universität wien

DISSERTATION | DOCTORAL THESIS

Titel | Title

Nonclassical Temporal Correlations Under Finite-Memory Constraints

verfasst von | submitted by

Lucas Vieira Barbosa

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 796 605 411
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Physik
Betreut von | Supervisor:	Univ.-Prof. Mag. Dr. Caslav Brukner
Mitbetreut von | Co-Supervisor:	Dr. Costantino Budroni

“O tempo só anda de ida.”

— Manoel de Barros

Kurzfassung

Jeder Informationsverarbeitungsvorgang kann als eine sequentielle Manipulation von Information über die Zeit hinweg betrachtet werden, bei der Eingaben empfangen und Ausgaben erzeugt werden. Zu diesem Zweck muss Information in einem physischen System, das als Speichermedium dient, gespeichert werden und es ist intuitiv einsichtig, dass eine begrenzte Speicherkapazität fundamentale Einschränkungen für die Art der durchführbaren Aufgaben nach sich zieht.

Aus physikalsicher Perspektive betrachtet hat diese abstrakte Einsicht tiefgreifende Konsequenzen. Dies wird deutlich, wenn man in Betracht zieht, dass das Ziel jeder informationstheoretischen Aufgabe im Wesentlichen darin besteht, spezifische zeitliche Korrelationen zwischen Eingaben und Ausgaben zu erzeugen. Diese Erkenntnis veranlasst uns, die physikalische Erzeugung von zeitlichen Korrelationen als eine eigenständige Aufgabe ohne darüber hinausgehende Interpretation zu betrachten, bei der die Erzeugung von Zuständen, sowohl wie deren Transformation und Messung als Speicheroperationen fungieren.

Die vorliegende Dissertation folgt dieser Sichtweise, indem sie zeitliche Korrelationen untersucht, die durch sequentielle Messung eines einzelnen System entstehen, wobei jede Messung eine Ausgabe erzeugt. Insbesondere untersuchen wir die fundamentalen Beschränkungen der daraus resultierenden zeitlichen Korrelationen, wenn dieses System eine endliche Dimension aufweist, was einer begrenzten Speicherkapazität entspricht. Zu diesem Zweck verwenden wir das Modell endlicher Zustandsmaschinen, das vor Kurzem für die Untersuchung zeitlicher Korrelationen, die ohne die häufig gemachte Annahme nichtinvasiver klassischer Messungen entstehen können, eingeführt wurde. Dieser Ansatz ermöglicht den Vergleich von Quantenmechanik und klassischer Physik unter gleichen Bedingungen, sodass ein präzises Konzept von *echten nichtklassischen* zeitlichen Korrelationen etabliert werden kann.

Wir beginnen diese Betrachtung mit der Aufgabe, spezifische Sequenzen von Ausgaben deterministisch zu erzeugen, und führen den Begriff der *deterministischen Komplexität* ein, um die minimale Dimension zu beziffern, die dafür notwendig ist, sei es für ein klassisches oder ein quantenmechanisches System. Unterhalb dieser Grenze and Speicherkapazität können die entsprechenden Sequenzen nur probabilistisch erzeugt werden, und wir führen eine umfassende Untersuchung der entsprechenden maximalen Erzeugungswahrscheinlichkeiten durch, sowohl im klassischen, wie im quantenmechanischen Fall. Ein Vergleich der beiden Theorien zeigt mehrere unerwartete Strukturen und fundamentale Unterschiede zwischen klassischen und quantenmechanischen Korrelationen auf.

Im Anschluss untersuchen wir die Anwendung zeitlicher Korrelationen als Zeuge für die Dimension der Umgebung eines offenen Quantensystems. Wir formulieren eine Hierarchie von semidefiniten Programmen, die in der Lage sind, obere Schranken für zeitliche Korrelationen zu berechnen, wobei eine dünnbesetzte Darstellung dieses Problems erforderlich war, um es zu lösen. Diese dünnbesetzte Darstellung ergibt sich aus einer neuen heuristischen Methode, die wir ebenfalls einführen, und die unnötige Variablen und Einschränkungen im Problem entfernt.

Danach untersuchen wir, ob verschränkungsbrechende (“entanglement breaking”) Kanäle als klassisches Speichermedium in einem zeitlichen Szenario angesehen werden können, wie dies oft in der Untersuchung der Nicht-Markovianität in Quantenprozessen, die zu mehreren Zeitpunkten gemessen werden, angenommen wird. Wir zeigen, dass dies nicht der Fall ist, indem wir explizite quantenmechanische Modelle präsentieren, die genauen klassischen Schranken verletzen, und dadurch subtile nichtklassische zeitliche Effekte aufdecken, die vormals aufgrund unzutreffender Analogien mit räumlichen Korrelationen übersehen wurden.

Zuletzt führen wir eine Verallgemeinerung der deterministischen Komplexität für den Fall von Eingabe-Ausgabe-Sequenzen ein und stellen Verbindungen zwischen zeitlichen Korrelationen und Gebieten der Graphentheorie sowie der theoretischen Informatik her.

Abstract

Any concrete information processing task can be viewed as the sequential manipulation of information over time, receiving inputs and producing outputs. To accomplish this, information must be stored in a physical system serving as a memory resource, and it is intuitively understood that a finite amount of memory imposes fundamental limitations on the tasks that can be performed.

These abstract notions have profound consequences when viewed from a physics perspective, made clear by recognizing that the goal of an information-theoretical task is, essentially, to generate specific temporal correlations between inputs and outputs. This insight prompts us to investigate the physical realization of temporal correlations as a task in its own right, forgoing any additional interpretation, with state preparations, transformations, and measurements acting as memory operations.

This thesis explores this perspective by investigating temporal correlations arising from sequential measurements on a single system, each measurement producing an output. In particular, we investigate how the resulting temporal correlations are fundamentally constrained when this system is finite-dimensional, corresponding to a finite amount of memory. To this end we employ the framework of finite-state machines, recently introduced in the study of temporal correlations without the traditional assumption of noninvasive classical measurements. This approach allows classical and quantum theory to be compared on equal terms, such that a precise notion of *genuinely nonclassical* temporal correlations can be established.

We begin by considering the task of deterministically generating specific sequences of outputs, introducing the notion of *deterministic complexity* to quantify the minimum dimension where this can be achieved, either by a classical or quantum system. Below this memory threshold, each sequence can only be generated probabilistically, and we perform an extensive survey of the corresponding maximum probabilities in both classical and quantum theory. A comparison reveals several unexpected structures and fundamental differences between classical and quantum temporal correlations.

Next, we investigate the application of temporal correlations in dimension witnesses for the environment of an open quantum system. We formulate a hierarchy of semidefinite programs capable of computing upper bounds on temporal correlations, which required a sparse representation in order to be solved. This sparse representation is obtained through a novel heuristic method, which we also introduce, that removes unnecessary variables and constraints in the problem.

We then investigate whether entanglement-breaking channels can be considered a classical memory resource in a temporal scenario, as is often assumed in the study of non-Markovianity in multi-time quantum processes. We answer in the negative by providing explicit quantum models violating the exact classical bounds, thereby uncovering subtle nonclassical temporal effects that have been overlooked due to improper analogies with spatial correlations.

Lastly, we introduce a generalization of the deterministic complexity to the case of sequences of inputs and outputs, establishing connections between temporal correlations and topics in graph theory and theoretical computer science.

Keywords: Temporal correlations, classicality and nonclassicality, information processing, memory, finite-state machines, Mealy machines, complexity, causality, Arrow-of-Time, open quantum systems, dimension witnesses, semidefinite programming, sparsity.

Acknowledgements

This thesis is the culmination of a long personal journey, so far spanning over 17 years and 10 000 km, and it would not have been possible without the support of my dear family, friends, and colleagues.

First, I would like to acknowledge my supervisors. I am grateful to Prof. Časlav Brukner, for his support, generosity, and wisdom throughout this process. Most importantly, I would like to express my deepest gratitude to my de-facto supervisor, Costantino Budroni, who gave me the opportunity to come to Vienna and work on such endlessly fascinating topics. I am eternally grateful for your trust, kindness, and patience. Your clarity of thought has been an immense source of inspiration.

I am also grateful to my other former YIRG colleagues: Yelena Guryanova, for your thoughtfulness and sharp sense of humor; Ämin Baumeler, for the wonderful discussions about life, computation, and everything in between; Giuseppe Vitagliano, for your creative insights and unwavering positive attitude; Paul Erker, for your admirable joy for physics and life; Lefteris Tselentis, for the camaraderie in navigating the challenges and bureaucracies of the PhD student life; and last, but not least, Simon Milz, for all your guidance, knowledge, friendship, and cheerful outlook on things.

Next, I would like to thank my Brazilian friends (and colleagues) during my time at IQOQI: Jessica Bavaresco, for your kindness, joyfulness, and incredibly sharp mind; Marco Túlio Quintino, for your seemingly infinite supply of both knowledge and good humor, and for helping me manage my *saudade* of Minas Gerais; and Isadora Veeren, for your wonderful insights, and the enthusiasm for life, physics, and good horror movies. Um enorme abraço pra vocês!

I am also thankful to Huan-Yu Ku, whose cleverness, sense of humor, generosity, and enthusiasm have all been inspiring. I must also express my sincere gratitude to Borivoje Dakić, whose mentorship gave me confidence when I needed it the most. I would also like to thank Joshua Morris and Miguel Navascués for valuable discussions throughout my research, and all the IQOQI administrative staff for their help and patience during my stay.

The work in this thesis was only possible through the generous financial support of the Austrian Science Fund (FWF), project ZK 3 (Zukunftskolleg).

On a personal level, I would first like to thank Bri, Sophie, and Martin. You have been like a family to me here in Austria, and I would not have made it this far without your generosity and companionship. I am also immensely grateful to Bianca, whose love, kindness, humor, and beautiful perspective on the world have all brought me so much joy.

A special thanks to my friends, who encouraged and cheered me on: Amber, Ria, and Renata, to whom I have no words to express my gratitude; from UFMG: Ludmila, Anna, Stela, André, Bonnie, Rodolfo, Clara, Izabel, Luiz, Prof. Elmo, Prof. Reinaldo, Thiago, Marcello, and Diego; and from Uberlândia: Bruno, Lorena, Spy, Beta, and Camila, with a heartfelt thanks to Lélia and Lara, who reminded me to be kind to myself.

Most importantly, I would like to thank my family, whom I love dearly: Maria Lúcia, Armando, Isis, Candida, and João. I am so lucky to have all of you, and I am forever grateful.

Vienna, 12 of June, 2024.

"Get busy living, or get busy dying." — Andy Dufresne

List of Publications

Peer reviewed

[198] *Temporal correlations in the simplest measurement sequences*
Lucas B. Vieira and Costantino Budroni
Quantum 6, p. 623 (2022) — Accepted 2022-01-09
Subject of Chapter 2

[201] *Witnessing environment dimension through temporal correlations*
Lucas B. Vieira, Simon Milz, Giuseppe Vitagliano, and Costantino Budroni
Quantum 8, p. 1224 (2024) — Accepted 2023-12-04
Subject of Chapter 3, and part of Chapter 5

Under review

[200] *Entanglement-breaking channels are a quantum memory resource*
Lucas B. Vieira, Huan-Yu Ku, Costantino Budroni
arXiv — Submitted 2024-02-06
Subject of Chapter 4

Contents

List of Figures

List of Tables

Preamble

Classical physics is based on our everyday experience with the macroscopic world, where physical systems and their properties seem to exist independently of observation. In this view, measurements can be understood as a means of uncovering information about an objective reality. Any probabilistic behavior can be attributed to uncertainty, possibly arising from imprecise measurements or incomplete knowledge about the underlying physical states.

While reasonable at the macroscopic scale, at the microscopic scale these notions have been challenged by quantum mechanics in profound ways. Not only do systems appear to have indefinite properties before measurements, but probabilistic behaviors also seem to emerge from a fundamental degree of indeterminacy in Nature. The predictions of quantum mechanics have achieved unprecedented success, being vindicated against competing explanations in numerous experimental tests. These results have forced us completely reevaluate our understanding of physical reality.

One of the most surprising predictions of quantum mechanics is the possibility of correlations between measurements that cannot be explained by classical theories. These *nonclassical correlations* have been the subject of intense research, mostly in the case of spatially separated systems. In their landmark paper, Einstein, Podolsky, and Rosen (EPR) [62] discuss how quantum theory predicts spatial correlations which would violate the principle of locality. To address this, they postulated that quantum theory might be incomplete, and that these scenarios could be described in terms of classical probabilities involving additional parameters, now known as *local hidden variables*. This concept was later addressed by Bell [11, 13], who formulated an experimentally testable inequality establishing bounds on spatial correlations achievable by any theory based on local hidden variables, but capable of being violated by quantum mechanics. Over the following decades, numerous experimental Bell tests were devised and performed, validating the predictions of quantum mechanics and establishing the existence of nonclassical spatial correlations beyond any model of local hidden variables.

However, physical phenomena unfold not only in space, but also in time. What do we know about nonclassical *temporal* correlations? Given the fundamental role of time in physics, one might expect a similar amount of research devoted to the subject. Unfortunately, that does not seem to be the case. There seems to be comparatively little research directly targeting nonclassical temporal correlations, with existing results either based on very narrow notions of "classicality", or relying on direct analogies to spatial correlations. Both approaches are unsatisfactory, as we will show throughout this thesis.

One way to directly investigate nonclassical temporal correlations is to consider physical scenarios involving a finite and discrete sequence of measurements on a single system. Early research using this idea can be traced back to the work of Leggett and Garg [116], who proposed a way to test for the existence of macroscopic quantum coherence by means of temporal correlations. In a similar spirit as Bell's inequality, Leggett and Garg formulated inequalities which would be satisfied by any theory obeying a certain set of assumptions expected of macroscopic systems. An experimental violation of a Leggett-Garg inequality would then force us to abandon at least one of these assumptions.

The first assumption is that of macroscopic realism, the idea that any macroscopic system should

always be found in only one of its macroscopically distinct states at any given moment. The second assumption is that of noninvasive measurability, meaning that any properties of this macroscopic state could (in principle) be measured without changing the result of subsequent measurements. Here, implicitly, is also the assumption of causality: the outcome of measurements in the present is not influenced by future choices of measurements. A theory obeying these assumptions is said to be "macrorealist".

While most research into Leggett-Garg inequalities revolves around tests of macrorealism, these assumptions have also become the standard way to formulate "nonclassicality" for temporal correlations, with the notion of "classicality" largely being conflated with the notion of noninvasive measurability. Realism, macroscopic or otherwise, is usually left implicit. This may have been reasonable for macroscopic systems, but it seems unreasonable as a general physical principle to distinguish between classical and nonclassical correlations at a fundamental level. Indeed, classical invasive measurements can easily violate a Leggett-Garg inequality, regardless of whether or not they could have "in principle" been made noninvasive. Therefore, we must abandon the notion of noninvasive measurability if the goal is to identify *genuinely* nonclassical temporal correlations.

A proper characterization of temporal correlations has only began to be pursued in recent years [52, 1, 219, 88, 180, 183, 182, 124, 34, 31, 32], and this thesis adds to this research effort. To this end, we reject the notion of invasiveness as an "unwanted disturbance", to be minimized or eliminated, instead considering it a deliberate manipulation of information in a physical state. In this way, sequential measurements on a system can be understood as information processing, with the physical system acting as a *memory resource*, storing information as it is manipulated over time. Nonclassical temporal correlations can then be identified using sequential tasks where quantum memories provide an advantage over the corresponding classical memories.

This perspective highlights fundamental differences between spatial and temporal correlations. In the spatial scenario, the principle of locality requires correlations to satisfy the No-Signaling conditions, which forbid two-way superluminal signaling between distant parties. In the temporal scenario, however, the principle of causality requires correlations to satisfy the Arrow-of-Time constraints [52], which only forbid signaling from the future to the past.

Since information can still be sent from the past to the future, any temporal correlation can be achieved with either classical or quantum memories if the amount of information is not restricted [67, 88, 183, 182, 124]. This is intuitively true, as we could imagine simply storing in memory the outcomes of all possible future measurements in advance. What this observation tells us is that, in the context of temporal correlations, it is also essential to have a detailed accounting of all memory resources available. This undertaking can be simplified by considering a single system acting as memory, in complete isolation, such that the system itself is the only available carrier of any information from the past to the future [88, 32, 182]. The amount of information can then be bounded by assuming a finite-dimensional system.

These considerations motivate studying temporal correlations through the framework of *finite-state machines* (FSMs) [90, 32], originally introduced in theoretical computer science as simple models of computation involving finite memory. A FSM is an abstract device that processes sequential inputs to generate the corresponding outputs. It consists of a finite number of states and a fixed transition rule, dictating how it behaves depending on its current state, the current input, and the output to be generated. While originally formulated in terms of deterministic classical devices, FSMs can be naturally extended [32] to the case of general probabilistic theories (GPTs) [123, 83, 44], which include classical and quantum theory as special cases. The "amount of memory" can then be directly quantified by the dimension of the FSM's memory, i.e., the maximum number of perfectly distinguishable states.

The main advantage of the FSM framework is that it allows a fair comparison to be made between different theories, such that the boundary of their corresponding sets of temporal correlations can be sharply

established. This application of FSMs was first formalized by Budroni et al. [32], where the authors derive temporal inequalities that can distinguish between classical, quantum, or genuinely GPT temporal correlations. FSMs have also been used for quantum dimension witnesses [88, 183], in establishing the memory cost of simulating extreme temporal correlations and their convex mixtures [182, 124], and quantum advantages enhancing the performance of time-keeping devices [33]. These results showcase the power and versatility of this approach.

This thesis is devoted to a systematic characterization of temporal correlations through the framework of finite-state machines. In particular, we focus on the problem of establishing sharp bounds on classical and quantum temporal correlations, obtained via sequential measurements on a finite-dimensional system. These bounds can be used in many applications, such as witnesses for the dimension or nonclassicality of an underlying system, which will be discussed in detail.

As mentioned earlier, sequential invasive measurements can be interpreted as an information processing task. Nevertheless, we will generally refrain from imposing any particular meaning to such tasks (e.g., an implementation of a specific protocol, or investigating correlations within some sort of adversarial game). Instead, we will treat the physical realization of a specific temporal correlation as the task in and of itself, with only causality and finite memory as constraints.

Note that this should not be confused with the problem of *statistical inference* when modeling or learning some existing time series data, as is done in the study of stochastic processes with hidden Markov models [158] or in computational mechanics with ε-machines [55, 119]. While all these approaches use a similar formalism as our FSMs[1], here we focus on the fundamental limits of classical and quantum theory in generating an arbitrary temporal correlation.

Existing research on temporal correlations based on finite-state machines is very recent, with most relevant results having only been published within the last decade. It is therefore important to provide the historical and technical context leading to its development. This is the main purpose of Chapter 1, which provides a background on Leggett-Garg inequalities and the subsequent use of noninvasive measurability as a definition of "classicality", highlighting its shortcomings. We then review the recent approach on temporal correlations based on invasive measurements and finite-dimensional systems, which this thesis builds upon.

We discuss the essential role of the Arrow-of-Time constraints and the Arrow-of-Time polytope in the context of temporal correlations, and how its extreme points correspond to deterministic distributions. Importantly, each extreme point requires a minimum dimension for its physical realization, which is the same for both classical and quantum systems. Thus, all temporal correlations can be realized by classical or quantum systems of sufficient size: an important difference between temporal and spatial correlations.

These results inspire us to interpret the system as a *memory resource*, with its dimension directly quantifying the "amount of memory", whether classical or quantum. In other words, the necessary assumption of a finite dimensional system can be understood as a *finite-memory constraint* on temporal correlations. This motivates the use of finite-state machines to analyze temporal correlations as information-processing tasks, with measurement settings being understood as "inputs" and outcomes as "outputs". The chapter concludes with a formal definition of these classical and quantum finite-state machines, followed by a short review of the numerical optimization methods used in the rest of the thesis.

Realizing that very little was known about temporal correlations, either classical or quantum, we performed a general survey in search of new insights. This is the subject of Chapter 2, where we investigate

[1] Our approach is from a physics foundations perspective, thereby requiring the most general FSMs possible, as discussed in Sec. 1.3. We also avoid many of the typical assumptions, such as stationarity or unifilarity, as they all would correspond to artificial constraints to the set of temporal correlations.

temporal correlations in the simplest scenario possible: a single system being repeatedly measured by a fixed instrument, producing dichotomous outcomes.

We formulate this problem in terms of individual sequences of outputs $\boldsymbol{a} = a_1 a_2 \ldots a_L$, for $a_t \in \{0, 1\}$ and $t = 1, \ldots, L$, studying the maximum probabilities $p(\boldsymbol{a}|d)$ that can be obtained by classical and quantum d-dimensional systems. Since this maximum probability is trivially 1 if d is sufficiently large, we introduce the *deterministic complexity* (DC) for quantifying this memory threshold. The DC is a computable complexity measure for output sequences, rooted on physical principles, and we provide an efficient algorithm to compute it. Since $p(\boldsymbol{a}|d) = 1$ is only achievable by classical or quantum systems if $d \geq \mathrm{DC}(\boldsymbol{a})$, differences between the two will only appear if $d < \mathrm{DC}(\boldsymbol{a})$, with each theory providing a nontrivial maximum for this probability. Thus, these "sub-deterministic" scenarios provide a sharp boundary between classical and nonclassical temporal correlations.

We investigate these maximum probabilities by means of gradient descent techniques, optimizing over the space of classical and quantum FSM models for every sequence of length up to $L = 10$. While the resulting optimal models only give a lower bound for the maxima, they still provide useful insights into how a finite memory resource can be used optimally within classical and quantum theory. Indeed, several nontrivial results emerged from this numerical survey. In particular, a family of sequences, which we call *one-tick sequences*, appear to have special significance, and we conjecture they may be used to upper-bound all other sequences. If proven, this conjecture implies an even more remarkable result: the existence of a universal upper bound of $1/\mathrm{e} \approx 37\%$ on the probability of *any* sequence in a sub-deterministic classical scenario. In contrast, we show that no nontrivial universal bound seems to exist in the quantum case.

The maximum probabilities for sub-deterministic scenarios can be used to construct dimension witnesses. This is the subject of Chapter 3, where we develop witnesses for the minimum dimension of an environment interacting with an open quantum system. We consider a scenario where a small "probe" system is measured after each interaction with its environment. With the system always being discarded and reprepared before each interaction, we can focus on the environment acting as a quantum memory resource. Thus, the resulting temporal correlations, obtained by measuring the probe, will be fundamentally restricted by the dimension of its environment.

For simplicity, we assume the system-environment interaction always occurs through the same joint unitary at every time step, which can be reasonably justified on physical grounds. As in Ch. 2, here we focus on the maximum probabilities of individual sequences of outcomes. To obtain upper bounds, we develop a hierarchy of semidefinite programs (SDPs) that are guaranteed to converge to the exact maximum. The maximum probability can be obtained by a global optimization over all unitaries, which is a problem that cannot be solved directly as an SDP. Using the formalism of quantum supermaps and the quantum de Finetti theorem, we were able to formulate semidefinite relaxations of this global optimization problem.

Nevertheless, the resulting SDP is numerically intractable even in the simplest nontrivial scenario, i.e., involving a qubit system and environment, and three time steps. More precisely, the dense SDP would require 3 TB of RAM to be solved with standard algorithms, even after being written in its most compact symmetric representation. To overcome this, we developed a novel heuristic method that converts the dense problem into an equivalent sparse problem, with less than 1% of the original number of variables and constraints. With this sparsity heuristic, we were able to compute upper bounds on quantum temporal correlations for select cases. Our heuristic method to obtain sparse SDPs is very general, and could conceivably be applied to many large-scale SDPs, both in quantum information and beyond. The technique is currently being developed further, with a dedicated manuscript under preparation. This is the subject of Chapter 5.

Beyond the Leggett-Garg approach, alternative notions of *classicality* for temporal correlations have

largely been based on flawed analogies from spatial correlations. In particular, this has been the case in the study of memory effects (i.e., non-Markovianity) in multi-time quantum processes, as analyzed through the formalism of quantum supermaps [135, 72, 14, 190]. Within this formalism, the propagation of the memory state from one time step to the next appears explicitly as quantum channel, with many authors assuming that the inclusion of a generic entanglement-breaking (EB) channel would result in a classical memory, and therefore, classical temporal correlations.

In Chapter 4 we show that such a claim is not generally accurate, by providing explicit examples where the action of a generic EB channel still leads to nonclassical temporal correlations. In particular, we show that a d-dimensional quantum memory under the action of an EB channel can still generate nonclassical correlations, as compared to the equivalent d-dimensional classical memory. The discrepancy is subtle, but we were able to uncover it for several reasons. First, FSMs can characterize classical temporal correlations *exactly*, as they describe correlations in terms of *manifestly classical* systems, i.e., not based on quantum representations being artificially restricted. A precise distinction between classical and quantum memory effects should emerge from explicit constructions within each theory, and to the best of our knowledge, we are the first to accurately characterize classical correlations in this manner. Secondly, the nonclassicality only seems to occur beyond qubits and in temporal correlations involving four or more time steps, both being scenarios which have received very little attention in previous research.

Despite being nonclassical, the resulting correlations still appear intimately related to classical correlations in unexpected ways. In particular, the maximum probability for quantum memories under the effect of an EB channel quickly approaches the universal classical bound of $1/\mathrm{e}$ (as discussed in Chapter 2), even violating it, but only by a tiny margin. This puzzling result is still unexplained.

The FSM framework also provides a bridge between temporal correlations and theoretical computer science. This is explored in Chapter 6 where the DC, originally introduced for sequences of outputs, is generalized for sequential measurements involving both inputs and outputs, thereby providing a concrete characterization of the memory cost of realizing any extreme point of the Arrow-of-Time polytope. We further establish connections between the DC and the problem of *maximum independent set* in graph theory, which enables the generalization of DC for scenarios where certain deterministic outputs can be ignored. This leads to the introduction of the partial DC (PDC) and conditional DC (CDC), both of which have interesting parallels with existing complexity measures in computer science.

We conclude the thesis with an overall discussion on the underappreciated subject of temporal correlations and how the framework of finite-state machines provides a new way forward for their investigation.

Chapter 1

Preliminary notions

1.1 Memory and information processing tasks

Since the notion of "memory" as a finite physical resource is a central theme underlying this thesis, our working definition of the term should be made precise.

Any information processing task can be understood as the storage, manipulation, and retrieval of information over a period of time. We informally conceptualize the capability of storing information with the notion of a "memory", i.e., a system capable of keeping track ("remembering") its internal state, with this state being broadly understood as the information "stored" in the memory[1].

These notions of information and memory are familiar to us, at least within classical theory, but its physical underpinnings are usually shrouded by computational abstractions. A bit, the fundamental unit of information in digital computers, stands for a memory unit with two distinct states: 0 or 1, *on* or *off*, etc. While a bit offers very little in terms of information storage (in fact the minimum amount possible), it is also the simplest memory to construct, thus enabling us to have many bits acting together as a single memory device. In this way, a memory with n bits has a total of $d = 2^n$ distinct *memory states* at our disposal for storing information. Nevertheless, we usually understand n as the "amount of memory" in such devices.

It is also generally understood that memory is an essential resource for performing any information processing task, with more complex tasks becoming feasible as more memory is made available. Conversely, it is natural to posit that a finite amount of memory imposes fundamental limitations on what sort of tasks are possible. By studying how these limitations emerge when memory is considered a finite physical resource, we may thus obtain insights into how it can be used optimally. At the same time, such memory-restricted scenarios can highlight fundamental differences between physical theories, arising from their different descriptions of how information can be stored and manipulated.

In the following chapters, these ideas will be explored from a physics foundations perspective, within both quantum and classical theories. To this end, we shall treat the memory as a single *indivisible* physical system with a finite number d of perfectly distinguishable states. We will adopt d, not $n = \log_2 d$, as the "amount" or "size" of the memory, i.e., the total dimension of the memory's state space. This will correspond to the dimension of the Hilbert space of a quantum system (a qudit), or the number of distinct states in a classical system (which we call a "d-it"[2]). The notion of a "memory resource" will correspond not only to the quantity d, but also the classical or quantum properties of the memory, i.e., the possible memory states and the physical transformations available on such states.

Note that this presents a slight conceptual departure from the usual notion of "amount of memory"

[1] We will refrain from any deeper discussion on what constitutes "information", however.

[2] We adopt "d-it" in analogy to "qudit", as "dit" is already commonly used for $d = 10$.

in terms of bits (or qubits). In practice, since each bit requires a finite amount of physical resources to be implemented, and bits are nearly always utilized in very large groups, the quantity d is rarely insightful. It is usually more practical to describe the amount of memory or information in terms of the number of bits (or its multiples, bytes, kilobytes, etc.), leading to the more familiar notions of "amount of memory" (i.e., $n = \log_2 d$) and "information" (e.g., information content $-\log_2 p$ for a probability p), both of which quantify the resources required when *encoding* information in terms of bits.

This "encoding" of information into bits is a convenient abstraction, largely motivated by engineering concerns, but it also implicitly assumes a notion of spatially separated bits (or qubits) and operations acting jointly upon them. Since we wish to focus on strictly temporal phenomena, involving a single indivisible system, this perspective is not really helpful, if not a little misguided[3]. In any case, as we shall see, the quantity d arises naturally in the sequential tasks we will be considering in this thesis.

1.2 Temporal correlations

Quantum mechanics predicts the existence of correlations beyond any correlation achievable by classical systems. While this phenomenon has been extensively studied in the case of spatial correlations [62, 29], the equally important subject of temporal correlations has historically received significantly less attention. To make matters worse, previous research on the subject has predominantly focused on definitions of "classicality" which are arguably too restrictive to be insightful about the differences between classical or quantum temporal effects [26, 203].

This thesis contributes to a recent research effort seeking a more definitive characterization of temporal correlations—whether classical, quantum, or beyond—by investigating the probability distributions resulting from sequential measurements on *finite-dimensional* physical systems. Importantly, these measurements are considered in their most general form as permitted by each theory, avoiding artificial constraints that would otherwise be imposed on the resulting correlations.

This section provides an overview of the subject of temporal correlations and how their characterization has been pursued so far. First, we briefly review the traditional approach based on Leggett-Garg inequalities, highlighting its shortcomings. Detailed reviews on the broader subject surrounding Leggett-Garg inequalities, such as tests of macrorealism or applications of the inequalities to temporal correlations, can be found in Refs. [64, 203]. Following this, we discuss the more recent approach involving finite-dimensional systems [219, 26] and its formulation in terms of finite-state machines [88, 32, 182], which are the basis for the remaining of this thesis.

1.2.1 Macroscopic realism and Leggett-Garg inequalities

Early investigations into temporal correlations can be traced back to the work of Leggett and Garg [116], where correlations arising from repeated measurements on a single system were used for investigating the existence of macroscopic quantum coherence. An extrapolation of quantum theory to the macroscopic scale suggests such phenomenon should exist, in great contradiction with our everyday experience. This seemingly absurd prediction of quantum theory was already remarked by Schrödinger in his burlesque "Schrödinger's cat" thought experiment [171].

The issue arises because macroscopic (i.e., "classical") systems appear to exist in well-defined states at any given moment, with outcomes of any measurements seemingly preexisting prior to (and independent of) the measurements. This notion is typically referred to as *realism*[4], and captures the expectation that physical states should provide a complete description of a system.

[3] While the extra $\log_2$ is mathematically inconsequential, in our context it is artificial and meaningless.

[4] The notion of "realism" in physics is more subtle than this, however; see Refs. [150, 108, 73].

While reasonable at the macroscopic scale, at the microscopic scale this notion of realism has been challenged by quantum theory in various ways, e.g., with the early observation that certain physical quantities, such as position and momentum, cannot be jointly measured to arbitrary precision [87, 172], or the necessity of contextuality in any hidden-variable model of quantum theory, as shown by Bell [12] and Kochen and Specker [175, 35]. This quantum indeterminacy is now recognized as a fundamental aspect of quantum theory. That being said, nothing in the theory suggests a scale where these properties ought to disappear.

Leggett and Garg approached this conundrum by proposing a minimal set of reasonable properties expected from any theory compatible with our intuitions about macroscopic systems. Such *macrorealist theories*, as they have been called, were defined in terms of two assumptions which Leggett and Garg described as [116]:

MR Macroscopic realism: A macroscopic system with two or more macroscopically distinct states available to it will at all times be in one or the other of these states.

NIM Noninvasive measurability: It is possible, in principle, to determine the state of the system with arbitrarily small perturbation on its subsequent dynamics.

The "subsequent dynamics" in NIM was later recognized as containing the implicit assumption of causality, now properly regarded as a third assumption [103, 115, 67, 64, 203]:

IND Induction: The outcome of measurements in the present is not influenced by future choices of measurements.

Leggett and Garg derived a class of inequalities which would be satisfied by any theory obeying these assumptions, and proposed experiments to test them. An irrefutable experimental violation of these *Leggett-Garg inequalities* (LGIs) would force us to abandon any realist description of the system at a macroscopic scale.

The simplest LGI can be derived from the experimental scenario shown in Fig. 1.1, where a single system is randomly chosen to be measured in two out of three possible times t_i, for $i \in \{1, 2, 3\}$, via an observable $Q(t_i)$ with dichotomous outcomes $q_i = \pm 1$. Over multiple runs of this experiment one obtains the empirical distributions $p_{ij}(q_i, q_j)$ for the pairs $(i, j) \in \{(1, 2), (1, 3), (2, 3)\}$, which can be used to define the two-time correlators

$$\langle Q(t_i)Q(t_j)\rangle := \sum_{q_i, q_j = \pm 1} q_i q_j \, p_{ij}(q_i, q_j). \tag{1.1}$$

With this, we may write the Leggett-Garg inequality,

$$\langle Q(t_1)Q(t_2)\rangle + \langle Q(t_2)Q(t_3)\rangle - \langle Q(t_1)Q(t_3)\rangle \leq 1, \tag{1.2}$$

which is satisfied for any system obeying the three macrorealist assumptions. To see how, first note that under the MR assumption each $Q(t_i)$ has a well-defined value, implying the existence of a joint probability distribution $p_{123}(q_1, q_2, q_3)$ describing the predetermined outcomes of each $Q(t_i)$ in any run of the experiment.

Next, the NIM and IND assumptions posit that past measurements—regardless of whether they have been performed or not—cannot influence future ones, meaning the result of any measurement can be safely discarded without changing the remaining statistics. This implies we can define the two-time distributions $p_{ij}(q_i, q_j)$ as marginals of a global joint distribution $p_{123}(q_1, q_2, q_3)$,

$$p_{ij}(q_i, q_j) = \sum_{q_k;\, k \neq i,j} p_{123}(q_1, q_2, q_3), \tag{1.3}$$

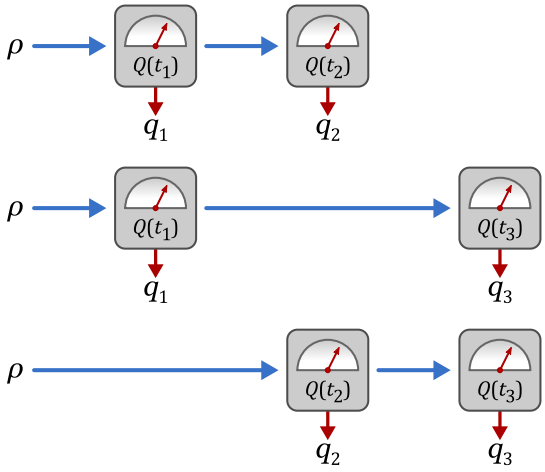

Figure 1.1: The basic Leggett-Garg scenario involves time-dependent measurements $Q(t)$ to be performed at three times $t_1 < t_2 < t_3$, obtaining outcomes $q_i = \pm 1$ for $i \in \{1, 2, 3\}$. For each run of the experiment, the system is prepared in the state ρ, then measurements are performed sequentially in only two time steps $i, j \in \{1, 2, 3\}$, with $i < j$.

such that the two-time correlators can be rewritten as

$$\langle Q(t_i)Q(t_j)\rangle = \sum_{q_k;\, k\neq i,j} \sum_{q_i,q_j} q_i q_j\, p_{123}(q_1, q_2, q_3). \tag{1.4}$$

Using Eq. (1.4), we can then write

$$\begin{aligned} K &= \langle Q(t_1)Q(t_2)\rangle + \langle Q(t_2)Q(t_3)\rangle - \langle Q(t_1)Q(t_3)\rangle \\ &= \sum_{q_1,q_2,q_3} (q_1q_2 + q_2q_3 - q_1q_3)\; p_{123}(q_1, q_2, q_3) \\ &\leq \max_{q_1,q_2,q_3} (q_1q_2 + q_2q_3 - q_1q_3) = 1, \end{aligned} \tag{1.5}$$

which establishes the inequality. Here, we have used the fact the q_i can be chosen independently, and that $p_{123}(q_1, q_2, q_3) \geq 0$ and $\sum_{q_1,q_2,q_3} p_{123}(q_1, q_2, q_3) = 1$. This LGI is easily violated in quantum mechanics, e.g., by using a two-level quantum system undergoing coherent oscillations [64, 203], which achieves a maximum value of $K = 3/2$.

While LGIs were designed as tests for *macro*realism, experiments are still largely confined to *micro*scopic scales, with considerable effort being directed towards devising better and larger-scale LGI tests [64]. Nevertheless, some authors have pointed out that there is nothing inherently *macroscopic* about the assumptions underlying LGIs [125, 64]. Since the notion of "macroscopicity" will be irrelevant to us, we shall simply refer to realism as the relevant assumption from now on.

A full discussion on LGIs and macrorealism is beyond the scope of this thesis; see Refs. [64, 203] for detailed reviews instead. For now, we simply note LGIs are also subject to many experimental loopholes, similar to those in Bell inequalities, but also much more challenging to address; see Refs. [29, 64, 203]. As we will see next, some of these loopholes are intimately related to the temporal character of LGIs and the fundamental differences between spatial and temporal correlations.

1.2.2 Temporal vs. spatial correlations

Beyond macrorealism, physicists quickly noted that LGIs could also be adopted as tests for the "nonclassicality" of temporal correlations. This is partially inspired by the fact LGIs share the same mathematical structure as Bell inequalities [64], which establish bounds on *spatial* correlations. Indeed the two inequalities are closely related, with LGIs even being referred to as "temporal Bell inequalities" [152]. However, such a direct analogy overlooks fundamental differences between spatial and temporal correlations in several important ways.

On the one hand, Bell inequalities treat measurements as happening on spatially separated systems, and are built upon the notions of realism and locality [150, 13, 29]. In experimental Bell tests, locality can be ensured by an appropriate spatial separation between the two systems such that no signaling is possible between measurements. Any experiment failing to do so would be subject to the "communication loophole" [50, 109], also known as "locality loophole", where violations of the inequality could be attributed to some form of signaling having occurred.

On the other hand, LGIs consider measurements on a single system at multiple moments in time, and are built upon the notions of realism and noninvasive measurability. A valid test of LGIs requires careful experimental control to ensure measurements are not invasive, as otherwise the experiment would be subject to the "clumsiness loophole" [210], and any violations of the LGI could be attributed to inadvertently invasive measurements having occurred.

Both loopholes can be understood in terms of some form of signaling occurring between measurements. In fact, the respective no-signaling conditions are essentially identical from a mathematical standpoint. It is instructive to make this comparison explicitly.

For the Bell inequality we can consider Clauser-Horne-Shimony-Holt (CHSH) scenario [51], with Alice and Bob in two space-like separated laboratories. Alice and Bob may each independently choose between two possible measurements, $\{A_x\}_x$ for settings $x = \{0, 1\}$ in Alice's case, and $\{B_y\}_x$ for settings $y = \{0, 1\}$ in Bob's case, with corresponding outcomes given by $a_x, b_y = \pm 1$. Defining the correlators

$$\langle A_x B_y \rangle := \sum_{a_x, b_y = \pm 1} a_x b_y \, p(a_x, b_y), \tag{1.6}$$

we may write the corresponding Bell-CHSH inequality as

$$\langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle \leq 2. \tag{1.7}$$

Under the assumptions of realism and locality, there exists [66] a global joint distribution, given by $p(a_0, a_1, b_0, b_1)$, such that

$$p(a_x, b_y) = \sum_{\substack{a_{x'}, b_{y'} = \pm 1 \\ x' \neq x \text{ and } y' \neq x}} p(a_0, a_1, b_0, b_1). \tag{1.8}$$

The marginals in Eq. (1.8) embody the No-Signaling (NS) condition between Alice and Bob, demanded by locality, which informs us that both Alice's and Bob's outcomes must be statistically independent from any measurements performed by the other party.

A comparable LGI can be constructed using four sequential measurements, but is otherwise similar to Eq. (1.2) derived from Fig. 1.1:

$$\langle Q(t_1) Q(t_2) \rangle + \langle Q(t_2) Q(t_3) \rangle + \langle Q(t_3) Q(t_4) \rangle - \langle Q(t_1) Q(t_4) \rangle \leq 2. \tag{1.9}$$

As before, the LGI assumptions imply the existence of a global joint distribution $p_{1234}(q_1, q_2, q_3, q_4)$, such that

$$p_{ij}(q_i, q_j) = \sum_{\substack{q_k, q_\ell = \pm 1 \\ k, \ell \neq i, j \text{ and } k \neq \ell}} p_{1234}(q_1, q_2, q_3, q_4), \tag{1.10}$$

meaning measurements at times t_i and t_j are unaffected by the other two, i.e., there is no signaling occurring between measurements performed at different times.

At first glance, the similarities between Eqs. (1.8) and (1.10) suggest Bell and Leggett-Garg inequalities are analogous. While the two make use of realism for their notions of classicality, their fundamental difference arises when considering the physical basis for the lack of signaling.

In the temporal case of LGI, the existence of a joint distribution $p_{1234}(q_1, q_2, q_3, q_4)$ is already guaranteed by the realism and NIM assumptions, and could be obtained by directly measuring at the four times. The spatial case of the Bell-CHSH inequality differs in that the joint distribution $p(a_0, a_1, b_0, b_1)$ cannot be obtained directly: since A_0 and A_1, or B_0 and B_1, are incompatible measurements, it is impossible to perform the necessary joint measurements [203]. Instead, under realism and locality such a distribution would arise from an appropriate local hidden-variable (LHV) model [66], e.g.,

$$p(a_0, a_1, b_0, b_1) = \sum_{\lambda} p(\lambda)\, p(a_x|x;\lambda)\, p(b_x|y;\lambda), \tag{1.11}$$

which embodies the locality constraint, thereby automatically satisfying the spatial no-signaling conditions in the resulting global distribution.

The principle of locality is motivated by the postulates of special relativity, and is generally regarded as a fundamental property of Nature. In contrast, NIM is not based on any fundamental physical principle, instead being simply an intuitive property expected of macroscopic classical systems. This discrepancy suggests Bell and Leggett-Garg tests should not be considered methodologically comparable [126], despite their mathematical similarities.

Thus, while closing the communication loophole in Bell inequalities can be achieved in a model-independent way, in terms of fundamental physical postulates, closing the clumsiness loophole in LGIs is not straightforward. Much of the literature surrounding LGIs has been dedicated to closing the clumsiness loophole by means of increasingly sophisticated protocols [64, 203]. Furthermore, while numerous loophole-free Bell tests have been proposed [29], even the mere possibility of a loophole-free Leggett-Garg test remains an open question [64, 203]. As a final remark, additional loopholes exist for both Bell and Leggett-Garg inequalities (see Refs. [64, 109, 203]), but they are not relevant for this discussion.

1.2.3 Temporal correlations and Leggett-Garg inequalities

Despite these challenges, LGIs have become the standard notion of nonclassical temporal correlations [203], with "classicality" largely being used interchangeably with the notion of noninvasive measurability (with realism left implicit). Indeed, the same assumption underlies the applications of the Kolmogorov extension theorem [104, 24] in defining classical stochastic processes and "classical memory" in a quantum mechanical setting [134, 133, 189, 169, 185, 160]. In this thesis we explicitly reject this notion of classicality, instead treating realism, in and of itself, as the defining feature of classicality in a temporal context.

To justify this choice, let us step back and review the three assumptions behind LGIs. We ask: What are the minimal assumptions expected of a general classical description of temporal correlations?

First and foremost, the IND assumption seems essential in the description of any temporal phenomenon, as it embodies the notion of causality observed in Nature, and applies equally to the case of quantum systems[5]. In what follows, we simply refer to causality as the explicit assumption.

As noted earlier, MR is essentially just realism, and is the same feature of classical mechanics invoked in Bell and noncontextuality [12, 175] inequalities. The fundamental indeterminacy in quantum theory, ultimately resulting from coherences, already provides a clear criterion to distinguish between classical and quantum theory in sequential temporal scenarios.

[5] Of course, the IND assumption is violated in models involving exotic causal structures, but their existence is disputed. This is a topic beyond the scope of this thesis, however.

The NIM assumption is the one that clearly stands out from the rest. The notion that classical measurements can always "in principle" be made without any disturbance seems of little use *in practice*, as there appears to be no irrefutable way to achieve this in the case of a single indivisible system, i.e., in a manner similar to locality in a Bell test by means of space-like separated measurements.

Furthermore, while measurement invasiveness is recognized as a fundamental feature of quantum mechanics, invasiveness itself is not necessarily a uniquely quantum effect: It is possible to devise models of classical invasive measurements [32, 160]. Since the presence of invasiveness can easily violate LGIs, the NIM assumption will invariably fail to discern between classical and quantum temporal correlations.

We are forced to conclude NIM is an unreasonably strong assumption, if not an unnecessary one, and several relaxations of the assumption have been proposed which have led to far more insightful results. For the remaining of this thesis all measurements (classical or quantum) will be allowed to be invasive.

1.2.4 Nonclassical temporal correlations beyond Leggett-Garg

Clearly, if future measurements can be influenced by earlier ones then information must have been relayed between them, and the system itself must be the carrier of this information. This naturally motivates an information processing perspective for temporal correlations: The system is understood as a *memory resource* (see Sec. 1.1), which is being used for the task of *generating* temporal correlations.

From this point of view, invasiveness need not be associated with an "unwanted disturbance" of the state, which we attempt to eliminate. Instead, it can be interpreted as the deliberate manipulation of information stored in a memory to achieve a specific task. Since each physical theory provides different descriptions of what memory states and their transformations are possible, their inherent limitations in achieving certain tasks can be utilized for distinguishing between theories. This naturally motivates treating *finite memory* as a constraint.

Different notions of nonclassical temporal correlations have recently been introduced based on similar principles. An algorithmic understanding of sequential operations was first proposed by Żukowski [219], where the assumption of a finite-dimensional system was employed in conjunction with a sequential information processing task for obtaining temporal inequalities capable of distinguishing between classical and quantum systems.

The application of finite-dimensional systems for investigating temporal correlations was later refined by Hoffmann et al. [88] and Budroni et al. [32], where the system's dimension was interpreted as the *memory cost* of generating temporal correlations, formalized through the framework of finite-state machines (FSMs). In particular, Budroni et al. [32] used FSMs for investigating temporal correlations in general probability theories (GPTs), which include classical and quantum theories as special cases. Within a unified framework, the authors were able to derive temporal inequalities capable of discriminating between classical, quantum, and genuine GPT correlations. These results highlight the versatility of the FSM framework and its capability of sharply characterizing temporal correlations. The framework has also been employed in investigating quantum advantages for time-keeping devices [33].

A related notion of nonclassical temporal correlations was introduced by Brierley et al. [26] in terms of communication cost, as opposed to memory cost. Correlations are assumed to be generated by a d-dimensional quantum system being sequentially measured by multiple parties, separated in both space and time. A correlation is said to be nonclassical if the parties cannot simulate it by communicating using at most a d-it of classical information between them. Nonclassical temporal correlations have also been investigated in a variety of other contexts, such as quantum random-access codes [209, 9, 10, 191, 131], dimension witnessing [28, 5, 20], and classical simulation of quantum contextuality [102, 65].

A unifying feature in all these approaches is the assumption of a finite-dimensional system. As we discuss in the following, this assumption emerges naturally in sequential temporal scenarios.

1.2.5 Inputs, outputs, and symbolic sequences

This thesis investigates temporal correlations arising from finite and discrete sequences of measurements on finite-dimensional systems. Aiming for a more abstract perspective in terms of information processing tasks, measurement settings will be referred to as "inputs" and measurement outcomes as "outputs".

Inputs and outputs will be labeled by symbols from the sets $\mathcal{X}$ and $\mathcal{A}$, respectively, which we refer to as *alphabets*, with sizes denoted by $X = |\mathcal{X}|$ and $A = |\mathcal{A}|$. In Chs. 2 to 4, unless stated otherwise, we take $\mathcal{A} = \{0, 1\}$ and $\mathcal{X} = \{0\}$ (i.e., trivial inputs), with Ch. 6 addressing the general $\mathcal{X}$ and $\mathcal{A}$ case.

A sequence of symbols $a_1 a_2 \dots a_L$ from an alphabet $\mathcal{A}$ will be denoted in bold, e.g. $\boldsymbol{a}$, with $|\boldsymbol{a}| = L$ the sequence length. The set of all sequences of length L will be denoted by $\mathcal{A}^L$, and the set of all sequences of lengths between ℓ and L by $\mathcal{A}^{\ell:L}$. We use $\mathcal{A}^* := \mathcal{A}^{0:\infty}$. It will also be convenient to specify portions of a sequence using the subscript notation $\boldsymbol{a}_{\ell:k} = a_\ell a_{\ell+1} \dots a_{k-1} a_k$, and concatenation of sequences by juxtaposition, so that $\boldsymbol{a} = \boldsymbol{a}_{1:\ell-1}\boldsymbol{a}_{\ell:L}$. Repetitions will be denoted with superscripts, e.g., $0^3 1 = 0001$, and the empty sequence will be denoted by ε.

1.2.6 The Arrow-of-Time constraints

The term *temporal correlations* will generally refer to the probability distributions $p(\boldsymbol{a}|\boldsymbol{x})$ obtained from repeated measurements, i.e., the distributions over outputs $\boldsymbol{a} \in \mathcal{A}^L$ conditioned on inputs $\boldsymbol{x} \in \mathcal{X}^L$ for a fixed length L, such that:

$$p(\boldsymbol{a}|\boldsymbol{x}) \geq 0, \quad \text{and} \quad \sum_{\boldsymbol{a}\in\mathcal{A}^L} p(\boldsymbol{a}|\boldsymbol{x}) = 1, \quad \text{for all } \boldsymbol{x} \in \mathcal{X}^L. \tag{1.12}$$

In the case of no inputs, we simply refer to $p(\boldsymbol{a})$.

The essential assumption of causality requires that all temporal correlations must satisfy the Arrow-of-Time (AoT) constraints [52]. For $L = 3$, they can be written as:

$$\begin{aligned} \sum_{a_2,a_3} p(a_1a_2a_3|x_1x_2x_3) &= \sum_{a_2,a_3} p(a_1a_2a_3|x_1x_2'x_3'), \\ \sum_{a_3} p(a_1a_2a_3|x_1x_2x_3) &= \sum_{a_3} p(a_1a_2a_3|x_1x_2x_3'), \end{aligned} \tag{1.13}$$

for all $x_1, x_2, x_2', x_3, x_3' \in \mathcal{X}$ and $a_1, a_2, a_3 \in \mathcal{A}$. The $L > 3$ case is analogous, leading to a larger set of equalities. The constraints in Eq. (1.13) specify that, at every time step, future inputs cannot influence past outputs, i.e., signaling is forbidden from future to the past. As such, the AoT constraints apply regardless of the underlying theory describing the correlations.

In this way, the AoT constraints are analogous to the No-Signaling conditions in the case of spatial correlations [29], with the important difference that signaling is only forbidden in one direction. This distinction leads to stark differences between spatial and temporal correlations, as we discuss in the following.

1.2.7 The Arrow-of-Time polytope

The AoT constraints in Eq. (1.13) imply we can decompose temporal correlations by recursively conditioning the probabilities at each time step on their respective past inputs and outputs, e.g.,

$$\begin{aligned} p(a_1a_2a_3|x_1x_2x_3) &= p(a_1|x_1)\, p(a_2a_3|x_1x_2x_3, a_1) \\ &= p(a_1|x_1)\, p(a_2|x_1x_2, a_1)\, p(a_3|x_1x_2x_3, a_1a_2), \end{aligned} \tag{1.14}$$

where we assume $p(a_2a_3|x_1x_2x_3, a_1) := p(a_1a_2a_3|x_1x_2x_3)/p(a_1|x_1)$ is zero if $p(a_1|x_1) = 0$, and similarly for other time steps. To make this temporal conditioning explicit, we adopt the convention that semicolons

separate present from past conditioning parameters. This *Arrow-of-Time decomposition* is then

$$p(a_1a_2a_3\ldots|x_1x_2x_3\ldots) := p(a_1|x_1)\,p(a_2|x_2;x_1,a_1)\,p(a_3|x_3;x_1x_2,a_1a_2)\cdots, \tag{1.15}$$

where each term in the decomposition can be understood as a *local-in-time* distribution [124]. Equation (1.15) not only makes the past-to-future signaling structure explicit, but also hints at the role of the physical system as a memory resource storing information about the past inputs and outputs.

Any correlation $p(\boldsymbol{a}|\boldsymbol{x})$ obeying the AoT constraints can be written as in Eq. (1.15) by an appropriate choice of local-in-time distributions. Since these distributions can be chosen to be deterministic (i.e., probabilities 0 or 1), and the AoT constraints are linear, we may also decompose any correlation satisfying the AoT constraints as a convex mixture of deterministic local-in-time distributions $\hat{p}_\lambda$:

$$\begin{aligned} p(a_1a_2a_3\ldots|x_1x_2x_3\ldots) &= \sum_i q(\lambda)\,\hat{p}_\lambda(a_1|x_1)\,\hat{p}_\lambda(a_2|x_2;x_1,a_1)\,\hat{p}_\lambda(a_3|x_3;x_1x_2,a_1a_2)\cdots \\ &= \sum_\lambda q(\lambda)\,f_\lambda(a_1a_2a_3\ldots|x_1x_2x_3\ldots) \end{aligned} \tag{1.16}$$

with $q(\lambda) \geq 0$ and $\sum_\lambda q(\lambda) = 1$, and f_λ denoting the resulting deterministic multi-time distribution. Equation (1.16) reveals that the set of correlations obeying the AoT constraints forms a convex polytope[6], known as the *Arrow-of-Time polytope* [52], with the deterministic distributions

$$f_\lambda(a_1a_2a_3\ldots|x_1x_2x_3\ldots) = \hat{p}_\lambda(a_1|x_1)\,\hat{p}_\lambda(a_2|x_2;x_1,a_1)\,\hat{p}_\lambda(a_3|x_3;x_1x_2,a_1a_2)\cdots, \tag{1.17}$$

corresponding exactly to its extreme points [1, 88].

The extreme points of the AoT polytope have only recently been characterized [1, 88, 182]. In particular, Spee at al. [182] has provided a detailed analysis of the symmetries of the AoT polytope, and introduced a criterion determining the minimum dimension required for the realization of its extreme points. These notions will be explored in more detail in Chs. 2 and 6.

As a final remark, since the AoT constraints concern only future *inputs* affecting the past, they do not fully capture the causal constraints in temporal correlations. This is especially evident in scenarios involving only outputs, where the AoT decomposition of $p(\boldsymbol{a})$ is simply

$$p(\boldsymbol{a}) = p(a_1)\,p(a_2|a_1)\,p(a_3|a_1a_2)\cdots p(a_L|a_1\ldots a_{L-1}), \tag{1.18}$$

which can always be written for an arbitrary joint distribution, i.e., any temporal ordering for the conditioning could have been chosen. This highlights the important role of the system as a carrier of past-to-future signaling information, something not apparent when looking strictly at the level of input-output distributions. A fully device-independent characterization of temporal correlations is therefore not generally possible, and further assumptions must be made about the system. As we have seen, the dimension of the system is a natural choice.

1.2.8 Dimension as a memory resource

Equation (1.16) also reveals that all temporal correlations in the AoT polytope can be generated by either a classical or quantum system, provided enough memory is available to store the relevant information about past inputs and outputs. This is a fundamental difference between temporal and spatial correlations, as even quantum theory is unable to generate all correlations in the spatial analogue of the AoT polytope, the No-Signaling polytope [48, 29].

[6] There are conflicting definitions of "polytope" in the literature as to whether boundedness or convexity are implied or not. In this thesis we assume polytopes are bounded.

It was quickly realized that the appropriate quantifier of memory for sequential tasks coincides with dimension of the system [88, 32], i.e., the maximum number of perfectly distinguishable states. Furthermore, since classical and quantum theory require the same dimension to realize a given extreme point of the AoT polytope [180], the classicality or nonclassicality of temporal correlations can only be investigated when resources are appropriately constrained [88, 32]. This motivates the question:

What are the *minimum* memory resources to physically realize $p(\boldsymbol{a}|\boldsymbol{x})$?

Here, the memory resource is to be understood not only as the dimension (the "amount" of memory), but also its classical or quantum characteristics; see Sec. 1.1. Broader notions of a "resource" are also possible by considering, e.g., time-dependent operations, initial randomness [32, 182], mixed initial states [55], or even conditioning on past inputs and outputs via a classical side-channel [124]. All of these can provide an additional advantage in generating temporal correlations[7]. That being said, this thesis focuses on scenarios without any of these additional resources, i.e., the system is the sole resource available to generate correlations. This approach allows classical and quantum memories to be compared directly on equal terms.

Nevertheless, exactly determining the minimum resources needed for arbitrary temporal correlations appears to be exceptionally difficult, due to the non-convexity of the set of correlations under finite-memory constraints [32, 124] and the general non-linearity which arises from sequential operations.

A good starting point is to focus on the extreme points of the AoT polytope, i.e., the deterministic distributions f_λ (Eq. (1.17)), where storing all past inputs and outputs is clearly not necessarily optimal. Spee at al. [182] introduced a criterion for determining the minimum dimension needed to realize these extremal points, and also investigated the symmetries of the AoT polytope under classical post-processing. In this thesis, we refine and greatly expand upon these ideas by introducing the deterministic complexity (DC), a computable complexity measure relevant in temporal correlations. The DC was first introduced in Ref. [198] in the context of output sequences, which is the subject of Ch. 2. In Ch. 6, the DC is generalized for sequences of inputs and outputs, while also establishing some connections to topics in graph theory and computer science.

The takeaway from these results is that the dimension d directly quantifies both classical and quantum memory resources, in particular due to the fact deterministic quantum strategies do not require coherences [182], i.e., a d-it can simulate a qudit. This naturally motivates the application of finite-state machines to model finite memory and sequential operations in a unified way [88, 32], which is the framework used throughout this thesis.

1.3 Finite-state machines

This thesis investigates temporal correlations through the framework of finite-state machines (FSMs), an application pioneered by the works in Refs. [88, 102, 65, 32, 182, 124], which we build upon. This section briefly formalizes the notion of FSMs used within this context of temporal correlations, while also highlighting how it diverges from the conventional uses of FSMs found elsewhere.

Finite-state machines were originally introduced in computer science as mathematical models of computation where only a fixed and finite amount of memory is available, as quantified by its number of states. At each moment the machine is found in exactly one of its possible states. Upon receiving an input symbol $x \in \mathcal{X}$, the machine performs a transition between its internal states and produces an output

[7] It seems possible to quantify the amount of memory "smuggled" by each of these alternatives, thereby unifying all resources into one formalism. This will be the subject of future research.

symbol $a \in \mathcal{A}$ following a time-independent rule. Despite their simplicity FSMs are just as powerful as a typical computer[8], but their usefulness as a model of general computation is very limited[9].

Here, we employ FSMs to model physical devices capable of manipulating a single finite-dimensional (classical or quantum) system acting as its internal memory. We assume discrete time steps, instantaneous measurements and transitions, and no free time evolution between measurements[10]. Thus, all state transitions occur due to the effects of measurements performed on the memory state, with outcomes corresponding to the machine's output. We will further assume the machine is reset to an initial state before each sequence of measurements, meaning no assumptions is made about stationary behaviors.

Since we will consider probabilistic FSMs and their quantum generalizations, the notion of classical and quantum memory states and transition rules will differ between the two. We formalize the two cases in the following sections.

Automata

A FSM that does not accept any inputs can be thought of as machine acting autonomously as a *sequence generator*. We will call such a FSM a *finite-state automaton*, or simply automaton. Automata can be interpreted as a physical device with a trivial input (i.e., $X = 1$), understood as a simple "measure" instruction producing a single output (the measurement outcome). If all transitions (and thus outputs) occur deterministically, the machine is referred to as *deterministic finite-state* automaton (DFA).

By repeatedly resetting and running the automaton for any desired number of steps, we obtain a probability distribution $p(\boldsymbol{a})$ over all sequences $\boldsymbol{a} \in \mathcal{A}^L$, for each length L. The finite amount of memory in the automaton is the sole resource available to realize this distribution. Figure 1.2 provides an intuitive picture of an automaton with $\mathcal{A} = \{0, 1\}$.

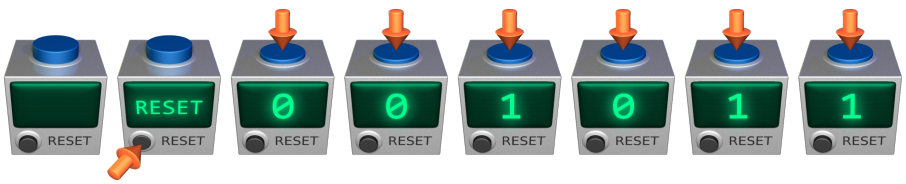

Figure 1.2: An automaton can be understood as a device with a finite amount of internal memory, an output display, a "reset" button, and a single "measure" button (on top). Each measurement produces an output while potentially changing the memory state. In this example, the sequence $\boldsymbol{a} = 001011$ was generated.

As a side note, while we consider automata exclusively as *generators* (i.e., only outputs), the reader might be more familiar with their usual applications in computer science and automata theory, which instead treat them as *acceptors*: machines taking a sequence of inputs and having states labeled as either "accepting" or "rejecting", with this label being read after the entire input sequence is processed. There are important connections and fundamental differences between the two approaches, but that discussion is beyond the scope of this thesis.

[8] A typical *physically real* computer has a finite number of inputs, outputs, and memory, thus a finite (but incomprehensibly large) number of states and transitions.

[9] It is difficult to think about algorithms in the "monolithic" picture of computation provided by FSMs, which is why Turing machines and other models are used instead [90]. Nevertheless, the approach does find important applications, e.g., in model checking [49], and in automatic parallelization of sequential programs [205].

[10] In principle, time evolution may still be included provided it is always given by the same channel at each step. It can then be directly embedded in the transition model.

Transducers

If the FSM has both inputs and outputs, it can be thought of as machine which sequentially translates input sequences into output sequences, one symbol at a time, a process known as *transduction*. Such a machine is referred to as a *finite-state transducer*, or simply transducer. In our context, an automaton is equivalent to a transducer with trivial inputs (i.e., $X = 1$). If all transitions (and thus input-output sequences) occur deterministically, the machine is referred to as *deterministic finite-state transducer* (DFT).

A transducer generates a conditional distribution $p(\boldsymbol{a}|\boldsymbol{x})$ over all possible inputs $\boldsymbol{x} \in \mathcal{X}^L$ and outputs $\boldsymbol{a} \in \mathcal{A}^L$, such that $\sum_{\boldsymbol{a}\in\mathcal{A}^L} p(\boldsymbol{a}|\boldsymbol{x}) = 1$ for every $\boldsymbol{x} \in \mathcal{X}^L$ and length L. The finite amount of memory in the transducer is the sole resource available to realize this distribution. Figure 1.3 provides an intuitive picture of a transducer with $\mathcal{X} = \mathcal{A} = \{0, 1\}$.

