherent measurement cost may outweigh the very distilled yield in a significant fraction of cases—thus rendering distillation wasteful; (2) possibilities of a tradeoff between the coherent measurement budget and the distilled yield.

In the rest of this section, we will provide a brief background and motivation for our research

problem, followed by a summary of our main results. We strive to keep the discussion high-level; but it will, of necessity, get gradually more technical by the end of the section. We will often use some technical terms and notation without interrupting the flow of presentation with their formal definitions, which we defer until the next section.

1.1 Coherence relative to bases

Quantum coherence refers to the presence of superposition effects in quantum mechanics. But of course, this begs the question “a superposition of what sort of entities?” Indeed, coherence can be given different formal definitions based on what we consider the “unsuperposed” objects; examples include eigenstates of conserved quantities [11], orthogonal subspaces induced by measurements [3, 12], and even mutually-nonorthogonal elements, such as the classical states of bosonic modes [13]. For our purposes it suffices to consider the unsuperposed objects to be mutually-orthogonal subspaces of the Hilbert space of quantum states, which we shall for brevity call elementary subspaces.

Within this notion, there is a broad distinction between two subtypes, according to what happens when two systems A and B are brought together:

1. The elementary subspaces $\mathcal{V}^{AB}$ of the composite AB are all and only the pairwise tensor products of those of A and B: $\mathcal{V}_{j,k}^{AB} = \mathcal{V}_j^A \otimes \mathcal{V}_k^B$. Under this composition rule, two distinct elementary subspaces $\mathcal{V}_{j_1}^A, \mathcal{V}_{j_2}^A$ of the same subsystem never contribute to the same elementary subspace of the composite.
2. The elementary subspaces of AB depend on those of A and B in a manner different from the simple tensor-product composition. An example of this is when the elementary subspaces are eigenspaces of operators (e.g., local Hamiltonians): AB may have degenerate eigenspaces that combine distinct $\mathcal{V}_{j_1}^A, \mathcal{V}_{j_2}^A$ (in the Hamiltonian example, when the energies of these eigenspaces are related to those of a pair of B eigenspaces through $E_{j_1}^A + E_{k_1}^B = E_{j_2}^A + E_{k_2}^B$).

We will restrict our attention to the first case, where the composition of two or more systems

is governed by a simple tensor product. Furthermore, we will only consider the case of one-dimensional elementary subspaces; it is evident that this dimensionality is preserved under tensor-product composition. In other words, the coherence of interest to us is relative to a certain fixed *orthogonal basis* on each system, and the corresponding tensor-product bases of compositions thereof. This special case has been studied extensively under the resource-theoretic paradigm, which we will review in a later section.

1.2 Operational aspects of coherence

The *static* form of coherence, present in quantum states, is usually most familiar and easily understood: for instance, a quantum bit (qubit) in one of the states $|\pm\rangle := (|0\rangle \pm |1\rangle)/\sqrt{2}$ contains coherence relative to the so-called computational basis $\{|0\rangle, |1\rangle\}$. But there is also a *dynamical* aspect to coherence, manifested in the action of processes that can transform an incoherent state to a coherent one. In addition, a *mensural* aspect of coherence is embodied by the ability of a measurement device to detect coherent superpositions. An elementary prototype of dynamical coherence is the unitary process effected by the qubit Hadamard gate, given by its action $H|0\rangle = |+\rangle$ and $H|1\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$ on the computational basis states. In turn, the resulting so-called Hadamard basis states $|\pm\rangle$ can be considered elementary prototypes of static coherence, and a measurement in this basis one of measurement coherence.

Under the tensor-product composition model within which we are working, these three aspects of coherence are not equal in their operational power: indeed, dynamical coherence is strictly more powerful than the other two, as it can be used to elevate both incoherent states and incoherent measurements to their respective coherent counterparts. Meanwhile, static coherence cannot be used to simulate either dynamical or measurement coherence, and measurement coherence is likewise restricted. Note that this restriction may not present itself under other composition rules—notably, when the elementary subspaces are eigenspaces of conserved quantities, the composition of two or more systems can admit static coherence to be turned dynamical [14].

Some important clarifications are in order regarding our use of the term “coherent measure-

ment”. Firstly, by “measurement”, we do not mean necessarily an action *intended* as a measurement, or one accompanied by a readout or wavefunction collapse. Rather, we are referring to a property on the mathematical level of description, which on the operational level corresponds to elementary features of the interaction between the system and the apparatus carrying out a process thereon; these elementary interactions necessarily cause certain pieces of information about the system’s state to impinge on the apparatus’—*effectively* a measurement of the system by the apparatus. Furthermore, when we say a process requires coherent measurement over k levels of the system, we mean that at least some of the information inevitably imprinted on the apparatus can distinguish k -fold superposed states of the system from other states. In other words, regardless of whether an agent sets up the apparatus for the express purpose of such measurement, carrying out the process entails the apparatus *effectively* measuring (in the above sense) relative to k -fold superpositions over the system’s computational states. This is a distinctly “measurement-flavoured” coherent capability that the apparatus *must* possess for carrying out the process, in contrast to static and dynamical coherence. Consider the Hadamard basis measurement mentioned above (with or without readout): one way of implementing it is to simply project the input onto the Hadamard basis states (corresponding to the Kraus operators $|\pm\rangle\langle\pm|$); in this case, the measuring process has dynamical coherence, since it can (probabilistically) map incoherent inputs to coherent outputs. But another way to implement the same measurement is to use incoherent “flags” for the two outcomes (e.g., via the Kraus operators $|0\rangle\langle+|$ and $|1\rangle\langle-|$), in which case the process lacks dynamical coherence. Measurement coherence is that operational capability which the process possesses in both cases alike.

Just as the static coherence consumed by a process is considered a cost, so should the requisite measurement coherence for a task (such as coherence distillation), even though we do not usually think of measurement as a “consumable”. A spiritually kindred concept is that of *query complexity*: the number of times a certain operational element needs to be activated in an algorithm. A further justification for considering it a cost is

that both static and measurement coherence can be derived from dynamical coherence—indeed, they always are in practice, though their operational distinction from dynamical coherence is an interesting feature of the quantum formalism (this point will be discussed at length in Section 6).

Before moving on to the next section, we note that our notion of measurement coherence is closely related to that formalized by Kim and Lee [15, 16]. But their notion is defined on POVMs—a formalization that does not incorporate the dynamical details of what happens to the system post-measurement. Also, POVMs are used in contexts with operational *intent* to measure (in contrast with ours, where operational *capability* takes precedence over intent). There are also differences in the finer technical details of Kim and Lee’s treatment besides these. Another notion in the literature distinct from ours is *coherence relative to measurements*, also called POVM coherence [3, 12]: while we are concerned with the coherence (relative to the fixed incoherent basis) *within* the effective measurement action in the Kraus operators, this notion of POVM coherence focuses on coherence *relative to* the block structure induced by the POVM effects themselves. This is, in a sense, complementary to our notion and that of Kim and Lee.

1.3 Resource theories of coherence

Various formalizations of the concept of coherence have been explored under the broad umbrella of *resource theories*; for a general exposition on coherence resource theories, we direct the reader to [1]. Here we will provide a brief introduction to certain specific resource theories of coherence, adequate for our purposes.

A resource theory formalizes the study of a quantum resource by identifying the operational capabilities required to create or proliferate it. Such capabilities are axiomatically forbidden, leaving only certain constrained actions that can be performed, called the “free operations”. The theory then endeavours to chart out what can and cannot be done using only the free operations—spiritually akin to determining, say, the plane figures that can be constructed using only a compass and a straightedge¹. Typically, the

¹ We are indebted to Gilad Gour for this evocative anal-

free operations in quantum resource theories are a class of quantum (sub)channels, including the preparation of so-called “free states”². All states that are not free are called resource states.

In the resource theories we will consider, the resource is coherence relative to a fixed orthogonal basis of the Hilbert space of a given quantum system. This basis is variously termed computational, canonical, classical, etc.; we will simply call it the *incoherent basis*. Furthermore, we will work in the paradigm where the incoherent basis of a composite system is just the tensor product of its subsystems’ incoherent bases. This notion of coherence falls under what has been called *speakeable coherence* in the literature [11]. It is operationally relevant for, e.g., gate-based quantum computing, where every elementary system has a computational basis and where tensor products of computational-basis states are easy to prepare.

The free states in these coherence resource theories are the incoherent states, i.e. the states whose density matrices are diagonal in the incoherent basis. We will refer to all other states—the resource states—as *coherent*. Most of the extensively studied resource theories are primarily concerned with static resources, and this is true also of coherence resource theories. Nevertheless, dynamical and mensural considerations do play a part in axiomatizing the class of free operations.

Adhering to the basic tenets of the resource-theoretic paradigm, the free operations in all of these resource theories are constrained to be incapable of creating coherence. But as it turns out, there are diverse ways to choose families of free operations obeying this constraint, spawning a veritable zoo of distinct coherence resource theories. Amongst these, we will focus on the resource theory whose free operations are the so-called *incoherent operations*³ (IO) [4]. Informally, an IO is a quantum process that can be broken down into sub-processes that *may detect* coherence (i.e., measure relative to coherent states, in the sense explained in Section 1.2) but *must not*

ogy.

² We use the term “states” to refer not only to pure states (associated with wavefunctions spanning a Hilbert space), but also to mixed states (formally identified with density operators acting on the Hilbert space).

³ We will adhere to the term “incoherent operations” established in the literature, notwithstanding its regrettable inspecificity.

create coherence when acting on incoherent input states. In other words, each of the sub-processes that constitute the process is devoid of dynamical coherence but may contain measurement coherence.

1.4 Distillation of coherence-resource

The main motivation for this work comes from the resource-theoretic concept of *distillation*: the task of converting an arbitrary resource state to a standard form. Resource distillation is often an essential part of applications [17]; for example, coherence distillation is closely related to the task of randomness extraction using incoherent measurements [18]. Beyond this direct value, the study of resource distillation also offers valuable insight into the structure of a resource theory.

In all resource theories of coherence, the standard form of coherence-resource is a pure state containing a uniform superposition of some number of incoherent basis elements, e.g. $|\Psi_M\rangle := M^{-1/2} \sum_{m \in [M]} |m\rangle$. This choice is justified by several factors. Firstly, these states are usually optimal for applications (e.g., phase estimation) requiring coherence. Secondly, the free operations can produce any required state from a single copy of one of these. A third justification comes from the *asymptotic* or *independent and identically-distributed* (i.i.d.) limit of the resource theory, where the free operations act on large numbers of independent copies of identical states. In this limit, the standard coherent states of different M values admit reversible⁴ “currency exchange” at a rate proportional to $\log_2 M$, which is the equivalent number of standard coherent bits (or *cobits*) $|\Psi_2\rangle$ (more familiar as the Hadamard state $|+\rangle$ discussed above). The cobit thus functions as a convenient unit for quantifying coherence. For the same reason, the Hadamard gate makes a good standard unit for dynamical coherence, and the Hadamard measurement one for measurement coherence.

The asymptotic limit of the IO resource theory affords an added feature: copies of *any* coherent state—pure or mixed—can be converted (albeit *not* reversibly—more on this later) by IO to cobits at a rate that is maximal in a resource-

⁴ Note that this reversibility is to leading order in the number of copies; in this work we will not consider higher-order effects [19].

theoretic sense [20]. In other words, coherence is asymptotically *universally distillable* by IO. But at the heart of this universal distillability lies the central question that motivates our work.

1.5 What powers coherence distillation?

Recall that IO can be implemented using components that do not create coherence, but may nevertheless detect it. *Strictly incoherent operations* (SIO) are the sub-class of IO that use only components that *cannot even detect* coherence [6]. This restriction ends up breaking the asymptotic universal distillability seen under IO. Indeed, SIO exhibit a particularly severe form of non-distillable, or “bound”, coherence: any number—however large—of copies of certain coherent states cannot be converted, even approximately, to even a *single* cobit [7].

In summary, the *unbounded* measurement coherence of IO enables universal distillation, while the strictly coherence-non-detecting SIO are too constrained to distill universally. But what lies between these two extremes? Our paper is an attempt to understand this intervening operational landscape, by answering questions such as:

1. How much measurement coherence (quantified in a way that will be discussed later) is necessary to recover the maximal distillability afforded by IO?
2. What are the corresponding costs for non-maximal and approximate distillation?
3. How does this coherent measurement cost behave in the asymptotic limit?
4. What are the conditions for the lower bound on the cost to be attainable? Are these conditions met in the asymptotic limit?

We approach these questions using a construction that we call the *target effect*: a measurement associated with a given quantum process, containing information about the efficacy of the process at mapping arbitrary inputs to a desired target output, as well as about the coherence-detecting power of the process. Among other things, we show that the coherent measurement cost of distillation is bounded by a particular property of the target effect, which we will now discuss.

1.6 Irretrievable coherence

In a resource theory, a real-valued function of states is called a *resource measure* if it satisfies the following two conditions: (1) It is a non-increasing monotone under the free operations; (2) It is *faithful*, i.e. takes nonzero values on all and only the resource (non-free) states. The answers to our central questions turn out to involve some important measures of coherence.

Given a state ρ , its *relative entropy of coherence* is defined as

$$C_r(\rho) = \min \{S(\rho\|\sigma) : \Delta[\sigma] = \sigma\}, \quad (1)$$

where σ takes values of density operators, $S(\cdot\|\cdot)$ is the Umegaki quantum relative entropy [21], and $\Delta(\cdot)$ denotes the diagonal part (in the incoherent basis representation) of the argument. Thus, the minimization is over all diagonal states σ —in other words, the free states. Conveniently, the minimization evaluates to $C_r(\rho) = S(\rho\|\Delta[\rho]) = S(\Delta[\rho]) - S(\rho)$, where $S(\cdot)$ is the von Neumann entropy. Meanwhile, ρ ’s *coherence of formation* is given by the so-called *convex-roof extension* of the restriction of C_r to pure states:

$$C_f(\rho) = \min_{p_x \geq 0; \sum_x p_x |\phi_x\rangle\langle\phi_x| = \rho} \sum_x p_x C_r(|\phi_x\rangle\langle\phi_x|), \quad (2)$$

where the minimization is over all convex decompositions of ρ into pure states. Notice that $C_r(\psi) = C_f(\psi) = S(\Delta[\psi]) = H(\mathbf{p})$ (where H is the Shannon entropy) for a pure state $\psi \equiv |\psi\rangle\langle\psi|$ ⁵ with incoherent-basis distribution $|\langle i|\psi\rangle|^2 = p_i$. In particular, $C_r(\Psi_M) = C_f(\Psi_M) = \log_2 M$ for the standard resources.

These measures have operational significance in the IO resource theory. Firstly, $C_r(\rho)$ is the regularized asymptotic *distillable coherence* under IO, defined as the maximum asymptotic rate at which cobits can be distilled from copies of ρ by IO. That is, $C_r(\rho)$ is the largest $r \in \mathbb{R}$ such that the transformation $\rho^{\otimes n} \mapsto \Psi_2^{\otimes rn}$ can be achieved by IO to an arbitrarily good approximation as $n \rightarrow \infty$. Likewise, $C_f(\rho)$ is the regularized asymptotic *coherence cost* under IO: the minimum asymptotic rate at which cobits must be *consumed* to *prepare* copies of ρ by IO, in an operational task called *resource dilution* or *formation*—the opposite of distillation. Mathemat-

⁵ We will use this shorthand for rank-1 projectors, where it is possible without ambiguity.

ically, $C_f(\rho)$ is the smallest $r \in \mathbb{R}$ such that the transformation $\Psi_2^{\otimes rn} \mapsto \rho^{\otimes n}$ can be achieved arbitrarily well as $n \rightarrow \infty$.

For almost all states ρ (in a measure-theoretic sense), C_f is strictly larger than C_r [5]. Hence, the coherence distillable by IO from a given input is generically smaller than that required to prepare the same input. Thus, IO presents an instance of *irreversibility* in resource theories, a topic of current interest [22, 23]. IO’s is a particularly strong form of irreversibility, since it persists even in the asymptotic limit and is, moreover, already present in the lowest order (i.e., between the regularized distillation and formation rates). Nevertheless, as we alluded to above, the IO-*distillable* coherence is in fact maximal under general resource-theoretic constraints [20]; therefore, the culprit behind the irreversibility is the inflated *cost* of resource formation under IO. Incidentally (as the reader may have anticipated from our earlier statements), the SIO resource theory is even more irreversible—this additional disparity owing solely to SIO’s inferior *distillable* coherence compared to IO’s, the two theories’ coherence costs being equal!

In our main results, these two coherence measures feature in the form of their difference $\ell(\rho) := C_f(\rho) - C_r(\rho)$. Because of the operational meaning of this quantity vis-à-vis the asymptotic irreversibility of the IO resource theory, we christen it the *irretrievable coherence*. It has, in fact, been encountered (though not named) in the literature in a different operational context: it quantifies the difference between the quantum and the classical values of the so-called *intrinsic randomness* of a state [24, 25]. It is worth noting that, while the irretrievable coherence is determined by two coherence measures and is itself a *signature* of coherence—it is nonzero only for coherent states—it is *not* a coherence measure in the resource-theoretic sense. It fails to be faithful, as can be seen from the case of pure states. But more importantly, it is not a monotone under IO or, indeed, any reasonable class of free operations.

1.7 Clues in the literature

Recall that distillation in the asymptotic limit is the task of converting many copies of a given input to an output close to a standard resource. Formally, for every $n \in \mathbb{Z}_+$, the input is $\varrho_n \equiv$

$\rho^{\otimes n}$ and is to be mapped approximately to $\Psi_2^{\otimes m_n}$ for some m_n . “Asymptotic” refers to the limit $n \rightarrow \infty$, and the asymptotic rate of distillation is the value $r = \lim_{n \rightarrow \infty} (m_n/n)$. As we alluded to earlier, the highest achievable rate for a given ρ is $r = C_r(\rho)$.

Winter and Yang [5] constructed an IO protocol achieving this maximal distillation rate. A high-level examination of the protocol already hints at connections between asymptotic irreversibility and the object of our interest, viz. the coherent measurement cost of distillation. Crucially, the protocol consists (apart from some asymptotically-inconsequential measurements) of just a unitary transformation of the input followed by a partial trace. Considering the purity required of the output (distillate), the effect of the protocol before the final partial trace can be summarized approximately as

$$\varrho_n^A \xrightarrow{\mathcal{U}} \tau^S \otimes \Psi_M^M. \quad (3)$$

Here the superscript A labels the input system, M the output system, and S the part that will be traced out. The question of how much coherent measurement the IO needs translates to how coherently this unitary channel $\mathcal{U}$ must act. Since the unitary does not involve any additional systems, the systems’ dimensionalities (which we denote by italicizing the corresponding labels) satisfy $A = SM$. Let us now make some heuristic estimates for these numbers, appealing to (an extremely crude form of) asymptotic typicality [26]; for brevity, we will omit qualifiers like “approximate” and “typical part” in the following statements, but stress that these qualifications are implicit.

Consider the input $\varrho_n \equiv \rho^{\otimes n}$: its rank (by unitarity, also the rank of τ) is⁶ $S_0 := \exp_2[n S(\rho)]$, due to the asymptotic equipartition property (AEP). Applying AEP on the diagonal part $\Delta(\varrho_n)$, which is in fact $[\Delta(\rho)]^{\otimes n}$, we see that the relevant dimensionality of the input (covering all the incoherent basis labels that occur with nonzero amplitudes) is $A = \exp_2[n S(\Delta[\rho])]$. Finally, the size of the maximal distillate is $M = \exp_2[n C_r(\rho)] = \exp_2[n (S(\Delta(\rho)) - S[\rho])]$. Notice that $M = A/S_0$, and therefore, $S_0 = S$. A basic consequence of this is that coherent measurements of rank S are sufficient for maximal

⁶ Throughout this paper, we will use the notation $\exp_2(\cdot) \equiv 2^{(\cdot)}$.

distillation; this therefore serves as a reference value against which to evaluate nontrivial bounds on the coherent measurement cost, which is our main concern.

Noting (again from equipartition) that the input's spectrum must be flat, we conclude that τ must be maximally mixed. In particular, this means that the subsystem S is discarded in an incoherent state, uncorrelated with the distillate M . Now let us view the protocol in reverse: we take Ψ_M , append to it an auxiliary system S in the *incoherent* state τ , and apply the unitary channel $\mathcal{U}^\dagger$ to map the composite SM to $\varrho_n^{\text{A}\equiv\text{SM}}$. How much coherence must $\mathcal{U}^\dagger$ generate for this? If it had had to act on a fully incoherent input, it would have needed to create all of the coherence in ϱ_n from scratch; considering that it has the Ψ_M to begin with, it still needs to account for the deficit—a hint at the difference between the formation cost and the distillable coherence.

To be sure, how much coherence an operation needs to generate to prepare a required resource is not an unambiguous concept: it depends on the class of operations considered. The Winter–Yang IO protocol's reversal, which we considered, is not itself an IO; nor is it in any of the other classes of incoherent operations defined in the literature. Besides, this maximal distillation protocol is but one possibility; in general, a protocol may use auxiliary systems, instead of acting unitarily on just the input. In this connection, a further difficulty is that there is not yet an operational understanding of IO in terms of their unitary implementations (or *dilations*): IO are only understood in terms of the abstract mathematical objects called Kraus operators. In contrast, SIO (for example) are understood through both their Kraus operators and their dilations.

But another hint in the same direction—again from the near-maximally-mixed fate of S and the near-unitarity of the protocol—is that IO distillation seems to entail fully “resolving” the space on which the input is supported, so as to retain in the distillate not only all of the input's coherence but also all of its *purity*. As such, the requisite coherent measurement of distillation translates to that of fully resolving the input's support. As a simple illustration of what we mean, consider the 2-dimensional subspace of a ququart (4-dimensional system) spanned by the vectors

$$|v_0\rangle \propto |+\rangle + |2\rangle; \quad |v_1\rangle \propto |-\rangle + |3\rangle, \quad (4)$$

where $|\pm\rangle \propto |0\rangle \pm |1\rangle$ as usual. An IO can distill one cobit from this subspace, e.g. using the Kraus operators $K_0 = |0\rangle\langle+| + |1\rangle\langle2|$ and $K_1 = |0\rangle\langle-| + |1\rangle\langle3|$. But to do so it must necessarily act coherently on the $|\pm\rangle$ part.

Let us now try to estimate the coherent measurement rank required to resolve (in this sense) the support of ϱ_n . Asymptotically, the projector onto this support is close to ϱ_n itself, as the latter's spectrum flattens out. Let $\varrho_n = \sum_j q_j \phi_j$ be some convex decomposition into pure components. Now consider the typical letter strings (i.e., incoherent basis indices) that occur in ϱ_n (or, equivalently, along its diagonal). As noted above, these are $A = \exp_2[nS(\Delta[\rho])]$ in number; and by asymptotic typicality, each of them carries a weight of A^{-1} in the overall input. On the other hand, the weight that any of these strings accrues by virtue of its occurrence in any single ϕ_j is $\lesssim (S \exp_2[nC_f(\rho)])^{-1}$: the S factor is the dimensionality of the support, while $\exp_2[nC_f(\rho)]$ lower-bounds the number of (approximately equally-superposed) typical strings in each pure component, as can be seen from the definition (2). Thus, in order to account for all of a string's weight in ϱ_n , it must occur in no less than

$$\frac{S 2^{nC_f(\rho)}}{A} = \exp_2(n[C_f(\rho) - C_r(\rho)]) = 2^{n\ell(\rho)} \quad (5)$$

distinct ϕ_j 's. This bound applies alike to all of the typical strings. Therefore, a measurement that resolves the entire support of ϱ_n can be expected to involve coherence over blocks no smaller than this size.

Though these hints are based on loose intuition and crude estimates, they proved helpful in our project, directing us towards more rigorous investigations and methods. In particular, they inspired us to consider a non-asymptotic idealization of the above “crudely-typicalized” case of maximal distillation, yielding a result (Theorem 1 below) that somewhat validates the hints and informs our conjectures 1 and 2 on the asymptotic case.

1.8 Summary of contributions

While the technical details in the following summary are included for the benefit of the expert reader, the essence of our contributions can be appreciated in light of the foregoing background

discussion. In particular, one may note that our main results provide lower bounds on the coherent measurement cost of various instances of coherence distillation. Also notable is that the irretrievable coherence and its variants feature among these bounds, as our intuitive arguments above had suggested.

Our first contribution is to establish a formal connection between the coherent measurement cost of distillation and a construction that we call the *target effect* (Section 3.2). For any given quantum process (channel) $\mathcal{E}$ and designated target state $|\alpha\rangle$, we define an associated target effect $T_{\mathcal{E}}^{\alpha}$, constructed to capture the probability with which $\mathcal{E}$ maps an arbitrary input ρ to the target α . We then show that the requisite coherent measurement action in any implementation of $\mathcal{E}$ is quantified by the *static* coherence of $T_{\mathcal{E}}^{\alpha}$ considered as a state. We then use this result as the foundation to study the coherent measurement cost in several forms of the distillation task. Although the construction is motivated by our interest in the gap between SIO and IO, its applicability is general enough that all our results apply to the most general class of free operations in coherence resource theories, the *coherence–non-creating channels* (established in the literature as “maximal incoherent operations”, or MIO).

As mentioned above, we first consider a certain single-shot (i.e., finite-sized and non-asymptotic) variant containing idealized versions of the AEP-related features encountered in maximal asymptotic distillation (Section 4.1):

Theorem 1. *Any coherence–non-creating channel that deterministically maps a rank- S state ρ^A to a standard coherent resource Ψ_M with $M = A/S$ must involve coherent measurement over at least*

$$M^{-1}r_C(\tau_{\rho}) \geq M^{-1}\exp_2 C_f(\tau_{\rho}) \geq \exp_2 \ell(\tau_{\rho}) \quad (6)$$

elements of A ’s incoherent basis, where $\tau_{\rho} := \mathbb{1}_{\rho}/S$.

Here, r_C denotes the *coherence rank* of a state, defined for pure states as $r_C(\psi) = \text{rank}\Delta(\psi)$ and for mixed states as

$$r_C(\rho) = \min_{p_x \geq 0; \sum_x p_x \phi_x = \rho} \max_x r_C(\phi_x). \quad (7)$$

We prove Theorem 1 by showing that the target effect associated with any channel achieving

such a transformation must be proportional to τ_{ρ} . When we present our results in detail, we will see that the condition $M = A/S$ is always associated with the distilled resource’ being maximal. In general, if an IO can distill Ψ_M from a rank- S state in A dimensions, then $M \leq A/S$. Our next result (Section 4.2) applies to exact *non-maximal* distillation, again in the single-shot regime.