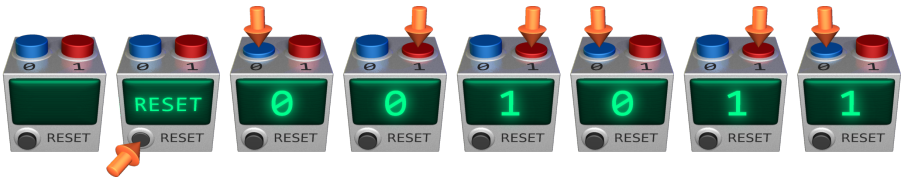

Figure 1.3: A transducer can be understood as a device with a finite amount of internal memory, an output display, a "reset" button, and multiple "measure" buttons (on top) corresponding to the possible inputs. Each measurement produces an output while potentially changing the memory state. In this example, the input sequence $\boldsymbol{x} = 011010$ resulted in the output sequence $\boldsymbol{a} = 001011$.

1.3.1 Classical finite-state machines

For classical FSMs we consider probabilistic Mealy-type (see Sec. 1.3.3) automata and weighted (or stochastic) transducers [151, 61], i.e., we assume outputs are generated through probabilistic transitions between states; see Sec. 1.3.3. We will refer to these simply as *probabilistic FSMs*[11].

The classical memory is a system which can be in one of d perfectly distinguishable states at any given time. Measurements upon this system lead to probabilistic state transitions and outcomes, such that the overall behavior of the device is stochastic. The FSM model we adopt for this device is defined as follows.

Definition 1. A *classical finite-state machine* is a tuple $(\mathcal{X}, \mathcal{A}, \mathcal{S}, s_0, \mathcal{T})$, where $\mathcal{X}$ is the input alphabet, $\mathcal{A}$ is the output alphabet, $\mathcal{S} := \{1, \ldots, d\}$ is a finite set of d states, s_0 is an initial "reset" distribution over $\mathcal{S}$, and $\mathcal{T} : \mathcal{X} \times \mathcal{S} \to \mathcal{A} \times \mathcal{S} \times [0, 1]$ is a stochastic transition rule.

Concretely, we specify $\mathcal{T}$ in terms of sub-stochastic *transition matrices* $T_{a|x} \in \mathbb{R}^{d\times d}$, such that $[T_{a|x}]_{ij}$ is the probability of the machine transitioning from state i to j while outputting a, when given an input x. We thus require that

$$[T_{a|x}]_{ij} \geq 0, \quad \text{and} \quad \sum_{a\in\mathcal{A}} \sum_{j=1}^{d} [T_{a|x}]_{ij} = 1 \quad \text{for all } x \in \mathcal{X}, i \in \mathcal{S}, \tag{1.19}$$

which enforces conservation of total probability. These matrices are collected into the classical *instruments* $T_x := (T_{a|x})_{a\in\mathcal{A}}$, giving the transition *model* $T := (T_x)_{x\in\mathcal{X}}$. Memory states are written as d-dimensional

[11] There are *several* alternative (and often conflicting) definitions and terms for "probabilistic FSMs" in the literature, depending on which context they have been applied to. Here, we simply consider finite-state machines (automata or transducers) with probabilistic transitions (and thus outputs). Unfortunately, the lack of consistent terminology is present even in the deterministic case [151, 130]. Also note that we do not require the usual notion of "accepting states", often found in the FSM literature, as we are only interested in the input-output correlations generated at every step. In this sense, our applications is closer to the way FSMs are used in the natural language processing literature [130].

stochastic row vectors π, i.e., where $\pi_j \geq 0$ for all $j = 1, \ldots, d$, and $\sum_{j=1}^{d} \pi_j = 1$. The input-output probability distribution generated by the FSM can then be written succinctly in terms of

$$p(\boldsymbol{a}|\boldsymbol{x}) := \pi_0 T_{a_1|x_1} T_{a_2|x_2} \cdots T_{a_L|x_L} \eta = \pi_0 T_{\boldsymbol{a}|\boldsymbol{x}} \eta, \tag{1.20}$$

where π_0 is the initial reset state according to s_0, and $\eta = [1, 1, \ldots, 1]^{\mathrm{T}}$ is used to sum over all final states. Naturally, in the case of an automaton this simplifies to

$$p(\boldsymbol{a}) := \pi_0 T_{a_1} T_{a_2} \cdots T_{a_L} \eta = \pi_0 T_{\boldsymbol{a}} \eta. \tag{1.21}$$

Equations (1.20) and (1.21) can be interpreted as follows. At each moment in time the machine can be found in exactly one of its d distinct states according to some distribution. The transition matrices $T_{a|x}$ describe the behavior of a random walk following state-dependent input-output rules. The matrix entries $[T_{\boldsymbol{a}|\boldsymbol{x}}]_{ij}$ represent the total probability of starting from state i and ending at the state j, over all possible execution histories which successfully output $\boldsymbol{a}$ given $\boldsymbol{x}$ as input. Since we do not care about the final state of the machine, η is used to obtain the final probability $p(\boldsymbol{a}|\boldsymbol{x})$ over all successful histories.

1.3.2 Quantum finite-state machines

Quantum FSMs can be understood as models for sequential quantum measurements performed on the same d-level quantum system, making it necessary that we consider the post-measurement states explicitly. Importantly, we do not restrict measurements to be projective as these impose artificial restrictions on the set of correlations achievable by a quantum system of a given dimension [34, 31, 32]. Instead, we consider the most general measurements in quantum theory, as described by *quantum instruments*.

Let $\mathcal{H}$ be a finite-dimensional Hilbert space with dimension $d = \dim \mathcal{H}$, and

$$\mathcal{S}_{\mathcal{H}} := \{\ \rho \in \mathcal{L}(\mathcal{H}) \ \mid\ \rho \geq 0, \text{ and } \mathrm{Tr}\,[\rho] = 1\ \} \tag{1.22}$$

denote the corresponding set of quantum states, where $\mathcal{L}(\mathcal{H})$ is the set of linear operators acting on $\mathcal{H}$. We define quantum FSMs as follows.

Definition 2. A quantum finite-state machine is a tuple $(\mathcal{X}, \mathcal{A}, \mathcal{S}_{\mathcal{H}}, \rho_0, \mathcal{T})$, where $\mathcal{X}$ is the input alphabet, $\mathcal{A}$ is the output alphabet, $\mathcal{S}_{\mathcal{H}}$ is the space of memory states, $\rho_0 \in \mathcal{S}_{\mathcal{H}}$ is an initial "reset" state, and $\mathcal{T} : \mathcal{X} \times \mathcal{S}_{\mathcal{H}} \to \mathcal{A} \times \mathcal{S}_{\mathcal{H}} \times [0, 1]$ is a transition rule.

We specify $\mathcal{T}$ in terms of quantum instruments $\mathcal{I}_x := (\mathcal{I}_{a|x})_{a \in \mathcal{A}}$, where the effects $\mathcal{I}_{a|x}$ are completely positive (CP) trace non-increasing maps such that $\sum_{a \in \mathcal{A}} \mathcal{I}_{a|x}$ is completely positive and trace preserving (CPTP) for every $x \in \mathcal{X}$. The overall quantum transition model will be denoted by $\mathcal{I} := (\mathcal{I}_x)_{x \in \mathcal{X}}$. Thus, $\mathrm{Tr}\left[\mathcal{I}_{a|x}(\rho)\right]$ is the probability of outputting a given input x when the machine is in the state $\rho \in \mathcal{S}_{\mathcal{H}}$.

Concretely, the effects $\mathcal{I}_{a|x}$ can be written using their Kraus decomposition,

$$\mathcal{I}_{a|x}(\rho) := \sum_{k=1}^{n_{a|x}} (K_{a|x}^k)\, \rho\, (K_{a|x}^k)^\dagger, \qquad \sum_{a \in \mathcal{A}} \sum_{k=1}^{n_{a|x}} (K_{a|x}^k)^\dagger (K_{a|x}^k) = \mathbb{1}, \tag{1.23}$$

where $n_{a|x} \leq d^2$ is the number of Kraus operators $K_{a|x}^k \in \mathbb{C}^{d \times d}$ for each effect, and the Kraus condition in Eq. (1.23) ensures $\sum_{a \in \mathcal{A}} \mathcal{I}_{a|x}$ is trace preserving. The probability distribution generated by the FSM can then be written succinctly in terms of

$$p(\boldsymbol{a}|\boldsymbol{x}) := \mathrm{Tr}\left[\mathcal{I}_{a_L|x_L} \circ \cdots \circ \mathcal{I}_{a_2|x_2} \circ \mathcal{I}_{a_1|x_1}(\rho_0)\right] = \mathrm{Tr}\left[\mathcal{I}_{\boldsymbol{a}|\boldsymbol{x}}(\rho_0)\right], \tag{1.24}$$

or, in the case of an automaton,

$$p(\boldsymbol{a}) := \mathrm{Tr}\left[\mathcal{I}_{a_L} \circ \cdots \circ \mathcal{I}_{a_2} \circ \mathcal{I}_{a_1}(\rho_0)\right] = \mathrm{Tr}\left[\mathcal{I}_{\boldsymbol{a}}(\rho_0)\right]. \tag{1.25}$$

Quantum-to-classical correspondence

A quantum FSM is capable of simulating any classical FSM of the same size. Given a classical FSM with d states, let $T := (T_x)_{x\in\mathcal{X}}$ with $T_x := (T_{a|x})_{a\in\mathcal{A}}$ be its transition model and π_0 its reset state. We can convert it into an equivalent quantum FSM, with dimension d, by choosing an orthonormal basis $\{|i\rangle\}_{i=1}^{d}$, with Kraus operators for $\mathcal{I}_{a|x}$ acting as

$$K_{a|x}^{k}\,|i\rangle := \sqrt{[T_{a|x}]_{ij}}\,|j\rangle\,, \tag{1.26}$$

which can be done with $n_{a|x} \leq d$ operators for each effect $\mathcal{I}_{a|x}$. Choosing $\rho_0 = \sum_{i=1}^{d}[\pi_0]_i\,|i\rangle\langle i|$, it is then easy to check that

$$p(\boldsymbol{a}|\boldsymbol{x}) = \pi_0 T_{\boldsymbol{a}|\boldsymbol{x}}\eta = \mathrm{Tr}\left[\mathcal{I}_{\boldsymbol{a}|\boldsymbol{x}}(\rho_0)\right]. \tag{1.27}$$

1.3.3 Moore vs. Mealy machines

FSMs can be broadly put into two categories, depending on how transitions and outputs are related:

- Mealy [127] or "edge emitting" machines, where outputs are assigned to transitions. The transition rules are of the form used earlier: $\mathcal{T} : \mathcal{X}\times\mathcal{S} \to \mathcal{A}\times\mathcal{S}\times[0,1]$.

- Moore [136] or "state emitting" machines, where outputs are assigned to states. This requires an additional "emission" rule $\mathcal{E}$ specifying how outputs are generated for each state. These rules are of the form: $\mathcal{T} : \mathcal{X}\times\mathcal{S} \to \mathcal{S}\times[0,1]$ and $\mathcal{E} : \mathcal{S} \to \mathcal{A}\times[0,1]$.

In this thesis we will exclusively use Mealy machines, whereas Moore machines are far more prevalent in most of the adjacent literature[12]. For this reason, a brief discussion is warranted regarding our choice of Mealy machines, as they have many desirable properties for our particular goals within theoretical physics.

First, while both types of machines are equivalent in many applications, for our purposes they are not. While every Moore machine can be converted into an equivalent Mealy machine of the same size—i.e., both generate the same distribution $p(\boldsymbol{a}|\boldsymbol{x})$—the conversion of a Mealy machine into an equivalent Moore machine will generally require more states [114, 188]. Since our goal is to understand fundamental constraints on which correlations can be realized with a finite amount of classical or quantum memory, the choice of Mealy machines is more natural.

The choice of Mealy machines is also motivated by physical principles. Any quantum measurement which reveals information about a system's state will necessarily disturb the state [36, 131]. In other words, informative measurements are *invasive*. From this perspective, projective measurements are maximally invasive (no subsequent measurement provides additional information), whereas noninvasive measurements are trivial [36]. This is an important consideration when using generalized quantum measurements, as they present a fundamental trade-off between information gain and disturbance.

In light of this, a fair comparison between classical and quantum memory effects must consider the possibility of general invasive classical measurements. This is precisely the defining feature of Mealy FSMs, where outputs are associated with state transitions. Thus, the choice of Mealy machines allows classical and quantum memory resources to be compared on equal terms: the same amount of memory d, being used in the most general way allowed within each theory, for the same task of generating a given distribution $p(\boldsymbol{a}|\boldsymbol{x})$ or $p(\boldsymbol{a})$.

[12] Examples: in computer science and automata theory in the study of regular and stochastic languages, the use of hidden Markov models for stochastic processes, the design of finite-state controllers, and in certain models of partially observable Markov decision process. While Mealy machines also occur occasionally, many results for Moore machines do not translate well to Mealy machines, as they rely on significant simplifications provided by the uncoupling of transitions and outputs.

In contrast, Moore machines are more restricted. In a classical Moore FSM the instruments must factorize as $T_{a|x} := T_x E_a$, for a stochastic transition matrix T_x and a sub-stochastic emission matrix E_a, corresponding to (possibly *non*-invasive) classical measurements. The equivalent quantum Moore machine would require quantum channels $\mathcal{J}_x$ and instruments $\mathcal{I}_a$ such that $\mathcal{I}_{a|x} := \mathcal{I}_a \circ \mathcal{J}_x$. This forced factorization of $T_{a|x}$ and $\mathcal{I}_{a|x}$ imposes artificial and incomparable restrictions on the possible behaviors of the FSMs, whereas no such restrictions arise in the Mealy case.

1.4 Numerical optimization

This thesis makes extensive use of numerical optimization, with two particular approaches being used. Explicit constructions and lower bounds on temporal correlations were found via gradient descent techniques. In the context of open quantum systems, upper bounds were obtained using outer approximations computed through semidefinite programming techniques.

1.4.1 Gradient descent

Gradient descent (GD) is a simple and flexible optimization method suitable for non-linear optimization problems involving differentiable functions. While the method can only guarantee a local optimum, it is often feasible to run multiple trials for different initial conditions, improving the likelihood of a good solution. In this section we give a brief outline of the GD methods used in this thesis, which we have used to obtain explicit classical and quantum models generating temporal correlations. More information about GD methods can be found in Refs. [22, 74, 166].

Let $f : \mathbb{R}^n \to \mathbb{R}$ denote a continuous differentiable function, the *objective function*, and ∇f denote its gradient. The goal is to find the optimal solution $x^* \in \mathbb{R}^n$, a *global minimizer*, where

$$x^* = \arg\min_{x \in \mathbb{R}^n} f(x), \tag{1.28}$$

such that $\omega^* = f(x^*)$ is the global minimum.

Starting from a random initial guess x_0, the basic GD method iteratively updates the estimate solution x_t by moving opposite to the gradient's direction, thus obtaining a sequence of candidate solutions

$$x_{t+1} = x_t - \alpha_t \nabla f(x_t). \tag{1.29}$$

The scalar value $\alpha_t > 0$ is often referred to as the *learning rate* and controls how large of a "step" is taken in the direction of *steepest descent* $-\nabla f(x_t)$. The algorithm usually stops after a certain maximum number of iterations, or if some desired accuracy can be determined to have been reached.

In this basic form GD converges at stationary points, i.e., where $\nabla f(x_t) = 0$, but these may correspond to saddle points instead of local extrema, an issue that often arises in practice [166, 74]. Furthermore, convergence of the basic GD can be quite slow, and it may even overshoot an optimum if α_t is not chosen to be sufficiently small for a given iteration t. Thus, performance of GD can be quite sensitive to the choice of learning rate [74]. A compendium of convergence theorems for gradient methods can be found in Ref. [71].

Several improvements to the basic GD method are possible [166, 74]. A learning rate α_t can be adjusted independently and adaptively for each dimension, e.g., based on estimates of local curvature. Convergence is usually greatly improved through the use of an intermediate "momentum" term v_t computed as an exponential moving average, i.e.,

$$v_{t+1} = \beta v_t + \alpha_t \nabla f(x_t) \tag{1.30}$$

$$x_{t+1} = x_t - v_{t+1} \tag{1.31}$$

where the scalar $\beta \in [0, 1]$, typically around $\beta = 0.9$, controls the exponential decay factor. This simple adjustment greatly accelerates movement in favorable directions, and helps the algorithm evade saddle points and local minima.

A relatively new algorithm called Adam [100] merges momentum with various previous techniques to achieve significant improvements in a wide range of problems [100, 166, 74]. The name Adam is derived from "adaptive moment estimation", as the algorithm computes an adaptive and independent learning rate for each dimension using running estimates of first and second moments of the gradient. Adam is well suited for stochastic gradient descent, where the gradient is sampled in only a subset of directions at each iteration, and has thus found widespread applications in the training of large neural networks [74, 166].

In this thesis, since we have direct access to $f(x)$ and are able to compute gradients numerically using finite differences, all terms of the gradient were used instead, which we found to improve results. Furthermore, Adam showed better performance than any GD variant we have tested, while also finding better optima than alternative optimization methods[13].

Finally, while GD is an unconstrained optimization method, constraints can be imposed in a number of ways, e.g., by continuously projecting x_t onto the feasible region, or the gradient ∇f onto the tangent space of the feasible region [74]. Regularization techniques using penalty or barrier functions can also be employed to discourage or forbid solutions violating any desired constraint [74, 22]. In this thesis, constraints were introduced by an appropriate transformation of optimization variables, as will be described in later chapters.

1.4.2 Semidefinite programming

Semidefinite programming (SDP) is a powerful convex optimization technique for problems involving linear objective functions, linear equality and inequality constraints, and positive semidefinite constraints. We only briefly outline SDPs in this section; for more details, see Refs. [197, 22, 211].

A SDP in standard form can be written as

$$\begin{aligned} \textbf{Given:} \quad & F,\ \{C_k\}_{k=1}^m,\ \{c_k\}_{k=1}^m \\ \textbf{Find:} \quad & \omega^* := \max_X \langle F, X\rangle \\ \textbf{Subject to:} \quad & \langle C_k, X\rangle = c_k, \quad k = 1, \dots, m \\ & X \geq 0, \end{aligned} \tag{1.32}$$

where $\langle A, B\rangle = \mathrm{Tr}\left[A^\dagger B\right]$ denotes the Frobenius inner product between matrices A and B, and F, C_k, and X are $n \times n$ complex-valued matrices, with $X = X^\dagger$. Equation (1.32) is referred to as the *primal* form of the SDP, with its *dual* SDP given by

$$\begin{aligned} \textbf{Given:} \quad & F,\ \{C_k\}_{k=1}^m,\ \{c_k\}_{k=1}^m \\ \textbf{Find:} \quad & \tau^* := \min_y\ y^{\mathrm{T}} c = \min_y\ \sum_{k=1}^m y_k c_k \\ \textbf{Subject to:} \quad & \sum_{k=1}^m y_k C_k \geq F \end{aligned} \tag{1.33}$$

where $y_k \in \mathbb{R}$ for $k = 1, \dots, m$ denotes the dual optimization variable. It's always possible to convert from one form into the other. An important feature of this SDP duality is that any feasible solution in the dual problem provides an upper bound for the primal problem, and vice versa. To see why, we simply evaluate

[13] Different methods tried were Nelder-Mead, differential evolution, and simulated annealing, all in their Wolfram Mathematica implementations.

$\tau^* - \omega^*$, obtaining:

$$y^{\mathrm{T}} c - \langle F, X \rangle = \sum_k y_k c_k - \langle F, X \rangle = \sum_k y_k \langle C_k, X \rangle - \langle F, X \rangle = \langle \sum_k y_k C_k - F, X \rangle \geq 0, \tag{1.34}$$

where we use the fact $\sum_{k=1}^m y_k C_k \geq F$ and $X \geq 0$. This result is known as the *weak duality theorem*, and the quantity $\tau^* - \omega^*$ is referred to as the *duality gap*.

When the optimal values of primal and dual coincide, the SDP is said to satisfy the *strong duality* property. Existence of strong duality is established in the following result [178]:

Theorem 1. (Strong duality) The primal-dual SDPs will satisfy strong duality, i.e. $\omega^* = \tau^*$, if either one of the following conditions hold:

1. The primal problem is strictly feasible, i.e., there exists $X > 0$ such that $\langle C_k, X \rangle = c_k$ for all $k = 1, \ldots, m$.

2. The dual problem is strictly feasible, i.e., there exits y such that $\sum_{k=1}^m y_k C_k > F$.

Not all SDPs feature strong duality, but these conditions are often satisfied in practice. Thus, by simultaneously solving the primal and dual problems one obtains convergence guarantees to any desirable precision.

While SDPs describe linear convex optimization problems, several relaxation techniques exist that can transform non-linear optimization problems into SDPs. Some examples are global polynomial optimization [148, 97, 112, 113], rank-constrained optimization [216], solving partial differential equations [129], and constrained bilinear optimization [16]. Given their flexibility, SDPs have found widespread applications in numerous scientific disciplines, and are particularly ubiquitous in quantum information science [177, 192]. Furthermore, efficient solvers are also widely and freely available [137, 145].

Chapter 2

Temporal correlations in the simplest measurement sequences

The main results in this chapter were published in Ref. [198]:

> *Temporal correlations in the simplest measurement sequences*
> **Lucas B. Vieira** and Costantino Budroni
> *Quantum* 6, p. 623 (2022) — Accepted 2022-01-09

Author contribution: In this work, the doctoral candidate significantly contributed to the conception of the research topic. The candidate was fully responsible for writing all of the gradient descent optimization code for the classical and quantum surveys (excluding the data in Fig. 2.13), analyzing the resulting survey data, formulating the main conjectures, and developing the deterministic complexity algorithm. The candidate also wrote the majority of the manuscript, and was solely responsible for producing all its graphical assets.

The contents of this chapter were adapted from the original text. Additional previously unpublished results have also been included, and will be noted when they appear.

2.1 Introduction

As discussed in Sec. 1.2.8, the problem of bounding the set of temporal correlations under the assumption of finite memory is exceptionally difficult in general, due to its highly non-convex structure and inherent high-dimensionality. As a consequence, the observation that the full set of temporal correlations — classical or quantum — forms the Arrow-of-Time polytope (Sec. 1.2.7) leads naturally to the idea of investigating its extremal points, which correspond to the deterministic distributions [67, 88, 182]. If initial randomness is an available resource, scenarios generating the extremal distributions may be combined into any distribution in their the convex hull, providing a linear characterization of the set of correlations. This idea was discussed in previous works [88, 124], where it is generally understood that much of the problem's structure is lost with this approach [182, 124].

In paving the way towards a deeper understanding of the problem, while preserving as much of its structure as we can, this chapter focuses instead on the simplest scenario possible for investigating temporal correlations: A single finite-dimensional system being sequentially subjected to the same dichotomous measurement over multiple time steps, generating a sequence of outcomes. In particular, we quantify the maximum probability for individual sequences to be generated using finite memory resources, i.e., by systems of bounded dimension.

Before we begin, however, it should be noted that the techniques developed in this chapter are sufficiently general to be applied to the investigation of more general scenarios, such as approximating entire distributions using a finite amount of memory. Here, we have focused on the properties of individual sequences as a further simplification, framing their generation as an information processing task.

The chapter is organized as follows. In Sec. 2.3, we introduce the physical scenario investigated, and briefly review the basic concepts related to finite-state machines required. In Sec. 2.5, the minimal dimension necessary for deterministic generation of a sequence is investigated, and the deterministic complexity defined. Approaches to investigate sub-deterministic scenarios numerically are discussed in Sec. 2.6, with Secs. 2.7 and 2.8 discussing the results of our classical and quantum numerical surveys, respectively. Finally, Sec. 2.9 closes the chapter with a discussion of the results.

2.2 Summary of main results

As a first step, we investigate the minimum dimension d required to generate each finite sequence $\boldsymbol{a} \in \mathcal{A}^L$ deterministically, in either classical or quantum theory, which we define as the sequence's *deterministic complexity* (DC), and provide an efficient algorithm computing it. The notion of a minimum dimension for generating deterministic temporal correlations was introduced in Ref. [182], in the context of deterministic input-output distributions. In this thesis, we conceptualize it instead as new computable measure of computational complexity for symbolic sequences, rooted on physical principles, which provides numerous new insights and generalizations. These ideas are discussed in detail and generalized in Chapter 6.

Given a sequence $\boldsymbol{a}$ and its deterministic complexity $\mathrm{DC}(\boldsymbol{a})$, the central observation is that below this dimension, i.e., for $d < \mathrm{DC}(\boldsymbol{a})$, any classical or quantum physical realization of $\boldsymbol{a}$ must be probabilistic, implying the existence of a nontrivial upper bound on these probabilities. Similar bounds on temporal correlations have been investigated in other contexts [88, 32], and generally differ between classical, quantum, or general probability theories [32, 33]. Therefore, these bounds provide a clear method for characterizing the correlations realizable by different physical theories.

Despite earlier results exploring finite-memory temporal correlations in various settings [88, 32, 182, 124, 33], not much was known about *how* classical and quantum memory resources can be used optimally for generating temporal correlations. In order to address this gap in our knowledge, we have employed numerical optimization techniques to comprehensively survey the optimal behaviors for classical and quantum models, for all sequences up to length $L = 10$ in all of their sub-deterministic scenarios, i.e., when $d < \mathrm{DC}(\boldsymbol{a})$. Note that, while the ultimate goal is to obtain upper bounds on the probabilities for each sequence, the numerical methods utilized in this chapter only provide lower bounds through explicit constructions, as will be discussed in Sec. 2.6. Nevertheless, these results still led to many novel insights.

The surveys revealed a highly nontrivial behavior for both classical and quantum models, even in this simplified scenario, as well as many puzzling features shared among very dissimilar sequences and respective optimal models. The problem appears to be rich in emergent structures, many of which remain unexplained.

A particularly striking observation from our numerical surveys suggests that the maximum probability for any sequence can be upper bounded by the maximum probability for a special family of sequences, which we call *one-tick sequences*. This result hinges on a remarkable connection between sub-deterministic temporal correlations and the deterministic complexity, as will be discussed in Sec. 2.7.3. Despite considerable efforts for a proof, this result remains as a conjecture. However, its proof immediately implies an even more surprising result: The existence of a universal upper bound of $1/\mathrm{e}$ for the probability of any sequence in sub-deterministic classical scenarios.

The results of the quantum survey have also provided many interesting insights. While they confirm the intuition that quantum memories should outperform classical ones, we have also found evidence

that there may exist memory-restricted scenarios where this advantage disappears (Sec. 2.8.5). Furthermore, for the task of maximizing the probability of individual sequences, optimal quantum strategies are achieved entirely with pure states. Our results also suggest that there is no nontrivial universal upper bound in the sub-deterministic quantum scenarios.

2.3 Sequential measurements

The physical scenario investigated in this chapter can be conceptualized as a device performing the same dichotomous measurement on its d-dimensional (classical or quantum) internal memory, as in Fig. 2.1. For each trial of an experiment, the device is first reset to a fixed initial state, then a finite sequence of L measurements is performed obtaining outcomes $\boldsymbol{a} \in \mathcal{A}^L$, with $\mathcal{A} = \{0, 1\}$. Over multiple trials, we may estimate a probability distribution $p(\boldsymbol{a}) := p(a_1 a_2 \dots a_L)$ over all possible sequences the device generates using only its limited internal memory as a resource[1].

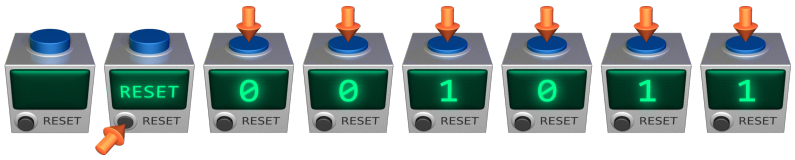

Figure 2.1: A depiction of our physical scenario, as a device with a single "measure" button (on top), a "reset" button (in front), and a display for the outcomes of measurements performed in the device's internal memory. Performing L sequential measurements produces an output sequence, e.g., $\boldsymbol{a} = 001011$. An empirical distribution $p(\boldsymbol{a})$ over all sequences generated by the device can be obtained by resetting it and performing multiple trials. Here, time moves forward left-to-right.

Importantly, we assume the measurements are time-independent and that the device has no access to external resources, e.g., a wall clock or a side channel carrying information about past outcomes. In this way, the sole resource available for generating the sequences is the internal memory, which is completely characterized by the dimension d and its classical or quantum nature. Ideally, we wish to bound the set of distributions which can be realized by classical or quantum d-dimensional memories. For this reason, we shall use the notation $p(\boldsymbol{a}|d)$ to make this relationship explicit.

Analyzing an entire probability distribution at once, however, is already quite challenging. Previous works have resorted to studying deterministic distributions or carefully constructing bounds for specific scenarios [88, 183, 32, 124]. We shall follow a different approach, by instead exhaustively characterizing sequences in isolation, such that general statements can be made about distributions by analyzing the respective probabilities of each sequence. In particular, our strategy will be to investigate the maximum probability of generating each sequence, given a classical or quantum memory of dimension d. In the process, we uncover new insights on the emergent structure of the problem.

A family of sequences will be of particular interest to us. These are the sequences of the form $\boldsymbol{a}^L_{\text{ot}} := 0^{L-1}1$, which we call *one-tick sequences*. They have previously appeared in the investigation of classical and quantum clocks [33], as the sequence corresponding to one tick of the clock after $L-1$ time steps without a tick, hence their name.

[1]Note that this is merely a conceptual picture for the physical scenario we consider, as in this thesis we will disregard the effects of finite statistics, focusing instead on idealized probability distributions.

2.4 Classical and quantum models

In order to properly characterize its temporal correlations, it is essential that we model this scenario through the most general physical states and operations allowed within each theory. If this is not carefully ensured, a restriction on classical operations (e.g. the non-invasiveness assumption of Leggett and Garg) could lead to classical correlations being incorrectly interpreted as non-classical, or the converse, whereby a restriction on quantum operations (e.g. considering only projective measurements [34, 32]) could lead to certain quantum advantages not being apparent. Chapter 4 investigates a similar issue, where quantum correlations could be incorrectly characterized as classical in protocols involving entanglement-breaking (or measure-and-prepare) channels.

With this in mind, the physical behavior in our scenario can be modeled through the formalism of probabilistic finite-state machines (FSMs) and their quantum generalizations, as introduced in Sec. 1.3. In particular, as there are no inputs (or, equivalently, $X = 1$), we may also refer to these machines as *finite-state automata*, here acting as autonomous *generators* of the sequences $\boldsymbol{a}$. This terminology will be used as to distinguish them from *transducers* (see Sec. 1.3), which operate between inputs and outputs. Recall that a (Mealy) finite-state automaton is a FSM which, at each time step, performs a state transition while simultaneously generating an output belonging to some alphabet $\mathcal{A}$, following a time-independent rule. We assume the machine has d perfectly distinguishable internal states (classical or quantum), which we may refer simply as the "amount of memory" available; see Sec. 1.2.8.

2.4.1 Classical models

Departing from the usual approach of Leggett and Garg, the classical case corresponds to a sequence of measurements which are allowed to be invasive, up to a certain finite amount defined by d. The most general classical behavior is modeled by a probabilistic automaton [151, 157], described by a pair of matrices $T := (T_0, T_1)$, where T_a are row sub-stochastic transition matrices such that $T_0 + T_1$ is row stochastic, i.e., with nonnegative entries, $[T_a]_{ij} \geq 0$ for $a \in \mathcal{A}$, and such that $\sum_{aj} [T_a]_{ij} = 1$ for all i, indicating an overall stochastic behavior which preserves probability. We refer to T as a *classical model*.

Given a d-dimensional classical model T, the probability of a sequence $\boldsymbol{a}$ being generated can be computed as

$$p(\boldsymbol{a}|T, d) := \pi T_{a_1} T_{a_2} \dots T_{a_L} \eta, \tag{2.1}$$

where π is the initial distribution over states (i.e., the "reset state"), with $\pi_i \geq 0$ and $\sum_i \pi_i = 1$, and where $\eta := (1, \dots, 1)^{\mathrm{T}}$ provides a sum over all possible final states. Equation (2.1) offers a compact description of the overall stochastic behavior of the model, where the internal state is modified upon each measurement. This description is not unique, as states may be relabeled in any permutation. In the intuitive classical macrorealist interpretation, the automaton is always in a well-defined internal state, and its behavior over time corresponds to a random walk over its state space. Since Eq. (2.1) is explicitly causally ordered, $p(\boldsymbol{a})$ automatically satisfies the AoT constraints.

State diagrams

Classical models can be visualized using a *state diagram* (Fig. 2.2), a directed multigraph where the nodes (disks) correspond to states, and the edges (arrows) to the various transitions, to each of which we assign an output symbol a and a probability q, giving the label as $a|q$. Stochasticity of $\sum_a T_a$ can be seen as the probability of all outgoing transitions adding up to 1 for each state.

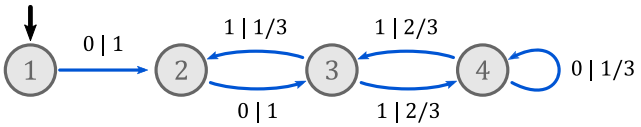

Figure 2.2: A state diagram for the probabilistic automaton of a $d = 4$ classical model. A label $a|q$ indicates a transition occurs with probability q while outputting a. Importantly, since we use Mealy automata, both probabilities and outputs are associated with transitions, not states, allowing for minimal classical and quantum memories to be compared on equal terms. The bold arrow indicates a pure initial state, i.e., an initial distribution $\pi = (1, 0, 0, 0)$. Note that state 3 can transition to states 2 or 4 with output $a = 1$, indicating this is a non-unifilar model.

Comparison with computational mechanics and ε-machines

Here, it is worth mentioning that similar probabilistic automata generating symbolic sequences also appear in the study of optimal *predictive* models for stochastic processes — usually assumed bi-infinite and stationary — in the form of ε-machines from computational mechanics [55, 173]. These machines are constructed by creating equivalence classes between all observed pasts producing the same observed future, which then define the ε-machine's *causal states*. In this way, ε-machines encode the minimal predictive causal structure of the observed process, with the Shannon entropy of the stationary distribution over causal states, known as the *statistical complexity* of the process, providing a quantifiable measure of its inherent structure.

However, ε-machines do not describe the *generative* dynamics of the underlying process. In particular, they necessarily obey the property of *unifilarity*[2], where no two outgoing transitions from any given state produce the same output — a direct consequence of the definition of causal states. While unifilarity can be proven to be optimal for predictive models [55, 173], it is known to be sub-optimal for generative models [167], which we are focusing on. In fact, ε-machines describing (infinite length) generative processes typically require an infinite number of causal states [55, 96].

Since we require the most general operations allowed by classical theory, we impose no restrictions on the structure of our models. In particular, there is no physical justification for unifilarity, and we must explicitly consider non-stationary memory effects in our analysis.

2.4.2 Quantum models

The quantum case corresponds to repeated applications of the same quantum measurement, with outcomes $a \in \mathcal{A}$. The sequential nature of the problem requires that we consider the post-measurement states explicitly, so that to each outcome a will be associated a state update rule. Importantly, we do not assume measurements to be projective or of measure-and-prepare type, as these impose artificial restrictions on the set of correlations achievable by a system of a given dimension [34, 31, 32].

Instead, we consider the most general measurements in quantum theory, described by *quantum instruments*. The output probability and state transition will be described[3] by an instrument $\mathcal{I} = (\mathcal{I}_0, \mathcal{I}_1)$, where the effects $\mathcal{I}_a$ for $a \in \mathcal{A}$ are completely positive (CP) trace non-increasing maps updating the state upon each outcome. It is further required that $\mathcal{I}_0 + \mathcal{I}_1$ be trace preserving, corresponding to the conservation of probability. We refer to $\mathcal{I}$ as a *quantum model*.

Given a quantum model acting on a d-dimensional Hilbert space, the probability for a sequence $\boldsymbol{a}$ can

[2]Meaning "single file" or "single thread", suggestive of how, from any starting state, there is a unique trajectory over the state space for any observed sequence of outputs.

[3]In particular, given a normalized state ρ_t we have $p(a) = \mathrm{Tr}\left[\mathcal{I}_a(\rho_t)\right]$ and a resulting subnormalized state $\rho_{t+1} = \mathcal{I}_a(\rho_t)$.

be computed, in the Schrödinger picture, as

$$p(\boldsymbol{a}|\mathcal{I}, d) := \mathrm{Tr}\left[\mathcal{I}_{a_L} \circ \mathcal{I}_{a_{L-1}} \circ \ldots \mathcal{I}_{a_1}(\rho)\right], \tag{2.2}$$

where ρ is the initial "reset" quantum state. The effects $\mathcal{I}_a$ can be written concretely in their Kraus decomposition,

$$\mathcal{I}_a(\rho) = \sum_{k=1}^{n_a} K_{ak}\rho K_{ak}^\dagger, \qquad \text{with} \sum_{a\in\mathcal{A}}\sum_{k=1}^{n_a} K_{ak}^\dagger K_{ak} = \mathbb{1}, \tag{2.3}$$

where we assign $n_a \leq d^2$ Kraus operators to each outcome, with a total of $N = \sum_a n_a$. The Kraus condition in Eq. (2.3) ensures the full set of Kraus operators preserves total probability. As in the classical case, models are symmetric under permutations of the states. Unfortunately there is no standard — or useful — diagrammatic way to represent quantum models, so their behavior is less intuitive.

Importantly, classical models form a strict a subset of quantum models, which can be seen with the following construction. Let T be an arbitrary classical model with d states. We may construct an equivalent quantum model, also on d states, through the a of Kraus operators $\{K_{ak}\}_{a,k}$ acting as

$$K_{ak}\,|i\rangle = \sqrt{[T_a]_{ij}}\,|j\rangle\,. \tag{2.4}$$

Letting $\rho = \sum_i \pi_i\,|i\rangle\langle i|$, we then have

$$\pi T_{a_1} T_{a_2} \ldots T_{a_L}\eta = \mathrm{Tr}\left[\mathcal{I}_{a_L} \circ \mathcal{I}_{a_{L-1}} \circ \ldots \mathcal{I}_{a_1}(\rho)\right]. \tag{2.5}$$

This direct correspondence is a major reason why we specifically consider probabilistic Mealy automata for the classical models, as it places both theories on equal terms. This is true not only with respect to their description of states and their transformations, but also in how the same dimension d is treated as a fundamental resource; see Sec. 1.2.8.

2.5 Deterministic scenarios

Having a concrete representation of the finite-dimension models, we can now define the main optimization problems of this chapter. The maximum classical and quantum probabilities of each sequence can be defined as:

$$\Omega_C(\boldsymbol{a}, d) := \sup_T p(\boldsymbol{a}|T, d), \qquad \Omega_Q(\boldsymbol{a}, d) := \sup_{\mathcal{I}} p(\boldsymbol{a}|\mathcal{I}, d), \tag{2.6}$$

The memory of the system, either classical or quantum, is the sole resource available to generate a given sequence, and is fundamentally constrained by the dimension d. Clearly, both maxima will increase monotonically towards unity as d increases, i.e.,

$$0 \leq \Omega(\boldsymbol{a}, 1) \leq \Omega(\boldsymbol{a}, 2) \leq \cdots \leq \Omega(\boldsymbol{a}, d) \leq 1, \tag{2.7}$$

where the subscript was omitted as we refer to either case.

An important observation is that, for sufficiently large d, any finite sequence can be generated deterministically (i.e., $p(\boldsymbol{a}|d) = 1$) in either classical or quantum theory [88, 182], outlining a stark difference between temporal and spatial correlations[4]. However, d may be larger than strictly required to achieve this, in which case the model could be making inefficient use of the memory resource, e.g., by having two different transitions which produce identical future behaviors. In such cases, the number of states d could be reduced without affecting the observed outcomes. This naturally leads us to investigate, precisely, the minimum dimension d needed for a sequence to be generated deterministically. This deterministic

[4]Notwithstanding the importance of this fact, it is still somewhat underappreciated among physicists, which has led to common misconceptions emerging from flawed analogies between spatial and temporal correlations. This is the subject of Ch. 4.

threshold will play a central role in the study of temporal correlations, and in our efforts to identify the differences between classical and quantum memories.

A key result is that both classical and quantum theory agree on this minimum dimension. As shown in Ref. [182], quantum states producing difference sequences with probability one must be orthogonal, such that the respective quantum model can be written as deterministic transitions between orthogonal states, therefore requiring no coherence effects. This, in turn, implies classical models can simulate the same behavior with the same number of states. It is therefore sufficient to think in terms of classical deterministic models to study this deterministic memory threshold for each sequence.

These observations also highlight several important differences between the formalism of finite-state machines and other models which have been used for studying temporal correlations, such as the original Leggett-Garg formulation [116, 64], which assumes noninvasive measurability, quantum contextuality for sequential measurements [101, 80, 35], which assumes some form of compatible measurements, other forms of temporal quantum correlations assuming projective measurements [34, 31, 170, 161, 179], or a combination of spatial and temporal correlations [70, 181, 19]. By allowing the most general classical and quantum behaviors to be compared on equal terms, the finite-state machine formalism provides an unambiguous and precise way to distinguish between *manifestly* classical and quantum memory effects. Ref. [203] provides an up-to-date review on temporal correlations, from Leggett-Garg inequalities to finite-state machines, with a detailed discussion on these important differences.

2.5.1 Deterministic Complexity

We are now ready to formalize the deterministic memory threshold. We begin by introducing the following complexity measure over finite sequences of symbols.

Definition 3. (Deterministic Complexity, DC). Let $\mathcal{A}$ be an alphabet of arbitrary size, and $\boldsymbol{a} \in \mathcal{A}^L$ a finite sequence. The *deterministic complexity* of the sequence $\boldsymbol{a}$, denoted by $\mathrm{DC}(\boldsymbol{a})$, is the minimal number of states d such that there exists a classical model T with $p(\boldsymbol{a}|T, d) = 1$.

As a direct consequence of this definition, we establish the following trivial — but *crucial* — observation for our investigations on temporal correlations:

Observation 1. For any sequence $\boldsymbol{a} \in \mathcal{A}^L$:

$$d < \mathrm{DC}(\boldsymbol{a}) \quad \Longrightarrow \quad \Omega_C(\boldsymbol{a}, d) < 1 \quad \text{and} \quad \Omega_Q(\boldsymbol{a}, d) < 1. \tag{2.8}$$

In words, $d < \mathrm{DC}(\boldsymbol{a})$ implies classical and quantum theory must obey nontrivial bounds on the maximum probability of generating $\boldsymbol{a}$. We refer to these as *sub-deterministic scenarios* as to emphasize this fundamental memory constraint.

A trivial upper bound for DC is given by the sequence length, i.e., $\mathrm{DC}(\boldsymbol{a}) \leq |\boldsymbol{a}|$. This bound is generally not tight. The general criterion for counting the minimal number of states needed for the deterministic realization of a sequence was formulated in Ref. [182], in the broader context of extremal points of the AoT polytope. The idea is to assign a unique candidate state at every time step, then group these candidate states into equivalence classes sharing the same future behaviors. The number of unique equivalence classes, or *nonequivalent futures*, corresponds to the minimal number of states. This criterion can be adapted into an efficient algorithm allowing for the computation of the deterministic complexity for an arbitrary sequence, as will be discussed shortly in Sec. 2.5.3.

Each deterministic sequence can be realized by a model consisting of transitions between orthogonal states, and furthermore, it is sufficient to consider pure states [182]. A simple argument can be used to show that all minimal models must be of this form, i.e., that they must involve only transitions between

pure orthogonal states. Since minimal classical and quantum models have the same number of states, it suffices to show this for the classical case.

Observation 2. If a sequence $\boldsymbol{a}$ is generated deterministically, then its minimal classical model T, with $d = \mathrm{DC}(\boldsymbol{a})$ states, involves only deterministic transitions, i.e.,

$$p(\boldsymbol{a}|T, d = \mathrm{DC}(\boldsymbol{a})) = 1 \implies [T_a]_{ij} \in \{0, 1\}, \ \forall\, i, j, a. \tag{2.9}$$

Proof. The proof is by contradiction. Suppose that at the n-th transition, $n < L$, instead of moving deterministically, the automaton transitions to the state s with probability q or to the state s' with probability $(1-q)$. The total probability can then be written as $1 = p = qp_1 + (1-q)p_2$, where p_1 and p_2 are the probabilities for the two paths, conditioned on that probabilistic transition at the n-th step. Since $0 \leq q \leq 1$ and $p_i \leq 1$, $p = 1$ implies $p_1 = p_2 = 1$. Hence, it is sufficient to follow the path going through, e.g., s to generate the sequence with probability 1. One may, then, simply remove the state s' and put $q = 1$, in contradiction with the assumption that the dimension was minimal. The argument can be applied iteratively, if more than one probabilistic transition exists. The only remaining case is that of a transition happening in the last step, i.e., $n = L$, in which case the transition is irrelevant for the model. In the terminology above, $1 = p = p_1(q + 1 - q) = 1$. □

Since the argument uses only the probability over paths in the state space, it is also valid for the case of finite-state machines involving both inputs and outputs, as in [88, 32, 182]. We develop this idea further in Ch. 6, where the deterministic complexity is generalized to input and output scenarios.

The previous criterion also implies that, for binary sequences, $\mathrm{DC}(\boldsymbol{a}) = |\boldsymbol{a}|$ is saturated only in the case of one-tick sequences $\boldsymbol{a}^L_{\mathrm{ot}} := 0^{L-1}1$, and the equivalent one obtained by the substitution $0 \leftrightarrow 1$. Many more sequences will saturate this bound for an alphabet of size A and sequence length L, and a simple combinatorial argument provides the exact number

$$(A-1)^{L-1}A. \tag{2.10}$$

The idea is to pick one symbol out of A available for a_L, then using any combination of the other $A-1$ symbols for the remaining $L-1$ positions. Since a_L occurs uniquely at the end, all futures are nonequivalent at each step, thus $\mathrm{DC}(\boldsymbol{a}) = |\boldsymbol{a}|$. This small observation was not discussed in Ref. [198], but we include it here for completeness.

In summary, the deterministic complexity of a sequence provides us with the threshold where quantum memories might provide an advantage over classical ones, namely, the sub-deterministic $d < \mathrm{DC}(\boldsymbol{a})$ scenarios where $\boldsymbol{a}$ can only occur probabilistically. The DC is also connected to what appears to be an upper bound for these probabilities, which depends only the pair $(\mathrm{DC}(\boldsymbol{a}), d)$ and can be computed in terms of the one-tick sequence $\boldsymbol{a}^{\mathrm{DC}(\boldsymbol{a})}_{\mathrm{ot}}$. This conjectured upper bound is discussed in Secs. 2.7.3 and 2.8.2.

2.5.2 Minimal DFAs and patterns

Deterministic models would more typically[5] be referred to as *deterministic finite-state automata* (DFAs). The previous results imply that all minimal DFAs follow a special form, which allows for a simple characterization. The key observation is considering what happens if one keeps measuring after all outputs have been generated by a minimal DFA for the sequence $\boldsymbol{a}$. Since the transitions are deterministic, this DFA will simply continue to generate new outputs, creating an infinite family[6] of sequences, one for each length L, with all length $L > |\boldsymbol{a}|$ sequences sharing $\boldsymbol{a}$ as a prefix.

[5] Although most of the literature deals with DFA as acceptors, not generators as we do.

[6] In computer science one could say a "language" instead of "family of sequences", but this level of jargon seems unnecessary here.

We can describe all such families by what we call a *pattern*, which perhaps is easier to explain with an example. The sequence $\boldsymbol{a} = 000110110$ can be seen as one of the infinite sequences generated by the pattern `00(011)`, where the initial `00` is a prefix and `(011)` denotes that the suffix `011` occurs at least once, then is repeated indefinitely, possibly with truncation. The notation for patterns was inspired by the notation for regular expressions in modern programming languages, where `00(011)+` would match all strings with prefix `00` followed by one or more[7] *full* repetitions of the suffix `011`. In our problem, however, we must also account for any truncation of the repeating suffix.

In the following, we will refer to the length of the pattern as the number of symbols $a \in \mathcal{A}$ appearing in it. With this, we introduce the following notion:

Definition 4. (Minimal pattern). A pattern is said to be *minimal* if it is the shortest pattern generating its infinite family of sequences.

There is a direct correspondence between minimal patterns and the minimal DFAs generating our sequences. To see this, note that each pattern can be directly translated to a DFA, by assigning a unique state to each output symbol in the pattern, with that output being emitted in the transition to the subsequent state, corresponding to the next output symbol in the pattern. At the end of the pattern, the final transition returns to the state associated to the first symbol within the cyclic part, inside the parenthesis. Therefore, we establish the following:

Observation 3. Every minimal deterministic model is characterized by a "*tail*" and a "*cycle*", which together completely describe the state transitions and induce a minimal pattern. The minimal number of states needed to describe the behavior of a pattern is precisely the length of the pattern.

In this way, a minimal DFA for a sequence will correspond to the shortest pattern generating that sequence. This structure is general: Since the number of states d is finite, for $L \geq d$ the system must at some point transition back to a previously used state and repeat the cycle thereafter, as in Fig. 2.3. Note that, due to the possibility of truncation in the recurrence of its cyclic part, more than one minimal pattern may describe a given sequence, but all such patterns will share the same length.

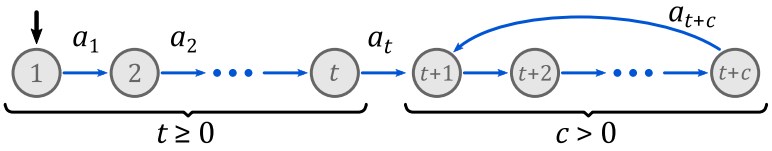

Figure 2.3: The general structure of minimal DFAs corresponding to minimal patterns. It consists of a tail, with $t \geq 0$ states, which leads to a cycle, with $c > 0$ states.

2.5.3 The DCPatterns algorithm

The previous observations regarding minimal DFAs and patterns suggests a simple algorithm to compute the DC of an arbitrary sequence $\boldsymbol{a} \in \mathcal{A}^L$: It is enough to compute the length of a minimal pattern that generates it. Concretely, for a sequence $\boldsymbol{a}$, its minimal patterns are characterized by two numbers, the length of the tail t and the length of the cycle c, such that $\mathrm{DC}(\boldsymbol{a}) = t + c$. Thus, to find the minimal patterns and the deterministic complexity of a given sequence $\boldsymbol{a}$ of length L, we proceed as follows:

1. Start by assuming a pattern length $\ell = 1$.

2. Iterate over the tail lengths $t = 0, \ldots, \ell - 1$, giving a cycle length $c = \ell - l_1$.

[7] Known as "Kleene plus" in computer science.

3. For each t, if all outcomes a_i for $i = \ell + 1, \ldots, L$ can be interpreted as repetitions of the cycle subsequence, $(a_{t+1}, \ldots, a_\ell)$, then (t, c) defines a valid pattern for $\boldsymbol{a}$ of length ℓ, and thus, $\mathrm{DC}(\boldsymbol{a}) = \ell$. We can either halt, or store (t, c) and continue to find more patterns.

4. If, however, no match is found for any t, increment ℓ and start over from step (2).

5. If we also wish to find all patterns compatible with $\boldsymbol{a}$, we may continue checking the remaining values of t for the same ℓ, then return the list of valid patterns.

By starting from $\ell = 1$ and incrementing only after all patterns of a given length have been exhausted, we are sure to obtain a minimal pattern. In this way, we need to test at most $L(L+1)/2$ patterns, each of which we compare with the original sequence, corresponding to $O(L^3)$ operations. Algorithm 1 can be used to efficiently compute a sequence's deterministic complexity, as well as finding all of its minimal patterns. Notice that the algorithm is not restricted to a binary alphabet, and works the same way for an arbitrary one.

Algorithm 1 Deterministic Complexity and Patterns (DCPatterns). The idea is to assume the sequences have the form (tail)+(cycle), with respective lengths t and c, such that $\mathrm{DC}(\boldsymbol{a}) = t + c$. We thus test all such patterns counting up from DC, which ensures we find the minimal-state representation as early as possible.

```
1:  procedure DCPatterns(a)
2:      L ← Length(a)
3:      patterns ← {}                          ▷ We start with an empty list of patterns.
4:      dc ← 0                                 ▷ We count from the bottom up in DC, stopping as early as possible.
5:      found ← False                          ▷ When we find any patterns, we can stop.
6:      while not found do
7:          dc ← dc + 1
8:          for t in 0, ..., dc − 1 do         ▷ For each possible tail length
9:              c ← dc − t                     ▷ ... assume the rest is a cycle.
10:             match ← True                   ▷ Assume this is a valid pattern.
11:             for i in 0, ..., (L − dc) do   ▷ For every other symbol beyond DC
12:                 if a[dc + i] ≠ a[t + (i mod c)] then   ▷ we test whether it can be extrapolated.
13:                     match ← False          ▷ If it can't, this pattern fails and we stop here.
14:                     break
15:                 end if
16:             end for
17:             if match then                  ▷ If the pattern matches the entire sequence...
18:                 patterns ← patterns + {(t, c)}   ▷ ... we add the tuple (t, c) to the list of patterns.
19:                 found ← True               ▷ We stop at this dc, but continue searching for patterns.
20:             end if
21:         end for
22:     end while
23:     return dc, patterns                    ▷ Return the optimum dc and the list of patterns
24: end procedure
```

As pointed out previously, the tail part may have zero length if the entire sequence is itself a (possibly truncated) repeating pattern, e.g., $\boldsymbol{a} = 01001001001$ giving the pattern (010), where the truncation occurs in the last 0 of the 4^{th} repetition. Under such analysis, one-tick sequences (and their $0 \leftrightarrow 1$ symmetric counterparts) are, once again, easily identified as the unique sequences saturating $\mathrm{DC}(\boldsymbol{a}) = |\boldsymbol{a}|$ for all sequences of length L and a binary alphabet $\mathcal{A}$. These sequences also possess the maximum number (L) of potential patterns for any sequence, as for example: $\boldsymbol{a} = 00001$ gives $(00001) \cong 0(0001) \cong 00(001) \cong 000(01) \cong 0000(1)$.

2.5.4 Properties of patterns

Arbitrary patterns are not necessarily unique or minimal in length. As a concrete example, the patterns $00(00)$ and $01(0101)$, of lengths 4 and 6 respectively, are equivalent to the minimal patterns (0) and (01), of lengths 1 and 2. This motivates a natural question of enumerating all minimal patterns of a given length.

Interestingly, the space of unique minimal (binary) patterns using exactly d states can be directly related to the unique minimal unary deterministic finite automata (uDFAs) with exactly d states. An analysis and enumeration of these uDFAs has been presented in Refs. [141, 60]. For $d = 1, 2, 3, \ldots$ there are[8]

$$N_{\mathrm{uDFA}}(d) = 2,\ 4,\ 12,\ 30,\ 78,\ 180,\ 432, \ldots \tag{2.11}$$

such uDFAs, and thus unique minimal patterns with exactly d states.

The argument for uDFAs is easily adapted to the general case, and can be understood as follows. For a pattern of length ℓ to be minimal, it must satisfy two obvious conditions: (i) the tail is minimal, and (ii) the cycle is minimal. Suppose the "tail" and "cycle" parts of the pattern are given by the sequences (or in computer science jargon, *words*) $\boldsymbol{t}$ and $\boldsymbol{c}$, respectively, and let $\boldsymbol{tc}$ be their concatenation. For condition (i) to be valid, the last symbol of $\boldsymbol{c}$ must be different from the last symbol of $\boldsymbol{t}$, otherwise, one could include the last transition of the tail into the cycle while removing the last transition of the cycle, obtaining a pattern of length $\ell - 1$ generating the same infinite sequence. For instance, the pattern $01(001)$ is not minimal because it could be reduced to $0(100)$.

Condition (ii) implies $\boldsymbol{c}$ is a so-called *primitive word* [7], i.e., it must be non-empty and not be expressible as $= \boldsymbol{w}^m$ for $\boldsymbol{w}$ a smaller word and $m \in \mathbb{N}$. As a concrete example, 0101 is not a primitive word of length 4, as it is a repetition of a shorter word 01 of length 2. Clearly, the number of primitive words of length n relates to the divisors d of n. For a given alphabet of k symbols, the number of primitive words $\psi_k(n)$ of length n may be computed [154, 141, 60, 7] in terms of the Möbius function $\mu(d)$:

$$\mu(d) = \begin{cases} 0, & \text{if } d \text{ is divisible by some } x^2 > 1, \text{ with } x \in \mathbb{N}; \\ (-1)^s, & \text{if } d = p_1 p_2 \ldots p_s, \text{ where } p_i \text{ are distinct primes.} \end{cases} \tag{2.12}$$

The number of primitive words is then given by[9]:

$$\psi_k(n) = \sum_{d|n} \mu(d) k^{n/d}. \tag{2.13}$$

The expression for the number of minimal patterns of length ℓ over a k-symbol alphabet is, thus,

$$N_k(\ell) = \psi_k(\ell) + \sum_{i=1}^{\ell-1} (k-1) k^{i-1} \psi_k(\ell - i), \tag{2.14}$$

where the term $\psi_k(\ell)$ counts the number of length-ℓ cycles (i.e., case of no tail), whereas $(k-1)k^{i-1}\psi_k(\ell-i)$ counts the number of length-$(\ell - i)$ cycles together with length-i tails, where one element of the tail is constrained by condition (i) above, giving the $(k-1)k^{i-1}$ factor. In light of these results, we may refer to minimal patterns simply as patterns, disregarding the non-minimal case.

2.6 Sub-deterministic scenarios

Having characterized the deterministic threshold, we now know precisely where to look for the sub-deterministic scenarios, where classical and quantum memories can be distinguished. Our goal now is

[8] Entry A059412 in the On-Line Encyclopedia of Integer Sequences [146].
[9] Entry A143324 in the On-Line Encyclopedia of Integer Sequences [146].

find the maximum probabilities $\Omega(\boldsymbol{a}, d)$ that a sequence $\boldsymbol{a}$ can be generated when $d < \mathrm{DC}(\boldsymbol{a})$, for both classical and quantum memories. The general problem is likely intractable in closed form, so we resort to numerical optimization techniques. Here, two general approaches are available:

- Obtaining upper bounds $\widehat{\omega}(\boldsymbol{a}, d)$ through outer approximations, or
- Obtaining lower bounds $\omega(\boldsymbol{a}, d)$ by optimizing over explicit models.

Used together, the two approaches establish a range of possible values for the maximum,

$$\omega(\boldsymbol{a}, d) \leq \Omega(\boldsymbol{a}, d) \leq \widehat{\omega}(\boldsymbol{a}, d), \tag{2.15}$$

such that, if we manage to find $\omega(\boldsymbol{a}, d) = \widehat{\omega}(\boldsymbol{a}, d)$, we can certify the exact value of $\Omega(\boldsymbol{a}, d)$. On this point, it is worth mentioning that a loose classical upper bound is of limited use, as a violation of the inequality $p(\boldsymbol{a}|d) \leq \widehat{\omega}_C(\boldsymbol{a}, d)$ would not necessarily reveal the presence of non-classical memory effects. Therefore, the two approaches are best used in conjunction.

The numerical results in this chapter, as published in Ref. [198], relied exclusively on computing lower bounds. Nevertheless, this section will provide a broader discussion of the general problem, including connections to recent developments involving the computation of upper bounds. We first discuss the technical aspects of these approaches, and how the classical and quantum problems require slightly different strategies to be solved. We then present the technical details of the specific implementations we have used for computing the lower bounds $\omega(\boldsymbol{a}, d)$. The results of our numerical optimizations, performed over all sub-deterministic classical and quantum scenarios up to length $L = 10$, will be discussed in great detail in Secs. 2.7 and 2.8.

2.6.1 Computing upper bounds

Ideally, we would like to obtain the exact maxima $\Omega(\boldsymbol{a}, d)$ in Eq. (2.6), which provide a sharp characterization of the achievable correlations for classical or quantum memories for each d. These are global optimization problems, involving non-linear objective functions with nontrivial constraints. Furthermore, the problems are also inherently high-dimensional, with Ad^2 real-valued parameters for classical models, and Nd^2 complex-valued parameters for quantum models[10]. Unfortunately, these facts suggest that computing the exact values will, in general, be very difficult.

A straightforward approach (at least conceptually) is through the branch-and-bound method, where a systematic brute-force search of the parameter space is performed. First, the space is split it into multiple regions, each with its own lower and upper bounds for the objective function[11]. Unfeasible regions can be discarded by comparing their bounds with the current best value known for the objective, while feasible regions are split into finer subregions. The process repeats until a desirable accuracy is reached. This method is quite general and "embarrassingly parallel"[12], making it suitable for computations using modern GPUs. It was used in Ref. [33], in the context of ticking clocks, for investigating the optimal classical and quantum bounds for one-tick sequences and $d = 2$. However, the number of regions grows exponentially with the dimension of the search space (i.e., the number of parameters in the model), which quickly renders this approach computationally intractable.

The alternative is to formulate outer approximations of the problem, allowing us to (in principle) obtain upper bounds $\widehat{\omega}(\boldsymbol{a}, d)$.

[10] This is without considering the constraints. For example, the number of degrees of freedom in classical models is $Ad^2 - d$ due to the stochasticity constraints $\sum_{a,j}[T_a]_{ij} = 1$ for each i. For the quantum case, each of the $N = \sum_a n_a$ Kraus operators has d^2 degrees of freedom, but the non-linear constraints $\sum_{a,k} K_{ak}^\dagger K_{ak} = \mathbb{1}$ reduce this down to $(N-1)d^2$.

[11] There are multiple ways to establish these bounds, e.g., bounding the absolute value of the objective function's gradient, as was done in Ref. [33], or through the use of Bernstein coefficients, as extensively discussed in Ref. [194]. These bounds are typically loose, but in order to be useful they must be easier to compute and converge uniformly as the region shrinks.

[12] A term used in the parallel computing literature, referring to problems which are trivial to parallelize.

Classical case

In the classical case, the probability in Eq. (2.1) expands into a positive and homogeneous multivariate polynomial of degree L in Ad^2 variables, with Ad^2 inequality constraints, $[T_a]_{ij} \geq 0$ for all a, i, j, and d equality constraints, $\sum_{aj}[T_a]_{ij} = 1$ for all i. The method of Lasserre and Parrilo [110, 148] is applicable, providing a hierarchy of SDP problems for $\widehat{\omega}(\boldsymbol{a}, d)$ which converge to the global maximum (see also Ref [113] for a survey).

In Ref. [207], a hybrid method based on dynamic programming principles was proposed, which is capable of bounding general time-ordered processes. Using their method, the authors were able to certify as exact some of the classical bounds found in Refs. [33, 198], also discussed in Sec. 2.7.2.

However, all of these approaches are extremely computationally demanding already for $d \geq 3$, so their application is limited. It is uncertain whether the specific structure of our problem can be exploited in order to simplify these computations.

Quantum case

In Ref. [201], which is the subject of Ch. 3, we formulate a hierarchy of SDPs capable of bounding — in principle, exactly — the set of temporal correlations in the scenario involving sequential measurements on an open quantum system. While that work addresses a much more general quantum scenario than the one discussed in this chapter, the techniques therein can — and were — used to obtain $\widehat{\omega}_Q(\boldsymbol{a}, d = 2)$ for short sequences.

This problem was also exceptionally difficult to solve numerically, and originally deemed intractable due to its computational requirements, in particular the amount of memory. Fortunately, we overcame these challenges by developing a novel technique to induce sparsity in SDPs, greatly reducing the number of variables and constraints, which rendered the problem tractable. The technique is very general and was developed further into Ref. [199], which is the subject of Ch. 5.

2.6.2 Computing lower bounds

Before this work, not much was known about the detailed structure of our problem, let alone about the optimal models and the maximum probabilities they achieve when generating each sequence. With the aim of gathering useful insights, we performed comprehensive numerical surveys of all sub-deterministic scenarios (classical and quantum), up to a length $L = 10$, obtaining lower bounds $\omega(\boldsymbol{a}, d)$ together with their respective models.

The surveys were performed using gradient descent[13] (GD) techniques, in particular, through a custom Python implementation of the Adam algorithm [100] using NumPy [84], SciPy [202], and further optimized through Numba [107]. Note that the classical survey originally performed in Ref. [198] used Adam as implemented in PyTorch [149], with only the quantum survey relying on a custom implementation. For this chapter, we refined all previous classical and quantum results by performing new and more accurate optimizations using a custom implementation.

All sequences of length $3 \leq L \leq 10$ with $2 \leq d < \mathrm{DC}(\boldsymbol{a})$ were investigated, for a total of 4 328 nontrivial scenarios. Note that $d = 1$ scenarios do not require optimization (see Sec. 2.6.3), hence $L \geq 3$. Due to the $0 \leftrightarrow 1$ relabeling symmetry of the problem, we consider only half of the sequences by assuming $a_1 = 0$. Together, the classical and quantum surveys took just under three weeks of continuous computation, running in parallel on 16 CPU cores.

Because our sequences are short, the issues of vanishing gradients and numerical underflow were not a major concern. In general, we found optimal (classical) models to be sparse, and with transition

[13] Or, more appropriate in our case, "gradient ascent", although it is a less common term.

probabilities (or amplitudes) on the order of 10^{-1}. Therefore, the survey was well within the numerical precision available using 64-bit floating-point numbers ($\approx 10^{-16}$).

Next, we explain the specific formulation of the optimization problems in the classical and quantum cases.

Classical case

The classical case consists of the following constrained optimization problem:

$$\begin{aligned} \textbf{Optimize:} \qquad & \omega_C(\boldsymbol{a}, d) := \max_T p(\boldsymbol{a}|T, d) = \max_T \pi T_{a_1} \cdots T_{a_L} \eta \\ \textbf{Subject to:} \qquad & \sum_{a\in\mathcal{A}} \sum_{j=1}^{d} [T_a]_{ij} = 1, \ \forall\, i, \quad \text{and} \quad [T_a]_{ij} \geq 0, \ \forall a, i, j, \end{aligned} \tag{2.16}$$

Note that, since the objective function is convex with respect to the initial state π, without loss of generality we can assume $\pi = (1, 0, \ldots, 0)$.

To make the problem amenable to the GD method, we convert it into an equivalent unconstrained problem as follows. First, we define the real-valued $d \times d$ matrices B_0 and B_1, which are unconstrained. We can convert these into T_0 and T_1 in the constrained problem via the normalization procedure:

$$[T_a]_{ij} = \frac{[B_a]_{ij}^2}{\sum_{a,\ell} [B_a]_{i\ell}^2}. \tag{2.17}$$

This adjustment is all that is required in the classical case. The results of the classical survey are discussed in Sec. 2.7.

Quantum case

The construction of explicit quantum models is a more involved task. Recall that, in the quantum case, transitions are described by the instrument $\mathcal{I} = (\mathcal{I}_0, \mathcal{I}_1)$, where each effect $\mathcal{I}_a$ can be written concretely in their Kraus decomposition:

$$\mathcal{I}_a(\rho) = \sum_{k=1}^{n_a} K_{ak} \rho K_{ak}^\dagger, \qquad \text{with} \ \sum_{a\in\mathcal{A}} \sum_{k=1}^{n_a} K_{ak}^\dagger K_{ak} = \mathbb{1}. \tag{2.18}$$

The quantum case then consists of the following constrained optimization problem:

$$\begin{aligned} \textbf{Optimize:} \qquad & \omega_Q(\boldsymbol{a}, d) := \max_{\mathcal{I}} p(\boldsymbol{a}|\mathcal{I}, d) = \max_{\mathcal{I}} \operatorname{Tr}\left[\mathcal{I}_{a_L} \circ \cdots \circ \mathcal{I}_{a_1}(\rho)\right] \\ \textbf{Subject to:} \qquad & \sum_{a\in\mathcal{A}} \sum_{k=1}^{n_a} K_{ak}^\dagger K_{ak} = \mathbb{1}, \end{aligned} \tag{2.19}$$

For simplicity, we assume $n_0 = n_1 = n$. As before, since the objective function is convex with respect to the initial state ρ, we may assume $\rho = |0\rangle\langle 0|$.

An unconstrained problem can be obtained as follows. First, define arbitrary $d \times d$ complex matrices B_{ak} as our optimization variables, then compute the matrix

$$E = \sum_{a,k} B_{ak}^\dagger B_{ak}, \tag{2.20}$$

which by construction is positive semidefinite. Next, let $\lambda_{\max}$ be the maximum eigenvalue of E. We can now define the (approximate) Kraus operators by the normalization,

$$\widetilde{K}_{ak} := \frac{B_{ak}}{\sqrt{\lambda_{\max}}}, \tag{2.21}$$

which then automatically satisfies the weaker constraint

$$\widetilde{E} := \sum_{a\in\mathcal{A}} \sum_{k=1}^{n_a} K_{ak}^\dagger K_{ak} \leq \mathbb{1}. \tag{2.22}$$

While there is no guarantee that the equality in Eq. (2.22) will be satisfied, this is not strictly necessary for our optimization purposes as the resulting gradient will always naturally favor $\widetilde{E} \approx \mathbb{1}$. This can be interpreted as if our experiment contained a new potential output, $\perp$, whose effect $\mathcal{I}_\perp$ corresponds to the missing Kraus operators required to complete the Kraus condition exactly. Since the output $\perp$ does not occur in the sequences we are considering, the contribution of $\mathcal{I}_\perp$ will vanish in the optimization of probability for these sequences. Note that the objective function now requires computation of the maximum eigenvalue of E at every evaluation, which is a nontrivial mathematical operation. This made a custom implementation of the Adam algorithm a requirement for this survey.

An alternative to this approximation is performing a constrained optimization directly over the space of Kraus operators, which is an example of Riemannian optimization on the Stiefel manifold [186]. We have not pursued such refinements, and the approximation above converged up to numerical precision in all scenarios investigated.

The results of the quantum survey are discussed in Sec. 2.8.

2.6.3 Exact value for $d = 1$

When $d = 1$, the memory cannot store any information, i.e., the scenario is *memoryless*, and classical and quantum maxima coincide. While the case is sufficiently simple as to be solved in closed form with elementary techniques, it still provides some valuable insights. The ideas in this subsection were originally discussed in Ref. [201] (subject of Ch. 3), but we present it here instead for completeness of this section.

We formulate the proof in the classical case. Recall that the probability of a sequence can be written as

$$p(\boldsymbol{a}|T, d) = \pi T_{a_1} \cdots T_{a_L} \eta. \tag{2.23}$$

But if $d = 1$, the matrices T_a reduce to scalars, i.e., $[T_a]_{11} = q_a \geq 0$, corresponding directly to the probability for each output, with $\sum_a q_a = 1$. Similarly, we have $\pi = \eta = 1$, so that

$$p(\boldsymbol{a}|d = 1) = \prod_{t=1}^{L} q_{a_t} = \prod_{a\in\mathcal{A}} q_a^{\ell_a}, \tag{2.24}$$

where $\ell_a \in \mathbb{N}$ is the number of occurrences of the symbol a, with $\sum_a \ell_a = L$. For simplicity, we may assume that $\boldsymbol{a}$ contains every symbol of $\mathcal{A}$ at least once, such that $\ell_a, q_a > 0$, otherwise we could simply assume a smaller $\mathcal{A}$ where this is the case. The maximum probability $\Omega(\boldsymbol{a}, d = 1)$ can be found with the standard technique of Lagrange multipliers. Using the Lagrangian

$$\mathcal{L} = \prod_{a\in\mathcal{A}} q_a^{\ell_a} - \alpha \left(\sum_{a\in\mathcal{A}} q_a - 1 \right), \tag{2.25}$$

we calculate the partial derivatives for each q_a, and equate them to zero:

$$\frac{\partial \mathcal{L}}{\partial q_a} = \left(\frac{\ell_a}{q_a} \right) p - \alpha = 0, \ \forall a. \tag{2.26}$$

Since $q_a > 0$ for all a, $p > 0$, we can rewrite

$$\frac{\ell_a}{q_a} = \frac{\alpha}{p} = \gamma \quad \rightarrow \quad q_a = \frac{\ell_a}{\gamma}, \ \forall a. \tag{2.27}$$

Finally, summing over a and applying the constraints $\sum_a q_a = 1$ and $\sum_a \ell_a = L$, we obtain:

$$\sum_a q_a = \frac{1}{\gamma} \sum_a \ell_a \quad \rightarrow \quad \gamma = L. \tag{2.28}$$

Thus, the (classical or quantum) maximum probability is

$$\Omega(\boldsymbol{a}, d = 1) = \prod_{a \in \mathcal{A}} \left(\frac{\ell_a}{L} \right)^{\ell_a}, \qquad \text{for} \quad q_a = \frac{\ell_a}{L}, \ \forall a \tag{2.29}$$

As this solution is unique if $q_a > 0$ and $p(\boldsymbol{a}) > 0$, this is indeed the global maximum. Given Eq. (2.7), the result of Eq. (2.29) establishes a nontrivial lower bound for all $\Omega(\boldsymbol{a}, d)$.

2.7 Survey of classical models

In the classical case, gradient descent optimization was performed for the $d > 1$ cases, using hyperparameters $\alpha = 0.05$, $\beta_1 = 0.99$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-10}$; see Ref. [100]. Each scenario was optimized multiple times in order to obtain 20 samples of the optimal probability. Each sample consisted of multiple trials, with each trial randomly initializing the matrices B_a with entries uniformly sampled from the interval $[-1, 1)$, then followed by 20 000 iterations of the Adam algorithm. A trial is deemed successful, producing a new sample, if its optimum is $\geq 90\%$ the value of the current best sample, and unsuccessful otherwise. In this way, trials were run indefinitely until 20 successful samples were collected, with each improved sample enforcing stricter thresholds on subsequent ones.

2.7.1 Results for general sequences

Under finite-memory constraints, models must make nontrivial use of the memory resources available. This gives rise to complex behaviors specific to each sequence, as in Fig. 2.4.

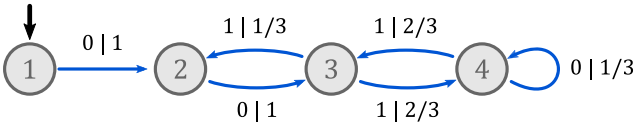

Figure 2.4: The optimal model found for the sequence $\boldsymbol{a} = 001011$ ($\mathrm{DC}(\boldsymbol{a}) = 5$) and $d = 4$, which results in $p(\boldsymbol{a}|d = 4) = 8/27 = 0.\overline{296}$. The model uses a nontrivial combination of probabilistic and deterministic transitions. Note also the non-unifilar behavior, as seen by the two transitions out of state 3 outputting $a = 1$.

Table 2.1 shows the behavior of various $d = \mathrm{DC}(\boldsymbol{a}) - 1$ models found in the survey. Optimal models are typically sparse, suggesting the possibility of a strategy for selecting optimal transition topologies in advance, before optimizing their probabilities. This sophisticated approach would significantly reduce the number of variables for large d, possibly enabling the computation of exact values for the maximum probabilities, e.g., with the techniques discussed in Sec. 2.6.1. However, this requires a deeper understanding of how optimal transition topologies emerge, based on some inherent structure of each sequence.

Up to numerical accuracy, most optimal probabilities found seem to be simple rational numbers (e.g. $1/4, 8/27, 512/3125$), but this is not always the case. For example, the optimal probability for $\boldsymbol{a} = 00100$ with $d = 2$ results from a model with the following general structure,

$$T_0 = \begin{bmatrix} x & 1-x \\ 0 & y \end{bmatrix}, \qquad T_1 = \begin{bmatrix} 0 & 0 \\ 1-y & 0 \end{bmatrix}, \tag{2.30}$$

which is sufficiently simple to be solved analytically. Using a computer algebra software, the unique global maximum (Fig. 2.5) can be found to be the irrational value

$$\omega_C(00100, 2) = \frac{4\left(411 - 41\sqrt{41}\right)}{3125} = 0.19004404001952...., \tag{2.31}$$

which matches the result obtained via gradient descent (Table 2.1). Given the small number of states in this scenario, this value is likely the exact maximum $\Omega_C(00100, 2)$.

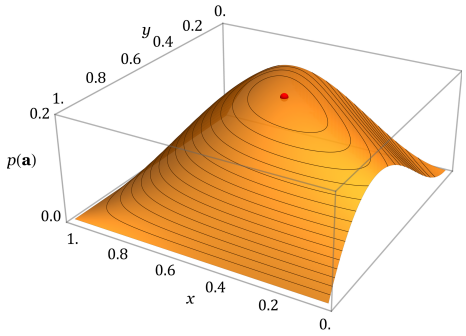

Figure 2.5: The global maximum (red dot) for the specific model in Eq. (2.30) is an irrational number. This value matches the optimal probability $\omega_C(00100, 2)$ found numerically.

On this note, an interesting observation is that, despite the large variety of structure in these models and the respective sequences, nearly all optimal probabilities seem to fall into a few equivalence classes. This seems to hint at the existence of some important shared property between sequences, which, despite all of our efforts, has yet to be uncovered. We leave these investigations for future research.

The results of the survey pointed towards a special property of the one-tick sequences, which required further investigation: The one-tick sequence (possibly truncated to DC) always outperformed all other sequences for the same number of states, offering an upper bound for any (L, d) combination. This important property will be discussed in detail in the next sections.

Finally, notwithstanding our ignorance about these hinted structures, the fact all sequences perform worse than the one-tick sequence may be interpreted as follows: There is a nontrivial trade-off occurring when models attempt to reach a higher probability with less resources. The way other sequences switch between 0 and 1 multiple times forces the memory resources to be spread over multiple incompatible transitions, leading to a worse performance overall. In contrast, one-tick sequences are almost uniform, allowing transitions to display a certain degree of redundancy.

2.7.2 Optimal classical models for one-tick sequences

Given their special role in the problem at hand, the optimal models and probabilities for one-tick sequences warranted a detailed analysis. In Ref. [33], a class of models was proposed as optimal for the one-tick sequences. We have since found a further generalization providing higher probabilities in certain scenarios.

Due to their sparse structure, these models are unusually difficult to find through unconstrained numerical optimizations. In the following, we investigate the properties of these optimal models.

Analysis of one-tick sequences

In addition to the general survey using gradient descent, the one-tick sequences were also analyzed in greater detail, both with analytical and numerical methods. These scenarios were already investigated in Ref. [33], where a specific model referred to as the *multicyclic model* was shown to be optimal in some

cases, and originally thought to be optimal in general. This model included, as special cases, the *one-way model* and the *cyclic model*, as shown in Fig. 2.6.

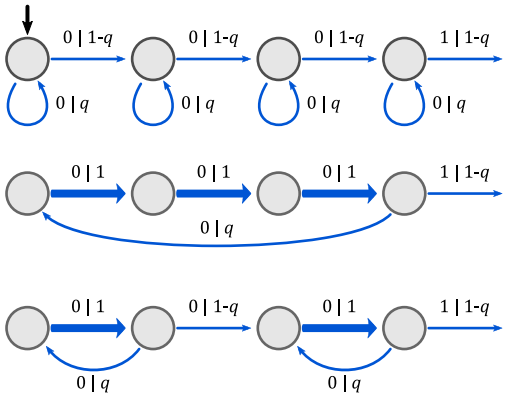

Figure 2.6: Different models for the one-tick sequences in $d = 4$, from top to bottom: one-way model, cyclic model, and multicyclic model (with cycles of length 2). Deterministic transitions are emphasized with thick arrows. In the one-tick sequence, the state after the final transition on outcome $a_L = 1$ is irrelevant. The initial state can be any state within the first cycle.

The optimal probability for the one-way model, for $d < L$, can be written in closed form as [33, 198]:

$$F_{\text{ow}}(L, d) := \binom{L-1}{d-1}\left(1 - \frac{d}{L}\right)^{L-d}\left(\frac{d}{L}\right)^{d}, \tag{2.32}$$

where d/L corresponds to the probability of forward transitions and $1 - d/L$ of self-transitions. One-way models were found to be generally best for $d = L - 1$, whereas the (multi)cyclic models were found to outperform it for certain $d < L - 1$. The multicyclic model divides the states into equal cycles, with probabilistic transitions only between cycles and deterministic transitions within. These restrictions imply that the total number of cycles n and the size of the cycles k must obey $d = nk$.

In Ref. [198] we have shown that the multicyclic model can be improved upon with a slight modification, providing better lower bounds on the maximum probability. We have named these *enhanced multicyclic models* (EMCMs), in analogy to the multicyclic model in Ref. [33]. While the optimality of these new models for all sub-deterministic scenarios involving one-tick sequences is still conjectured, we believe the results to be discussed in Sec. 2.7.2 provide a strong argument that this is the case.

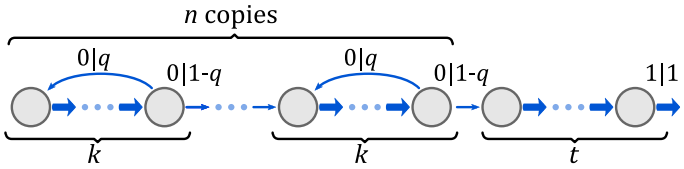

Figure 2.7: The structure of enhanced multicyclic models (EMCMs), consisting of n identical probabilistic cycle blocks of size $k \in \mathbb{N}$, followed by a "tail" of $t \geq 0$ states such that $d = nk+t$. Deterministic transitions, within each cycle and at the tail, are shown as thick arrows. The initial state (not shown) can chosen to be any of the first k states, in the first block, and this freedom provides an additional advantage in a few scenarios.

A schematic representation of EMCMs is presented in Fig. 2.7. The model consists of n blocks of size k, each forming its own cycle, followed by $t \geq 0$ deterministic transitions, such that $d = nk+t$. The transitions within each cycle block are also deterministic, thus incurring no penalty in the final probability. In the last state of each cycle, the machine can either cycle through the block with probability q, or transition to

the next one with probability $1-q$. All these transition are associated with the output 0, and once the last state of the last cycle is reached, if $t > 0$, the machine can either cycle again with probability q, or step into the final "tail" block with deterministic transitions. At the end of the final t deterministic transitions, the machine emits the outcome 1 with certainty, after which the subsequent state transition is irrelevant. In the special case where $t = 0$, i.e., no deterministic transitions at the end, the output 1 is associated to the forward transition of the last cycle. As a concrete example, an EMCM with $k = 2, n = 2, t = 1$ is given by the following:

$$T_0 = \left[\begin{array}{cc|cc|c} 0 & 1 & 0 & 0 & 0 \\ q & 0 & 1-q & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & q & 0 & 1-q \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right], \quad T_1\eta = \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{array}\right], \tag{2.33}$$

where we use $T_1\eta$ to emphasize that the final transition may go to an arbitrary state. Note that the initial state must be chosen within the first block, otherwise the states of previous blocks are never used, in contradiction with the assumption of a minimal model. The special case $t = 0$ reduces to the earlier models, with $n = d$ and $k = 1$ corresponding to the one-way model, $d = nk$ the multicyclic model, and $n = 1$, $k = d$ the cyclic model.

The optimal probability q depends on the structure of the model and the length of the sequence. An EMCM structure can be described with five parameters

$$\begin{aligned} L &\quad \text{the sequence length for the model,} \\ n &\quad \text{the number of cycle blocks,} \\ k &\quad \text{size of the cycle blocks,} \\ t &\quad \text{size of the deterministic block,} \\ z &\quad \text{initial state shift,} \end{aligned} \tag{2.34}$$

with $L, n, k, t, z \in \mathbb{N}$, $k > 0$, $n = (d-t)/k$, and $z \leq k-1$, where $z = 0$ means we start from the initial state, $z = 1$ from the second, and so on. Note that only k and z are independent parameters of the model, whereas n and t depend jointly on the model and the sequence. We take Eq. (2.34) as a *definition* of the EMCM, and will denote optimal EMCMs (in the given context) simply by E, instead of T.

Denoting a generic EMCM $T^{(L,n,k,t,z)}$, we can compute the probability for the $\boldsymbol{a}_{\text{ot}}^L$ sequence as

$$\begin{aligned} p(\boldsymbol{a}_{\text{ot}}^L | T^{(L,n,k,t,z)}, d) &= F_{\text{ow}}(\lambda, \delta) \\ \text{with} \quad \lambda = \frac{L-t+z}{k}, &\quad \delta = \frac{d-t}{k} \end{aligned} \tag{2.35}$$

where, in addition to previous constraints, z satisfies $(L-t+z)/k \in \mathbb{N}$. Given Eq. (2.35), we may thus optimize directly the parameters k and z and obtain (L, n, k, t, z) for any (L, d). Next, we discuss further details and intuitions for these results, in particular the meaning of (λ, δ).

The structure and optimality of EMCMs were found by a systematic analysis of a generalization of the multicyclic model from Ref. [33]. This analysis is discussed in Sec. 2.10, in this chapter's appendix.

Reducible sequences and models

There is an intuitive way to understand the advantage provided by EMCMs. The initial shift z can be interpreted as "increasing" the sequence length (without affecting d): Since the transitions within each cycle block are deterministic, starting from the state $z+1$ is equivalent to lengthening the sequence by z steps and starting on the first state. Similarly, the final sequence of t deterministic transitions does not alter the probability, defined by the cycle blocks. Thus, one may "trim" these deterministic transitions by removing

t steps in the sequence (all outputting 0), together with t states. The result of both transformations is a multicyclic model, starting at the first state, with the same probability as before.

Finally, each block of size k is equivalent to a single state of an one-way model of reduced dimension, effectively a time-scaling by a factor of $1/k$. In other words, the initial state z and the final deterministic block t both work together to perfectly synchronize the cycles with the number of transitions. In summary, the optimal probability for an EMCM with d states and one-tick sequence of length L can, then, easily be computed by performing the following transformations, all of which leave the probability unchanged:

1. For a given L, d, pick $k, z \in \mathbb{N}$ obeying $1 \leq k \leq d$ and $0 \leq z \leq k-1$. Compute $n = \lfloor d/k \rfloor$ and $t = d - nk$.

2. If $t > 0$, we may discard the entire final deterministic block, effectively reducing the dimension, if we also shorten the one-tick sequence by the same amount t (removing t 0s). In the process, we incorporate the deterministic output 1 into the final probabilistic transition forwards in the last state of the last cycle. In Table 2.4, this corresponds to moving diagonally towards the top-left corner by t cells.

3. If the initial state is not the first state, i.e., $z > 0$, we may increase the sequence length by adding z 0s, as the transitions within cycle blocks are deterministic. This effectively shifts the initial state to the first state of the first block. This corresponds to moving down in Table 2.4 by z cells.

4. Once these transformations have been performed, we have transformed the problem into the $L \mapsto L - t + z$, $d \mapsto d - t$ scenario with a simple multicyclic model.

5. Finally, since all blocks have the same length k with $k-1$ deterministic transitions, we may divide both the new L and d by k, obtaining a one-way model with length $\lambda = \frac{L-t+z}{k}$ and number of states $\delta = \frac{d-t}{k}$, with probability $F_{\text{ow}}(\lambda, \delta)$.

Thus, we then have the model and sequence in an irreducible canonical form: A one-way model with δ states, generating a one-tick sequence of length λ with the same probability as the original sequence and EMCM. We may then directly compute this probability using Eq. (2.32). The only free parameters in such an approach are the optimal k and z, which require a straightforward numerical optimization on the order of $O(d^2)$ to be found.

This optimality of enhanced multicyclic cycles for one-tick sequences, due to their reducibility, leads to an alternative physical interpretation of these models. The resulting uniform multicyclic models with cycle lengths k represent a "time-scaling symmetry". Since every $k-1$ transitions occur deterministically, they may be considered as mere *delays* which introduce no relevant dynamics to the behavior of the model. The result of this reduction is a "time scaling" by a factor of $1/k$, which produces an irreducible version of the model that captures all of its dynamics in the smallest number of transitions and states.

Note: The following contains additional material not originally included in Ref. [198].

Inspired by this interpretation, a systematic analysis of the optimal EMCMs for various (L, d) scenarios suggests that their optimal reduced (λ, δ) can be calculated directly using

$$\lambda = \begin{cases} \left\lceil \frac{L}{d} \right\rceil & \text{if } {}^{d}/_{L} < {}^{1}/_{2} \\ \left\lceil \frac{d+1}{L-d} \right\rceil & \text{if } {}^{d}/_{L} \geq {}^{1}/_{2} \end{cases}, \quad \text{and} \quad \delta = \begin{cases} 1 & \text{if } {}^{d}/_{L} < {}^{1}/_{2} \\ \left\lceil \frac{d+1}{L-d} \right\rceil - 1 & \text{if } {}^{d}/_{L} \geq {}^{1}/_{2} \end{cases} \tag{2.36}$$

where $\lceil x \rceil$ is the ceiling function[14], which gives the optimal EMCM probability for any given (L, d), in

[14] $\lceil x \rceil$ rounds x up to the nearest integer.

closed form, as:

$$F(L,d) := \begin{cases} 1, & {}^{d}/_{L} \geq 1 \\ F_{\text{ow}}\left(\left\lceil \frac{d+1}{L-d} \right\rceil, \left\lceil \frac{d+1}{L-d} \right\rceil - 1\right), & {}^{1}/_{2} \leq {}^{d}/_{L} < 1 \\ F_{\text{ow}}\left(\left\lceil \frac{L}{d} \right\rceil, 1\right), & {}^{d}/_{L} < {}^{1}/_{2} \end{cases} . \tag{2.37}$$

While this expression was obtained empirically[15] it can be intuitively understood as follows. If ${}^{d}/_{L} < {}^{1}/_{2}$, then the optimal EMCM minimizes states with probabilistic transitions, leading to a cyclic model which reduces to a single state. Otherwise, if ${}^{d}/_{L} > {}^{1}/_{2}$, the optimal EMCM maximizes states with probabilistic transitions, reducing to a one-way model with $\delta = \lambda + 1$, which provides the maximum probability for a given $\delta < \lambda$. For ${}^{d}/_{L} = {}^{1}/_{2}$, both strategies lead to $\lambda = 2$ and $\delta = 1$.

2.7.3 Conjectured upper bounds

Upper bounds for sub-deterministic scenarios

With the significant improvements obtained by optimal EMCMs, all the data available thus far seems to confirm a striking property of one-tick sequences: They seem to achieve the highest probability out of any sequence if $d < \mathrm{DC}(\boldsymbol{a})$ and $L \leq \mathrm{DC}(\boldsymbol{a})$. In particular, the optimal probability for the one-tick sequence of length $\mathrm{DC}(\boldsymbol{a})$ at dimension d seems to act as an upper-bound for all sequences with the same DC and d. While a proof of this is yet to be found, the current results lead us to formulate the following conjecture:

Conjecture 1. The maximum probability of any sequence $\boldsymbol{a}$ occurring, in any classical scenario with $d < \mathrm{DC}(\boldsymbol{a})$, is upper-bounded by the probability of the one-tick sequence $\boldsymbol{a}_{\text{ot}}^{L}$ with length $L = \mathrm{DC}(\boldsymbol{a})$, using its optimal EMCM on d states, i.e.:

$$p(\boldsymbol{a}|T,d) \leq p(\boldsymbol{a}_{\text{ot}}^{\mathrm{DC}(\boldsymbol{a})}|E,d), \quad \forall \boldsymbol{a}, T, d < \mathrm{DC}(\boldsymbol{a}). \tag{2.38}$$

This would be a remarkable result, establishing a direct connection between deterministic and sub-deterministic scenarios through the DC of a sequence. In particular, the left-hand side of Eq. (2.38) involves an arbitrarily complex sequence being generated by an arbitrary classical model T on d states, whereas the right-hand side subsumes all of this structure into a single number, $\mathrm{DC}(\boldsymbol{a})$, thereby providing an upper-bound on all possible classical behaviors. Furthermore, with the results of Eq. (2.37), we could compute this upper bound directly with $F(\mathrm{DC}(\boldsymbol{a}), d)$.

A graphical representation of the survey results, together with this conjectured bound, is shown in Fig. 2.8, where it is clear that even tighter bounds might be possible. This would require identifying additional structures in the sequences, and the origin of the probability equivalence classes which appear as "plateaus".

Universal classical bound

If proven, the previous conjecture leads to a more surprising statement regarding classical scenarios, which is presented in the following.

Observation 4. If Conj. 1 holds, then it follows that:

- $p(\boldsymbol{a}|T,d) < 1/\mathrm{e}, \quad \forall\, \boldsymbol{a} \in \mathcal{A}^{L}, T$, if $d < \mathrm{DC}(\boldsymbol{a})$,

  or equivalently,

- $p(\boldsymbol{a}|T,d) \geq 1/\mathrm{e} \implies d \geq \mathrm{DC}(\boldsymbol{a})$.

[15] Verified up to $L = 500$ using arbitrary precision arithmetic.

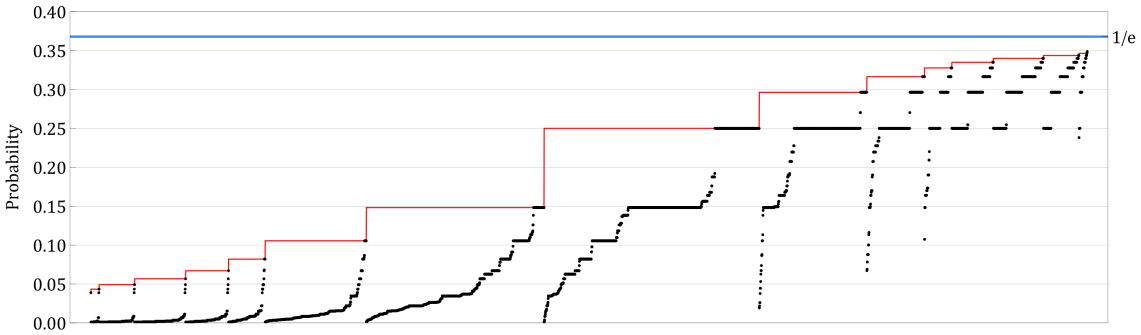

Figure 2.8: Optimal probabilities for all sub-deterministic scenarios in the classical survey (black dots). The red line is the EMCM upper-bound as described in Conj. 1. For clarity, the scenarios were sorted first by increasing EMCM bound, then by increasing probability. At the top, the conjectured universal classical bound $1/\mathrm{e}$ is shown (see Sec. 2.7.3).

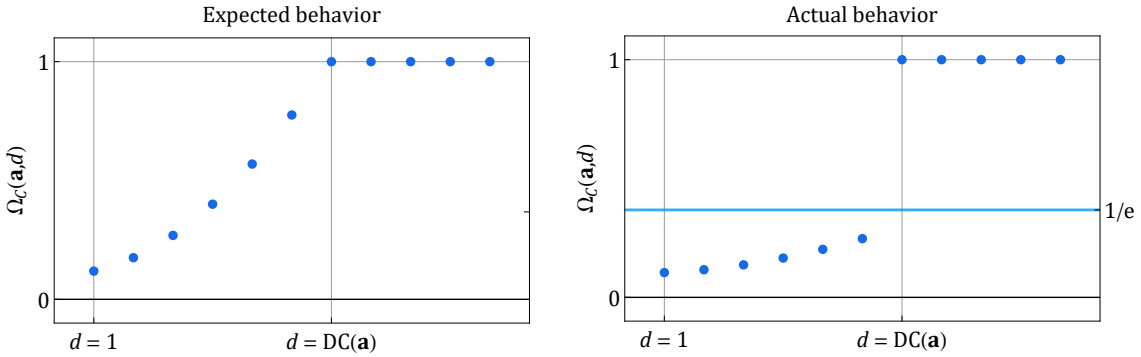

Figure 2.9: The expectation that $\Omega_C(\boldsymbol{a}, d)$ should increase gradually towards 1 as d increases is contradicted by our results, which reveal a sharp transition between the sub-deterministic and deterministic scenarios.

Proof. To show this, it is sufficient to notice that

$$\begin{aligned} p(\boldsymbol{a}_{\mathrm{ot}}^{L}|E,d) &\leq \sup_{L,\text{ and } d<L} F(L,d) = \sup_{L} F(L,L-1) \\ &= \lim_{L\to\infty} \left(1-\tfrac{1}{L}\right)^{L} = \frac{1}{\mathrm{e}}\,, \end{aligned} \tag{2.39}$$

where we used Eq. (2.37) as a value for the EMCM, the fact $F_{\mathrm{ow}}(L,d)$ is monotonically increasing in d, and that the convergence of $(1-1/L)^L$ is monotone, i.e., $1/\mathrm{e}$ is the upper bound for all L. □

Therefore, provided the memory is classical, any probabilistic realization with $p(\boldsymbol{a}) \geq 1/\mathrm{e}$ requires at least the same dimension as a deterministic realization. In other words: There is a *universal upper bound* of $1/\mathrm{e}$ on the probability of generating any sequence in its sub-deterministic scenarios with classical memory. These ideas are concisely shown in Fig. 2.8.

This result was surprising, especially when considering Eq. (2.7), as one might expect that $\Omega_C(\boldsymbol{a}, d)$ should increase gradually towards 1 with increasing d. What our results have shown, however, is that there seems to be a sharp transition between sub-deterministic and deterministic classical scenarios, as illustrated in Fig. 2.9.

Despite these promising results, it remains to be proven that the one-tick sequence and EMCMs provide such upper bounds. As we will see next, quantum models can easily violate these bounds, and no universal bound seems to exist in the quantum case.

2.8 Survey of quantum models

The quantum survey took considerably longer than the classical one, as quantum models involve a larger number variables and more computationally demanding constraints. As in the classical case, the quantum survey covered all nontrivial scenarios up to $L = 10$, using hyper-parameters $\alpha = 0.07$, $\beta_1 = 0.99$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-10}$, but only in the $n = 1$ case, i.e., a single Kraus operator per effect. We have also investigated possible advantages when using more than one Kraus operator, but this was restricted to $L \leq 5$ due to its computational cost, and the fact no advantage was observed up to that point.

Each scenario was optimized 20 times in order to obtain multiple samples for the optimal probability. As opposed to the classical case, each sample consisted of a single trial, each starting with matrices B_{ak} with random entries, where real and imaginary values were independently sampled from the interval $[-1, 1)$. This was followed by 15 000 iterations of the Adam algorithm. The best sample was used as the optimum.

2.8.1 Results for general sequences

Quantum advantages over the classical models were observed for nearly all scenarios, and many displayed a violation of the conjectured universal classical upper bound of $1/\mathrm{e}$. A sample of our results is shown in Fig. 2.10. These results indicate that the one-tick sequence also outperforms all other sequences in the quantum case, while at the same time providing the greatest advantage over its classical counterpart. Surprisingly, the results also show no benefit in using more than a one Kraus operator per $\mathcal{I}_a$. This will be discussed in more detail in Sec. 2.8.4.

As expected, if d is reduced further from the deterministic threshold, i.e., if $d < \mathrm{DC}(\boldsymbol{a}) - 1$, we generally see a worse performance in both classical and quantum models. Nevertheless, quantum models still outperform their classical counterparts in nearly all scenarios, including violations of the $1/\mathrm{e}$ bound. The results of our surveys also point towards the possibility of nontrivial scenarios where quantum memories offer no advantage, as shown in Fig. 2.11. This will be discussed further in Sec. 2.8.5.

2.8.2 Conjectured upper-bounds in the quantum case

Note: This section contains additional material not originally included in Ref. [198].

A natural question to ask is whether the conjecture for classical memories in Sec. 2.7.3 also applies to the quantum case. As we do not have a good understanding of optimal quantum models for one-tick sequences, here we analyze the quantum version of the conjecture in terms of the models found numerically. The result, as shown in Fig. 2.12, seems to indicate that a similar conjecture should hold for the quantum case, i.e.,

$$p(\boldsymbol{a}|\mathcal{I}, d) \leq \Omega_Q(\boldsymbol{a}_{\mathrm{ot}}^{\mathrm{DC}(\boldsymbol{a})}, d), \quad \forall\, \boldsymbol{a}, \mathcal{I}, d < \mathrm{DC}(\boldsymbol{a}). \tag{2.40}$$

As the structure of the optimal quantum models is typically very complex, we leave these questions open for future research.

2.8.3 No nontrivial universal quantum bound

Similarly, we also investigate whether a universal bound appears in the quantum case. To obtain some intuition on this problem, we analyzed the one-tick sequences for the special case $d = L - 1$. Despite the complexity of the exact problem, especially in high dimensions, the existence of a universal upper bound in the quantum case can be investigated using an appropriate *ansatz*, as we describe next.

For simplicity, we opt for using a single Kraus operator representation for each effect, each with one degree of freedom in the following parameterization. To automatically satisfy the constraints, the Kraus

operators are written in the polar decomposition $K_a = U_a\sqrt{E_a}$, with $E_a \geq 0$ and $E_0 + E_1 = \mathbb{1}$ and U_a a unitary matrix. We can choose the operators E_a to be diagonal and tailored for the target sequence (in our case, $\boldsymbol{a}^L_{\rm ot}$), based on the optimal classical model found, i.e., $E_a = \mathrm{diag}(\eta_a)$, where $\eta_a = T_a\eta$ for the optimal T_a of the classical model. As in [33], the one-parameter unitaries are defined in terms of a Fourier transform of the computational basis, namely,

$$U_a = \mathrm{e}^{-\mathrm{i}H\theta_a}, \quad H = F\left(\sum_{i=0}^{d-1} k\,|k\rangle\langle k|\right)F^\dagger, \quad \text{with } [F]_{jk} = \frac{1}{\sqrt{d}}\mathrm{e}^{\frac{2\pi\mathrm{i}}{d}jk}. \tag{2.41}$$

Thus, the only parameter of the model is the angle θ_0 appearing in Eq. (2.41), with E_0 fixed to be the diagonal matrix $\mathrm{diag}(1,1,\ldots,1,q)$, and where q taken as in the optimal classical model previously described, i.e., $q = 1 - d/(d+1)$.

Let us formulate the problem in this simplified form. Fixing the dimension by the condition $d = L-1$ and an initial state $\rho = |0\rangle\langle 0|$, without loss of generality, the probability can be written as

$$\begin{aligned} p(\boldsymbol{a}^L_{\rm ot}) &= \mathrm{Tr}\left[K_0^d\,|0\rangle\langle 0|\,(K_0^\dagger)^d E_1\right] = \mathrm{Tr}\left[|0\rangle\langle 0|\,(K_0^\dagger)^d E_1 K_0^d\right] \\ &= (1-q)\,\mathrm{Tr}\left[|0\rangle\langle 0|\,(K_0^\dagger)^d\,|d-1\rangle\langle d-1|\,K_0^d\right] \\ &= (1-q)\left|\langle d-1|\,K_0^d\,|0\rangle\right|^2, \end{aligned} \tag{2.42}$$

where we used the cyclic property of the trace, and the definition of E_1 as $\mathbb{1} - E_0 = (1-q)\,|d-1\rangle\langle d-1|$. Using the optimal E_0 and E_1, corresponding to $q = 1 - d/L = 1 - d/(d+1)$, we are left with only one parameter θ_0 to optimize. For large d, the optimal probability is highly sensitive on the value of θ_0 (see Fig. 2.14). A good performance, however, is given by the angle $\theta_0 = (2\pi/d)(1-1/d)$, suggesting that it is asymptotically optimal, even if for small d some deviation from the true optimum can be found. To guess this particular value, we analyzed the action of the unitary U_0. We recall that applying d times the unitary U_0 gives

$$U_0^d = \mathrm{e}^{-\mathrm{i}\theta_0 dH} = \mathrm{e}^{-\mathrm{i}\theta_0 dF\left(\sum_{k=0}^{d-1} k|k\rangle\langle k|\right)F^\dagger} = F\left(\sum_{k=0}^{d-1}\mathrm{e}^{-\mathrm{i}\theta_0 dk}\,|k\rangle\langle k|\right)F^\dagger. \tag{2.43}$$

This gives a transition probability between the last and the first state

$$\begin{aligned} \left|\langle d-1|\,U_0^d\,|0\rangle\right|^2 &= \left|\sum_{k=0}^{d-1}[F]_{d-1,k}\,\mathrm{e}^{-\mathrm{i}kd\theta_0}\,[F^\dagger]_{k,0}\right|^2 \\ &= \frac{1}{d^2}\left|\sum_{k=0}^{d-1}\exp\left[\tfrac{2\pi\mathrm{i}}{d}(d-1)k - \mathrm{i}kd\theta_0\right]\right|^2. \end{aligned} \tag{2.44}$$

The maximum occurs when all the amplitudes are in phase, which implies the slowest rotating amplitude ($k=1$) has to reach zero phase, i.e., when $\frac{2\pi}{d}(d-1) - d\theta_0 = 0$. This gives $\theta_0 = (2\pi/d)(1-1/d)$, as desired[16]. The matrix E_0 has been ignored here, which is at the origin of the deviation for small d. However, from the result of numerical calculation, see Fig. 2.13, it seems that our approximation is good enough.

The results of this numerical optimization, shown in Fig. 2.13, suggest that in the quantum case probability 1 can be asymptotically reached. Of course, this fact cannot be proven rigorously by these numerical optimization methods, as any concrete computation, in a given dimension, results in a probability strictly smaller than 1. To approach such questions a more sophisticated construction of quantum models is warranted.

For a fixed d, the probability as a function of θ_0 is characterized by a series of d equally-spaced peaks, which become sharper with increasing d (Fig. 2.14), indicating that the optimal value requires fine-tuning.

[16] In the general (L,d) case, we obtain instead $\theta_0 = \frac{2\pi}{d}\left(1 - \frac{L-d}{L-1}\right)$.

This simplification made the problem tractable, namely, we were able to compute the probability for high dimension, without the need of an optimization over the parameters θ_0 and q. This result, however, slightly deviates from the true optimal θ_0 for small d. For large d, it is asymptotically close to the true optimal. The optimal probability tends towards 1 as the dimension (and the length of the sequence) increases (Fig. 2.13), suggesting that there is no nontrivial upper bound, strictly smaller than 1, in the quantum case.

The above *ansatz* provides an interesting comparison and intuition between the classical and quantum behaviors in these one-way models. In the classical case, each transition produces a statistical mixture of states, in such a way that the average behavior is the model transitioning forward d states after $d+1$ transitions, introducing a delay of one state at the end of the sequence. These transitions always result in a spreading of the probability over many states. In the quantum case, the dynamics is governed by the unitary, which produces a coherent superposition of states. The model is still transitioning forward d states after $d+1$ transitions, but the coherence allows the probability to concentrate at the final state after an initial spread. The two behaviors are shown in Fig. 2.15, where the "concentration" of probability can be observed in the quantum case.

2.8.4 Number of Kraus operators

Note: This section contains additional results not originally included in Ref. [198].

Surprisingly, the numerical results also seem to show that there is no benefit in using more than one Kraus operator per effect $\mathcal{I}_a$. In all scenarios investigated up to $L = 5$ and $n = 3$, the optimization algorithm eventually converges with all n Kraus operators of each effect being scalar multiples of one another, which means they can be combined into a single operator. Therefore, the same probabilities were obtained in all such cases, up to numerical accuracy of 10^{-6}.

Naturally, classical models can be simulated by quantum ones (see Sec. 2.4.2), but classical probabilistic models transform pure states into mixed states, a transformation which requires the use of multiple Kraus operators to be written in a quantum model. The fact quantum models can achieve the same probability, and beyond, with only coherent transformations on a pure state, highlights the power and versatility of quantum memory effects.

A possible explanation of this result is the observation that any quantum advantage would arise from the use of coherences between orthogonal memory states, and that such coherence effects will be reduced if the state becomes mixed. Therefore, additional Kraus operators could only lower the performance of the model, by bringing their behavior closer towards the classical probabilistic behavior.

2.8.5 Nontrivial scenarios with no quantum advantage

Note: This section contains additional material not originally included in Ref. [198].

As can be seen in Fig. 2.11, our results seem to indicate the existence of nontrivial scenarios in which quantum memories offer no advantage over classical ones. Naturally, no quantum advantage is possible for the scenarios with $d = 1$, but several nontrivial scenarios with $2 \leq d \leq 5$ have been found. Table 2.2 includes all such scenarios up to length 7, with many more scenarios found for $L > 7$ and up to $d = 5$.

All scenarios with no quantum advantage required $d < \mathrm{DC}(\boldsymbol{a}) - 1$, which could indicate the existence of a memory threshold for nonclassicality to emerge, i.e., a dimension $d_Q(\boldsymbol{a})$ such that:

$$d < d_Q(\boldsymbol{a}) \quad \Longrightarrow \quad \Omega_Q(\boldsymbol{a}, d) = \Omega_C(\boldsymbol{a}, d). \tag{2.45}$$

We leave the investigation of this possibility to future research.

2.9 Conclusions and outlook

In this chapter we explored the optimal deterministic and probabilistic realizations of symbolic sequences, as rooted on fundamental physical principles. In particular, we consider each symbol as being generated, one at a time, by sequential operations performed on a physical system of finite dimension, acting as the sole memory resource for the task of generating the sequence. The finite-state machine framework we have employed to describe this physical setup places classical and quantum theory on equal terms, enabling a very precise comparison between optimal behaviors allowed within each theory.

We introduced the notion of deterministic complexity (Sec. 2.5.1) for quantifying the minimal dimension required for deterministic realizations of a sequence, where differences between classical and quantum behaviors become irrelevant, and provided an algorithm to efficiently compute it. In the sub-deterministic scenarios, i.e., $d < \mathrm{DC}(\boldsymbol{a})$, we surveyed optimal (classical and quantum) realizations of sequences up to length $L = 10$ via gradient descent techniques, namely, the Adam algorithm [100]. This survey has led to several novel insights on the structure of temporal correlations under finite-memory constraints. An indication of the good performance of the gradient descent method is that it was able to consistently find very sparse optimal solutions.

Of note, the detailed survey led to an improvement upon a previously known model for the one-tick sequence [33], with the discovery of the enhanced multicyclic models, or EMCMs (Sec. 2.7.2). Moreover, our results suggest that, for any given sequence $\boldsymbol{a}$ and any dimension $d < \mathrm{DC}(\boldsymbol{a})$, the optimal EMCM for the sequence $\boldsymbol{a}_{\mathrm{ot}}^{\mathrm{DC}(\boldsymbol{a})}$ provides an upper bound for $p(\boldsymbol{a}|T, d)$ over all its classical realizations. In particular, the investigation of optimal probabilities of EMCMs suggests the existence of a universal upper bound of $1/\mathrm{e}$ for all classical sub-deterministic realizations of any sequence, i.e., $d < \mathrm{DC}(\boldsymbol{a})$. That being said, this result remains a conjecture.

In the quantum case, we have shown how even simple models, with a small amount of memory as resource, are already able to violate this conjectured universal classical bound, and consequently, to outperform any classical scenario in which the realization of a sequence must necessarily be probabilistic. Moreover, an analysis of the performance of such quantum models for the one-tick sequences in high dimension, i.e., up to $d = 500$, suggests that quantum mechanics features no nontrivial universal upper bound.

Notwithstanding two fundamental limitations of our approach, namely, the analysis of only a finite set of sequences and the use of gradient descent methods, which guarantee only a local maximum in our case (and thus lower bounds), we believe our findings provide a strong support for our conjectures. In fact, the total number of scenarios analyzed, counting also their realizations in different dimensions and multiple optimization trials, was on the order of tens of thousands, with consistent results obtained for the optimal solutions. Furthermore, the more accurate optimizations performed for this thesis have corroborated all previous results reported in the original work (Ref. [198]). It seems unlikely that any additional properties of sequences and models would emerge, which could invalidate these bounds. For the quantum bound, similarly to the classical case, it would be very surprising if the limit of the one-tick sequence for $d = L-1$ and $L \to \infty$ would converge to some value strictly smaller than one.

Our results stimulate future research in several directions. First, it would be interesting to understand whether the general upper bound based on the EMCM fails for longer sequences or if it is valid in general. This will also have consequences to the universal upper bound. Without proof of its optimality, it may be possible that for certain longer sequences a nontrivial model may outperform its respective EMCM. The optimality of EMCMs themselves, for one-tick sequences, is also only conjectured at this time, but search for more general models have shown no evidence that EMCMs can be outperformed by uneven cycle lengths or independent probabilities, as described in Sec. 2.10. Furthermore, we believe the results in Sec. 2.7.2 provide a strong argument that EMCMs are indeed optimal, as they seem to map directly

either to the scenario with $d = 1$, where the exact maximum can be found analytically (see Sec. 2.6.3), or to the scenario with $d = L-1$, where one-way models seem to provide the exact maximum in all scenarios that have been verified with outer approximations [207, 200]. Understanding the origin of the empirical results in Eqs. (2.36) and (2.37) might provide insights into these questions.

Moving away from the one-tick sequences, the optimal models for general sequences have proven to be exceptionally complex, showing a rich nontrivial behavior (e.g. Fig. 2.4 and Table 2.1). Despite this, nearly all optimal models found by our numerical search seem to fall into a small number of equivalence classes with same probability, typically given by rational numbers. These same optimal probabilities seem to occur despite the dissimilarity between the scenarios (sequence length and structure, DC, number of states, Hamming weight, tail and cycle lengths, etc.), as shown in Fig. 2.8 and Table 2.1. The origin of these shared structures is as of yet unknown, but their existence suggests that tighter upper bounds may be established by exploiting additional structures of the sequences, beyond their deterministic complexity.

Our results for the quantum case also indicate that, at least when optimizing for a single sequence, we may restrict our attention to instruments with effects defined by single Kraus operators (Sec. 2.8.4), i.e., having $n_a = 1$ for all $a \in \mathcal{A}$. This significantly reduces the search space in the optimization. The optimality of $n_a = 1$ is also conjectured, but has been observed in all scenarios we have tried, up to numerical accuracy. It would be interesting to know whether there are sequential tasks where more than one Kraus operator per effect can be a valuable resource, and under which circumstances it can be exploited as such.

A question also remains of whether there are sequences and nontrivial dimensions for which classical and quantum models provide the same maximum probability. These could lead to Leggett-Garg-type temporal inequalities which are not violated by quantum systems, in analogy to the Guess-Your-Neighbor's-Input non-local game [8]. Our results seem to suggest such scenarios are possible (Sec. 2.8.5), but this requires further investigation.

The results and techniques utilized here can be generalized to arbitrary output alphabets, as well as the inclusion of inputs, which may reveal new information processing applications for finite-state machine methods. Chapter 6 provides an initial foray in this direction, discussing the generalization of the deterministic complexity to the case of inputs and output functions, which establishes some connections between questions in physics foundations and computer science. This will be the subject of future research.

Finally, it is important to remark that the investigation of temporal correlations within the formalism of finite-state machines is a relatively new area of research, with only a few techniques having been tried or developed to address this problem. Despite the theoretical limitations of the gradient descent surveys we have performed, the structure of temporal correlations under finite-memory constraints appears to be sufficiently rich as to offer new insights, even in the simplest scenario we have focused on in this chapter.

2.10 Appendix: Generalized multicyclic models

The multicyclic model introduced in Ref. [33] assumes $d = nk$ and an initial state shift z, such that the n identical cycles of length k must perfectly partition the d states. A natural question arises of whether there could be an advantage if the cycles were not forced to be identical. In the following, we discuss the systematic investigation of this question, and how it eventually led to the structure of the *enhanced multicyclic models* (EMCMs).

In order to fully generalize the multicyclic model, we should allow any possible decomposition of the d states into distinct cycles, each with its own independent cycle probability. We refer to this as a *generalized multicyclic model* (GMCM), where the T_0 transition matrix consists of n cycles with internal deterministic transitions, and a probabilistic transition to the initial state of the cycle, with probability q_i, or to the first

state in the next cycle, with probability $1 - q_i$, for the i-th cycle. Once the last state is reached, we have a nonzero probability, i.e., $1 - q_n$, for the output 1.

An explicit example of T_0 for a $d = 9$ model is Eq. (2.46), and a general GMCM structure in shown in Fig. 2.16. The T_1 matrix is $T_1 = \mathrm{diag}(0, \ldots, 0, 1 - q_n)$. Note that, since the last transition is irrelevant, in T_1 we can choose any position in the last row for the non-zero entry $1 - q_n$.

$$T_0 = \left[\begin{array}{cccc|ccc|cc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ q_1 & 0 & 0 & 0 & 1-q_1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & q_2 & 0 & 0 & 1-q_2 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & q_3 & 0 \end{array}\right] \tag{2.46}$$

Such a GMCM is defined by several parameters: a *signature* $\mathrm{k} = (k_1, \ldots, k_n)$ denoting the size of each of the n cycles, with $k_i \in \mathbb{N}^+$ and $\sum_i k_i = d$, and $\mathrm{q} = (q_1, \ldots, q_n)$ denoting the probability of each cycling transition to occur, with $0 \leq q_i < 1$ (the strict inequality is important as having one of the $q_i = 1$ would give probability zero for the one-tick sequence). Note that there are in total 2^{d-1} potential signatures for a given d, corresponding to the number of integer compositions[17] of the number d. A priori, none of these could have been ruled out, so an exhaustive numerical survey of these models was performed for $L = 3, \ldots, 10$ and $d = 1, \ldots, L - 1$. The optimal models found are shown in Table 2.3, where a few structures become apparent: due to optimal models having the same probabilities for the same block sizes, signatures largely obey permutation symmetry, but favoring bigger blocks at the beginning.

Despite the fact that many signatures of the optimal models found specify non-uniform cycle lengths (e.g. $(1, 2, 2)$), in all such cases the smallest cycles found had $q_i = 0$, that is, they do not behave like cycles at all, but instead acting as a series of deterministic transitions. This renders the models given by $(1, 1, 4)$ and $(2, 4)$ for $L = 10$ identical, for instance. Furthermore, only the largest cycles required $q_i > 0$, and in all cases the optimal models display $q_i = q_j$ if $k_i = k_j$. Finally, in all cases where a signature did not satisfy permutation symmetry, the optimal model contained the deterministic transitions at the end. This suggests that we can always permute the model so that all deterministic transitions occur at the end, as a single deterministic block, while all identical probabilistic cycles occur at the beginning, thus unifying all such optimal models within the same structure.

This inspires us to the following simplified model: we only need to consider models of the form $\mathrm{k} = (k, \ldots, k, t)$, where $d = nk + t$, with a number n of k-sized blocks, i.e., the vast majority of signatures can be dismissed. In such models, we always have $n = \lfloor d/k \rfloor$, and $t = d - nk$, and an initial state $z = 0, \ldots, k - 1$. These are the enhanced multicyclic models (EMCMs). They can be fully specified by the 5-tuple of parameters (L, n, k, t, z). Table 2.4 details the parameters describing various optimal EMCMs.

In addition to the comprehensive low-dimension survey of GMCMs, which led us to the EMCMs, we have also investigated some higher-dimensional cases up to $d = 20$, and $20 < L \leq 50$ using both GMCMs and unrestricted models obtained by gradient descent. In all cases, the best model found was always in the form of an EMCM, which further supports our claim that EMCMs are optimal.

[17] Also known as ordered partitions. For example, $d = 4$ has 8 compositions: $(1, 1, 1, 1)$, $(1, 1, 2)$, $(1, 2, 1)$, $(2, 1, 1)$, $(2, 2)$, $(1, 3)$, $(3, 1)$, and (4). These compositions can be easily enumerated.

DC	d	Sequence	Probability	Model
3	2	001	0.296296	

DC	d	Sequence	Probability	Model
4	3	0001	0.316406	
3	2	0010	0.296296	
3	2	0100	0.250000	
3	2	0011	0.250000	
3	2	0110	0.250000	

DC	d	Sequence	Probability	Model
5	4	00001	0.327680	
4	3	00010	0.316406	
4	3	01011	0.296296	
4	3	00011	0.296296	
4	3	01100	0.296296	
4	3	01110	0.296296	
4	3	00110	0.296296	
3	2	00111	0.250000	
3	2	01000	0.250000	
3	2	01101	0.250000	
3	2	00100	0.190044	
3	2	01001	0.163840	
3	2	00101	0.163840	

DC	d	Sequence	Probability	Model
6	5	000001	0.334898	
5	4	000010	0.327680	
5	4	000011	0.316406	
5	4	011110	0.316406	
5	4	000110	0.316406	
5	4	011100	0.316406	
5	4	001110	0.296296	
5	4	010100	0.296296	
5	4	001011	0.296296	
5	4	010011	0.296296	
4	3	000100	0.316406	
4	3	001000	0.296296	
4	3	000111	0.296296	
4	3	011001	0.296296	
4	3	011101	0.296296	

DC	d	Sequence	Probability	Model
4	3	011000	0.250000	
4	3	010111	0.250000	
4	3	011010	0.250000	
4	3	010110	0.250000	
4	3	001101	0.250000	
4	3	000101	0.250000	
4	3	001100	0.250000	
4	3	010001	0.250000	
3	2	001111	0.250000	
3	2	010000	0.250000	
3	2	001010	0.163840	
3	2	010010	0.163840	
3	2	011011	0.148148	
3	2	001001	0.087791	

Probabilities observed

Closed form	Decimal value
$\frac{32}{729}$	0.0438957476...
$\frac{729}{16\,384}$	0.0444946289...
$\frac{256}{3125}$	0.08192
$\frac{64}{729}$	0.0877914952...
$\frac{27}{256}$	0.10546875
$\frac{4}{27}$	0.148148...

Closed form	Decimal value
$\frac{512}{3125}$	0.16384
$\frac{4\,(411-41\,\sqrt{41})}{3125}$	0.19004404001...
$\frac{1}{4}$	0.25
$\frac{8}{27}$	0.296296...
$\frac{81}{256}$	0.31640625
$\frac{1024}{3125}$	0.32768

Table 2.1: Optimal models found for $L = 3, \ldots, 6$ sequences, grouped by decreasing DC and sorted by decreasing probability. Only $d = \mathrm{DC}(\boldsymbol{a}) - 1$ cases are shown. The pictorial representation of the model matrices (T_0, T_1) has 0 entries as white and 1 as black, with intermediate values in gray. Notice that many optimal probabilities coincide, despite the varied structure of the respective sequences and models. On the bottom right, the distinct probabilities observed are shown along with their closed form representation, valid up to the numerical accuracy.

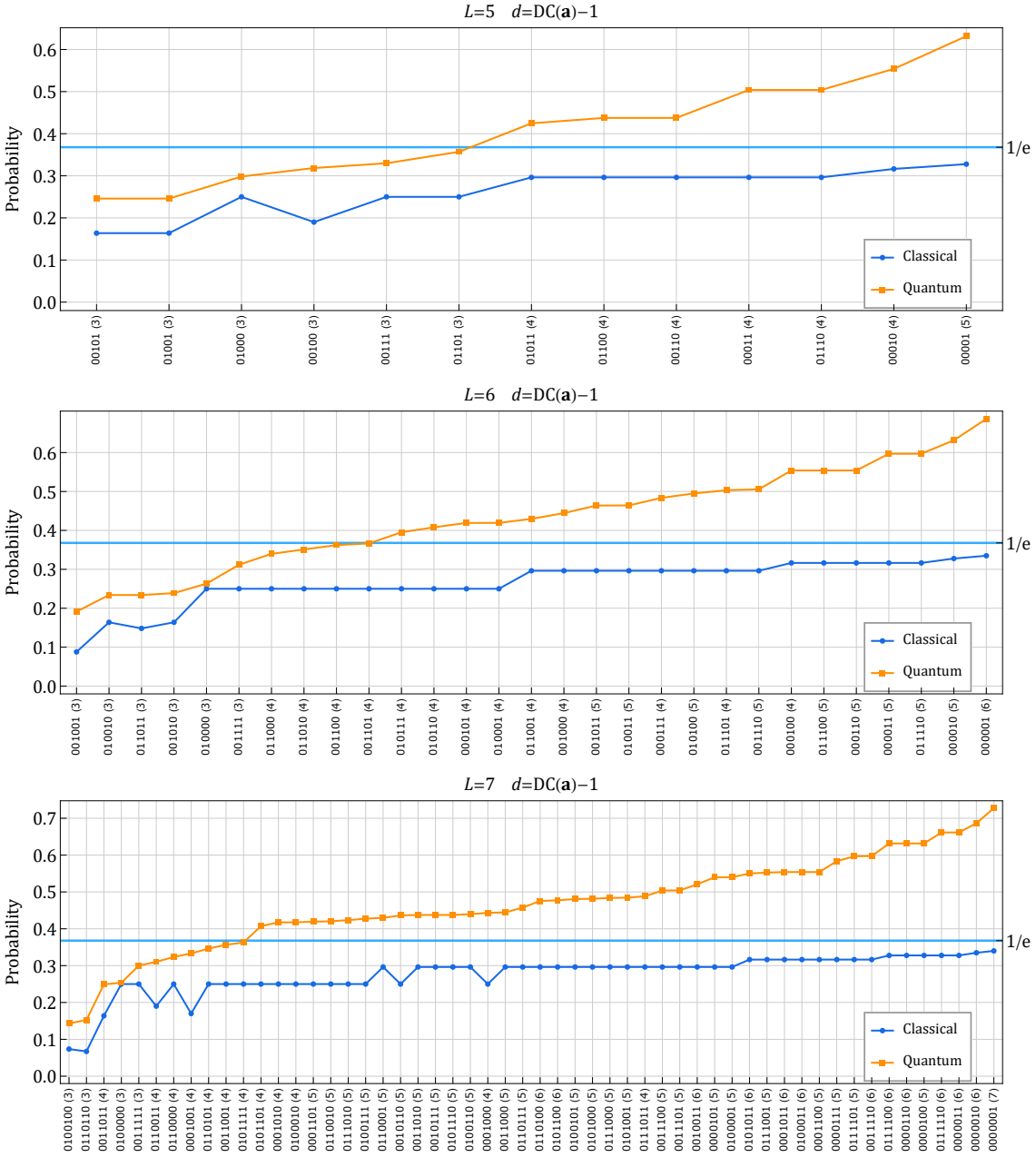

Figure 2.10: Comparison between the optimal probabilities found for classical and quantum models, for all $L = 5, 6$ and 7 and $d = \mathrm{DC}(\boldsymbol{a}) - 1 \geq 2$ scenarios. Sequences are sorted by quantum probability, with each sequence's DC noted inside parentheses. Quantum models can easily violate the conjectured universal classical bound or $1/\mathrm{e}$. Despite the discrete nature of the data shown, points were joined by lines in order to help with the comparison.