Theorem 2. *Any coherence–non-creating channel that deterministically maps a rank- S state ρ^A to a standard coherent resource Ψ_M with $M = pA/S$ must involve coherent measurement over at least $M^{-1}r_{C;p}(\tau_{\rho}) \geq M^{-1}\exp_2 C_{f;p}(\tau_{\rho}) \geq \exp_2 \ell_{;p}(\tau_{\rho})$ elements of A ’s incoherent basis, where $\tau_{\rho} := \mathbb{1}_{\rho}/S$ and the bounds are defined as follows:*

$$r_{C;p}(\tau) := \min_{\gamma \in \mathcal{T}_p^A(\tau)} r_C(\gamma); \quad (8)$$

$$C_{f;p}(\tau) := \min_{\gamma \in \mathcal{T}_p^A(\tau)} C_f(\gamma); \quad (9)$$

$$\ell_{;p}(\tau) := C_{f;p}(\tau) - C_r(\tau) \quad (10)$$

with

$$\mathcal{T}_p^A(\tau) := \left\{ \gamma = p\tau + (1-p)\tau_{\perp} : \Delta(\gamma) = \frac{\mathbb{1}^A}{A} \right\}, \quad (11)$$

$\tau_{\perp}$ taking values as density operators on the subspace complementary to τ ’s support.

We arrive at this result by methods similar to those of Theorem 1: showing that the target effect of any viable channel is associated with a state σ containing a fraction p of τ_{ρ} mixed with some state orthogonal thereto.

Moving on, we have an *approximate* version of the maximal case—a lower bound on the coherent measurement cost of mapping a given input to an output that is close enough (i.e., has high enough fidelity) to a near-maximal standard resource (Section 4.3).

Theorem 3. *Any coherence–non-creating channel that deterministically maps a rank- S state ρ^A , satisfying $r_{\min}\mathbb{1}^A \leq \rho^A \leq r_{\max}\mathbb{1}^A$, to an output σ^M such that $F(\sigma, \Psi_M) \geq 1 - \epsilon$ for $M = A/\tilde{S}$ must involve coherent measurement over at least*

$$\frac{\exp_2 \left[C_f(\tau_{\rho}) - \delta \log_2 A - (1 + \delta)h\left(\frac{\delta}{1+\delta}\right) \right]}{M} \quad (12)$$

elements of A 's incoherent basis, where $\tau_\rho := \mathbb{1}_\rho/S$ and $\delta := \sqrt{2 \left(1 - \frac{S_{\rho,\epsilon}}{\sqrt{SS}}\right)}$ with

$$S_{\rho,\epsilon} := \max \left\{ \frac{1-\epsilon}{r_{\max}}, S \left(1 - \frac{\epsilon}{r_{\min}}\right) \right\}. \quad (13)$$

For small ϵ and near-maximal distillate (i.e., $\tilde{S} \approx S$), δ becomes small and we recover Theorem 1 as a limiting case. Although we do not delve into approximate *non*-maximal distillation, our methods can be suitably extended to address this case. Theorem 3 follows by applying the asymptotic continuity of the coherence of formation [5] to approximate versions of Theorem 1's conditions.

Next, in Section 5, we look at distillation in the asymptotic limit, where the input is of the form $\varrho_n \equiv \rho^{\otimes n}$. Based on our result on approximate distillation, and on the fact that ϱ_n for large n is approximately maximally-mixed on a large part of its support (due to a statistical effect called asymptotic typicality—see Appendix C), we make the following conjecture.

Conjecture 1. *Suppose a sequence $\mathcal{E}_n$ of MIO channels is maximally-distilling on copies of an input ρ . That is, $\mathcal{E}_n$ acting on $\varrho_n \equiv \rho^{\otimes n}$ achieves $F[\mathcal{E}_n(\varrho_n), \Psi_{M_n}] \geq 1 - o(1)$ for $\log_2 M_n = n[C_r(\rho) - o(1)]$. Then any Kraus operator decomposition of $\mathcal{E}_n$ involves coherent measurement over at least L_n elements of the input's incoherent basis, where*

$$\log_2 L_n \geq n[\ell(\rho) - o(1)]. \quad (14)$$

This rank bound applies both to the single most coherent measurement element and to the average (of the rank's logarithm) over all involved measurements under the distribution induced by the input.

We make some progress towards proving this conjecture. The proof sketch proceeds very similarly to the approximate single-shot case, with the approximation threshold dictated by n -dependent parameters associated with asymptotic equipartition. Essentially, we show that with increasing n the task gets closer to the idealized maximal instance of Theorem 1—a formalization of the observations we made in Section 1.7. But unfortunately, some subtleties of asymptotic typicality pose obstacles in completing our proof. Nevertheless, in the course of

our attempts, we show that the target effects in asymptotically maximal distillation saturate certain scaling laws enforced by the generalized quantum Stein's lemma [9, 10]—a lemma with far-reaching consequences in quantum statistics and resource theories.

So far, we showed or conjectured the *necessity* of a certain coherent measurement cost for distillation. In the asymptotic limit, we conjecture that the cost scaling in Conjecture 1 is also *sufficient*.

Conjecture 2. *For any ρ , there exists a sequence $\mathcal{E}_n$ of maximally-distilling IO channels with all but an asymptotically-vanishing fraction of measurements individually attaining the bound of Conjecture 1, as well as attaining it on average.*

This is motivated by certain special properties of the IO distillation protocol constructed by Winter and Yang [5]. We also make partial progress towards proving this conjecture, including a general recipe for constructing maximally-distilling IO channels (Appendix D, Observation D.1). But overall, it turns out to be rather more involved than the other direction, requiring putting together several pieces:

1. A construction for a decomposition of ϱ_n that
 - asymptotically approaches the defining bound (2) of the coherence of formation and
 - possesses some symmetries (thanks to ϱ_n 's asymptotic typicality properties), whereby the coherence of each component in the decomposition approaches the overall average value (i.e., ϱ_n 's coherence of formation).
2. A sequence of maximally-distilling (candidate) IO subchannels $\mathcal{F}_n$ based on
 - filtering the above decomposition to further “typicalize” or “flatten” the coherence in each pure component,
 - truncating the remaining components to get rid of parts more coherent than a threshold that asymptotically scales favourably, and

- adapting from Winter and Yang’s IO distillation protocol [5] a certain “pooling” of the classical labels to construct potential maximally-distilling IO channels by connecting their target effects to approximate convex decompositions of ϱ_n .

By virtue of the above truncation, the resulting $\mathcal{F}_n$ use measurement coherence bounded by the claimed scaling.

3. Showing that, despite the above filtering and truncation, the IO subchannel sequence $\mathcal{F}_n$ asymptotically converges towards trace preservation, so that the maximal distillate it produces is asymptotically deterministic.
4. Finally, showing that the (asymptotically negligible but nonzero) trace deficit remaining can be fulfilled by completing each $\mathcal{F}_n$ to a full channel $\mathcal{E}_n = \mathcal{F}_n + \mathcal{G}_n$ using a subchannel $\mathcal{G}_n$ that is also IO.

We encounter hurdles related to those of Conjecture 1, but also other, unrelated ones. Due to the added difficulties, we are inclined to place less confidence in Conjecture 2.

In Section 6 we build on these conjectures to speculate on the requisite coherent measurement cost of *non*-maximal distillation in the asymptotic limit, and on a possible tradeoff between the amount of coherent measurement used and the distilled yield.

Apart from the results summarized above, we also make some progress in understanding cases where the target state is a nonuniform superposition (Section 4.2), a semidefinite programming (SDP)–based approach to estimating the coherent measurement cost (Appendix E), the behaviour of certain SIO monotones under constrained IO (Appendix F), and connections between coherence distillation and certain linear-algebraic structures that we call *decoupling schemes* (Appendix G).

With this summary, we are now ready to present our work in full detail.

2 Technical preliminaries

Throughout this paper, we will use upright Roman symbols (e.g., A) for system labels; these labels will be rendered as superscripts where we

deem them necessary, and omitted altogether where clear from context. For a system A , $\mathcal{H}^A$ will denote its associated Hilbert space and A the dimensionality thereof. We will use a for the classical symbols labelling A ’s incoherent basis vectors $|a\rangle$, and $\mathcal{A}$ for the collection thereof (i.e., A ’s “classical alphabet”). For a generic space $\mathcal{V} \subseteq \mathcal{H}^A$, we will denote the space of its linear automorphisms by $\mathcal{L}(\mathcal{V})$ and the projector onto $\mathcal{V}$ by $\mathbb{1}_{\mathcal{V}}$, but use the abbreviation $\mathbb{1}^A \equiv \mathbb{1}_{\mathcal{H}^A}$ in the case of the entire Hilbert space associated with some system, $\mathbb{1}_{\mathcal{I}}$ for the projector onto $\text{span}\{|a\rangle : a \in \mathcal{I} \subseteq \mathcal{A}\}$, and $\mathbb{1}_{\rho} \equiv \mathbb{1}_{\text{supp}\rho}$ in the case of the support of a positive-semidefinite operator. All vector spaces mentioned will be implicitly assumed to be finite-dimensional. We will use the term “basis” to mean specifically “orthonormal basis”. We will use the established notation $S(\cdot)$ for the von Neumann entropy, $H(\cdot)$ for the Shannon entropy, and $S(\cdot\|\cdot)$ for the Umegaki quantum relative entropy. Though we will also name some variables S , this latter use of the symbol will be clear from context, with no scope for confusion. We will abusively express von Neumann (Shannon) entropies with their arguments *density operators* (*distributions*) instead of the systems (random variables) distributed thereby; e.g., $H(\mathbf{p}) \equiv H(X)_{(p_x)_x}$. For a Hermitian operator T , we will use the notation $\|T\|_p$ (with $p \in \mathbb{R}$) for its p –Schatten norm, defined as the ℓ^p norm of the vector of its singular values. We will usually put the arguments of a map / function in round parentheses, e.g. $f(x)$, but sometimes use square brackets in the context of nested parenthesization, e.g. $(x, f[x])$, generally preferring to alternate between square and round in a nested sequence.

In the foregoing discussion, we used the term “(sub)channel” a few times. A channel is a quantum operation, or state transformation, induced locally on a quantum system by a unitary interaction with an auxiliary system with which it is initially uncorrelated. Mathematically, a channel mapping states of some system A to states of M ⁷ can be identified with a completely-positive (CP) trace-preserving (TP) map $\mathcal{E} : \mathcal{L}(\mathcal{H}^A) \rightarrow \mathcal{L}(\mathcal{H}^M)$; we will often identify the input and out-

⁷ We will use M for the output system since we focus on distillation, a context where denoting the output M seems common practice.

put systems with the shorthand $\mathcal{E}^{A \rightarrow M}$, and further abbreviate as $\mathcal{E}^A \equiv \mathcal{E}^{A \rightarrow A}$ when the systems are identical. A subchannel is a CP trace-*nonincreasing* map, and corresponds with the action of a quantum operation conditioned on one of several possible outcomes, each occurring with some (initial state-dependent) probability; a channel is a special case with only one outcome, occurring deterministically.

Any subchannel can be specified operationally through a so-called *dilation*: a (non-unique) specification of an auxiliary system, a unitary interaction, and an auxiliary measurement, that collectively implement it. Alternately, it can be described somewhat more abstractly through a set (also non-unique) of operators called *Kraus operators*. In the remainder, we will assume the reader has a basic background of these concepts; for a detailed introduction, see [26]. For our purposes, we will need the following definitions for classes of quantum operations. We will primarily have IO in consideration, but most of our results will apply to the largest class of free operations, MIO.

Definition 1 (Coherence-non-creating channel). $\mathcal{E}^{A \rightarrow M}$ is a *coherence-non-creating channel* if it maps all incoherent inputs to incoherent outputs; that is, for all $a \in \mathcal{A}$,

$$\mathcal{E}^{A \rightarrow M}(|a\rangle\langle a|) = \Delta^M \circ \mathcal{E}^{A \rightarrow M}(|a\rangle\langle a|). \quad (15)$$

Since the class of such operations has been established in the literature as the *maximal (class of) incoherent operations* (MIO), we will also use the abbreviation MIO.

Remark 2.1. The operations in the MIO class have occasionally been called “maximally incoherent operations”—a maximally irresponsible misnomer, considering that MIO are arguably *minimally* incoherent! To avoid such awkwardness, we take the liberty to use the descriptor “coherence-non-creating”. Note also that in this paper we use the abbreviation MIO to refer exclusively to *channels* (i.e., trace-preserving operations).

Definition 2 (Incoherent operation). A (sub) channel $\mathcal{E}^{A \rightarrow M}$ is an *incoherent operation* (IO) if it admits a Kraus operator decomposition $\mathcal{E}(\cdot) = \sum_{c \in \mathcal{C}} K_c(\cdot) K_c^\dagger$ such that $\forall c \in \mathcal{C}$ and $a \in \mathcal{A}$,

$$K_c |a\rangle \propto |m \equiv g_c(a)\rangle, \quad (16)$$

where $g_c : \mathcal{A} \rightarrow \mathcal{M}$ is a function.

Remark 2.2. For a generic channel $\mathcal{E}^{A \rightarrow M}(\cdot) \equiv \sum_c K_c(\cdot) K_c^\dagger$, each of its Kraus operators can be expressed (in the incoherent basis) in terms of its rows $|w_{c,m}\rangle^A$ as

$$K_c = \sum_{m \in \mathcal{M}} |m\rangle^M \langle w_{c,m}|^A. \quad (17)$$

In particular, IO Kraus operators have rows of the form

$$|w_{c,m}\rangle = \sum_{a \in g_c^{-1}(m)} \xi_{ca} |a\rangle. \quad (18)$$

Notice that, by virtue of the IO condition (16), the $|w_{c,m}\rangle$ for any fixed c and distinct m values involve disjoint subsets of the incoherent basis:

$$\Delta(|w_{c,m}\rangle\langle w_{c,m}|) \perp \Delta(|w_{c,m'}\rangle\langle w_{c,m'}|). \quad (19)$$

This also implies, of course, that $\langle w_{c,m} | w_{c,m'} \rangle \propto \delta_{mm'}$.

Definition 3 (Strictly incoherent operation). An IO $\mathcal{E}^{A \rightarrow M}$ is a *strictly incoherent operation* (SIO) if it admits a Kraus operator decomposition $\mathcal{E}(\cdot) = \sum_{c \in \mathcal{C}} K_c(\cdot) K_c^\dagger$ that, in addition to satisfying (16), has every g_c invertible on its image. In other words, $\forall c \in \mathcal{C}$ and $m \in g_c(\mathcal{A})$,

$$K_c^\dagger |m\rangle \propto |a \equiv g_c^{-1}(m)\rangle, \quad (20)$$

where $g_c^{-1}(m)$ is unique and well-defined.

Remark 2.3. Recall that we used $\Delta(\cdot)$ earlier for the diagonal part of the argument in the incoherent basis. In fact, this mapping is a channel—the dephasing channel—realizable by SIO: $\Delta(\cdot) = \sum_a |a\rangle\langle a| (\cdot) |a\rangle\langle a|$ (where a runs over the incoherent labels). Under our assumption that the incoherent basis of a composite system is the tensor product of the constituent systems’ incoherent bases, the dephasing channel inherits this convenient multiplicativity: $\Delta^{AB} = \Delta^A \otimes \Delta^B$.

Finally, the formal definition of the quantity that figures in our main results:

Definition 4 (Irretrievable coherence). We define the *irretrievable coherence* of a state ρ as

$$\ell(\rho) := C_f(\rho) - C_r(\rho), \quad (21)$$

where $C_r(\rho) := \min_{\sigma: \Delta[\sigma]=\rho} S(\rho\|\sigma) = S[\rho\|\Delta(\rho)] = S[\Delta(\rho)] - S(\rho)$ is the *relative entropy of coherence* and

$$C_f(\rho) = \min_{p_x \geq 0; \sum_x p_x \phi_x = \rho} \sum_x p_x C_r(\phi_x) \quad (22)$$

its convex-roof extension, the *coherence of formation*.

In addition to these coherence measures, we will also use the *coherence rank* r_C , defined for pure states as $r_C(\psi) = \text{rank} \Delta(\psi)$ and for mixed states as

$$r_C(\rho) = \min_{p_x \geq 0; \sum_x p_x \phi_x = \rho} \max_x r_C(\phi_x). \quad (23)$$

Note that $\log_2 r_C(\psi) \geq C_r(\psi)$ and therefore $\log_2 r_C(\rho) \geq C_f(\rho)$ in general.

3 The coherent measurement cost and target effects

As we develop various technical tools and results below, we shall bear in mind our ultimate aim—namely, to estimate the minimum measurement coherence required for performing certain tasks. To this end, let us first formalize this concept.

3.1 Operationalizing the cost

For a given channel in the form (17) in terms of a Kraus operator decomposition, consider the vectors $|w_{c,m}\rangle^A$. These capture the coherently-measured input components mapped to different incoherent output labels m (cf. the qutrit example in Section 1.7). Via a suitable Stinespring dilation, they can be operationally interpreted as some part of the requisite coherent measurement when the channel is implemented through (1) a scheme for preparing incoherent basis states $|m\rangle \otimes |c\rangle$ *coherently conditioned* on a destructive measurement in some orthonormal basis extending $\{|w_{c,m}\rangle\}_{c,m}$ (i.e., some $\{|\tilde{w}_{c,m}^A\rangle = |w_{c,m}\rangle \oplus |v_{c,m}\rangle\}_{c,m}$ with a suitable extension $\tilde{A}$ of A), followed by (2) an ancillary measurement of the index c , effectively decohering the output system relative to this index. Our deliberate choice to expand the dilation isometry in the $|m\rangle \otimes |c\rangle$ basis on its output side

is in keeping with the resource-theoretic principle of disallowing the preparation of coherent states; of course, the isometry is ultimately just some coherent dynamical process, and its interpretation as “coherently-conditioned incoherent-state preparation” is meant only to capture the resource-theoretic spirit in operational terms.

Moving on, note that we can use different measures of coherence to quantify this coherent measurement resource. To keep our theorems clear, we will use the following:

Definition 5 (Rank-based coherent measurement cost). We will say that a channel’s implementation associated with Kraus operators K_c of the form standardized in Remark 2.2 *involves coherent measurement over at least r elements of the input’s incoherent basis*, where

$$r := \max_{c,m} r_C(\phi_{c,m}). \quad (24)$$

This definition employs the coherence rank as the underlying coherence measure and, furthermore, applies to specific operator-sum decompositions of channels (rather than to the channels themselves). However, we note the following: (1) the technical lemmas underlying our theorems will provide *measure-agnostic* constraints on the structure of the $|w_{c,m}\rangle^A$; (2) moreover, the constraints will hold for *any* channel accomplishing the relevant distillation task and *any* Kraus operator decomposition thereof. For these reasons, we can meaningfully interpret the constraints as determining the coherent measurement cost of the given task, despite their being derived from Kraus operators instead of from operational descriptions such as Stinespring dilations.

While additional coherent measurement action may be required on auxiliary systems, our method accounts for all such action *in the space of the input* A . On this note, we also assume that A is just big enough to contain the incoherent basis elements occurring with nonzero amplitudes in the input under consideration. This brings about no loss in generality, since any channel action outside this subspace is irrelevant for matters concerning the coherent measurement cost of distilling from the given input (except in possibly inflating the cost superfluously).

3.2 The target effect construction

The following construction will prove useful in characterizing the coherence of the vectors $|w_{c,m}\rangle$ and, thereby, the coherent measurement cost of the channel they implement.

Definition 6 (Target effect). Given a channel $\mathcal{E}^{A \rightarrow M}$ and a “target” state α^M , define the *target effect*

$$T_{\mathcal{E}}^{\alpha} := \mathcal{E}^{\dagger}(\alpha), \quad (25)$$

where $\mathcal{E}^{\dagger}$ is the *Hilbert–Schmidt adjoint map* of $\mathcal{E}$. Consequently, for any operator $X \in \mathcal{L}(\mathcal{H}^A)$,

$$\text{Tr}(XT_{\mathcal{E}}^{\alpha}) = \text{Tr}[\alpha \mathcal{E}(X)]. \quad (26)$$

In particular, by choosing X to be a density operator, we see that $T_{\mathcal{E}}^{\alpha}$ is an effect⁸ whose expectation value under any given state quantifies the probability with which $\mathcal{E}$ maps the state to the target—hence its name. As an effect, it satisfies $0 \leq T_{\mathcal{E}}^{\alpha} \leq \mathbb{1}^A$.

Remark 3.1. Hereafter, we will consider only *pure* target states, which is the relevant case for distillation.

Observation 3.2. Given a channel $\mathcal{E}^{A \rightarrow M}$, a decomposition $\mathcal{E}(\cdot) = \sum_c K_c(\cdot) K_c^{\dagger}$ with Kraus operators $K_c = \sum_m |m\rangle \langle w_{c,m}|$, and a pure target $|\alpha\rangle = \sum_m \alpha_m |m\rangle$, define

$$|w_c^{\alpha}\rangle := \sum_m \alpha_m |w_{c,m}\rangle \quad (27)$$

for each c . Then, the associated target effect satisfies

$$T_{\mathcal{E}}^{\alpha} = \sum_c |w_c^{\alpha}\rangle \langle w_c^{\alpha}|, \quad (28)$$

irrespective of the Kraus operator decomposition chosen.

Proof. Using the definition of K_c in terms of the $|w_{c,m}\rangle$,

$$\begin{aligned} T_{\mathcal{E}}^{\alpha} &= \mathcal{E}^{\dagger}(\alpha) = \sum_c K_c^{\dagger} |\alpha\rangle \langle \alpha| K_c \\ &= \sum_{c,m,m'} |w_{c,m}\rangle \langle m| \alpha_m |m\rangle \langle m'| \alpha_{m'}^* |m'\rangle \langle w_{c,m'}| \\ &= \sum_{c,m,m'} \alpha_m |w_{c,m}\rangle \langle w_{c,m'}| \alpha_{m'}^* = \sum_c |w_c^{\alpha}\rangle \langle w_c^{\alpha}|. \end{aligned} \quad (29)$$

□

⁸ A positive-semidefinite Hermitian operator that induces probabilities on density operators is called a *positive operator-valued measure (POVM) effect*, or simply an effect.

In the IO case, as noted in Remark 2.2, the various $|w_{c,m}\rangle$ for a given c involve disjoint incoherent subbases. As such, their coherence is fully reflected in that of the $|w_c^{\alpha}\rangle$, enabling us to use the latter to put lower bounds on various measures of the $|w_{c,m}\rangle$ ’s coherence. While this is no longer true for more general channels, lower bounds on coherence measures of the $|w_c^{\alpha}\rangle$ place *some* constraints on the collective structure of the $|w_{c,m}\rangle$. In particular, the *coherence rank* of $|w_c^{\alpha}\rangle$ cannot be larger than the sum of those of the $|w_{c,m}\rangle$ ’s; this will allow us to put a lower bound on the average row coherence rank even under the more general MIO class of operations.

The $|w_c^{\alpha}\rangle$, in turn, constitute a convex decomposition of $T_{\mathcal{E}}^{\alpha}$, whereby coherence quantifiers based on minimizing over all convex decompositions (e.g., C_f) applied on the latter yield lower bounds on the values that corresponding pure-state coherence quantifiers (e.g., C_r) take on the $|w_c^{\alpha}\rangle$.

Observation 3.3. For an MIO $\mathcal{E}^{A \rightarrow M}$ and target $|\alpha\rangle$, the associated target effect $T_{\mathcal{E}}^{\alpha}$ satisfies

$$\begin{aligned} \langle a| T_{\mathcal{E}}^{\alpha} |a\rangle &= \text{Tr}[\mathcal{E}(|a\rangle \langle a|) \Delta(\alpha)] \\ &= \langle a| \mathcal{E}^{\dagger} \circ \Delta(\alpha) |a\rangle \end{aligned} \quad (30)$$

$\forall a \in \mathcal{A}$.

Proof. Since $\mathcal{E}$ is an MIO, $\mathcal{E}(|a\rangle \langle a|) = \Delta \circ \mathcal{E}(|a\rangle \langle a|)$ for any incoherent basis element $|a\rangle^A$. Therefore,

$$\begin{aligned} \langle a| T_{\mathcal{E}}^{\alpha} |a\rangle &= \text{Tr}[|a\rangle \langle a| \mathcal{E}^{\dagger}(\alpha)] \\ &= \text{Tr}[\mathcal{E}(|a\rangle \langle a|) \alpha] \\ &= \text{Tr}[\Delta \circ \mathcal{E}(|a\rangle \langle a|) \alpha] \\ &= \text{Tr}[\mathcal{E}(|a\rangle \langle a|) \Delta(\alpha)] \\ &= \text{Tr}[|a\rangle \langle a| \mathcal{E}^{\dagger} \circ \Delta(\alpha)]. \end{aligned} \quad (31)$$

□

Remark 3.4. By summing over $a \in \mathcal{A}$, we get $\text{Tr} T_{\mathcal{E}}^{\alpha} = \text{Tr} \mathcal{E}^{\dagger}[\Delta(\alpha)]$. But note that this trace condition holds for any channel $\mathcal{E}$ that maps the identity to an incoherent output: $\mathcal{E}(\mathbb{1}) = \Delta \circ \mathcal{E}(\mathbb{1})$. Our forthcoming results can all be adapted to apply to all channels with this property (e.g., all unital channels) by loosening every application of Observation 3.3 to one of this trace condition.

Remark 3.5. In all our subsequent use of this construction, the target $|\alpha\rangle$ will be clear from the context, and therefore we will use the simplified notation $|w_c\rangle$ and $T_{\mathcal{E}}$.

4 Single-shot distillation

We are now ready to apply the tool of target effects to the study of distillation. We first consider the task of distilling in the single-shot regime, i.e. from finite-sized inputs away from the asymptotic limit. Let us start with exact distillation: an MIO $\mathcal{E}$, acting on an input ρ^A , is required to map it exactly to a desired target $\alpha^M = |\alpha\rangle\langle\alpha|$. Since the desired output is pure and the transformation is required to be exact, $\mathcal{E}$ must in fact uniformly map the entire space $\mathcal{L}(\mathcal{V})$, with $\mathcal{V} \equiv \text{supp}\rho \subseteq \mathcal{H}^A$, to α (up to a scalar factor)⁹; as such, the only relevant property of ρ is the structure of its support $\mathcal{V}$.

Lemma 4.1. *If a channel $\mathcal{E}^{A \rightarrow M}$ maps a subspace $\mathcal{V} \subset \mathcal{H}^A$ exactly to a target α , then the target effect $T_{\mathcal{E}}$ has the structure*

$$T_{\mathcal{E}} = \mathbb{1}_{\mathcal{V}} + T_{\perp}, \quad (32)$$

where $T_{\perp}$ is supported entirely on $\mathcal{V}_{\perp} \subset \mathcal{H}^A$, the subspace complementary to $\mathcal{V}$.