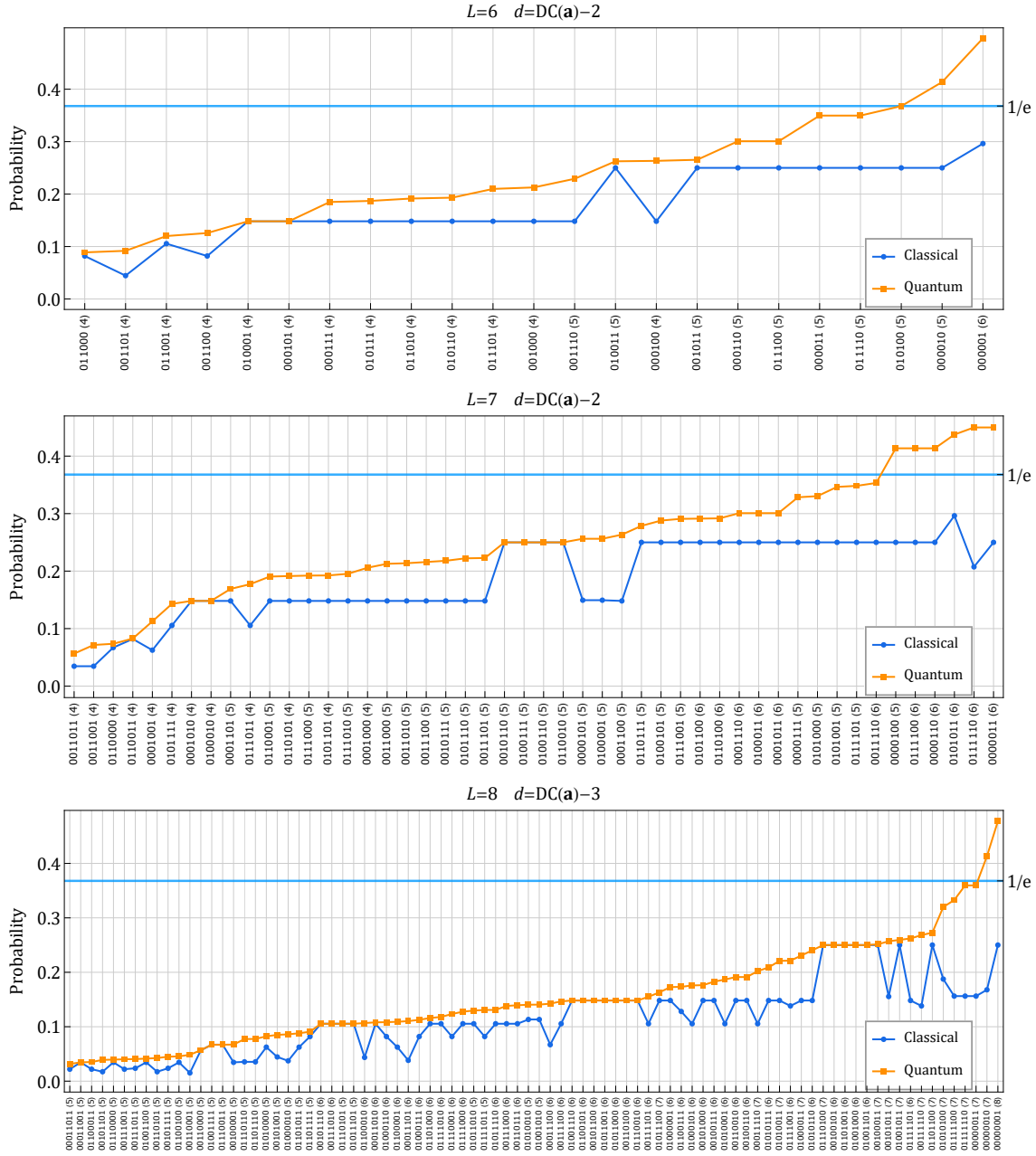

Figure 2.11: If the amount of memory is reduced further, i.e., $d < \mathrm{DC}(\boldsymbol{a}) - 1$, both classical and quantum models perform worse. Quantum models still outperform classical ones in nearly all scenarios, with our results hinting at the existence of scenarios where the quantum advantage disappears, shown here as overlaps between the two lines; see Sec. 2.8.5. As before, only $d \geq 2$ scenarios are shown.

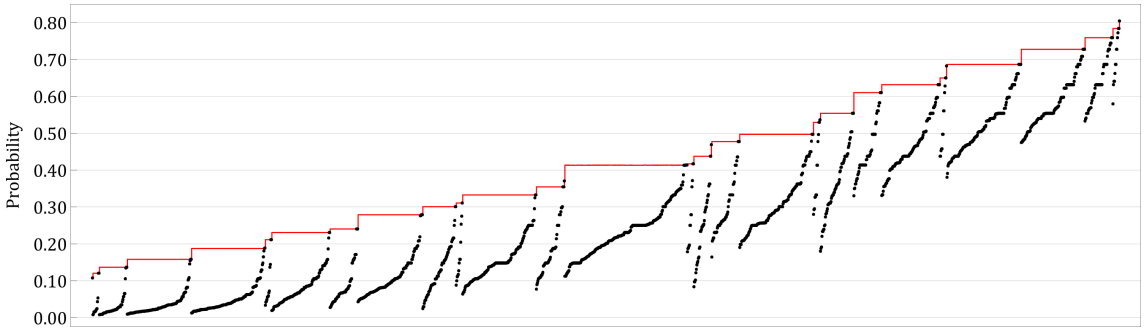

Figure 2.12: Optimal probabilities for all sub-deterministic scenarios in the quantum survey (black dots). The red line is the upper bound provided by the optimal quantum models found for the one-tick sequences $\boldsymbol{a}_{\text{ot}}^{\text{DC}(\boldsymbol{a})}$ with the same d. For clarity, the scenarios were sorted first by increasing conjectured bound, then by increasing probability.

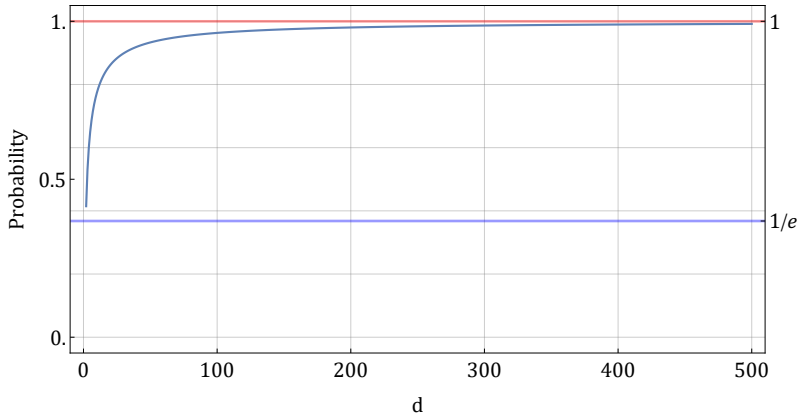

Figure 2.13: Results of the numerical optimization for the probability of the $\boldsymbol{a}_{\text{ot}}^{L}$ sequence for the case $L = d + 1$ or $d = \text{DC}(\boldsymbol{a}_{\text{ot}}^{L}) - 1$, for $d = 2, \ldots, 500$. The probability seems to converge to the trivial upper bound of 1. These models also easily outperform the $1/\text{e}$ bound of the classical scenario.

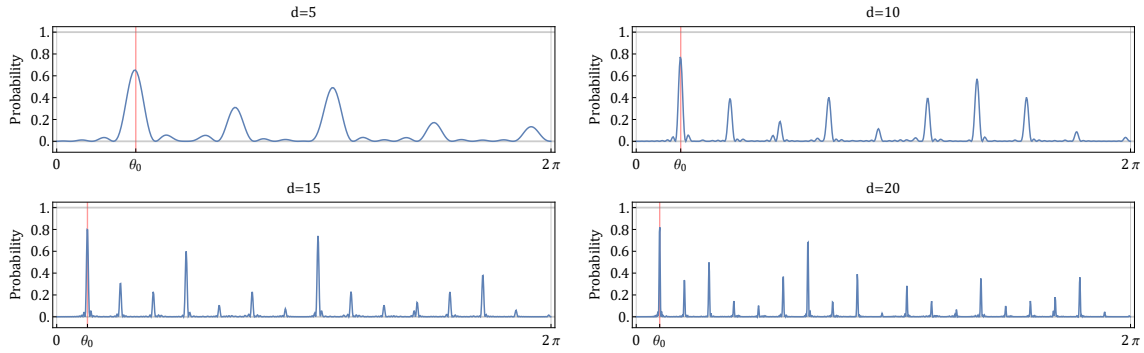

Figure 2.14: The probability $p(\theta_0|d)$ of the quantum one-way model for increasing d. The various peaks become increasingly sharp for higher dimensions. The value $\theta_0 = (2\pi/d)(1 - 1/d)$ is shown, which appears to be asymptotically optimal.

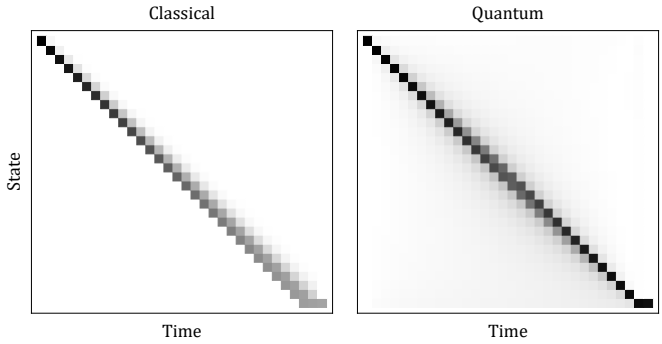

Figure 2.15: The overall behavior of classical and quantum one-way models, showing the evolution of the probability distributions over the d states. In this diagram, each column represents the states at a given time step in the sequence, starting from the left with $a_1 = 0$ and ending at the right with $a_L = 1$. The first state is at the top. In the quantum case, the distribution corresponds to the diagonal entries in the density matrix. White corresponds to zero probability, black to 1.

DC	Sequence	d	Probability
4	01011	2	1/4
4	000101	2	4/27
4	010001	2	4/27
5	011100	2	27/256
4	0001010	2	4/27
4	0100010	2	4/27
5	0010110	3	1/4
5	0100110	3	1/4
5	0100111	2	256/3125
5	0100111	3	1/4

DC	Sequence	d	Probability
4	0110011	2	256/3125
5	0101000	2	4/27
5	0101110	2	27/256
5	0110010	3	1/4
5	0111001	2	256/3125
5	0111010	2	27/256
6	0101011	2	4/27
6	0101100	3	4/27
6	0110100	3	1/4
6	0111100	2	256/3125

Table 2.2: Scenarios with $L \leq 7$ and $d \geq 2$ where no quantum model was found to perform better than the corresponding classical model. The fractions approximate the probabilities up to numerical accuracy. It is unclear whether $\Omega_C = \Omega_Q$ for these cases, as suggested by the results of the survey, or whether this is an artifact of the local optima found by the gradient descent approach.

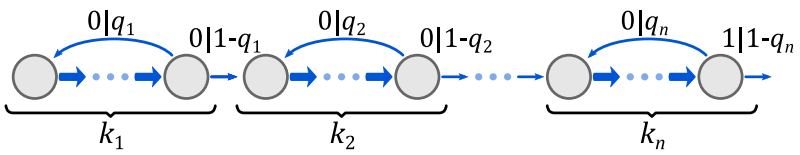

Figure 2.16: The structure of generalized multicyclic models, which consist of splitting the d states into n blocks of various sizes k_i, each block corresponding to an internally-deterministic cycle with independent probability q_i of cycling, and $1 - q_i$ of moving to the next cycle. Only the last state of the last block has a probability $1 - q_n$ of outputting 1. All optimal models for the one-tick sequence investigated are of this form.

$L = 3$		
d	$p(\boldsymbol{a}^L_{\rm ot}\|G,d)$	k
1	0.148148	$(1)_0$
2	0.296296	$(1,1)_0$

$L = 4$		
d	$p(\boldsymbol{a}^L_{\rm ot}\|G,d)$	k
1	0.105469	$(1)_0$
2	0.250000	$(2)_0$
3	0.316406	$(1,1,1)_0$

$L = 5$		
d	$p(\boldsymbol{a}^L_{\rm ot}\|G,d)$	k
1	0.081920	$(1)_0$
2	0.148148	$(2)_1$
3	0.250000	$(1,2)^*_0, (3)_1$
4	0.327680	$(1,1,1,1)_0$

$L = 6$		
d	$p(\boldsymbol{a}^L_{\rm ot}\|G,d)$	k
1	0.066980	$(1)_0$
2	0.148148	$(2)_0$
3	0.250000	$(3)_0$
4	0.296296	$(2,2)_0$
5	0.334898	$(1,1,1,1,1)_0$

L=7		
d	$p(\boldsymbol{a}^L_{\rm ot}\|G,d)$	k
1	0.056653	$(1)_0$
2	0.105469	$(2)_1$
3	0.148148	$(1,2)^*_0,\ (3)_2$
4	0.250000	$(1,3)^*_0,\ (4)_1$
5	0.296296	$(1,2,2)^*_0$
6	0.339917	$(1,1,1,1,1,1)_0$

L=8		
d	$p(\boldsymbol{a}^L_{\rm ot}\|G,d)$	k
1	0.049087	$(1)_0$
2	0.105469	$(2)_0$
3	0.148148	$(3)_1$
4	0.250000	$(4)_0$
5	0.250000	$(1,1,3)^*_0,\ (2,3)^*_0,\ (4,1)_1,\ (5)_2$
6	0.316406	$(2,2,2)_0$
7	0.343609	$(1,1,1,1,1,1,1)_0$

L=9		
d	$p(\boldsymbol{a}^L_{\rm ot}\|G,d)$	k
1	0.043305	$(1)_0$
2	0.081920	$(2)_1$
3	0.148148	$(3)_0$
4	0.148148	$(3,1)_1,\ (4)_3$
5	0.250000	$(1,4)^*_0,\ (5)_1$
6	0.296296	$(3,3)_0$
7	0.316406	$(1,2,2,2)^*_0$
8	0.346439	$(1,1,1,1,1,1,1,1)_0$

L=10		
d	$p(\boldsymbol{a}^L_{\rm ot}\|G,d)$	k
1	0.038742	$(1)_0$
2	0.081920	$(2)_0$
3	0.105469	$(3)_2$
4	0.148148	$(4)_2,\ (1,3)^*_0$
5	0.250000	$(5)_0$
6	0.250000	$(1,1,4)^*_0,\ (2,4)^*_0,\ (5,1)_1,\ (6)_2$
7	0.296296	$(1,3,3)^*_0$
8	0.327680	$(2,2,2,2)_0$
9	0.348678	$(1,1,1,1,1,1,1,1,1)_0$

Table 2.3: All optimal generalized multicyclic models (G) for $3 \le L \le 10$, with their probabilities and signatures k. Subscripts on signatures indicate the optimal initial state z (starting from 0), and an asterisk indicates that all permutations of the signature were found to be equivalent. Several optimal models appear with non-uniform signatures, improving upon results in Ref. [33].

L \ d	1	2	3	4	5	6	7	8	9
2	$(1,1,0,0)$								
3	$(1,1,0,0)$	$(2,1,0,0)$							
4	$(1,1,0,0)$	$(1,2,0,0)$	$(3,1,0,0)$						
5	$(1,1,0,0)$	$(1,2,0,1)$	$(1,3,0,1)$	$(4,1,0,0)$					
6	$(1,1,0,0)$	$(1,2,0,0)$	$(1,3,0,0)$	$(2,2,0,0)$	$(5,1,0,0)$				
7	$(1,1,0,0)$	$(1,2,0,1)$	$(1,3,0,2)$	$(1,4,0,1)$	$(2,2,1,0)$	$(6,1,0,0)$			
8	$(1,1,0,0)$	$(1,2,0,0)$	$(1,3,0,1)$	$(1,4,0,0)$	$(1,5,0,2)$	$(3,2,0,0)$	$(7,1,0,0)$		
9	$(1,1,0,0)$	$(1,2,0,1)$	$(1,3,0,0)$	$(1,4,0,3)$	$(1,5,0,1)$	$(2,3,0,0)$	$(3,2,1,0)$	$(8,1,0,0)$	
10	$(1,1,0,0)$	$(1,2,0,0)$	$(1,3,0,2)$	$(1,4,0,2)$	$(1,5,0,0)$	$(1,6,0,2)$	$(2,3,1,0)$	$(4,2,0,0)$	$(9,1,0,0)$

Table 2.4: Optimal (n,k,t,z) parameters for EMCMs for various L and d. In the case multiple sets of parameters resulted in the same probability, the set with the smallest t was chosen to highlight the cases in which the EMCM structure is strictly required. Including L, the five parameters fully specify the probability for a given (L,d).

Chapter 3

Witnessing environment dimensions through temporal correlations

The main results in this chapter were published in Ref. [201]:

Witnessing environment dimension through temporal correlations
Lucas B. Vieira, Simon Milz, Giuseppe Vitagliano, and Costantino Budroni
Quantum 8, p. 1224 (2024) — Accepted 2023-12-04

Author contribution: In this work, the doctoral candidate significantly contributed to the conception of the research topic, the formulation of the SDP relaxation, and to several of its proofs. The candidate is fully responsible for writing of all the SDP optimization code, the analysis of the results, the generalization of symmetric space representations and operations in Sec. 3.7.4 and, especially, to the creation of the sparsity heuristic (adapted into Ref. [199] and Ch. 5). The candidate also contributed with the writing of nearly all the manuscript, and was solely responsible for the production of all its graphical assets.

The contents of this chapter were adapted from the original text. Some of its results have been relocated to Ch. 2 (Sec. 2.6.3) and Ch. 5 (Sec. 3.7.3) in order to avoid redundancy in this thesis.

3.1 Introduction

Physical systems are never truly isolated, and inevitably interact with their surrounding environment [24, 162]. As a consequence, information within the system leaks away into its surroundings, leading to entanglement between the system and the environment. In many instances, this leaked information may be partially recovered at a later time, leading to non-Markovian dynamics, i.e., non-negligible memory effects and complex correlations in time [163, 25, 132].

Like their spatial counterpart, temporal correlations in quantum mechanics fundamentally differ from those that can be observed in the classical case. Such differences between classical and quantum temporal correlations have been noted since the works of Leggett and Garg [116, 115, 64, 203]. In the presence of memory, a clear distinction between an underlying process – carrying temporal correlations – and the measurement process – probing said correlations – can be obtained by employing correlation kernels [118, 3] or higher order quantum maps [155] for their description.

Broadly speaking, the dimension of a physical system, i.e., the number of perfectly distinguishable states, imposes fundamental constraints over the temporal correlations it is able to produce [32]. In this sense, the physical dimension acts as a memory resource, restricting the amount of information stored about the past that is capable of affecting the future. This memory constraint leads to different behaviors

(i.e., different achievable temporal correlations) in classical, quantum, or more general physical theories [67, 88, 32, 182, 198, 124]. In particular quantum memories are known to allow for a larger set of correlations than classical ones of the same size [32, 33, 198]. However, this advantage only holds for restricted memory size, i.e., if the memory size is *unrestricted*, all correlations compatible with a time-ordered causal structure may be achieved with either classical or quantum memories [67, 88]. Therefore, understanding the nature of dimensional constraints on observable correlations sheds light on fundamental differences between our descriptions of physical systems, as well as their connection with causality.

These limitations imposed by the dimensionality of physical systems have been exploited for the construction of "dimension witnesses", inequalities which, when violated, certify the minimum dimension of the system compatible with observations [69, 28, 81, 31, 170, 182, 179]. In a similar spirit, the approach described in this chapter allows for the computation of upper bounds on the temporal correlations achievable in an open system dynamics with an environment of bounded dimension, so that any violations of these bounds certify the minimum dimension of the effective environment. Employing techniques from entanglement detection [59, 140], these bounds are obtained by exploiting the inherent symmetries of the problem. In particular, we relax the *a priori* non-linear problem of computing maximum joint probabilities in sequential measurements to a numerically tractable hierarchy of semidefinite programs (SDPs) [22]. Both the computational accessibility of the final formulation of the problem, as well as the non-triviality of the resulting bounds, are then demonstrated for paradigmatic examples, showing that, indeed, joint probability distributions obtained from probing an open system alone provide viable means to deduce dimensional properties of the a priori experimentally inaccessible environment it is coupled to.

This chapter is organized as follows. Section 3.2 provides an outline of the main results presented in this chapter, and the original work it was adapted from (Ref. [201]). In Sec. 3.3, we introduce the sequential measurement protocol in the case of open quantum systems, the notion of temporal correlations, and their mathematical description. Section 3.4 discusses the application of temporal correlations to the task of characterizing open systems and their environment. Section 3.5 describes the semidefinite program we constructed to bound temporal correlations, with Sec. 3.6 presenting the numerical results we obtained in their optimization. In Sec. 3.7 we comment on the various challenges involved in realizing these numerical optimizations, and how we have overcome them. In Sec. 3.8 we comment on several additional technical details involving our work, with Sec. 3.9 covering our conclusions and future outlook.

3.2 Summary of main results

In this work, we investigate applications of bounds on temporal correlations to the problem of witnessing the dimension of an environment interacting with an open quantum system. Since the environment acts as the memory resource in the study of non-Markovian dynamics, the results in Ref. [198] (Ch. 2) suggested that upper bounds on temporal correlations could offer new insights into the study of open quantum systems. In contrast with this previous work, which resorted to computing lower bounds on classical and quantum temporal correlations, here we focus on the problem of obtaining upper bounds for the open quantum system case, which is much more general.

To obtain these upper bounds, we formulate the problem of sequential measurements and repeated system-environment unitary interactions in the formalism of quantum supermaps [41, 156, 155], using the Choi-Jamiołkowski isomorphism [46, 94]. In this way, the repeated system-environment unitary interactions are mapped to symmetric and separable pure states in a larger Hilbert space, where techniques from entanglement detection [58, 59, 140], bilinear optimization [16], and rank-constrained optimization [216] can be applied.

The inherent symmetry of the problem allows for an efficient representation directly in the symmetric subspace [85], which required a generalization of earlier results known for qubits [184] to systems

of arbitrary dimension (Sec. 3.7.4). Furthermore, this symmetry enables a direct application of the quantum de Finetti theorem [39, 47] for establishing a hierarchy of SDP relaxations which converge exactly to the original problem [59]. Nevertheless, this is a very computationally demanding technique, which renders the problem numerically intractable even in the simplest nontrivial scenarios, despite being already represented directly in the symmetric subspace.

To overcome this challenge, for this work we have developed a novel heuristic algorithm for obtaining sparse representations of large-scale SDPs, by exploiting their inherent *effective sparsity*, which emerges naturally from its objective function and constraints. This heuristic was particularly well suited for our problem, allowing for a reduction to less than 1% of the original number of variables and constraints in its symmetric representation (Sec. 3.7.3). This sparsity heuristic is very general, and was adapted into its own paper in Ref. [199], which we discuss in detail in Ch. 5.

Thanks to this novel sparsity heuristic, an exact sparse representation of the SDP relaxations could be obtained allowing us to compute upper bounds on the maximum probability of a few sequences for the case of a qubit environment. The methods we have developed can be used to obtain similar upper bounds in a variety of other open quantum system scenarios, each acting as a witness for the environment dimension.

3.3 The sequential measurement protocol

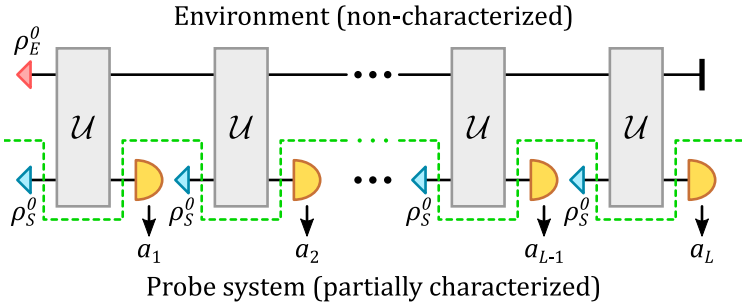

Figure 3.1: Diagram of the sequential measurement protocol, involving a partially characterized probe system and a non-characterized environment, here separated by the dashed line. The sequence of measurement outcomes $\boldsymbol{a} = a_1 \ldots a_L$ is obtained by repeated preparations of a probe state ρ_{S}^0, fixed and the same at every time step, which is left to interact with the environment, then followed by measurements (semicircles). The environment acts as a *memory* resource, capable of establishing long-term correlations between measurements.

We assume that an experimenter is able to prepare a "probe" system in a known state and perform measurements in a certain basis. Its dynamics, on the other hand, are not under experimental control, and are governed by its inevitable interaction with the environment. Typically, this environment is a much larger system, generally inaccessible, and featuring complex dynamics. Together, system and environment undergo closed, i.e., unitary evolution. Their interaction leads to an imprinting of information on the probe system, which can be used to learn something about the environment by means of the probe alone. In fact, this is a common way of making indirect measurement on (typically large many-body) systems that are not fully controlled [23, 37]. For example, a small probe can be used for estimating an unknown parameter of a larger (many-body) environment [56], in particular temperature [6, 168, 54, 128].

We now introduce the details of our scenario and the notation used. By $\mathcal{H}_{\mathrm{S}}$ and $\mathcal{H}_{\mathrm{E}}$ we denote the finite-dimensional Hilbert spaces of system and environment, with $d_{\mathrm{S}} = \dim \mathcal{H}_{\mathrm{S}}$ and $d_{\mathrm{E}} = \dim \mathcal{H}_{\mathrm{E}}$, and their joint space as $\mathcal{H}_{\mathrm{ES}} = \mathcal{H}_{\mathrm{E}} \otimes \mathcal{H}_{\mathrm{S}}$, with $d_{\mathrm{ES}} = d_{\mathrm{E}} d_{\mathrm{S}}$. The experiment involves L identical measurements on the probe system, at discrete time steps, each outcome $a_\ell \in \mathcal{A}$ collected into a sequence

$\boldsymbol{a} = a_1 \dots a_L$. For simplicity, we consider $\mathcal{A} = \{0, 1\}$ in the following; the generalization to arbitrary outcomes is straightforward. Measurements on the system are described by a Positive Operator-Valued Measure (POVM) $\{E_S^a\}_a$, i.e., $E_S^a \geq 0$ and $\sum_{a\in\mathcal{A}} E_S^a = \mathbb{1}_S$.

The system is initially prepared in the state ρ_{S}^0, left to interact with the environment via the global unitary $\mathcal{U}(\cdot) = U \cdot U^\dagger$, then measured and reprepared in the state ρ_{S}^0. This procedure is repeated a total of L times; see Fig. 3.1. The unitary operation is assumed to be the same at each iteration (corresponding to, for example the situation of a time-independent system-environment Hamiltonian and temporally equidistant measurements); see Sec. 3.8.1 for more details.

In the following, we consider a measure-and-prepare operation of the form

$$\mathcal{M}_a(\rho_{ES}) = \mathrm{Tr}_S\left[\rho_{ES} \cdot (\mathbb{1}_E \otimes E_S^a)\right] \otimes \rho_{\mathrm{S}}^0, \tag{3.1}$$

with the generalization to arbitrary operations being straightforward. The probability of a sequence of outcomes can then be written as

$$\begin{aligned} p(\boldsymbol{a}|d_{\mathrm{E}}) &= p(a_1, \dots, a_L|d_{\mathrm{E}}) \\ &= \mathrm{Tr}\left[\mathcal{M}_{a_L} \circ \mathcal{U} \circ \cdots \circ \mathcal{M}_{a_1} \circ \mathcal{U}(\rho_{\mathrm{E}}^0 \otimes \rho_{\mathrm{S}}^0)\right], \end{aligned} \tag{3.2}$$

where $p(\boldsymbol{a}|d_{\mathrm{E}})$ denotes that it is obtained with an environment of dimension d_{E}. Our task is to establish upper bounds on such probabilities given a finite amount of "memory", i.e., the dimension d_{E}. The assumption of identical measurements is essential for capturing this notion of memory, as arbitrary time-dependent operations would impose no nontrivial restriction on the probabilities.

For simplicity, we consider the maximization of the probability of a given sequence $\boldsymbol{a}$, namely, to find[1]

$$\Omega(\boldsymbol{a}, d_{\mathrm{E}}) := \sup p(\boldsymbol{a}|d_{\mathrm{E}}) \tag{3.3}$$

such that $p(\boldsymbol{a}|d_{\mathrm{E}}) \leq \Omega(\boldsymbol{a}, d_{\mathrm{E}})$, for all possible correlations generated by an environment of dimension d_{E}. The generalization to arbitrary linear functions of the distributions $(p(\boldsymbol{a}|d_{\mathrm{E}}))_{\boldsymbol{a}}$ is straightforward. Since obtaining the exact maximum $\Omega(\boldsymbol{a}, d_{\mathrm{E}})$ is very difficult, in the following we resort to obtaining upper bounds $\Omega(\boldsymbol{a}, d_{\mathrm{E}}) \leq \widehat{\omega}(\boldsymbol{a}, d_{\mathrm{E}})$ via a relaxation of the problem.

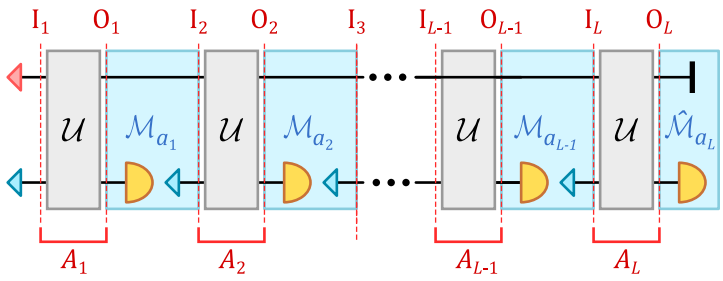

Figure 3.2: The protocol with each time step having its own set of input and output Hilbert spaces, denoted by the vertical dotted lines. Note that the input and output spaces of the $\mathcal{U}$ and $\mathcal{M}_a$ are interleaved: $\mathcal{U}$ has inputs I_ℓ and outputs O_ℓ, but for $\mathcal{M}_{a_\ell}$, inputs are O_ℓ and outputs $\mathrm{I}_{\ell+1}$.

We formulate the problem via the Choi-Jamiołkowski (CJ) isomorphism [94, 46], which involves taking multiple copies of the original Hilbert space $\mathcal{H}_{\mathrm{ES}}$ to describe its time evolution; see Fig. 3.2. The ℓ-th unitary acts as $\mathcal{U} : \mathrm{I}_\ell \to \mathrm{O}_\ell$, while the ℓ-th measurement, placed between unitaries, acts instead as $\mathcal{M}_a : \mathrm{O}_\ell \to \mathrm{I}_{\ell+1}$. All spaces are isomorphic, i.e., $\mathrm{O} \cong \mathrm{I} \cong \mathcal{H}_{\mathrm{ES}}$. For convenience, the ℓ-th unitary's input and output spaces are jointly referred to as $\mathrm{A}_\ell = \mathrm{I}_\ell \otimes \mathrm{O}_\ell$. It is important to emphasize that the local input

[1] The notation for these bounds is the same as used in Ch. 2, but since in this chapter we only consider the quantum case we omit the Q subscript.

and output spaces of the $\mathcal{U}$ and $\mathcal{M}_a$ are interleaved, in the sense that the outputs of one are the inputs of the other (Fig. 3.2). For clarity, we shall use the cursive $\mathcal{I}$ and $\mathcal{O}$ to generically refer to each map's local input and output spaces: for unitaries, $\mathcal{I} = \mathrm{I}_\ell$ and $\mathcal{O} = \mathrm{O}_\ell$, and for measurements, $\mathcal{I} = \mathrm{O}_\ell$ and $\mathcal{O} = \mathrm{I}_{\ell+1}$.

Given a linear map $\Lambda : \mathcal{I} \to \mathcal{O}$ between input and output Hilbert spaces $\mathcal{I}$ and $\mathcal{O}$, with $d_\mathcal{I} = \dim \mathcal{I}$ and $d_\mathcal{O} = \dim \mathcal{O}$, its Choi-Jamiołkowski representation [94, 46] is given by the matrix

$$\mathrm{CJ}[\Lambda] := \sum_{i,j=1}^{d_\mathcal{I}} |i\rangle\langle j| \otimes \Lambda(|i\rangle\langle j|) = \mathrm{id}_\mathcal{I} \otimes \Lambda(|\Psi\rangle\langle\Psi|), \tag{3.4}$$

where $|\Psi\rangle = \sum_{i=0}^{d_\mathcal{I}-1} |ii\rangle$ is a non-normalized maximally entangled state. The matrix $\mathrm{CJ}[\Lambda]$ is typically referred to as a Choi matrix for the map Λ and, if normalized, also as its Choi state. Note that $\mathrm{CJ}[\Lambda]$ is of size $d_\mathcal{I} d_\mathcal{O} \times d_\mathcal{I} d_\mathcal{O}$, and since in our case input and output spaces are isomorphic we have $d_\mathcal{I} = d_\mathcal{O} = d_\mathrm{ES}$. We note that the decomposition of sequential maps in Eq. (3.2) is equivalent to the standard formulation of temporal scenarios in the quantum comb [43], process matrix [147, 45], and process tensor [155] formalism.

3.4 Witnesses for open system dynamics

Given a probe system S, the first question we may ask is whether or not it is open, i.e., if it is interacting with an environment at all, or equivalently, whether $d_\mathrm{E} > 1$. To address this question, we can compute $\Omega(\boldsymbol{a}, 1)$ and consider the inequality

$$p(\boldsymbol{a}|d_\mathrm{E} = 1) \leq \Omega(\boldsymbol{a}, 1). \tag{3.5}$$

As an example, the maximum for the sequence $\boldsymbol{a} = 00101$ is $\Omega(\boldsymbol{a}, 1) = (3/5)^3(2/5)^2 = 0.03456$, which we explain how to obtain shortly. If we perform an experiment and observe $p(\boldsymbol{a} = 00101) = 0.5$, then we must conclude the system is open. In other words, the inequality Eq. (3.5) acts as a witness for open systems. However, if no violation is observed, the experiment is inconclusive.

To compute this maximum, we first note that, for our choice of $\mathcal{M}_a$, since both ρ_S^0 and E_S^a are the same at each measurement, all outcomes are independent and identically distributed. Writing $q_a = \mathrm{Tr}\left[\rho_\mathrm{S}^0 E_S^a\right]$ as the probability of outcome a, with $\sum_a q_a = 1$, and using n_a as the number of occurrences of a symbol a in $\boldsymbol{a}$, we have

$$p(\boldsymbol{a}|d_\mathrm{E} = 1) = \prod_{\ell=1}^{L} q_{a_\ell} = \prod_{a\in\mathcal{A}} q_a^{n_a}, \qquad \sum_{a\in\mathcal{A}} n_a = L. \tag{3.6}$$

The global maximum $\Omega(\boldsymbol{a}, 1)$ can be found analytically as was done in Sec. 2.6.3, and is given by:

$$p(\boldsymbol{a}|d_\mathrm{E} = 1) \leq \Omega(\boldsymbol{a}, 1) = \prod_{a\in\mathcal{A}} \left(\frac{n_a}{L}\right)^{n_a}, \tag{3.7}$$

with the convention $q_a^{n_a} = 1$ for $q_a = n_a = 0$. The maximum is independent of d_S and is obtained with $E_S^a = \frac{n_a}{L}\mathbb{1}_S$ and any ρ_S^0. Note that this direct calculation of $\Omega(\boldsymbol{a}, 1)$ crucially depends on the specific form of the instrument we employ. If the respective outcomes do *not* correspond to a $\mathcal{M}_a$ of measure-and-prepare form, the joint probabilities $p(\boldsymbol{a}|d_\mathrm{E})$ do not factorize as in Eq. (3.6), and an alternative approach must be used.

A similar principle to Eq. (3.5) applies when developing witnesses for the case of $d_\mathrm{E} > 1$, although the calculation of the maximum becomes nontrivial, as the probability no longer factorizes as in Eq. (3.6). Concretely, we want to obtain the tightest bounds of the form

$$p(\boldsymbol{a}|d_\mathrm{E}) \leq \Omega(\boldsymbol{a}, d_\mathrm{E}), \tag{3.8}$$

which hold for all possible realizations of the sequence $\boldsymbol{a}$ in an experiment. If a violation of Eq. (3.8) is observed, we can certify that the dimension is greater than d_{E}.

In the simple case of $d_{\mathrm{E}} = 1$, we were able to optimize over ρ_{S}^0 and $\{E_S^a\}_a$. In the general case, however, optimizing over all possible protocols (i.e., all possible ρ_{S}^0, $\{E_S^a\}_a$ and $\mathcal{U}$) is difficult. Nevertheless, some general considerations can be made about how the bounds depend on these parameters; see Sec. 3.8.3. We shall assume fixed and well characterized preparations and measurements, in agreement with our initial assumptions on the probe system. We then wish to obtain the maximum

$$\Omega(\boldsymbol{a}, d_{\mathrm{E}}) := \sup_{\mathcal{U}} p(\boldsymbol{a}|d_{\mathrm{E}}), \tag{3.9}$$

where assumptions about the initial state of the environment can be removed by convexity and symmetry arguments; see Sec. 3.8.2. In the following section, we explain how to compute upper bounds $\widehat{\omega}(\boldsymbol{a}, d_{\mathrm{E}})$ on these nontrivial maxima, i.e., $\Omega(\boldsymbol{a}, d_{\mathrm{E}}) \leq \widehat{\omega}(\boldsymbol{a}, d_{\mathrm{E}})$, via convex optimization methods.

3.5 A hierarchy of semidefinite programs bounding temporal correlations

The maxima in Eq. (3.9) require that the probe state ρ_{S}^0 and measurements $\{E_S^a\}_a$ are characterized and fixed throughout the experiment, as discussed in the previous section. To obtain the upper bounds $\widehat{\omega}(\boldsymbol{a}, d_{\mathrm{E}})$, we formulate the problem as a hierarchy of semidefinite programs (SDPs) [22]. This class of optimization problems are ubiquitous in quantum information theory, offering strict convergence of solutions for many classes of problems, as well as having efficient algorithms widely available for solving them [22, 211]; see Sec. 1.4.2 for a brief introduction. For a recent compendium on SDPs in quantum information science, see [177, 192]. In the following, we explain the general steps taken in formulating our problem as an SDP, which can be solved numerically and, importantly, efficiently. A more detailed step-by-step formulation is included in Sec. 3.5.1.

Under the CJ representation, the repeated applications of the unitary and measurements can each be written as single operators existing in a larger space, encompassing multiple time steps, with their input and output spaces interleaved. This is illustrated in Fig. 3.2. Via Eq. (3.4), we define

$$\begin{aligned} C_U &:= \tfrac{1}{d_{ES}} \mathrm{CJ}[\mathcal{U}], \\ M_a &:= \mathrm{CJ}[\mathcal{M}_a], \quad \text{and} \quad \widehat{M}_a := \mathrm{Tr}_{\mathcal{O}}[M_a]\,. \end{aligned} \tag{3.10}$$

With this, we can write the L repeated applications of $\mathcal{U}$ as $C_U^{\otimes L}$, and the measure-and-prepare operations as a single operator defining our objective function,

$$F := d_{\mathrm{ES}}^L (\rho_{\mathrm{E}}^0 \otimes \rho_{\mathrm{S}}^0) \otimes M_{a_1} \otimes M_{a_2} \otimes \cdots \otimes \widehat{M}_{a_L}, \tag{3.11}$$

where $\widehat{M}_a$ is simply the final measurement without a re-preparation, obtained by partially tracing M_a over its output space. As the goal is to obtain a maximum, the initial environment state ρ_{E}^0 can be chosen to be pure by convexity arguments, fixed to be $|0\rangle\langle 0|_E$ by unitary invariance, and considered as part of the experimental setup in F without loss of generality; see Sec. 3.8.2. We can then express the probability of $\boldsymbol{a}$ as

$$p(\boldsymbol{a}|d_{\mathrm{E}}) = \mathrm{Tr}\left[F^{\mathrm{T}} C_U^{\otimes L}\right]. \tag{3.12}$$

Note that by Eq. (3.4) and Eq. (3.10) C_U is positive, rank-1, and normalized, which makes $C_U^{\otimes L}$ a pure symmetric separable quantum state. In addition, each C_U satisfies $\mathrm{Tr}_{\mathcal{O}}[C_U] = \mathbb{1}_{\mathcal{I}}/d_{\mathrm{ES}}$, which arises from

the trace-preserving property of each individual unitary. With all of these observations, we reformulate Eq. (3.9) as the equivalent optimization problem:

$$\begin{aligned}
\textbf{Given:}\quad & d_{\mathrm{E}},\ d_{\mathrm{S}},\ \rho_{\mathrm{E}}^{0},\ \rho_{\mathrm{S}}^{0},\ \{M_a\}_a,\ \boldsymbol{a}, \\
\text{with}\quad & F := d_{\mathrm{ES}}^{L}(\rho_{\mathrm{E}}^{0}\otimes\rho_{\mathrm{S}}^{0})\otimes M_{a_1}\otimes\cdots\otimes M_{a_{L-1}}\otimes\widehat{M}_{a_L}, \\
\textbf{Find:}\quad & \Omega(\boldsymbol{a}, d_{\mathrm{E}}) := \max_{C_U} \operatorname{Tr}\left[F^{\mathrm{T}} C_U^{\otimes L}\right] \\
\textbf{Subject to:}\quad & C_U \geq 0, \quad \operatorname{rank} C_U = 1, \\
& \operatorname{Tr}_{\mathcal{O}}[C_U] = \mathbb{1}_{\mathcal{I}}/d_{\mathrm{ES}}.
\end{aligned} \tag{3.13}$$

As C_U enters in the objective function as a tensor power, and we further require it to be rank-1, the problem is both non-linear and non-convex with respect to C_U, and thus the above problem cannot be directly solved as an SDP. To reach an SDP *relaxation* of Eq. (3.13), we first transform it into a chain of equivalent problems.

The problem can be made convex by replacing $C_U^{\otimes L}$ with a separable state on the symmetric subspace

$$\widetilde{X}_{\mathrm{A}_1^L} := \sum_i p_i \left|\varphi_i\right\rangle\!\left\langle\varphi_i\right|^{\otimes L}, \tag{3.14}$$

with $p_i \geq 0$, $\sum_i p_i = 1$. By convexity, the maximum will be achieved for a rank-1 $\widetilde{X}_{\mathrm{A}_1^L}$, therefore this relaxation leaves the optimal value unchanged. Thus, we replace $C_U^{\otimes L}$ in Eq. (3.13) with $\widetilde{X}_{\mathrm{A}_1^L}$ satisfying

$$\widetilde{X}_{\mathrm{A}_1^L} \in \mathrm{SEP}_L, \qquad \mathrm{P}_L^{+}\widetilde{X}_{\mathrm{A}_1^L} = \widetilde{X}_{\mathrm{A}_1^L}, \tag{3.15}$$

and the partial trace constraint of Eq. (3.13), where SEP_L is the set of fully separable L-partite states, and P_L^{+} is the projector onto the symmetric subspace of L systems (see the definition in Eq. (3.28) in Sec. 3.7.4). In fact, the constraints in Eq. (3.15) are exact, as it can be shown that all separable states in the symmetric subspace are of the form in Eq. (3.14); see Sec. 3.5.1 for a proof.

So far, we transformed the original problem into an equivalent one, but due to the condition $\widetilde{X}_{\mathrm{A}_1^L} \in \mathrm{SEP}_L$ it is not yet an SDP. This constraint can be addressed by relaxing the original problem via the quantum de Finetti theorem [39] (see also [47]), which tells us that $\widetilde{X}_{\mathrm{A}_1^L}$ can be approximated as the reduced state of a larger symmetric (and potentially entangled) state $X_{\mathrm{A}_1^N}$, such that $\widetilde{X}_{\mathrm{A}_1^L} \approx \operatorname{Tr}_{\mathrm{A}_{L+1}^N}[X_{\mathrm{A}_1^N}]$, where $\operatorname{Tr}_{\mathrm{A}_{L+1}^N}$ is the trace over systems $L+1,\ldots,N$.

While this relaxation *a priori* only provides an upper bound for the original problem, it establishes a hierarchy of approximate solutions, which are known to converge exactly to the separable set as N increases [59]. The de Finetti theorem, and analogous results for permutationally invariant (or exchangeable), rather than symmetric, operators have found broad applications in quantum information theory, from entanglement detection [58, 59, 140] to more general optimization problems, such as evaluating convex roofs of entanglement measures [195], constrained bilinear optimization [16], dimension-bounded quantum games [95], as well as rank-constrained optimization [216]. All the above mentioned results exploit the basic idea of using either the symmetric subspace or permutation invariance to relax nonlinear constraints. In particular, Ref. [16] analyzed the optimization over pairs of quantum channels, basing their construction on permutation invariant operators, and showing the partial trace constraint necessary to translate the usual procedure (e.g., in the entanglement detection approach) from states to channels. Moreover, [216] introduced the idea of a rank-constrained optimization based on the symmetric subspace, which is central to impose the rank constraint in Eq. (3.13). Our construction is inspired by these works and could be derived starting from some of these results. However, it is more straightforward to provide a direct construction in terms of the de Finetti theorem; see Sec. 3.5.1.

Ultimately, we arrive at an SDP relaxation of the original problem:

$$\begin{aligned}
\textbf{Given:}\quad & d_{\mathrm{E}},\ d_{\mathrm{S}},\ \rho_{\mathrm{E}}^{0},\ \rho_{\mathrm{S}}^{0},\ \{M_a\}_a,\ \boldsymbol{a},\ N,\\
\text{with}\quad & F := d_{\mathrm{ES}}^{L}(\rho_{\mathrm{E}}^{0}\otimes\rho_{\mathrm{S}}^{0})\otimes M_{a_1}\otimes\cdots\otimes M_{a_{L-1}}\otimes\widehat{M}_{a_L},\\
\text{and}\quad & \widehat{M}_{a_L} := \mathrm{Tr}_{\mathcal{O}}\big[M_{a_L}\big]\\
\textbf{Find:}\quad & \widehat{\omega}^{N}(\boldsymbol{a},d_{\mathrm{E}}) := \max_{X_{\mathrm{A}_1^N}} \mathrm{Tr}\left[(F^{\mathrm{T}}\otimes\mathbb{1}_{\mathrm{A}_{L+1}^N})X_{\mathrm{A}_1^N}\right]\\
\textbf{Subject to:}\quad & X_{\mathrm{A}_1^N}\geq 0,\\
& \mathrm{P}_N^{+}X_{\mathrm{A}_1^N} = X_{\mathrm{A}_1^N},\ \ \mathrm{Tr}\left[X_{\mathrm{A}_1^N}\right]=1,\\
& \mathrm{Tr}_{\mathrm{O}_1}\left[X_{\mathrm{A}_1^N}\right] = \frac{\mathbb{1}_{\mathrm{I}_1}}{d_{\mathrm{ES}}}\otimes\mathrm{Tr}_{\mathrm{A}_1}\left[X_{\mathrm{A}_1^N}\right].
\end{aligned} \tag{3.16}$$

We now include the superscript N in $\widehat{\omega}^{N}(\boldsymbol{a},d_{\mathrm{E}})$ as to emphasize it is part of a hierarchy of upper bounds. Here, $N\geq L$ is the size of the symmetric state $X_{\mathrm{A}_1^N}$ used in the approximation of the separable L-partite state $X_{\mathrm{A}_1^L}$. The last constraint in Eq. (3.16) is required to ensure that $X_{\mathrm{A}_1^N}$ represents a sequence of trace-preserving maps, and it can be enforced on a single step due to the symmetry constraint on $X_{\mathrm{A}_1^N}$; see Sec. 3.5.1. Optionally, one can also add further linear constraints to the SDP to improve its approximation to the separable set, e.g., entanglement witnesses [193, 79] or the positive partial transpose (PPT) criterion [153] $X_{\mathrm{A}_1^N}^{\mathrm{T}_\alpha}\geq 0,\forall\alpha$, where $X_{\mathrm{A}_1^N}^{\mathrm{T}_\alpha}$ is the partial transpose with respect to a bipartition α. These conditions are not necessary for convergence, but they may provide better results [140] at an extra computational cost.

Once approximate solutions $\widehat{\omega}^{N}(\boldsymbol{a},d_{\mathrm{E}})$ are obtained from Eq. (3.16), they can be used to establish a convergent hierarchy of upper bounds on the temporal correlations for each sequence, i.e.,

$$\begin{aligned}
&\Omega(\boldsymbol{a},d_{\mathrm{E}})\leq\cdots\leq\widehat{\omega}^{N+1}(\boldsymbol{a},d_{\mathrm{E}})\leq\widehat{\omega}^{N}(\boldsymbol{a},d_{\mathrm{E}})\leq\cdots\leq\widehat{\omega}^{L}(\boldsymbol{a},d_{\mathrm{E}}),\\
&\text{with}\quad \lim_{N\to\infty}\widehat{\omega}^{N}(\boldsymbol{a},d_{\mathrm{E}})=\Omega(\boldsymbol{a},d_{\mathrm{E}}).
\end{aligned} \tag{3.17}$$

Consequently, the above formulates a sequence of SDPs capable of approximating the maxima $\Omega(\boldsymbol{a},d_{\mathrm{E}})$ to arbitrary precision. Numerical results obtained from *any* of these SDPs (i.e., for any $N\geq L$) can be used to construct dimensional witnesses, which, when violated, certify the minimum dimension of the effective environment interacting with the system. A software implementation of the SDP in Eq. (3.16), however, is not straightforward even for small values of $\{d_{\mathrm{S}},d_{\mathrm{E}},L,N\}$, as the memory requirements quickly render the problem computationally intractable. Therefore, obtaining the numerical results presented in the next section, required a significant amount of optimization; see Sec. 3.7 for details. A schematic outline of all steps undertaken to formulate the SDP and obtain the numerical results is presented in Fig. 3.3.

The asymptotic convergence of the hierarchy of bounds in Eq. (3.17) for the SDP in Eq. (3.16), without PPT constraints, is given by [42]

$$|\Omega(\boldsymbol{a},d_{\mathrm{E}})-\widehat{\omega}^{N}(\boldsymbol{a},d_{\mathrm{E}})|\leq\frac{L(L+d_{\mathrm{ES}}^2-1)}{N+d_{\mathrm{ES}}^2}. \tag{3.18}$$

If PPT constraints are included, we only have partial analytical results for the asymptotic error. Numerical evidence and previous results involving PPT constraints [140] lead us to conjecture an asymptotic scaling of the form

$$|\Omega(\boldsymbol{a},d_{\mathrm{E}})-\widehat{\omega}^{N}(\boldsymbol{a},d_{\mathrm{E}})|\leq f(L,d_{\mathrm{ES}})\,O\left(\frac{1}{N^2}\right), \tag{3.19}$$

where $f(L,d_{\mathrm{ES}})$ is a function of L and d_{E}. We leave the analysis of these asymptotic error bounds to a future investigation.

		Formulation of the problem	Mathematical representation		
Exact	Not SDP	Original problem	$\max_U p(\boldsymbol{a}\|d_\mathrm{E})$		
		Choi-Jamiołkowski representation	$\max_{C_U} \mathrm{tr}[F^\mathrm{T}(C_U)^{\otimes L}]$		
		Convex relaxation	$\max_{C_U^L} \mathrm{tr}[F^\mathrm{T} C_U^L]$	$C_U^L = \sum_i p_i \cdot (C_U^{(i)})^{\otimes L}$	
		Symmetric separable states relaxation	$\max_{X_{\mathrm{A}_1^L}} \mathrm{tr}[F^\mathrm{T} X_{\mathrm{A}_1^L}]$	$X_{\mathrm{A}_1^L} \in \mathcal{P}_L^+ \cap \mathrm{SEP}_L$	
Outer Approximation	SDP	Separable approximation (Quantum de Finetti)	$\max_{X_{\mathrm{A}_1^N}} \mathrm{tr}[(F^\mathrm{T} \otimes \mathbb{1}_{\mathrm{A}_{L+1}^N}) X_{\mathrm{A}_1^N}]$	$X_{\mathrm{A}_1^N} \in \mathcal{P}_N^+$	Intractable
		Software implementation			
		Symmetric representation	$\max_x \mathrm{tr}[fx]$	$x \in \mathcal{S}_N$	
		Sparse implementation	$\max_x \mathrm{tr}[fx]$	$x_{\boldsymbol{t},\boldsymbol{t}'} = 0 \ \forall (\boldsymbol{t},\boldsymbol{t}') \notin E^*$	Tractable
		Numerical results			

Figure 3.3: Schematic of all steps undertaken for computing an upper bound for $\widehat{\omega}(\boldsymbol{a}, d_\mathrm{E})$ by formulating and solving the SDP. Detailed descriptions of each step are covered in Secs. 3.5.1 and 3.7.4. The tractable/intractable labels, on the right, refer to the case $d_\mathrm{E} = 2$.

3.5.1 Constructing the SDP relaxation

This section explains in detail the various steps to analytically formulate the maximization problem $\Omega(\boldsymbol{a}, d_\mathrm{E}) := \sup p(\boldsymbol{a}|d_\mathrm{E})$ as a semidefinite program, with Sec. 3.7.4 focusing on its software implementation. Figure 3.3 provides a schematic outline of our approach.

Let $\ell = 1, \ldots, L$ enumerate the time steps, and let I_ℓ and O_ℓ be the input and output spaces of the ℓ-th unitary evolution $\mathcal{U}$, respectively. For convenience, we write $A_\ell := \mathrm{I}_\ell \otimes \mathrm{O}_\ell$ for the joint space at step ℓ, and $\mathrm{A}_a^b := A_a \otimes A_{a+1} \otimes \cdots \otimes A_{b-1} \otimes A_b$ for the sequential spaces from a to b. We assume that all spaces are isomorphic, i.e., $\mathrm{I}_\ell \cong \mathrm{O}_\ell \cong \mathcal{H}_\mathrm{ES}$ for all ℓ, but we preserve labels for clarity. We can thus write the respective maps in the Choi-Jamiołkowski representation (see Eq. (3.4)) as:

$$C_U := \tfrac{1}{d_\mathrm{ES}} \mathrm{CJ}[\mathcal{U}], \quad M_a := \mathrm{CJ}[\mathcal{M}_a], \quad \widehat{M}_a := \mathrm{Tr}_\mathcal{O}[M_a], \tag{3.20}$$

where $\mathrm{CJ}[\Lambda]$ is the Choi matrix of the map Λ and C_U is normalized such that it corresponds to a quantum state, which is useful during implementation of the symmetric representation of the problem (Sec. 3.7.4). As before, we highlight that the local input ($\mathcal{I}$) and output ($\mathcal{O}$) spaces of these maps are interleaved: For the ℓ-th $\mathcal{U}$, $\mathcal{I} = \mathrm{I}_\ell$ and $\mathcal{O} = \mathrm{O}_\ell$, but for $\mathcal{M}_{a_\ell}$ we have $\mathcal{I} = \mathrm{O}_\ell$ and $\mathcal{O} = \mathrm{I}_{\ell+1}$. Since the final output state is discarded (i.e., we are only concerned with the *probability* of outcome sequences), the final $\mathcal{O}$ is traced out, yielding $\widehat{M}_a$ in the equation above. This structure of the spaces is illustrated in Fig. 3.4.

We may now specify the probability of the sequence $\boldsymbol{a}$ as

$$p(\boldsymbol{a}|d_\mathrm{E}) = \mathrm{Tr}\left[F^\mathrm{T}\left(\bigotimes_{\ell=1}^{L} C_U^{\mathrm{I}_\ell \mathrm{O}_\ell}\right)\right] = \mathrm{Tr}\left[F^\mathrm{T}(C_U)^{\otimes L}\right] \tag{3.21}$$

$$F := d_\mathrm{ES}^L \cdot (\rho_\mathrm{E}^0 \otimes \rho_\mathrm{S}^0)^{I_1} \otimes M_{a_1}^{O_1 I_2} \otimes \cdots \otimes M_{a_{L-1}}^{O_{L-1} I_L} \otimes \widehat{M}_{a_L}^{O_L}. \tag{3.22}$$

Where the correcting normalization factor d_ES^L was incorporated into F (to make up for the normalization of C_U), and F^T denotes the transpose with respect to the basis chosen for the isomorphism. In order to obtain an upper-bound for $p(\boldsymbol{a}|d_\mathrm{E})$, our goal is to optimize Eq. (3.21) over all possible unitaries, in terms of C_U. As per the above definitions, it follows that C_U must satisfy the constraints:

- $C_U \geq 0$ and $\mathrm{rank}\, C_U = 1$, as it represents a unitary channel,
- $\mathrm{Tr}_\mathcal{O}[C_U] = \mathbb{1}_\mathcal{I}/d_\mathrm{ES}$, as it is a trace preserving map

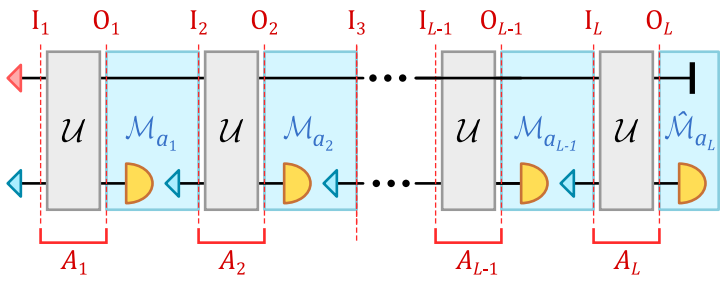

Figure 3.4: Diagram of the specific protocol discussed in this work, with each time step written in terms of distinct input and output Hilbert spaces, denoted by the dotted lines. Note that the input and output spaces of the $\mathcal{U}$ and $\mathcal{M}_a$ are interleaved: $\mathcal{U}$ has inputs I_ℓ and outputs I_ℓ, but for $\mathcal{M}_{a_\ell}$, inputs are O_ℓ and outputs $\mathrm{I}_{\ell+1}$.

We may thus define our optimization problem as

Optimization Problem 1. The initial formulation of the problem.

$$\begin{aligned} \textbf{Given:} \quad & \rho_{\mathrm{E}}^0,\ \rho_{\mathrm{S}}^0,\ \{M_a\}_a,\ \boldsymbol{a} \\ \textbf{Find:} \quad & \Omega(\boldsymbol{a}, d_{\mathrm{E}}) = \max_{C_U} \ \mathrm{Tr}\left[F^{\mathrm{T}}(C_U)^{\otimes L}\right] \\ \textbf{Subject to:} \quad & C_U \geq 0, \quad \mathrm{rank}\, C_U = 1, \quad \mathrm{Tr}_{\mathcal{O}}[C_U] = \mathbb{1}_{\mathcal{I}}/d_{\mathrm{ES}}. \end{aligned} \tag{3.23}$$

This may be immediately relaxed to a convex form, without affecting the maximum of the objective function, as

Optimization Problem 2. The convex relaxation of the original problem.

$$\begin{aligned} \textbf{Given:} \quad & \rho_{\mathrm{E}}^0,\ \rho_{\mathrm{S}}^0,\ \{M_a\}_a,\ \boldsymbol{a} \\ \textbf{Find:} \quad & \Omega(\boldsymbol{a}, d_{\mathrm{E}}) = \max_{\{|\varphi_i\rangle\}_i} \mathrm{Tr}\left[F^{\mathrm{T}} \widetilde{X}_{\mathrm{A}_1^L}\right], \\ \text{with} \quad & \widetilde{X}_{\mathrm{A}_1^L} = \sum_i p_i \, |\varphi_i\rangle\langle\varphi_i|^{\otimes L},\ p_i \geq 0,\ \sum_i p_i = 1 \\ \textbf{Subject to:} \quad & \mathrm{Tr}_{\mathcal{O}}[|\varphi_i\rangle\langle\varphi_i|] = \mathbb{1}_{\mathcal{I}}/d_{\mathrm{ES}}, \quad |\varphi_i\rangle \in \mathcal{H}_{\mathrm{ES}} \otimes \mathcal{H}_{\mathrm{ES}}. \end{aligned} \tag{3.24}$$

However, this optimization problem is non-linear, not only due to the tensor product in $|\varphi_i\rangle\langle\varphi_i|^{\otimes L}$, but also on the rank-1 nature of $|\varphi_i\rangle\langle\varphi_i|$. Let us now define our *target set* $\mathcal{T}$ as

$$\widetilde{X}_{\mathrm{A}_1^L} \in \mathcal{T} := \left\{ \sum_i p_i \, |\varphi_i\rangle\langle\varphi_i|^{\otimes L} \;\middle|\; |\varphi_i\rangle \in \mathcal{H}_{\mathrm{ES}} \otimes \mathcal{H}_{\mathrm{ES}},\ p_i \geq 0,\ \sum_i p_i = 1,\ \mathrm{Tr}_{\mathcal{O}}[|\varphi_i\rangle\langle\varphi_i|] = \mathbb{1}_{\mathcal{I}}/d_{\mathrm{ES}} \right\}. \tag{3.25}$$

Elements of this set will not generally be rank-1, but the optimal solutions of Probs. 1 and 2 belong to this set. Our goal now is to approach $\mathcal{T}$ by means of further relaxations, which is achieved by exploiting the symmetry of $\widetilde{X}_{\mathrm{A}_1^L}$.

The unitary channel constraints

The symmetric structure of $\widetilde{X}_{\mathrm{A}_1^L}$ requires all subspaces to be in the same local state. We can relax this by considering instead the separable set on L parties,

$$X_{\mathrm{A}_1^L} \in \mathrm{SEP}_L := \mathrm{conv}\left\{ \bigotimes_{\ell=1}^{L} Q_\ell \;\middle|\; Q_\ell \in \mathcal{B}(\mathcal{H}_{\mathrm{ES}} \otimes \mathcal{H}_{\mathrm{ES}}),\ Q_\ell \geq 0,\ \mathrm{Tr}\,[Q_\ell] = 1 \right\}, \tag{3.26}$$

such that all local spaces become independent. Here, conv denotes the convex hull of the set, and $\mathcal{B}$ the set of bounded operators. We must now find ways to restore rank-1 and permutation invariance constraints for the optima over this set. For now, we highlight that $\mathcal{T} \subset \mathrm{SEP}_L$, but simply switching from $\widetilde{X}_{\mathrm{A}_1^L} \in \mathcal{T}$ to $X_{\mathrm{A}_1^L} \in \mathrm{SEP}_L$ *does* change our problem, so that we must impose further constraints to the set SEP_L to restore the original problem in $\mathcal{T}$. The first step is restoring symmetry and optimality of rank-1 to $X_{\mathrm{A}_1^L}$. After this, we may restore the partial trace constraint.

Let $\sigma \in \mathfrak{S}_n$ be a permutation from the set of all permutations on n symbols $\mathfrak{S}_n$. We define

$$V_\sigma := \sum_{i_1,\cdots,i_n=1}^{d} |i_{\sigma^{-1}(1)},\cdots,i_{\sigma^{-1}(n)}\rangle\langle i_1,\cdots,i_n|, \tag{3.27}$$

$$P_n^+ := \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n} V_\sigma \tag{3.28}$$

as, respectively, the permutation operator and projector onto the symmetric subspace. The symmetric subspace on n copies of $\mathcal{H}$ is defined as

$$\mathrm{Sym}_n(\mathcal{H}) := \left\{ |\psi\rangle \in \mathcal{H}^{\otimes n} \;\middle|\; P_n^+ |\psi\rangle = |\psi\rangle \right\}, \tag{3.29}$$

and the space of symmetric operators is given by the set [85]:

$$\mathrm{Sym}_n(\mathcal{B}(\mathcal{H})) := \{A \in \mathcal{B}(\mathcal{H}^{\otimes n}) \mid P_n^+ A = A\}. \tag{3.30}$$

For a Hermitian operator ρ, the symmetry condition can be equivalently written either as $P_n^+\rho = \rho$ or $P_n^+\rho P_n^+ = \rho$. In fact, if $P_n^+\rho = \rho$, then $\rho^\dagger = (P_n^+\rho)^\dagger = \rho^\dagger P^+ = \rho P^+ = \rho \Rightarrow P_n^+\rho P_n^+ = \rho$. Conversely, $\rho = P_n^+\rho P_n^+ = (P_n^+)^2\rho P_n^+ = P_n^+\rho$. More details about the symmetric subspace can be found in Sec. 3.7.4.

Ensuring symmetry and rank-1 at the optimum

By utilizing $X_{\mathrm{A}_1^L}$ as our variable as in Eq. (3.26), we have lost both symmetry and rank-1 guarantees at the optimum. However, both can be restored by ensuring $X_{\mathrm{A}_1^L}$ is in the symmetric subset of SEP_L; see, e.g., [105]. For completeness, we provide in the following an elementary proof that

$$\mathrm{Sym}_n(\mathcal{B}(\mathcal{H})) \cap \mathrm{SEP}_n = \left\{ \sum_i p_i |\varphi_i\rangle\langle\varphi_i|^{\otimes n} \;\middle|\; |\varphi_i\rangle \in \mathcal{H}_{\mathrm{ES}} \otimes \mathcal{H}_{\mathrm{ES}},\ \langle\varphi_i|\varphi_i\rangle = 1,\ p_i \geq 0,\ \sum_i p_i = 1 \right\}, \tag{3.31}$$

which implies $\mathcal{T} \subset \mathrm{Sym}_L(\mathcal{B}(\mathcal{H})) \cap \mathrm{SEP}_L$. In fact, consider $\rho \in \mathrm{Sym}_n(\mathcal{B}(\mathcal{H})) \cap \mathrm{SEP}_n$. Since it is separable, we can write it as a convex mixture of pure product states, i.e.,

$$\rho = \sum_i p_i |\Phi_i\rangle\langle\Phi_i|, \tag{3.32}$$

with $|\Phi_i\rangle$ a n-party pure product state, i.e., $|\Phi_i\rangle = \bigotimes_{j=1}^n |\varphi_{i,j}\rangle$, with the $\{|\varphi_{i,j}\rangle\}_j$ not necessarily equal for a given i. However, $\rho \in \mathrm{Sym}_n(\mathcal{B}(\mathcal{H}))$ implies $\mathrm{ran}(\rho) \subseteq \mathrm{Sym}_n(\mathcal{H})$, and if we prove that $\mathrm{ran}(|\Phi_i\rangle\langle\Phi_i|) \subseteq \mathrm{Sym}_n(\mathcal{H})$ for all i, then we find that $\{|\varphi_{i,j}\rangle\}_j$ must be identical for each i, otherwise $|\Phi_i\rangle$ fails to be symmetric, namely, $|\Phi_i\rangle\langle\Phi_i| = |\varphi_i\rangle\langle\varphi_i|^{\otimes n}$ and, thus, $\rho = \sum_i p_i |\varphi_i\rangle\langle\varphi_i|^{\otimes n}$.