Proof. Since $\mathcal{E}$ is required to map all of $\mathcal{V}$ to the pure state α , Definition 6 entails $\langle v|T_{\mathcal{E}}|v\rangle = \langle v|v\rangle$ for all $|v\rangle \in \mathcal{V}$, whereby $\mathbb{1}_{\mathcal{V}}T_{\mathcal{E}}\mathbb{1}_{\mathcal{V}} = \mathbb{1}_{\mathcal{V}}$. Furthermore, $0 \leq T_{\mathcal{E}} \leq \mathbb{1}^A$ (also noted in Definition 6) then implies $\mathbb{1}_{\mathcal{V}}T_{\mathcal{E}}\mathbb{1}_{\mathcal{V}_{\perp}} = 0$. $\square$

All our main results will follow by applying the above simple result (or its variants) to progressively more complex cases of distillation.

4.1 Maximal distillation

Let us start with the “idealized maximal distillation” case motivated in Section 1.7.

Lemma 4.2. *If an MIO channel $\mathcal{E}^{A \rightarrow M}$ maps an S -dimensional subspace $\mathcal{V} \subset \mathcal{H}^A$ exactly to the M -fold standard coherent state Ψ_M with $S = A/M$ (the divisibility being assumed), then*

$$T_{\mathcal{E}} = \mathbb{1}_{\mathcal{V}}. \quad (33)$$

⁹ We will hereafter abbreviate this condition, abusively, as $\mathcal{E}$ “mapping $\mathcal{V}$ to α ”.

Proof. By virtue of Lemma 4.1, it remains only to show that $T_{\perp} = 0$. Applying Observation 3.3 to $|\alpha\rangle \equiv |\Psi_M\rangle$,

$$\text{Tr } T_{\mathcal{E}} = \text{Tr } \mathcal{E}^{\dagger} [\Delta(\Psi_M)] = \text{Tr } \mathcal{E}^{\dagger} \left(\frac{\mathbb{1}^M}{M} \right). \quad (34)$$

Since A contains no $\mathcal{V}$ -extraneous incoherent labels (see Remark 2.2), and since all of $\mathcal{V}$ must be mapped to Ψ_M , the image of $\mathcal{E}$ is supported on the M -dimensional space spanned by the incoherent components of $|\Psi_M\rangle$. The TP property of $\mathcal{E}$ implies that $\mathcal{E}^{\dagger}$ is unital on this image: $\mathcal{E}^{\dagger}(\mathbb{1}^M) = \mathbb{1}^A$. Therefore,

$$\text{Tr } T_{\mathcal{E}} = \text{Tr } \frac{\mathbb{1}^A}{M} = \frac{A}{M} = S. \quad (35)$$

Thus, $\text{Tr } T_{\perp} = 0$. Since $T_{\mathcal{E}} \geq 0$, this implies $T_{\perp} = 0$. $\square$

By applying coherence measures on Lemma 4.2, we have the following precursor to our first main result.

Proposition 4.3. *Let $\mathcal{E}^{A \rightarrow M}$, with Kraus operators $K_c := \sum_{m \in \mathcal{M}} |m\rangle^M \langle w_{c,m}|^A$, be an MIO channel that distills Ψ_M from a subspace $\mathcal{V} \subseteq \mathcal{H}^A$ of dimensionality $S = A/M$. Denoting $t_{c,m} := \langle w_{c,m}|w_{c,m}\rangle$ and $t_c := M^{-1} \sum_m t_{c,m}$, define the normalized states $|\phi_{c,m}\rangle := t_{c,m}^{-1/2} |w_{c,m}\rangle$, $|\phi_c\rangle := (Mt_c)^{-1/2} \sum_m |w_{c,m}\rangle = \sum_m \sqrt{\frac{t_{c,m}}{Mt_c}} |\phi_{c,m}\rangle$, and $\tau_{\mathcal{V}} := \mathbb{1}_{\mathcal{V}}/S$; note that $(t_{c,m}/[Mt_c])_m$ and $(t_c/S)_c$ are normalized distributions. Then, the following must hold:*

1. $\sum_c \frac{t_c}{S} C_r(\phi_c) \geq C_f(\tau_{\mathcal{V}})$;
2. $\max_c r_C(\phi_c) \geq r_C(\tau_{\mathcal{V}})$;
3. $\sum_m r_C(\phi_{c,m}) \geq r_C(\phi_c)$ for all c ;
4. $\log_2 M \leq C_r(\tau_{\mathcal{V}})$.

Consequently,

$$\begin{aligned} \max_{c,m} r_C(\phi_{c,m}) &\geq \frac{\max_c \sum_m r_C(\phi_{c,m})}{M} \geq \frac{\max_c r_C(\phi_c)}{M} \\ &\geq \left\{ \left\{ \frac{\sum_c \frac{t_c}{S} \log_2 r_C(\phi_c)}{2^{\frac{r_C(\tau_V)}{M}}} \right\} \geq \frac{\sum_c \frac{t_c}{S} C_r(\phi_c)}{M} \right\} \geq \frac{2^{C_f(\tau_V)}}{M} \geq 2^{\ell(\tau_V)}. \end{aligned} \quad (36)$$

Furthermore, if some K_c is also IO, then

$$\log_2 M + \sum_{m \in \mathcal{M}} \frac{t_{c,m}}{M t_c} C_r(\phi_{c,m}) \geq C_r(\phi_c), \quad (37)$$

which adds more detail to the inequality chains in (36).

Proof. Applying the appropriate normalization factors to (33), $\tau_V = \sum_c (t_c/S) \phi_c$, whence points 1 and 2 follow. Point 3 follows from the sub-additivity of the coherence rank under vector addition. For point 4, since $\mathcal{E}(\tau_\rho) = \Psi_M$ and the relative entropy of coherence is a monotone under MIO, $C_r(\tau_\rho) \geq C_r(\Psi_M) = \log_2 M$.

Now suppose the K_c are IO Kraus operators, so that for any given c , the distinct $|\phi_{c,m}\rangle$ don't overlap in their classical symbols. Then, $\Delta(\phi_c) \cong \bigoplus_m \frac{t_{c,m}}{M t_c} \Delta(\phi_{c,m})$ and therefore

$$\begin{aligned} C_r(\phi_c) &= S[\Delta(\phi_c)] \\ &= H\left[\left(\frac{t_{c,m}}{M t_c}\right)_m\right] + \sum_{m \in \mathcal{M}} \frac{t_{c,m}}{M t_c} S[\Delta(\phi_{c,m})] \\ &\leq \log_2 M + \sum_{m \in \mathcal{M}} \frac{t_{c,m}}{M t_c} C_r(\phi_{c,m}). \end{aligned} \quad (38)$$

□

Remark 4.4. In fact, in the IO case, $t_{c,m} = t_c$ for all (c, m) and, furthermore,

$$C_r(\phi_c) = \log_2 M + \sum_{m \in \mathcal{M}} M^{-1} C_r(\phi_{c,m}). \quad (39)$$

To see this, note that since $T_{\mathcal{E}} - |w_c\rangle\langle w_c| \geq 0$, $|w_c\rangle \in \mathcal{V} \forall c$. Exploiting the action of the K_c 's on

$\mathcal{V}$,

$$\begin{aligned} K_c |w_c\rangle &\propto |\Psi_M\rangle \\ \Rightarrow \left(\sum_{m_1} |m_1\rangle \langle w_{c,m_1}| \right) \left(\sum_{m_2} |w_{c,m_2}\rangle \right) &\propto \sum_m |m\rangle \\ \Rightarrow \sum_m |m\rangle \langle w_{c,m}| w_{c,m}\rangle &\propto \sum_m |m\rangle, \end{aligned} \quad (40)$$

the latter owing again to $\langle w_{c,m_1}| w_{c,m_2}\rangle \propto \delta_{m_1 m_2}$. Consequently, $\langle w_{c,m}| w_{c,m}\rangle$ must be independent of m , which by normalization yields $t_{c,m} = t_c$ for all (c, m) . Hence, we can refine the observation preceding (38) to $\Delta(\phi_c) \cong \bigoplus_m M^{-1} \Delta(\phi_{c,m})$, in turn refining (38) to

$$C_r(\phi_c) = H\left[\left(M^{-1}\right)_m\right] + \sum_{m \in \mathcal{M}} \frac{S[\Delta(\phi_{c,m})]}{M}, \quad (41)$$

thence obtaining (39). Nevertheless, the essence of Proposition 4.3 is in bounding various measures of the collective coherence of the $\phi_{c,m}$ in any Kraus operator representation of the channel in question. We will later prove a version of this result for more general cases, where the output is not a maximal Ψ_M . In that context, we will see that a variant of the average coherence inequality (36) still holds, even though the properties discussed in this remark do not.

Recalling Definition 5 for the requisite coherent measurement rank, Proposition 4.3 immediately yields our first main result as a corollary, which we state without proof.

Theorem 1. *Any coherence-non-creating channel that deterministically maps a rank- S state ρ^A to a standard coherent resource Ψ_M with $M = A/S$ must involve coherent measurement over at*

least

$$M^{-1}r_C(\tau_\rho) \geq M^{-1}\exp_2 C_f(\tau_\rho) \geq \exp_2 \ell(\tau_\rho) \quad (42)$$

elements of A 's incoherent basis, where $\tau_\rho := \mathbb{1}_\rho/S$.

Remark 4.5. Proposition 4.3 is stronger than Theorem 1: it puts bounds not only on the maximal coherent measurement rank, but also on various collective or average coherence properties of the $|\phi_{c,m}\rangle$. While the distribution governing this averaging is operationally appropriate only for an input satisfying $\tau_\rho = \rho$ (i.e., one with a flat spectrum) and not otherwise, it is expected to approach the actual distribution of coherent measurement resources in the asymptotic limit (Section 5). As such, our bounds affect not only some outlying measurements occurring with small probabilities, but even the collective statistics of all involved measurements, especially in the asymptotic limit.

It is also notable that we can get variants of Proposition 4.3 by using different coherence quantifiers, although generic quantifiers may not admit a clean splitting of the coherence of each ϕ_c into separate terms for Ψ_M and the $\phi_{c,m}$. The essential structure is captured by Lemma 4.2, whereof each coherence quantifier illuminates certain specific facets.

Nevertheless, we have explicated and highlighted the weaker form in Theorem 1 for two reasons. Firstly, applying Proposition 4.3 to an instance requires a detailed specification of all the Kraus operators, which may be arbitrarily numerous. On the other hand, Theorem 1 only requires specifying the largest coherent measurement rank used in implementing a channel. Secondly, we will see in Section 5 that in the asymptotic limit, the above-noted operational interpretation of the average coherence is expected to be not only necessary but also sufficient for maximal distillation.

We shall now extend the above results to non-maximal distillation tasks.

4.2 Non-maximal distillation

Consider the distillation of a generic $|\alpha\rangle = \sum_m \alpha_m |m\rangle$ (without loss of generality, we can assume $\alpha_m \in \mathbb{R}$ and $\alpha_m > 0$). First, we have a counterpart to Lemma 4.2.

Lemma 4.6. *If $\mathcal{E}^{A \rightarrow M}$ is an MIO that distills α^M from a subspace $\mathcal{V} \subseteq \mathcal{H}^A$, then*

$$\alpha_{\min}^2 \mathbb{1}^A \leq \Delta(T_{\mathcal{E}}) \leq \alpha_{\max}^2 \mathbb{1}^A, \quad (43)$$

where $\alpha_{\min/\max} := \min/\max_m \alpha_m$.

Proof. First, note that $\Delta(\alpha) = \sum_m \alpha_m^2 |m\rangle \langle m|$. Thus, $\alpha_{\min}^2 \mathbb{1}^M \leq \Delta(|\alpha\rangle \langle \alpha|) \leq \alpha_{\max}^2 \mathbb{1}^M$, whence the result follows from Observation 3.3 and the TP property of $\mathcal{E}$. $\square$

This allows us to derive a non-maximal variant of Proposition 4.3.

Proposition 4.7. *Suppose an MIO $\mathcal{E}^{A \rightarrow M}$, with Kraus operators $K_c := \sum_{m \in \mathcal{M}} |m\rangle^M \langle w_{c,m}|^A$, distills the state α from a subspace $\mathcal{V} \subseteq \mathcal{H}^A$. Define $S := \text{Tr } T_{\mathcal{E}}$ for the associated target effect, and $\tau_{\mathcal{V}} := \mathbb{1}_{\mathcal{V}}/\dim \mathcal{V}$. Denoting $t_{c,m} := \langle w_{c,m} | w_{c,m} \rangle$ and $t_c := \sum_m \alpha_m^2 t_{c,m}$, define the normalized states $|\phi_{c,m}\rangle := t_{c,m}^{-1/2} |w_{c,m}\rangle$ and $|\phi_c\rangle := t_c^{-1/2} \sum_m \alpha_m |w_{c,m}\rangle = \sum_m \alpha_m \sqrt{t_{c,m}/t_c} |\phi_{c,m}\rangle$; note that $(\alpha_m^2 t_{c,m}/t_c)_m$ and $(t_c/S)_c$ are now normalized distributions by Lemma 4.6. Define the set $\mathcal{T} := \{T \in \mathcal{L}(\mathcal{H}^A) : \text{Conditions (44)}\}$, wherein the conditions are as follows:*

$$\begin{aligned} T^\dagger &= T; \\ \mathbb{1}_{\mathcal{V}} &\leq T \leq \mathbb{1}^A; \\ (T - \mathbb{1}_{\mathcal{V}}) \mathbb{1}_{\mathcal{V}} &= 0; \\ \alpha_{\min}^2 \mathbb{1}^A &\leq \Delta(T) \leq \alpha_{\max}^2 \mathbb{1}^A. \end{aligned} \quad (44)$$

Through this set, define

$$r_0 := \min_{T \in \mathcal{T}} r_C(T/\text{Tr } T); \quad (45)$$

$$C_0 := \min_{T \in \mathcal{T}} C_f(T/\text{Tr } T). \quad (46)$$

Then,

$$\max_{c,m} r_C(\phi_{c,m}) \geq \frac{\max_c \sum_m r_C(\phi_{c,m})}{M} \geq \frac{\max_c r_C(\phi_c)}{M} \geq \left\{ \left\{ \frac{\sum_c \frac{t_c}{S} \log_2 r_C(\phi_c)}{2^{\frac{1}{c}} \frac{M}{\max_c C_r(\phi_c)}} \right\} \geq \frac{\sum_c \frac{t_c}{S} C_r(\phi_c)}{M} \right\} \geq \frac{2^{C_0}}{M}. \quad (47)$$

Furthermore, if some K_c is also IO, then

$$\log_2 M + \sum_{m \in \mathcal{M}} \frac{\alpha_m^2 t_{c,m}}{t_c} C_r(\phi_{c,m}) \geq C_r(\phi_c). \quad (48)$$

We omit the proof, since it follows exactly like that of Proposition 4.3. We leave further investigations concerning general $|\alpha\rangle$'s for future work. In the remainder of this paper, we will restrict ourselves to standard outputs $|\alpha\rangle = |\Psi_M\rangle$, for which $\alpha_{\min} = \alpha_{\max} = M^{-1/2}$. The following definitions will be useful for these cases.

Definition 7. Given a density operator τ^A and $p \in [0, 1]$, let

$$\mathcal{T}_p^A(\tau) := \left\{ p\tau + (1-p)\sigma : \tau \perp \sigma, \Delta(\gamma) = \frac{\mathbb{1}^A}{A} \right\}, \quad (49)$$

where σ takes values as density operators on A acting on the subspace complementary to τ 's support. Note that, unlike the $\mathcal{T}$ of Proposition 4.7, this set contains only normalized density operators. Using it, define

$$r_{C;p}(\tau) := \min_{\gamma \in \mathcal{T}_p^A(\tau)} r_C(\gamma); \quad (50)$$

$$C_{f;p}(\tau) := \min_{\gamma \in \mathcal{T}_p^A(\tau)} C_f(\gamma); \quad (51)$$

$$\ell_{;p}(\tau) := C_{f;p}(\tau) - C_r(\tau). \quad (52)$$

Corollary 4.7.1. Suppose an MIO $\mathcal{E}^{A \rightarrow M}$, with Kraus operators $K_c := \sum_{m \in \mathcal{M}} |m\rangle^M \langle w_{c,m}|^A$, distills Ψ_M from a subspace $\mathcal{V} \subseteq \mathcal{H}^A$. For the associated target effect $T_{\mathcal{E}}$, let $S := \text{Tr} T_{\mathcal{E}}$ and define $\tau_{\mathcal{V}} := \mathbb{1}_{\mathcal{V}} / \dim \mathcal{V}$; note that Lemma 4.6 implies $S = A/M$, while Lemma 4.1 entails $S \geq \dim \mathcal{V}$. Denoting $t_{c,m} := \langle w_{c,m} | w_{c,m} \rangle$ and $t_c := M^{-1} \sum_m t_{c,m}$, define the normalized states $|\phi_{c,m}\rangle := t_{c,m}^{-1/2} |w_{c,m}\rangle$ and $|\phi_c\rangle := (Mt_c)^{-1/2} \sum_m |w_{c,m}\rangle = \sum_m \sqrt{\frac{t_{c,m}}{Mt_c}} |\phi_{c,m}\rangle$. Then, the inequality chains in (47) hold with $r_0 = r_{C;p}(\tau_{\mathcal{V}})$ and $C_0 = C_{f;p}(\tau_{\mathcal{V}})$, where $p := \frac{\dim \mathcal{V}}{S}$. Furthermore, if some K_c is also IO, then (37) holds for it.

Our next main result follows from this in much the same elementary way as Theorem 1 from Proposition 4.3.

Theorem 2. Any coherence-non-creating channel that deterministically maps a rank- S state ρ^A to a standard coherent resource Ψ_M with $M = pA/S$ must involve coherent measurement over at least $M^{-1} r_{C;p}(\tau_{\rho}) \geq M^{-1} \exp_2 C_{f;p}(\tau_{\rho}) \geq \exp_2 \ell_{;p}(\tau_{\rho})$ elements of A 's incoherent basis, where $\tau_{\rho} := \mathbb{1}_{\rho}/S$.

Remark 4.8. Our results on maximal distillation are in fact corollaries of the non-maximal versions. But we deemed the former to be of sufficient importance, and derivable through sufficiently simpler means, to warrant the order of presentation that we have chosen.

4.3 Approximate distillation

Let us now consider the task of approximate distillation, where we only require an output that is close enough to a standard resource. Formally, given an input ρ^A and an error tolerance $\epsilon \in [0, 1]$, we shall require that the action of a channel $\mathcal{E}$ satisfy

$$F[\mathcal{E}(\rho), \Psi_M] \geq 1 - \epsilon, \quad (53)$$

where

$$F(\sigma, \tau) := \left(\text{Tr} \sqrt{\sqrt{\sigma} \tau \sqrt{\sigma}} \right)^2 \quad (54)$$

is the *Uhlmann–Jozsa fidelity*. Since Ψ_M is pure, the condition (53) simplifies as $F[\mathcal{E}(\rho), \Psi_M] = \langle \Psi_M | \mathcal{E}(\rho) | \Psi_M \rangle = \text{Tr}(\rho T_{\mathcal{E}})$. Thus,

$$\text{Tr}(\rho T_{\mathcal{E}}) \geq 1 - \epsilon. \quad (55)$$

Notably, unlike in exact distillation where only the space $\text{supp} \rho$ mattered, here the detailed structure of ρ must be taken into account. Say $\rho = \sum_s r_s \psi_s$ is an eigendecomposition. Denoting $r_{\max/\min} := \max/\min_s r_s$, we have $\mathbb{1}_{\rho} = \sum_s \psi_s \geq \sum_s (r_s/r_{\max}) \psi_s = \rho/r_{\max}$. Therefore, (55) implies

$$\text{Tr}(\mathbb{1}_{\rho} T_{\mathcal{E}}) \geq \frac{1 - \epsilon}{r_{\max}}. \quad (56)$$

Meanwhile, another consequence of (55) is as follows: since each s term is weighted by an r_s factor, $\langle \psi_s | T_{\mathcal{E}} | \psi_s \rangle$ can afford to be further away from the ideal value 1 for those s whose r_s are smaller. A lower bound on the smallest possible single $\langle \psi_s | T_{\mathcal{E}} | \psi_s \rangle$ is $F_{\min}$, defined through $r_{\min} F_{\min} + (1 - r_{\min}) \cdot 1 = 1 - \epsilon$. This is solved by $F_{\min} = 1 - \epsilon/r_{\min}$, and so

$$\text{Tr}(\mathbb{1}_{\rho} T_{\mathcal{E}}) \geq S \cdot \left(1 - \frac{\epsilon}{r_{\min}}\right), \quad (57)$$

where $S := \text{rank} \rho$ (note the departure from the notation of Section 4.2). Combining (56) and (57), we have

$$\text{Tr}(\mathbb{1}_{\rho} T_{\mathcal{E}}) \geq S_{\rho, \epsilon}, \quad (58)$$

where

$$S_{\rho, \epsilon} := \max \left\{ \frac{1 - \epsilon}{r_{\max}}, S \left(1 - \frac{\epsilon}{r_{\min}}\right) \right\}. \quad (59)$$

When ρ is nearly maximally-mixed on its support, i.e. $r_{\max} \approx r_{\min} \approx 1/S$ (as in the asymptotic case that we will soon take up), the first bound is tighter: $\frac{1 - \epsilon}{r_{\max}} \approx S(1 - \epsilon)$, whereas $S \left(1 - \frac{\epsilon}{r_{\min}}\right) \approx S(1 - S\epsilon)$. The second bound is more useful when ρ is far from maximally-mixed (i.e., $r_{\max} \gg 1/S$) and $\epsilon \ll r_{\min}$: then, $\frac{1 - \epsilon}{r_{\max}} \ll S$, while $S \left(1 - \frac{\epsilon}{r_{\min}}\right) \approx S$.

These constraints, together with those of Observation 3.3, can be used to find lower bounds on the coherent measurement cost, as we did in the exact case. There we had $T_{\mathcal{E}}^{\rho} \equiv \mathbb{1}_{\rho} T_{\mathcal{E}} \mathbb{1}_{\rho} = \mathbb{1}_{\rho}$, leading to the exact block structure (32). In the approximate case, for small enough ϵ we should be able to bound the amplitude of the cross-block parts. We leave this line of inquiry and the pursuit of “good” bounds for future work, here contenting ourselves with a crude bound that suffices for our analysis of maximal asymptotic distillation (Section 5). For this bound, we will show that the normalized density operator $\tau_{\mathcal{E}} := T_{\mathcal{E}}/\tilde{S}$ (where $\tilde{S} := \text{Tr} T_{\mathcal{E}} = A/M$) is “not too different from” $\tau_{\rho} := \mathbb{1}_{\rho}/S$.

Let $\tau_{\mathcal{E}}^{\rho} := \mathbb{1}_{\rho} \tau_{\mathcal{E}} \mathbb{1}_{\rho}$, and define its normalized version $\tau_{\mathcal{E}}^{|\rho} := \tau_{\mathcal{E}}^{\rho} / \text{Tr} \tau_{\mathcal{E}}^{\rho}$. Note that $\text{Tr} \tau_{\mathcal{E}}^{\rho} \geq$

$S_{\rho, \epsilon} / \tilde{S}$. Then,

$$\begin{aligned} F(\tau_{\mathcal{E}}, \tau_{\rho}) &= \left(\text{Tr} \sqrt{\sqrt{\tau_{\rho}} \tau_{\mathcal{E}} \sqrt{\tau_{\rho}}} \right)^2 \\ &= \frac{1}{S} \left(\text{Tr} \sqrt{\mathbb{1}_{\rho} \tau_{\mathcal{E}} \mathbb{1}_{\rho}} \right)^2 \\ &= \frac{1}{S} \left(\text{Tr} \sqrt{\tau_{\mathcal{E}}^{\rho}} \right)^2 \geq \frac{S_{\rho, \epsilon}}{S \tilde{S}} \left\| \sqrt{\tau_{\mathcal{E}}^{|\rho}} \right\|_1^2. \end{aligned} \quad (60)$$

Since $T_{\mathcal{E}} \leq \mathbb{1}$, also $T_{\mathcal{E}}^{\rho} \leq \mathbb{1}$. Thus, $\tau_{\mathcal{E}}^{|\rho} = T_{\mathcal{E}}^{\rho} / \text{Tr} T_{\mathcal{E}}^{\rho} \leq \mathbb{1} / S_{\rho, \epsilon}$. Meanwhile, $\left\| \sqrt{\tau_{\mathcal{E}}^{|\rho}} \right\|_2 = \sqrt{\text{Tr} \tau_{\mathcal{E}}^{\rho}} = 1$. Therefore,

$$\begin{aligned} \left\| \sqrt{\tau_{\mathcal{E}}^{|\rho}} \right\|_1 &\geq \min_{\mathbf{x} \in \mathbb{R}^S} \left\{ \|\mathbf{x}\|_1 : \|\mathbf{x}\|_2 = 1, \|\mathbf{x}\|_{\infty} \leq \frac{1}{\sqrt{S_{\rho, \epsilon}}} \right\} \\ &\geq \sqrt{S_{\rho, \epsilon}}. \end{aligned} \quad (61)$$

Combining this with the bound in (60) and expressing the result in terms of the Bures distance

$$B(\sigma, \tau) := \sqrt{2 \left[1 - \sqrt{F(\sigma, \tau)} \right]},$$

$$B(\tau_{\mathcal{E}}, \tau_{\rho}) \leq \sqrt{2 \left(1 - \frac{S_{\rho, \epsilon}}{\sqrt{S \tilde{S}}} \right)} =: \delta. \quad (62)$$

This δ depends on ϵ , A , M , and ρ ; we suppress its dependencies to avoid clutter. Notice that δ approaches 0 when $SM \approx A$ and $\epsilon \ll 1/S$, which is the case we will encounter in maximal asymptotic distillation. We will skip an approximate analog to Proposition 4.3 and proceed directly to an analog to Theorem 1, which follows by applying the asymptotic continuity of the coherence of formation [5] (Lemma B.2 in Appendix B) on (62) and repeating the arguments of Theorem 1.

Theorem 3. *Any coherence-non-creating channel that deterministically maps a rank- S state ρ^A , satisfying $r_{\min} \mathbb{1}^A \leq \rho^A \leq r_{\max} \mathbb{1}^A$, to an output σ^M such that $F(\sigma, \Psi_M) \geq 1 - \epsilon$ for $M = A/\tilde{S}$ must involve coherent measurement over at least*

$$\frac{\exp_2 \left[C_f(\tau_{\rho}) - \delta \log_2 A - (1 + \delta) h \left(\frac{\delta}{1 + \delta} \right) \right]}{M} \quad (63)$$

elements of A 's incoherent basis, where $\tau_{\rho} := \mathbb{1}_{\rho}/S$ and $\delta := \sqrt{2 \left(1 - \frac{S_{\rho, \epsilon}}{\sqrt{S \tilde{S}}} \right)}$ with

$$S_{\rho, \epsilon} := \max \left\{ \frac{1 - \epsilon}{r_{\max}}, S \left(1 - \frac{\epsilon}{r_{\min}} \right) \right\}. \quad (64)$$

This bound is good for the case where M is near-maximal, i.e. $\tilde{S} \approx S$. We leave a more careful analysis of the non-maximal approximate case for future work. Moving on, we shall derive a slight variant that performs well for nearly-maximally-mixed ρ ; it will simplify our task in the asymptotic case.

Our approach above was to show that $\tau_{\mathcal{E}}$ is close to τ_{ρ} . When ρ is close to maximally-mixed on its support, we can use ρ itself instead of τ_{ρ} . Noting that $\rho \geq r_{\min} \mathbb{1}_{\rho}$, we can modify (60) to

$$\begin{aligned} F(\tau_{\mathcal{E}}, \rho) &\geq r_{\min} \left(\text{Tr} \sqrt{\mathbb{1}_{\rho} \tau_{\mathcal{E}} \mathbb{1}_{\rho}} \right)^2 \\ &\geq \frac{r_{\min} S_{\rho, \epsilon}}{\tilde{S}} \left\| \sqrt{\tau_{\mathcal{E}}} \right\|_1^2 \geq \frac{r_{\min} S_{\rho, \epsilon}^2}{\tilde{S}}. \end{aligned} \quad (65)$$

Repeating the rest of the steps as above, we have:

Lemma 4.9. *An MIO channel mapping a state ρ^A , satisfying $r_{\min} \mathbb{1}^A \leq \rho^A \leq r_{\max} \mathbb{1}^A$, to an output σ^M such that $F(\sigma, \Psi_M) \geq 1 - \epsilon$ for $M = A/\tilde{S}$ must involve coherent measurement over at least*

$$\frac{\exp_2 \left[C_f(\rho) - \delta \log_2 A - (1 + \delta) h \left(\frac{\delta}{1 + \delta} \right) \right]}{M} \quad (66)$$

elements of A 's incoherent basis, where

$$\delta := \sqrt{2 \left(1 - S_{\rho, \epsilon} \sqrt{\frac{r_{\min}}{\tilde{S}}} \right)} \quad (67)$$

with $S_{\rho, \epsilon} := (1 - \epsilon) / r_{\max}$.

Remark 4.10. Our analysis of approximate distillation incorporated only the overall trace condition on $T_{\mathcal{E}}$; we have not found a way to work its parent constraint $\Delta(T_{\mathcal{E}}) = \mathbb{1}^A/M$ (resulting from Observation 3.3) into the reckoning. Considering the level of detail this would add, we expect its inclusion to significantly improve the relevant results (including those pertaining to asymptotic distillation, studied in the next section). On the other hand, our reliance solely on the trace condition expands the scope of these results' applicability beyond MIO (see Remark 3.4).

5 Asymptotic distillation

The asymptotic limit (see Section 1.7 for a brief background) is an important window into a resource theory. The behaviour of resource-theoretic quantities in this limit is aptly compared with the classic laws of thermodynamics:

asymptotic equipartition leads to certain near-universal features across diverse resource theories, such as extensivity of resource distillation yields and formation costs. We will now present some evidence suggesting the extensivity of the coherent measurement cost (quantified as the requisite number of elementary coherent gates such as the qubit Hadamard gate) of asymptotically maximal coherence distillation.

Chitambar [27] showed that the resource theory of coherence is asymptotically reversible under the class of free operations called dephasing-covariant operations (DIO), with the asymptotic rate of interconversion given by the relative entropy of coherence C_r . A consequence of this fact is that the rate of distillation of copies of Ψ_2 from those of a state ρ under even the largest class of free operations—the coherence-non-creating operations MIO—is bounded above by $C_r(\rho)$. Both DIO and IO, although strict subclasses of MIO, achieve this distillation rate [5].

For MIO achieving this maximal rate, we have the following conjectures.

Conjecture 1. *Suppose a sequence $\mathcal{E}_n$ of MIO channels is maximally-distilling on copies of an input ρ . That is, $\mathcal{E}_n$ acting on $\varrho_n \equiv \rho^{\otimes n}$ achieves $F[\mathcal{E}_n(\varrho_n), \Psi_{M_n}] \geq 1 - o(1)$ for $\log_2 M_n = n[C_r(\rho) - o(1)]$. Then any Kraus operator decomposition of $\mathcal{E}_n$ involves coherent measurement over at least L_n elements of the input's incoherent basis, where*

$$\log_2 L_n \geq n[\ell(\rho) - o(1)]. \quad (68)$$

This rank bound applies both to the single most coherent measurement element and to the average (of the rank's logarithm) over all involved measurements under the distribution induced by the input.

Conjecture 2. *For any ρ , there exists a sequence $\mathcal{E}_n$ of maximally-distilling IO channels with all but an asymptotically-vanishing fraction of measurements individually attaining the bound of Conjecture 1, as well as attaining it on average.*

5.1 Towards bounding the asymptotic cost

We will now attempt to prove Conjecture 1, essentially through a formalization of our crude typicality-based statements in Section 1.7. The

crudeness, when examined deeper, turns out unfortunately to conceal some finer features of asymptotic typicality that thwart our efforts at completing the proof. Nevertheless, we hope that the following account of our proof attempt elicits a more complete treatment from the community, thereby either proving or refuting our conjecture.

Remark 5.1. In the following, we will make use of the triangle inequality property of the *Fubini-Study metric* (also called the *fidelity angle*) $\theta(\rho, \sigma) := \arccos \sqrt{F(\rho, \sigma)}$; namely,

$$\theta(\rho, \sigma) + \theta(\sigma, \tau) \geq \theta(\tau, \rho). \quad (69)$$

In our calculations, the approximation parameters will be associated with $1 - F(\cdot, \cdot)$, whereby they are the *squared sines* of the associated fidelity angles. For convenience in applying the angle triangle inequality in their terms, we will use the shorthand

$$\begin{aligned} \epsilon \boxplus \delta &:= \sin^2 \left(\arcsin \sqrt{\epsilon} + \arcsin \sqrt{\delta} \right) \\ &= \left(\sqrt{\epsilon(1-\delta)} + \sqrt{\delta(1-\epsilon)} \right)^2. \end{aligned} \quad (70)$$

Note that $\epsilon \boxplus \delta \leq (\sqrt{\epsilon} + \sqrt{\delta})^2 \leq 4 \max\{\epsilon, \delta\}$ in general; but if $\delta \rightarrow 0$ while ϵ is held fixed, $\epsilon \boxplus \delta \rightarrow \epsilon$.

Proof sketch for Conjecture 1. We shall build on Lemma 4.9, with^{10,11}

$$“M” \equiv M_n = \exp_2 \left(n \left[C_r(\rho) - \epsilon_n^{(0)} \right] \right). \quad (71)$$

To estimate the other relevant parameters, we will use the quantum asymptotic equipartition property (AEP) reviewed in Appendix C.

For some $\delta_S > 0$, let $\varrho_n^{\delta_S}$ denote the unnormalized projection of ϱ_n onto its δ_S -weakly-typical subspace, and $\varrho_n^{|\delta_S}$ the normalized version thereof; the analog to “ ρ ” will be $\varrho_n^{|\delta_S}$. Applying Lemma C.4, $\text{Tr} \varrho_n^{\delta_S} \geq 1 - \delta_S$ and $2^{-n[S(\rho) + \delta_S]} \mathbb{1} \leq \varrho_n^{\delta_S} \leq 2^{-n[S(\rho) - \delta_S]} \mathbb{1}$, so that

$$2^{-n[S(\rho) + \delta_S]} \mathbb{1} \leq \varrho_n^{|\delta_S} \leq \frac{2^{-n[S(\rho) - \delta_S]} \mathbb{1}}{1 - \delta_S}. \quad (72)$$

¹⁰ We use quotes to indicate a variable from Lemma 4.9 whose analog here has a different notation, and whose Lemma 4.9 notation may here mean something different.

¹¹ Any ϵ_n 's we introduce shall be understood to be asymptotically-vanishing as $n \rightarrow \infty$ and the relevant AEP-related δ 's $\rightarrow 0$.

The lower bound above functions as “ $r_{\min}$ ”; we will presently define “ $r_{\max}$ ”, slightly differently from how we did in Lemma 4.9. Meanwhile, let us apply typicality on the classical sequences $\mathbf{a} \equiv (a_1 \dots a_n)$ formed by the incoherent basis labels occurring in ϱ_n , which are distributed according to $\Delta(\varrho_n) \equiv [\Delta(\rho)]^{\otimes n}$. For any $\delta_A > 0$ we can identify the δ_A -weakly-typical subalphabet $\mathcal{A}_n^{\delta_A}$. The corresponding system $A_n^{\delta_A}$ will be the analog to Lemma 4.9's “A”; by the AEP, its dimensionality

$$A_n^{\delta_A} \leq \exp_2 [n (S[\Delta(\rho)] + \delta_A)], \quad (73)$$

yielding the bound

$$“\tilde{S}” = A_n^{\delta_A} / M_n \leq \exp_2 \left(n \left[S(\rho) + \epsilon_n^{(0)} + \delta_A \right] \right). \quad (74)$$

We will now work towards bounding “ $S_{\rho, \epsilon}$ ”.

Let $\varrho_n^{\delta_A}$ denote the unnormalized projection of ϱ_n on this subalphabet, and $\varrho_n^{\backslash \delta_A}$ that on the complement thereof, so that $\text{Tr} (\varrho_n^{\delta_A} + \varrho_n^{\backslash \delta_A}) = 1$. We shall extend this superscript notation to projections of any operator. Furthermore, we will use superscripts combining “ δ_S ”, “ δ_A ”, and “ $\backslash$ ” to denote the results of successive projections from the inside out. For example, “ $\backslash \delta_S, \delta_A$ ” will denote a projection on the complement of the δ_S -typical subspace followed by one on the δ_A -typical subalphabet. As before, we will denote the corresponding normalized density operators by preceding the superscripts with “ $|$ ”. With this notational arrangement, first note that $\varrho_n^{\delta_S} + \varrho_n^{\backslash \delta_S} = \varrho_n$, since these projections are defined via ϱ_n 's eigenspaces. Note also that the δ_A -projections commute with the dephasing channel Δ , so that $\text{Tr} \varrho_n^{\backslash \delta_A} = \text{Tr} (\varrho_n^{\backslash \delta_A}) = \text{Tr} [\Delta(\varrho_n)]^{\backslash \delta_A} \leq \delta_A$, the last inequality following from the properties of the δ_A -typical subalphabet. Using these facts,

$$\begin{aligned} \text{Tr} (\varrho_n^{\delta_S, \backslash \delta_A} + \varrho_n^{\backslash \delta_S, \backslash \delta_A}) &= \text{Tr} \varrho_n^{\backslash \delta_A} \leq \delta_A \\ &\Rightarrow \text{Tr} \varrho_n^{\delta_S, \backslash \delta_A} \leq \delta_A \\ &\Rightarrow \text{Tr} \varrho_n^{\delta_S, \delta_A} = \text{Tr} (\varrho_n^{\delta_S} - \varrho_n^{\delta_S, \backslash \delta_A}) \geq 1 - \delta_S - \delta_A. \end{aligned} \quad (75)$$

As a step towards determining “ $S_{\rho, \epsilon}$ ”, we shall now show that ϱ_n is close to the normalized ver-

sion $\varrho_n^{|\delta_S, \delta_A}$ of the operator in the last line above.

$$\begin{aligned} F(\varrho_n, \varrho_n^{|\delta_S, \delta_A}) &= (\text{Tr} \varrho_n^{\delta_S, \delta_A})^{-1} F(\varrho_n, \varrho_n^{\delta_S, \delta_A}) \\ &= (\text{Tr} \varrho_n^{\delta_S, \delta_A})^{-1} F(\varrho_n^{\delta_A}, \varrho_n^{\delta_S, \delta_A}) \\ &\geq (\text{Tr} \varrho_n^{\delta_S, \delta_A})^{-1} \left(\frac{\text{Tr} \varrho_n^{\delta_S, \delta_A}}{\text{Tr} \varrho_n^{\delta_A}} \right)^2 \\ &\geq 1 - \delta_S - \delta_A, \end{aligned} \quad (76)$$

where we again used $\varrho_n = \varrho_n^{\delta_S} + \varrho_n^{\delta_S}$, this time to apply point 5 of Appendix A on $\varrho_n^{\delta_A} = \varrho_n^{\delta_S, \delta_A} + \varrho_n^{\delta_S, \delta_A}$. In a minor variation of the method of Lemma 4.9, we will make $\|\varrho_n^{|\delta_S, \delta_A}\|_\infty$ the analog of “ $r_{\max}$ ”—which is possible since we identify Lemma 4.9’s “A” with $A_n^{\delta_A}$. To bound this quantity, we first note that

$$\varrho_n^{\delta_S, \delta_A} \leq \varrho_n^{\delta_S} \leq \exp_2(-n[S(\rho) - \delta_S]) \quad (77)$$

by the quantum AEP. Since $\text{Tr} \varrho_n^{\delta_S, \delta_A} \geq 1 - \delta_S - \delta_A$

as noted in (75), this implies

$$\varrho_n^{|\delta_S, \delta_A} \leq \frac{\exp_2(-n[S(\rho) - \delta_S])}{1 - \delta_S - \delta_A}. \quad (78)$$

By assumption, the sequence $\mathcal{E}_n$ of channels achieves $F[\mathcal{E}_n(\varrho_n), \Psi_{M_n}] \geq 1 - \epsilon_n$. By (76), $F(\varrho_n, \varrho_n^{|\delta_S, \delta_A}) \geq 1 - \delta_S - \delta_A$, so that (by contractivity and the fidelity-angle triangle inequality) $F[\mathcal{E}_n(\varrho_n^{|\delta_S, \delta_A}), \Psi_{M_n}] \geq 1 - \epsilon_n^{(1)}$ with $\epsilon_n^{(1)} := \epsilon_n \boxplus (\delta_S + \delta_A)$. The associated target effects $T_n \equiv T_{\mathcal{E}_n}$ must therefore satisfy $\text{Tr}(T_n \varrho_n^{|\delta_S, \delta_A}) \geq 1 - \epsilon_n^{(1)}$. Using an argument similar to Lemma 4.9’s based on (78),

$$\begin{aligned} \text{Tr} T_n^{\delta_A, \delta_S} &\geq \frac{1 - \epsilon_n^{(1)}}{\|\varrho_n^{|\delta_S, \delta_A}\|_\infty} \\ &\geq [1 - \epsilon_n^{(1)}] (1 - \delta_S - \delta_A) 2^{n[S(\rho) - \delta_S]}. \end{aligned} \quad (79)$$

This bound is analogous to “ $S_{\rho, \epsilon}$ ”. Note the order of projections in $T_n^{\delta_A, \delta_S}$: we are essentially using $T_n^{\delta_A}$ as “ $\mathcal{T}_\mathcal{E}$ ” and $\varrho_n^{\delta_S}$ as “ ρ ”, whereby $T_n^{\delta_A, \delta_S}$ plays the role of “ $\mathcal{T}_\mathcal{E}^\rho$ ”. Putting the pieces together as in (65),

$$F(\tau_n^{|\delta_A}, \varrho_n^{|\delta_S}) \geq [1 - \epsilon_n^{(1)}]^2 (1 - \delta_S - \delta_A)^2 \exp_2(-n[\epsilon_n^{(0)} + 3\delta_S + \delta_A]) =: 1 - \epsilon_n^{(2)}. \quad (80)$$

We have now arrived at our primary obstacle: for this $\epsilon_n^{(2)}$ to be a vanishing sequence, we would need the exponents to all vanish. The $n\epsilon_n^{(0)}$ is obviously menacing, as the definition of maximal distillation makes no stipulation whatsoever on how fast the $\epsilon_n^{(0)}$ must decay. As for the $n\delta$ ’s: naively, we might expect that eventually taking the limits $\delta_S, \delta_A \rightarrow 0$ would at least get rid of these exponents. However, these parameters cannot be taken to zero *independently of n* : for any given δ_S , the bounds in (72) are guaranteed only “for large enough n ”; likewise for δ_A and (73). Indeed, the requisite n to validate these bounds scales as δ^{-2} , whereby $\exp_2(-n\delta) \sim \exp_2(-\sqrt{n})$. This puts a definitive end to any prospects of a bound like (80) succeeding.

Let us now pretend we did not encounter the above problem, and continue our proof sketch as

if $\epsilon_n^{(2)} \in o(1)$. Using a projection-related property of the fidelity (point 4 of Appendix A), $F(\varrho_n, \varrho_n^{|\delta_S}) = \text{Tr} \varrho_n^{\delta_S} \geq 1 - \delta_S$. Combining this with (80) through the angle triangle inequality,

$$F(\tau_n^{|\delta_A}, \varrho_n) \geq 1 - \epsilon_n^{(3)}, \quad (81)$$

where $\epsilon_n^{(3)} := \epsilon_n^{(2)} \boxplus \delta_S$. Finally, “ δ ”, for which we can conveniently use the same symbol, is given by

$$\delta := \sqrt{2 \left(1 - \sqrt{1 - \epsilon_n^{(3)}} \right)}. \quad (82)$$

Thus, the coherent measurement rank for the action of $\mathcal{E}_n$ on $A_n^{\delta_A}$ is no less than

$$\begin{aligned} L_n &:= M_n^{-1} \exp_2[C_f(\tau_n^{|\delta_A})] \\ &\geq \frac{2^{C_f(\varrho_n) - \delta \log_2 A_n - (1+\delta)h(\frac{\delta}{1+\delta})}}{M_n}, \end{aligned} \quad (83)$$

where $A_n \equiv A^n$, with $A := \text{rank} \Delta(\rho)$, is the dimensionality of the entire Hilbert space where ϱ_n acts. By the additivity of the coherence of formation under tensor products, $C_f(\varrho_n) = nC_f(\rho)$, while $\log_2 M_n = n [C_r(\rho) - \epsilon_n^{(0)}] = n [C_r(\rho) - o(1)]$. We now take the limit as both δ_S and δ_A approach zero, whereupon $\epsilon_n^{(1)} \rightarrow \epsilon_n$, $\epsilon_n^{(2)} \leq 2\epsilon_n + n\epsilon_n^{(0)}$, and $\epsilon_n^{(3)} \leq 2\epsilon_n + n\epsilon_n^{(0)}$ as well. Thus, if we resolve to ignore the $n\epsilon_n^{(0)}$, we get $\delta \leq \sqrt{2(1 - \sqrt{1 - 2\epsilon_n})} \leq \sqrt{2\epsilon_n} \in o(1)$, and so

$$\begin{aligned} L_n &\geq \exp_2(n [C_f(\rho) - C_r(\rho) - o(1)] - o(1)) \\ &\geq \exp_2(n [\ell(\rho) - o(1)]). \end{aligned} \quad (84)$$

Finally, in light of (81), we conclude that the bound applies asymptotically also to the *average* logarithmic measurement coherence rank under the distribution induced by the input ϱ_n . $\square$

We attempted to prove the conjecture via the asymptotic continuity of the coherence of formation (Lemma B.2 in Appendix B). As such, we set ourselves the tall order of showing that the normalized target effect τ_n is close in fidelity to ϱ_n . In retrospect, the expectation that factors as large as $2^{nS(\rho)}$ mutually cancel to leave something close to 1 was naively optimistic: the normalization involves exponential factors with the AEP-related δ 's, not to mention the $\epsilon_n^{(0)}$ from M_n that we already had to contend with. Indeed, even if τ_n were exactly proportional to $\mathbb{1}_{\varrho_n^{\delta_S}}$ —which is the most complete characterization we got in the non-asymptotic case—we would then be required to show that the latter is close to $\varrho_n^{\delta_S}$. While AEP does guarantee that $\varrho_n^{\delta_S}$ gets rather flat in its spectrum, it is not nearly flat enough to have high fidelity with the maximally-mixed state on its support.

A more direct approach might try to avoid having to precisely trade exponentially large factors. For example, there might be a modified “asymptotic continuity” property that holds for density operators that are not necessarily close in fidelity. Of course there can be no such property for fully general pairs of operators σ^{A_n} and τ^{A_n} . But in our context, the operators come from sequences of $T_n \leq \mathbb{1}$ and $\varrho_n = \rho^{\otimes n}$ that satisfy $\text{Tr}(T_n \varrho_n) \rightarrow 1$; the T_n also satisfy $\text{Tr} T_n^{\delta_A} \approx A_n^{\delta_A} / M_n$ (which we used) and the stronger condition (following from Observation 3.3) $\langle \mathbf{a} | T_n | \mathbf{a} \rangle = M_n^{-1}$ for all $\mathbf{a} \in \mathcal{A}_n$ (which could yet be incorporated, though we could not find a way to).

This latter property reveals an interesting connection between the target effect and the generalized quantum Stein’s lemma, an important result in quantum statistics and the general overarching theory of resource theories. The lemma was initially believed to have been proved by Brandão and Plenio [28], but a gap in their proof was discovered by Berta *et al.* [29], who were able to prove the lemma by alternate means for a subclass of resource theories, including those of coherence. Remarkably, Hayashi and Yamasaki [9] and Lami [10] independently proved the lemma recently in full generality. The connection to our problem, evident from the previous paragraph’s discussion, is as follows:

Observation 5.2. *Given an MIO sequence $\mathcal{E}_n$ that asymptotically distills maximally from copies of an input ρ , the sequence of associated target effects T_n saturates the type-II error exponent scaling bound set by the generalized quantum Stein’s lemma. In fact, $\text{Tr}(\sigma T_n) = M_n^{-1}$ identically for all free states σ , with M_n saturating the scaling bound. Consequently, the coherent measurement cost (quantified by any measure) of maximal asymptotic distillation is bounded below by the infimum of the values attained by the measure over the set of hypothesis-testing effect sequences saturating the lemma’s bound.*

All this additional structure could admit a modified notion of asymptotic continuity. As a signature of the special structure, we can observe from our proof sketch that $F(\tau_n, \varrho_n)$ can only decay sub-exponentially, which is not a generic property among pairs of state sequences. Unfortunately, this property alone is not enough to ensure that $C_f(\tau_n)$ differs at most sub-extensively from $C_f(\tau_n)$: a simple counterexample is $\tau_n = \epsilon_n \varrho_n + (1 - \epsilon_n) \sigma_n$ for any coherent ρ , incoherent σ_n , and sub-exponentially-decaying ϵ_n . All the same, considered together with the other properties discussed above, it lends credence to our conjecture.

Our problem might also benefit from the tools and methods of the smooth entropy calculus [30]. These tools are often used to derive AEPs for entropy-like quantities (see, e.g., [31]). An example relevant to our problem is the AEP

$$\lim_{\epsilon \rightarrow 0^+} \lim_{n \rightarrow \infty} \frac{\log_2 r_C^\epsilon(\rho^{\otimes n})}{n} = C_f(\rho), \quad (85)$$

due to Zhao *et al.* [32]; here, r_C^ϵ is a so-called *smoothed* variant of r_C , with ϵ a real smoothing parameter. This property is reflected in one of the “flattening” steps we are able to accomplish (point 3) in our construction in Section 5.2. Indeed, $C_f(\rho)$ is a good candidate for a universal AEP limit for all reasonable coherence measures based on optimizations over convex decompositions. The bounds encountered in our problem are all measures of this kind, though evaluated not on the input state $\varrho_n \equiv \rho^{\otimes n}$ itself, but instead on the maximally mixed state τ_{ϱ_n} on its support (or variations thereof). Yet, considering that ϱ_n flattens out spectrally as n grows, it is not unreasonable to expect $C_f(\rho)$ to emerge as the limiting rate even for measures evaluated on τ_{ϱ_n} .

A possible way to put such cases on an equal footing with those like (85) is to use a normalized version of ϱ_n^α , for $\alpha \in \mathbb{R}$, as the argument; in this landscape, $\tau_{\varrho_n} \propto \varrho_n^0$. We could suitably define smoothed counterparts of the measures; Definition 7 is an example of a possible way to effect the smoothing. It is plausible then that there is an AEP of the form

$$\lim_{\epsilon \rightarrow 0^+} \lim_{n \rightarrow \infty} \frac{1}{n} C^\epsilon \left(\frac{\varrho_n^\alpha}{\text{Tr} \varrho_n^\alpha} \right) = C_f(\rho) \quad (86)$$

for a whole class of measures C . If true, this could give us a way of relating our $r_C(\tau_n)$ with $C_f(\rho)$, thereby potentially proving Conjecture 1.

Of course, it is a different matter that all such quantities—thanks to their involving optimizations over all convex decompositions—would likely be hard to compute. Nevertheless, there might still be a way to *formally* establish an AEP and relate all of these (computationally intractable) measures evaluated on ϱ_n to the (also computationally intractable, but better-understood) $C_f(\rho)$.

Finally, a caveat is also in order concerning the claim “[t]his rank bound applies both to the single most coherent measurement element and to the *average* (of the rank’s logarithm) over all involved measurements under the distribution induced by the input” in Conjecture 1: the failure of (81) deals a blow to the average clause, even were the rest of the conjecture valid. Nevertheless, by virtue of the logarithms, we remain hopeful that any future technique capable of proving the latter would also be up to the more exacting task of proving the whole conjecture.

5.2 How to attain the bounds

From putting a lower bound on the asymptotic coherent measurement cost, we now turn to attaining the bound (Conjecture 2). One way of proving Conjecture 2 would be to explicitly construct bound-attaining distillation channels. We now formulate some general guiding principles towards such constructions. Let us take (for example) the result of Proposition 4.7 and inspect the chain (47) of inequalities therein:

$$\begin{aligned} \log_2 M + \sum_{c,m} \frac{\alpha_m^2 t_{c,m}}{\tilde{S}} C_r(\phi_{c,m}) \\ \geq \sum_c \frac{t_c \left(H \left[\left(\frac{\alpha_m^2 t_{c,m}}{t_c} \right)_m \right] + \sum_m \frac{\alpha_m^2 t_{c,m}}{t_c} C_r[\phi_{c,m}] \right)}{S} \\ = \sum_c \frac{t_c}{S} C_r(\phi_c) \geq C_f(\tau_\mathcal{E}) \geq C_0. \end{aligned} \quad (87)$$

For the last inequality to be saturated, we need $\dim \mathcal{V} = S$ —i.e., our “maximal distillation” condition $M \dim \mathcal{V} = A$. For the one before, the decomposition $\sum_c (t_c/\tilde{S}) \phi_c$ of $\tau_\mathcal{E}$ must be one that attains the bound in the definition of $C_f(\tau_\mathcal{E})$ —a so-called optimal decomposition. Finally, the inequality in the middle line is saturated when $(\alpha_m^2 t_{c,m}/t_c)_m$ is uniform for each c . This is the case for a maximal uniform distillate, i.e. $|\alpha\rangle = |\Psi_M\rangle$, as mentioned in Remark 4.4; we are not aware of weaker conditions where it still holds.