It remains to prove that $\mathrm{ran}(|\Phi_i\rangle\langle\Phi_i|) \subseteq \mathrm{Sym}_n(\mathcal{H})$. To show this, we note that by Eq. (3.32) and positivity of ρ and $|\Phi_i\rangle\langle\Phi_i|$, we have $\mathrm{Ker}(\rho) = \cap_i \mathrm{Ker}(|\Phi_i\rangle\langle\Phi_i|)$. By Hermiticity, it follows that $\mathrm{ran}(\rho) = (\mathrm{Ker}(\rho))^\perp = \mathrm{span}(\cup_i \mathrm{ran}(|\Phi_i\rangle\langle\Phi_i|))$, which concludes the proof.

Importantly, Eq. (3.31), restores the symmetry and rank-1 properties for optimum solutions of Prob. 2 when $X_{\mathrm{A}_1^L}$ is restricted to $\mathrm{Sym}_n(\mathcal{B}(\mathcal{H})) \cap \mathrm{SEP}_n$. In light of this, in the following we use $\mathrm{SymSEP}_n := \mathrm{Sym}_n(\mathcal{B}(\mathcal{H})) \cap \mathrm{SEP}_n$.

The trace-preserving constraint

We now restore the trace-preserving constraint $\mathrm{Tr}_{\mathcal{O}}[|\varphi_i\rangle\langle\varphi_i|] = \mathbb{1}_{\mathcal{I}}/d_{\mathrm{ES}}$ for states expressed as in Eq. (3.31). This is established by proving that it is sufficient to satisfy this condition for a single party in the convex mixture.

Proposition 1. For any $X_{\mathrm{A}_1^L} \in \mathrm{SymSEP}_L$:

$$\mathrm{Tr}_{\mathcal{O}_1}\left[X_{\mathrm{A}_1^L}\right] = \frac{\mathbb{1}_{\mathcal{I}_1}}{d_{\mathrm{ES}}} \otimes \mathrm{Tr}_{\mathrm{A}_1}\left[X_{\mathrm{A}_1^L}\right] \iff \mathrm{Tr}_{\mathcal{O}_\ell}\left[|\varphi_i\rangle\langle\varphi_i|_{\mathrm{A}_\ell}\right] = \frac{\mathbb{1}_{\mathcal{I}_\ell}}{d_{\mathrm{ES}}} \otimes |\varphi_i\rangle\langle\varphi_i|^{\otimes L-1} \quad \forall \ell. \tag{3.33}$$

Proof. We provide a generalization of the arguments in [216, App. A]. From Eq. (3.31), if $\mathrm{P}_L^+ X_{\mathrm{A}_1^L} = X_{\mathrm{A}_1^L}$, we know it admits a decomposition of the form

$$X_{\mathrm{A}_1^L} = \sum_i p_i \, |\varphi_i\rangle\langle\varphi_i|^{\otimes L}, \qquad p_i \geq 0, \quad \sum_i p_i = 1. \tag{3.34}$$

We introduce the one-party auxiliary map $\mathcal{E}(\cdot) = \mathrm{Tr}_{\mathcal{O}}[\cdot] - \mathrm{Tr}\,[\cdot]\,\mathbb{1}_{\mathcal{I}}/d_{\mathrm{ES}}$ implementing our partial trace constraint in $\mathcal{T}$, such that $(\mathcal{E}_{\mathrm{A}_1} \otimes \mathrm{id}_{\mathrm{A}_2} \otimes \mathrm{id}_{\mathrm{A}_3^L})(X_{\mathrm{A}_1^L}) = 0$ for any $X_{\mathrm{A}_1^L} \in \mathrm{SymSEP}_L \cap \mathcal{T}$, and, due to symmetry, the same is true if the map had been applied to the second party instead. Defining $\tilde{\mathcal{E}}(\cdot) = [\mathcal{E}(\cdot)]^\dagger$, we may then write $(\mathcal{E}_{\mathrm{A}_1} \otimes \tilde{\mathcal{E}}_{\mathrm{A}_2} \otimes \mathrm{id}_{\mathrm{A}_3^L})(X_{\mathrm{A}_1^L}) = 0$. As $\mathcal{E}$ acts on Hermitian operators and is Hermiticity-preserving, we have $\mathcal{E} = \tilde{\mathcal{E}}$ and the explicit distinction is only made for added clarity in the following proof. Applying the map to the convex mixture in Eq. (3.34), we obtain

$$\begin{aligned} &\sum_i p_i \mathcal{E}_{\mathrm{A}_1}(|\varphi_i\rangle\langle\varphi_i|_{\mathrm{A}_1}) \otimes \tilde{\mathcal{E}}_{\mathrm{A}_1}(|\varphi_i\rangle\langle\varphi_i|_{\mathrm{A}_2}) \otimes |\varphi_i\rangle\langle\varphi_i|^{\otimes L-2} \\ &= \sum_i p_i E_i \otimes E_i^\dagger \otimes |\varphi_i\rangle\langle\varphi_i|^{\otimes L-2} = 0, \end{aligned} \tag{3.35}$$

where $E_i = \mathcal{E}(|\varphi_i\rangle\langle\varphi_i|)$. We need this to be true for all E_i if we want to ensure $X_{\mathrm{A}_1^L}$ obeys our partial trace constraint globally in the convex mixture. To prove this is the case, let $G \in \mathbb{C}^{d_{\mathrm{ES}}^2 \times d_{\mathrm{ES}}^2}$, so that

$$\begin{aligned} &\mathrm{Tr}\left[(G \otimes G^\dagger \otimes \mathbb{1}_{\mathrm{A}_3^L}) \left(\sum_i p_i E_i \otimes E_i^\dagger \otimes |\varphi_i\rangle\langle\varphi_i|^{\otimes L-2}\right)\right] \\ &= \sum_i p_i |\mathrm{Tr}\,[G E_i]|^2 \,\mathrm{Tr}\left[|\varphi_i\rangle\langle\varphi_i|^{\otimes L-2}\right] = 0, \end{aligned} \tag{3.36}$$

which must hold for any G. Since $p_i \geq 0$, $|\mathrm{Tr}\,[G E_i]| \geq 0$, and $\mathrm{Tr}\left[|\varphi_i\rangle\langle\varphi_i|^{\otimes L-2}\right] = 1$, this can only be true for all G if all $E_i = 0$. Furthermore, by the symmetry of each term in the mixture, it is then sufficient to ensure the constraint is satisfied in a single party. From Eq. (3.31), the converse statement follows trivially. □

Thus, any state $X_{\mathrm{A}_1^L} \in \mathrm{SymSEP}_L$ satisfying $\mathrm{Tr}_{\mathrm{O}_1}\left[X_{\mathrm{A}_1^L}\right] = (\mathbb{1}_{\mathrm{I}_1}/d_{\mathrm{ES}}) \otimes \mathrm{Tr}_{\mathrm{A}_1}\left[X_{\mathrm{A}_1^L}\right]$ is part of the target set $\mathcal{T}$ and vice versa.

The final rank-constrained problem over SEP

Given all of the previous results, we have now established a relation between the sets $\mathcal{T}$ and SEP_L,

$$\mathcal{T} = \left\{ X_{\mathrm{A}_1^L} \in \mathrm{SEP}_L \;\middle|\; \mathrm{P}_L^+ X_{\mathrm{A}_1^L} = X_{\mathrm{A}_1^L}, \quad \mathrm{Tr}_{\mathrm{O}_1}\left[X_{\mathrm{A}_1^L}\right] = (\mathbb{1}_{\mathrm{I}_1}/d_{\mathrm{ES}}) \otimes \mathrm{Tr}_{\mathrm{A}_1}\left[X_{\mathrm{A}_1^L}\right] \right\}, \tag{3.37}$$

which allows us to remove the non-linear dependence on the tensor product $(C_U)^{\otimes L}$, by replacing the search space with SymSEP_L, while obeying the linear constraints on the partial trace of one party. We thus obtain:

Optimization Problem 3. Rank-1 and symmetric optimum (exact)

$$\begin{aligned} \textbf{Given:} \quad & \rho_{\mathrm{E}}^{0},\ \rho_{\mathrm{S}}^{0},\ \{M_a\}_a,\ \boldsymbol{a} \\ \textbf{Find:} \quad & \Omega(\boldsymbol{a}, d_{\mathrm{E}}) = \max_{X_{\mathrm{A}_1^L} \in \mathrm{SEP}_L} \mathrm{Tr}\left[F^{\mathrm{T}} X_{\mathrm{A}_1^L}\right] \\ \textbf{Subject to:} \quad & X_{\mathrm{A}_1^L} \geq 0, \quad \mathrm{Tr}\left[X_{\mathrm{A}_1^L}\right] = 1, \quad \mathrm{P}_L^{+} X_{\mathrm{A}_1^L} = X_{\mathrm{A}_1^L}, \\ & \mathrm{Tr}_{\mathrm{O}_1}\left[X_{\mathrm{A}_1^L}\right] = \frac{\mathbb{1}_{\mathrm{I}_1}}{d_{\mathrm{ES}}} \otimes \mathrm{Tr}_{\mathrm{A}_1}\left[X_{\mathrm{A}_1^L}\right] \end{aligned} \tag{3.38}$$

However, $X_{\mathrm{A}_1^L} \in \mathrm{SEP}_L$ is still not a linear constraint and highly nontrivial to be enforced, as a full characterization of the separable set is NP-Hard [82]. Therefore, while a priori appearing more manageable, the above problem is still not in a form that can be tackled numerically. Fortunately, the symmetric constraint arising from our problem can also be exploited to obtain approximations of the symmetric separable states.

The SEP constraint

Approximation of SEP via the quantum de Finetti theorem

Ensuring $X_{\mathrm{A}_1^L} \in \mathrm{SEP}_L$ cannot be done exactly, so as it stands the problem is still numerically intractable. However, it is possible to define arbitrarily precise outer approximations of the separable set through symmetric extensions and the quantum de Finetti theorem [39, 47]. Concretely, we can approximate, in principle to any desirable precision, a $X_{\mathrm{A}_1^L} \in \mathrm{SEP}_L$ by a marginal over a larger symmetric (and not necessarily separable) state $X_{\mathrm{A}_1^N}$, provided N is sufficiently large. For any finite N, this leads to an outer approximation of $\mathcal{T}$, which establishes a convergent hierarchy of outer approximations, such that for any $N \geq L$, we have

$$\begin{aligned} & \Omega(\boldsymbol{a}, d_{\mathrm{E}}) \leq \cdots \leq \widehat{\omega}^{N+1}(\boldsymbol{a}, d_{\mathrm{E}}) \leq \widehat{\omega}^{N}(\boldsymbol{a}, d_{\mathrm{E}}) \leq \cdots \leq \widehat{\omega}^{L}(\boldsymbol{a}, d_{\mathrm{E}}), \\ & \text{with} \quad \lim_{N\to\infty} \widehat{\omega}^{N}(\boldsymbol{a}, d_{\mathrm{E}}) = \Omega(\boldsymbol{a}, d_{\mathrm{E}}). \end{aligned} \tag{3.39}$$

More precisely, using the quantum de Finetti theorem we can establish asymptotic error bounds on this approximation [42, Cor.1]. In terms of the trace norm, this asymptotic error bound is given by

$$\left\| X_{\mathrm{A}_1^L} - \mathrm{Tr}_{\mathrm{A}_{L+1}^N}\left[X_{\mathrm{A}_1^N}\right] \right\|_1 \leq \frac{2L(L + d_{\mathrm{ES}}^2 - 1)}{N + d_{\mathrm{ES}}^2}. \tag{3.40}$$

As the symmetry requirement applies to both $X_{\mathrm{A}_1^L}$ and $X_{\mathrm{A}_1^N}$, this application of the quantum de Finetti theorem for approximating $X_{\mathrm{A}_1^L} \in \mathrm{SEP}_L$ is straightforward in the SDP in Eq. (3.16), only requiring the use of a larger state $X_{\mathrm{A}_1^N}$ as the optimization variable, and a suitable adjustment of the objective function using $(F^{\mathrm{T}} \otimes \mathbb{1}_{\mathrm{A}_{L+1}^N})$.

Additional separability constraints

To improve the approximation, at an extra computational cost, we may also include additional separability constraints. In our case, we consider the constraints of positive partial transpose (PPT) on $X_{\mathrm{A}_1^N}$:

$$X_{\mathrm{A}_1^N}^{\mathrm{T}_\alpha} \geq 0, \quad \forall \text{ non-equivalent bipartitions } \alpha. \tag{3.41}$$

Here, "non-equivalent bipartitions" refers to the fact that it is unnecessary to include all bipartitions, as the state $X_{\mathrm{A}_1^N}$ is symmetric, and thus permutation invariant. Therefore, the bipartitions needed are α_k,

where, e.g., we transpose only the first k subspaces, for $1 \leq k \leq \lfloor N/2 \rfloor$, with symmetry taking care of the remaining constraints.

The existence of entangled states which have positive partial transpose [92] means these constraints reduce the feasible convex set to that of PPT states, not the separable states. While entangled PPT states are not suitable solutions to the original problem, they are still adequate approximations and provide upper bounds for the actual optimal values. Inclusion of PPT constraints will perform at least as well as optimizing without them (i.e., the solution can only be improved), but convergence is significantly improved in certain cases [140] to $O(1/N^2)$ as opposed to $O(1/N)$ in the absence of these constraints. As shown in Table 3.1, for the $d_{\mathrm{E}} = 1$ case PPT constraints were sufficient to provide the exact analytical bounds.

SDP for SEP relaxation

By all of the above results, we arrive at the following SDP relaxation of the original problem:

Optimization Problem 4. Final outer approximation (SDP).

$$\begin{aligned}
\textbf{Given:}\quad & d_{\mathrm{E}},\ d_{\mathrm{S}},\ \rho_{\mathrm{E}}^{0},\ \rho_{\mathrm{S}}^{0},\ \{M_a\}_a,\ \boldsymbol{a} \\
\textbf{Find:}\quad & \widehat{\omega}^{N}(\boldsymbol{a}, d_{\mathrm{E}}) := \max_{X_{\mathrm{A}_1^N}} \mathrm{Tr}\left[(F^{\mathrm{T}} \otimes \mathbb{1}_{\mathrm{A}_{L+1}^N}) X_{\mathrm{A}_1^N}\right] \\
\textbf{Subject to:}\quad & X_{\mathrm{A}_1^N} \geq 0, \quad \mathrm{Tr}\left[X_{\mathrm{A}_1^N}\right] = 1, \quad \mathrm{P}_N^{+} X_{\mathrm{A}_1^N} = X_{\mathrm{A}_1^N}, \\
& \mathrm{Tr}_{\mathrm{O}_1}\left[X_{\mathrm{A}_1^N}\right] = \frac{\mathbb{1}_{\mathrm{I}_1}}{d_{\mathrm{ES}}} \otimes \mathrm{Tr}_{\mathrm{A}_1}\left[X_{\mathrm{A}_1^N}\right] \\
& X_{\mathrm{A}_1^N}^{\mathrm{T}_\alpha} \geq 0, \forall \alpha \in \mathfrak{A},
\end{aligned} \tag{3.42}$$

with $\mathfrak{A}$ denoting the set of non-equivalent bipartitions (Sec. 3.5.1). The solutions of this SDP provide upper-bounds for the maxima $\Omega(\boldsymbol{a}, d_{\mathrm{E}})$.

3.6 Numerical results

In this section, we discuss the numerical results we obtained for $\widehat{\omega}^{N}(\boldsymbol{a}, d_{\mathrm{E}})$. The SDPs were run with CVXPY [57, 4] using the solver SCS [144, 145], on a compute server with an Intel Xeon Gold 5218 24-core processor at 2.294 GHz, and with 128 GB of RAM. For $d_{\mathrm{E}} = 1$, optimization time was in the order of seconds, whereas $d_{\mathrm{E}} = 2$ scenarios required from 4 up to 14 hours to complete.

For the explicit computation, we chose the following measurement protocol: $\mathcal{A} = \{0, 1\}$, $\rho_{\mathrm{E}}^{0} = |0\rangle\langle 0|_E$, $d_{\mathrm{S}} = 2$, $\rho_{\mathrm{S}}^{0} = |0\rangle\langle 0|_S$ and $E_S^a = |a\rangle\langle a|_S$, so that

$$\mathcal{M}_a(\rho_{\mathrm{ES}}) = \mathrm{Tr}\left[\rho_{\mathrm{ES}} \cdot \mathbb{1}_E \otimes |a\rangle\langle a|_S\right] \otimes |0\rangle\langle 0|_S\,. \tag{3.43}$$

For this particular choice, the bounds $\widehat{\omega}(\boldsymbol{a}, d_{\mathrm{E}})$ are the same as for the isolated system case studied in [198], where explicit time evolutions were found through gradient descent techniques. This allows us to use these known values as benchmarks and investigate the convergence properties of our SDP.

The analytical maxima for $d_{\mathrm{E}} = 1$ described in Sec. 3.4 also serve as a useful test for the soundness of our approach. We ran the SDP for all binary sequences (up to $0 \leftrightarrow 1$ relabeling symmetry), for $L \in \{2, 3, 4\}$, and $N \in \{L, L+1\}$. Every case was tested with and without the additional PPT constraints for comparison. Results for $L = 2, 3$ are shown in Table 3.1, in which it can be observed that, for $d_{\mathrm{E}} = 1$, either the PPT constraints or a symmetric extension of a single extra system appear to be sufficient for achieving the exact analytical maximum.

L	N	$\boldsymbol{a}$	Without PPT	With PPT	$\Omega(\boldsymbol{a}, 1)$
2	2	00	0.999996	0.999999	1
2	3	00	1.000000	1.000000	1
2	2	01	0.500000	0.250002	1/4
2	3	01	0.250007	0.250001	1/4
3	3	000	1.000010	1.000000	1
3	4	000	1.000000	0.999991	1
3	3	001	0.250005	0.148149	4/27
3	4	001	0.148148	0.148148	4/27
3	3	010	0.250005	0.148149	4/27
3	4	010	0.148148	0.148148	4/27
3	3	011	0.250005	0.148149	4/27
3	4	011	0.148148	0.148148	4/27

Table 3.1: Results of the SDP for $d_{\mathrm{E}} = 1$, compared with the analytical maxima $\Omega(\boldsymbol{a}, 1)$. Note that $4/27 = 0.\overline{148}$. It can be seen that either PPT constraints or an extension of one extra system (i.e., $N = L + 1$) was sufficient to achieve the analytical maximum in this case. Similar results were obtained for $L = 4$, omitted here for conciseness.

We have also solved the SDP for the case $d_{\mathrm{S}} = d_{\mathrm{E}} = 2$, for various sequences without a trivial maximum probability (i.e., $\Omega(\boldsymbol{a}, d_{\mathrm{E}}) < 1$), but only without the additional PPT constraint, as the requirements needed for its addition would have exceeded the available memory. Additionally, since the involved SDPs are only numerically tractable for $N \approx L$, and no convergence between lower and upper bounds was observed for the values we managed to compute, we can only claim to have obtained upper bounds for the maximum $\Omega(\boldsymbol{a}, d_{\mathrm{E}})$, as in Eq. (3.17).

As we show in Sec. 3.8.3, our choice of probe state and measurements (Eq. (3.43)) for $d_{\mathrm{S}} = 2$ reproduces the quantum scenarios previously investigated in [198] (see Sec. 2.8), where lower bounds were obtained through explicit time evolutions found via gradient descent techniques. Therefore, we can compare the upper bounds obtained through our SDP against these previous lower bounds, allowing us to estimate the range of values containing the maxima $\Omega(\boldsymbol{a}, d_{\mathrm{E}} = 2)$. These comparisons are shown in Table 3.2, where, for instance, we see $0.437341 \leq \Omega(001, 2) \leq 0.521219$.

As a concrete example of a violation of this bound, for $d_{\mathrm{S}} = 2$ and $d_{\mathrm{E}} = 3$, we may construct a unitary in $\mathcal{H}_{\mathrm{ES}}$ as follows:

$$\begin{aligned} U = {} & |1\rangle\langle 0|_E \otimes |0\rangle\langle 0|_S + |2\rangle\langle 1|_E \otimes |0\rangle\langle 0|_S + |2\rangle\langle 0|_E \otimes |1\rangle\langle 1|_S \\ & + |1\rangle\langle 1|_E \otimes |1\rangle\langle 1|_S + |0\rangle\langle 2|_E \otimes |0\rangle\langle 1|_S + |0\rangle\langle 2|_E \otimes |1\rangle\langle 0|_S \,. \end{aligned} \tag{3.44}$$

For the ρ_{E}^0, ρ_{S}^0 and E_S^a we have chosen, the above unitary gives $p(001|d_{\mathrm{E}} = 3) = 1 > \Omega(001, 2)$, implying that the bound found for $d_{\mathrm{E}} = 2$ is not only nontrivial, but actually witnesses environment dimension since it can indeed be violated by means of a larger environment ($d_{\mathrm{E}} = 3$ in this case). In fact, this example certifies that $\Omega(001, 3) = 1$; see also Sec. 3.8.3.

L	N	$\boldsymbol{a}$	Lower bound (GD)	Upper bound (SDP)
3	3	001	0.437341	0.683477
3	4	001	0.437341	0.521219
4	4	0010	0.437142	0.512220
4	4	0011	0.362047	0.487058
4	4	0100	0.333147	0.494499
4	4	0110	0.361968	0.488837
4	4	0001	0.300545	0.492088

Table 3.2: Comparison between SDP upper bounds and lower bounds previously known for the maximum $\Omega(\boldsymbol{a}, d_{\mathrm{E}})$, for $d_{\mathrm{E}} = 2$. Lower bounds were obtained through gradient descent (GD) in Ref. [198], but are not suitable as witnesses. All results are in agreement, and for $\boldsymbol{a} = 001$, we observe convergence of the bounds as N increases.

3.7 Implementation

This section outlines the concrete steps we have taken to relax the original problem to a hierarchy of SDPs, and to make these SDPs numerically tractable; see Sec. 3.7.4 for more technical details. As noted previously, a straightforward implementation is not computationally tractable, even for short sequences, due to the large number of variables and constraints involved.

We have addressed these additional numerical challenges by exploiting several properties of the problem. Firstly, the symmetric constraint imposed on $X_{\mathrm{A}_1^N}$ can be satisfied automatically by expressing $X_{\mathrm{A}_1^N}$ in terms of a basis for the symmetric subspace in the numerical implementation of the SDP. While this provides a significant reduction in the total number of involved variables and constraints, it is by itself insufficient to make the problem tractable.

However, thanks to our specific choice of initial states and measurements, we were able to exploit the sparsity of the objective function (as defined by F) to significantly reduce the number of variables and constraints, by eliminating all of those that do not affect the objective, either directly or indirectly. This sparse representation was obtained through a novel heuristic algorithm, developed for this problem, which we have since expanded upon in Ref. [199]. This heuristic is the subject of Ch. 5.

Note that, while generally such relaxations would result in an outer approximation of the original problem, here an exact sparse representation can be obtained using our heuristic. See Sec. 5.4.1 for the details.

3.7.1 Choice of parameters

Before a concrete implementation, one must first choose the initial environment state ρ_{E}^0, as well as fixing the probe state and measurements. As the objective function is convex on ρ_{E}^0, and the optimization is over all unitaries, we may – without loss of generality – fix a pure initial state $\rho_{\mathrm{E}}^0 = |0\rangle\langle 0|_E$ and eliminate any assumptions on the environment state; see Sec. 3.8.2. On the other hand, as assumptions on d_{S}, ρ_{S}^0 and $\{E_S^a\}_a$ are experiment-dependent, one must search for "proper" probe states and measurements on a case-by-case basis, but optimal choices, leading to a larger bound, can also be addressed in general terms; see Sec. 3.8.3.

For the choice of sequence, in particular, if it is too simple relative to d_{E}, the maximum attainable probability may be trivial (i.e., equal to one), such that no optimization is required, and no witness can

be constructed. Therefore, it is important to choose sequences which have a nontrivial maximum for a given d_{E}. For the case of repeated measurements on isolated systems, this problem has been solved via the notion of "deterministic complexity" introduced in Ref. [198], and discussed in Sec. 2.5.1, which defines the minimum requirements for a sequence to be able to occur deterministically. In the open system case, the conditions for determinism involve not only the available memory (i.e., the environment), but also the dimension of the probe system. We elaborate on these conditions in Sec. 3.8.3.

3.7.2 Symmetric representation

Effectively solving any of the SDPs in the hierarchy requires a significant amount of optimization, as a naive implementation quickly becomes computationally intractable. As a rough example, without any simplifications, $X_{\mathrm{A}_1^N}$ is a square matrix of size $(d_{\mathrm{ES}}^2)^N$, which in the simplest non-trivial scenario, $d_{\mathrm{E}} = d_{\mathrm{S}} = 2$ and $N = L = 3$, already results in a 4096×4096 matrix, with over 16 million complex scalar variables. The partial trace constraint alone involves over 1 million linear equations between these variables, making the SDP numerically intractable in this naïve formulation.

In practice, several simplifications can be made; see Sec. 3.7.4 for full details. $X_{\mathrm{A}_1^N}$ and $F \otimes \mathbb{1}_{\mathrm{A}_{L+1}^N}$ may be written directly in terms of an operator basis for the symmetric subspace, thus automatically satisfying the symmetry constraint and significantly reducing the number of variables. Defining an isometry between the symmetric subspace of A_1^N, denoted by $\mathcal{S}_N$, and a canonical basis for $\mathcal{S}_N$,

$$V_N = \sum_{\mathrm{t}} |\mathrm{t}\rangle_{\mathcal{S}_N} \langle \mathrm{sym}(\mathrm{t})|_{\mathrm{A}_1^N}, \tag{3.45}$$

we may exploit the fact the problem lies entirely within the symmetric subspace and cast all matrices as $\dim \mathcal{S}_N \times \dim \mathcal{S}_N$ matrices:

$$\begin{aligned} X_{\mathrm{A}_1^N} &= \sum_{\mathrm{t},\mathrm{t}'} x_{\mathrm{t},\mathrm{t}'} \, |\mathrm{sym}(\mathrm{t})\rangle\langle \mathrm{sym}(\mathrm{t}')|_{\mathrm{A}_1^N}, \\ x = V_N X_{\mathrm{A}_1^N} V_N^\dagger, \quad &\text{and} \quad f = V_N (F^{\mathrm{T}} \otimes \mathbb{1}_{\mathrm{A}_{L+1}^N}) V_N^\dagger, \end{aligned} \tag{3.46}$$

where t denotes a "type" for the canonical representation of symmetric states $|\mathrm{sym}(\mathrm{t})\rangle_{\mathrm{A}_1^N}$; see Sec. 3.7.4. Under this representation, the objective function becomes

$$\begin{aligned} \mathrm{Tr}\left[(F^{\mathrm{T}} \otimes \mathbb{1}_{\mathrm{A}_{L+1}^N}) X_{\mathrm{A}_1^N}\right] &= \mathrm{Tr}\left[(F^{\mathrm{T}} \otimes \mathbb{1}_{\mathrm{A}_{L+1}^N}) {V_N}^\dagger x V_N\right] \\ &= \mathrm{Tr}\left[V_N (F^{\mathrm{T}} \otimes \mathbb{1}_{\mathrm{A}_{L+1}^N}) {V_N}^\dagger x\right] = \mathrm{Tr}\left[fx\right], \end{aligned} \tag{3.47}$$

and the constraints may be formulated in terms of $x_{\mathrm{t},\mathrm{t}'}$ directly, without resorting to the full space A_1^N. Note that $VV^\dagger = \mathbb{1}_{\mathcal{S}_N}$, but $V^\dagger V \neq \mathbb{1}_{\mathrm{A}_1^N}$, i.e., $V^\dagger V$ is the projector onto the symmetric subspace of A_1^N. As the dimension of $\mathcal{S}_N$ is [85]

$$\dim \mathcal{S}_N = \binom{N + d_{\mathrm{ES}}^2 - 1}{N}, \tag{3.48}$$

for $d_{\mathrm{ES}} = 4$ and $N = 3$ the matrix x is of size 816×816, with around 600 thousand variables: a reduction to 4% of the original 16 million. With further work, the partial trace constraints can also be efficiently written directly in this representation, leading to $(d_{\mathrm{ES}} \cdot \binom{(N-1)+d_{\mathrm{ES}}^2-1}{N-1})^2$ equality constraints, resulting in approximately 300 thousand constraints in the $d_{\mathrm{ES}} = 4$ and $N = 3$ scenario: a reduction to 28% of the original. This symmetric representation is explained in more detail in Sec. 3.7.4.

3.7.3 Sparse implementation

Note: This section addresses the sparsity heuristic originally developed for and presented in Ref. [201]. This heuristic has since been expanded into its own manuscript, which is the subject of Ch. 5.

While these are significant improvements, they still proved to be insufficient, as the dense representation of the symmetric problem involves too many variables and constraints to be solvable. As a concrete example, for the simplest case of $\boldsymbol{a} = 001$, $d_{\mathrm{E}} = 2$ and $L = N = 3$, the SDP solver SCS attempts to allocate a dense array of over 3 TB in size.

To address this, we have developed an heuristic algorithm which exploits any sparsity of the objective function (now defined by f), and—whenever possible—constructs a sparse relaxation of the original problem. This is achieved by iteratively selecting which variables of x and which of its constraints are strictly necessary to solve the problem, due to their direct or indirect influence in the objective function.

For the state and measurements considered, our technique was capable of reducing the number of variables and constraints immensely, to less than 1% of the symmetric case (see Table 3.3). This allowed us to successfully compute upper bounds for various sequences up to $N = 4$ and $d_{\mathrm{E}} = 2$ (Table 3.2). While further optimizations are still possible, we have not pursued them in this work.

N		Symmetric	Sparse	Reduction to
3	Variables	665 856	3 566	0.54%
	Constraints	295 937	2 809	0.95%
4	Variables	15 023 376	35 688	0.24%
	Constraints	10 653 697	40 441	0.38%

Table 3.3: Comparison between the number of variables and linear constraints in the symmetric problem vs. its sparse implementation, for the sequence $\boldsymbol{a} = 001$. Our heuristic algorithm achieves a vast reduction in the number of variables and constraints, while still allowing exact solutions for our problem. Note that these numbers precede any further optimization, which could remove any remaining redundant constraints or variables.

As can be observed in Table 3.3, the sparse problems can clearly be simplified further, by eliminating redundant variables and constraints. We opted for leaving such task to the numerical pre-solver, as this only took a few minutes of computing time. Solving the SDP for $N = 3$ was possible within minutes, but for $N = 4$, from 4 up to 14 hours were needed, depending on the sequence.

3.7.4 Technical details

Note: This section discusses the more technical details required for the implementation of the SDP, without which it would not have been numerically tractable. Its contents were originally included as an appendix in Ref. [201]. The details on the sparse heuristic are discussed in Ch. 5 instead.

While in principle numerically accessible, the SDP in Prob. 4 still contains too many variables to allow for the computation of upper bounds, even for low-dimensional cases. This problem can be tackled by exploiting the fact that all appearing objects can be expressed within the symmetric subspace, as well as the (potential) sparsity of the problem.

In our case, a further simplification is possible by first noticing that both F and P_N^+ are real-valued. If we additionally consider a real-valued basis for the partial trace, we conclude that, for any feasible $X_{\mathrm{A}_1^N}$, its entry-wise complex conjugate $X^*_{\mathrm{A}_1^N}$ is also feasible, while providing the same value for the objective function. Therefore, we may perform the optimization using a real-valued $X_{\mathrm{A}_1^N}$. Next, we explain how to exploit the symmetry property.

Symmetric representation

The symmetry constraint can be satisfied automatically by expressing $X_{\mathrm{A}_1^N}$ directly in terms of a basis for the symmetric subspace. This allows for a great reduction in the number of variables and linear constraints in the SDP, which is required for its efficient numerical optimization. In the following, we describe how to construct this basis for the symmetric subspace, largely based on [85], and how to adapt the remaining constraints to act directly in this symmetric representation, based on a generalization of results provided in [184] for the qubit case.

We construct an explicit basis for the relevant subspace of the problem, such that the symmetry constraints are satisfied automatically. We begin by defining a canonical basis for the symmetric subspace. With the convention $\mathbb{N} = \{1, 2, 3, \ldots\}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, for $d, n \in \mathbb{N}$, let

$$\mathfrak{T}_d^n := \left\{ (t_1, t_2, \ldots, t_d) \;\middle|\; t_i \in \mathbb{N}_0, \quad \sum_{i=1}^{d} t_i = n \right\} \tag{3.49}$$

be the set of weak integer compositions for n into exactly d parts, possibly of zero size, which we refer to as *types*. These types form a canonical labeling for the basis of the symmetric subspace. Now, defining $[d] := \{1, 2, \ldots, d\}$, let $\mathrm{u} \in [d]^n$, i.e., $\mathrm{u} = (u_1, \ldots, u_n)$ with $u_\ell \in [d]$, and $T(\mathrm{u}) = \mathrm{t}$ be the type of the vector u, where t_i counts the number of instances where $u_\ell = i$ holds. As a concrete example, if $d = 6$ and $n = 8$, we might have:

$$\mathrm{u} = (1, 4, 1, 2, 3, 2, 2, 6) \qquad \longrightarrow \qquad T(\mathrm{u}) = (2, 3, 1, 1, 0, 1). \tag{3.50}$$

In words, types count how many times a number $1, \ldots, d$ occurs in u, and therefore are invariant under permutations, i.e., $T(\mathrm{u}) = T(P_\sigma \mathrm{u})$ for any permutation operator P_σ acting on entries of u. Given a Hilbert space $\mathcal{H}$ with $d = \dim \mathcal{H}$, we may now define a basis for $\mathrm{Sym}_n(\mathcal{H})$ in terms of the types $\mathrm{t} \in \mathfrak{T}_d^n$ by constructing the non-normalized and normalized orthogonal basis vectors for the symmetric subspace [85], respectively, as

$$|\mathrm{Sym}(\mathrm{t})\rangle := \sum_{\mathrm{u}; T(\mathrm{u})=\mathrm{t}} |u_1, u_2, \cdots, u_n\rangle, \qquad |\mathrm{sym}(\mathrm{t})\rangle := (\mathrm{t})_!^{-1/2} |\mathrm{Sym}(\mathrm{t})\rangle, \tag{3.51}$$

where $(\mathrm{t})_! = \frac{(\sum_i t_i)!}{t_1! \cdot t_2! \cdots t_d!}$ denotes the multinomial coefficient for the normalization, and, for convenience in notation, we adopt the canonical basis $\{|u\rangle\}_{u=1}^d$ for each individual u_i. Equation (3.51) clearly defines a symmetric state, and thus, for any permutation operator V_σ, we have $V_\sigma |\mathrm{sym}(\mathrm{t})\rangle = |\mathrm{sym}(\mathrm{t})\rangle$, and similarly for the non-normalized $|\mathrm{Sym}(\mathrm{t})\rangle$.

From Eq. (3.31), any symmetric separable state can be written as $\sum_i p_i |\varphi_i\rangle\langle\varphi_i|^{\otimes n}$. Therefore, by expressing $|\varphi_i\rangle^{\otimes n}$ in terms of the symmetric basis in Eq. (3.51), we may write any symmetric operator by indexing the degrees of freedom of $X_{\mathrm{A}_1^N}$ by the types t, t'. In our current problem, with a local dimension $d = (d_\mathrm{E} d_\mathrm{S})^2 = d_\mathrm{ES}^2$, the symmetric space has total dimension [85]

$$D_\mathcal{S} = \dim \mathrm{Sym}_N(\mathcal{H}) = \left(\!\!\binom{d_\mathrm{ES}^2}{N}\!\!\right) = \binom{N + d_\mathrm{ES}^2 - 1}{N}, \tag{3.52}$$

with $\left(\!\binom{d}{n}\!\right)$ the multiset notation, i.e., the number of ways of picking n elements out of d, with repetitions allowed, which corresponds to the cardinality of $\mathfrak{T}_d^n$. This is a significant reduction from the original $\dim \mathrm{A}_1^N = (d_\mathrm{ES})^{2N}$.

For simplicity in notation, we consider the dependence of $D_\mathcal{S}$ on d_ES and N implicit in the following. To concretely exploit this reduction of the number of variables, i.e., in order to write $X_{\mathrm{A}_1^N}$ in terms of a smaller matrix, we re-express the problem as follows. Let $\mathcal{S}_N := \mathbb{C}^{D_\mathcal{S} \otimes D_\mathcal{S}}$, and $x \in \mathcal{S}_N$, where the mapping $X_{\mathrm{A}_1^N} \mapsto x$ is given by Eqs. (3.46) and (3.47).

We then specify the elements $x_{\mathrm{t},\mathrm{t}'}$, with $\mathrm{t}, \mathrm{t}' \in \mathfrak{T}_{d_\mathrm{ES}^2}^N$, by a canonical orthonormal basis $\{|\mathrm{t}\rangle_{\mathcal{S}_N}\}_\mathrm{t}$ over the symmetric types. Positivity of $X_{\mathrm{A}_1^N}$ can be guaranteed by requiring $x \geq 0$ directly, as x and $X_{\mathrm{A}_1^N}$ have

the same non-zero eigenvalues. With the chosen normalization of C_U, we may also write $\mathrm{Tr}\,[x] = 1$. In practice, it may be sufficient to numerically compute f directly by means of Eq. (3.46), as computing the projection of $F^{\mathrm{T}} \otimes \mathbb{1}_{\mathrm{A}_{L+1}^N}$ onto the symmetric subspace analytically can be cumbersome for more general F, and this computation only needs to be performed once before any numerical optimization.

To rewrite the partial trace constraints of the SDP directly in this symmetric basis, we must find a way to express partial traces in terms of the symmetric representation. In the following, we generalize some results known for qubits [184] to arbitrary local dimension. But first, recall that the constraint we wish to rewrite is given by

$$\mathrm{Tr}_{\mathrm{O}_1}\left[X_{\mathrm{A}_1^N}\right] = \frac{\mathbb{1}_{\mathrm{I}_1}}{d_{\mathrm{ES}}} \otimes \mathrm{Tr}_{\mathrm{A}_1}\left[X_{\mathrm{A}_1^N}\right]. \tag{3.53}$$

Treating types as ordinary vectors, we may define addition between types with the same d, such that if $\mathrm{r} \in \mathfrak{T}_d^n$ and $\mathrm{s} \in \mathfrak{T}_d^m$, then $\mathrm{r} + \mathrm{s} = \mathrm{t} \in \mathfrak{T}_d^{n+m}$. This corresponds to the fact that $\mathcal{S}_{n+m} \subset \mathcal{S}_n \otimes \mathcal{S}_m$ [184]. Generalizing upon this idea, we may split an n-partite symmetric state into m parts of sizes k_i, with $\sum_{i=1}^m k_i = n$, through the decomposition

$$|\mathrm{Sym}(\mathrm{t})\rangle_{\mathrm{A}_1^N} = \sum_{\mathrm{r}_1+\cdots+\mathrm{r}_m=\mathrm{t}} |\mathrm{Sym}(\mathrm{r}_1)\rangle_{\mathrm{B}_1} \cdots |\mathrm{Sym}(\mathrm{r}_m)\rangle_{\mathrm{B}_m}, \tag{3.54}$$

where the sum is over all tuples $(\mathrm{r}_1, \ldots, \mathrm{r}_m) \in (\mathfrak{T}_d^{k_1} \times \cdots \times \mathfrak{T}_d^{k_m})$ satisfying $\mathrm{r}_1 + \cdots + \mathrm{r}_m = \mathrm{t}$. Here, $\mathrm{B}_\ell := \mathrm{A}_{b_\ell+1}^{b_\ell+k_\ell}$, with $b_\ell = \sum_{i=1}^{\ell-1} k_i$, corresponding to the subspace of the ℓ-th part of the decomposition. Importantly, this decomposition is performed in the full space with non-normalized vectors. For completion, the normalized version of the decomposition in Eq. (3.54) is given by

$$|\mathrm{sym}(\mathrm{t})\rangle_{\mathrm{A}_1^N} = \sum_{\mathrm{r}_1+\cdots+\mathrm{r}_m=\mathrm{t}} \left(\frac{(\mathrm{t})_!}{(\mathrm{r}_1)_! \cdots (\mathrm{r}_m)_!}\right)^{-1/2} |\mathrm{sym}(\mathrm{r}_1)\rangle_{\mathrm{B}_1} \cdots |\mathrm{sym}(\mathrm{r}_m)\rangle_{\mathrm{B}_m}. \tag{3.55}$$

However, working under such normalization is cumbersome and inefficient due to the several coefficients involved. Instead, we have chosen to normalize directly in terms of the original types t, t', which makes normalization straightforward. To adapt the partial trace constraint, we first write the state $X_{\mathrm{A}_1^N}$ as in Eq. (3.46), and using Eq. (3.54) split the basis into two parts of sizes $(1, N-1)$, obtaining

$$X_{\mathrm{A}_1^N} = \sum_{\mathrm{t},\mathrm{t}'} \sum_{\mathrm{r}+\mathrm{s}=\mathrm{t}} \sum_{\mathrm{r}'+\mathrm{s}'=\mathrm{t}'} \hat{x}_{\mathrm{t},\mathrm{t}'} \, |\mathrm{Sym}(\mathrm{r})\rangle\langle\mathrm{Sym}(\mathrm{r}')|_{\mathrm{A}_1} \otimes |\mathrm{Sym}(\mathrm{s})\rangle\langle\mathrm{Sym}(\mathrm{s}')|_{\mathrm{A}_2^N}, \tag{3.56}$$

where the proper normalization is now taken care of by defining $\hat{x}_{\mathrm{t},\mathrm{t}'} := ((\mathrm{t})_!(\mathrm{t}')_!)^{-1/2} x_{\mathrm{t},\mathrm{t}'}$. Since r and r' are types for a single party, we have that $|\mathrm{Sym}(\mathrm{r})\rangle_{\mathrm{A}_1}$ are simply vectors in the canonical basis $\{|u\rangle\}_{u=1}^d$. Thus, we will adopt the notation $|u(\mathrm{r})\rangle_{\mathrm{A}_1} = |\mathrm{Sym}(\mathrm{r})\rangle_{\mathrm{A}_1}$ in what follows. The tensor product form allows for the simplified application of the partial traces in Eq. (3.53). By linearity of the partial trace, we can treat each term of Eq. (3.56) separately, so that the terms on the left (Y_L) and right (Y_R) hand sides of Eq. (3.53) can be written as

$$\begin{aligned} Y_L &:= \mathrm{Tr}_{\mathrm{O}_1}\left[|u(\mathrm{r})\rangle\langle u(\mathrm{r}')|_{\mathrm{A}_1} \otimes |\mathrm{Sym}(\mathrm{s})\rangle\langle\mathrm{Sym}(\mathrm{s}')|_{\mathrm{A}_2^N}\right] \\ &= \mathrm{Tr}_{\mathrm{O}_1}\left[|u(\mathrm{r})\rangle\langle u(\mathrm{r}')|_{\mathrm{A}_1}\right] \otimes |\mathrm{Sym}(\mathrm{s})\rangle\langle\mathrm{Sym}(\mathrm{s}')|_{\mathrm{A}_2^N} \end{aligned} \tag{3.57}$$

$$\begin{aligned} Y_R &:= \frac{\mathbb{1}_{\mathrm{I}_1}}{d_{\mathrm{ES}}} \otimes \mathrm{Tr}_{\mathrm{A}_1}\left[|u(\mathrm{r})\rangle\langle u(\mathrm{r}')|_{\mathrm{A}_1} \otimes |\mathrm{Sym}(\mathrm{s})\rangle\langle\mathrm{Sym}(\mathrm{s}')|_{\mathrm{A}_2^N}\right] \\ &= \mathrm{Tr}\left[|u(\mathrm{r})\rangle\langle u(\mathrm{r}')|_{\mathrm{A}_1}\right] \cdot \frac{\mathbb{1}_{\mathrm{I}_1}}{d_{\mathrm{ES}}} \otimes |\mathrm{Sym}(\mathrm{s})\rangle\langle\mathrm{Sym}(\mathrm{s}')|_{\mathrm{A}_2^N}. \end{aligned} \tag{3.58}$$

A few simplifications are now evident. First, we observe that $\mathrm{Tr}\left[|u(\mathrm{r})\rangle\langle u(\mathrm{r}')|_{\mathrm{A}_1}\right] = \delta_{\mathrm{r},\mathrm{r}'}$, where $\delta_{\mathrm{r},\mathrm{r}'}$ is the Kronecker delta. Secondly, we may split the states $|u(\mathrm{r})\rangle_{\mathrm{A}_1}$ into the local input and output spaces

$$|u(\mathrm{r})\rangle_{\mathrm{A}_1} = |i(\mathrm{r})\rangle_{\mathrm{I}_1} |o(\mathrm{r})\rangle_{\mathrm{O}_1}, \tag{3.59}$$

so that we obtain

$$\begin{aligned}\mathrm{Tr}_{\mathrm{O}_1}\left[|u(\mathrm{r})\rangle\langle u(\mathrm{r}')|_{\mathrm{A}_1}\right] &= \mathrm{Tr}_{\mathrm{O}_1}\left[|i(\mathrm{r})\rangle\langle i(\mathrm{r}')|_{\mathrm{I}_1}\otimes|o(\mathrm{r})\rangle\langle o(\mathrm{r}')|_{\mathrm{O}_1}\right]\\ &= |i(\mathrm{r})\rangle\langle i(\mathrm{r}')|_{\mathrm{I}_1}\cdot\mathrm{Tr}\left[|o(\mathrm{r})\rangle\langle o(\mathrm{r}')|_{\mathrm{O}_1}\right]\\ &= |i(\mathrm{r})\rangle\langle i(\mathrm{r}')|_{\mathrm{I}_1}\cdot\delta_{o(\mathrm{r}),o(\mathrm{r}')}.\end{aligned} \tag{3.60}$$

Using the above results, and correcting for the missing normalizations, we can apply the isometry $\mathbb{1}_{\mathrm{I}_1}\otimes V_{N-1}$, so that Eqs. (3.57) and (3.58) are written in terms of:

$$\begin{aligned}Y_L &\to ((\mathrm{s})_!(\mathrm{s}')_!)^{1/2}\cdot\delta_{o(\mathrm{r}),o(\mathrm{r}')}\cdot|i(\mathrm{r})\rangle\langle i(\mathrm{r}')|_{\mathrm{I}_1}\otimes|\mathrm{s}\rangle\langle\mathrm{s}'|_{\mathcal{S}_{N-1}}\\ Y_R &\to ((\mathrm{s})_!(\mathrm{s}')_!)^{1/2}\cdot\frac{\delta_{\mathrm{r},\mathrm{r}'}}{d_{\mathrm{ES}}}\cdot\sum_{i=1}^{d_{\mathrm{ES}}}|i\rangle\langle i|_{\mathrm{I}_1}\otimes|\mathrm{s}\rangle\langle\mathrm{s}'|_{\mathcal{S}_{N-1}}.\end{aligned} \tag{3.61}$$

Here, we have written $\mathbb{1}_{\mathrm{I}_1}=\sum_{i=1}^{d_{\mathrm{ES}}}|i\rangle\langle i|_{\mathrm{I}_1}$ as to make the $\mathrm{I}_1\otimes\mathcal{S}_{N-1}$ decomposition explicit in both expressions. By inserting Eq. (3.61) back into the sum of Eq. (3.56), we can appreciate the fact that Eq. (3.61) neatly separates each term of the sum as square matrices of size $d_{\mathrm{ES}}\times\left(\!\!\binom{d_{\mathrm{ES}}^2}{N-1}\!\!\right)$, written in terms of $|i\rangle\langle i'|_{\mathrm{I}_1}\otimes|\mathrm{s}\rangle\langle\mathrm{s}'|_{\mathcal{S}_{N-1}}$. The constraint of Eq. (3.53) then tells us we must sum over all t,t' on both sides, where we can then apply the equality constraint element-wise by simply matching the resulting $(i,\mathrm{s},i',\mathrm{s}')$ entries. As the extra normalization $((\mathrm{s})_!(\mathrm{s}')_!)^{1/2}$ of Eq. (3.61) is always equal between these elements, it cancels out in their element-wise equality constraint. Thus, it is sufficient to use the normalization of $\hat{x}_{\mathrm{t},\mathrm{t}'}$.

During implementation in software, the summations in Eq. (3.56) can be efficiently performed by taking into account the fact that valid decompositions of the form $\mathrm{t}=\mathrm{r}+\mathrm{s}$ are very restricted in number, and can be efficiently enumerated, grouped, filtered and counted. By assigning a tuple (i,o,s) to each t, the tracing over inputs and output spaces, and applications of the Kronecker deltas, can thus be efficiently computed. Therefore, the linear constraints between variables $x_{\mathrm{t},\mathrm{t}'}$ can be constructed by grouping terms $x_{\mathrm{t},\mathrm{t}'}$ by matching $|i\rangle\langle i'|_{\mathrm{I}_1}\otimes|\mathrm{s}\rangle\langle\mathrm{s}'|_{\mathcal{S}_{N-1}}$, and normalization is simplified by performing it with respect to $x_{\mathrm{t},\mathrm{t}'}$ at the very end, i.e., using $\hat{x}_{\mathrm{t},\mathrm{t}'}$.

A similar strategy as above may be employed to define partial transpositions through the symmetric representation, as was done in [184] for the qubit case, in order to implement the PPT condition. However, for $d_{\mathrm{E}}=2$ in our case, this approach would render the matrices and the number of constraints too large for a viable computation, once again, and therefore we did not pursue a generalization of this idea. The optimizations we have performed for $d_{\mathrm{E}}=1$ involving PPT constraints, as shown in Table 3.1, used $X_{\mathrm{A}_1^N}$ directly following Eq. (3.46), which was still tractable in this case.

While the above steps lead to a significantly smaller representation of the problem, it is not generally a sufficient reduction in variables and constraints for the SDP to be numerically tractable. For concreteness, in the smallest nontrivial case of $d_{\mathrm{E}}=d_{\mathrm{S}}=2$, this gives for $N=2,\ldots,5$ a symmetric space of size $D_{\mathcal{S}}=136,\ 816,\ 3\,876,\ 15\,504,\ \ldots$, such that the SDP would be written in terms of $(D_{\mathcal{S}})^2=18\,496,\ 665\,856,\ 15\,023\,376,\ 240\,374\,016,\ \ldots$ variables. These examples illustrate how quickly our problem can become numerically intractable, even with symmetry taken into account.

3.8 Discussion

3.8.1 Repeated unitaries

In the following, we discuss in which cases the condition of repeated unitaries is physically justified. To see this, imagine that the evolution of the system is governed by a time-independent Hamiltonian H_{SE}. The corresponding unitaries will be of the form $U(t)=\mathrm{e}^{-\mathrm{i}Ht}$. This assumption of time-independence is always possible, as the environment may include any source of time dependence. Ideally, then, we would

have the same unitary if we perform the measure-and-prepare operations equally spaced in time, say, by an interval Δt. One should take into account, however, some uncertainty in the time measurement. Let us then assume that our choice of time for performing a measurement is distributed according to a distribution $q(t)$, centered around Δt. This means that instead of transforming our system according to the unitary $U(\Delta t)$, we transform it, on average, according to the mapping $\Lambda := \int \mathcal{U}(t)q(t)dt$, seemingly breaking our assumption of repeated unitaries. Since Λ is a valid completely positive trace preserving (CPTP) map, it can be dilated by means of a larger environment into a unitary $\mathcal{U}'$, namely $\Lambda(\rho_{\mathrm{SE}}) = \mathrm{Tr}_{\mathrm{E}_1}\big[\mathcal{U}'(\rho_{\mathrm{SE}} \otimes |0\rangle\langle 0|_{\mathrm{E}_1})\big]$. To complete the argument that this situation can still be described by repeated unitaries, it is enough to show that the same can be done for multiple copies of it, namely, that the repeated operation Λ^L can be dilated to some unitary $\widetilde{\mathcal{U}}^L$. To do so, it is enough to provide at each time-step i a new environment E_i prepared in the correct initial state $|0\rangle\langle 0|_{\mathrm{E}_i}$. Let us compute explicitly the case for two time steps, the general case is straightforward. We define $W_{\mathrm{E}_1\mathrm{E}_2}$ the swap operator between systems E_1 and E_2, and we set $\widetilde{\mathcal{U}} := W_{\mathrm{E}_1\mathrm{E}_2} \circ \mathcal{U}'_{\mathrm{SEE}_1}$. Let us calculate, for simplicity

$$\begin{aligned}
&\mathrm{Tr}_{\mathrm{E}_1\mathrm{E}_2}\big[\mathcal{U}'_{\mathrm{SEE}_1} \circ W_{\mathrm{E}_1\mathrm{E}_2} \circ \mathcal{U}'_{\mathrm{SEE}_1}(\rho_{\mathrm{SE}} \otimes |0\rangle\langle 0|_{\mathrm{E}_1} \otimes |0\rangle\langle 0|_{\mathrm{E}_2})\big] \\
&\quad = \mathrm{Tr}_{\mathrm{E}_1\mathrm{E}_2}\big[\mathcal{U}'_{\mathrm{SEE}_1} \circ W_{\mathrm{E}_1\mathrm{E}_2}(\sigma_{\mathrm{SEE}_1} \otimes |0\rangle\langle 0|_{\mathrm{E}_2})\big] \\
&\quad = \mathrm{Tr}_{\mathrm{E}_1}\big[\mathcal{U}'_{\mathrm{SEE}_1}(\mathrm{Tr}_{\mathrm{E}_2}\big[\sigma_{\mathrm{SEE}_2}\big] \otimes |0\rangle\langle 0|_{\mathrm{E}_1})\big] \\
&\quad = \mathrm{Tr}_{\mathrm{E}_1}\big[\mathcal{U}'_{\mathrm{SEE}_1}(\Lambda(\rho_{\mathrm{SE}}) \otimes |0\rangle\langle 0|_{\mathrm{E}_1})\big] = \Lambda^2(\rho_{\mathrm{SE}}),
\end{aligned} \tag{3.62}$$

where $\sigma_{\mathrm{SEE}_1} := \mathcal{U}'_{\mathrm{SEE}_1}(\rho_{\mathrm{SE}} \otimes |0\rangle\langle 0|_{\mathrm{E}_1})$. The effect of the final $W_{\mathrm{E}_1\mathrm{E}_2}$ operation is to swap the space E_1 and E_2 that are irrelevant after the operations on SE have been performed. Evidently, this argument can straightforwardly be extended to more time steps. In summary, this means that the condition of "the same unitary" is satisfied as long as the time choice is always drawn from the same distribution. In other words, it is sufficient to always "probabilistically repeat" the same operation.

Finally, we remark that even if this procedure may use a very large environment, we are only interested in showing that we can always assume there *is* a unitary dynamics, with the notion of an effective environment, then, taking care of estimating the environment consistent with the observed statistics. Moreover, we also remark that time-independent operations are necessary, as unrestricted time-dependent operations can achieve arbitrarily long temporal correlations even with bounded memory size [124].

3.8.2 Effective environment and initial state

The environment of a quantum system, intended as all the physical systems surrounding and possibly interacting with it, is typically a very high-dimensional system, if not directly assumed to be infinite-dimensional. From this perspective, we want to make sense of the notion of *effective* environment. Consider a global transformation of the system and environment. As a first approximation, we can say that the global unitary is of the form $U_{\mathrm{SE}} \otimes U_{\mathrm{E}'}$, where U_{SE} is an entangling unitary between the system and the effective environment and $U_{\mathrm{E}'}$ is acting on the rest of the environment.

At the same time, an evolution of the form $U_{\mathrm{SE}} \otimes U_{\mathrm{E}'}$ is just an approximation of the full evolution of the environment, as we expect its state to thermalize after some time interval. Nevertheless, from a physical perspective this approximation is still valid if the time required for a single run of the experiment, i.e., the measurement of the temporal sequence, is much shorter than the time needed for the effective environment to thermalize. This is to be expected, as the environment is composed of many particles that may interact with each other with different strengths. Underlying this expectation is the assumption — known as Markovian embedding [30, 213, 214] and frequently employed in the description and modeling of open quantum system dynamics [187, 120, 121, 122] — that the environment can always be split into two parts: a far environment, leading to irretrievable, memoryless information loss, and an effective environment, that can transport memory. Due to the irretrievable information loss, the dynamics of the

system and the effective environment is then described by a general (non-unitary) CPTP map, a situation which, as discussed in the previous section, can again be dilated to repeated unitaries.

Reversing this perspective, namely, looking at the problem of characterizing the effective environment from temporal correlation experiments, we may say that this characterization is inherently dependent on the typical time-scales of a single experimental run. This difference in time scales and the eventual thermalization of the environment, however, are essential for repeating the experiment for collecting statistics, as it is required that the initial state of the environment is always (probabilistically) described by the same state ρ_{E}^0 in each experimental run. As previously mentioned, this is typically a thermal state, but it does not need to be characterized in our approach. In fact, since the SDP maximizes the temporal correlations, which are linear in the initial state of the environment, we know that the maximum is always achieved with a pure state. Up to local unitaries, we can thus always assume it to be $\rho_{\mathrm{E}}^0 = |0\rangle\langle 0|_E$. The bound calculated for this state is, then, valid for any possible initial state of the environment.

Finally, somewhat independent of the explicit experimental situation, we may see our setup as a question of simulation resources: What is the smallest dimension an environment coupling to a known system must have in order to reproduce observed statistics in a unitary way? Seen in this way, the results we present attribute a "simulation hardness" (in the sense of required environment dimension) to each sequence, that is agnostic with respect to concrete time scales or experimental limitations, but rather inherent to the respective sequence.

3.8.3 Conditions for deterministic realizations

The maximum probability for a given sequence increases as more memory is available, i.e.,

$$0 \leq \Omega(\boldsymbol{a}, 1) \leq \Omega(\boldsymbol{a}, 2) \leq \cdots \leq \Omega(\boldsymbol{a}, d_{\mathrm{E}}) \leq 1, \tag{3.63}$$

as larger environments can always simulate the dynamics of smaller ones. Therefore, if the sequence $\boldsymbol{a}$ is sufficiently simple, its maximum may be trivial for a given d_{E}, i.e., $\Omega(\boldsymbol{a}, d_{\mathrm{E}}) = 1$.

The maxima $\Omega(\boldsymbol{a}, d_{\mathrm{E}})$ depend on the sequence, as not all sequences are equally "difficult" to produce with a given amount of memory d_{E}. For example, it is easy to see that the sequence 000 can *always* be produced with unit probability, while the sequence 001 cannot be produced with unit probability if the environment is not of sufficient size [198]. This observation suggests a relevant notion of "complexity" of sequences, which offers a more fundamental relationship between $\boldsymbol{a}$ and d_{E}. Since our goal is to establish nontrivial maxima on the probabilities of sequences, the natural question to ask is how large should d_{E} be such that $\boldsymbol{a}$ can occur deterministically?

Understanding the conditions where this occurs allows us to pick only scenarios featuring non-trivial maxima. In Ref. [198], the notion of deterministic complexity (DC) of a sequence was introduced to address this question, and was discussed in detail in Sec. 2.5.1. Previously, the DC was applied to the scenario involving repeated measurements on a single isolated system, which is a deviation from the scenario we are currently considering. Nevertheless, since $\mathrm{DC}(\boldsymbol{a})$ is a property of the sequences $\boldsymbol{a}$, and is independent of the model (quantum or classical), this notion is also relevant here, albeit with some modifications due to the multipartite nature of the current problem.

First, we establish a correspondence between the single system and open system scenarios. Let $d_{\mathrm{S}} = |\mathcal{A}|$, so that we may choose $E_S^a = |a\rangle\langle a|_S$ and $\rho_{\mathrm{S}}^0 = |0\rangle\langle 0|_S$. Then, the changes in the environment state due to the unitary and measurements can be written succinctly in terms of the completely positive trace non-increasing maps

$$\mathcal{I}_a(\rho_{\mathrm{E}}) = \mathrm{Tr}_S\left[U(\rho_{\mathrm{E}} \otimes |0\rangle\langle 0|_S)U^\dagger \cdot \mathbb{1}_E \otimes |a\rangle\langle a|_S\right]. \tag{3.64}$$

Therefore, under the action of $(\mathcal{I}_a)_{a\in\mathcal{A}}$, we may interpret the *environment* as a single isolated "memory" system in which the maps $\mathcal{I}_a$ act sequentially, as in Ref. [198] and Ch. 2. The converse mapping, from the

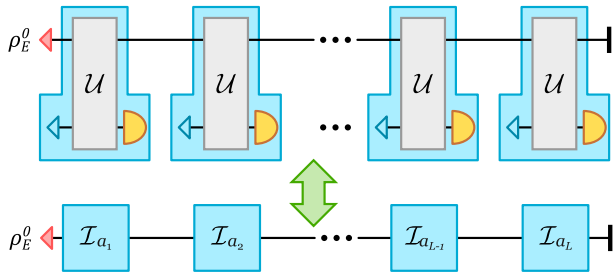

Figure 3.5: The open system and environment scenario can be equivalently interpreted as the dilation of the scenario involving sequential measurements $\mathcal{I}_a$ on an isolated environment system.

sequential measurements on an isolated memory system to the open system scenario, is the dilation map with the probe system acting as the ancilla and the environment as the memory. This construction is well known, but it is worth expanding it in detail in order to understand what are the conditions on the ancilla (i.e., the probe system) needed to implement it.

Let $n(\boldsymbol{a})$ denote the number of unique symbols appearing in $\boldsymbol{a}$, with $n(\boldsymbol{a}) \leq |\mathcal{A}|$, and let us assume $d_{\mathrm{S}} \geq n(\boldsymbol{a})$. Now, consider an instrument given by the maps $\tilde{\mathcal{I}}_a$ of the form $\tilde{\mathcal{I}}_a(\rho_{\mathrm{E}}) = K_a \rho_{\mathrm{E}} K_a^\dagger$, for $(K_a)_{a\in\mathcal{A}}$ Kraus operators, i.e., $\sum_{a\in\mathcal{A}} K_a^\dagger K_a = \mathbb{1}_E$. The corresponding unitary arises from the $d_{\mathrm{E}} d_{\mathrm{S}} \times d_{\mathrm{E}}$ isometry matrix

$$Q = \sum_{a\in\mathcal{A}} K_a \otimes |a\rangle_S \,, \tag{3.65}$$

which can always be completed into a unitary matrix U. Thus $\tilde{\mathcal{I}}_a$ can be written in the form of Eq. (3.64). As we have chosen $d_{\mathrm{S}} = |\mathcal{A}|$, $\rho_{\mathrm{S}}^0 = |0\rangle\langle 0|_S$ and $E_S^a = |a\rangle\langle a|_S$ in our implementation of the SDP (Sec. 3.7), there is a direct correspondence between the two scenarios; see Fig. 3.5. Therefore, the upper bounds obtained by the SDP in Eq. (3.16) can be compared with the known achievable values from [198] (Sec. 2.8), as shown in Table 3.2 and discussed in Sec. 3.6.

This construction applies to deterministic models as they have deterministic transitions between states, i.e., each is described by a single Kraus operator per outcome. We thus have

Proposition 2. If both $d_{\mathrm{E}} \geq \mathrm{DC}(\boldsymbol{a})$ and $d_{\mathrm{S}} \geq n(\boldsymbol{a})$, then there is a choice of probe state and measurements such that $\Omega(\boldsymbol{a}, d_{\mathrm{E}}) = 1$.