These conditions would suffice, in principle, for the loosest bound of Proposition 4.7 (effectively Proposition 4.3, considering the above) to be attained without necessarily saturating the inequality in Theorem 1. The latter would further require

1. each of the $C_r(\phi_c)$ to be equal to $C_f(\tau_\rho)$;
2. each of the $C_r(\phi_{c,m})$ to be equal, in turn, to $C_f(\tau_\rho) - \log_2 M$; and moreover,
3. each $\phi_{c,m}$ to be a uniform superposition.

Condition 1 occurs when τ_ρ is a so-called *flat-roof point* for the convex-roof function in question [33]. There does not seem to be any work in the literature on flat-roof points for the function C_f . As for the further conditions 2 and 3, suppose

- 1 holds;

- the Kraus operators are IO;
- each $(\alpha_m^2 t_{c,m}/t_c)_m$ is uniform; and furthermore,
- each ϕ_c is a uniform superposition.

This would automatically ensure both 2 and 3; once again, we are not aware of situations where these conditions could be met without those in the previous sentence. Finally, attaining the rank bound $\ell(\tau_\rho)$ would necessitate $C_r(\tau_\rho) = \log_2 M$ —i.e., the distillate is also maximal in terms of C_r .

For each inequality in these chains, the condition for its saturation is essentially some sort of “flattening” of features. We are not aware of any noteworthy situations in which some of the bounds are tight while others are not; in particular, whether the “average coherence” bound of Proposition 4.3 can be attained in cases where the maximal rank bound of Theorem 1 cannot is an intriguing open question. There is one situation, though, where all manner of flattening tendencies collude: maximal asymptotic distillation. We use the above observations to attempt a systematic construction (summarized towards the end of Section 1.8) of a distillation protocol that asymptotically attains the ℓ -based rank bound of Conjecture 1. The construction is summarized as follows:

1. In our proof sketch for Conjecture 1, we attempted to show that the target effect-based density operators $\tau_n \equiv \tau_{\mathcal{E}_n}$ associated with any sequence of asymptotically maximally-distilling MIO channels $\mathcal{E}_n$ must approach ϱ_n [see (81)]. Any given set of Kraus operators of $\mathcal{E}_n$ correspond to the pure states in a certain convex decomposition of τ_n (Observation 3.2), which would then be an approximate convex decomposition of ϱ_n . For the converse, we adapt Winter and Yang’s maximal IO distillation protocol [5] to derive conditions under which a given approximation sequence τ_n corresponds to some maximally-distilling channel sequence $\mathcal{E}_n$; we also show that *any* convex decomposition of such a τ_n yields IO Kraus operators for $\mathcal{E}_n$. We start our construction with ϱ_n itself, progressively working towards a viable τ_n .
2. If we choose an optimal decomposition attaining $C_f(\varrho_n)$ from the convex roof of ϱ_n ,

we already obtain Kraus operators that attain the measurement coherence bound in the *average* sense of Proposition 4.3. It remains to flatten out further to attain the bound on a per- $|\phi_{c,m}\rangle$ basis.

3. First, to flatten relative to c , we construct a decomposition wherein all but an asymptotically-vanishing weight is carried by pure states whose individual C_r values are close to $C_f(\varrho_n)$: simply take a decomposition $\rho = \sum_j q_j \phi_j$ that is optimal for ρ , and decompose (most of) ϱ_n into pure states of the form $|\Phi_{c \equiv \mathbf{j}}\rangle \equiv \bigotimes_{k=1}^n |\phi_{j_k}\rangle$ where $\mathbf{j} \equiv (j_k)_k$ is a strongly-typical sequence under $\mathbf{q}^{\otimes n}$ (see Appendix C for background on asymptotic typicality)¹². These pure components then have $C_r(\Phi_{\mathbf{j}}) \approx C_f(\varrho_n) = nC_f(\rho)$.
4. To ensure that the final target effect is bounded as $T_n \leq \mathbb{1}$ (as required for the associated map to be a valid subchannel), we project all remaining $|\Phi_{\mathbf{j}}\rangle$ onto a strongly-typical subspace of ϱ_n , resulting in $|\Phi_{\mathbf{j}}^{\delta_S}\rangle$.
5. To flatten relative to m , we first show that each $|\Phi_{\mathbf{j}}^{\delta_S}\rangle$ can be made arbitrarily close to a near-uniform superposition $|\Phi_{\mathbf{j}}^{\delta_S, \delta_A}\rangle$.
6. We then draw again on Winter–Yang’s construction to show that we can discard an asymptotically-vanishing fraction of the remaining vectors to leave only ones almost entirely contained (with an asymptotically-vanishing error) in a subspace $\mathcal{V}$ with the following property: any $|w\rangle \in \mathcal{V}$ can be decomposed as $|w\rangle = M_n^{-1/2} \sum_m |w_m\rangle$, with $\langle w_m | w_m \rangle = \langle w | w \rangle \forall m$, such that $|w_m\rangle \in \text{span}\{|\mathbf{a}\rangle : \mathbf{a} \in \mathcal{I}_m\}$, where $\{\mathcal{I}_m\}_m$ is a *fixed* (i.e., $|w\rangle$ -independent) disjoint partitioning of $\mathcal{A}_n \equiv \mathcal{A}^n$. The latter property enables the vectors to be used in constructing IO Kraus operators, while the former ensures that they are weighted equally over m . Thus, we now have $|\Phi_{\mathbf{j}}^{\delta_S, \delta_A}\rangle \approx M_n^{-1/2} \sum_m |\Phi_{\mathbf{j},m}^{\delta_S, \delta_A}\rangle$

¹² This works except when $\mathbf{q}$ is uniform. For this case, we can use this construction for arbitrarily small non-uniform perturbations of $\mathbf{q}$, resulting in corresponding perturbations of ρ . The properties of convex roofs ensure that the perturbed $\mathbf{q}$ decomposition is optimal for the perturbed ρ [34].

with $\langle \Phi_{j,m}^{\delta_S, \delta_A} | \Phi_{j,m}^{\delta_S, \delta_A} \rangle$ close to a uniform distribution over m .

7. We then show that we can delete a vanishing fraction of $|\Phi_{j,m}^{\delta_S, \delta_A}\rangle$ from each remaining $|\Phi_j^{\delta_S, \delta_A}\rangle$ such that the remaining have logarithmic coherence rank tightly concentrated around $C_r(\Phi_j^{\delta_S, \delta_A}) - \log_2 M_n \approx n[C_f(\rho) - C_r(\rho)] = n\ell(\rho)$. We thus obtain IO Kraus operator fragments $|w_{c,m}\rangle \propto |\Phi_{j,m}^{\delta_S, \delta_A}\rangle$ that individually approach the measurement coherence rank bound of Proposition 1.
8. Finally, we show that the “IO Kraus operators” (in quotes because the resulting map may fail to be trace-non-increasing) constructed using these fragments asymptotically output an approximation to Ψ_{M_n} near-deterministically. In the case that the map is a sub-channel, we show that it can be completed to a channel with an IO residual sub-channel incurring a measurement coherence overhead asymptotically vanishing in relation to the rest.

Appendix D provides a detailed survey of the above construction and an attempt to thereby prove Conjecture 2. The overall difficulties here are much more involved than in the case of Conjecture 1. This suggests that Conjecture 2 may stand a bleaker chance of holding up. We now close this section and turn to the implications of our conjectures, were they to hold.

6 Ramifications of our conjectures

If true, conjectures 1 and 2 would pin down the coherent measurement cost of asymptotically maximal distillation, up to subextensive terms. We shall now explore some implications that would then follow. We will first discuss the operational implications for maximal distillation, and then sketch some possibilities for an asymptotic tradeoff between the coherent measurement budget and the distilled yield.

All statements in the remainder of this section will be premised on conjectures 1 and 2.

6.1 Maximal distillation may be a net loss

Recall our observation from Section 1.2 that dynamical coherence can be used to endue both

incoherent states and incoherent measurements with (respectively, static and mensural) coherence. Indeed, although the latter forms of coherence are well-defined operational primitives in the formalism, in practice we always derive them from dynamical coherence—we never find ourselves in possession of the operational capability to prepare or measure coherently but not to implement coherent dynamics. Thus, it is justified to consider all aspects of coherence involved in a task as ultimately dynamical. In this spirit, we now weigh the coherence used in implementing coherent measurements (whose quantification has been our main preoccupation) against that eventually distilled in the form of standard resource states.

According to Conjecture 1, the coherent measurements used in any maximal distillation protocol (on average, up to subextensive terms) would be equivalent to a number of Hadamard measurements no smaller than $n\ell(\rho)$, where n is the number of copies of the input ρ used. Meanwhile, the distilled yield is equivalent to $nC_r(\rho)$ cobits. Could the coherent measurement cost exceed the yield, i.e. $\ell(\rho) > C_r(\rho)$?

Indeed, this does sometimes happen. While no general method is known for efficiently computing the coherence of formation, it does have a closed-form expression for states of a single qubit [24]:

$$C_f(\rho) = h\left(\frac{1 + \sqrt{1 - 4|\langle 0|\rho|1\rangle|^2}}{2}\right). \quad (88)$$

Using this formula, we computed the excess coherence cost $\ell(\rho) - C_r(\rho)$ for a representative sample of qubit states (Fig. 1). The results suggest that a nonzero measure of states incur an excess cost for maximal distillation, and moreover, that this excess cost can be an arbitrarily large multiple of the distilled yield!

Coherence distillation is sometimes an instrumental or incidental outcome in a larger task, e.g. entanglement distillation using incoherent local operations [35]. But if the task is coherence distillation itself, then our results call into question its operational utility in situations where $\ell(\rho) \geq C_r(\rho)$: if we are able to implement $n\ell(\rho)$ Hadamard gates, we should rather use them to simply prepare as many fresh cobits than squander them in distilling only $nC_r(\rho)$.

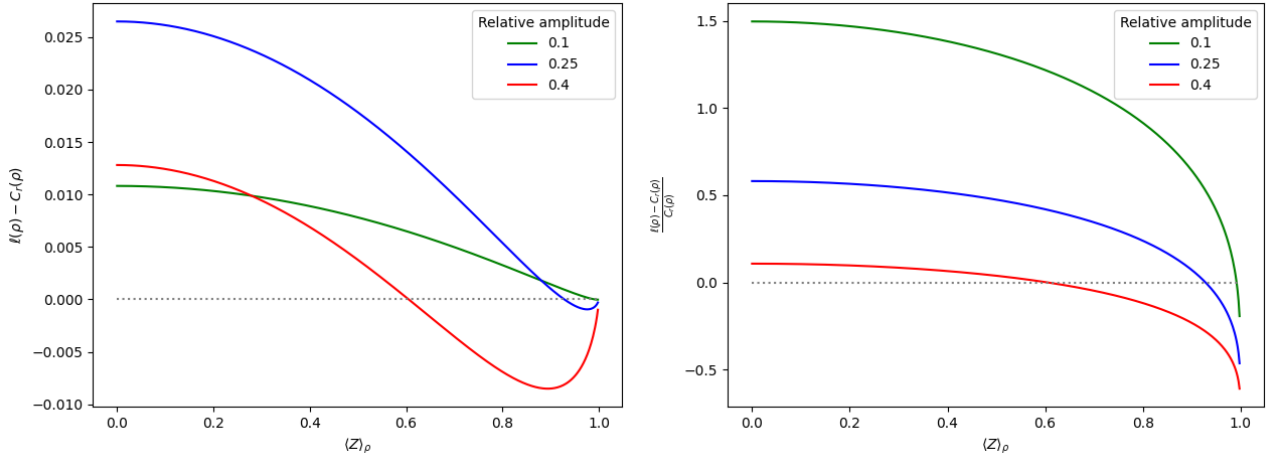


Figure 1: (Left) Excess coherent measurement cost, i.e. the difference between the asymptotic coherent measurement cost $\ell(\rho)$ and the distillable coherence $C_r(\rho)$, for qubit states with different values of the relative off-diagonal amplitude $|\langle 0|\rho|1\rangle|/\sqrt{\langle 0|\rho|0\rangle\langle 1|\rho|1\rangle}$: the cost generically exceeds the yield for relative amplitude $\lesssim 0.4$. (Right) The ratio of the excess coherent measurement cost $\ell(\rho) - C_r(\rho)$ to the distillable coherence $C_r(\rho)$: the ratio attains arbitrarily large values for low enough relative amplitude.

6.2 Asymptotic cost–yield tradeoff

An obvious question that arises is what the coherent measurement cost of asymptotically–non-maximal distillation is. In this section we will make some conjectures on this question, restricting our consideration to IO. Since the coherence rank $M_n \approx \exp_2[n C_r(\rho)]$ of the maximal distillate and that of the conjectured requisite measurement, $L_n \approx \exp_2[n \ell(\rho)]$, both scale exponentially in the number of copies, one might naively expect that the best one can do with a measurement rank scaling $\tilde{L}_n \approx (L_n)^t$ with $0 \leq t < 1$ is to distill maximally from a fraction t of the input’s copies, thereby achieving $\tilde{M}_n \approx (M_n)^t$, corresponding to the distillation rate $t C_r(\rho)$.

However, this would only be a loose lower bound on the achievable rate, as can already be seen by putting Conjecture 2 together with past results on SIO distillation [7, 8]. Recall that SIO are just IO that don’t use coherent measurements, i.e. the case $t = 0$. We know that the maximal SIO-distillable rate is nonzero for certain inputs whose maximal IO rate is strictly larger—thus, in these cases, although $\log_2 L_n \in \Omega(n)$ for maximal IO distillation (assuming Conjecture 1), $\tilde{L}_n = 1 \in O[(L_n)^{t=0}]$ is nevertheless able to achieve $\Omega(n) \ni \log_2 \tilde{M}_n \notin O[tn C_r(\rho)]$.

Can we go further and find the exact asymptotic tradeoff between the cost and the yield?

Definition 8 (Feasible cost–yield pair). $(l, r) \in \mathbb{R}^2$ is a *feasible cost–yield pair* for ρ if there exists a sequence $\mathcal{E}_n$ of IO channels that respectively distill from $\rho^{\otimes n}$ at the asymptotic rate r and coherent measurement cost l , defined as

$$l = \lim_{n \rightarrow \infty} \frac{\log_2 L_n}{n} \quad (89)$$

with L_n the cost of $\mathcal{E}_n$ both on (logarithmic) average and on a per-measurement basis (for almost all measurements involved)—the two being identical under Conjecture 2.

For example, $(l, r) = (\ell[\rho], C_r[\rho])$ is a feasible pair for ρ according to Conjecture 2. Taking Theorem 2 and conjectures 1 and 2 as clues, we now make the informed guess that some quantity like the ones defined in Definition 7 would quantify the necessary coherent measurement cost of non-maximal asymptotic distillation. If so, the freedom we have in optimizing a distillation strategy would consist of choosing an appropriate density operator $\tau_\perp$ orthogonal to ϱ_n , such as to minimize the mixture’s C_f . Intuitively, for $\tau_\perp$ to help reduce the final C_f , its incoherent alphabet must overlap with ϱ_n ’s as much as possible.

One possibility suggested by this intuition is to pick $\tau_\perp$ so that the mixture becomes $\sigma^{\otimes n}$ for some state σ satisfying $\Delta(\sigma) = \Delta(\rho)$. In particular, define the set

$$\mathcal{D}(\rho) := \left\{ \sigma : \Delta(\sigma) = \Delta(\rho), \rho \xrightarrow{\text{SIO}} \sigma \right\}. \quad (90)$$

As per Conjecture 2, $(l, r) = (\ell[\sigma], C_r[\sigma])$ for any $\sigma \in \mathcal{D}(\rho)$ would also be a feasible pair for ρ , since each copy of ρ could first be mapped to one of σ without requiring coherent measurements. A special case of this is when σ is Lami's $\bar{\rho}$ [7, 8], the state consisting only of the pure diagonal blocks of ρ —we then get the SIO-feasible pair $(\ell[\bar{\rho}], C_r[\bar{\rho}]) = (0, Q[\rho])$, where $Q[\rho]$ is Lami's *quintessential coherence*.

We can achieve any convex combination of a set of feasible pairs by performing the respective protocols achieving them on appropriate fractions of the total number of input copies. Finally, the feasibility of (l, r) trivially implies that of any $(l, r_1 < r)$. Thus, Conjecture 2 implies that any pair in the following set is feasible for ρ :

$$\mathcal{C}(\rho) := \text{cvx} \{(\ell[\sigma], r) : r \in [0, C_r(\sigma)], \sigma \in \mathcal{D}(\rho)\}. \quad (91)$$

The symmetry and typicality properties of the asymptotic limit suggest that there might be no nontrivial feasible pairs besides these. We discussed above why we believe it would not help to expand $\mathcal{D}(\rho)$ to include states with a diagonal part different from the input's. As for the possibility of collective SIO pre-processing on many copies of ρ , Lami's results on SIO distillation hint against the prospect of this producing new feasible pairs. There would of course be IO strategies that deviate from maximal distillation in ways other than an SIO pre-processing. But based on our preliminary studies of distillation in the dilation picture (Appendix G), we are inclined to believe that these variations do not expand the feasible set either. Hence, we have the following:

Conjecture 3. *The $\mathcal{C}(\rho)$ defined through (90) and (91) is the set of all feasible cost–yield pairs for ρ .*

In the above line of reasoning we considered using the coherence budget l only in implementing an IO acting on the given input. But allowing the coherent action to be used possibly also in preparing fresh coherent states (as discussed in Section 6.1) is arguably more operationally meaningful. In that case, (l, l) should also be considered a feasible pair for any $l \geq 0$ and any ρ . If we wish to include these in our reckoning, we could modify the definition (91) by replacing $C_r(\sigma)$ with $\max\{\ell(\sigma), C_r(\sigma)\}$.

In general, the problem of computing the largest achievable rate r for a given coherent

measurement budget l , or inversely, that of computing the least l achieving a desired rate r , is likely to be very hard (even if our conjectures hold)—after all, we have no known efficient way to compute $\ell(\rho)$ itself, saying nothing of the difficulty of finding the set $\mathcal{D}(\rho)$. But we expect there to be a nontrivial cost–yield tradeoff landscape, finding which would be both fundamentally and operationally important.

7 Conclusion

We laid out a framework for quantifying the cost of coherent measurements involved in distilling coherent states from arbitrary inputs using coherence–non-creating channels (MIO). The framework uses a construction we call the *target effect*—a measurement effect that quantifies the probability with which a given channel maps an input to a desired target state. This object also doubles up as a quantifier of the requisite coherent measurement action in implementing the channel. We derived conditions on the target effect—and thereby, lower bounds on the coherent measurement cost—for exact (maximal and non-maximal) and approximate maximal coherence distillation from finite-sized inputs.

Based on our result on the approximate case, we conjectured a scaling law for the coherent measurement cost of asymptotic distillation at the maximal rate—namely, that the necessary and sufficient cost (in an equivalent number of qubit Hadamard gates) is extensive in the number of input copies, with a rate given by the input's irretrievable coherence $\ell(\rho) = C_f(\rho) - C_r(\rho)$. We went through detailed and rigorous proof sketches for both the necessity and the sufficiency of this cost, and discussed the difficulties we encountered in completing the proofs. As a byproduct, we showed a connection between our problem and the generalized quantum Stein's lemma. By virtue of this connection, the coherent measurement cost of maximal asymptotic distillation can be bounded by the coherence of hypothesis-testing effects that saturate the error exponent scaling bound in the lemma—a potential tactic for future attempts at our conjectures.

We then discussed some implications of our conjectures. First, we noted that our conjectured coherent measurement cost exceeds the very distilled yield in a nonzero measure of qubit in-

stances of maximal distillation. Thus, maximal coherence distillation as a standalone operational task may sometimes amount to a net attrition of coherence-resource and should therefore be superseded by coherent state preparation. We then made some speculations on an asymptotic trade-off between the coherent measurement budget and distilled yield.

A question of possible interest for future work is whether the appearance of the irretrievable coherence—a signature of irreversibility—is accidental, or if, rather, there is a deeper connection between resource-theoretic irreversibility and auxiliary (or otherwise hidden) costs of distillation. The fact that our results all apply to MIO, which do not exhibit irreversibility, supports the accident hypothesis, but there may yet be some subtleties that we have missed.

Our results and conjectures have foundational significance in understanding the resource theory of coherence, but also operational implications in applications that use the resource. For example, combining our results with those of Ref. [18] directly implies that our coherent measurement cost bounds apply also to the task of private incoherent randomness extraction from the same input state; our Conjecture 1, if true, would further imply that, given an operational coherence budget, more private randomness can sometimes be extracted by ignoring the input and instead simply preparing fresh coherent states (to be incoherently measured for generating randomness). Further investigation on our conjectures may shed light on information-theoretic aspects of coherence manipulation, e.g. a possible asymptotic equipartition property of convex-roof extensions of entropic coherence quantifiers. Developing methods and heuristics to actually compute such quantifiers would be a challenging project in itself. Other natural avenues for inquiry would be in situations involving the interplay of coherence with other resources, e.g. multipartite settings with local incoherent operations [35].

Naively, one might expect non-SIO IOs to be implementable as SIO augmented by the consumption of coherent states (a basic treatment of whose asymptotic distilling power was done by Lami [8]). However, nontrivial coherent state-augmented SIO may not even be IO—at any rate, the effective Kraus operators induced by an SIO Kraus operator representation of the

original SIO channel fail to be IO. This necessitates other methods, such as ours, for studying the coherent measurement cost of tasks.

Besides the target effect-based approach detailed in this paper, we also attempted some alternative approaches towards quantifying the coherent measurement cost of distillation. In Appendix E, we present some preliminary results towards a semidefinite programming (SDP)-based approach. In Appendix F, we describe an approach wherein we studied the behaviour under constrained IO of certain SIO monotones constructed by Lami [8], in the process defining a generalization of the measures. A better understanding of how Lami’s monotones, and generalizations thereof, behave under tensor products (especially in the asymptotic limit) seems essential for studying the coherent measurement cost either directly through constrained IO (as we do in Appendix F) or by relating it with the dynamical coherence [36, 37] of unitary dilations.

Lami relates these coherence measures to the information-theoretic properties of a classical variable labelling the pure diagonal blocks of the input state—i.e., classical information encoded in a certain way in coherent quantum states. Generalizations such as ours may be operationally related to corresponding generalized encoding tasks. This suggests as-yet unexplored connections between Shannon theory and the resource theory of coherence.

In Appendix G, we present our preliminary investigations on what we call decoupling schemes: certain linear-algebraic structures that emerge when distillation is framed in the dilation picture. These suggest yet other fruitful lines of inquiry, including possible connections with established notions of decoupling [38, 39, 40, 41].

Acknowledgements

We thank Anurag Anshu, Kishor Bharti, Dagmar Bruß, Ian George, Ludovico Lami, Alessandro Luongo, Iman Marvian, Ryuji Takagi, Thomas Theurer and Benjamin Yadin for helpful discussions. Special thanks are due to Eric Chitambar, Mile Gu and Yunlong Xiao for particularly detailed and engaging discussions. This project was originally conceived during a collaboration with Francesco Buscemi and Bartosz Regula. We express our sincerest gratitude for their

References

- [1] Alexander Streltsov, Gerardo Adesso, and Martin B Plenio. “Colloquium: Quantum coherence as a resource”. *Reviews of Modern Physics* **89**, 041003 (2017).
- [2] Johan Åberg. “Subspace preservation, subspace locality, and gluing of completely positive maps”. *Annals of Physics* **313**, 326–367 (2004).
- [3] Johan Åberg. “Quantifying superposition” (2006).
- [4] Tillmann Baumgratz, Marcus Cramer, and Martin B Plenio. “Quantifying coherence”. *Physical review letters* **113**, 140401 (2014).
- [5] Andreas Winter and Dong Yang. “Operational resource theory of coherence”. *Physical review letters* **116**, 120404 (2016).
- [6] Benjamin Yadin, Jiajun Ma, Davide Girolami, Mile Gu, and Vlatko Vedral. “Quantum processes which do not use coherence”. *Physical Review X* **6**, 041028 (2016).
- [7] Ludovico Lami, Bartosz Regula, and Gerardo Adesso. “Generic Bound Coherence under Strictly Incoherent Operations”. *Phys. Rev. Lett.* **122**, 150402 (2019).
- [8] Ludovico Lami. “Completing the grand tour of asymptotic quantum coherence manipulation”. *IEEE Transactions on Information Theory* **66**, 2165–2183 (2019).
- [9] Masahito Hayashi and Hayata Yamasaki. “Generalized quantum stein’s lemma and second law of quantum resource theories” (2024).
- [10] Ludovico Lami. “A solution of the generalised quantum stein’s lemma”. *IEEE Transactions on Information Theory* (2025).
- [11] Iman Marvian and Robert W Spekkens. “How to quantify coherence: Distinguishing speakable and unspeakable notions”. *Physical Review A* **94**, 052324 (2016).
- [12] Felix Bischof, Hermann Kampermann, and Dagmar Bruß. “Quantifying coherence with respect to general quantum measurements”. *Physical Review A* **103**, 032429 (2021).
- [13] Thomas Theurer, Nathan Killoran, Dario Egloff, and Martin B Plenio. “Resource theory of superposition”. *Physical review letters* **119**, 230401 (2017).
- [14] Johan Åberg. “Catalytic coherence”. *Physical review letters* **113**, 150402 (2014).
- [15] Ho-Joon Kim and Soojoon Lee. “Relation between quantum coherence and quantum entanglement in quantum measurements”. *Physical Review A* **106**, 022401 (2022).
- [16] Ho-Joon Kim and Soojoon Lee. “Maneuvering measurement-coherence into measurement-entanglement” (2024).
- [17] Eric Chitambar and Gilad Gour. “Quantum resource theories”. *Reviews of modern physics* **91**, 025001 (2019).
- [18] Masahito Hayashi, Kun Fang, and Kun Wang. “Finite block length analysis on quantum coherence distillation and incoherent randomness extraction”. *IEEE Transactions on Information Theory* **67**, 3926–3944 (2021).
- [19] Kamil Korzekwa, Christopher T Chubb, and Marco Tomamichel. “Avoiding irreversibility: Engineering resonant conversions of quantum resources”. *Physical Review Letters* **122**, 110403 (2019).
- [20] Fernando GSL Brandao and Gilad Gour. “Reversible framework for quantum resource theories”. *Physical review letters* **115**, 070503 (2015).
- [21] Hisaharu Umegaki. “Conditional expectation in an operator algebra, iii”. In *Kodai Mathematical Seminar Reports. Volume 11, pages 51–64*. Department of Mathematics, Tokyo Institute of Technology (1959).
- [22] Ludovico Lami and Bartosz Regula. “No second law of entanglement manipulation after all”. *Nature Physics* **19**, 184–189 (2023).
- [23] Ludovico Lami, Bartosz Regula, and Alexander Streltsov. “No-go theorem for entanglement distillation using catalysis”. *Physical Review A* **109**, L050401 (2024).