Note that the condition $d_{\mathrm{S}} \geq n(\boldsymbol{a})$ is strictly required for deterministic production of $\boldsymbol{a}$. In fact, if $p(\boldsymbol{a}|d_{\mathrm{E}}) = 1$, every symbol must occur deterministically. Then, at each step the system must be in a state ρ_{S}^a such that $\mathrm{Tr}\left[\rho_{\mathrm{S}}^a E_S^a\right] = 1$. Thus, the measurements E_S^a must be able to perfectly discriminate between the states $\{\rho_{\mathrm{S}}^a\}_a$, which is possible only if $d_{\mathrm{S}} \geq n(\boldsymbol{a})$. We have established:

Proposition 3. If $d_{\mathrm{S}} < n(\boldsymbol{a})$, then $\Omega(\boldsymbol{a}, d_{\mathrm{E}}) < 1$ for any choice of probe state or measurements, and any environment dimension d_{E}.

In conclusion, these results, together with those presented in Ref. [198] and Ch. 2, tell us that nontrivial bounds appear for $d_{\mathrm{E}} < \mathrm{DC}(\boldsymbol{a})$, that we can compare these bound with the single-system scenario for $d_{\mathrm{S}} \geq n(\boldsymbol{a})$, and that $|\mathcal{A}| = 2$ and $L = 3$ is the smallest scenario displaying nontrivial memory effects.

3.8.4 Choice of sequence

As briefly noted in Section 3.8.3, in order to construct a witness for $d_{\mathrm{E}} > d$, it is vital to choose a sequence $\boldsymbol{a}$ with sufficient deterministic complexity. If the sequence chosen is too short or too simple (i.e., if $\mathrm{DC}(\boldsymbol{a}) \leq d$), then the bound $\widehat{\omega}(\boldsymbol{a}, d)$ is trivial and the dimension d_{E} cannot be tested with such sequence.

However, due to the difficulty in computing these bounds, and the fact longer sequences generally become less probable—and therefore less likely to violate the witness inequality—generally one should choose sequences to be as short as possible while having deterministic complexity of the same size as the minimum environment dimension one wishes to witness.

As a concrete example, if we wish to witness $d_{\mathrm{E}} > 3$, we should choose a sequence $\boldsymbol{a}$ with deterministic complexity $\mathrm{DC}(\boldsymbol{a}) = 4$, e.g., $\boldsymbol{a} = 0001$, which is the shortest sequence satisfying this requirement. Then, a violation of the bound $\widehat{\omega}(0001, 3)$ informs us that, in fact, $d_{\mathrm{E}} \geq 4$.

3.9 Conclusions and outlook

We presented a method to lower bound the dimension of the environment interacting with a probe system, based only on the statistics of measurements performed on the (partially characterized) system, and without any assumption on the environment or the dynamics. This is achieved via a hierarchy of semidefinite programs that upper bound temporal correlations achievable in various experimental scenarios, under the assumption of finite memory. Such bounds can be applied to the detection of the effective environment size in the dynamics of open systems, as well as a certification of the minimum size of an environment's dimension compatible with observations.

To keep the discussion simpler, we applied the optimization for a single sequence. It is straightforward to adapt the objective function to arbitrary linear functions of the full probability distribution $(p(\boldsymbol{a}|d_{\mathrm{E}}))_{\boldsymbol{a}}$, as those appearing in the temporal inequalities derived in, e.g., [88, 32, 182, 124]. This may, in principle, lead to better witnesses. We leave the numerical explorations of this problem to a future investigation.

We assumed a joint unitary evolution between system and environment, which leads to a CJ representation of these maps given in terms of symmetric states. As explained in Sec. 3.8.1, this assumption can usually be justified on physical grounds. Nevertheless, a natural question to ask is how do our results change if we consider arbitrary CPTP maps? In that case, the SDP should be modified by replacing the symmetry constraint with permutation invariance, i.e., $V_\sigma X_{\mathrm{A}_1^N} V_\sigma^\dagger = X_{\mathrm{A}_1^N}$ for all permutations $\sigma \in \mathfrak{S}_N$; see Eq. (3.27). We expect that bounds for CPTP maps may be larger than those for unitary channels of the same dimension, especially for objective functions involving more than a single sequence, but numerical optimization is significantly more costly in this case, as the permutation invariant operator basis is of size [206, Ch. 7] $\binom{N+d_{\mathrm{ES}}^4-1}{N}$, in contrast to $\binom{N+d_{\mathrm{ES}}^2-1}{N}^2$ in the symmetric case. While it is also possible to write such CPTP maps in terms a dilation of the environment, this approach would likely result in more extraneous variables in the SDP, i.e., the terms in the subspace orthogonal to the dilation ancilla's initial state, possibly rendering the optimization intractable.

Additionally, the SDP in Eq. (3.16) and its related implementation techniques are, in fact, quite general, and can be applied to a wide range of scenarios beyond what we have considered here. As the operator F from Eq. (3.11) is fixed, any choice of intermediate operations between each unitary could be chosen. E.g., the initial system and environment states could be correlated, and the intermediate maps $\mathcal{M}_a$ could each be replaced by arbitrary joint operations, even time-dependent ones. Such approaches could then, for example, be used to bound observables for specific types of processes, e.g., processes with only classical memory [72], by assuming a completely dephasing channel on the environment; see also Ch. 4. Therefore, provided the problem is numerically tractable, our techniques are independent of what explicit measurements are chosen.

While the high dimensionality of the current formulation of the SDP quickly renders general numerical implementations intractable, our approach still offers new avenues for the subject of bounding temporal correlations, and their relationship to open-system dynamics. Ultimately, the techniques developed herein should be taken as a proof-of-concept for future developments and improvements. It remains to be seen whether more efficient numerical techniques, or even alternative outer approximations, are better suited

for addressing such problems. Further investigation of these avenues is subject to future work.

Nevertheless, the success of our approach highlights the wealth of information contained in temporal correlations and the potential of new techniques for characterizing large complex systems by means of a small probe alone, by exploiting nontrivial properties of the temporal correlations achieved by systems of bounded size.

Chapter 4

Entanglement-breaking channels are a quantum memory resource

The main results in this chapter were included in the manuscript Ref. [200], awaiting peer review:

Entanglement-breaking channels are a quantum memory resource
Lucas B. Vieira, Huan-Yu Ku, Costantino Budroni
arXiv — Submitted 2024-02-06

Author contribution: In this work, the doctoral candidate significantly contributed to the conception of the research topic. The candidate was solely responsible for writing all the gradient descent optimization code, the equiangular tight frame construction, and the production of all graphical assets. The author contributed significantly to the analysis and interpretation of the results, and was responsible for writing the majority of the manuscript.

The contents of this chapter were adapted from the original text. Additional previously unpublished results have also been included, and will be noted when they appear.

4.1 Introduction

Any information processing task consists of manipulating information stored in a memory over time. This is achieved by sequential operations performed on the memory, with classical and quantum systems offering vastly different capabilities on what memory states and their transformation are possible. Since a finite amount of memory imposes fundamental limitations on which tasks can be achieved, investigating these limits not only provides insights into foundational questions in physics, but also establishes physical constraints to what can be achieved with a certain amount (and type) of memory resource.

In this context, the notion of classical memory, in the sense of storage of information, is familiar to us from our everyday life. When we talk about the number of bits a device has, we refer to the logarithm of the number of distinguishable internal states, where information can be stored and retrieved. This quantity represents the number of logical memory units available to our control, each realized through a concrete physical implementation, providing a direct connection between the abstract information resources and the underlying physical ones.

In contrast, our understanding of quantum memories is still incomplete. With the advent of quantum information and quantum computation, distinguishing between classical and quantum memory effects is of utmost importance for identifying scenarios where quantum systems offer an advantage over classical ones [165, 72]. An approach towards this goal is by establishing bounds on information processing

tasks achievable with classical memory alone. A violation of such bounds then provides a certificate of non-classicality, often also providing a quantifiable metric for the *quantumness* of the underlying memory resource [67, 32, 134, 33, 217, 106, 2, 40]. In recent years, the formalism of quantum supermaps [41, 156, 155], also known as process tensors or quantum combs, has enabled a thorough investigation of temporal quantum protocols, as it is naturally well-suited for the study of multi-time statistics [134, 14, 190]. Within this framework, many quantum-inspired notions have been employed as conflicting *definitions* of "classical memory" [165, 14, 190, 72]; see Ref. [190] for a detailed characterization.

A typical approach involves entanglement-breaking (EB) channels, which are known to be exactly equivalent to measure-and-prepare (M&P) operations [91]. As the input quantum state is destroyed by the measurement operation and replaced by a classical label (albeit encoded in the prepared output quantum state), one may be inclined to claim that such channels are essentially equivalent to classical channels [165, 217, 106, 2, 14], i.e., that they can only establish classical correlations between their inputs and outputs. EB channels have been used in this manner, e.g., to develop resource theories of quantum memories [165], the benchmarking and application of quantum devices [217], characterizing non-classical memory effects in quantum processes [72, 135, 14], bosonic Gaussian channels [93, 2], and in establishing a hierarchy between competing notions of classical memory [190]. Figure 4.1 depicts such an application of EB channels in the context of quantum supermaps, as seen, e.g., in Refs. [135, 72, 14, 190].

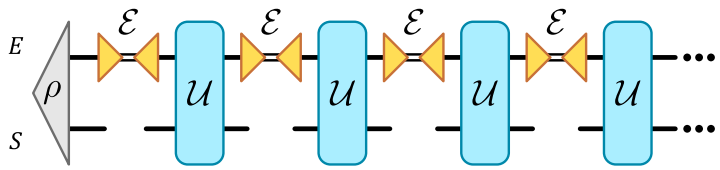

Figure 4.1: A typical application of entanglement-breaking channels in the context of quantum supermaps. The channel $\mathcal{E}$ is of measure-and-prepare form, thus breaking the entanglement between system (S) and environment (E). Since information is entirely relayed through the intermediate classical space (double lines) of $\mathcal{E}$, it is often assumed that the temporal correlations obtained from the resulting supermap can be considered classical. In this chapter, we show that this interpretation can be misleading.

A potential issue with such entanglement-inspired definitions is that they still operate within the quantum formalism, and therefore, may not truly characterize the limitations inherent to classical systems. Clearly, a precise distinction between classical and quantum memory effects must emerge from explicit constructions within each theory.

In this chapter, we show that the previous interpretation and applications of EB channels can be misleading in the case of temporal scenarios involving sequential measurements. We address the following question: is the memory of a "d-it"[1] (a classical system with d states) equivalent to that of a *qudit* (d-level quantum system) passing through an EB channel? We answer it in the negative by presenting explicit examples where a qudit passing through an EB channel outperforms a d-it of classical memory in the task of generating correlations in time. We conclude that EB channels cannot be used to define or characterize classical memories in the most general settings.

The chapter is organized as follows. Section 4.3 covers preliminary notions required to understand the results of the chapter. Section 4.4 introduces the sequential measurement protocol covered in this chapter, similar to that of Ch. 2, which enables the comparison between classical and quantum memory effects in the presence of EB channels. Section 4.5 discusses an explicit construction showing a quantum advantage even in the presence of EB channels, with Sec. 4.5.2 discussing further improvements on this advantage. In light of these results, in Sec. 4.6 we discuss some conditions where quantum advantages

[1] We use "d-it", analogous to "qudit", as the term *dit* is already standard for "decimal digit", i.e., $d = 10$.

actually disappear, while Sec. 4.5.3 explores how some degree of classicality is still apparent despite these advantages. Finally, Sec. 4.7 provides a discussion of the results and future outlook.

4.2 Summary of main results

We investigate a common claim that EB channels act as classical memory resources, by choosing an information processing task where classical and quantum memories can be compared on equal terms: the generation of temporal correlations through sequential measurements.

As has been done throughout this thesis, we achieved this through the framework of finite-state machines. Building upon earlier results from Refs. [33, 198], and exact maxima Ω_C obtained with techniques developed in Ref. [207], we are able to perfectly distinguish between classical and nonclassical temporal correlations involving certain one-tick sequences.

Since EB channels are equivalent to M&P channels, described in terms of m intermediate classical labels, a generic d-dimension EB channel may involve $m > d$ intermediate classical labels. We show that if $m > d$ and $d > 2$, then it is possible to construct an EB channel and a quantum instrument such that nonclassical correlations remain even in the presence of EB channels. This is shown by an explicit quantum model, based on a construction using harmonic equiangular tight frames, as well as explicit models obtained numerically via gradient descent.

Our results show that a very common application of EB channels to the definition of classical memories within quantum supermaps is flawed, and is based on an improper analogy between temporal and spatial correlations. Furthermore, these results highlight the potential of the FSM framework in uncovering novel (and subtle) distinctions between classical and nonclassical temporal correlations.

4.3 Preliminary notions

A channel $\mathcal{E}_A$ is said to be entanglement breaking if $(\mathcal{E}_A \otimes \mathrm{id}_B)(\rho_{AB})$ is separable for any bipartite state ρ_{AB}. A well-known result establishes that any EB channel $\mathcal{E}$ can be written as a measure-and-prepare (M&P) channel in the *Holevo form* [89, 91]

$$\mathcal{E}(\rho) = \sum_{i=0}^{m} \mathrm{Tr}\left[\rho E_i\right] \sigma_i, \tag{4.1}$$

for $\{E_i\}_{i=1}^m$ a Positive Operator-Valued Measure (POVM), i.e., $E_i \geq 0$ and $\sum_{i=1}^m E_i = \mathbb{1}$, and σ_i quantum states. Here, m can be interpreted as the number of labels—in principle, arbitrarily large—required in the channel's intermediate classical memory in order to describe its operation [91, 165]. EB channels can be fully characterized [91] by the fact their Choi matrix [94, 46], $C_{\mathcal{E}} := (\mathcal{E} \otimes \mathrm{id})(|\Phi\rangle\langle\Phi|)$ with $|\Phi\rangle$ a maximally entangled state, is separable and with rank $r \geq d$, which also implies any such channel can be rewritten with at most $m = d^2$ operators E_i and σ_i. These channels can be understood in terms of a quantum-classical-quantum operation:

1. The input quantum state ρ is first measured, producing a classical outcome label i, out of a total m, with probability $\mathrm{Tr}\left[\rho E_i\right]$,

2. Conditioned on i, the corresponding quantum state σ_i is passed forward as output; see Fig. 4.3b.

While these channels are still quantum operations, the input-output transmission of the classical label can be understood as an intermediate classical channel, suggesting that EB (M&P) channels should behave in a classical manner.

To properly distinguish between classical and quantum memories, we investigate their capabilities when generating temporal correlations in the simplest scenario, i.e., that of a single, finite-dimensional

system subject to the same dichotomous measurement at multiple times, as shown in Fig. 4.2. This is similar to how we have done in Ch. 2, where we employ the formalism of finite-state machines [32, 198] to describe the scenario.

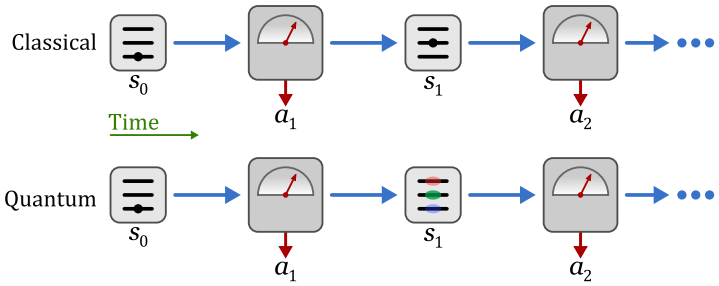

Figure 4.2: Temporal correlations can be investigated by a classical or quantum system with d internal states, in the initial state s_0, which is repeatedly measured by the same instrument at multiple times. The measurement of state s_t produces the outcome a_{t+1} and changes the state to s_{t+1}.

In the classical case, the most general description of its behavior is by an initial stochastic vector π over its d distinct states (i.e., a d-it of memory), and a pair of (row) sub-stochastic transition matrices $T = (T_0, T_1)$ describing its dynamics upon each outcome $a \in \{0, 1\}$, with $T_0 + T_1$ row stochastic. The probability of a length-L sequence of outcomes $\boldsymbol{a} = a_1 a_2 \cdots a_L$ is then given by

$$p(\boldsymbol{a}|T, d) = \pi T_{a_1} \cdots T_{a_L} \eta, \tag{4.2}$$

with $\eta = (1, \ldots, 1)^{\mathrm{T}}$. The quantum case is defined analogously, with a single d-dimensional quantum system (a qudit), initially at state ρ_0, being repeatedly measured by a quantum instrument $\mathcal{I} = (\mathcal{I}_0, \mathcal{I}_1)$, with effects $\mathcal{I}_a$ being completely positive (CP) trace non-increasing maps such that $\mathcal{I}_0 + \mathcal{I}_1$ is completely positive trace-preserving (CPTP). The probability of a sequence $\boldsymbol{a}$ is given by

$$p(\boldsymbol{a}|\mathcal{I}, d) = \mathrm{Tr}\left[\mathcal{I}_{a_L} \circ \cdots \circ \mathcal{I}_{a_1}(\rho_0)\right]. \tag{4.3}$$

In both cases, the system's state is the sole resource available to generate the probability distributions, and any dynamics is entirely governed by the effects associated with an outcome a_i. In this way, classical and quantum memories are placed on equal footing as a memory resource. Crucially, memory is the central resource for this task, as any arbitrary temporal correlation can be generated by a classical system with enough memory, i.e., enough internal states [67, 52, 182]. Consequently, the system's dimension establishes fundamental limitations on the set of achievable temporal correlations [88, 32, 182, 33, 198], providing a clear notion for distinguishing between *manifestly* classical and quantum memories. Note that we do not allow for external storage of information (e.g., by conditioning on past outcomes, or invoking an external reference clock), so that the total amount of memory can be quantified by the number of states d of the system.

One possible approach to distinguish between classical and quantum memory effects is by establishing upper bounds on the maximum probability for individual sequences to be produced within classical and quantum mechanics [32, 33, 198]. The notion of *deterministic complexity* of a sequence $\boldsymbol{a}$, denoted by $\mathrm{DC}(\boldsymbol{a})$, was introduced in Ref. [198] (see Sec. 2.5.1) as the minimum dimension d for a system (classical or quantum) to be capable of generating $\boldsymbol{a}$ with probability one. Thus, if $d < \mathrm{DC}(\boldsymbol{a})$, then *no* classical or quantum system can produce the sequence $\boldsymbol{a}$ deterministically, implying both classical and quantum theory must obey nontrivial upper bounds for this maximum probability. Since $d = 2$ is the smallest nontrivial amount of memory, these observations also highlight the importance of studying temporal correlations over three or more time steps, as every sequence of length two can be produced deterministically with either a bit or qubit.

As the underlying memory resource can be explicitly confined within each theory, such non-trivial upper bounds provide a sharp demarcation between classical and quantum memories and their capabilities, but they are generally difficult to obtain precisely [198, 207, 201]. In Ref. [198], a universal upper bound of $1/\mathrm{e}$ was conjectured for all sub-deterministic classical memory scenarios (Sec. 2.7.3), whereas no nontrivial universal quantum bound seems to exist (Sec. 2.8.3). As it turns out, the universal classical bound still seems to play a role even in the quantum case involving EB channels, as will be discussed in Sec. 4.5.3.

4.4 The sequential measurement protocol

We investigate the non-classicality of entanglement-breaking channels by inserting an EB channel $\mathcal{E}$, as in Eq. (4.1), before each quantum measurement in Eq. (4.3) (depicted in Fig. 4.3a), obtaining

$$\begin{aligned} p(\boldsymbol{a}|\mathcal{I},\mathcal{E},d) &= \mathrm{Tr}\left[\mathcal{I}_{a_L}\circ\mathcal{E}\circ\cdots\circ\mathcal{I}_{a_2}\circ\mathcal{E}\circ\mathcal{I}_{a_1}\circ\mathcal{E}(\rho_0)\right] \\ &= \sum_{i_1\ldots i_L=1}^{m}\mathrm{Tr}\left[\rho_0 E_{i_1}\right]\mathrm{Tr}\left[\mathcal{I}_{a_1}(\sigma_{i_1})E_{i_2}\right]\cdots\mathrm{Tr}\left[\mathcal{I}_{a_L}(\sigma_{i_L})\mathbb{1}\right] \\ &= \sum_{i_1\ldots i_{L+1}=1}^{m}\pi_{i_1}[T_{a_1}]_{i_1 i_2}\cdots[T_{a_L}]_{i_L i_{L+1}}\eta_{i_{L+1}}, \end{aligned} \tag{4.4}$$

where we use $[T_a]_{ij} := \mathrm{Tr}\left[\mathcal{I}_a(\sigma_i)E_j\right]$, $\pi_i := \mathrm{Tr}\left[\rho_0 E_i\right]$, and $\eta = (1,\ldots,1)^{\mathrm{T}}$ as in Eq. (4.2). Note that, since the E_i are POVM elements, each trace term denoted a probability. We have thus collected all such probabilities in the matrices T_a such that

$$\sum_{a,j}[T_a]_{ij} = \mathrm{Tr}\left[\textstyle\sum_a\mathcal{I}_a(\sigma_i)\cdot\sum_j E_j\right] = \mathrm{Tr}\left[\hat{\sigma}_i\mathbb{1}\right] = 1, \tag{4.5}$$

where we use the fact E_i form a POVM and $\sum_a\mathcal{I}_a$ is CPTP, thus $\hat{\sigma}_i = \sum_a\mathcal{I}_a(\sigma_i)$ has unit trace. Furthermore, we can define an initial probability vector π arising from the first application of the channel, with $\sum_i\pi_i = 1$. The above results imply we can model the protocol classically, by recognizing T_a as $m\times m$ sub-stochastic transition matrices and π as an initial stochastic vector (Fig. 4.3c). Equation (4.4) can thus be equivalently written as in Eq. (4.2), with m instead of d states.

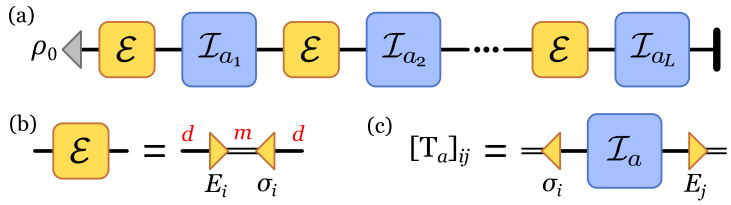

Figure 4.3: (a) The sequential measurement protocol considered in this work. An isolated quantum system in an initial state ρ_0 is repeatedly passed through an entanglement-breaking channel $\mathcal{E}$ before being measured by a quantum instrument $\mathcal{I}$, obtaining a sequence of outcomes $\boldsymbol{a} = a_1 a_2 \ldots a_L$. (b) The entanglement-breaking channel $\mathcal{E}$ understood as a measure-and-prepare operation on a d-dimensional quantum system, with a m-dimensional intermediate classical space. (c) The $m\times m$ transition matrix of the effective classical model.

We are led to conclude that the repeated inclusion of an arbitrary entanglement-breaking channel before the quantum instrument allows the resulting temporal correlations to be, indeed, described classically. This observation seems to justify the association of EB channels with classical memories. Nevertheless, as we already discussed, any temporal correlation can be obtained via a classical system if enough memory

is available, such that a meaningful comparison must be made between systems of *bounded memory*, i.e., bounded dimension.

Here, a central observation is that the effective classical model thus obtained is over m "virtual" classical states (i.e., the number of outcomes in the measure-and-prepare description of the channel), not d as in the dimension of the underlying Hilbert space. This suggests that a qudit passing through an EB channel may be able to generate stronger correlations, i.e., carry more information, than a classical d-it. Clearly, the exception is the case $m \leq d$, where any quantum model can be converted into a classical one of the same dimension (and vice versa) by choosing E_i, σ_i and $\mathcal{I}_a$ acting diagonally on the same basis; see Sec. 4.6.1 for the explicit construction.

4.5 Quantum advantages with EB channels

Here, we show that EB channels with $m > d$ are able to generate non-classical memory effects. This can be investigated by choosing a sequence $\boldsymbol{a}$ with $d < \mathrm{DC}(\boldsymbol{a})$, then computing the maximum probability achievable over all classical models with d states:

$$\Omega_C(\boldsymbol{a}, d) = \max_T \; p(\boldsymbol{a}|T, d). \tag{4.6}$$

Analogously, using a d-dimensional quantum system, we look for instruments $\mathcal{I}$ and EB channels $\mathcal{E}$ with $m > d$ such that

$$p(\boldsymbol{a}|\mathcal{E}, \mathcal{I}, d) > \Omega_C(\boldsymbol{a}, d). \tag{4.7}$$

Crucially, the maxima $\Omega_C(\boldsymbol{a}, d)$ must be found explicitly within classical theory, either by exact computation or by obtaining an upper bound, whereas for the quantum value of $p(\boldsymbol{a}|\mathcal{I}, \mathcal{E}, d)$ an explicit model $(\mathcal{I}, \mathcal{E})$ must be found, obtaining a lower bound for the quantum maximum $\Omega_Q^{\mathcal{E}}(\boldsymbol{a}, d)$. If the classical memory bound has been violated even in the presence of an EB channel $\mathcal{E}$, as in Eq. (4.7), then the system still acts as a quantum memory resource for the task of generating temporal correlations.

To achieve this goal, we consider the *one-tick sequences* [33, 198] of length L, defined as $\boldsymbol{a}_{\mathrm{ot}}^L := 0^{L-1}1$ and having $\mathrm{DC}(\boldsymbol{a}_{\mathrm{ot}}^L) = L$, under the scenario $d = L - 1$. As the case of $L = 2$ is trivial, we consider the smallest nontrivial cases $L \geq 3$. For the classical model, analytical upper bounds can be computed via semidefinite programming (SDP) [207]. The case $L = 3$ and 4 were already solved in [207], and here we computed explicitly the case $L = 5$ with the same method. In all these cases, the upper bounds computed via SDP coincide, up to the numerical precision, with the values obtained via the *one-way model* [33], which is conjectured to be optimal for all $d = L - 1$ scenarios [33, 198]. We are therefore sure to have obtained the exact bounds for the cases $L = 3, 4$, and 5.

4.5.1 A quantum model violating classical bounds

An explicit quantum model violating these classical bounds can be obtained with the following construction. Consider the deterministic classical model on m states given by $[T_0]_{i,i+1} = 1$ for $i = 1, \ldots, m-1$, $[T_1]_{m,1} = 1$, zeros for all other entries, and an initial state $\pi = (1, 0, \ldots, 0)$. The idea is to approximate this model by emulating $m = d + 1$ nearly-orthogonal virtual states with the available d-dimensional quantum system; see Fig. 4.5.

With $\zeta = \mathrm{e}^{2\pi \mathrm{i}/(m-1)}$, we define $\mathcal{E}$ using

$$|\psi_n\rangle = \frac{1}{\sqrt{d-1}} \sum_{k=1}^{d-1} \zeta^{nk} \, |k\rangle \,, \quad n = 1, \ldots, m-1, \tag{4.8}$$

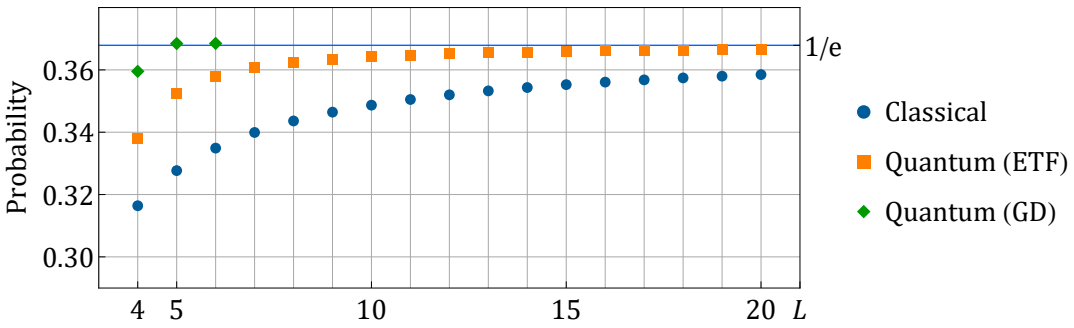

Figure 4.4: Probabilities of one-tick sequences $\boldsymbol{a}^L_{\text{ot}}$ for classical one-way models, the quantum construction using ETFs, and the models obtained numerically with gradient descent (GD), all using $d = L - 1$ and $m = d + 1$. One-way models are conjectured to be optimal for all $L = d - 1$, and are known to achieve the exact upper bounds $\Omega_C(\boldsymbol{a}^L_{\text{ot}}, d)$ for $L = 3, 4$, and 5. The exact bounds for $L = 4$ and 5 are violated by the explicit quantum construction using ETFs, conclusively showing non-classical memory effects of EB channels. Note that both classical and quantum ETF models seem to converge to the $1/\text{e}$ universal classical bound (conjectured in Sec. 2.7.3), except the quantum model converges faster. Quantum models obtained via GD outperform the ETF construction, and even violate the $1/\text{e}$ bound by a small margin for $L = 5$ and 6.

then let $\sigma_n = |\psi_n\rangle\langle\psi_n|$ and $E_{n+1} = \frac{d-1}{d}\sigma_n$ for $n = 1, \ldots, m-1$, and finally setting $\rho_0 = \sigma_m = E_1 = |0\rangle\langle 0|$. The effects of the instrument are defined as $\mathcal{I}_a(\rho) = K_a \rho K_a^\dagger$, with $K_0 = \text{diag}(1, \ldots, 1, 0)$ and $K_1 = \mathbb{1} - K_0$. This quantum model corresponds to isolating the $(d-1)$-dimensional subspace orthogonal to $|0\rangle$, assembling a harmonic complex *equiangular tight frame* (ETF) [196] of size $m - 1$ within this subspace, then using it for the virtual states in the transitions of T_0. ETFs are useful for this task as they provide a generalization of orthonormal bases, saturating the Welch bounds for minimum overlaps [208].

As seen in Fig. 4.4, this quantum ETF construction clearly violates the exact classical bounds for $L = 4$ and 5, conclusively showing that general EB channels cannot be applied to characterize classical memory effects.

4.5.2 Numerical optimization

The quantum model based on ETFs provides an explicit violation of the classical bounds, but this is not necessarily the maximum one can achieve in the quantum case. A natural question to ask is just how much quantum advantage is possible in this scenario, i.e., we wish to obtain the quantum maximum:

$$\Omega^{\mathcal{E}}_Q := \sup_{\mathcal{E},\mathcal{I}} p(\boldsymbol{a}|\mathcal{E}, \mathcal{I}, d). \tag{4.9}$$

While the techniques we presented in Chs. 3 and 5 can address this question, in principle, in practice the problem will not be numerically tractable beyond the $d = 2$ and $L = 3$ case. Therefore, to investigate the quantum maximum with EB channels, we have also performed numerical optimization via gradient descent methods, obtaining lower bounds for $\Omega^{\mathcal{E}}_Q$. Results are summarized in Table 4.1. The quantum models found numerically (also with $m = d + 1$) slightly outperform the previous ETF construction. See Sec. 4.5.2 for the explicit models found.

For $\boldsymbol{a}^L_{\text{ot}} = 0^{L-1}1$ and $d = L-1$, the best classical bounds known are achieved by the one-way models [32, 198], as discussed in Sec. 2.7.2, given explicitly by the matrices

$$[T_0]_{i,i} = 1/L, \quad [T_0]_{i,i+1} = [T_1]_{d,1} = 1 - 1/L, \tag{4.10}$$

with zeroes everywhere else, giving $p(\boldsymbol{a}^L_{\text{ot}}|T, d) = (1 - {}^1\!/_L)^L$. These models are conjectured to give the optimal probability for a given $d = L - 1$ in any sub-deterministic classical scenario (i.e., $d < \text{DC}(\boldsymbol{a})$). For

L	d	**Classical** (Ω_C)	**Quantum** ($\Omega_Q^{\mathcal{E}}$)	$\Omega_Q^{\mathcal{E}}/\Omega_C$	$\Omega_Q^{\mathcal{E}}/(1/\mathrm{e})$
3	2	$(2/3)^3 = 0.\overline{296}$	$\stackrel{?}{=} 0.296296\ldots$	$\stackrel{?}{=} 1$	$\stackrel{?}{=} 0.8054168\ldots$
4	3	$(3/4)^4 = 0.31640625$	≥ 0.359523	≥ 1.136270	≥ 0.9772848
5	4	$(4/5)^5 = 0.32768$	≥ 0.368445	≥ 1.124405	≥ 1.0015381
6	5	$(5/6)^6 \stackrel{?}{=} 0.33489798\ldots$	≥ 0.368483	≥ 1.100285	≥ 1.0016421

Table 4.1: Results from numerical optimization for various one-tick sequences of length L, and Hilbert space dimension $d = L - 1$. The quantum models found numerically for $L = 4$ and 5 outperform the previous ETF construction, and are included in Sec. 4.5.2. No violations were found for $d = 2$, and the $d = 4$ and 5 cases violate even the conjectured universal classical upper bound of $1/\mathrm{e}$, albeit only by a very small amount; see Sec. 4.5.3.

$d = 2, 3$, and 4, these upper bounds are known to be exact [207], with the $d = 4$ case obtained with the same techniques for Ref. [200].

In the quantum case, numerical optimization was performed via gradient descent techniques using the Adam algorithm [100], as done in Ch. 2. In this scenario, however, we required far more precision to obtain reliable results. This was achieved with an additional exponentially decreasing learning rate starting after 10 000 iterations, from 0.07 to 10^{-12} over the subsequent 40 000 iterations, and many trials with random initial values. We have found sufficient to optimize for rank-1 E_i and σ_i, as all optimal models converged to this form. No further assumptions were made about the quantum model.

Each effect $\mathcal{I}_a$ was defined with a single Kraus operator, as no advantage was observed in using more. As the probability in Eq. (4.4) is convex for ρ_0, and we maximize for a single sequence, this initial state can be assumed pure, and w.l.o.g., set to $\rho_0 = |0\rangle\langle 0|$. We have the following constrained optimization problem:

$$\begin{aligned} \textbf{Find:} \quad & \max_{\mathcal{E},\mathcal{I}} p(\boldsymbol{a}|\mathcal{E},\mathcal{I},d) \\ \textbf{Such that:} \quad & E_i, \sigma_i \geq 0 \ \forall i, \\ & \sum_i E_i = \mathbb{1}, \ \mathrm{Tr}\left[\sigma_i\right] = 1, \ \forall i, \\ & \sum_a K_a^\dagger K_a = \mathbb{1}, \end{aligned} \tag{4.11}$$

with $\mathcal{E}(\rho) = \sum_{i=1}^m \mathrm{Tr}\left[\rho E_i\right] \sigma_i$ and $\mathcal{I}_a(\rho) = K_a \rho K_a^\dagger$.

This can be relaxed into an unconstrained optimization by using arbitrary $d \times d$ complex matrices A_i, B_i, C_a, then defining $\sigma_i := A_i^\dagger A_i / \mathrm{tr}[A_i^\dagger A_i]$ for the states. For the POVM, we first define $\tilde{B}_i := B_i^\dagger B_i$ and compute $\lambda_{\max}^E := \max \mathrm{eigs}(\sum_i \tilde{B}_i)$, such that letting $E_i := \tilde{B}_i / \lambda_{\max}^E$ gives $\sum_i E_i \leq \mathbb{1}$. For the instrument we compute $\lambda_{\max}^{\mathcal{I}} = \max \mathrm{eigs}(\sum_a C_a^\dagger C_a)$ and define $K_a := C_a / \sqrt{\lambda_{\max}^{\mathcal{I}}}$, giving $\sum_a K_a^\dagger K_a \leq \mathbb{1}$. In this way, the gradient of the objective function will naturally favor solutions satisfying $\sum_i E_i = \mathbb{1}$ and $\sum_a K_a^\dagger K_a = \mathbb{1}$. The above construction is general, but can be easily adapted for rank-1 σ_i and E_i. The best quantum models found for $L = 4$ and 5 are provided in Sec. 4.5.2.

Note: The remaining of this subsection was not originally included in Ref. [200].

Using $[T_a]_{ij} := \mathrm{Tr}\left[\mathcal{I}_a(\sigma_i) E_j\right]$, we can compare the deterministic classical model with the effective classical model obtained from the ETF construction and the model obtained via GD. A visual comparison is shown in 4.5. Optimal quantum models can be understood as noisy approximations of the deterministic one, with the ETF construction spreading the error uniformly over the σ_i and E_i, whereas the optimal model obtained with GD concentrates the overlaps further via a sophisticated interplay of coherences.

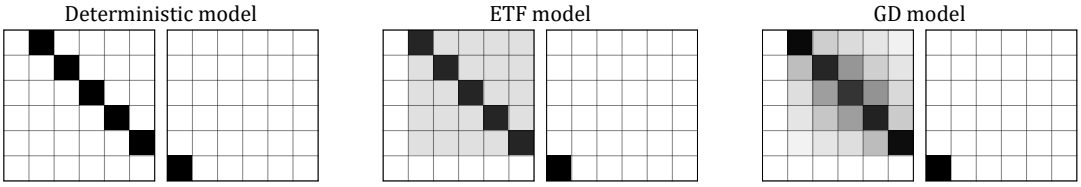

Figure 4.5: Visualization of T_0 and T_1 matrices of the classical deterministic model, and the effective classical models obtained with the ETF construction and gradient descent (GD), for the $d = 5$ and $L = m = 6$ case. Matrix entries (i.e., transition probabilities) are shown with 0 as white, 1 as black, and intermediate values in shades of gray.

Examples violating the classical bound

The following are the best quantum models found for $L = 4$ and 5 using gradient descent techniques. The optimal models for a given dimension and sequence length required high accuracy numerical optimization to be found, and rely on several significant figures in order to adequately satisfy the constraints in Eq. (4.11). Despite considerable efforts, no closed form description of these models has been found.

Since all optimal E_i and σ_i found were rank-1, we include the vectors such that $E_i = |e_i\rangle\langle e_i|$ and $\sigma_i = |\varphi_i\rangle\langle\varphi_i|$. We provide only 5 significant digits of precision, for simplicity, but convergence was obtained up to 10 significant digits. Both cases assume $\rho_0 = |0\rangle\langle 0|$ and $\mathcal{I}_a(\rho) = K_a \rho K_a^\dagger$.

Model for $L = 4$:

$$|e_1\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad |\varphi_1\rangle = \begin{bmatrix} 0 \\ -0.08293 - 0.35949i \\ 0.92946 \end{bmatrix},$$

$$|e_2\rangle = \begin{bmatrix} 0 \\ -0.09692 - 0.41924i \\ 0.74404 \end{bmatrix}, \qquad |\varphi_2\rangle = \begin{bmatrix} 0 \\ 0.09971 + 0.4288i \\ 0.89788 \end{bmatrix},$$

$$|e_3\rangle = \begin{bmatrix} 0 \\ 0.07184 + 0.30898i \\ 0.65475 \end{bmatrix}, \qquad |\varphi_3\rangle = \begin{bmatrix} 0 \\ 0.96462 \\ 0.05913 - 0.25695i \end{bmatrix},$$

$$|e_4\rangle = \begin{bmatrix} 0 \\ -0.18847 - 0.82383i \\ -0.13307 \end{bmatrix}, \qquad |\varphi_4\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

$$K_0 = \mathrm{diag}(0, 1, 1), \qquad K_1 = \mathrm{diag}(1, 0, 0).$$

Model for $L = 5$:

$$|e_1\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad |\varphi_1\rangle = \begin{bmatrix} 0 \\ -0.25627 - 0.10917i \\ -0.29951 - 0.06935i \\ 0.90988 \end{bmatrix},$$

$$|e_2\rangle = \begin{bmatrix} 0 \\ -0.18168 - 0.03287i \\ -0.33582 - 0.01918i \\ 0.83859 \end{bmatrix}, \qquad |\varphi_2\rangle = \begin{bmatrix} 0 \\ 0.78968 \\ 0.17386 + 0.37575i \\ -0.02481 + 0.45209i \end{bmatrix},$$

$$|e_3\rangle = \begin{bmatrix} 0 \\ 0.06342 + 0.62645i \\ -0.36075 + 0.0699i \\ -0.32475 \end{bmatrix}, \qquad |\varphi_3\rangle = \begin{bmatrix} 0 \\ 0.73235 \\ 0.46844 - 0.13164i \\ 0.35015 - 0.32293i \end{bmatrix},$$

$$|e_4\rangle = \begin{bmatrix} 0 \\ -0.40607 - 0.42989i \\ -0.28345 - 0.24945i \\ -0.4251 \end{bmatrix}, \qquad |\varphi_4\rangle = \begin{bmatrix} 0 \\ -0.34778 + 0.40375i \\ 0.82555 \\ 0.14552 + 0.11544i \end{bmatrix},$$

$$|e_5\rangle = \begin{bmatrix} 0 \\ -0.3906 + 0.2592i \\ 0.78055 \\ 0.05633 + 0.08616i \end{bmatrix}, \qquad |\varphi_5\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$K_0 = \mathrm{diag}(0, 1, 1, 1), \qquad K_1 = \mathrm{diag}(1, 0, 0, 0).$$

For $L = 3$ and $d = 2$, we were unable to find any violation of the classical bound, strongly suggesting EB-based temporal non-classicality only emerges beyond two-level systems. Furthermore, since $\mathrm{DC}(\boldsymbol{a}) \leq L$ for any sequence of length L, either a qubit or a bit are sufficient to produce any correlations on two time steps. It is therefore essential to also look beyond two-time steps in order to reveal fundamental differences between classical and quantum systems in this sequential measurement scenario.

We have also performed optimizations for the case $m = d + 2 = L + 1$, but no additional advantage was observed up to the available numerical accuracy.

4.5.3 Relation with the conjectured universal classical bound

Note: This section contains additional material not originally included in Ref. [200].

From Fig. 4.4, it is interesting to note that despite quantum memories offering an advantage they still seem to be intimately related to the $1/\mathrm{e}$ universal classical bound, conjectured in Ref. [198] (see Sec. 2.7.3), with the quantum models merely converging faster. The probability resulting from the ETF quantum models is difficult to obtain in closed form, but numerical evaluation suggests the ETF model also converges uniformly to the $1/\mathrm{e}$ bound as L goes to infinity (with $d = L - 1$ and $m = d + 1$).

An even more surprising result occurs for the optimal models found with gradient descent, as seen in Table 4.1. In particular, the optimal probabilities for $L = 5$ and 6 violate the universal classical bound, but only by a little over 0.15%. This tiny violation is extremely sensitive to the parameters in the model, requiring many trials and high precision in order to be found. We have refined these optimal models further using arbitrary precision arithmetic [138], certifying this violation up to several significant figures in each case. Our numerical results suggest that

$$\Omega_Q^{\mathcal{E}}(\boldsymbol{a}_{\mathrm{ot}}^5, 4) \geq 1.001538121192536 \cdot \frac{1}{\mathrm{e}}. \tag{4.12}$$

A similar violation was also observed for $L = 6$, with

$$\Omega_Q^{\mathcal{E}}(\boldsymbol{a}_{\mathrm{ot}}^6, 5) \geq 1.001642112752928 \cdot \frac{1}{\mathrm{e}}, \tag{4.13}$$

but the model is too cumbersome to include explicitly here, requiring many digits of precision to be specified accurately. Despite our efforts we were unable to find a closed form for these models or the resulting probabilities, but the structure shown in Fig. 4.5 may offer insights into how this could be achieved. We leave the investigation of the quantum upper bounds for future research.

Notwithstanding the fact the $1/\mathrm{e}$ bound is still conjectured, these results point towards a deeper connection between quantum memory effects under the action of EB (M&P) channels and the limits of classical memories, which could provide a bridge between our understanding of classical and quantum memory effects and their role as resources in information processing tasks.

4.6 Conditions for classicality

Since generic EB channels can violate bounds for classical memories, it is desirable to know the conditions under which non-classical memory effects are actually erased by their application, as often assumed, i.e., such that no choice of quantum instrument $\mathcal{I}$ can lead to non-classical temporal correlations.

A full investigation of these conditions is beyond the scope of this work. Nevertheless, understanding some sufficient conditions offers insights on the strength of the implicit assumptions being made when applying EB channels as means of characterizing classical memory effects. In this section we provide some initial results relating to these questions.

4.6.1 Correspondence between classical and quantum models for $m \leq d$

Using $\mathcal{A}$ for the set of outcomes, let $T = (T_a)_{a\in\mathcal{A}}$ and $\{|i\rangle\}_{i=1}^{d}$ denote an arbitrary classical model and orthonormal basis, respectively. We consider a generic quantum instrument with effects written in their Kraus decomposition

$$\mathcal{I}_a(\rho) = \sum_{k=1}^{n_a} K_{ak}\rho K_{ak}^\dagger, \quad \text{with} \quad \sum_{a\in\mathcal{A}}\sum_{k=1}^{n_a} K_{ak}^\dagger K_{ak} = \mathbb{1}, \tag{4.14}$$

such that $\sum_a \mathcal{I}_a$ is trace preserving. We can define a Kraus operator for each non-zero $[T_a]_{ij}$, via the action $K_{aij}|i\rangle = \sqrt{[T_a]_{ij}}\,|j\rangle$, such that $\mathcal{I}$ acts diagonally in this basis. Now, by choosing $E_i = \sigma_i = |i\rangle\langle i|$, we have a completely dephasing channel $\mathcal{E}_{\mathrm{CD}}$ in the same basis, $\mathcal{E}_{\mathrm{CD}}(\rho) := \sum_{i=1}^{m}\langle i|\rho|i\rangle\,|i\rangle\langle i|$. Finally, by using the initial state $\rho_0 = \sum_i \pi_i |i\rangle\langle i|$, we conclude

$$\mathrm{Tr}\left[\mathcal{I}_{a_L}\circ\mathcal{E}\circ\cdots\circ\mathcal{I}_{a_1}\circ\mathcal{E}(\rho_0)\right] = \pi T_{a_1}\cdots T_{a_L}\eta. \tag{4.15}$$

Together with the results from Eqs. (4.4) and (4.5), this establishes the correspondence between classical and quantum scenarios for $m \leq d$, and motivates the investigation of the $m > d$ case, as we have done.

4.6.2 Sufficient conditions for $m > d$

For the case $m > d$, we present two independent sufficient conditions: either the states $\{\sigma_i\}_{i=1}^{m}$ commute, and/or the POVM elements $\{E_i\}_{i=1}^{m}$ commute.

If $[\sigma_i, \sigma_j] = 0$, $\forall i, j$, then a common eigenbasis exists in which all operators can be written as $\sigma_i = \mathrm{diag}(s_i^0, \ldots, s_i^{d-1})$. Without loss of generality, we can express the channel in this basis as

$$\begin{aligned}\mathcal{E}(\rho) &= \sum_{i=1}^{m}\mathrm{Tr}\left[E_i\rho\right]\sigma_i = \sum_{i=1}^{m}\sum_{\ell=0}^{d-1}\mathrm{Tr}\left[E_i\rho\right]s_i^\ell\,|\ell\rangle\langle\ell| \\ &= \sum_{\ell=0}^{d-1}\mathrm{Tr}\left[\left(\sum_{i=1}^{m}s_i^\ell E_i\right)\rho\right]|\ell\rangle\langle\ell| = \sum_{\ell=0}^{d-1}\mathrm{Tr}\left[F_\ell\rho\right]|\ell\rangle\langle\ell|,\end{aligned} \tag{4.16}$$

where $F_\ell := \sum_i s_i^\ell E_i$ and $\{F_\ell\}_{\ell=1}^{d}$ is a valid POVM. In fact, $s_i^\ell \geq 0$ since $\sigma_i \geq 0$ and $\sum_\ell s_i^\ell = 1$ since $\mathrm{Tr}\left[\sigma_i\right] = 1$, which implies $F_\ell \geq 0$ and $\sum_\ell F_\ell = \sum_{i,\ell} s_i^\ell E_i = \sum_i E_i = \mathbb{1}$. We thus have a new definition

of the channel $\mathcal{E}$ which uses only d outcomes and states. But from Sec. 4.6.1, this implies there exists an equivalent classical model with d states. The case for $[E_i, E_j] = 0$ is analogous.

4.7 Conclusions and outlook

Entanglement breaking channels are universally recognized as classical memories [165, 217, 14, 72, 190], and, indeed they are classical under many points of view. Quite surprisingly, however, we found that there exist scenarios where they still act as genuine quantum memories. More precisely, we answer in the negative the following question: is a qudit passing through an entanglement-breaking channel equivalent to a classical d-it?

To address this question, we have investigated a task where memory is the central resource, namely, the generation of correlations in time. Indeed, it has been shown that any temporal correlation can be reproduced by a classical system if enough memory is available [67, 52, 182], such that differences between classical and quantum temporal correlations only emerge if memory is bounded [32, 32, 198]. As classical memory is understood in terms of the number of internal states d of the system, a fair comparison between classical and quantum theory must involve systems of the same size. Differences between quantum and classical memories can be found by investigating the most general set of memory operations allowed in each theory, then establishing bounds on their capabilities.

The violations of classical bounds we have found showcase the importance of investigating classical memory effects in terms of manifestly classical descriptions, instead of relying on quantum-inspired analogies, as otherwise unexpected quantum effects may persist. As we have shown, the common presumption that EB or M&P channels act as classical resource carries additional assumptions, which could be rather strong; see discussion in Sec. 4.6. Furthermore, our results emphasize that, in the context of temporal correlations, it is essential to look beyond qubits and two-time scenarios, as otherwise important distinctions between classical and quantum systems are not apparent, as also recognized in Ref. [190]; see also Sec. 4.5.2.

Our findings inspire several future research directions. Understanding the minimal requirements for an EB channel to truly act as a classical memory could offer insights into the origins of the quantum advantage in the context of temporal correlations, allowing a more definitive understanding of classical and quantum memory effects in non-Markovian processes. Furthermore, since the quantum advantage seen in this work stems from the non-orthogonality of the states σ_i, an investigation of the quantum upper bounds in the presence of EB channels could lead to new insights into the optimal encoding of classical information within quantum systems for various sequential tasks, with potential applications to, e.g., quantum random-access codes [209, 9, 191, 131] or quantum compression of classical predictive models for stochastic processes [17, 63, 212]. These possibilities are particularly promising, considering the close relationship found between the universal classical bound of $1/\mathrm{e}$ and the EB channel upper bound, as discussed in Sec. 4.5.3.

Naturally, there are many distinct features one could wish from a "quantum memory", e.g., preserving coherence or entanglement [217, 106, 40], or enabling nonclassical multi-time statistics [134, 190]. If the quantum memory is required to preserve entanglement, then EB channels are obviously useless as a quantum resource. However, quantum effects are not restricted to spatial correlations, and in considering the temporal character inherent to any information processing task we are required to also consider the fundamental differences between classical and quantum temporal correlations.

Our results indicate a more careful and detailed investigation of quantum memory resources is warranted, such that their precise origin and advantage over classical memories can be properly understood, quantified, and successfully applied in emergent quantum technologies.

Chapter 5

Inducing sparsity in semidefinite programs with sparse objectives

The sparsity heuristic discussed in this chapter was first introduced and applied in Ref. [201] (subject of Ch. 3). These results have since been generalized further, and a more detailed analysis of the problem is present in a separate manuscript, currently in preparation:

> *Inducing sparsity in semidefinite programs with sparse objectives* (working title)
> **Lucas B. Vieira** and Costantino Budroni
> *In preparation*

These expanded results, some not originally present in Ref. [201], are included in this chapter.

Author contribution: The doctoral candidate was fully responsible for the creation of the sparsity heuristic, the writing of all optimization codes, the analysis of the results, and the writing of the majority of the manuscript. The candidate is fully responsible for the production of all its graphical assets. The candidate contributed significantly to the proofs present in this work.

5.1 Introduction

Large-scale problems in semidefinite programming (SDP) are ubiquitous in science, often arising as relaxations of a harder underlying problem, e.g., global polynomial optimization [148, 97, 112, 113], entanglement detection [59, 192], rank-constrained optimization [216], integer programming [159], solving partial differential equations [129], sensor network localization [18, 98], control systems [21, 148], algebraic geometry [75], and combinatorial optimization [21, 111], to name a few.

Solving these problems can be quite challenging due to their scale, usually requiring efficient sparse representations before they become numerically tractable. The usual approach for developing a sparse representation leverages the "aggregate sparsity" of the problem [68, 112, 204, 99, 218], obtained directly from its objective function together with all of its linear constraints. Additionally, despite the variety of constraints appearing in such problems their objective function is often very sparse, enabling further optimizations [97, 112, 139]. In this context, it is worth noting that since these relaxations may fail to provide a feasible solution for the original problem, one typically prioritizes obtaining an optimal value for the objective over its corresponding solution in the relaxation.

Inspired by these observations, in this chapter we present a heuristic method for obtaining sparse representations applicable to arbitrary semidefinite programs. Our method works by iteratively assembling a self-sufficient subset of variables and constraints which, directly or indirectly, affect the objective function. In contrast to the usual approach using aggregate sparsity, which includes all constraints of

the problem, our method *discards* irrelevant variables and constraints by finding the "effective sparsity" implicit in a problem, not directly apparent from its full definition, but which emerges naturally from the structure of its objective and constraints.

We provide two different methods based on this heuristic. The first is capable of providing an exact sparse representation, while the second can be used to obtain a hierarchy of sparse relaxations. Of particular note, our heuristic method is problem-agnostic and can be used in conjunction with existing techniques, possibly as a general pre-optimization step.

The chapter is organized as follows. Section 5.2 provides a summary of the results, and Sec. 5.3 introduces the basic notions of graph theory and semidefinite matrix completion required. Section 5.4 outlines the ideas underlying our heuristic method, including Sec. 5.4.1 describing the heuristic method to obtain an exact effective sparsity of a problem, and Sec. 5.4.2 an alternative method for obtaining sparse relaxations. In Sec. 5.5, we discuss how our heuristic compares with existing methods. Section 5.6 showcases the efficacy of our heuristic in its original application, with Sec. 5.7 discussing some potential refinements. Finally, Sec. 5.8 provides an overview of the results and final remarks.

5.2 Summary of main results

For Ref. [201] (subject of Ch. 3), we have developed a convergent hierarchy of semidefinite programs providing upper bounds on correlations arising from sequential operations on open quantum systems. We applied this to the problem of witnessing the dimension of an environment interacting with a probe system being sequentially measured. The goal was to use the resulting temporal correlations on the sequence of outcomes, $p(\boldsymbol{a})$ for $\boldsymbol{a} \in \mathcal{A}^L$, in order to obtain information about the inaccessible environment.

However, the simplest nontrivial scenario—a qubit system and qubit environment with three sequential measurements—was already numerically intractable. Even after all of the problem's symmetries had been accounted for, it would still require 3 TB of RAM to be solved with standard algorithms. We overcame this by developing a heuristic technique which automatically obtains an *exact* sparse representation of the SDP through an iterative approach, resulting in the elimination of variables and constraints. The technique exploits the sparsity of the objective function and constraints separately, iteratively assembling a self-sufficient set of variables and constraints where the original problem can be solved exactly. For this particular problem our novel heuristic was extremely successful, reducing the number of variables and constraints to less then 1% of their original number.

It was also realized that this heuristic was not only quite general, being applicable to several other problems in quantum information theory and beyond, but also very flexible, allowing hierarchies of relaxations to be obtained automatically for arbitrary semidefinite problems, in a problem-agnostic way. We have also found that our heuristic differs significantly from many existing methods that attempt to obtain sparse representations of large-scale SDPs.

Our successful results with this heuristic are currently under preparation as a separate manuscript, in order to showcase its strengths and flexibility to a wider community, and on a variety of problems. This chapter is partially adapted from this manuscript, in preparation, and contains some material originally included in Ref. [201], where our technique was first introduced.

5.3 Preliminary notions

5.3.1 Graph theory

A graph G is defined by a pair (V, E), where $V = \{1, \ldots, n\}$ is a set, the *vertices*, and $E \subseteq V \times V$ a set of pairs (i, j), the *edges*, for some $i, j \in V$. If $(i, j) \in E$, the two vertices $i, j \in V$ are said to be *adjacent*,

with the edge (i, j) being *incident* to them. The *degree* of a vertex is the number of edges incident to it. We consider only *undirected* graphs, meaning that the pairs in E are not ordered, i.e., $(i, j) = (j, i)$. We admit the possibility of edges connecting vertices with themselves, i.e., $(i, i) \in E$ and we call such edges *self-edges* or *loops* (not to be confused with cycles, see below). If all vertices of a graph are adjacent, the graph is said to be *complete*.

Given a subset of vertices $V' \subseteq V$, the graph $G' = (V', E')$, with edges $E' = E \cap (V' \times V')$, is referred to as an *induced subgraph* of G. A *path* is a sequence of vertices $v_0, v_1, \ldots, v_n$ connected by edges, i.e., $(v_i, v_i + 1) \in E$ for $i = 0, \ldots, n - 1$. If $v_0 = v_n$, and all vertices are otherwise distinct, we say that the path is a *cycle* of length n (not to be confused with a loop, see above). A graph is said to be *connected* if there is a path (i.e., a sequence of edges) joining any two of its vertices. The *connected components* of G are the disjoint connected subgraphs $\{G_\ell\}_\ell$ such that $G = \bigcup_\ell G_\ell$, and no two G_ℓ share a vertex.

If an induced subgraph G' is a complete graph, it is referred to a *clique* of G. A *maximal clique* is a clique which is not a subgraph of a larger clique. A graph is said to be *chordal* if for every cycle of length $n \geq 4$ there is a *chord*, i.e., an edge between two non-adjacent vertices in the cycle. Given any graph G, we can add extra edges such that the obtained graph G' is chordal. We refer to G' as the chordal completion of G. Moreover, the set of cliques of a chordal graph satisfy the *running intersection property* [68], namely, there is an ordering of them $\{V_1, \ldots, V_m\}$ such that

$$V_i \cap (V_1 \cup \ldots \cup V_{i-1}) \subset V_j \text{ for } j < i. \tag{5.1}$$

One can verify that a graph is chordal via the Lex-BFS algorithm [164, 53]. Finally, a graph $G = (V, E)$ can be represented by an *adjacency matrix*, e.g., a $n \times n$ matrix A, for $n = |V|$, where $A_{ij} = 1$ if $(i, j) \in E$ and 0 otherwise. Since we consider only undirected graphs, adjacency matrices will be symmetric. Figure 5.1 provides a concrete example of these concepts.

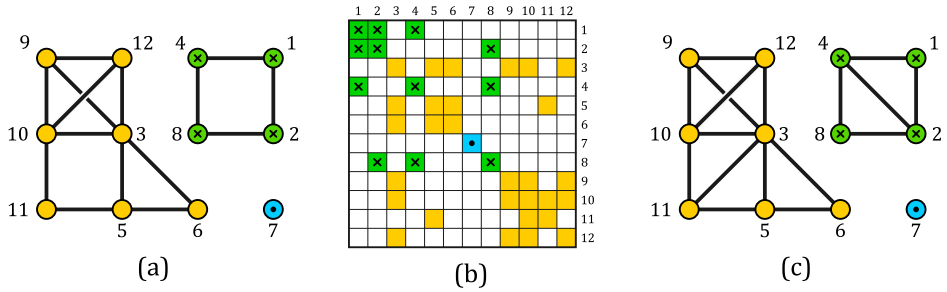

Figure 5.1: (a) A non-chordal graph on 12 vertices with three connected components — {3,5,6,9,10,11,12}, {1,2,4,8}, and {7}. Non-chordality is due to the chord-less length-4 cycles (3,5,11,10) and (1,2,8,4). The graph has 12 cliques of size 1 (vertices), 15 of size 2 (edges), 5 of size 3, and 1 of size 4. Three cliques are maximal: {3,9,10,12}, {3,5,6}, and {7}. (b) A visual depiction of the graph's adjacency matrix. Non-zero entries are depicted with the colors (and markers) of the components they belong to. Self-edges or loops, corresponding to diagonal entries, are omitted in the graph. (c) A chordal completion of the graph, now with 7 maximal cliques, obtained by adding edges (3,11) and (2,4).

5.3.2 Partial matrices and completions

Let $\mathcal{M}_n$ denote the set of $n \times n$ Hermitian matrices and $\mathcal{M}_n^+$ its space of positive semidefinite (PSD) matrices. Given a graph $G = (V, E)$ for $V = \{1, \ldots, n\}$, we define the *G-partial Hermitian matrix* (PHM) $X[G]$ as a matrix X such that only the entries X_{ij} corresponding to the edges (i, j) of G are specified, and $X_{ij} = X_{ji}^*$. An *Hermitian completion* of $X[G]$ is a matrix $M \in \mathcal{M}_n$ such that $M_{ij} = X_{ij}$ whenever $(i, j) \in E$.

Moreover, given a subgraph G' of G, we define $X[G']$ as the PHM obtained from X by discarding (i.e., leaving unspecified) all entries corresponding to vertices and edges of G that do not appear in G'.

Similarly, we say that the partial Hermitian matrix $A[G]$ is a *G-partial positive semidefinite matrix* (PPSDM) if $A[G]$ is a PHM, and for all cliques V_i of G the (full) matrix $A[V_i]$ is positive semidefinite. Given a PPSDM $A[G]$, $M \in \mathcal{M}_n^+$ is a *positive semidefinite completion* of $A[G]$ whenever M is an Hermitian completion of A that is positive semidefinite[1].

Theorem 2. (Grone et al. [76, Theorem 7]) The PPSDM $A[G]$ for $G = (V, E)$ has a positive semidefinite completion X if the graph G is chordal.

Theorem 2 is obtained by Grone et al. by a direct construction of the extension and convexity arguments. This key result is the basis for many approaches to obtain sparse representations of semidefinite problems [68, 97, 204, 112, 113, 99, 218].

Given a matrix $A \in \mathcal{M}_n$, we define its *sparsity pattern* as

$$\mathcal{S}(A) := \{\ (i,j) \in V \times V \ \mid\ A_{ij} \neq 0\ \}, \tag{5.2}$$

where $V := \{1, \ldots, n\}$ is the set of row/column indices. Since A is Hermitian, we can interpret the pair $(V, E = \mathcal{S}(A))$ as an undirected graph. Given (V, E), we define the *vertex support* of E, denoted by $\mathcal{V}(E)$, as the subset of vertices from V with non-zero degree, i.e.,

$$\mathcal{V}(E) := \{\ i \in V \ \mid\ \deg_E(i) > 0\ \}, \tag{5.3}$$

where $\deg_E(i)$ denotes the degree of the vertex $i \in V$ given E. Therefore, $\mathcal{V}(\mathcal{S}(A))$ is the set of indices (row *or* column) in the non-zero entries of A.

5.4 Main results

We consider optimization problems in the form of semidefinite programs (SDPs), written as

$$\begin{aligned} \textbf{Given:}\quad & F,\ \{C_k\}_{k=1}^m,\ \{c_k\}_{k=1}^m \\ \textbf{Find:}\quad & f := \max_X \langle F, X\rangle \\ \textbf{Subject to:}\quad & \langle C_k, X\rangle = c_k, \quad k = 1, \ldots, m \\ & X \geq 0 \end{aligned} \tag{P1}$$

with $F, X, C_k \in \mathcal{M}_n$ and $c_k \in \mathbb{C}$, and where we denote by $\langle A, B\rangle = \mathrm{tr}[A^\dagger B]$ the Frobenius inner product. Our goal is to obtain a sparse representation of this problem, exactly or as a relaxation.

These sparse representations are defined by a sparsity pattern E^*, allowing the structure of the problem to be mapped onto a graph $G = (V, E^*)$ where each edge $(i, j) \in E^*$ is associated with a scalar optimization variable X_{ij}. For this reason, and with a slight abuse of notation, we may refer to $X_{ij} \in E$ to indicate $(i, j) \in E$. There are several approaches that exploit the *aggregate sparsity* of the problem [68, 112, 204, 99, 218], which takes into account, at the same time, the sparsity of the objective function F and the linear constraints C_k. Here, we show that the sparsity of the problem can be, in principle, further increased by exploiting the individual sparsity structures of the objective function and the constraints. A comparison with the usual methods in the literature is provided in Sec. 5.5.

[1] All results mentioned hold both in the positive definite and positive semidefinite cases. To keep the notation lighter, we simply discuss the positive semidefinite case.

5.4.1 Exact effective sparsity

Our first method transforms the original problem into an equivalent one with more sparsity, i.e., with the same optimum and a feasible solution, but involving less variables and constraints. First, we separate the homogeneous and non-homogeneous constraints. Letting $K = \{1, 2, \ldots, m\}$ denote the indices of the m constraints, we define the sets

$$K_{\mathrm{nh}} := \{\, k \in K \;\mid\; c_k \neq 0 \,\}, \qquad K_{\mathrm{h}} := \{\, k \in K \;\mid\; c_k = 0 \,\}. \tag{5.5}$$

We begin by defining the initial *base sparsity* E^0_{base} of the problem via

$$E^0 := \mathcal{S}(F) \cup \bigcup_{k \in K_{\mathrm{nh}}} \mathcal{S}(C_k), \qquad E^0_{\mathrm{base}} := E^0 \cup \big\{\, (i,i) \;\big|\; i \in \mathcal{V}(E^0) \,\big\}, \tag{5.6}$$

In other words, we include all the variables X_{ij} appearing in F and in non-homogeneous constraints C_k, as well as the corresponding diagonal variables X_{ii} and X_{jj}.

Since the variables $X_{ij} \notin E^0_{\mathrm{base}}$ do not appear in the objective function, they do not contribute directly to its value. Nevertheless, some of them may still contribute indirectly through the other linear constraints or the positivity constraint. The central idea of our method is to identify which such variables and constraints are relevant for the optimization.

Let $\kappa \subseteq K$ denote the subset of constraints we will keep, which we initialize with $\kappa \leftarrow K_{\mathrm{nh}}$ for this method. Using Proc. 1 with E^0_{base} and this initial κ, we obtain the corresponding *extended sparsity* E^0_{ext} and updated κ.

Procedure 1. Obtaining an extended sparsity from a base sparsity.

Input: Base sparsity E_{base}, current set of constraints κ.

Output: Extended sparsity E_{ext}, updated set of constraints κ.

1. Initialize $\varepsilon \leftarrow E_{\mathrm{base}}$ as the current candidate for E_{ext}.
2. Set $K' = \{\, k \in K_{\mathrm{h}} \;\mid\; \mathcal{S}(C_k) \cap \varepsilon \neq \emptyset \,\}$, and update $\kappa \leftarrow \kappa \cup K'$.
3. Set $\varepsilon' = \varepsilon \cup \bigcup_{k \in K'} \mathcal{S}(C_k)$.
4. If $\varepsilon' = \varepsilon$, then return $E_{\mathrm{ext}} = \varepsilon$ and κ.
5. Else, set $\varepsilon \leftarrow \varepsilon'$, then repeat from step (2).

The resulting extended sparsity E^0_{ext} is an initial guess for the minimum subset of variables $X_{ij} \in E^0_{\mathrm{ext}}$ which should be considered in the problem, and κ the relevant subset of constraints. Let $G^0 := (V, E^0_{\mathrm{ext}})$ be a graph with connected components $G_\ell = (V_\ell, E_\ell)$. Positivity of $X[G^0]$ can be ensured by treating it as a (possibly permuted) block matrix, with each positive semidefinite block corresponding to a component G_ℓ made into a complete graph:

$$E^0_{\mathrm{comp}} := \bigcup_{\ell} V_\ell \times V_\ell. \tag{5.7}$$

The completed sparsity E^0_{comp} will contain all variables required to exactly solve a *relaxation* of Prob. 1 involving only the constraints in κ. If the completion procedure in Eq. (5.7) led to the inclusion of new variables, however, then these variables might involve additional constraints beyond κ, which now must also be taken into account.

Therefore, we define a new base sparsity $E^1_{\mathrm{base}} = E^0_{\mathrm{comp}}$, and repeat all of the previous steps. If after s iterations we find that $E^s_{\mathrm{base}} = E^s_{\mathrm{comp}}$, then we have found a self-sufficient subset of variables and constraints completely independent from the rest. We refer to these as the *effective sparsity* E^* and the *effective constraints* κ^* of the problem, as they emerged naturally from its structure. The overall procedure is outlined in Proc. 2.

Procedure 2. Obtaining an exact effective sparsity for an SDP.

Input: Objective matrix F, constraint matrices $\{C_k\}_k$.

Output: Exact effective sparsity E^* and subset of constraints κ^*.

1. Initialize iteration $s \leftarrow 0$, base sparsity E^0_{base} as in Eq. (5.6), and constraints $\kappa \leftarrow K_{\mathrm{nh}}$.
2. Obtain the extended sparsity E^s_{ext} and updated κ using Proc. 1.
3. Find the connected components $\{G_\ell\}_\ell$ of $G = (V, E^s_{\mathrm{ext}})$, with $G_\ell = (V_\ell, E_\ell)$.
4. Set $E^s_{\mathrm{comp}} = \bigcup_{\ell=1}^{L} V_\ell \times V_\ell$ as the completed sparsity.
5. If $E^s_{\mathrm{base}} = E^s_{\mathrm{comp}}$, then return $E^* = E^s_{\mathrm{base}}$ and $\kappa^* = \kappa$.
6. Else, set $E^{s+1}_{\mathrm{base}} \leftarrow E^s_{\mathrm{comp}}$, increment s, then repeat from step (2).

The resulting sparse SDP is then:

$$\begin{aligned} \textbf{Given:} \quad & F,\ \{C_k\}_{k\in\kappa^*},\ \{c_k\}_{k\in\kappa^*} \\ \textbf{Find:} \quad & f := \max_{\{X_\ell\}_\ell} \sum_\ell \langle F_\ell, X_\ell\rangle \\ \textbf{Subject to:} \quad & \sum_\ell \langle C^\ell_k, X_\ell\rangle = c_k,\ \forall\, k \in \kappa^* \\ & X_\ell \geq 0,\ \forall\, \ell. \end{aligned} \tag{P2}$$

Here, we used that $\{F_\ell\}_\ell$ are the blocks of the matrix $F' = PFP^{\mathrm{T}} = \bigoplus_\ell F_\ell$ and $\{C^\ell_k\}_\ell$ the blocks of the matrix $C'_k = PC_kP^{\mathrm{T}} = \bigoplus_\ell C^\ell_k$. Note that such a decomposition exists since the sparse variable X admits the decomposition $PXP^{\mathrm{T}} = \bigoplus_\ell X_\ell$.

We can now prove the first result.

Theorem 3. The optimization problems in Prob. 1 and Prob. 2 are equivalent, namely, from every feasible solution of one a feasible solution of the other can be constructed with the same value for the objective function.

Proof. The first observation is that the original variable X in Prob. 1 can be split into two terms $X = X^* + X_\perp$, where X^* contains the variables X_{ij} appearing in E^*, and zero everywhere else, corresponding up to a permutation to $\bigoplus_\ell X_\ell$, whereas $X_\perp$ contains the variables X_{ij} not appearing in E^*, and zero everywhere else. Now, for the objective function, we have that

$$\langle F, X\rangle = \langle F, X^*\rangle + \langle F, X_\perp\rangle = \langle F, X^*\rangle + 0 = \langle F, X^*\rangle\,, \tag{5.9}$$

i.e., the variables in $X_\perp$ do not contribute to the objective function. Thus, substituting X with X^* does not change the objective function. Moreover, $X^* = P^{\mathrm{T}}(\bigoplus_\ell X_\ell)P$ constructed from $\{X_\ell\}_\ell$, a feasible solution of Prob. 2, satisfies only a subset of the constraints. As a consequence, Prob. 2 provides a relaxation of the original problem. A more careful analysis, however, shows that X^* actually provides feasible solution of the original problem. This is simply done by defining $X = X^* + X_\perp$ with $X_\perp = 0$. The positivity $X \geq 0$ is guaranteed by the block structure of X^*, and the missing constraints, i.e., $\{\langle C_k, X\rangle = 0\}_{k\in K_{\mathrm{h}}\setminus\kappa^*}$ are all satisfied. In fact, they involve only variables in $X_\perp$, i.e., $\langle C_k, X\rangle = \langle C_k, X_\perp\rangle$ for all $k \in K_{\mathrm{h}} \setminus \kappa^*$, which are all homogeneous linear constraints. □

In summary, any sparse solution $X^* \in E^*$ is also a valid solution to the dense problem, with the same optimal value f for the objective: the sparse representation is exact. In the following, we show how we can further increase the sparsity of the problem at the price of defining a relaxation of it.

5.4.2 Approximate effective sparsity

Our second method combines the results of Thm. 2 on the positive extensions of G-PPSDM, as exploited also by [68], with the iterative elements of the exact effective sparsity approach described in Sec. 5.4.1. We begin by defining the initial base sparsity as,

$$E^0_{\text{base}} := \mathcal{S}(F) \cup \{\ (i,i)\ \mid\ i \in \mathcal{V}(\mathcal{S}(F))\ \}, \tag{5.10}$$

and κ as initially empty. The procedure for obtaining the extended sparsity E^0_{ext} and updated κ is the same as before (Proc. 1). However, to obtain the completed sparsity, we now perform a chordal completion of each component G_ℓ to obtain the chordal graph $G^{\text{ch}}_\ell = (V_\ell, E^{\text{ch}}_\ell)$. This provides the (chordal) completed sparsity

$$E^0_{\text{comp}} := \bigcup_{\ell=1}^{L} E^{\text{ch}}_\ell, \tag{5.11}$$

which we then use as E^1_{base} in order to iteratively collect further homogeneous and non-homogeneous constraints, repeating until the effective sparsity E^* and constraints κ^* are obtained. Dropping the ch superscripts, let $G = (V, E^*)$ be the final graph of the effective sparsity, with each connected component $G_\ell = (V_\ell, E_\ell)$ being a chordal graph. Just as in Sec. 5.4.1, each component can be written as a partial matrix $X_\ell[G_\ell]$, with the constraints that each block corresponding to a clique in E_ℓ is positive semidefinite. By Thm. 2, this guarantees that $X_\ell[G_\ell]$ has a positive semidefinite completion $X_\ell \succeq 0$.

Let us fix a connected component ℓ and define $\{V^\ell_1, \ldots, V^\ell_m\}$ as the maximal cliques in G_ℓ. We define a dense matrix Y^ℓ_r for each clique V^ℓ_r and denote by $W^\ell_{rr'} := V^\ell_r \cap V^\ell_{r'}$ the vertices shared by two cliques r and r'. The overlap between cliques implies that the matrices $\{Y^\ell_r\}_r$ do not represent independent variables, but are related by the equality constraints $[Y^\ell_r]_{ij} := [Y^\ell_{r'}]_{ij}$ for all $i, j \in W^\ell_{rr'}$, and for all cliques r, r' of every connected component ℓ. By Thm. 2, it is sufficient to impose positivity to each Y^ℓ_r to guarantee the existence of a completion $X_\ell \succeq 0$. Note, however, that the necessity of these equality constraints creates a trade-off between the number of cliques and the size of the cliques; see Ref. [68] for a discussion.

The overall procedure is outlined in Proc. 3.

Procedure 3. Obtaining an approximate effective sparsity for an SDP.

Input: Objective matrix F, constraint matrices $\{C_k\}_k$.

Output: Approximate effective sparsity E^* and subset of constraints κ^*.

1. Initialize iteration $s \leftarrow 0$, base sparsity E^0_{base} as in Eq. (5.6), and constraints κ as empty.

2. Obtain the extended sparsity E^s_{ext} and updated κ using Proc. 1.

3. Find the connected components $\{G_\ell\}_{\ell=1}^L$ of $G := (V, E^s_{\text{ext}})$, with $G_\ell = (V_\ell, E_\ell)$.

4. Find a chordal completion $G^{\text{ch}}_\ell = (V_\ell, E^{\text{ch}}_\ell)$ of each G_ℓ.

5. Set $E^s_{\text{comp}} = \bigcup_{\ell=1}^L E^{\text{ch}}_\ell$ as the (chordal) completed sparsity.

6. If $E^s_{\text{base}} = E^s_{\text{comp}}$, then return $E^* = E^s_{\text{base}}$ and $\kappa^* = \kappa$.

7. Else, set $E^{s+1}_{\text{base}} \leftarrow E^s_{\text{comp}}$, increment s, then repeat from step (2).

We thus obtain the sparse SDP:

$$\begin{aligned}
\textbf{Given:} \quad & F,\ \{C_k\}_{k\in\kappa^*},\ \{c_k\}_{k\in\kappa^*} \\
\textbf{Find:} \quad & f := \max_{\{Y_r^\ell\}_{\ell,r}} \sum_{r,\ell} \langle F^{(r,\ell)}, Y_r^\ell \rangle \\
\textbf{Subject to:} \quad & \sum_{r,\ell} \langle C_k^{(r,\ell)}, Y_r^\ell \rangle = c_k,\ \forall\, k \in \kappa^* \\
& Y_r^\ell \succeq 0,\ \forall\, \ell, r \\
& [Y_r^\ell]_{ij} = [Y_{r'}^\ell]_{ij}, \quad \forall\, i,j \in W_{rr'}^\ell \text{ and } \forall\, \ell, r, r'.
\end{aligned} \tag{P3}$$

Here, the matrix $F^{(r,\ell)}$ is obtained first by splitting it into connected components F_ℓ, as in Prob. 2, then constructing a smaller matrix $F^{(r,\ell)}$ for each clique V_r^ℓ, corresponding to the matrices Y_r^ℓ. Such a splitting is not enough, however, since one has to take into account also the overlap between cliques. A possible solution is that in case of overlap, i.e., $i,j \in W_{rr'}^\ell$, one put to zero $[F^{(r,\ell)}]_{ij}$ for all matrices but one. The same construction is applied to generate the matrices $C_k^{(r,l)}$.

We can then state our second result:

Theorem 4. The optimization in Prob. 3 provides a relaxation of Prob. 1.

Proof. To prove this, it is sufficient to show that from every feasible solution of Prob. 1 one can construct a feasible solution of Prob. 3. To see how, given a matrix X as a feasible solution of Prob. 1, we can construct the submatrices $\{Y_r^\ell\}_{\ell,r}$. By construction, such matrices are positive semidefinite, and provide the same value for the objective function, i.e., $\langle F, X\rangle = \sum_{r,\ell}\langle F^{(r,\ell)}, Y_r^\ell\rangle$, while satisfying all linear constraints, i.e., $\langle C_k, X\rangle = \sum_{r,\ell}\langle C_k^{(r,\ell)}, Y_r^\ell\rangle = c_k$ for all $k \in \kappa^*$. □

Note that, since multiple chordal completions may be available for a given graph, the solutions obtained by this method will not generally be unique. Furthermore, since the effective sparsity is built iteratively based on such completions, the size of the final effective sparsity may be sensitive to the particular choice of completions performed at every iteration. We will return to this point in the next sections.

5.5 Comparison with methods using aggregate sparsity

The typical approach to obtain a sparse representation of an SDP exploits the *aggregate sparsity* of the problem [68, 112, 204, 99, 218], defined as:

$$E_{\text{agg}} := \mathcal{S}(F) \cup \bigcup_{k=1}^{m} \mathcal{S}(C_k) \cup \{\ (i,i)\ \mid\ i \in V\ \}\,. \tag{5.13}$$

In words, the aggregate sparsity assumes every single constraint and every diagonal term must be included in the sparsity pattern by default. This sparsity gives the graph $G = (V, E_{\text{agg}})$, for which a chordal completion $G' = (V, E_{\text{agg}}^*)$ is obtained directly. Positivity of X can then be imposed by requiring positivity on the smaller matrices $\{X_r\}_r$, each corresponding to one of the cliques in the completed chordal graph. The resulting SDP is then given by

$$\begin{aligned}
\textbf{Given:} \quad & F,\ \{C_k\}_{k\in K},\ \{c_k\}_{k\in K} \\
\textbf{Find:} \quad & f := \max_{\{X_r\}_r} \sum_{r} \langle F_r, X_r \rangle \\
\textbf{Subject to:} \quad & \sum_{r} \langle C_k^r, X_r \rangle = c_k,\ \forall\, k \in K \\
& X_r \succeq 0,\ \forall\, r \\
& [X_r]_{ij} = [X_{r'}]_{ij}, \quad \forall\, i,j \in W_{rr'},\ \text{and } \forall\, r, r',
\end{aligned} \tag{P4}$$

where the matrices C_k^r and F_r are obtained for each clique V_r, corresponding to the matrices X_r. Just as in Prob. 3, one has to account for the overlaps between cliques $W_{rr'} := V_r \cap V_{r'}$, fixing $[X_r]_{ij} = [X_{r'}]_{ij}$ for all $i, j \in W_{rr'}$ and ensuring only one $[F^r]_{ij}$ is non-zero in the overlapping cliques.

Despite the success of this approach, for certain problems the constraint matrices C_k may involve too many non-zero entries when considered together, such that we obtain $E_{\text{agg}} \approx V \times V$ and this sparse representation is not tractable. Therefore, a problem-specific transformation [68, 97, 99, 218] or some regularization method [142] is typically used instead.

The aggregate sparsity method would have been entirely useless in our original problem in Ref. [201] (Chapter 3), for which we first developed our heuristic. In this problem the aggregate sparsity offers practically no advantage, whereas our heuristic provides an immense degree of optimization while still offering an exact solution; see Sec. 5.6.

Note that, as our method is heuristic in nature, a sparse representation is not guaranteed to be obtained, in which case an alternative approach must be pursued anyway. As of yet, it is unclear which properties of a problem make our heuristic work well. This will be subject of future research. However, since our method is problem-agnostic and simply returns the original problem in its worst case scenario, it can be useful if adopted as a general pre-processing step, possibly preceding subsequent optimizations. We elaborate on these ideas in Sec. 5.7.

Finally, we note that it is always possible to transform an SDP into an equivalent one with no constraints by means of an appropriate problem-specific affine transform. In this form, our exact heuristic ceases to apply, and the approximate heuristic falls back to the usual aggregate sparsity strategy. However, for moderate scale SDPs this transformation might be costly to obtain, and will likely obfuscate the original structure of the problem. This could possibly hinder a clear interpretation of its results.

5.6 Examples

Our heuristic method was first applied in Ref. [201] (Chapter 3), where a hierarchy of SDPs was developed to upper bound the maximum probability of obtaining specific sequences of measurement outcomes. The sequences arise from repeated preparations and then measurements on an open quantum system interacting with a quantum environment of finite dimension. The SDP was constructed by describing the physical scenario in the formalism of quantum supermaps [41], where the repeated unitary interactions between a qubit system and a d-dimensional environment are represented by a tensor product of $N \geq 3$ copies of a $(2d)^2 \times (2d)^2$ matrix (Eq. (3.13)).

Even exploiting the highly symmetric structure of the problem (Sec. 3.7.4), the smallest non-trivial scenarios with $d = 2$ already involve far too many variables and constraints to be numerically tractable. Furthermore, despite the sparse objective, the problem's aggregate sparsity is nearly fully dense (Table 5.1) due to a partial trace constraint required by the supermap representation of quantum channels (Eq. (3.16)). These features demanded a novel approach. Using our exact heuristic for effective sparsity, as described in Sec. 5.4.1, the problem could be vastly simplified into a sparse representation that was numerically tractable. Table 5.1 compares the scale of the problem before and after applying our method, providing a compelling case for the power and efficacy of our iterative heuristic.

5.7 Refinements

While we have presented our sparsity heuristic in a problem-agnostic way, many refinements are possible by exploiting additional structures of a problem, or when considering limited computational resources. In this section we discuss a few ways in which our method can be adapted to specific scenarios.

N		Dense problem (intractable)	Aggregate sparsity (intractable)	Effective sparsity (exact & tractable)
3	Variables	665 856	574 016 (86%)	3 566 (0.54%)
	Constraints	295 937	295 937 (100%)	2 809 (0.95%)
4	Variables	15 023 376	14 250 276 (95%)	35 688 (0.24%)
	Constraints	10 653 697	10 653 697 (100%)	40 441 (0.38%)

Table 5.1: Comparison between the number of variables and constraints in the symmetric version of the problem, comparing its dense form against its sparse representations. The aggregate sparsity method can only eliminate a few variables while keeping all constraints, whereas our method achieves a vast reduction in the number of both variables and constraints, while still providing exact solutions.

5.7.1 Metaheuristics

First, we note that the iterative completion procedure can be stopped at any moment before a stable sparsity E^* and set of constraints κ^* are found, thereby obtaining an automatic "hierarchy of relaxations" of a problem based on its inherent structure. This strategy can be used, e.g., when computational resources are limited and a maximum problem size is set beforehand, thus allowing the available resources to tackle a given problem to whatever level of complexity is tractable.

Several strategies can be built upon this idea. In many optimization problems, it is natural to consider some constraints as more important than others, such that it could be useful to include them directly in the base sparsity. In the same spirit, if a priority can be assigned to the constraints, then a metaheuristic approach becomes available wherein the completed sparsity procedure (Proc. 1) only includes the highest priority constraints at each step for E^s_{comp}, until some stopping condition is reached. This could ensure that the automatic hierarchy of relaxations is capturing the essential features of the problem first.

The choice between full completion (Sec. 5.4.1) vs. chordal completion (Sec. 5.4.2) of each connected component can also be made on a case-by-case basis, perhaps based on a similar notion of priorities, or based on a threshold on the size of the component to be completed. Since chordal completions G^{ch}_ℓ for a given G_ℓ are not unique, and finding a completion with the minimum number of edges is a NP-Complete problem [215], various algorithms exist for obtaining or enumerating reasonably small completions in practical scenarios [86, 27]. However, in our heuristic approach the inclusion of new constraints or variables (i.e., edges in E^s_{comp}) is performed iteratively, such that the subsequent chain of variables and constraints to be included can be unpredictable. Thus, there might not be, *a priori*, a notion of "optimal" chordal completion at each step. In this case, it might be helpful to use an enumerative approach instead (see Ref. [27]), obtaining multiple candidate sequences of completions, then choosing the sequence resulting in the smallest number of edges.

5.7.2 Optimizing the procedure for extended sparsity

While we have presented Proc. 1 constructing E^s_{ext} iteratively, this can also be achieved in single step with some pre-processing. First, construct a graph relating the overlaps between constraints, i.e., we let $G_K := (K, E_K)$, where $E_K := \{ (k,k') \in K \times K \mid \mathcal{S}(C_k) \cap \mathcal{S}(C_{k'}) \neq \emptyset \}$. Then, the connected components of G_K each correspond to a set of constraints (and corresponding variables) which would be iteratively added to E^s_{base} to obtain E^s_{ext}. The procedure then reduces to checking whether E^s_{base} overlaps with any constraint C_k, finding which component of G_K contains k, and adding all of its constraints to κ and variables to E^s_{ext}. However, this non-iterative procedure might involve a considerable amount of pre-computation for defining G_K, which might never be used in full, and could also be in conflict with some of the metaheuristics previously discussed.

5.7.3 Effective sparsity as a pre-optimization

Finally, since our heuristic is problem-agnostic and entirely structural—i.e., it only eliminates variables and constraints, with no further transformations of the input problem—it can be considered as a general pre-optimization step, thus providing a reduced version of a problem which is more amenable to other simplifications.

As an example, the sparse problem could allow for Gaussian elimination (or some other transformation) to be performed, automatically reducing the number of variables and constraints further (possibly at the cost of sparsity). Nevertheless, we leave such investigations for future work.

5.8 Conclusions and outlook

We have developed a heuristic method that is capable of significantly reducing the size of large-scale semidefinite programs. Unlike the traditional approach based on a problem's aggregate sparsity, our method eliminates both variables and constraints by identifying the *effective sparsity* structure that inherently satisfies constraints, and naturally emerges from the problem itself. This method is automatic and problem-agnostic, making it suitable for a variety of applications.

While our heuristic has been very successful in producing sparse versions of certain problem, there is no general guarantee that the final problem will be sparse. It may be the case that our method eventually includes all variables and constraints. When this happens, alternative strategies must be employed, such as stopping before achieving a stable sparsity pattern (typically resulting in an outer approximation) or leveraging additional structures of the original problem.

The specific features of a problem that lead to an effective sparsity are still unclear. Given that our method is relatively new, further research is needed to evaluate its performance across different scenarios. These topics will be the focus of future research.

Chapter 6

Deterministic Complexity

This chapter includes unpublished material expanding upon the notion of deterministic complexity (DC), as introduced in Ref. [198] and discussed in Sec. 2.5.1. Whereas the DC was first defined for sequences of outputs, here we generalize it to the case of sequences of inputs and outputs from alphabets of arbitrary size.

In contrast with the previous chapters here we veer towards a computational perspective, which offers new insights into how the physics of temporal correlations relates to fundamental limitations in resource-bounded computational and information processing tasks. These topics will be explored in future research.

The chapter is organized as follows. Sec. 6.1 offers a summary of the results, while Secs. 6.2 and 6.3 provide a review of finite-state machines, and the Arrow-of-Time constraints together with its polytope, introduced earlier in the thesis. In Sec. 6.4 we define "Arrow-of-Time functions" in order to reformulate the problem of generating deterministic temporal correlations as computational tasks. In Sec. 6.5 we introduce the generalization of the deterministic complexity to the case of input-outputs, and in Sec. 6.6 we investigate upper bounds on its value. Section 6.7 introduces the partial deterministic complexity (PDC) and the conditional deterministic complexity (CDC), further generalizations of DC based on graph-theoretical principles. Finally, Sec. 6.8 presents an overall discussion of these results and their connections with topics in computer science.

Author contribution: In this work, the doctoral candidate significantly contributed to the conception of the DC, and the development of the proofs in Obs. 7 and 8. The candidate is responsible for writing the majority of the proofs, and is fully responsible for the maximum DC upper bound, the results based on graph theory, the development of the partial and conditional DCs, the writing of all the text within this chapter, and the production of all its graphical assets.

6.1 Summary of results

The general principles underlying the deterministic complexity were introduced by Spee et al. [182] in the context of simulating extremal points of the Arrow-of-Time (AoT) polytope, which correspond to the deterministic conditional distributions $p(\boldsymbol{a}|\boldsymbol{x})$; see Secs. 1.2.6 and 1.2.7. Several properties of the polytope were also investigated in this work (e.g. its symmetries under relabeling, asymptotic behaviors), but a detailed and concrete characterization of arbitrary extremal points was missing.

In Ref. [198], which is the subject of in Ch. 2, these principles were formalized in the context of output sequences $\boldsymbol{a} \in \mathcal{A}^L$ (i.e., without inputs, or with $X = 1$), as a computational complexity measure over the physical realization of a symbolic sequence, one symbol at a time, by means of a deterministic automaton. While this previous work was restricted to studying sequences of outputs, the computational perspective

provided by the deterministic complexity can be immediately generalized to the case of inputs and outputs, with alphabets of arbitrary size.

As we will show, this generalization can be understood as a measure of computational complexity rooted on fundamental physical principles, namely: causality, finite time, finite memory resources, and the physical encoding of information in physical systems.

This chapter formalizes this generalization, and briefly remarks on some of its mathematical, physical, and computational aspects. We first build upon Refs. [182, 198] and generalize Alg. 1 for the input-output case with Alg. 2, allowing an efficient and direct computation of the minimum dimension required for the realization of any extremal point of the AoT polytope.

This is achieved by mapping deterministic distributions obeying the AoT constraints into what we call "Arrow-of-Time Functions" (AoTFs), a subset of functions $F : \mathcal{X}^L \to \mathcal{A}^L$ which satisfy the AoT constraints in their specific way. AoTFs can be related to tree graphs representing a computation tree, providing a neat conceptual description of the problem. While similar notions were already used in Ref. [182], here we shift the focus towards the computational aspects of this problem, and formulate our results in terms of computational and graph-theoretical principles.

We then establish a direct connection between the DC and the problem of finding the *maximum independent vertex set* in graph theory. This graph-theoretical approach is further employed in generalizing the deterministic complexity to partial AoTFs, where certain input-output pairs can be left unspecified, leading to the definition of the partial deterministic complexity (PDC). The PDC is then applied to define the conditional deterministic complexity (CDC), which establishes the memory cost of deterministic generation of an output sequence when an input sequence is given as an additional information resource.

We conclude the chapter with a brief discussion on how the deterministic complexity relates to similar complexity measures in computer science.

6.2 Input-output temporal correlations

We follow the notation introduced in Sec. 1.2.5. Throughout this chapter, let $\mathcal{X}$ and $\mathcal{A}$ be fixed arbitrary input and output alphabets, respectively, with sizes $X = |\mathcal{X}|$ and $A = |\mathcal{A}|$. While in principle it is possible to have a separate output alphabet $\mathcal{A}_x$ associated to each input $x \in \mathcal{X}$, for simplicity we will assume that all inputs share the same output alphabets.

The physical scenario we consider can be conceptualized as a device with a finite-dimensional physical system acting as its internal memory. The device features a "reset" button, which prepares the memory in a fixed initial state (here assumed pure), a display screen, and input buttons labeled $x \in \mathcal{X}$. Upon pressing an input button, a "measurement" is performed on the internal memory and its outcome is displayed on the screen as the output $a \in \mathcal{A}$.

We assume the device is always reset before each sequence of inputs. Thus, the device generates a conditional input-output distribution $p(\boldsymbol{a}|\boldsymbol{x})$ (see Sec. 1.2.6), and its internal dynamics can be modeled as a finite-state transducer, as introduced in Sec. 1.3.

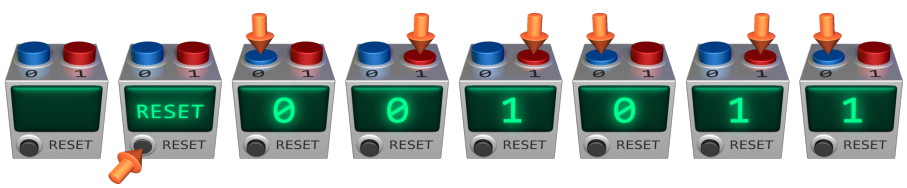

Figure 6.1: A device with a "reset" button, $X = 2$ different inputs, and $A = 2$ possible outputs. In this example, the input sequence $\boldsymbol{x} = 011010$ results in the output sequence $\boldsymbol{a} = 001011$.

Importantly, as discussed in Sec. 1.2.8, we assume the device is isolated, having no access to an external reference clock or to information about past inputs/outputs stored in some external memory resource. The measurements are thus time-independent, with the d-dimensional memory being the sole resource available for generating $p(\boldsymbol{a}|\boldsymbol{x})$.

6.3 Arrow-of-Time constraints and polytope

As discussed in Sec. 1.3, temporal correlations $p(\boldsymbol{a}|\boldsymbol{x})$ can be written in terms of an appropriate FSM model, i.e.,

$$p(\boldsymbol{a}|\boldsymbol{x}) := \pi_0 T_{a_1|x_1} T_{a_2|x_2} \cdots T_{a_L|x_L} \eta = \pi_0 T_{\boldsymbol{a}|\boldsymbol{x}} \eta, \tag{6.1}$$

for a classical model T, and

$$p(\boldsymbol{a}|\boldsymbol{x}) := \mathrm{Tr}\left[\mathcal{I}_{a_L|x_L} \circ \cdots \circ \mathcal{I}_{a_2|x_2} \circ \mathcal{I}_{a_1|x_1}(\rho_0)\right] = \mathrm{Tr}\left[\mathcal{I}_{\boldsymbol{a}|\boldsymbol{x}}(\rho_0)\right], \tag{6.2}$$

for a quantum model $\mathcal{I}$. The probability distributions arising from Eqs. (6.1) and (6.2) automatically satisfy the Arrow-of-Time (AoT) constraints (see Sec. 1.2.6), with the d-dimensional system corresponding to the finite memory resource enabling a limited amount of information to be relayed (or *signaled*) from the past to the future.

The set of temporal correlations allowing only past-to-future signaling forms the Arrow-of-Time polytope (see Sec. 1.2.7), denoted by $\mathcal{P}^L_{X,A}$ for length-L sequences, which is convex and bounded, and therefore can be defined as the convex hull of its extreme points [78]. In particular, the extreme points correspond to deterministic distributions $f_\lambda(\boldsymbol{a}|\boldsymbol{x})$ [1, 88], so that any distribution $p(\boldsymbol{a}|\boldsymbol{x})$ obeying the AoT constraints can be written as a convex mixture:

$$p(\boldsymbol{a}|\boldsymbol{x}) = \sum_\lambda q(\lambda)\, f_\lambda(\boldsymbol{a}|\boldsymbol{x}), \quad \text{with } q(\lambda) \geq 0 \text{ and } \sum_\lambda q(\lambda) = 1. \tag{6.3}$$

Here, the probability distribution $q(\lambda)$ can be interpreted as *initial randomness*, i.e., upon being reset, the device randomly picks a deterministic behavior λ with probability $q(\lambda)$.

A crucial observation is that realizing the deterministic distributions f_λ is not a trivial task, both physically and computationally, and requires a minimum amount of memory (i.e., dimension) to be achieved [88, 182]; see also Sec. 1.2.8. Furthermore, as shown in Ref. [182] and Obs. 2, this amount is the same for either classical and quantum systems, as the deterministic quantum behavior requires no coherence effects. Thus, one can think of Eq. (6.3) as a classical simulation of the set of temporal correlations through the deterministic realization of its extreme points, with initial randomness acting as an extra resource.

In this context, Ref. [182] established a general criterion for the minimum dimension for generating the extreme points f_λ. Further properties of the polytope were also studied, such as its symmetries over classical relabeling of input/outputs, and bounds on the dimension needed for realizing any extreme point, which we also investigate in Sec. 6.6.

Whether or not initial randomness is available as a resource, an exact characterization of the minimum memory resources needed for generating an arbitrary correlation $p(\boldsymbol{a}|\boldsymbol{x})$ poses significant challenges due to the complexity of probabilistic classical or quantum behaviors, especially in the presence of noise [124]. For these reasons, in this chapter we focus on a deeper and more concrete characterization of the deterministic distributions, while revealing some computational implications of the existence of the AoT polytope.

6.4 Arrow-of-Time Functions

Since we will focus exclusively on deterministic distributions, it is helpful to think beyond probabilities and focus on the computational task at hand, namely, that of sequentially mapping inputs $\boldsymbol{x} \in \mathcal{X}^L$ into

outputs $\boldsymbol{a} \in \mathcal{A}^L$, one symbol at a time. This is known as *transduction*.

In particular, we are interested in the subset $\{F_\lambda\}_\lambda$ of functions $F : \mathcal{X}^L \to \mathcal{A}^L$ such that

$$f_\lambda(\boldsymbol{a}|\boldsymbol{x}) = \delta[F_\lambda(\boldsymbol{x}) = \boldsymbol{a}], \tag{6.4}$$

where we use the notation $\delta[\cdot] = 1$ if the statement within brackets is true, and $\delta[\cdot] = 0$ if false. As we require f_λ to obey the AoT constraints, these constraints must also be satisfied by each F_λ. Thus, we introduce the following notion:

Definition 5. (Arrow-of-Time functions, AoTF). An *Arrow-of-Time Function* (AoTF) F of depth L is a function $F : \mathcal{X}^L \to \mathcal{A}^L$ which satisfies the Arrow-of-Time constraints, given by:

$$[F(\boldsymbol{x})]_{1:\ell} = [F(\boldsymbol{x}')]_{1:\ell} \quad \forall\, \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}^L \text{ with } \boldsymbol{x}_{1:\ell} = \boldsymbol{x}'_{1:\ell}, \text{ and } \forall\, \ell = 1, \ldots, L. \tag{6.5}$$

In words, each prefix of inputs $\boldsymbol{x}_{1:\ell}$ must produce the same prefix of outputs $\boldsymbol{a}_{1:\ell}$, for all inputs $\boldsymbol{x} \in \mathcal{X}^L$ and prefix lengths $\ell = 1, \ldots, L$. Due to this structure, any AoTF can be fully specified by a set of tuples $(\boldsymbol{x}, a_{\boldsymbol{x}})$ relating a single output $a_{\boldsymbol{x}} \in \mathcal{A}$ to every prefix $\boldsymbol{x} \in \mathcal{X}^{1:L}$ (see Table 6.1). This naturally motivates investigating AoTFs in terms of a tree structure, as will be discussed next, which is why we will refer to the "depth" of F when referring to the length of its input-output sequences.

$\boldsymbol{x}$	$F(\boldsymbol{x})$
0 0 0	$a_0\ a_{00}\ a_{000}$
0 0 1	$a_0\ a_{00}\ a_{001}$
0 1 0	$a_0\ a_{01}\ a_{010}$
0 1 1	$a_0\ a_{01}\ a_{011}$
1 0 0	$a_1\ a_{10}\ a_{100}$
1 0 1	$a_1\ a_{10}\ a_{101}$
1 1 0	$a_1\ a_{11}\ a_{110}$
1 1 1	$a_1\ a_{11}\ a_{111}$

Table 6.1: An Arrow-of-Time function with depth 3 and $\mathcal{X} = \{0, 1\}$, written as a table. The AoT constraints imply every prefix of inputs produces the same prefix of outputs, such that F can be fully specified by choosing an output $a_{\boldsymbol{x}} \in \mathcal{A}$ for every $\boldsymbol{x} \in \mathcal{X}^{1:L}$.

For a fixed $\mathcal{X}$, $\mathcal{A}$, and L, Eq. (6.4) establishes that AoTFs are in one-to-one correspondence to the deterministic distributions f_λ. Therefore, we may refer to $\mathcal{F}^L_{\mathcal{X},\mathcal{A}}$ as the set of AoT functions, each defining an extreme point in the corresponding AoT polytope. Nevertheless, the computational perspective of AoTFs will be useful in the following sections.

As a side note, although ideas related to AoTFs exist within automata theory (e.g. rational functions and sequential transductions [15]), they are typically explored in broader contexts far removed from the physical interpretations we require. Therefore, we introduce this term in order to emphasize both the physical (AoT) and computational (F) aspects of our particular problem.

6.5 Deterministic complexity with inputs and outputs

An AoTF can be directly interpreted as the computation performed by a deterministic finite-state transducer (DFT, see Sec. 1.3) up to the first L steps. This motivates the following definition.

Definition 6. (Deterministic Complexity of an AoTF). Let $\mathcal{X}$ and $\mathcal{A}$ be, respectively, input and output alphabets of arbitrary size, and let $F \in \mathcal{F}^L_{\mathcal{X},\mathcal{A}}$ be an Arrow-of-Time function of depth L. The *deterministic*

complexity of F, denoted by $\mathrm{DC}(F)$, is the minimum number of states such that there exists a deterministic transducer computing F.

This generalizes the previous output-only definition of deterministic complexity (Def. 3) to the case of inputs and outputs, while also framing it in terms of a computational task. In what follows, we will denote the previous output-only deterministic complexity by DC_{O}, and the general input-output case simply by DC. Just as the DC_{O} was a property of symbolic sequences, DC is a property of symbolic "sequential functions". Thus, while both are motivated by a concrete physical scenario they ultimately pertain to abstract sequential computational tasks.

6.5.1 Computation trees

To obtain the DC of an AoTF, we consider a hypothetical transducer computing F and unravel all its possible behaviors as a *computation tree*, i.e., the tree which *computes* F. As we do not know the minimal transducer in advance, each vertex of the tree represents an *unknown* state of the transducer at a given moment, while edges represent transitions between these unknown states. Importantly, because we require Mealy-type FSMs (see Sec. 1.3.3) outputs are assigned to the edges (transitions), not vertices (states).

Due to the AoT constraints and deterministic transitions, these can be represented as complete X-ary tree graphs, i.e., at every step we have X possible branches, one for each input, each resulting in an independent state transition and corresponding output. We use the formalism of graph theory (see Sec. 5.3.1 for an introduction to the relevant notions), adopting the following definition.

Definition 7. (Computation tree). A depth-L computation tree is a directed graph $G = (V, E, \alpha)$, where $V = \mathcal{X}^{0:L}$ is a set of vertices (the "nodes"), and $E = \{(\boldsymbol{x}_{0:|\boldsymbol{x}|-1}, \boldsymbol{x})\}_{\boldsymbol{x}}$ is a set of edges (the "branches"), indexed by an input prefix $\boldsymbol{x} \in \mathcal{X}^{1:L}$, with corresponding edge labels $\alpha = \{a_{\boldsymbol{x}}\}_{\boldsymbol{x}}$ with $a_{\boldsymbol{x}} \in \mathcal{A}$.

We will refer to these simply as "trees", denoting them by τ. A tree computing F can be constructed by choosing the outputs $a_{\boldsymbol{x}_{1:\ell}} = [F(\boldsymbol{x})]_\ell$, for every input $\boldsymbol{x} \in \mathcal{X}^L$ truncated to every prefix length $\ell = 1, \dots, L$. Therefore, in the following we assume τ and F have the same depth.

Given a tree τ, we will denote its depth by $|\tau|$, and refer to zero-depth trees as *empty*. We will use $\tau(\boldsymbol{x}) \in \mathcal{A}$ to denote the output associated with an input prefix $\boldsymbol{x} \in \mathcal{X}^{1:|\tau|}$. Since every input prefix $\boldsymbol{x} \in \mathcal{X}^{0:L}$ (i.e., including the empty prefix $\boldsymbol{x} = \varepsilon$) uniquely determines a vertex, we define the *subtree* of τ starting at the vertex $\boldsymbol{x}$ by $\tau[\boldsymbol{x}]$, with $\tau[\varepsilon] = \tau$. We denote by τ_ℓ the truncation of the tree τ to depth $\ell \leq |\tau|$, i.e., the tree where every vertex and edge with input prefix longer than ℓ is discarded, thus $|\tau_\ell| = \ell$. Naturally, truncating to the tree's own depth has no effect, i.e., $\tau_{|\tau|} = \tau$. The trees τ and τ' are equal if: (i) they share input and output alphabets (ii) they have the same depth, and (iii) $\tau(\boldsymbol{x}) = \tau'(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}^{1:|\tau|}$. Throughout this chapter we will always assume trees share alphabets. We will also require the following notion.

Definition 8. (Consistent trees). Let τ, τ' be two trees, and let $\ell = \min(|\tau|, |\tau'|)$. The two trees are said to be *consistent*, denoted by $\tau \sim \tau'$, if $\tau(\boldsymbol{x}) = \tau'(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}^{1:\ell}$ or, equivalently, if the two trees are equal up to truncation, i.e., $\tau_\ell = \tau'_\ell$.

Observation 5. Consistency relations between trees are reflexive ($\tau \sim \tau$) and symmetric ($\tau \sim \tau' \implies \tau' \sim \tau$), but they are *not* transitive: $\tau \sim \tau'$ and $\tau' \sim \tau''$ do not imply $\tau \sim \tau''$. Therefore $\sim$ is not an equivalence relation.

Proof. This is due to truncation. Suppose $|\tau| = |\tau''| = 2$, $|\tau'| = 1$, $\tau \neq \tau''$, and $\tau_1 = \tau'_1 = \tau''_1$. Then $\tau \sim \tau'$ and $\tau' \sim \tau''$ due truncation to length 1, but $\tau \nsim \tau''$. □

Observation 6. Given two trees τ and τ', if $\tau_\ell \nsim \tau'_\ell$ for some $\ell \leq \min(|\tau|, |\tau'|)$, then $\tau \nsim \tau'$.

Since branching is defined by the inputs, and an output must be associated to each transition, we may use the spatial disposition of the vertices in a diagram to represent inputs, and use edge labels to specify outputs. An example of such diagram is shown in Fig. 6.2, where we take the transducer as starting at the *root*, at the top, with sequences of inputs/outputs (and therefore time) read downwards. For this reason, the direction of edges is left implicit.

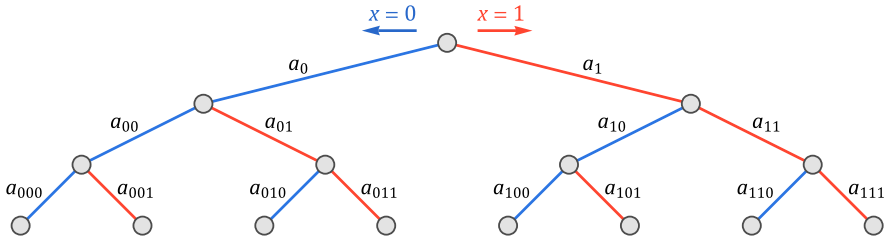

Figure 6.2: Computation tree for $\mathcal{X} = \{0, 1\}$ showing all possible behaviors of a transducer after $L = 3$ steps. Vertices (disks) correspond to the unknown states at each step, each associated to a unique input prefix $\boldsymbol{x} \in \mathcal{X}^{0:L}$ determining the "path" towards it. Each edge and its corresponding label (i.e., outputs $a_{\boldsymbol{x}} \in \mathcal{A}$) can be associated with an input prefix $\boldsymbol{x} \in \mathcal{X}^{1:L}$. Outputs for a given input sequence $\boldsymbol{x} \in \mathcal{X}^L$ are read one by one, starting at the root (on top) and collecting outputs downwards, moving left for $x = 0$ and right for $x = 1$, as depicted. In this example, the input sequence $\boldsymbol{x} = 010$ results in the output sequence $\boldsymbol{a} = a_0\, a_{01}\, a_{010}$. Edge colors follow the button colors in Fig. 6.1.

Note that the same notion of a tree was also utilized in Ref. [182], although from a physical perspective. Here, we make the distinction between the function F (the computation being performed) and a tree representing the behavior of a FSM computing F. We briefly elaborate on this distinction in the following section.

6.5.2 Computing the deterministic complexity

A tree τ represents *what* is required for any transducer to compute a given F, but it does not tell us *how* the computation is performed. Since the tree vertices correspond to unknown transducer states, a concrete description of a transducer's behavior can be obtained by assigning definite states $s_i \in \mathcal{S}$ to each vertex $i \in V$ (and its corresponding subtree $\tau[i]$), where $V := \mathcal{X}^{0:L}$. We will denote such assignments by $\sigma = (s_i)_{i \in V}$.

Importantly, this assignment must be consistent. If two vertices $i, j \in V$ are to represent the same state $s_i = s_j$, then a necessary condition is that $\tau[i] \sim \tau[j]$. The converse does not hold, however: $\tau[i] \sim \tau[j]$ might be true due to truncation, even though $s_i \neq s_j$. In such cases, the states s_i and s_j may also appear elsewhere in τ, assigned to larger subtrees. The example in Obs. 5 illustrates such a scenario. In our notation, this can be formalized as follows.

Definition 9. (Consistent assignment). Let $\sigma = (s_i)_{i \in V}$ be an assignment of states $s_i \in \mathcal{S}$ for the vertices of τ, and let $v(s) := \{\, i \in V \mid s_i = s \,\}$ denote the set of vertices assigned to a state $s \in \mathcal{S}$. The assignment σ is *consistent* if $\tau[i] \sim \tau[j]$ for all $i, j \in v(s)$ and all $s \in \mathcal{S}$.

Definition 10. (Optimal assignment). A consistent assignment is *optimal* if it uses the minimum number of states.

Proposition 4. $\mathrm{DC}(F)$ corresponds to the number of states in an optimal assignment for τ.

Proposition 5. Let $\sigma = (s_i)_{i \in V}$ be an optimal assignment and $v(s)$ the set of vertices assigned to $s \in \mathcal{S}$. Then for each state s we can pick a vertex $i \in v(s)$ constructing a set $S \subseteq V$, with $|S| = |\mathcal{S}|$, such that for every distinct $i, j \in S$ we have $s_i \neq s_j \iff \tau[i] \nsim \tau[j]$.

Proof. If the assignment is consistent, then for any two vertices $i, j \in V$ with $\tau[i] \nsim \tau[j]$ we have $s_i \neq s_j$. If the assignment is also optimal, then for every $s, s' \in \mathcal{S}$ with $s \neq s'$ there must exist vertices $i, j \in V$ with $s_i = s$ and $s_j = s'$ such that $\tau[i] \nsim \tau[j]$, otherwise the states are indistinguishable and could be merged into a single state, contradicting the assumption of optimality. Furthermore, by Obs. 6, for every $i, j \in V$ with $\tau[i] \nsim \tau[j]$ and $s_i = s$, there will exist a vertex $i' \in v(s)$ such that $|\tau[i]| \leq |\tau[i']|$ and $\tau[i'] \nsim \tau[j]$. Thus, we may pick $\tau[i']$ as a "representative tree" for the state s, and collect all vertices of representative trees into a set S.

□

Intuitively, an assignment is optimal if every state $s \in \mathcal{S}$ is assigned to a least one subtree (i.e., no state is unused) and every pair of distinct states can be distinguished using their corresponding subtrees. If this was not the case, the unused state could have been removed, or the two indistinguishable states could be merged into a single one, in either case reducing the total number of states. We conclude that an optimal assignment not only uses the least number of states (no state can be removed), but also has the maximum possible number of nonequivalent states (any extra state is redundant).

These criteria are essentially the same as the ones introduced by Spee et al. [182], as the number of "nonequivalent futures" determining the minimum dimension required to simulate an extreme point of the AoT polytope. Interestingly, despite the finite-length deterministic scenario never being considered, the same principles are also used when obtaining causal states of ε-machines (and ε-transducers) for modeling stochastic processes [55, 119]. In this chapter, we adopt the computational perspective on these ideas in order to generalize them further.

An algorithm for computing $\mathrm{DC}(F)$ is given in Alg. 2. Following Prop. 5, the algorithm corresponds to a breadth-first search of the tree τ, where each subtree is matched against a list of previous subtrees corresponding to nonequivalent states. If a match is found, the matched state is assigned to the current vertex, whereas if no matches are found the subtree at the current vertex is added to the list of known states.

Figure 6.3 illustrates the result of applying Alg. 2. Note that multiple assignments are possible (e.g., the vertex $\boldsymbol{x} = 010$ could have also been assigned states 2, 3, 4, or 5), but the algorithm returns only one. The algorithm can easily be adapted to enumerate all possible assignments—and, therefore, all minimal transducers—by finding all matching states instead of stopping at the first match (line 18).

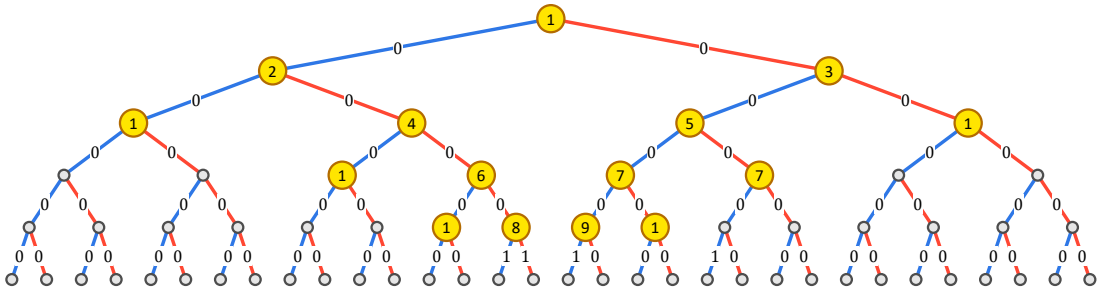

Figure 6.3: Example of a tree for an AoTF F, with $X = A = 2$ and $L = 5$, showcasing $\mathrm{DC}(F) = 8$ through an optimal assignment. The state assignments are shown in yellow and are labeled sequentially, in the order found by Alg. 2. Note that we omit redundant assignments, as once a subtree can be assigned to a previous state its subtrees are completely determined.

Algorithm 2 Deterministic complexity algorithm for an Arrow-of-Time function F.

Input: A tree τ computing F.

Outputs: The $\mathrm{DC}(F)$, a transducer model T, state assignments, and corresponding trees.

1: **procedure** FINDDC(τ)
2: $states \leftarrow \{ (1, \tau) \}$ ▷ Label and subtree of states found so far. τ itself is always the first state.
3: $dc \leftarrow 1$ ▷ Current number of states known.
4: $assignments \leftarrow \{ (\varepsilon, 1) \}$ ▷ Initialize list of $(\boldsymbol{x}, s)$ tuples, the state assignments at each prefix $\boldsymbol{x}$.
5: $transitions \leftarrow \{\}$ ▷ List of state transitions found, initially empty.
6: $queue \leftarrow \{\}$ ▷ Initialize a queue of candidate states.
7: **for** x **in** $\mathcal{X}$ **do** ▷ We add the first subtrees to the queue.
8: $queue.enqueue(\ (\varepsilon, x, \tau(x), 1, \tau[x])\ )$ ▷ Prefix, input, output, and state leading to the subtree $\tau[x]$.
9: **end for**
10: **while** $queue$ **not empty do** ▷ Continue for as long as we have trees to check.
11: $(\boldsymbol{y}, x, a, s, \tau) \leftarrow queue.dequeue()$ ▷ Get next element from queue as candidate tree, replacing τ.
12: $match \leftarrow$ False
13: **for** (s', τ') **in** $states$ **do** ▷ Scanning over the list of known states...
14: **if** $\tau \sim \tau'$ **then** ▷ ... we check if the current tree is consistent with a known state.
15: $match \leftarrow$ True ▷ A known state has been found.
16: $transitions.append(\ (x, a, s, s')\ )$ ▷ Save the transition $s \to s'$ with input x and output a.
17: $assignments.append(\ (\boldsymbol{y}x, s')\ )$ ▷ Assign the state s' for the vertex at prefix $\boldsymbol{x} = \boldsymbol{y}x$.
18: **break** ▷ We can stop looking for matches.
19: **end if**
20: **end for**
21: **if not** $match$ **then** ▷ If τ did not match any known state, it is a new state.
22: $dc = dc + 1$ ▷ Update total number of states. The new state will be labeled by dc.
23: $states.append(\ (dc, \tau)\ )$ ▷ Add current subtree τ and its label to our list of known states.
24: $transitions.append(\ (x, a, s, dc)\ )$ ▷ Save the transition $s \to dc$ with input x and output a.
25: $assignments.append(\ (\boldsymbol{y}x, dc)\ )$ ▷ Assign the state dc for the vertex at prefix $\boldsymbol{x} = \boldsymbol{y}x$.
26: **for** x **in** $\mathcal{X}$ **do** ▷ Add all subtrees of τ to our queue, searching the next depth.
27: $queue.enqueue(\ (x, \tau(x), dc, \tau[x])\ )$
28: **end for**
29: **end if**
30: **end while**
31: ▷ With queue empty, we are done. Assemble the final model T.
32: $T \leftarrow 0^{X \times A \times dc \times dc}$ ▷ Initialize a $X \times A \times dc \times dc$ transition model with zeros.
33: **for** (x, a, i, j) **in** $transitions$ **do** ▷ Construct the model following the transition structure found earlier.
34: $[T_{a|x}]_{ij} \leftarrow 1$
35: **end for**
36: **return** $dc, T, assignments, states$
37: **end procedure**

Similar to the results in Sec. 2.5.2, minimal transducers will not generally be unique, but all will have the same number of states. Furthermore, each minimal transducer computing a given F will also define an infinite family of AoTFs, obtained by extending the sequences beyond the original depth L. This is analogous to the result on patterns discussed in Sec. 2.5.2, where we can interpret trees, together with state assignments, as the generalization of patterns to the input-output case. For these reasons, we distinguish between AoTFs and their corresponding computation trees, as they represent distinct objects, e.g., a tree of depth L also computes functions of depths $\ell < L$.

To finish this section, we show that Alg. 2 results in the minimal number of states.

Observation 7. (No undercounting). Algorithm 2 outputs a valid model for its input tree.

Proof. After each new state is found, its direct descendants, corresponding to subtrees, are added to the queue (lines 21-28), which is recursively parsed until the entire tree is exhausted. Whether these descendants refer to a new state or previous states, in both cases the existing transition in the input tree is

accounted for in the final model (lines 16 and 24). Thus, the model T returned by the algorithm shares the same input-output structure as the entire tree. □

Observation 8. (No overcounting). Algorithm 2 outputs a model with the smallest number of states.

Proof. The proof is by contradiction. Let τ be the input tree, and $M = (d, T, \mathcal{K}, \mathcal{S})$ be the output of Alg. 2, with $d = |\mathcal{S}|$. Now, suppose there exists a tuple $M' = (d', T', \mathcal{K}', \mathcal{S}')$, with $d' = |\mathcal{S}'|$, corresponding to the same tree, but for which $d' < d$, i.e., M' describes the same tree, albeit with a simpler model with less states.

Let $s_i \in \mathcal{S}$ and $s'_i \in \mathcal{S}'$ refer to the i-th unique states in M and M', respectively, and σ_i and σ'_i their corresponding subtrees, obtained by traversing the tree according to the appropriate paths as described by the assignments $\mathcal{K}$ and $\mathcal{K}'$. Without loss of generality, we may refer to $s_1 = s'_1$ as the initial states in $\mathcal{S}$ and $\mathcal{S}'$ respectively, such that $\sigma_1 = \sigma'_1 = \tau$, since the initial states must correspond to the entire tree. Additionally, since the transition structure is entirely encoded in the models T and T', which are deterministic, we have that state labels are arbitrary, and we may also order s_i and s'_i in shortlex order[1] based on the shortest prefixes in $\mathcal{K}$ and $\mathcal{K}'$ assigned to s_i and s'_i. (Note: The parsing order of Alg. 2 automatically follows this ordering.)

If that is the case, then there is an input prefix $\boldsymbol{x} \in \mathcal{X}^{0:L}$ such that the subtree $\tau[\boldsymbol{x}]$ corresponds to a new state $s_k \in \mathcal{S}$ in M (as described by the model T), but to a previous state $s'_\ell \in \mathcal{S}'$ in M' (according to the model T'), with $\ell < k$. However, since Alg. 2 parses the tree breadth-first and lexicographically, this earlier state must have also been previously found in M as well, which contradicts the assumption that $\tau[\boldsymbol{x}]$ corresponds to a new state s_k. □

6.6 Maximum deterministic complexity

At this point, a most natural question to ask is: What is the maximum DC over all AoTFs for a given X, A, and L? We will refer to this quantity by $M(X, A, L)$, or simply M, and use $\underline{m}$ and $\overline{m}$ for its lower and upper bounds, respectively.

Knowledge of M has interesting implications. Computationally, it establishes one notion of maximum complexity for finite computational tasks. Physically, it quantifies the minimum dimension required to realize any extreme point in the corresponding AoT polytope $\mathcal{P}^L_{A,X}$, under either classical or quantum theory, so that a device with a classical d-it of memory and some initial randomness can generate *any* temporal correlation $p(\boldsymbol{a}|\boldsymbol{x})$ provided $d \geq M$.

A discussion on M can also be found in Spee et al. [182], where two methods were provided for obtaining lower bounds, as well as results estimating its asymptotic behavior. There, it was found that M scales at least exponentially in L. This section offers the complementary discussion. We provide a method that attempts to compute M directly, but ultimately provides an upper bound $\overline{m}$. The idea follows from a non-constructive combinatorial counting argument involving the total number of possible trees of a given size. Consider an arbitrary tree for a choice of parameters (X, A, L). As each layer ℓ of the tree contains X^ℓ vertices, the total number of vertices n_V and edges n_E are

$$n_V(L) = \sum_{\ell=0}^{L} X^\ell = \frac{X^{L+1} - 1}{X - 1}, \tag{6.6}$$

$$n_E(L) = n_V(L) - 1 = \frac{X(X^L - 1)}{X - 1}. \tag{6.7}$$

Since to each edge we may freely assign an output $a \in \mathcal{A}$, there are a total of

$$N_L = A^{n_E(L)} = A^{\frac{X(X^L-1)}{X-1}} \tag{6.8}$$

[1] Sorting by length, then lexicographically, e.g.: 0, 1, 00, 01, 10, 11, 000, 001, ...

distinct trees, a number that grows doubly-exponentially in L. This is precisely the number of extreme points in the AoT polytope [88, 182]. Note that many of the trees will be isomorphic under relabeling of inputs and outputs, what can be understood as a form of classical post-processing. A detailed analysis of these symmetries is available in Ref. [182]. That being said, for our purposes these symmetries will not be relevant.

As we have seen, computing $\mathrm{DC}(F)$ reduces to finding an optimal assignment of states to each vertex (and its subtree) in the corresponding tree for F, such that every distinct pair of states can be distinguished through an inconsistency between their associated subtrees. Furthermore, since the last vertices in a tree (its "leaf" vertices) correspond to empty subtrees their state assignments are arbitrary, reflecting the fact that the final transducer states are irrelevant when computing F. We can then state our first result.

Proposition 6. $M(X, A, L)$ is strictly upper bounded by the number of non-leaf vertices in the corresponding trees for $F \in \mathcal{F}^{L}_{\mathcal{X},\mathcal{A}}$, i.e.:

$$M(A, X, L) := \max_{F} \left\{ \mathrm{DC}(F) \;\middle|\; F \in \mathcal{F}^{L}_{\mathcal{X},\mathcal{A}} \right\} \leq n_V(L-1) = \frac{X^L - 1}{X - 1}. \tag{6.9}$$

This trivial bound can be saturated if a unique depth-1 subtree is assigned to each of the X^{L-1} second-to-last vertices. Since there are A^X depth-1 trees, the bound can be reached whenever $A^X \geq X^{L-1}$. However, since this bound grows exponentially with depth it is quickly rendered useless, motivating the search for tighter upper bounds.

For the remaining of this section, let $\hat{\tau}$ be the depth-L tree for an AoTF $\hat{F} \in \mathcal{F}^{L}_{\mathcal{X},\mathcal{A}}$ such that $\mathrm{DC}(\hat{F}) = M$. We will not attempt to construct $\hat{\tau}$ directly. Instead, our goal will be to count the maximum number of subtrees of $\hat{\tau}$ that can be made mutually-inconsistent, which is equivalent to the $\mathrm{DC}(\hat{F})$. We will require the following notion.

Definition 11. (Tree intersection, intersection depth). Given two trees τ and τ', their intersection $\tau \cap \tau'$ is the largest tree they have in common through truncation. More precisely, we can write

$$|\tau \cap \tau'| := \max \left\{ \ell \in \mathbb{N}_{\geq 0} \;\middle|\; \tau_\ell = \tau'_\ell \right\}, \tag{6.10}$$

where $|\tau \cap \tau'|$ is the *intersection depth* between the two trees.

With the notion of truncation and intersection, an inconsistency relation $\tau \nsim \tau'$ between two trees partitions their layers into three regions, as shown in Fig. 6.4. At the top we have $|\tau \cap \tau'| \geq 0$ layers corresponding to their intersection, followed by $\min(|\tau|, |\tau'|) - |\tau \cap \tau'| > 0$ layers where the trees can be found to be inconsistent. Lastly, we have the remaining layers of the larger tree, irrelevant due to truncation (if any).

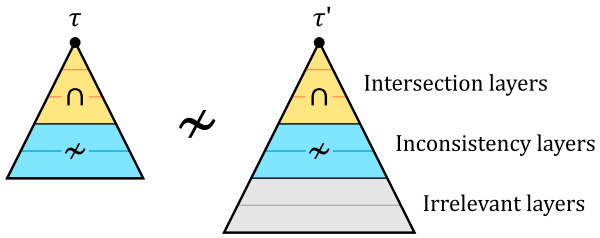

Figure 6.4: Diagrammatic representation of the inconsistency relation between two trees, here depicted as layered triangles, where we identify three separate regions. These diagrams and ideas will be useful in the following sections.

6.6.1 Outline of the counting argument

The idea for computing M non-constructively is to consider the behavior of Alg. 2 when computing $\mathrm{DC}(\hat{F})$. The algorithm will parse the subtrees $\hat{\tau}[i]$ at each vertex $i \in V$ of $\hat{\tau}$ breadth-first, i.e., layer by layer, starting at the root, while keeping track of all subtrees it has seen before. The current subtree $\hat{\tau}[i]$ is assigned a new state if it is inconsistent with all of the previous subtrees, otherwise it is assigned the state associated with a previous tree consistent with it. Therefore, in the case of $\hat{\tau}$, as many subtrees as possible must be mutually-inconsistent throughout this breadth-first parsing order.