- [24] Xiao Yuan, Hongyi Zhou, Zhu Cao, and Xiongfeng Ma. “Intrinsic randomness as a measure of quantum coherence”. *Physical Review A* **92**, 022124 (2015).
- [25] Xiao Yuan, Qi Zhao, Davide Girolami, and Xiongfeng Ma. “Quantum coherence and intrinsic randomness”. *Advanced Quantum Technologies* **2**, 1900053 (2019).
- [26] Mark M Wilde. “Quantum information theory”. *Cambridge University Press*. (2013).
- [27] Eric Chitambar. “Dephasing-covariant operations enable asymptotic reversibility of quantum resources”. *Physical Review A* **97**, 050301 (2018).
- [28] Fernando GSL Brandao and Martin B Plenio. “A generalization of quantum stein’s lemma”. *Communications in Mathematical Physics* **295**, 791–828 (2010).
- [29] Mario Berta, Fernando GSL Brandão, Gilad Gour, Ludovico Lami, Martin B Plenio, Bartosz Regula, and Marco Tomamichel. “On a gap in the proof of the generalised quantum stein’s lemma and its consequences for the reversibility of quantum resources”. *Quantum* **7**, 1103 (2023).
- [30] Marco Tomamichel. “Quantum information processing with finite resources: mathematical foundations”. *Volume 5*. Springer. (2015).
- [31] Marco Tomamichel, Roger Colbeck, and Renato Renner. “A fully quantum asymptotic equipartition property”. *IEEE Transactions on information theory* **55**, 5840–5847 (2009).
- [32] Qi Zhao, Yunchao Liu, Xiao Yuan, Eric Chitambar, and Xiongfeng Ma. “One-shot coherence dilution”. *Physical review letters* **120**, 070403 (2018).
- [33] Armin Uhlmann. “Roofs and convexity”. *Entropy* **12**, 1799–1832 (2010).
- [34] Bartosz Regula. “Convex geometry of quantum resource quantification”. *Journal of Physics A: Mathematical and Theoretical* **51**, 045303 (2017).
- [35] Eric Chitambar and Min-Hsiu Hsieh. “Relating the resource theories of entanglement and quantum coherence”. *Physical review letters* **117**, 020402 (2016).
- [36] Gaurav Saxena, Eric Chitambar, and Gilad Gour. “Dynamical resource theory of quantum coherence”. *Physical Review Research* **2**, 023298 (2020).
- [37] Masaya Takahashi, Swapna Rana, and Alexander Streltsov. “Creating and destroying coherence with quantum channels”. *Physical Review A* **105**, L060401 (2022).
- [38] Francesco Buscemi. “Private quantum decoupling and secure disposal of information”. *New Journal of Physics* **11**, 123002 (2009).
- [39] Frédéric Dupuis, Mario Berta, Jürg Wullschleger, and Renato Renner. “One-shot decoupling”. *Communications in Mathematical Physics* **328**, 251–284 (2014).
- [40] Omar Fawzi and Renato Renner. “Quantum conditional mutual information and approximate markov chains”. *Communications in Mathematical Physics* **340**, 575–611 (2015).
- [41] Anurag Anshu and Rahul Jain. “Efficient methods for one-shot quantum communication”. *npj Quantum Information* **8**, 97 (2022).
- [42] Igor Devetak and Andreas Winter. “Distillation of secret key and entanglement from quantum states”. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **461**, 207–235 (2005).
- [43] Martin Ringbauer, Thomas R Bromley, Marco Cianciaruso, Ludovico Lami, WY Sarah Lau, Gerardo Adesso, Andrew G White, Alessandro Fedrizzi, and Marco Piani. “Certification and quantification of multilevel quantum coherence”. *Physical Review X* **8**, 041007 (2018).

A Useful properties of the fidelity

In this paper we adhere to the *Uhlmann–Jozsa* (i.e., square root-free) definition of the fidelity, $F(\sigma, \tau) := \left(\text{Tr} \sqrt{\sqrt{\sigma} \tau \sqrt{\sigma}} \right)^2$, which we will apply on arbitrary positive-semidefinite arguments (not only density operators). We will make use of the following properties, referring to this section

when we do so:

1. One or both arguments pure: $F(\psi, \tau) = \langle \psi | \tau | \psi \rangle$.
2. Multiplication by nonnegative scalar factors: $F(t\sigma, \tau) = tF(\sigma, \tau)$ for $t \geq 0$.
3. Joint concavity of the square-root fidelity:

for $0 \leq p \leq 1$,

$$\begin{aligned} & \sqrt{F[p\sigma_1 + (1-p)\sigma_2, p\tau_1 + (1-p)\tau_2]} \\ & \geq p\sqrt{F(\sigma_1, \tau_1)} + (1-p)\sqrt{F(\sigma_2, \tau_2)}. \end{aligned} \quad (92)$$

4. Fidelity of an operator with a normalized projection thereof: for some space $\mathcal{V}$, if $\sigma^\mathcal{V} := \mathbb{1}_\mathcal{V}\sigma\mathbb{1}_\mathcal{V}$ and $\sigma^{|\mathcal{V}} := \sigma^\mathcal{V}/\text{Tr}\sigma^\mathcal{V}$, then

$$\begin{aligned} F(\sigma, \sigma^{|\mathcal{V}}) &= F(\sigma^\mathcal{V}, \sigma^{|\mathcal{V}}) \\ &= (\text{Tr}\sigma^\mathcal{V}) F(\sigma^{|\mathcal{V}}, \sigma^{|\mathcal{V}}) \\ &= \text{Tr}\sigma^\mathcal{V}. \end{aligned} \quad (93)$$

5. Fidelity of a density operator with a convex subcomponent thereof: if $\tau = p\tau_1 + (1-p)\tau_2$ is a density operator with $0 \leq p \leq 1$ and $\tau_{1/2}$ also density operators, then by joint concavity,

$$\begin{aligned} F(\tau, \tau_1) &= F[p\tau_1 + (1-p)\tau_2, p\tau_1 + (1-p)\tau_1] \\ &\geq \left[p\sqrt{F(\tau_1, \tau_1)} + (1-p)\sqrt{F(\tau_2, \tau_1)} \right]^2 \\ &\geq [p \cdot 1 + (1-p) \cdot 0]^2 = p^2. \end{aligned} \quad (94)$$

B Useful results from the literature

In our work we will make use of the following results from [5], namely the asymptotic continuity of the relative entropy of coherence C_r and the coherence of formation C_f .

Lemma B.1 ([5, Suppl. Material, Lemma 12]). *For two states τ_1^A and τ_2^A with $\|\tau_1 - \tau_2\|_1 \leq \delta$,*

$$|C_r(\tau_1) - C_r(\tau_2)| \leq \delta \log_2 A + 2h\left(\frac{\delta}{2}\right), \quad (95)$$

where $h(t) = -t \log_2 t - (1-t) \log_2(1-t)$ is the binary entropy function.

Lemma B.2 ([5, Suppl. Material, Lemma 15]). *For two states τ_1^A and τ_2^A with $B(\tau_1, \tau_2) \leq \delta$,*

$$|C_f(\tau_1) - C_f(\tau_2)| \leq \delta \log_2 A + (1+\delta)h\left(\frac{\delta}{1+\delta}\right). \quad (96)$$

C Asymptotic typicality

Here we give a brief self-contained review of the necessary background on asymptotic typicality for i.i.d. classical and quantum samples. For a detailed treatment, see [26].

C.1 Classical sources and typical sets

We will denote classical variables with uppercase letters and specific values they take with lowercase. Suppose a classical i.i.d. stochastic source outputs a variable X taking values in a finite alphabet $\mathcal{X}$ and distributed according to $\mathbf{p} \equiv [p(x)]_{x \in \mathcal{X}}$. Consider a length- n sample $\mathbf{X} \equiv X_1 X_2 \dots X_n$ from the source; it takes values of length- n sequences $\mathbf{x} \equiv x_1 x_2 \dots x_n \in \mathcal{X}^n$. By the source's i.i.d. property, the sample distribution is $\mathbf{p}_n(\mathbf{X}) = \mathbf{p}^{\otimes n}$. Recall that $H(\mathbf{p}) \equiv H(X)_\mathbf{p} = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$ denotes the Shannon entropy of the source.

Definition C.1 (Typical sequences and sets). For any real $\delta > 0$, a δ -weakly-typical (or entropy-typical) sequence $\mathbf{x}$ is one that satisfies

$$2^{-n[H(\mathbf{p})+\delta]} \leq p_n(\mathbf{x}) \leq 2^{-n[H(\mathbf{p})-\delta]}. \quad (97)$$

For any $x \in \mathcal{X}$, let $f_x(\mathbf{x})$ denote the frequency of occurrences of x in $\mathbf{x}$, i.e. $f_x(\mathbf{x}) = |\{j : x_j = x\}|/n$. A δ -strongly-typical (or letter-typical) sequence $\mathbf{x}$ is one that satisfies

$$\sum_{x \in \mathcal{X}} |f_x(\mathbf{x}) - p(x)| \leq \delta. \quad (98)$$

The set $\bar{\mathcal{T}}_n^\delta$ of all δ -weakly-typical sequences is called the δ -weakly-typical set, and the set $\mathcal{T}_n^\delta$ of all δ -strongly-typical sequences the δ -strongly-typical set.

Lemma C.2 (Asymp. equipartition property [AEP]). *For any $\delta > 0$ and large enough n , the weakly-typical set $\bar{\mathcal{T}}_n^\delta$ has the following properties:*

1. $\sum_{\mathbf{x} \in \bar{\mathcal{T}}_n^\delta} p_n(\mathbf{x}) \geq 1 - \delta$;
2. $(1 - \delta)2^{n[H(\mathbf{p})-\delta]} \leq |\bar{\mathcal{T}}_n^\delta| \leq 2^{n[H(\mathbf{p})+\delta]}$.

Furthermore, there exists a continuous real function $\eta(\delta)$, such that $\eta(\delta) \xrightarrow{\delta \rightarrow 0} 0$ and for large enough n , the strongly-typical set $\mathcal{T}_n^\delta$ has the following properties:

1. $2^{-n[H(\mathbf{p})+\eta(\delta)]} \leq p_n(\mathbf{x}) \leq 2^{-n[H(\mathbf{p})-\eta(\delta)]}$ for all $\mathbf{x} \in \mathcal{T}_n^\delta$. In other words, $\mathcal{T}_n^\delta \subset \bar{\mathcal{T}}_n^{\eta(\delta)}$;
2. $\sum_{\mathbf{x} \in \mathcal{T}_n^\delta} p_n(\mathbf{x}) \geq 1 - \delta$;
3. $(1 - \delta)2^{n[H(\mathbf{p})-\eta(\delta)]} \leq |\mathcal{T}_n^\delta| \leq 2^{n[H(\mathbf{p})+\eta(\delta)]}$.

Intuitively, we can understand AEP as follows: for a large enough sample from an i.i.d. source, the weakly- or strongly-typical set for any $\delta > 0$ supports all but δ of the probability. The cardinality of this set scales exponentially, roughly as $\exp_2[nH(\mathbf{p})]$; note that this is an exponentially small fraction of the cardinality $|\mathcal{X}|^n$ of all n -length sequences! Finally, the distribution of sequences within this set is nearly uniform.

C.2 Quantum sources and typical subspaces

As in the classical case, there is also a quantum AEP. Suppose a quantum source outputs i.i.d. copies of an elementary *finite-dimensional* system A , each prepared in the state ρ^A . Considering a length- n sample $\varrho_n \equiv \rho^{\otimes n}$, the quantum counterparts to the classical sequences $\mathbf{x}$ are directions in the Hilbert space of $A_n \equiv A^{\otimes n}$, and the quantum counterparts to sequence probabilities are the directional densities induced by ϱ_n . In particular, the eigenvectors of ϱ_n , and their associated eigenvalues, embody an informationally-complete description of its structure. If $\rho = \sum_{x \in [A]} r_x \psi_x$ is an eigendecomposition, then $|\Psi_{\mathbf{x}}\rangle := \bigotimes_{k=1}^n |\psi_{x_k}\rangle$, for $\mathbf{x} \equiv (x_k)_{k=1}^n \in [A]^n$, constitute a complete basis of eigenvectors of ϱ_n , with associated eigenvalues $R_{\mathbf{x}} := \prod_k r_{x_k}$. Based on these, we can define quantum counterparts of typical sequences and sets:

Definition C.3 (Typical eigenvectors and subspaces). For any real $\delta > 0$, a δ -weakly- (respectively, strongly-) typical eigenvector $|\Psi_{\mathbf{x}}\rangle$ is one whose associated label sequence $\mathbf{x}$ is δ -weakly- (resp. strongly-) typical under the distribution $\mathbf{r}^{\otimes n}$. The δ -weakly- (resp. strongly-) typical subspace $\tilde{\mathcal{V}}_n^\delta$ (resp. $\mathcal{V}_n^\delta$) is the span of the δ -weakly- (resp. strongly-) typical eigenvectors. Note that these subspaces do not depend on the choice of eigenbasis $\{|\psi_x\rangle\}_x$.

Lemma C.4 (Quantum AEP). *For any $\delta > 0$ and large enough n , the weakly-typical subspace $\tilde{\mathcal{V}}_n^\delta$ has the following properties:*

1. $\text{Tr}(\mathbb{1}_{\tilde{\mathcal{V}}_n^\delta} \varrho_n) \geq 1 - \delta$;
2. $(1 - \delta)2^{n[S(\rho) - \delta]} \leq \dim \tilde{\mathcal{V}}_n^\delta \leq 2^{n[S(\rho) + \delta]}$.

Again, there exists a continuous real-valued function $\eta(\delta)$, such that $\eta(\delta) \xrightarrow{\delta \rightarrow 0} 0$ and for large enough n , the strongly-typical subspace $\mathcal{V}_n^\delta$ has the following properties:

1. $2^{-n[S(\rho) + \eta(\delta)]} \leq \mathbb{1}_{\mathcal{V}_n^\delta} \varrho_n \mathbb{1}_{\mathcal{V}_n^\delta} \leq 2^{-n[S(\rho) - \eta(\delta)]}$;
2. $\text{Tr}(\mathbb{1}_{\mathcal{V}_n^\delta} \varrho_n) \geq 1 - \delta$;
3. $(1 - \delta)2^{n[S(\rho) - \eta(\delta)]} \leq \dim \mathcal{V}_n^\delta \leq 2^{n[S(\rho) + \eta(\delta)]}$.

D Proof sketch for Conjecture 2

We now present a detailed attempt at proving Conjecture 2 using the construction summarized in Section 2. Note that the ϵ 's and δ 's below are not the same as those in the proof sketch for Conjecture 1.

Proof sketch for Conjecture 2. We will now try to construct a sequence of viable IO channels based on the one from Winter and Yang's maximal distillation protocol [5]. Each of Winter–Yang's channels¹³ $\tilde{\mathcal{E}}_n$ has the following special property: it can be decomposed into IO Kraus operators $\tilde{K}_s$ that take the form

$$\tilde{K}_s = \sum_m |m\rangle \langle \tilde{w}_{s,m}| = \sqrt{M_n} \sum_m |m\rangle \langle \tilde{w}_s| \mathbb{1}_{\mathcal{I}_m}, \quad (99)$$

where $\{\mathcal{I}_m\}_m$ is a partitioning of $\mathcal{A}_n \equiv \mathcal{A}^n$ into $M_n = \exp_2(n[C_r(\rho) - \tilde{\epsilon}_n])$ disjoint subalphabets, and $\mathbb{1}_{\mathcal{I}_m}$ denotes the projector onto $\text{span}\{|\mathbf{a}\rangle : \mathbf{a} \in \mathcal{I}_m\}$. Importantly, the fragments $|\tilde{w}_{s,m}\rangle$ for a given m all lie within the subspace corresponding to $\mathcal{I}_m$, independent of s . Consider the target effect $\tilde{T}_n = \sum_s |\tilde{w}_s\rangle \langle \tilde{w}_s|$. The particular Kraus operators in Winter–Yang's construction happen to involve $|w_c\rangle \equiv |\tilde{w}_s\rangle$ that form a basis of ϱ_n 's eigenvectors. But thanks to the above-noted s -independent block structure, *any* convex decomposition of $\tilde{T}_n$ into arbitrary pure components $|w_c\rangle$ can as well be used to construct a set of IO Kraus operators (implementing the same $\tilde{\mathcal{E}}_n$), since $|w_{c,m}\rangle := \sqrt{M_n} \mathbb{1}_{\mathcal{I}_m} |w_c\rangle \in \text{span}\{|\mathbf{a}\rangle : \mathbf{a} \in \mathcal{I}_m\}$. More generally, we can even use some other T_n sharing $\tilde{T}_n$'s essential properties and construct IO Kraus operators from arbitrary pure decompositions thereof:

Observation D.1. *Suppose a sequence T_n of effects satisfies*

1. $M_n \mathbb{1}_{\mathcal{I}_m} T_n \mathbb{1}_{\mathcal{I}_m} \leq \mathbb{1}_{\mathcal{I}_m}$ for every value of m indexing a partitioning $\{\mathcal{I}_m\}_m$ of $\mathcal{A}_n$ into

¹³ Where there is scope for confusion, we use tildes to denote objects associated with Winter–Yang's construction.

$M_n = \exp_2(n[C_r(\rho) - o(1)])$ disjoint subalphabets;

2. $\text{Tr}(T_n \varrho_n) \rightarrow 1$ as $n \rightarrow \infty$.

Such a sequence can be used to construct a sequence of IO channels distilling asymptotically maximally from ϱ_n .

Proof. If $T_n = \sum_c |w_c\rangle \langle w_c|$ is a decomposition into pure components, we define the Kraus operators

$$K_c := \sum_m |m\rangle \langle w_{c,m}| := \sqrt{M_n} \sum_m |m\rangle \langle w_c| \mathbb{1}_{\mathcal{I}_m} \quad (100)$$

and the maps $\mathcal{F}_n(\cdot) := \sum_c K_c(\cdot) K_c^\dagger$. Note that

$$\sum_c K_c^\dagger K_c = M_n \sum_m \mathbb{1}_{\mathcal{I}_m} T_n \mathbb{1}_{\mathcal{I}_m} =: \mathbf{T}_n. \quad (101)$$

Therefore, as long as condition 1 is satisfied, we can use the construction of (100) on $T_n^\lambda := M_n^{-1}(\mathbb{1}^{A_n} - \mathbf{T}_n)$ to obtain IO subchannels $\mathcal{G}_n$ such that $\mathcal{E}_n := \mathcal{F}_n + \mathcal{G}_n$ are channels. With these conditions met, the resulting $\mathcal{E}_n$ are maximally-distilling channels if condition 2 is met, since the T_n are (by construction) target witnesses of the subchannels $\mathcal{F}_n$ with respect to the maximally-scaling targets Ψ_{M_n} . $\square$

This concludes our formalization of point 1 in Section 5.2. In our proof sketch for Conjecture 1, we tried to show that the target effect-based density operators τ_n satisfy $F(\tau_n, \varrho_n) \rightarrow 1$ asymptotically. One way of trying to construct bound-attaining distillation channels is to construct a sequence T_n satisfying the conditions of Observation D.1 starting from $\tau_n = \varrho_n$ itself. Since a necessity for condition 2 is $\text{Tr} T_n \geq 2^{n[S(\rho) - o(1)]}$, we can aim to satisfy this weaker condition through a ϱ_n -based construction.

As noted in point 2, we can choose any optimal decomposition of ϱ_n attaining $C_f(\varrho_n)$ [possibly after some pruning to uphold condition 1] to saturate the measurement coherence bound on average—but we can try to do better.

As summarized in point 3, we will flatten relative to c by decomposing ϱ_n in a special way. Let $\rho = \sum_j q_j \phi_j$ be an optimal decomposition attaining $C_f(\rho)$; such a decomposition exists, by virtue of the finitude of A . Since $\varrho_n = \left(\sum_j q_j \phi_j\right)^{\otimes n}$, it can of course be decomposed convexly into pure states of the form $|\Phi_{c \equiv \mathbf{j}}\rangle \equiv \bigotimes_{k=1}^n |\phi_{j_k}\rangle$, where

$\mathbf{j} \equiv (j_k)_k$. If we then take the strongly- δ_J -typical set $\mathcal{T}_n^{\delta_J}$ under the distribution $\mathbf{Q}_n \equiv \mathbf{q}^{\otimes n}$ (see Definition C.1 and Lemma C.2), the letter frequencies in any $\mathbf{j} \in \mathcal{T}_n^{\delta_J}$ satisfy $|f_j(\mathbf{j}) - q_j| \leq \delta_J$. By the additivity of the relative entropy of coherence under tensor products,

$$\begin{aligned} C_r(\Phi_{\mathbf{j}}) &= \sum_j [f_j(\mathbf{j}) n] C_r(\phi_j) \\ &= n \sum_j [q_j + O(\delta_J)] C_r(\phi_j) \\ &= n [C_f(\rho) + O(\delta_J)] \end{aligned} \quad (102)$$

for any of these $\mathbf{j}$. The part of ϱ_n composed thereof is $\varrho_n^{\delta_J} \equiv \sum_{\mathbf{j} \in \mathcal{T}_n^{\delta_J}} Q_n(\mathbf{j}) \Phi_{\mathbf{j}}$. By the AEP, it has weight

$$\text{Tr} \varrho_n^{\delta_J} = Q_n^{\delta_J} \geq 1 - \delta_J, \quad (103)$$

where $Q_n^{\delta_J} := \sum_{\mathbf{j} \in \mathcal{T}_n^{\delta_J}} Q_n(\mathbf{j})$. As before, letting $\varrho_n^{|\delta_J|}$ denote the normalized version thereof, the behaviour of the fidelity under convex mixtures (point 5 of Appendix A) yields

$$F(\varrho_n, \varrho_n^{|\delta_J|}) \geq \left(Q_n^{\delta_J}\right)^2 \geq 1 - 2\delta_J. \quad (104)$$

This concludes point 3. We will now subject the remaining $|\Phi_{\mathbf{j}}\rangle$ to a minor modification to ensure that our final construction can achieve $\text{Tr} T_n \approx \exp_2[nS(\rho)]$. Let $\mathcal{V}^{\delta_S}$ be ϱ_n 's δ_S -strongly-typical subspace. Define $|\Phi_{\mathbf{j}}^{\delta_S}\rangle := \mathbb{1}_{\mathcal{V}^{\delta_S}} |\Phi_{\mathbf{j}}\rangle$ and

$$\varrho_n^{\delta_J, \delta_S} := \sum_{\mathbf{j} \in \mathcal{T}_n^{\delta_J}} Q_n(\mathbf{j}) \Phi_{\mathbf{j}}^{\delta_S}. \quad (105)$$

By the quantum AEP (Lemma C.4),

$$2^{-n[S(\rho) + \eta(\delta_S)]} \mathbb{1} \leq \varrho_n^{\delta_S} \leq 2^{-n[S(\rho) - \eta(\delta_S)]} \mathbb{1}, \quad (106)$$

where $\eta(\delta) \xrightarrow{\delta \rightarrow 0} 0$. Noting that $\varrho_n^{\delta_S} = \varrho_n^{\delta_J, \delta_S} + \varrho_n^{\setminus \delta_J, \delta_S}$,

$$\varrho_n^{\delta_J, \delta_S} \leq \varrho_n^{\delta_S} \leq 2^{-n[S(\rho) - \eta(\delta_S)]}. \quad (107)$$

Since the $\mathbf{j}$'s we have retained are all strongly-typical, the further restriction to a strongly-typical subspace has a negligible effect on the distribution of incoherent label strings $\mathbf{a}$ conditioned on $\mathbf{j}$. To see this, let $\rho = \sum_i r_i \psi_i$ be an eigendecomposition, and suppose $|\phi_j\rangle = \sum_i \chi_{ji} |\psi_i\rangle$. Then, $\sum_j q_j |\chi_{ji}|^2 = r_i$, and therefore every $\mathbf{i}$ sequence that is strongly-typical *conditional* on $\mathbf{j}$ is also part of the *unconditional*

strongly-typical set that determines the typical subspace. As such, the distribution of $\mathbf{a}$ conditioned on $\mathbf{j}$ is unaffected except for a δ_S loss of total measure. We will therefore consider the distribution of $\mathbf{a}$ in $|\Phi_{\mathbf{j}}\rangle$.

Each $|\Phi_{\mathbf{j}}\rangle$ can be made arbitrarily close to a uniform superposition in the incoherent $|\mathbf{a}\rangle$ basis. Let $|\phi_j\rangle = \sum_a \sqrt{\xi_{a|j}} e^{i\varphi_{j,a}} |a\rangle$ for $\xi_{|j} \equiv (\xi_{a|j})_a$ a distribution and $\varphi_{j,a} \in \mathbb{R}$; note that $H(\xi_{|j}) = C_r(\phi_j)$. Then,

$$|\phi_j\rangle^{\otimes n_j} = \left(\sum_a \sqrt{\xi_{a|j}} e^{i\varphi_{j,a}} |a\rangle \right)^{\otimes n_j}. \quad (108)$$

We now apply (weak) asymptotic typicality on the sequences $\mathbf{a}_j \equiv (a_k)_{k=1}^{n_j}$ (where the subscript j in $\mathbf{a}_j$ signifies that the latter takes values in $\mathcal{A}_{n_j}$ and not $\mathcal{A}_n$) under the distribution $\xi_{|j}^{\otimes n_j}$. The number of δ_A -weakly-typical sequences is $\leq 2^{n_j[C_r(\phi_j) + \delta_A]}$, and they collectively garner an amplitude $\geq \sqrt{1 - \delta_A}$; this holds alike for all j . Note that the strongly-typical $|\Phi_{\mathbf{j}}\rangle$ described above can be obtained by applying subsystem permutations (which are incoherent unitaries) on $\bigotimes_j |\phi_j\rangle^{\otimes n_j \equiv [q_j + O(\delta_J)]^n}$. Therefore, if $|\Phi_{\mathbf{j}}^{\delta_A}\rangle$ denote $|\Phi_{\mathbf{j}}\rangle$ projected onto the span of $\mathbf{a} \in \mathcal{A}_n$ simultaneously δ_A -typicalized under all the $\xi_{|j}^{\otimes n_j}$,

$$r_C(\Phi_{\mathbf{j}}^{\delta_A}) \leq 2^{n[C_f(\rho) + \delta_A + O(\delta_J)]}; \quad (109)$$

$$|\langle \Phi_{\mathbf{j}}^{\delta_A} | \Phi_{\mathbf{j}} \rangle| \geq \sqrt{(1 - \delta_A)^J} = 1 - O(\delta_A), \quad (110)$$

where $\Phi_{\mathbf{j}}^{\delta_A}$ denotes the normalized $\Phi_{\mathbf{j}}^{\delta_A}$ and J is the number of components in the optimal decomposition $\rho = \sum_j q_j \phi_j$. Applying the same arguments to the subspace-typicalized $|\Phi_{\mathbf{j}}^{\delta_S, \delta_A}\rangle$,

$$r_C(\Phi_{\mathbf{j}}^{\delta_S, \delta_A}) \leq 2^{n[C_f(\rho) + \delta_A + O(\delta_J) + O(\delta_S)]}; \quad (111)$$

$$|\langle \Phi_{\mathbf{j}}^{\delta_S, \delta_A} | \Phi_{\mathbf{j}} \rangle| \geq 1 - O(\delta_S) - O(\delta_A). \quad (112)$$

Applying Lemma B.1 (the asymptotic continuity of the relative entropy of coherence [5]), we also have

$$\begin{aligned} C_r(\Phi_{\mathbf{j}}^{\delta_S, \delta_A}) \\ = n[C_f(\rho) + O(\delta_J) + O(\delta_S) + O(\delta_A)]. \end{aligned} \quad (113)$$

Now define $\varrho_n^{\delta_J, \delta_S, \delta_A} := \sum_{\mathbf{j} \in \mathcal{T}_n^{\delta_J}} Q_n(\mathbf{j}) \Phi_{\mathbf{j}}^{\delta_S, \delta_A}$ and let $\varrho_n^{\delta_J, \delta_S, \delta_A}$ denote its normalized version. The

joint concavity of the square-root fidelity implies

$$\begin{aligned} F(\varrho_n^{\delta_J}, \varrho_n^{\delta_J, \delta_S, \delta_A}) \\ \geq \left(\sum_{\mathbf{j} \in \mathcal{T}_n^{\delta_J}} \frac{Q_n(\mathbf{j})}{Q_n^{\delta_J}} |\langle \Phi_{\mathbf{j}}^{\delta_S, \delta_A} | \Phi_{\mathbf{j}} \rangle| \right)^2 \\ \geq 1 - O(\delta_S) - O(\delta_A), \end{aligned} \quad (114)$$

whence

$$F(\varrho_n, \varrho_n^{\delta_J, \delta_S, \delta_A}) \geq 1 - (2\delta_J \boxplus [O(\delta_S) + O(\delta_A)]). \quad (115)$$

For convenience, let $Q_n^{\delta_S, \delta_A}(\mathbf{j})$ denote the appropriate normalized measure such that

$$\varrho_n^{\delta_J, \delta_S, \delta_A} = \sum_{\mathbf{j} \in \mathcal{T}_n^{\delta_J}} Q_n^{\delta_S, \delta_A}(\mathbf{j}) \Phi_{\mathbf{j}}^{\delta_S, \delta_A}. \quad (116)$$

This $\varrho_n^{\delta_J, \delta_S, \delta_A}$ can be decomposed into the pure components $\Phi_{\mathbf{j}}^{\delta_S, \delta_A}$, each of whose coherence rank $\leq \exp_2(n[C_f(\rho) + \delta_A + O(\delta_J) + O(\delta_S)])$, thus accomplishing point 5 (point 4 is achieved by the $\mathcal{V}^{\delta_S}$ projection).

Next, as anticipated in point 6, we will show that the logarithm of the remaining coherence rank within each $\mathcal{I}_m$ is tightly concentrated around the value $n\ell(\rho)$. To this end, for any given $\mathbf{j} \in \mathcal{T}_n^{\delta_J}$, define

$$\mathcal{M}_n^{\mathbf{j}, \delta_R} := \left\{ m : r_C(\Phi_{\mathbf{j}, m}^{\delta_S, \delta_A}) \leq 2^{n[\ell(\rho) + \tilde{\epsilon}_n + \delta_R]} \right\}, \quad (117)$$

where $\Phi_{\mathbf{j}, m}^{\delta_S, \delta_A} := \mathbb{1}_{\mathcal{I}_m} \Phi_{\mathbf{j}, m}^{\delta_S, \delta_A} \mathbb{1}_{\mathcal{I}_m} / \text{Tr}(\mathbb{1}_{\mathcal{I}_m} \Phi_{\mathbf{j}, m}^{\delta_S, \delta_A})$ and $\delta_R > 0$. If $\mu_n^{\mathbf{j}, \delta_R} := |\mathcal{M}_n^{\mathbf{j}, \delta_R}| / M_n$,

$$\begin{aligned} 2^{n[C_f(\rho) + \delta_A + O(\delta_J) + O(\delta_S)]} \\ \geq r_C(\Phi_{\mathbf{j}}^{\delta_S, \delta_A}) \\ = \sum_m r_C(\Phi_{\mathbf{j}, m}^{\delta_S, \delta_A}) \\ \geq (1 - \mu_n^{\mathbf{j}, \delta_R}) M_n 2^{n[\ell(\rho) + \tilde{\epsilon}_n + \delta_R]}. \end{aligned} \quad (118)$$

Thus, if we choose $\delta_R \gg \delta_A + O(\delta_J) + O(\delta_S)$ in the exponent on the left,

$$1 - \mu_n^{\mathbf{j}, \delta_R} \lesssim \exp_2(-n\delta_R), \quad (119)$$

where we used the fact that $M_n 2^{n[\ell(\rho) + \tilde{\epsilon}_n]} = 2^{nC_f(\rho)}$. We have thus shown that the fraction of m values whose logarithmic coherence rank exceeds $n\ell(\rho)$ decays exponentially. However, this still does not rule out the possibility that such a

small fraction of m 's actually contribute most of the norm (and of the coherence) of $\Phi_j^{|\delta_S, \delta_A\rangle}$ —for, the actual measure over m is not the uniform distribution (which we used above) but the one induced by the ranks $r_C(\Phi_{j,m}^{|\delta_S, \delta_A\rangle})$ themselves. Intuitively, one expects that a measure concentration around a small fraction of m 's is inconsistent with the fact that this operator approximates ϱ_n , which in turn is known to be maximally-distillable under this block structure; we will now formalize this intuition.

Here we will exploit another special property of the Winter–Yang construction: each $\tilde{\mathcal{E}}_n$ can be dilated unitarily within $\mathcal{H}^{A_n}$, without appending any auxiliary system:

$$\tilde{\mathcal{E}}_n^{A_n \rightarrow M_n}(\varrho_n) = \text{Tr}^{S_n} \left[\mathcal{U}^{A_n \rightarrow S_n M_n}(\varrho_n) \right], \quad (120)$$

where S_n and M_n are suitable subsystems of A_n and $\mathcal{U}_n(\cdot) \equiv U_n(\cdot)U_n^\dagger$ for some unitary U_n . To be sure, the unitary is preceded by a type measurement and a “good vs. bad m ” measurement, but both of these have asymptotically-deterministic outcomes. Since the measurements are block-incoherent, the post-measurement unitary can be expanded to a fully-unitary protocol that subsumes the measurements, by controlling on the incoherent basis and defining an arbitrary unitary action (e.g., the identity) on the outlying outcomes.

Since $\tilde{\mathcal{E}}_n$ are valid maximal distillation channels, there is an asymptotically-vanishing sequence $\tilde{\epsilon}_n^{(1)}$ such that

$$\begin{aligned} F[\tilde{\mathcal{E}}_n(\varrho_n), \Psi_{M_n}] &\geq 1 - \tilde{\epsilon}_n^{(1)} \\ \Rightarrow \langle \Psi_{M_n} | \tilde{\mathcal{E}}_n(\varrho_n) | \Psi_{M_n} \rangle &\geq 1 - \tilde{\epsilon}_n^{(1)} \\ \Rightarrow \text{Tr} \left[(\mathbb{1}^{S_n} \otimes \Psi_{M_n}^{M_n}) \mathcal{U}_n(\varrho_n) \right] &\geq 1 - \tilde{\epsilon}_n^{(1)}. \end{aligned} \quad (121)$$

Let

$$\begin{aligned} \mathcal{V}' &:= (\mathbb{1}^{S_n} \otimes \Psi_{M_n}^{M_n}) [\text{supp } \mathcal{U}_n(\varrho_n)] \\ &=: \mathcal{W}^{S_n} \otimes |\Psi_{M_n}\rangle^{M_n}; \end{aligned} \quad (122)$$

note that $\mathcal{V}'$ and $\mathcal{W}$ are vector spaces by construction. For any $|w\rangle \in \mathcal{W}$,

$$|v\rangle := U_n^\dagger(|w\rangle \otimes |\Psi_{M_n}\rangle) = M_n^{-1/2} \sum_m |v_m\rangle \quad (123)$$

with $|v_m\rangle := U_n^\dagger(|w\rangle \otimes |m\rangle)$, whereby $\langle v_m | v_m \rangle = \langle v | v \rangle$ for all m . Moreover, since U_n induces the block structure discussed in point 1, $|v_m\rangle \in \text{span}\{|\mathbf{a}\rangle : \mathbf{a} \in \mathcal{I}_m\}$. The span of all such $|v\rangle$'s is $\mathcal{V} := U_n^\dagger \mathcal{V}'$. By unitarity, (121) implies

$$\text{Tr}(\mathbb{1}_{\mathcal{V}} \varrho_n) \geq 1 - \tilde{\epsilon}_n^{(1)}. \quad (124)$$

Combining this with (115), we have

$$\begin{aligned} \text{Tr}(\mathbb{1}_{\mathcal{V}} \varrho_n^{|\delta_S, \delta_A\rangle}) &\geq 1 - \epsilon_n^{(2)} \Rightarrow \\ \sum_{\mathbf{j} \in \mathcal{T}_n^{\delta_J}} Q_n^{\delta_S, \delta_A}(\mathbf{j}) \langle \Phi_{\mathbf{j}}^{|\delta_S, \delta_A; \mathcal{V}} | \Phi_{\mathbf{j}}^{|\delta_S, \delta_A; \mathcal{V}} \rangle &\geq 1 - \epsilon_n^{(2)}, \end{aligned} \quad (125)$$

where $|\Phi_{\mathbf{j}}^{|\delta_S, \delta_A; \mathcal{V}}\rangle := \mathbb{1}_{\mathcal{V}} |\Phi_{\mathbf{j}}^{|\delta_S, \delta_A}\rangle$ (unnormalized).

For any $\epsilon \in (\epsilon_n^{(2)}, 1]$, let

$$\begin{aligned} \mathcal{T}_n^{\delta_J, \epsilon} &:= \\ \{ \mathbf{j} \in \mathcal{T}_n^{\delta_J} : \langle \Phi_{\mathbf{j}}^{|\delta_S, \delta_A; \mathcal{V}} | \Phi_{\mathbf{j}}^{|\delta_S, \delta_A; \mathcal{V}} \rangle &\geq 1 - \epsilon \} \end{aligned} \quad (126)$$

and $Q_n^{\delta_J, \epsilon} := \sum_{\mathbf{j} \in \mathcal{T}_n^{\delta_J, \epsilon}} Q_n^{\delta_S, \delta_A}(\mathbf{j})$. Then,

$$\begin{aligned} Q_n^{\delta_J, \epsilon} \cdot 1 + (1 - Q_n^{\delta_J, \epsilon})(1 - \epsilon) &\geq 1 - \epsilon_n^{(2)} \\ \Rightarrow Q_n^{\delta_J, \epsilon} &\geq 1 - \frac{\epsilon_n^{(2)}}{\epsilon}. \end{aligned} \quad (127)$$

In particular, choosing $\epsilon \equiv \epsilon_n^{(3)} := \sqrt{\epsilon_n^{(2)}}$,

$$Q_n^{\delta_J, \epsilon^{(3)}} \geq 1 - \epsilon_n^{(3)}. \quad (128)$$

Consequently,

$$\text{Tr} \left[\sum_{\mathbf{j} \in \mathcal{T}_n^{\delta_J, \epsilon_n^{(3)}}} Q_n^{\delta_S, \delta_A}(\mathbf{j}) \Phi_{\mathbf{j}}^{|\delta_S, \delta_A; \mathcal{V}} \right] \geq Q_n^{\delta_J, \epsilon_n^{(3)}} \min_{\mathbf{j} \in \mathcal{T}_n^{\delta_J, \epsilon_n^{(3)}}} \langle \Phi_{\mathbf{j}}^{|\delta_S, \delta_A; \mathcal{V}} | \Phi_{\mathbf{j}}^{|\delta_S, \delta_A; \mathcal{V}} \rangle \geq [1 - \epsilon_n^{(3)}]^2 \geq 1 - 2\epsilon_n^{(3)}. \quad (129)$$

Henceforth we shall only consider $\mathbf{j} \in \mathcal{T}_n^{\delta_J, \epsilon_n^{(3)}}$. Define

$$\varrho_n^{|\delta_J, \delta_S, \delta_A, \epsilon_n^{(3)}|} := \frac{1}{Q_n^{\delta_J, \epsilon}} \sum_{\mathbf{j} \in \mathcal{T}_n^{\delta_J, \epsilon_n^{(3)}}} Q_n^{\delta_S, \delta_A}(\mathbf{j}) \Phi_{\mathbf{j}}^{|\delta_S, \delta_A|}. \quad (130)$$

Note that we have not projected the vectors onto $\mathcal{V}$ in this definition: we have merely restricted the values of $\mathbf{j}$ further. Due to (128),

$$F\left(\varrho_n^{|\delta_J, \delta_S, \delta_A, \epsilon_n^{(3)}|}, \varrho_n^{|\delta_J, \delta_S, \delta_A, \epsilon_n^{(3)}|}\right) \geq 1 - 2\epsilon_n^{(3)}, \quad (131)$$

which implies via (115) that

$$F\left(\varrho_n, \varrho_n^{|\delta_J, \delta_S, \delta_A, \epsilon_n^{(3)}|}\right) \geq 1 - \epsilon_n^{(4)}. \quad (132)$$

Now define $|\Phi_{\mathbf{j}}^{|\delta_S, \delta_A, \mathcal{V}}|\rangle$ by normalizing $|\Phi_{\mathbf{j}}^{|\delta_S, \delta_A|}\rangle$. By construction, for each remaining $\mathbf{j}$,

$$\begin{aligned} & \left| \langle \Phi_{\mathbf{j}}^{|\delta_S, \delta_A|} | \Phi_{\mathbf{j}}^{|\delta_S, \delta_A, \mathcal{V}} \rangle \right|^2 \geq 1 - \epsilon_n^{(3)} \\ \Rightarrow \sum_m & \left| \langle \Phi_{\mathbf{j}}^{|\delta_S, \delta_A|} | \mathbb{1}_{\mathcal{I}_m} | \Phi_{\mathbf{j}}^{|\delta_S, \delta_A, \mathcal{V}} \rangle \right|^2 \geq 1 - \epsilon_n^{(3)}. \end{aligned} \quad (133)$$

Since each $|\Phi_{\mathbf{j}}^{|\delta_S, \delta_A, \mathcal{V}}|\rangle$ is a unit vector in $\mathcal{V}$, its norm within each $\mathcal{I}_m$ is exactly M_n^{-1} [see (123)]:

$$\langle \Phi_{\mathbf{j}}^{|\delta_S, \delta_A, \mathcal{V}} | \mathbb{1}_{\mathcal{I}_m} | \Phi_{\mathbf{j}}^{|\delta_S, \delta_A, \mathcal{V}} \rangle = M_n^{-1}. \quad (134)$$

If $\alpha_m := \sqrt{\langle \Phi_{\mathbf{j}}^{|\delta_S, \delta_A|} | \mathbb{1}_{\mathcal{I}_m} | \Phi_{\mathbf{j}}^{|\delta_S, \delta_A|} \rangle}$ (for brevity, we suppress its dependency on $\mathbf{j}$ and other parameters), the Cauchy–Schwartz inequality implies

$$\left| \langle \Phi_{\mathbf{j}}^{|\delta_S, \delta_A|} | \mathbb{1}_{\mathcal{I}_m} | \Phi_{\mathbf{j}}^{|\delta_S, \delta_A, \mathcal{V}} \rangle \right|^2 \leq \frac{|\alpha_m|^2}{M_n}. \quad (135)$$

Thus, defining $|\alpha\rangle := \sum_m \alpha_m |m\rangle$, (133) begets

$$|\langle \alpha | \Psi_{M_n} \rangle|^2 \geq 1 - \epsilon_n^{(3)}. \quad (136)$$

Again applying Lemma B.1,

$$\begin{aligned} C_r(|\alpha\rangle \langle \alpha|) &= C_r(\Psi_{M_n}) + n\epsilon_n^{(5)} \\ &= \log_2 M_n + n\epsilon_n^{(5)} \\ &= n \left[C_r(\rho) + \epsilon_n^{(6)} \right]. \end{aligned} \quad (137)$$

Define $|\Phi_{\mathbf{j}, m}^{|\delta_S, \delta_A|}\rangle$ as the normalized $\mathbb{1}_{\mathcal{I}_m} |\Phi_{\mathbf{j}}^{|\delta_S, \delta_A|}\rangle$, and through these, the vector

$$|\bar{\Phi}_{\mathbf{j}}^{|\delta_S, \delta_A|}\rangle := M_n^{-1/2} \sum_m |\Phi_{\mathbf{j}, m}^{|\delta_S, \delta_A|}\rangle. \quad (138)$$

Although this vector is not necessarily in $\mathcal{V}$, it has even m amplitudes and, moreover, satisfies $\langle \bar{\Phi}_{\mathbf{j}}^{|\delta_S, \delta_A|} | \Phi_{\mathbf{j}}^{|\delta_S, \delta_A|} \rangle = \langle \alpha | \Psi_{M_n} \rangle$. Thus, applying Lemma B.1 again and using (113),

$$C_r\left(\bar{\Phi}_{\mathbf{j}}^{|\delta_S, \delta_A|}\right) = n \left[C_f(\rho) + \epsilon_n^{(7)} \right]. \quad (139)$$

Note that

$$\begin{aligned} C_r\left(\bar{\Phi}_{\mathbf{j}}^{|\delta_S, \delta_A|}\right) &= C_r(\Psi_{M_n}) + \frac{\sum_m C_r\left(\Phi_{\mathbf{j}, m}^{|\delta_S, \delta_A|}\right)}{M_n} \\ &= n \left[C_r(\rho) - \tilde{\epsilon}_n \right] + \frac{\sum_m C_r\left(\Phi_{\mathbf{j}, m}^{|\delta_S, \delta_A|}\right)}{M_n}. \end{aligned} \quad (140)$$

Thus,

$$M_n^{-1} \sum_m C_r\left(\Phi_{\mathbf{j}, m}^{|\delta_S, \delta_A|}\right) = n \left[\ell(\rho) + \epsilon_n^{(8)} \right]. \quad (141)$$

Now recall (119), where we showed that for $\delta_R \gg \delta_A + O(\delta_J) + O(\delta_S)$, the fraction of m values (under the uniform measure M_n^{-1}) with $r_C\left(\Phi_{\mathbf{j}, m}^{|\delta_S, \delta_A|}\right) > \exp_2(n[\ell(\rho) + \tilde{\epsilon}_n + \delta_R])$ is no larger than $\exp_2(-n\delta_R)$. Define

$$\bar{\mathcal{M}}_n^{\mathbf{j}, \delta_R} := \left\{ m : C_r\left(\Phi_{\mathbf{j}, m}^{|\delta_S, \delta_A|}\right) \leq n[\ell(\rho) + \tilde{\epsilon}_n + \delta_R] \right\} \quad (142)$$

Since $C_r \leq \log_2 r_C$,

$$1 - \frac{|\bar{\mathcal{M}}_n^{\mathbf{j}, \delta_R}|}{M_n} \leq \exp_2(-n\delta_R). \quad (143)$$

Thus, (141) implies

$$M_n^{-1} \sum_{m \in \bar{\mathcal{M}}_n^{\mathbf{j}, \delta_R}} C_r\left(\Phi_{\mathbf{j}, m}^{|\delta_S, \delta_A|}\right) \geq n \left[\ell(\rho) + \epsilon_n^{(9)} \right]. \quad (144)$$

As in the context of (129), this is again a situation where the average value of a function is close to an extremal value. Choosing $\delta_R \approx \epsilon_n^{(10)} := \sqrt{\epsilon_n^{(9)}}$ and making the other δ 's small enough, we can apply similar arguments to get

$$\frac{|\mathcal{M}_n^{\mathbf{j}, \pm \epsilon_n^{(10)}}|}{M_n} \geq 1 - \epsilon_n^{(10)}, \quad (145)$$

where

$$\mathcal{M}_n^{\mathbf{j}, \pm \epsilon_n^{(10)}} := \left\{ m : \left| \frac{C_r\left(\Phi_{\mathbf{j}, m}^{|\delta_S, \delta_A|}\right)}{n} - \ell(\rho) \right| \leq \epsilon_n^{(10)} \right\}. \quad (146)$$

Now defining

$$\left| \bar{\Phi}_{\mathbf{j}}^{\delta_S, \delta_A, \epsilon_n^{(10)}} \right\rangle := M_n^{-1/2} \sum_{m \in \mathcal{M}_n^{\mathbf{j}, \pm \epsilon_n^{(10)}}} \left| \Phi_{\mathbf{j}, m}^{\delta_S, \delta_A} \right\rangle, \quad (147)$$

it follows from (145) that

$$\left| \left\langle \bar{\Phi}_{\mathbf{j}}^{\delta_S, \delta_A, \epsilon_n^{(10)}} \middle| \bar{\Phi}_{\mathbf{j}}^{\delta_S, \delta_A} \right\rangle \right| \geq 1 - \epsilon_n^{(10)}. \quad (148)$$

Similarly defining $\left| \Phi_{\mathbf{j}}^{\delta_S, \delta_A, \epsilon_n^{(10)}} \right\rangle$ and then applying

$$\left| \left\langle \Phi_{\mathbf{j}}^{\delta_S, \delta_A} \middle| \bar{\Phi}_{\mathbf{j}}^{\delta_S, \delta_A} \right\rangle \right| = |\langle \alpha | \Psi_{M_n} \rangle| \geq \sqrt{1 - \epsilon_n^{(3)}} \quad (149)$$

twice, we have

$$\begin{aligned} \left| \left\langle \Phi_{\mathbf{j}}^{\delta_S, \delta_A, \epsilon_n^{(10)}} \middle| \Phi_{\mathbf{j}}^{\delta_S, \delta_A} \right\rangle \right| &\approx \left| \left\langle \Phi_{\mathbf{j}}^{\delta_S, \delta_A, \epsilon_n^{(10)}} \middle| \bar{\Phi}_{\mathbf{j}}^{\delta_S, \delta_A} \right\rangle \right| \\ &= \left| \left\langle \Phi_{\mathbf{j}}^{\delta_S, \delta_A} \middle| \bar{\Phi}_{\mathbf{j}}^{\delta_S, \delta_A, \epsilon_n^{(10)}} \right\rangle \right| \\ &\approx \left| \left\langle \bar{\Phi}_{\mathbf{j}}^{\delta_S, \delta_A} \middle| \bar{\Phi}_{\mathbf{j}}^{\delta_S, \delta_A, \epsilon_n^{(10)}} \right\rangle \right|. \end{aligned} \quad (150)$$

Thus,

$$\left| \left\langle \Phi_{\mathbf{j}}^{\delta_S, \delta_A, \epsilon_n^{(8)}} \middle| \Phi_{\mathbf{j}}^{\delta_S, \delta_A} \right\rangle \right| \geq 1 - \epsilon_n^{(11)}. \quad (151)$$

We now define

$$\tau_n := \text{norm} \sum_{\mathbf{j} \in \mathcal{T}_n^{\delta_J, \epsilon_n^{(3)}}} Q_n^{\delta_S, \delta_A}(\mathbf{j}) \Phi_{\mathbf{j}}^{\delta_S, \delta_A, \epsilon_n^{(10)}}, \quad (152)$$

where “norm” denotes normalization. The joint concavity of the square-root fidelity, together with (132) and (151), yields

$$F(\tau_n, \varrho_n) \geq 1 - \epsilon_n^{(12)}. \quad (153)$$

Furthermore, every modification since (107) has been among the following types:

1. Unnormalized projection of the entire operator;
2. Unnormalized projection of a pure component in a convex decomposition;
3. Renormalization of the entire operator by a factor close to 1.

Therefore,

$$\tau_n \leq \left(1 + \epsilon_n^{(13)}\right) 2^{-n[S(\rho) - \eta(\delta_S)]}. \quad (154)$$

Defining $\bar{S}_n := (\|\tau_n\|_\infty)^{-1}$,

$$T_n := \bar{S}_n \tau_n \quad (155)$$

satisfies $\text{Tr } T_n \approx \left(1 - \epsilon_n^{(13)}\right) 2^{n[S(\rho) - \eta(\delta_S)]}$ and $T_n \leq \mathbb{1}$. But this T_n may not satisfy condition 1, since in general

$$M_n \mathbb{1}_{\mathcal{I}_m} T_n \mathbb{1}_{\mathcal{I}_m} \not\leq \mathbb{1}_{\mathcal{I}_m}. \quad (156)$$

This adds to our growing list of obstacles against completing our proof. A possible remedy for this particular challenge might be to appeal to the symmetric and random nature of the choice of partitioning (see [5, Supplemental Material, Lemma 16]; [42, Prop. 2.4]) to prune out a small fraction of offending m values. Alternately, in the step (130) where (as we then explicated) we only restricted $\mathbf{j}$ to $\mathcal{T}_n^{\delta_J, \epsilon_n^{(3)}}$, we could additionally project each pure component $\left| \Phi_{\mathbf{j}}^{\delta_S, \delta_A} \right\rangle$ onto $\mathcal{V}$, which by its special structure automatically satisfies

$$M_n \mathbb{1}_{\mathcal{I}_m} \mathbb{1}_{\mathcal{V}} \mathbb{1}_{\mathcal{I}_m} \leq \mathbb{1}_{\mathcal{I}_m}. \quad (157)$$

If we therefore ensure that the overall T_n (at this projection step) satisfies $T_n \lesssim \mathbb{1}_{\mathcal{V}}$, the subsequent modifications would still maintain $M_n \mathbb{1}_{\mathcal{I}_m} T_n \mathbb{1}_{\mathcal{I}_m} \lesssim \mathbb{1}_{\mathcal{I}_m}$, even though the eventual T_n may not be supported on $\mathcal{V}$. This would enable us to uphold condition 1 by applying a normalization factor close to 1 (as in the other steps).

However, projecting $\left| \Phi_{\mathbf{j}}^{\delta_S, \delta_A} \right\rangle$ onto $\mathcal{V}$ brings another problem: while we were able to approximate each $\left| \Phi_{\mathbf{j}}^{\delta_S} \right\rangle$ with a near-uniform superposition $\left| \Phi_{\mathbf{j}}^{\delta_S, \delta_A} \right\rangle$ by appealing to the former’s tensor-product structure, no such structure is assured for $\left| \Phi_{\mathbf{j}}^{\delta_S, \delta_A, \mathcal{V}} \right\rangle$. Though we are able to show that this projected vector is close to the unprojected one, it is not close enough to assure the retention of the near-uniformity of the superposition. Here again, we speculate that the symmetry and randomness in the choice of partitioning might help argue for the existence of near-uniform approximations to the $\mathcal{V}$ -projected vectors.

Finally, if everything worked out, the residual $T_n^\backslash$ would have trace exponentially smaller than that of T_n . Thus, the additional coherent measurement cost from it, as well as its contribution to the average (measured in logarithmic rank), would be negligible. ?

E Potential SDP-based approach

Here we discuss a possible semidefinite programming (SDP)-based approach to the problem of estimating the coherent measurement cost of coherence distillation. We start with some definitions and results from Ringbauer *et al.* [43]. Define the set

$$\mathcal{C}_k := \{\text{density operators } \sigma : r_C(\sigma) \leq k\} \\ = \text{cvx} \{ \psi : \langle \psi | \psi \rangle = 1, r_C(\psi) \leq k \}. \quad (158)$$

The scope of the density operators is to be understood as determined by the underlying system Hilbert space, which we leave implicit.

Definition E.1 (Robustness of multilevel coherence). For a density operator ρ , its *generalized robustness of $(k+1)$ -coherence* is defined as

$$R_{C_k}(\rho) := \inf_{\sigma} \left\{ s \geq 0 : \frac{\rho + s\sigma}{1+s} \in \mathcal{C}_k \right\}, \quad (159)$$

where the infimum is over all density operators σ .

For a system A , denote

$$\mathcal{P}_k := \{\mathcal{I} \subseteq \mathcal{A} : |\mathcal{I}| = k\}. \quad (160)$$

Lemma E.2 ([43, Supplemental Material G]). *The robustness of multilevel coherence can be cast as the following SDP:*

$$R_{C_k}(\rho) = \min_{\text{s.t.}} \left\{ \begin{array}{l} \text{Tr} \left(\sum_{\mathcal{I} \in \mathcal{P}_k} \sigma_{\mathcal{I}} \right) - 1 \\ \sum_{\mathcal{I} \in \mathcal{P}_k} \sigma_{\mathcal{I}} \geq \rho \\ \sigma_{\mathcal{I}} \geq 0 \\ \mathbb{1}_{\mathcal{I}} \sigma_{\mathcal{I}} \mathbb{1}_{\mathcal{I}} = \sigma_{\mathcal{I}} \end{array} \right\} \forall \mathcal{I} \in \mathcal{P}_k. \quad (161)$$

We now note that this definition can be extended to arbitrary positive-semidefinite operators (beyond density operators), by defining

$$R_{C_k}(T) := \min_{\text{s.t.}} \left\{ \begin{array}{l} \text{Tr} \left(\sum_{\mathcal{I} \in \mathcal{P}_k} S_{\mathcal{I}} - T \right) \\ \sum_{\mathcal{I} \in \mathcal{P}_k} S_{\mathcal{I}} \geq T \\ S_{\mathcal{I}} \geq 0 \\ \mathbb{1}_{\mathcal{I}} S_{\mathcal{I}} \mathbb{1}_{\mathcal{I}} = S_{\mathcal{I}} \end{array} \right\} \forall \mathcal{I} \in \mathcal{P}_k, \quad (162)$$

also an SDP. The resulting “robustness” measure is itself not normalized; but it has the convenient property of being (non)zero iff the corresponding value evaluated on the normalized density operator $\tau \equiv T/\text{Tr}T$ is (non)zero. Now, note that

the coherence rank can be defined through the robustness measures as

$$r_C(T) := \min \{k \in \mathbb{Z}_+ : R_{C_k}(T) = 0\}. \quad (163)$$

This definition has the desirable property that $r_C(T) = r_C(xT) \forall x > 0$. Thereby, it lends itself well to use in quantifying the coherent measurement cost of a channel $\mathcal{E}$ in terms of the requisite measurement coherence rank $r_C(T_{\mathcal{E}})$. This quantity can be computed directly through (iterations over instances of) the latter SDP, without the tricky business of having to normalize $T_{\mathcal{E}}$. Thus, using the properties of the target effect construction, we have the following:

Observation E.3. *Given an input state ρ^A and some $M, k \in [A] + 1$ and $\epsilon \in \mathbb{R}_+$, distilling Ψ_M with fidelity $\geq 1 - \epsilon$ by MIO requires measurements of coherence rank $> k$ if $R_k^{\epsilon}(\rho, A, M) > 0$, where*

$$R_k^{\epsilon}(\rho, A, M) := \min_{\text{s.t.}} \left\{ \begin{array}{l} \text{Tr} \left(\sum_{\mathcal{I} \in \mathcal{P}_k} S_{\mathcal{I}} - T \right) \\ 0 \leq T \leq \mathbb{1} \\ \text{Tr}(T|a\rangle\langle a|) = M^{-1} \quad \forall a \in [A] \\ \text{Tr}(\rho T) \geq 1 - \epsilon \\ \sum_{\mathcal{I} \in \mathcal{P}_k} S_{\mathcal{I}} \geq T \\ S_{\mathcal{I}} \geq 0 \\ \mathbb{1}_{\mathcal{I}} S_{\mathcal{I}} \mathbb{1}_{\mathcal{I}} = S_{\mathcal{I}} \end{array} \right\} \quad \forall \mathcal{I} \in \mathcal{P}_k. \quad (164)$$

Therefore, the requisite coherence rank for the task is no smaller than

$$r_C^{\epsilon}(\rho, A, M) := \min \{k \in \mathbb{Z}_+ : R_k^{\epsilon}(\rho, A, M) = 0\}. \quad (165)$$

Given any specific tuple of $(A, M, \rho, \epsilon, k)$, the computation of $R_k^{\epsilon}(\rho, A, M)$ is an SDP. But it may not be efficient in terms of these parameters’ sizes: in particular, $|\mathcal{P}_k| = \binom{A}{k}$ grows roughly exponentially with A . Computing $r_C^{\epsilon}(\rho, A, M)$ would further entail iterating over many k values, potentially up to $k \approx A/M$.

Nevertheless, we hope that the above SDP formulation of $R_k^{\epsilon}(\rho, A, M)$ affords insights into the behaviour of $r_C^{\epsilon}(\rho, A, M)$, e.g. for the case of maximal asymptotic distillation. We leave this for future work as a potential approach towards settling our conjectures.

F Lami's SIO monotones under rank-constrained IO

In order to tease out SIO-distillable coherence, Lami [8] introduced a family of functions μ_k that are monotones under SIO but not under general IO. For any ρ^A , let

$$R^\rho := [\Delta(\rho)]^{-1/2} \rho [\Delta(\rho)]^{-1/2}, \quad (166)$$

$$\mathcal{P}_{k,L} := \left\{ \mathcal{J} \equiv \{\mathcal{I}_m \subseteq \mathcal{A}\}_{m \in [k]} : |\mathcal{I}_m| \leq L \text{ \& } |\mathcal{I}_m \cap \mathcal{I}_{m'}| \propto \delta_{mm'} \ \forall m, m' \in [k] \right\}. \quad (168)$$

At first glance, it may seem that $\mathcal{P}_{k,1} \cong \mathcal{P}_k$ in the notation of Appendix E (where we use “ $\cong$ ” to signify the identification of $\mathcal{J} \equiv \{\{j_m\}\}_{m \in [k]}$ with $\mathcal{I} \equiv \{j_m\}_{m \in [k]}$). But actually, since $|\mathcal{I}_m|$ is allowed to be $< L$ in the definition of $\mathcal{P}_{k,L}$, we have, instead, $\mathcal{P}_{k,1} \cong \bigcup_{l=0}^k \mathcal{P}_l$.

Remark F.2. A generic L -ary IO Kraus operator with an M -dimensional output has the form

$$K = \sum_m |m\rangle \langle v_m|, \quad (169)$$

where $|v_m\rangle \in \text{span}\{|i\rangle\}_{i \in \mathcal{I}_m}$ for some $\mathcal{J} \in \mathcal{P}_{M,L}$.

We will now perform an L -ary construction analogous to the definition (166). For any given $\mathcal{J} \in \mathcal{P}_{M,L}$, define

$$\Delta_{\mathcal{J}}(\cdot) := \sum_m \mathbb{1}_{\mathcal{I}_m}(\cdot) \mathbb{1}_{\mathcal{I}_m}, \quad (170)$$

and therewith, for any ρ ,

$$R_{\mathcal{J}}^\rho := [\Delta_{\mathcal{J}}(\rho)]^{-1/2} \rho [\Delta_{\mathcal{J}}(\rho)]^{-1/2}. \quad (171)$$

It is worth noting that

$$\mathbb{1}_{\mathcal{I}_m} R_{\mathcal{J}}^\rho \mathbb{1}_{\mathcal{I}_m} = \mathbb{1}_{\rho_{\mathcal{I}_m} \equiv \mathbb{1}_{\mathcal{I}_m} \rho \mathbb{1}_{\mathcal{I}_m}}. \quad (172)$$

Now define

$$\mu_{M,L}(\rho) := \max_{\mathcal{J} \in \mathcal{P}_{M,L}} \log_2 \|R_{\mathcal{J}}^\rho\|_\infty. \quad (173)$$

with the inverse defined on the support of $\Delta(\rho)$. Then,

$$\mu_k(\rho) := \max_{\mathcal{I} \subseteq \mathcal{A}: |\mathcal{I}|=k} \log_2 \|\mathbb{1}_{\mathcal{I}} R^\rho \mathbb{1}_{\mathcal{I}}\|_\infty. \quad (167)$$

One way of approaching the problem of quantifying the requisite coherent measurement cost of coherence distillation would be to study the behaviour of these SIO monotones under IO with constrained coherent measurement action.

Definition F.1 (L -ary IO). An IO $\mathcal{E}$ is L -ary if it can be decomposed in terms of IO Kraus operators each of whose rows has $\leq L$ nonzero entries.

For convenience, also define

Observation F.3. Given an input ρ ,

1. For an individual L -ary IO Kraus operator K that partitions its input according to some $\mathcal{J} \in \mathcal{P}_{M,L}$, let $\sigma := K\rho K^\dagger$. Then, $\|R^\sigma\|_\infty \leq \|R_{\mathcal{J}}^\rho\|_\infty$. By convexity, $\mu_M[\mathcal{E}(\rho)] \leq \mu_{M,L}(\rho)$ for any L -ary IO $\mathcal{E}$.
2. For each $\mathcal{J} \in \mathcal{P}_{M,L}$, there exists an L -ary IO Kraus operator K partitioning by $\mathcal{J}$ such that, for $\sigma := K\rho K^\dagger$, $\|R^\sigma\|_\infty = \|R_{\mathcal{J}}^\rho\|_\infty$.

Consequently,

$$\max_{L\text{-ary IO } \mathcal{E}} \mu_M[\mathcal{E}(\rho)] = \mu_{M,L}(\rho). \quad (174)$$

Proof. Let

$$K = \sum_m |m\rangle \langle v_m|, \quad (175)$$

where $|v_m\rangle \in \text{span}\{|a\rangle\}_{a \in \mathcal{I}_m}$ for some $\mathcal{J} \in \mathcal{P}_{M,L}$. For every m , define $|u_m\rangle := \sqrt{\rho_{\mathcal{I}_m}} |v_m\rangle \in \text{supp}(\rho_{\mathcal{I}_m})$. Then, for $\sigma := K\rho K^\dagger$,

$$\begin{aligned} \langle m_1 | \sigma | m_2 \rangle &= \langle v_{m_1} | \rho | v_{m_2} \rangle \\ &= \langle v_{m_1} | \sqrt{\rho_{\mathcal{I}_{m_1}}} R_{\mathcal{J}}^\rho \sqrt{\rho_{\mathcal{I}_{m_2}}} | v_{m_2} \rangle \\ &= \langle u_{m_1} | R_{\mathcal{J}}^\rho | u_{m_2} \rangle \end{aligned} \quad (176)$$

for all m_1, m_2 . Therefore,

$$\begin{aligned} R^\sigma &= \sum_{m_1, m_2} \frac{|m_1\rangle \langle m_1| \sigma |m_2\rangle \langle m_2|}{\sqrt{\langle m_1| \sigma |m_1\rangle \langle m_2| \sigma |m_2\rangle}} \\ &= \sum_{m_1, m_2} \frac{|m_1\rangle \langle u_{m_1}| R_{\mathcal{J}}^\rho |u_{m_2}\rangle \langle m_2|}{\sqrt{\langle u_{m_1}| R_{\mathcal{J}}^\rho |u_{m_1}\rangle \langle u_{m_2}| R_{\mathcal{J}}^\rho |u_{m_2}\rangle}} \\ &= \sum_{m_1, m_2} \frac{|m_1\rangle \langle u_{m_1}| R_{\mathcal{J}}^\rho |u_{m_2}\rangle \langle m_2|}{\sqrt{\langle u_{m_1}| u_{m_1}\rangle \langle u_{m_2}| u_{m_2}\rangle}}, \quad (177) \end{aligned}$$

the last line following from (172). Now, for any $|\phi\rangle \in \text{supp}(R^\sigma)$, define

$$|\psi\rangle := \sum_m \frac{|u_m\rangle \langle m| \phi\rangle}{\sqrt{\langle u_m| u_m\rangle}}. \quad (178)$$

$\langle \psi| \psi\rangle = \langle \phi| \phi\rangle$ holds by construction, while $\langle \psi| R_{\mathcal{J}}^\rho |\psi\rangle = \langle \phi| R^\sigma |\phi\rangle$ follows from (177). Thus, $\|R^\sigma\|_\infty \leq \|R_{\mathcal{J}}^\rho\|_\infty$.

Conversely, given any $\mathcal{J} \in \mathcal{P}_{M,L}$ and $|\psi\rangle := \sum_m |u_m\rangle$ with $|u_m\rangle \in \text{supp}(\rho_{\mathcal{I}_m})$, define $|v_m\rangle := \rho_{\mathcal{I}_m}^{-1/2} |u_m\rangle$ for each m . Then, $K := \sum_m |m\rangle \langle v_m|$ is an L -ary IO Kraus operator such that, for $\sigma := K\rho K^\dagger$, (177) holds. Defining

$$|\phi\rangle := \sum_m \sqrt{\langle u_m| u_m\rangle} |m\rangle, \quad (179)$$

we obtain $\langle \phi| \phi\rangle = \langle \psi| \psi\rangle$ and $\langle \phi| R^\sigma |\phi\rangle = \langle \psi| R_{\mathcal{J}}^\rho |\psi\rangle$. Thus, $\|R^\sigma\|_\infty \leq \|R_{\mathcal{J}}^\rho\|_\infty$ can be saturated by choosing $|\psi\rangle$ appropriately.

Finally, if $\mathcal{J}$ is a partitioning that attains the maximum on the right side of (173), and K a Kraus operator constructed as above to saturate $\|R^\sigma\|_\infty \leq \|R_{\mathcal{J}}^\rho\|_\infty$, we can complete this to an L -ary IO channel $\mathcal{E}$ with other Kraus operators whose output spaces don't overlap with that of K . This $\mathcal{E}$ then attains the maximum. $\square$

This bound is tight when the maximization is over *all* L -ary IO $\mathcal{E}$. But what is relevant in the context of distillation is the μ_M for M equalling the dimension of the channel's output. Clearly, this is a much more constrained quantity and would generically not attain the bound we have found. Specifically, the argument used in the last paragraph of the proof would fail when the output space is constrained to be M -dimensional. If we are to use this bound to pin down the distillation rate under L -ary IO, we need to understand the behaviour of $\mu_{M,L}$ under tensor product copies (i.e., an analog of [8, Proposition 16]). By virtue of asymptotic typicality, we might get by without having to incorporate the constraint mentioned above.

G Decoupling schemes

Here we will study a type of decoupling task, whereof coherence distillation is a special case. In particular, we will look at *exact deterministic decoupling of a pure output*. The input in this task is some state ρ , and the goal is to apply a channel $\mathcal{E}$ such that $\mathcal{E}(\rho) = |\alpha\rangle \langle \alpha|$ for some specified pure state $|\alpha\rangle := \sum_m \alpha_m |m\rangle$. On the face of it, it may appear superfluous to view this as decoupling—for example, $\mathcal{E}$ could just ignore the input and prepare the required output. But the utility of the decoupling perspective will become apparent presently.

Recall that exactly producing a pure output $|\alpha\rangle \langle \alpha|^M$ from ρ^A entails that the channel $\mathcal{E}$ map the entire space $\mathcal{L}(\mathcal{V} \equiv \text{supp} \rho)$ to (scalar multiples of) $|\alpha\rangle \langle \alpha|$. Let $S := \dim \mathcal{V}$ and $\{|v_s\rangle\}_{s \in S}$ be a $\mathcal{V}$ basis. Then, $\mathcal{E}$ must have a dilation V with the action

$$V^{A \rightarrow \text{CM}} |v_s\rangle^A = |u_s\rangle^C |\alpha\rangle^M \quad (180)$$

with orthonormal $\{|u_s\rangle\}_{s \in S}$. If V can be completed to a unitary within $\mathcal{H}^A$, we then have some well-defined $|v_{s,m}\rangle^A := V^\dagger (|u_s\rangle^C |m\rangle^M) \in \mathcal{H}^A$. Again by unitarity, the collection $\mathcal{V}_m := \{|v_{s,m}\rangle\}_{s \in S}$ for each $m \in \mathcal{M}$ must be orthonormal; in fact, the entire collection $\mathcal{V}_{\mathcal{M}} := \bigcup_{m \in \mathcal{M}} \mathcal{V}_m$ must be orthonormal.

The conditions above are already necessary for the existence of a unitary $V^{A \rightarrow \text{CM}}$ that accomplishes the task; the requirement that it be an IO dilation will bring in further constraints. But we will ignore these for the time being and try to learn what we can about unitary decoupling in general.

Definition G.1 (α -decoupling scheme). For a dimension- S subspace $\mathcal{V}$ of a Hilbert space $\mathcal{H}^A \cong \mathbb{C}^A$ and a unit vector $\alpha \equiv (\alpha_m)_{m \in \mathcal{M} \equiv [M]} \in \mathbb{C}^M$, a collection $\mathcal{V}_{\mathcal{M}} \equiv \left\{ |v_{s,m}\rangle^A \in \mathcal{H}^A \right\}_{m \in \mathcal{M}, s \in S} = \bigcup_m \left(\mathcal{V}_m \equiv \left\{ |v_{s,m}\rangle^A \right\}_s \right)$ of orthonormal vectors is an α -decoupling scheme for $\mathcal{V}$ in $\mathcal{H}^A$ if $\mathcal{V} \equiv \{|v_s\rangle = \sum_m \alpha_m |v_{s,m}\rangle\}_s$ is a basis for $\mathcal{V}$.

Observation G.2. For an α -decoupling scheme $\mathcal{V}_{\mathcal{M}}$ in $\mathcal{H}^A$ for $\mathcal{V}$, let $|\alpha\rangle := \sum_m \alpha_m |m\rangle$ and $\mathcal{V}_{\mathcal{M}} := \text{span} \mathcal{V}_{\mathcal{M}}$. For some system C of dimensionality $C \geq A/M$, define a unitary $V^{A \rightarrow \text{CM}}$

whose action on $\mathcal{V}_{\mathcal{M}}$ is given by

$$V_{\mathcal{V}_{\mathcal{M}}}^{A \rightarrow \text{CM}} = \sum_{s \in \mathcal{S}, m \in \mathcal{M}} |s\rangle^{\text{C}} |m\rangle^{\text{M}} \langle v_{s,m}|^{\text{A}}. \quad (181)$$

Then, for any $|v\rangle \equiv \sum_s \xi_s |v_s\rangle \in \mathcal{V}$,

$$V|v\rangle = \left(\sum_s \xi_s |s\rangle \right)^{\text{C}} \otimes |\alpha\rangle^{\text{M}}. \quad (182)$$

We omit a proof for the above observation, since it is straightforward. Nevertheless, it is instructive in explicating the operational motivation for our definition of the term “ α -decoupling scheme”. It also suggests that the decoupling perspective on the problem has potential nontrivial utility when the space $\mathcal{H}^{\text{A}}$ within which the channel can be dilated is explicitly considered.

So far we restricted ourselves to cases where it can be done through unitary action within $\mathcal{H}^{\text{A}}$. Such a unitary gave us a decoupling scheme consisting of orthogonal vectors $|v_{s,m}\rangle^{\text{A}}$. But in general, we need to consider decoupling isometries that may not be completable to a unitary within $\mathcal{H}^{\text{A}}$. The associated analogs to decoupling schemes could then be oblique and overcomplete. To understand these cases, let us once again examine the basic condition on an α -decoupling isometry V on a space $\mathcal{V}$:

$$V^{A \rightarrow \text{CM}} |v_s\rangle^{\text{A}} = |u_s\rangle^{\text{C}} |\alpha\rangle^{\text{M}}, \quad (183)$$

where $|v_s\rangle$ is a basis on $\mathcal{V}$. Now let us expand A to some $\bar{\text{A}}$, of dimensionality $\bar{\text{A}} := \text{CM}$, such that $V^{A \rightarrow \text{CM}}$ can be completed to the unitary $V^{\bar{\text{A}} \rightarrow \text{CM}}$. By assumption, $V^\dagger |u_s\rangle^{\text{C}} |\alpha\rangle^{\text{M}} = |v_s\rangle^{\text{A}} =: |\bar{v}_s\rangle^{\bar{\text{A}}}$. Indeed, if we complete $\{|u_s\rangle\}_{s \in \mathcal{S}}$ to an $\mathcal{H}^{\text{C}}$ basis $\{|u_c\rangle\}_{c \in \mathcal{C}}$ and thereby define $|\bar{v}_c\rangle^{\bar{\text{A}}} := V^\dagger |u_c\rangle^{\text{C}} |\alpha\rangle^{\text{M}}$ for all $c \in \mathcal{C}$, then by construction, $\bar{\mathcal{V}} := \text{span} \left[\bar{\mathcal{V}} \equiv \left\{ |\bar{v}_c\rangle^{\bar{\text{A}}} \right\}_{c \in \mathcal{C}} \right]$ is an α -decouplable subspace of $\mathcal{H}^{\bar{\text{A}}}$.

Recalling our ultimate aim to estimate the coherent measurement cost of an IO whose dilation is some V as above, let us inspect a set of Kraus operators that would implement the combined action of the embedding of A in $\bar{\text{A}}$, then the unitary V , and finally a C -partial trace; since we have kept C and $|u_c\rangle^{\text{C}}$ generic, we lose no generality in decomposing the partial trace in the canonical basis $\langle c|^{\text{C}}$, as in Remark 2.2. The Kraus operators, then, are

$$K_c^{A \rightarrow \text{M}} = \langle c|^{\text{C}} V^{\bar{\text{A}} \rightarrow \text{CM}} \mathbb{1}^{\text{A}} =: \sum_{m \in \mathcal{M}} |m\rangle^{\text{M}} \langle w_{c,m}|^{\text{A}}. \quad (184)$$

Here $|w_{c,m}\rangle^{\text{A}} = \mathbb{1}^{\text{A}} |\bar{w}_{c,m}\rangle^{\bar{\text{A}}}$, the latter defined through

$$\sum_{c,m} |c\rangle^{\text{C}} |m\rangle^{\text{M}} \langle \bar{w}_{c,m}|^{\bar{\text{A}}} = \sum_{c,m} |u_c\rangle^{\text{C}} |m\rangle^{\text{M}} \langle \bar{v}_{c,m}|^{\bar{\text{A}}} (= V). \quad (185)$$

Notice that $\bar{\mathcal{W}}_{\mathcal{M}} \equiv \left\{ |\bar{w}_{c,m}\rangle^{\bar{\text{A}}} \right\}_{c,m}$ is, just like $\bar{\mathcal{V}}_{\mathcal{M}}$, an α -decoupling scheme for $\bar{\mathcal{V}}$ in $\mathcal{H}^{\bar{\text{A}}}$; the schemes are associated, respectively, with the $\bar{\mathcal{V}}$ bases $\bar{\mathcal{W}} \equiv \left\{ |\bar{w}_c\rangle^{\bar{\text{A}}} := \sum_m \alpha_m |\bar{w}_{c,m}\rangle^{\bar{\text{A}}} \right\}_c$ and $\bar{\mathcal{V}}$. We summarize the above observations in the following... well, observation.

Observation G.3. *If $\mathcal{E}^{A \rightarrow \text{M}}$ is a channel (IO or otherwise) that deterministically maps a subspace $\mathcal{V} \in \mathcal{H}^{\text{A}}$ to $|\alpha\rangle \langle \alpha|^{\text{M}}$, any Kraus operator decomposition of $\mathcal{E}$ must involve operators of the form*

$$K_c = \sum_{m \in \mathcal{M}} |m\rangle^{\text{M}} \langle w_{c,m}|^{\text{A}} \quad (186)$$

with $|w_{c,m}\rangle^{\text{A}} = \mathbb{1}^{\text{A}} |\bar{w}_{c,m}\rangle^{\bar{\text{A}}}$ projections of an α -decoupling scheme $\bar{\mathcal{W}}_{\mathcal{M}} \equiv \left\{ |\bar{w}_{c,m}\rangle^{\bar{\text{A}}} \right\}_{c,m}$ for some $\bar{\mathcal{V}} \supset \mathcal{V}$ in some $\mathcal{H}^{\bar{\text{A}}} \supset \mathcal{H}^{\text{A}}$.

Thus, the problem of finding the least-coherent IO channel(s) executing a given decoupling task involves optimizing over all possible decoupling schemes within arbitrarily large $\bar{\text{A}}$ and $\bar{\mathcal{V}}$.

Since $\bar{\mathcal{W}}$ (introduced before Observation G.3) is a $\bar{\mathcal{V}}$ -basis and $\mathcal{V} \subset \bar{\mathcal{V}}$,

$$\begin{aligned} \mathbb{1}_{\bar{\mathcal{V}}} &= \sum_c |\bar{w}_c\rangle \langle \bar{w}_c| \geq \mathbb{1}_{\mathcal{V}} \\ \Rightarrow T_{\mathcal{E}} &= \mathbb{1}^{\text{A}} \left(\sum_c |\bar{w}_c\rangle \langle \bar{w}_c| \right) \mathbb{1}^{\text{A}} \geq \mathbb{1}^{\text{A}} \mathbb{1}_{\mathcal{V}} \mathbb{1}^{\text{A}} = \mathbb{1}_{\mathcal{V}}, \end{aligned} \quad (187)$$

where we used the fact that $\mathcal{V} \subset \mathcal{H}^{\text{A}}$. Meanwhile, since $\mathbb{1}_{\bar{\mathcal{V}}} \leq \mathbb{1}^{\bar{\text{A}}}$, $T_{\mathcal{E}} = \mathbb{1}^{\text{A}} \mathbb{1}_{\bar{\mathcal{V}}} \mathbb{1}^{\text{A}} \leq \mathbb{1}^{\text{A}} \mathbb{1}^{\bar{\text{A}}} \mathbb{1}^{\text{A}} = \mathbb{1}^{\text{A}}$. Thus, using the concept of decoupling schemes we have arrived at the result of Lemma 4.1 via a different route. More research into these structures may allow us to harness their properties to attack channel-related problems in the dilation picture.