If we knew $\hat{\tau}$ in advance, the parsing order of Alg. 2 provides a correct sequence of M mutually-inconsistent subtrees $\Delta^* := (\hat{\tau}[i_1], \hat{\tau}[i_2], \ldots, \hat{\tau}[i_M])$ which all fit together within $\hat{\tau}$. Since we *don't* know $\hat{\tau}$, we can instead imagine the finite sets of trees from which we *could* iteratively build the correct (but unknown) sequence Δ^*. This motivates a counting argument: pretend we are building Δ^* sequentially as Δ, and for each new state k we calculate the remaining number of trees R we can pick which are inconsistent with all previous trees currently in Δ.

A reasonable strategy is to treat each layer of vertices $\ell \in \{0, \ldots, L-1\}$ iteratively, starting at the root at $\ell = 0$, and assume a new state can be assigned to a vertex (and its subtree) as long as $R > 0$. This assumption is not entirely correct, as all subtrees must also be internally consistent within $\hat{\tau}$, but it will generally provide a good upper bound. Thus, once we find that no trees are available for a step $k = K+1$, we will conclude $\mathrm{DC}(\hat{F}) = M \leq K$.

For layer $\ell = 0$, we will have $R = N_L$ possible choices of trees from which we pick one (i.e., X^0): $\hat{\tau}[i_1] = \hat{\tau}$. In the next layer $\ell = 1$, we attempt to pick a total of X^1 trees, all mutually inconsistent. There are a total of N_{L-1} trees of depth $L-1$, but one will be excluded for being consistent with $\hat{\tau}[i_1]$, namely, $\hat{\tau}[i_1]$ truncated to depth $L-1$. Thus, there are $R = N_{L-1} - 1$ trees available for $\hat{\tau}[i_2]$. Picking one, we are left with $R = N_{L-1} - 2$ trees for $\hat{\tau}[i_3]$, and so on, repeating this process until all trees have been picked for $\ell = 1$ before moving on to the next layer.

Here, complications begin to arise. Starting from a total of N_{L-2} possible subtrees at layer $\ell = 2$, we need to exclude all subtrees consistent with the $|\Delta| = X^0 + X^1 = X + 1$ previous subtrees in the sequence $\Delta = (\hat{\tau}[i_1], \hat{\tau}[i_2], \ldots, \hat{\tau}[i_{X+1}])$. But using $R = N_{L-2} - |\Delta|$ underestimates the number of remaining trees, as it assumes every tree $\tau \in \Delta$ results in a distinct tree when truncated to depth $L-2$. It is perfectly possible that all trees in Δ would count as a single tree if they all share a sufficiently large intersection depth, leading to $R = N_{L-2} - 1$ instead.

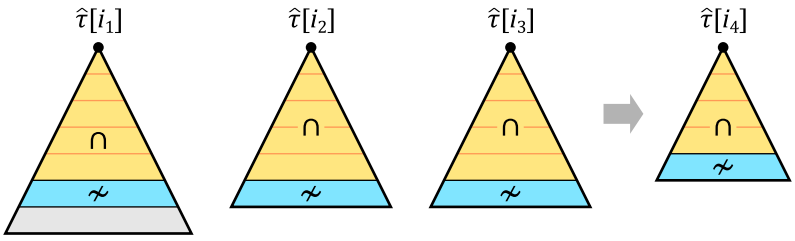

Figure 6.5: Suppose $X = 2$ and $L = 7$. Starting at layer $\ell = 2$, we would have already picked $\Delta = (\hat{\tau}[i_1], \hat{\tau}[i_2], \hat{\tau}[i_3])$, which could mutually have an intersection depth of 5. Therefore, all of them could count as a single tree to be excluded when picking $\hat{\tau}[i_4]$. It is then necessary to always maximize the intersection depth between trees in Δ in order to minimize the number of excluded trees at every step.

Importantly, if we wish to obtain and upper bound for M it is necessary to exclude the *minimum* possible number of trees from the total $N_{L-\ell}$, ensuring we always can pick a new tree if $k < M$. Therefore, all trees in Δ must have the largest intersection depth as possible at each step. Figure 6.5 illustrates these

ideas. For each vertex of layer ℓ, the exact number of trees we could pick from is then given by:

$$R = N_{L-\ell} - |\{\tau_{L-\ell} \mid \tau \in \Delta\}|, \tag{6.11}$$

where $|\{\tau_{L-\ell} \mid \tau \in \Delta\}|$ corresponds to the minimum number of distinct trees obtained by truncating every tree in Δ to the current depth $L - \ell$.

That being said, we do not know the correct Δ^* and have no prior knowledge about the structure of its subtrees $\hat{\tau}[i_k]$. As mentioned earlier, this counting argument assumes a new tree (and therefore a new state) will be picked at the current layer as long as $R > 0$, but this will not always be the case in the correct Δ^*. As all subtrees must be internally consistent, fitting within $\hat{\tau}$, picking a subtree one at a time is not something that can be done independently from the overall structure of $\hat{\tau}$. This counting argument ignores these additional constraints.

6.6.2 Upper bounds

We can circumvent the issues discussed in the previous section if we use an approach which is guaranteed to overestimate the number of remaining trees. A trivial way to do this is to simply assume a maximal intersection depth for all trees from the previous layers $\ell' < \ell$, as in Fig. 6.5, which corresponds to assuming $|\{\tau_{L-\ell} \mid \tau \in \Delta\}| = 1$. This gives, at each step,

$$R = N_{L-\ell} - 1 - n, \tag{6.12}$$

where n is the number of trees already picked at the current layer ℓ. This assumption is unrealistic, but it will never undercount the remaining number of trees. We can establish the following result.

Proposition 7. An upper bound for M is given by:

$$M(X, A, L) \leq \overline{m}(X, A, L) = \sum_{\ell=0}^{L-1} \min\left(X^\ell, N_{L-\ell} - 1\right) \tag{6.13}$$

Proof. Let L be the depth of $\hat{\tau}$ and consider we are picking a new tree at layer ℓ, with $\Delta_{<\ell}$ the sequence of trees picked for the previous layers $\ell' < \ell$, i.e., Δ excluding the n trees picked at the current layer. Assuming a maximum intersection depth for all trees in $\Delta_{<\ell}$, we have $|\tau \cap \tau'| = L - \ell$ for all $\tau, \tau' \in \Delta_{<\ell}$, implying $\tau_{L-\ell} = \tilde{\tau}$ for every $\tau \in \Delta_{<\ell}$, i.e., $\tilde{\tau}$ is their (assumed) common intersection (e.g., see Fig. 6.5). Furthermore, since we require every new tree to be inconsistent with the previous ones, we require $\hat{\tau}[i_k] \nsim \tau$ for every $\tau \in \Delta_{<\ell}$, and that all n trees picked for the current layer ℓ will be distinct. Since the exact number of remaining trees at any step at a layer ℓ is given by Eq. (6.11), we find:

$$\begin{aligned} R &= N_{L-\ell} - |\{\tau_{L-\ell} \mid \tau \in \Delta\}| \\ &= N_{L-\ell} - |\{\tau_{L-\ell} \mid \tau \in \Delta_{<\ell}\}| - n \\ &= N_{L-\ell} - 1 - n. \end{aligned} \tag{6.14}$$

We can pick a new tree $\hat{\tau}[i_k]$ (and therefore a new state) as long as $R > 0$. Therefore, the number n of trees we can pick at layer ℓ will be bounded by $n < N_{L-\ell} - 1$. Finally, since we pick at most X^ℓ trees per layer, we have $\min\left(X^\ell, N_{L-\ell} - 1\right)$ trees picked per layer ℓ. Summing over all layers $\ell = 0, \ldots, L-1$ completes the proof. □

Improved upper bounds are possible, but we leave this discussion for future research.

6.6.3 Maximal trees and comparison with lower bounds

Table 6.2 compares the lower bounds computed with the two approaches described in Spee et al. [182] with our upper bounds computed with Eq. (6.13). Figure 6.6 displays several maximal trees for $L = 3$ to 6.

L	$\underline{m}_1(2,2,L)$	$\underline{m}_2(2,2,L)$	$\overline{m}(2,2,L)$
1	1	1	1 (exact)
2	2	1	3 (exact)
3	4	3	6 (exact)
4	4	7	10 (exact)
5	8	15	18 (exact)
6	16	15	34
7	32	31	66
8	64	63	129
9	64	127	193
10	128	255	321

L	$\underline{m}_1(2,2,L)$	$\underline{m}_2(2,2,L)$	$\overline{m}(2,2,L)$
11	256	511	577
12	512	1 023	1 089
13	1 024	2 047	2 113
14	2 048	4 095	4 161
15	4 096	4 095	8 257
16	8 192	8 191	16 449
17	16 384	16 383	32 832
18	16 384	32 767	49 216
19	32 768	65 535	81 984
20	65 536	131 071	147 520

Table 6.2: Comparison between $\overline{m}(2,2,L)$ obtained with Eq. (6.13) and the lower bounds $\underline{m}_1(2,2,L)$ and $\underline{m}_2(2,2,L)$ following the arguments in Ref. [182]. For $L \leq 5$, we can verify $M(2,2,L) = \overline{m}(2,2,L)$ is exact by the trees given explicitly in Fig. 6.6.

6.7 Generalizations

The computational and graph-theoretical perspectives offered by the DC motivate several generalizations, which have connections to existing concepts in computer science. These connections will be discussed later. For now, we briefly introduce the generalizations.

6.7.1 Consistency graphs and independence number

In the following, let $G = (V, E)$ be a graph defined over vertices V and edges E. We assume undirected graphs, such that $(i, j) \in E$ implies $(j, i) \in E$. We begin by introducing a few graph-theoretical concepts.

Definition 12. (Independent sets). An *independent set* of G is a subset of vertices $I \subseteq V$ such that $(i, j) \notin E$ for all distinct $i, j \in I$. An independent set is *maximal* if it is not a proper subset of any other independent set, and it is a *maximum independent set* if it contains the largest possible number of vertices.

Definition 13. (Independence number). Let $I \subseteq V$ be a maximum independent set. The *independence number* of G is the cardinality of I, and is denoted by $\alpha(G) := |I|$.

Independent sets are complementary to graph cliques: an independent set in G is a clique in its graph complement[2] $\overline{G}$. The problem of finding a maximum independent set is therefore equivalent to finding a maximum clique. Both problems are NP-Hard for arbitrary graphs, but for restricted graphs one formulation may offer an advantage over the other [77, 176].

The generalizations of DC will be based on the following graph structure.

Definition 14. (Consistency graph). Let F be an AoTF and τ its corresponding tree. The *consistency graph* of F is the undirected graph $G = (V, E)$ with vertices $V = \mathcal{X}^{0:L-1}$, each $i \in V$ corresponding to a non-empty subtree $\tau[i]$, and edges $E = \{\ (i, j) \in V \times V \ \mid\ \tau[i] \sim \tau[j]\ \}$.

In words, the consistency graph represents all pairwise consistencies between non-empty subtrees of τ. We can now establish the following result.

[2] Given a graph $G = (V, E)$, its complement is the graph $\overline{G} = (V, \overline{E})$ with $\overline{E} := (V \times V) \setminus E$.

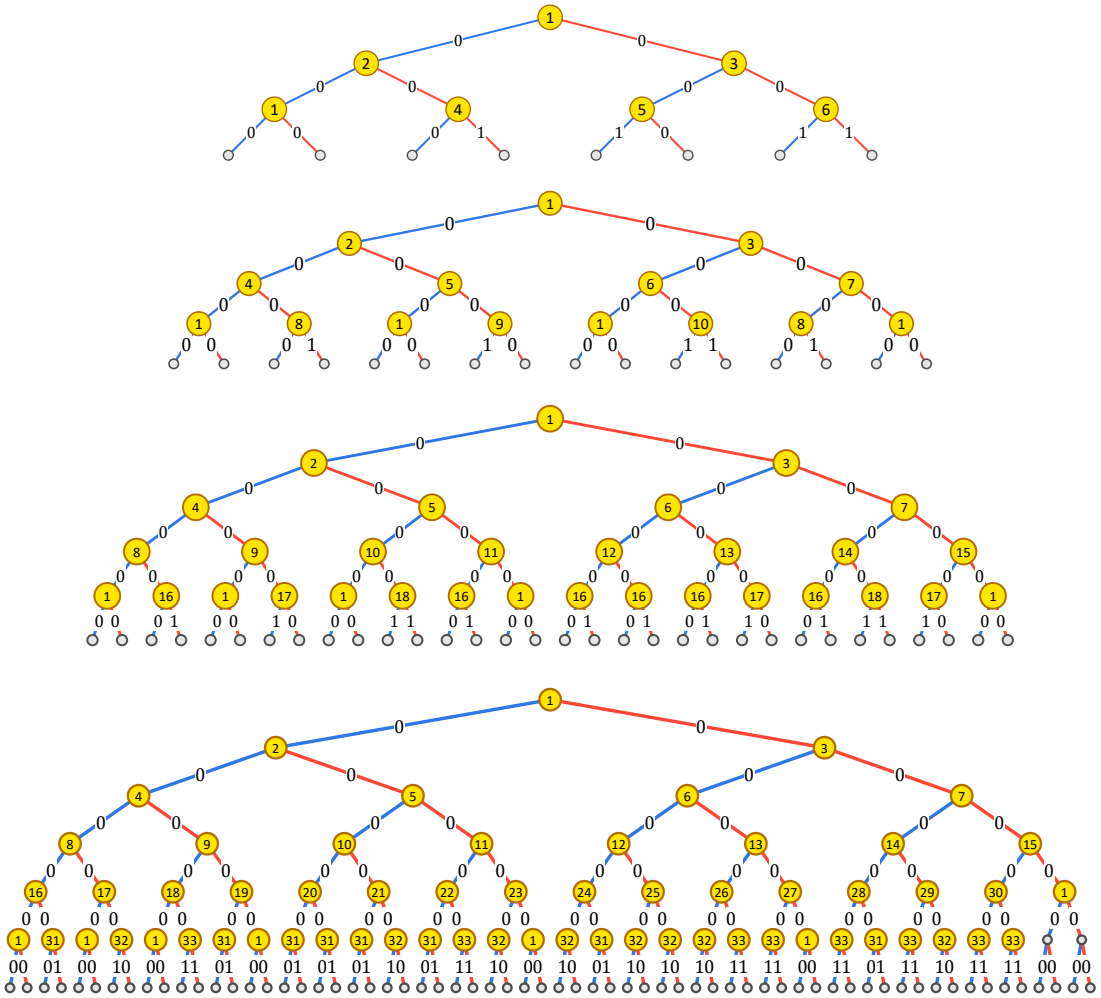

Figure 6.6: Examples of trees with maximum deterministic complexity, corresponding to (from top to bottom): $M(2,2,3) = 6$, $M(2,2,4) = 10$, $M(2,2,5) = 18$ and $M(2,2,6) \geq 33$. Yellow vertices with labels correspond to the distinct state assignments found with Alg. 2.

Theorem 5. Let F be an AoTF and G its consistency graph. Then:

$$\mathrm{DC}(F) = \alpha(G). \tag{6.15}$$

Proof. Let $I \subseteq V$ be a maximum independent set, with $\alpha(G) = |I|$, and let $\mathcal{S}$ be the set of states in the minimal transducer, with $\mathrm{DC}(F) = |\mathcal{S}|$. By definition: $i, j \in I \implies (i,j) \notin E \iff \tau[i] \nsim \tau[j]$. Thus, $s_i \neq s_j$ and $\alpha(G) \leq \mathrm{DC}(F)$. To prove the equality, recall that by Props. 4 and 5 DC is obtained by finding an optimal assignment. Therefore, there exists a subset of vertices $S \subseteq V$, with $|S| = |\mathcal{S}|$, such that for every distinct $i, j \in S$ we have that $s_i \neq s_j$ and $\tau[i] \nsim \tau[j]$, so that $(i,j) \notin E$. In other words, S is an independent set. To show that is a maximum independent set, note that by definition $|S| \leq |I|$ as I is a maximum independent set. Since $\mathrm{DC}(F) = |S|$ and $\alpha(G) = |I|$, we have proven $\mathrm{DC}(F) \leq \alpha(G)$. Together with the previous result, we conclude $\mathrm{DC}(F) = \alpha(G)$. □

While this result is interesting, it is not directly useful, since the DC algorithm in Alg. 2 is far more straightforward and efficient than attempting to first compute G, then $\alpha(G)$. Nevertheless, the result establishes a basis for the following generalizations.

6.7.2 Partial deterministic complexity

Until now we only considered total AoTFs, where a definite output is assigned for every input. A natural generalization is to consider partially-specified AoTFs, where the output for certain input prefixes is left unspecified. There are multiple ways to address this. One possibility is to keep the assumption of an output being generated deterministically at every step, but an unspecified output in F simply means we do not care about it. In this case, we can assign a wildcard output $*$ to any unspecified input prefix, acting as a "don't care" instruction.

Definition 15. (Partial Arrow-of-Time function). Let $*$ denote a special wildcard symbol, i.e., such that $* = a$ evaluates to *true* for any $a \in \mathcal{A}$. A depth L function $F : \mathcal{X}^L \to (\mathcal{A} \cup \{*\})^L$ obeying the AoT constraints is said to be a *partial Arrow-of-Time function* if a wildcard output is assigned to at least one prefix $\boldsymbol{x} \in \mathcal{X}^{1:L}$.

We will refer to the non-wildcard behavior of an AoTF as its *definite portion*. Note that we consider $* \notin \mathcal{A}$ as to preserve the notion of consistent trees. Instead, we introduce the following generalization.

Definition 16. (Partially consistent trees). Two trees τ and τ' are *partially consistent*, denoted by $\tau \sim_* \tau'$, if they are consistent possibly due to wildcard comparisons.

Consistency graphs can be naturally generalized using $\sim_*$.

Definition 17. (Partial deterministic complexity, PDC) The *partial deterministic complexity* (PDC) of F, denoted by $\mathrm{PDC}(F)$, is the minimum number of states such that there exists a deterministic transducer computing the definite portion of F.

Computing the PDC is equivalent to solving the combinatorial optimization problem given by

$$\mathrm{PDC}(F) = \min \left\{ \mathrm{DC}(F') \;\middle|\; F' \in \mathcal{F}^L_{\mathcal{X},\mathcal{A}} \text{ and } \tau \sim_* \tau' \right\}, \tag{6.16}$$

where τ and τ' denote corresponding trees for F and F'. In simpler terms, it is the minimum DC over all total AoTFs F' partially consistent with F. An example of a tree corresponding to a partial AoTF is shown in Fig. 6.7.

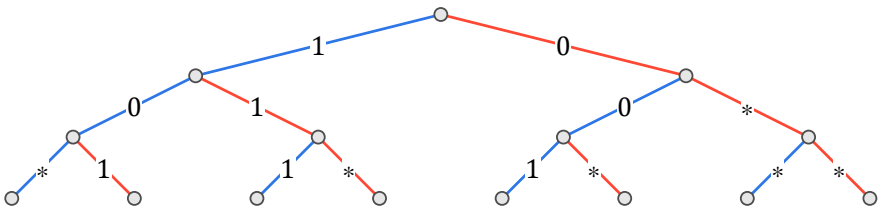

Figure 6.7: Tree for a partial AoTF with $\mathcal{X} = \mathcal{A} = \{0, 1\}$, and with wildcards assigned to some of its outputs. Its consistency graph is shown in Fig. 6.8.

Computing PDC is considerably more challenging than the regular DC. As can be seen from Eq. (6.16), it boils down to finding a definite assignment of outputs for every wildcard in F such that the final DC is minimized. Nevertheless, the graph theoretical perspective provides a lower bound, as we show next.

Proposition 8. Let F be a partial AoTF function and $G = (V, E)$ its consistency graph. Then:

$$\alpha(G) \leq \mathrm{PDC}(F). \tag{6.17}$$

Proof. As in Thm. 5, by definition: $i, j \in I \implies (i, j) \notin E \iff \tau[i] \nsim \tau[j]$. Thus, $s_i \neq s_j$ and $\alpha(G) \leq \mathrm{PDC}(F)$. To further show that equality does not necessarily hold, we provide an example of

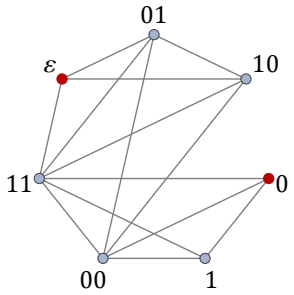

Figure 6.8: Consistency graph G for the function depicted in Fig. 6.7, with vertices labeled by their input prefix. The two highlighted vertices (ε and 0) form the unique maximum independent set in this graph, therefore the independence number is $\alpha(G) = 2$.

$\alpha(G) < \mathrm{PDC}(F)$. The partial AoTF in Fig. 6.7 has $\alpha(G) = 2$, and its number of wildcard outputs is sufficiently small that an exhaustive search can be performed verifying that $\mathrm{PDC}(F) = 3$. This can be obtained with the tree shown in Fig. 6.9. □

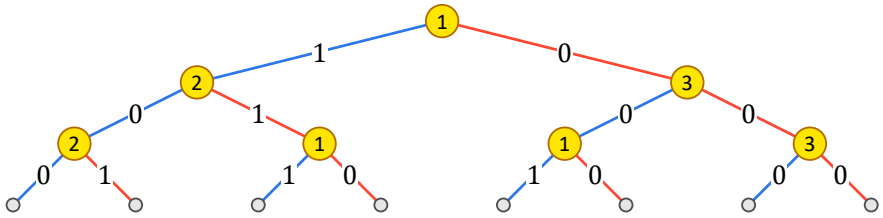

Figure 6.9: Tree for an AoTF F' partially consistent with the tree for F in Fig. 6.7, with a minimum $\mathrm{DC}(F') = 3$. A brute-force search reveals $\mathrm{PDC}(F) = 3$.

6.7.3 Conditional deterministic complexity

Using the PDC, we may also define the *conditional deterministic complexity* (CDC) as follows.

Definition 18. (Conditional Deterministic Complexity, CDC). Let $\boldsymbol{x} \in \mathcal{X}^L$ and $\boldsymbol{a} \in \mathcal{A}^L$ be an input and output sequence, respectively. The *conditional deterministic complexity* (CDC) of $\boldsymbol{a}$ given $\boldsymbol{x}$, denoted by $\mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x})$, is the minimum number of states in a transducer which outputs $\boldsymbol{a}$ given input $\boldsymbol{x}$.

Some basic properties of CDC are:

- $\mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x} = \boldsymbol{a}) = 1$ for any $\boldsymbol{a} \in \mathcal{A}^*$. The transducer simply outputs its input directly.
- $\mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x} = x^L) = \mathrm{DC}_\mathrm{O}(\boldsymbol{a})$ for any $x \in \mathcal{X}$, and $\boldsymbol{a} \in \mathcal{A}^L$. If $\boldsymbol{x}$ only uses one symbol, then the transducer reduces to an automaton and we restore the original definition of DC_O.
- $\mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x}) \leq \mathrm{DC}_\mathrm{O}(\boldsymbol{a})$. This follows from the fact $\boldsymbol{x}$ acts an extra resource, reducing the required memory.
- $\mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x} = (1, 2, \ldots, L)) = 1$, for $\mathcal{X} = \{1, 2, \ldots, L\}$. In other words, if the input acts as an external reference clock, then all transitions can be time-dependent and no information ever needs to be stored in memory.

Proposition 9. For $\boldsymbol{x} \in \mathcal{X}^L$ and $\boldsymbol{a} \in \mathcal{A}^L$, let τ be a depth-L tree such that $\tau(\boldsymbol{x}_{1:\ell}) = a_\ell$ for all $\ell = 1, \ldots, L$, and wildcards elsewhere, as show in Fig. 6.10. Then τ defines a partial AoTF $F_{\boldsymbol{x}\mapsto\boldsymbol{a}}$ such that:

$$\mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x}) = \mathrm{PDC}(F_{\boldsymbol{x}\mapsto\boldsymbol{a}}). \tag{6.18}$$

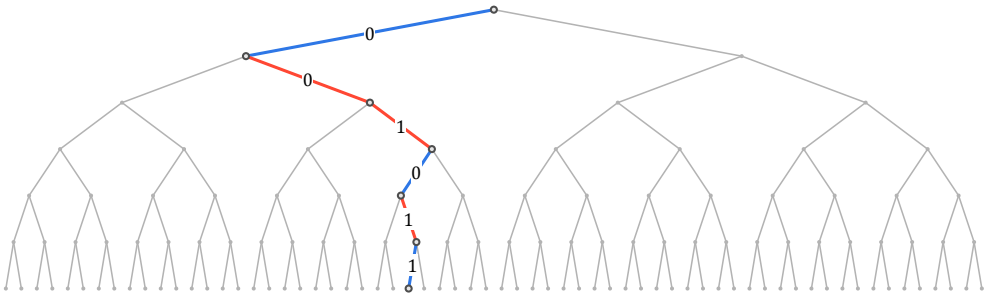

Figure 6.10: Tree for $F_{\boldsymbol{x}\mapsto\boldsymbol{a}}$ with $\boldsymbol{x} = 011010$ and $\boldsymbol{a} = 001011$ (as in Fig. 6.1), which can be used for computing CDC in terms of PDC. Edges with wildcard outputs are grayed out to avoid visual clutter.

Proposition 10. The CDC lower-bounds $\mathrm{DC}(F)$ for any AoTF F. Concretely:

$$\mathrm{DC}(F) \geq \max_{\boldsymbol{x}\in\mathcal{X}^L} \mathrm{CDC}(\boldsymbol{a} = F(\boldsymbol{x})|\boldsymbol{x}). \tag{6.19}$$

Proof. The result is intuitively obvious, as the minimal transducer must have enough states to compute each input-output relation. Nevertheless, the result follows immediately from Eq. (6.16) and Prop. 9. One simply has to consider that F itself will be one the functions F' appearing in Eq. (6.16). □

Despite the result in Prop. 8 which establishes that $\alpha(G) \leq \mathrm{PDC}(F)$ and Prop. 9 connecting CDC with PDC, numerical evidence suggests the following result.

Conjecture 2. For $\boldsymbol{x} \in \mathcal{X}^L$ and $\boldsymbol{a} \in \mathcal{A}^L$, let $F_{\boldsymbol{x}\mapsto\boldsymbol{a}}$ be defined as in Prop. 9 with G its consistency graph. Then:

$$\mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x}) = \mathrm{PDC}(F_{\boldsymbol{x}\mapsto\boldsymbol{a}}) = \alpha(G) \tag{6.20}$$

We leave the proof for future research. For now, we discuss the numerical results.

To estimate the CDC, we used the gradient descent techniques from Sec. 2.7 to search for transducers of increasing dimension d until a deterministic behavior was found, thereby obtaining $\alpha(G) \leq \mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x}) \leq d$. Since deterministic models are typically easy to find if memory is sufficiently large, this approach seemed reasonable. Numerical results were compared with a direct computation of $\alpha(G)$, obtained using Wolfram Mathematica. In almost every scenario $\alpha(G)$ was equal to d, confirming $\alpha(G) = \mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x})$.

Assuming the conjecture is true, a visualization of $\mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x})$ is shown in Fig. 6.11 for various lengths. The horizontal and vertical symmetries arise from the invariance under $0 \leftrightarrow 1$ relabeling, in both inputs and outputs, meaning only a quarter of cases needed to be checked. The dark diagonals correspond to $\mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x} = \boldsymbol{a}) = 1$. While much of the apparent "fractal" structure emerges from the lexicographic ordering of sequences, several interesting structures appear which seem consistent as sequence lengths increase.

6.7.4 Quantum PDC and quantum advantages

As we have seen, under the assumption of a total AoT function both classical and quantum theory will require the same amount of memory (i.e., dimension) to realize it. Partial AoTFs can be interpreted as the deterministic outputs of measurements simply being ignored, which is the scenario discussed in Sec. 6.7.2. In this interpretation, classical and quantum PDC will also coincide, as we are simply introducing the notion of a wildcard or "don't care" instruction.

A different interpretation is to treat unspecified outputs as operations without outputs, i.e., channels. From this perspective classical and quantum theory can differ drastically, as quantum coherence effects

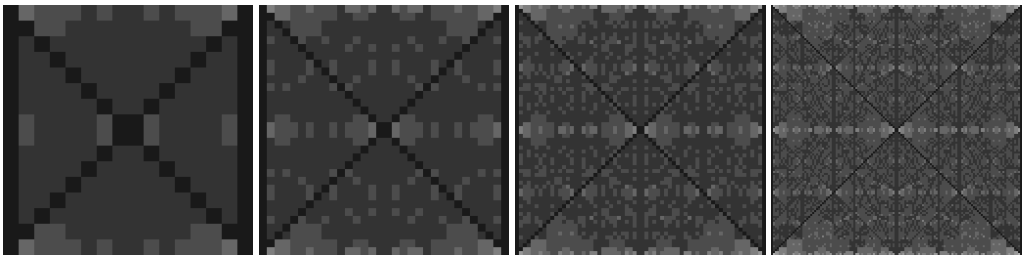

Figure 6.11: Visualization of $\mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x}) = \alpha(G)$ for $\mathcal{X} = \mathcal{A} = \{0, 1\}$ and lengths $L \in \{4, 5, 6, 7\}$ (left to right). Rows correspond to inputs $\boldsymbol{x} \in \mathcal{X}^L$ and columns to outputs $\boldsymbol{a} \in \mathcal{A}^L$, in lexicographic ordering, with larger $\mathrm{CDC}(\boldsymbol{a}|\boldsymbol{x})$ mapped to brighter colors.

can lead to deterministic behaviors beyond what can be achieved by classical systems of the same dimension. Thus, the "quantum PDC" can be lower than the corresponding "classical PDC" for certain partial AoT functions. Indeed, this discrepancy has recently been employed in protocols certifying quantum computation [143].

6.8 Conclusions and outlook

In this chapter we generalized the notion of deterministic complexity first introduced in Ref. [198], formulating it as a type of computational complexity measure. The deterministic complexity originates from the physical scenario involving sequential measurements, leading us to adopt Mealy-type finite-state machines as the appropriate model of computation. In this case, the number of states acted as the natural measure of complexity for an AoT function or, in the case of DC_{O}, of a sequence of outputs.

In other contexts, different notions of complexity could be applied. Within computer science and algorithmic information theory, the Kolmogorov complexity [117] is commonly used instead as the measure of complexity, defined as the size of the shortest possible program for a deterministic universal Turing machine that can generate a given object, such as a binary string. Essentially, it measures the minimum description length of a certain piece of information, or the highest degree it can be algorithmically compressed.

That being said, the Kolmogorov complexity is noncomputable, and therefore, its applications are restricted to theoretical results. This limitation has led to alternative computable complexity measures to be proposed using simpler computational models, including automata and transducers. Several parallels can be made between these complexity measures and the DC.

In a similar way to DC_{O}, a complexity measure based on the number of states is the automatic complexity introduced by Shallit and Wang [174]. Given an input sequence $\boldsymbol{x}$, its automatic complexity, or $\mathrm{AC}(\boldsymbol{x})$, is defined as the minimum number of states in a deterministic automaton which uniquely accepts $\boldsymbol{x}$, i.e., it rejects every other sequence of the same length. Since this automaton can be understood as a Moore-type FSM (i.e., we associate the outputs "accept" and "reject" to states, see Sec. 1.3.3), and since Moore machines can be converted to Mealy machines with the same number of states, it seems plausible that some formal connections will exist between the automatic complexity and the deterministic complexity (in particular, the PDC).

In terms of transducers, a complexity measure was introduced by Calude et al. [38] in analogy to the Kolmogorov complexity, although it differs significantly from the DC. Unlike universal Turing machines, there are no universal transducers [38], so the analogous complexity measure requires the complete description of the transducer itself in some standard encoding. This perspective does not seem relevant to our physically-motivated scenario. Nevertheless, the notion of *conditional* Kolmogorov complexity [117]

also exists, with a similar notion also introduced in Ref. [38] using transducers. It would be interesting to understand their relation to the conditional DC we have introduced in this chapter.

For now, we conclude this chapter by highlighting that the notion of deterministic complexity, and the analysis of temporal correlations through finite-state machines, was entirely motivated by fundamental physical principles: causality, and the fact finite-dimensional systems can only carry a finite amount of information from the past to the future. In this way, the computational perspective on temporal correlations may offer new insights into how these physical principles fundamentally constrain classical and quantum information processing under limited resources. These topics will be subject of future research.

Concluding discussion and outlook

In this thesis we investigated temporal correlations arising from sequences of measurements on finite-dimensional systems. Both classical and quantum scenarios were investigated, treating measurements in their most general form as described within each theory. In particular, we considered the possibility of classical measurements also being invasive.

Through this approach, we have interpreted the realization of temporal correlations as an information processing task, with the system being understood as a memory resource being used for achieving it. As we have seen, the system's dimension directly quantifies this resource. Furthermore, this amount must be finite, otherwise either classical and quantum theory can realize any temporal correlation. These observations motivated our investigation of temporal correlations through the framework of finite-state machines (FSMs).

In Chapter 2, we investigated classical and quantum temporal correlations in the simplest scenario possible, involving a fixed instrument with two possible outcomes: 0 or 1. We performed a numerical survey using gradient descent techniques in order to estimate the maximum probabilities of generating each sequence, for all sequences up to length 10. The notion of *deterministic complexity* (DC) was introduced, quantifying the memory threshold where both classical and quantum theory can generate a sequence deterministically. This allowed us to focus on the sub-deterministic scenarios, where each theory can only generate a sequence with a nontrivial maximum probability. This direct investigation of temporal correlations was lacking, especially in the classical case, but is essential for certifying genuine nonclassicality of any temporal correlation. The survey revealed a rich structure for temporal correlations, even in this simplest case. The results of the classical survey also led us to conjecture the existence of a universal upper bound of $1/\mathrm{e}$ for all sub-deterministic classical scenarios. Furthermore, no such universal upper bound seems to exist in the quantum case.

In Chapter 3, we investigated an application of temporal correlations to the task of dimension witnessing. In particular, we develop a convergent hierarchy of SDP relaxations that can tightly upper-bound temporal correlations in a scenario involving open quantum systems. Here, the environment acts as the memory resource, and these bounds can be used as dimension witnesses for its dimension. However, the resulting SDPs are very large and will generally be numerically intractable in their original form. To overcome this, we have developed a heuristic method which is capable of transforming an SDP into an equivalent sparse representation, with less than 1% of the original number of variables and constraints. This *effective sparsity* is obtained by removing variables and constraints which are unnecessary to solve the original problem. The heuristic method is quite general, and we have further developed it into a separate manuscript (in preparation), which was discussed in Chapter 5.

In Chapter 4, we showcase how FSMs can provide an *exact* characterization of temporal correlations, allowing an unambiguous distinction to be made between classical vs. genuinely nonclassical temporal correlations. In the study of non-Markovianity through quantum supermaps, entanglement-breaking (EB) channels are often used to define classical memories and classical temporal correlations. Here, we conclusively show that this approach is flawed. Our precise characterization of classical memories using FSMs

revealed nonclassicality of temporal correlations in quantum scenarios, even in the presence of EB channels. This is possible because FSMs can characterize classical correlations directly within classical theory, instead of through quantum-inspired analogies, as is often done. These nonclassical temporal effects are quite subtle, only occurring beyond 2-level quantum systems and at least four time steps—scenarios that are rarely investigated. Our results reveal that classical and nonclassical temporal correlations are still widely misunderstood, often due to improper and hasty analogies drawn from spatial correlations, and an insufficient treatment of the relevant memory resources.

In Chapter 6, we expand the deterministic complexity to the case of deterministic sequences of measurements involving inputs and outputs, thereby providing an exact characterization of an arbitrary extreme point in the Arrow-of-Time polytope. We further establish connections between the DC and the problem of *maximum independent set* in graph theory, enabling further generalizations. In particular, we introduce the partial DC and the conditional DC, which have interesting parallels with existing complexity measures in computer science. These results strengthen the bridge between foundational questions on temporal correlations and fundamental notions of complexity in theoretical computer science.

This thesis makes significant contributions to our understanding of temporal phenomena, providing one of the first systematic investigations into genuine nonclassicality of temporal correlations. In particular, our analysis of classical and quantum scenarios led to numerous novel insights on uniquely temporal effects, showcasing how classical and quantum theory behave differently under the constraints of causality and finite memory. The framework of finite-state machines was essential for these results. Its unique ability to sharply distinguish between classical and nonclassical temporal correlations highlights how little we actually know about this subject.

We have shown that many existing notions of nonclassicality are based on arbitrary restrictions on classical systems (Sec. 1.2), or improper analogies from spatial correlations applied to the temporal case (Ch. 4), both of which cannot adequately distinguish between classical and nonclassical temporal correlations. Furthermore, a significant portion of previous work on temporal correlations has relied on simple prepare-and-measure scenarios or two-time correlations. However, the results in this thesis revealed that several unique features of temporal correlations do not appear under these conditions, only emerging in scenarios beyond two-dimensional systems and at least three separate measurements. Our results indicate that a significant portion of genuinely temporal phenomena is currently being neglected. This important point has also been noted by other authors [190].

As discussed in Sec. 1.2.7, any temporal correlation can be realized by a classical or quantum system if enough memory is available. This is a widely acknowledged fact about temporal correlations, with the assumption of a bounded memory (i.e., bounded dimension) being part of many sequential protocols. That being said, there has not been a systematic analysis of the resources required for realizing arbitrary temporal correlations. This is a challenging task, as the space of dimension-bounded temporal correlations is very complex [182, 124]. Nevertheless, a characterization of the minimum resources needed to realize temporal correlations would be incredibly valuable. It would lead to new ways of certifying quantum information processing, and would most likely be based on a deeper understanding of how quantum coherence effects can be sequentially engineered to achieve specific tasks.

We have made significant contributions towards this goal by introducing the notion of deterministic complexity (DC), which characterizes the minimum dimension required for realizing deterministic correlations. Together with initial randomness, the DC allows an initial characterization of the resources needed to generate any correlation in the Arrow-of-Time polytope [32, 182, 124]. Furthermore, we formalized its connections to concrete computational tasks, establishing a bridge between fundamental notions in physics (causality and dimension) to those in theoretical computer science (time and memory).

The research in this thesis also inspires several future directions. As we have seen, both classical

and quantum theory can generate all temporal correlations if enough memory resources are available. Thus, a proper accounting of all resources is essential, otherwise no distinction can be made between classical and nonclassical temporal correlations. However, there are many ways in which such resources may be inadvertently included, e.g., time-dependent operations, initial randomness, a mixed initial state, and conditioning on past measurements. All of these can lead to a significant reduction in the amount of memory needed to realize arbitrary temporal correlations, and all have been independently treated as a free resource in previous works, essentially becoming a form of "memory smuggling". It would be useful to precisely characterize these resources in terms of their memory advantage, thereby unifying all resources available for temporal correlations within the same framework. We believe this is possible, and it will be the subject of future research.

Based on Ch. 3, we have seen that it is possible to use temporal correlations to obtain information about an inaccessible environment system through measurements on a smaller probe system. It would be interesting to investigate whether different temporal protocols can be devised targeting specific properties about the environment, or the system-environment interaction. For example, can we estimate the strength of the system-environment interaction using temporal correlations? This should be possible, as this interaction is intimately related to the amount of information that can be exchanged between system and environment.

There is considerable interest in understanding the origin of quantum advantages in quantum computing, and the FSM framework offers clear paradigmatic problems for investigating this question. Already in the simplest case investigated in Ch. 2, where only output sequences were considered, we have found several dimensionally-restricted scenarios where the quantum advantage seems to disappear, despite the classical and quantum models operating in completely different ways (Sec. 2.8.5). This suggests the existence of a memory threshold for quantum advantages to exist in specific computational tasks. By investigating such small-scale scenarios, we may be able to uncover new insights on the necessary conditions for quantum advantages to emerge in terms of information-theoretical principles.

Finally, we note that our investigations into temporal correlations have relied heavily on numerical techniques. Bounding temporal correlations appears inherently challenging, significantly more so than spatial correlations, with even the simplest scenarios leading to large-scale nonlinear optimization problems that are quickly rendered intractable (see Ch. 3). It is unclear whether alternative mathematical or numerical techniques exist that are better suited for this task.

Despite these difficulties, we are optimistic that the framework of finite-state machines will eventually lead to further developments and insights into temporal correlations. In particular, its unique capability of accurately identifying genuinely nonclassical temporal correlations is essential when certifying the existence of quantum advantages in information processing tasks.

As a final remark, we emphasize that temporal correlations are an underappreciated and underexplored subject. While temporal phenomena arising from finite-memory effects are ubiquitous, being implicitly used throughout all of quantum information science, there is a considerable lack of research directly and systematically investigating temporal correlations. We hope that our contributions can help bring much needed attention to this fascinating topic.

Bibliography

[1] Alastair A. Abbott et al. "Multipartite causal correlations: Polytopes and inequalities". In: *Phys. Rev. A* 94 (3 Sept. 2016), p. 032131. DOI: 10.1103/PhysRevA.94.032131.

[2] Paolo Abiuso. "Verification of continuous-variable quantum memories". In: *Quantum Science and Technology* 9.1 (Nov. 2023), 01lt02. DOI: 10.1088/2058-9565/ad097c.

[3] L. Accardi, A. Frigerio, and J. T. Lewis. "Quantum Stochastic Processes". In: *Publ. Rest. Inst. Math. Sci.* 18 (1982), pp. 97–133. DOI: 10.2977/prims/1195184017.

[4] Akshay Agrawal et al. "A rewriting system for convex optimization problems". In: *J. Control. Decis* 5.1 (2018), pp. 42–60. DOI: 10.1080/23307706.2017.1397554.

[5] Edgar A. Aguilar et al. "Certifying an Irreducible 1024-Dimensional Photonic State Using Refined Dimension Witnesses". In: *Phys. Rev. Lett.* 120 (23 June 2018), p. 230503. DOI: 10.1103/PhysRevLett.120.230503.

[6] S. Alipour, M. Mehboudi, and A. T. Rezakhani. "Quantum Metrology in Open Systems: Dissipative Cramér-Rao Bound". In: *Phys. Rev. Lett.* 112 (12 Mar. 2014), p. 120405. DOI: 10.1103/PhysRevLett.112.120405.

[7] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences: theory, applications, generalizations*. Cambridge University Press, 2003. DOI: 10.1017/CBO9780511546563.

[8] Mafalda L. Almeida et al. "Guess Your Neighbor's Input: A Multipartite Nonlocal Game with No Quantum Advantage". In: *Phys. Rev. Lett.* 104 (23 June 2010), p. 230404. DOI: 10.1103/PhysRevLett.104.230404.

[9] Andris Ambainis et al. "Dense quantum coding and a lower bound for 1-way quantum automata". In: *Proceedings of the thirty-first annual ACM symposium on Theory of computing*. 1999, pp. 376–383. DOI: 10.1145/301250.301347.

[10] Andris Ambainis et al. "Dense quantum coding and quantum finite automata". In: *Journal of the ACM (JACM)* 49.4 (2002), pp. 496–511. DOI: 10.1145/581771.581773.

[11] John S Bell. "On the Einstein-Podolsky-Rosen Paradox". In: *Physics 1, 195* (1964). DOI: 10.1103/PhysicsPhysiqueFizika.1.195.

[12] John S. Bell. "On the Problem of Hidden Variables in Quantum Mechanics". In: *Rev. Mod. Phys.* 38 (3 July 1966), pp. 447–452. DOI: 10.1103/RevModPhys.38.447.

[13] John Stewart Bell. *Speakable and unspeakable in quantum mechanics: Collected papers on quantum philosophy*. Cambridge university press, 2004. ISBN: 9780511815676. DOI: 10.1017/CBO9780511815676.

[14] Graeme D. Berk et al. "Resource theories of multi-time processes: A window into quantum non-Markovianity". In: *Quantum* 5 (Apr. 2021), p. 435. ISSN: 2521-327X. DOI: 10.22331/q-2021-04-20-435.

[15] Jean Berstel. *Transductions and context-free languages*. Springer-Verlag, 2013.

[16] Mario Berta et al. “Semidefinite programming hierarchies for constrained bilinear optimization”. In: *Math. Program.* 194.1 (July 2022), pp. 781–829. ISSN: 1436-4646. DOI: 10.1007/s10107-021-01650-1.

[17] Felix C. Binder, Jayne Thompson, and Mile Gu. “Practical Unitary Simulator for Non-Markovian Complex Processes”. In: *Phys. Rev. Lett.* 120 (24 June 2018), p. 240502. DOI: 10.1103/PhysRevLett.120.240502.

[18] Pratik Biswas and Yinyu Ye. “A Distributed Method for Solving Semidefinite Programs Arising from Ad Hoc Wireless Sensor Network Localization”. In: *Multiscale Optimization Methods and Applications*. Ed. by William W. Hager et al. Boston, MA: Springer US, 2006, pp. 69–84.

[19] Joseph Bowles, Flavio Baccari, and Alexia Salavrakos. “Bounding sets of sequential quantum correlations and device-independent randomness certification”. In: *Quantum* 4 (Oct. 2020), p. 344. ISSN: 2521-327x. DOI: 10.22331/q-2020-10-19-344.

[20] Joseph Bowles, Nicolas Brunner, and Marcin Pawłowski. “Testing dimension and nonclassicality in communication networks”. In: *Phys. Rev. A* 92 (2 Aug. 2015), p. 022351. DOI: 10.1103/PhysRevA.92.022351.

[21] Stephen Boyd and Lieven Vandenberghe. “Semidefinite Programming Relaxations of Non-Convex Problems in Control and Combinatorial Optimization”. In: *Communications, Computation, Control, and Signal Processing: a tribute to Thomas Kailath*. Ed. by Arogyaswami Paulraj, Vwani Roychowdhury, and Charles D. Schaper. Boston, MA: Springer US, 1997, pp. 279–287. ISBN: 978-1-4615-6281-8. DOI: 10.1007/978-1-4615-6281-8_15.

[22] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. ISBN: 9780521833783. DOI: 10.1017/cbo9780511804441. URL: https://web.stanford.edu/~boyd/cvxbook/.

[23] V. B. Braginsky and F. Y. Khalili. *Quantum Measurement*. Cambridge University Press, 1992. DOI: 10.1017/cbo9780511622748.

[24] Heinz-Peter Breuer and Francesco Petruccione. *The Theory of Open Quantum Systems*. Oxford University Press, 2002. ISBN: 978-0-198-52063-4. DOI: 10.1093/acprof:oso/9780199213900.001.0001.

[25] Heinz-Peter Breuer et al. “Colloquium: Non-Markovian dynamics in open quantum systems”. In: *Rev. Mod. Phys.* 88 (2 Apr. 2016), p. 021002. DOI: 10.1103/RevModPhys.88.021002.

[26] Stephen Brierley et al. “Nonclassicality of Temporal Correlations”. In: *Phys. Rev. Lett.* 115 (12 Sept. 2015), p. 120404. DOI: 10.1103/PhysRevLett.115.120404.

[27] Caroline Brosse, Vincent Limouzy, and Arnaud Mary. “Polynomial delay algorithm for minimal chordal completions”. In: (2021). arXiv: 2107.05972 [cs.DS].

[28] Nicolas Brunner, Miguel Navascués, and Tamás Vértesi. “Dimension Witnesses and Quantum State Discrimination”. In: *Phys. Rev. Lett.* 110 (15 Apr. 2013), p. 150501. DOI: 10.1103/PhysRevLett.110.150501.

[29] Nicolas Brunner et al. “Bell nonlocality”. In: *Rev. Mod. Phys.* 86 (2 Apr. 2014), pp. 419–478. DOI: 10.1103/RevModPhys.86.419.

[30] Adrián A. Budini. “Embedding non-Markovian quantum collisional models into bipartite Markovian dynamics”. In: *Phys. Rev. A* 88.3 (Sept. 2013), p. 032115. DOI: 10.1103/PhysRevA.88.032115.

[31] Costantino Budroni and Clive Emary. “Temporal Quantum Correlations and Leggett-Garg Inequalities in Multilevel Systems”. In: *Phys. Rev. Lett.* 113 (5 July 2014), p. 050401. DOI: 10.1103/PhysRevLett.113.050401.

[32] Costantino Budroni, Gabriel Fagundes, and Matthias Kleinmann. "Memory cost of temporal correlations". In: *New J. Phys.* 21.9 (Sept. 2019), p. 093018. DOI: 10.1088/1367-2630/ab3cb4.

[33] Costantino Budroni, Giuseppe Vitagliano, and Mischa P Woods. "Ticking-clock performance enhanced by nonclassical temporal correlations". In: *Phys. Rev. Research* 3.3 (3 July 2021), p. 033051. DOI: 10.1103/PhysRevResearch.3.033051.

[34] Costantino Budroni et al. "Bounding Temporal Quantum Correlations". In: *Phys. Rev. Lett.* 111 (2 July 2013), p. 020403. DOI: 10.1103/PhysRevLett.111.020403.

[35] Costantino Budroni et al. "Kochen-Specker contextuality". In: *Rev. Mod. Phys.* 94 (4 Dec. 2022), p. 045007. DOI: 10.1103/RevModPhys.94.045007.

[36] Paul Busch. "'No Information Without Disturbance': Quantum Limitations of Measurement". In: *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle: Essays in Honour of Abner Shimony*. Dordrecht: Springer Netherlands, 2009, pp. 229–256. ISBN: 978-1-4020-9107-0. DOI: 10.1007/978-1-4020-9107-0_13.

[37] Paul Busch, Pekka J. Lahti, and Peter Mittelstaedt. *The Quantum Theory of Measurement*. 2nd ed. Vol. 2. Lecture Notes in Physics Monographs. Springer-Verlag Berlin Heidelberg, 1996. DOI: 10.1007/978-3-540-37205-9.

[38] Cristian Calude, Kai Salomaa, and Tania Roblot. "Finite-State Complexity and the Size of Transducers". In: vol. 31. Open Publishing Association, Aug. 2010, pp. 38–47. DOI: 10.4204/eptcs.31.6.

[39] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack. "Unknown quantum states: The quantum de Finetti representation". In: *J. Math. Phys.* 43.9 (2002), pp. 4537–4559. DOI: 10.1063/1.1494475.

[40] Wan-Guan Chang et al. "Visually quantifying single-qubit quantum memory". In: (2023). arXiv: 2312.06939 [quant-ph].

[41] G. Chiribella, G. M. D'Ariano, and P. Perinotti. "Transforming quantum operations: Quantum supermaps". In: *Europhysics Letters* 83.3 (July 2008), p. 30004. DOI: 10.1209/0295-5075/83/30004.

[42] Giulio Chiribella. "On Quantum Estimation, Quantum Cloning and Finite Quantum de Finetti Theorems". In: *Theory of Quantum Computation, Communication, and Cryptography*. Ed. by Wim van Dam, Vivien M. Kendon, and Simone Severini. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 9–25. DOI: 10.1007/978-3-642-18073-6_2.

[43] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. "Theoretical framework for quantum networks". In: *Phys. Rev. A* 80 (2 Aug. 2009), p. 022339. DOI: 10.1103/PhysRevA.80.022339.

[44] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. "Informational derivation of quantum theory". In: *Phys. Rev. A* 84 (1 July 2011), p. 012311. DOI: 10.1103/PhysRevA.84.012311. URL: https://link.aps.org/doi/10.1103/PhysRevA.84.012311.

[45] Giulio Chiribella et al. "Quantum computations without definite causal structure". In: *Phys. Rev. A* 88 (2 Aug. 2013), p. 022318. DOI: 10.1103/PhysRevA.88.022318.

[46] Man-Duen Choi. "Completely positive linear maps on complex matrices". In: *Linear Algebra Its Appl.* 10.3 (1975), pp. 285–290. ISSN: 0024-3795. DOI: 10.1016/0024-3795(75)90075-0.

[47] Matthias Christandl et al. "One-and-a-half quantum de Finetti theorems". In: *Commun. Math. Phys.* 273.2 (2007), pp. 473–498. DOI: 10.1007/s00220-007-0189-3.

[48] B. S. Cirel'son. "Quantum generalizations of Bell's inequality". In: *Letters in Mathematical Physics* 4.2 (Mar. 1980), pp. 93–100. ISSN: 1573-0530. DOI: 10.1007/BF00417500.

[49] E.M. Clarke et al. *Model Checking*. 2nd. Cyber Physical Systems Series. MIT Press, 2018. ISBN: 9780262038836. URL: https://mitpress.mit.edu/9780262038836/model-checking/.

[50] J F Clauser and A Shimony. "Bell's theorem. Experimental tests and implications". In: *Reports on Progress in Physics* 41.12 (Dec. 1978), p. 1881. DOI: 10.1088/0034-4885/41/12/002.

[51] John F. Clauser et al. "Proposed Experiment to Test Local Hidden-Variable Theories". In: *Phys. Rev. Lett.* 23 (15 Oct. 1969), pp. 880–884. DOI: 10.1103/PhysRevLett.23.880.

[52] Lucas Clemente and Johannes Kofler. "No Fine Theorem for Macrorealism: Limitations of the Leggett-Garg Inequality". In: *Phys. Rev. Lett.* 116 (15 Apr. 2016), p. 150401. DOI: 10.1103/PhysRevLett.116.150401.

[53] Derek G. Corneil. "Lexicographic Breadth First Search – A Survey". In: *Graph-Theoretic Concepts in Computer Science*. Ed. by Juraj Hromkovič, Manfred Nagl, and Bernhard Westfechtel. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 1–19. ISBN: 978-3-540-30559-0.

[54] Luis A. Correa et al. "Individual Quantum Probes for Optimal Thermometry". In: *Phys. Rev. Lett.* 114 (22 June 2015), p. 220405. DOI: 10.1103/PhysRevLett.114.220405.

[55] James P. Crutchfield. "The calculi of emergence: computation, dynamics and induction". In: *Physica D: Nonlinear Phenomena* 75.1 (1994), pp. 11–54. ISSN: 0167-2789. DOI: 10.1016/0167-2789(94)90273-9.

[56] C. L. Degen, F. Reinhard, and P. Cappellaro. "Quantum sensing". In: *Rev. Mod. Phys.* 89 (3 July 2017), p. 035002. DOI: 10.1103/RevModPhys.89.035002.

[57] Steven Diamond and Stephen Boyd. "CVXPY: A Python-embedded modeling language for convex optimization". In: *J. Mach. Learn. Res* 17.83 (2016), pp. 1–5. DOI: 10.5555/2946645.3007036.

[58] A. C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. "Distinguishing Separable and Entangled States". In: *Phys. Rev. Lett.* 88 (18 Apr. 2002), p. 187904. DOI: 10.1103/PhysRevLett.88.187904.

[59] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. "Complete family of separability criteria". In: *Phys. Rev. A* 69 (2 Feb. 2004), p. 022308. DOI: 10.1103/PhysRevA.69.022308.

[60] Michael Domaratzki, Derek Kisman, and Jeffrey Shallit. "On the number of distinct languages accepted by finite automata with n states". In: *Journal of Automata, Languages and Combinatorics* 7.4 (2002), pp. 469–486. DOI: 10.5555/782466.782472.

[61] Manfred Droste, Werner Kuich, and Heiko Vogler. *Handbook of Weighted Automata*. 1st. Springer Publishing Company, Incorporated, 2009. ISBN: 3642014917. DOI: 10.5555/1667106.

[62] A. Einstein, B. Podolsky, and N. Rosen. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" In: *Physical Review* 47.10 (May 1935), pp. 777–780. DOI: 10.1103/PhysRev.47.777.

[63] Thomas J. Elliott et al. "Quantum Adaptive Agents with Efficient Long-Term Memories". In: *Phys. Rev. X* 12 (1 Jan. 2022), p. 011007. DOI: 10.1103/PhysRevX.12.011007.

[64] Clive Emary, Neill Lambert, and Franco Nori. "Leggett–Garg inequalities". In: *Rep. Prog. Phys.* 77.1 (Dec. 2013), p. 016001. ISSN: 0034-4885. DOI: 10.1088/0034-4885/77/1/016001.

[65] Gabriel Fagundes and Matthias Kleinmann. "Memory cost for simulating all quantum correlations from the Peres–Mermin scenario". In: *J. Phys. A* 50.32 (2017), p. 325302. DOI: 10.1088/1751-8121/aa7ab3.

[66] Arthur Fine. "Hidden Variables, Joint Probability, and the Bell Inequalities". In: *Phys. Rev. Lett.* 48 (5 Feb. 1982), pp. 291–295. DOI: 10.1103/PhysRevLett.48.291.

[67] Tobias Fritz. "Quantum correlations in the temporal Clauser–Horne–Shimony–Holt (CHSH) scenario". In: *New J. Phys.* 12.8 (2010), p. 083055. DOI: 10.1088/1367-2630/12/8/083055.

[68] Mituhiro Fukuda et al. “Exploiting sparsity in semidefinite programming via matrix completion I: General framework”. In: *SIAM J. Optim.* 11.3 (2001), pp. 647–674. DOI: 10.1137/s1052623400366218.

[69] Rodrigo Gallego et al. “Device-Independent Tests of Classical and Quantum Dimensions”. In: *Phys. Rev. Lett.* 105 (23 Nov. 2010), p. 230501. DOI: 10.1103/PhysRevLett.105.230501.

[70] Rodrigo Gallego et al. “Nonlocality in sequential correlation scenarios”. In: *New Journal of Physics* 16.3 (Mar. 2014), p. 033037. DOI: 10.1088/1367-2630/16/3/033037.

[71] Guillaume Garrigos and Robert M. Gower. *Handbook of Convergence Theorems for (Stochastic) Gradient Methods*. 2024. arXiv: 2301.11235 [math.OC].

[72] Christina Giarmatzi and Fabio Costa. “Witnessing quantum memory in non-Markovian processes”. In: *Quantum* 5 (Apr. 2021), p. 440. ISSN: 2521-327x. DOI: 10.22331/q-2021-04-26-440.

[73] S. Goldstein et al. “Bell’s theorem”. In: *Scholarpedia* 6.10 (2011). revision #91049, p. 8378. DOI: 10.4249/scholarpedia.8378.

[74] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. ISBN: 9780262035613. URL: http://www.deeplearningbook.org.

[75] Pablo A. Parrilo Grigoriy Blekherman and Rekha R. Thomas, eds. *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics, Dec. 2012. ISBN: 978-1-611-97229-0. DOI: 10.1137/1.9781611972290.

[76] Robert Grone et al. “Positive definite completions of partial Hermitian matrices”. In: *Linear Algebra and its Applications* 58 (1984), pp. 109–124. ISSN: 0024-3795. DOI: 10.1016/0024-3795(84)90207-6.

[77] Jonathan L Gross and Jay Yellen. *Handbook of graph theory*. CRC press, 2003.

[78] Branko Grünbaum et al. *Convex polytopes*. New York: Springer, 2003. ISBN: 9780387004242.

[79] Otfried Gühne and Géza Tóth. “Entanglement detection”. In: *Phys. Rep.* 474.1 (2009), pp. 1–75. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2009.02.004.

[80] Otfried Gühne et al. “Compatibility and noncontextuality for sequential measurements”. In: *Phys. Rev. A* 81 (2010), p. 022121. DOI: 10.1103/PhysRevA.81.022121.

[81] Otfried Gühne et al. “Bounding the quantum dimension with contextuality”. In: *Phys. Rev. A* 89 (6 June 2014), p. 062107. DOI: 10.1103/PhysRevA.89.062107.

[82] Leonid Gurvits. “Classical Deterministic Complexity of Edmonds’ Problem and Quantum Entanglement”. In: *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*. Stoc ’03. San Diego, CA, USA: Association for Computing Machinery, 2003, pp. 10–19. ISBN: 1581136749. DOI: 10.1145/780542.780545.

[83] Lucien Hardy. *Quantum Theory From Five Reasonable Axioms*. 2001. arXiv: quant-ph/0101012 [quant-ph].

[84] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

[85] Aram W Harrow. “The church of the symmetric subspace”. In: *arXiv:1308.6595* (2013). DOI: 10.48550/arXiv.1308.6595.

[86] Pinar Heggernes. “Minimal triangulations of graphs: A survey”. In: *Discrete Mathematics* 306.3 (2006). Minimal Separation and Minimal Triangulation, pp. 297–317. ISSN: 0012-365X. DOI: 10.1016/j.disc.2005.12.003.

[87] W. Heisenberg. “Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik”. In: *Zeitschrift für Physik* 43.3 (Mar. 1927), pp. 172–198. ISSN: 0044-3328. DOI: 10.1007/BF01397280.

[88] Jannik Hoffmann et al. “Structure of temporal correlations of a qubit”. In: *New J. Phys.* 20.10 (Oct. 2018), p. 102001. DOI: 10.1088/1367-2630/aae87f.

[89] A S Holevo. “Quantum coding theorems”. In: *Russian Mathematical Surveys* 53.6 (Dec. 1998), p. 1295. DOI: 10.1070/RM1998v053n06ABEH000091.

[90] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to automata theory, languages, and computation, 2nd edition.* Vol. 32. 1. New York, NY, USA: ACM Press, 2001, pp. 60–65. DOI: 10.1145/568438.568455.

[91] Michael Horodecki, Peter W. Shor, and Mary Beth Ruskai. “Entanglement Breaking Channels”. In: *Reviews in Mathematical Physics* 15.06 (2003), pp. 629–641. DOI: 10.1142/S0129055X03001709.

[92] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. “Mixed-State Entanglement and Distillation: Is there a “Bound” Entanglement in Nature?” In: *Phys. Rev. Lett* 80 (24 June 1998), pp. 5239–5242. DOI: 10.1103/PhysRevLett.80.5239.

[93] J. Solomon Ivan, Krishna Kumar Sabapathy, and R. Simon. “Nonclassicality breaking is the same as entanglement breaking for bosonic Gaussian channels”. In: *Phys. Rev. A* 88 (3 Sept. 2013), p. 032302. DOI: 10.1103/PhysRevA.88.032302.

[94] A. Jamiołkowski. “Linear transformations which preserve trace and positive semidefiniteness of operators”. In: *Rep. Math. Phys.* 3.4 (1972), pp. 275–278. ISSN: 0034-4877. DOI: 10.1016/0034-4877(72)90011-0.

[95] Hyejung H. Jee et al. “Quasi-Polynomial Time Algorithms for Free Quantum Games in Bounded Dimension”. In: *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021).* Ed. by Nikhil Bansal, Emanuela Merelli, and James Worrell. Vol. 198. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 82:1–82:20. ISBN: 978-3-95977-195-5. DOI: 10.4230/LIPIcs.ICALP.2021.82.

[96] Alexandra M Jurgens and James P Crutchfield. “Shannon entropy rate of hidden Markov processes”. In: *Journal of Statistical Physics* 183.2 (2021), p. 32. DOI: 10.1007/s10955-021-02769-3.

[97] Sunyoung Kim, Masakazu Kojima, and Hayato Waki. “Generalized Lagrangian Duals and Sums of Squares Relaxations of Sparse Polynomial Optimization Problems”. In: *SIAM Journal on Optimization* 15.3 (2005), pp. 697–719. DOI: 10.1137/030601260.

[98] Sunyoung Kim, Masakazu Kojima, and Hayato Waki. “Exploiting Sparsity in SDP Relaxation for Sensor Network Localization”. In: *SIAM Journal on Optimization* 20.1 (2009), pp. 192–215. DOI: 10.1137/080713380.

[99] Sunyoung Kim et al. “Exploiting sparsity in linear and nonlinear matrix inequalities via positive semidefinite matrix completion”. In: *Mathematical Programming* 129.1 (Sept. 2011), pp. 33–68. ISSN: 1436-4646. DOI: 10.1007/s10107-010-0402-6.

[100] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization.* 2017. arXiv: 1412.6980 [cs.LG].

[101] G. Kirchmair et al. “State-independent experimental test of quantum contextuality”. In: *Nature (London)* 460 (2009), pp. 494–497. DOI: 10.1038/nature08172.

[102] Matthias Kleinmann et al. “Memory cost of quantum contextuality”. In: *New J. Phys.* 13.11 (2011), p. 113011. DOI: 10.1088/1367-2630/13/11/113011.

[103] Johannes Kofler and Časlav Brukner. “Conditions for Quantum Violation of Macroscopic Realism”. In: *Phys. Rev. Lett.* 101 (9 Aug. 2008), p. 090403. DOI: 10.1103/PhysRevLett.101.090403.

[104] Andrey Kolmogoroff. “Grundbegriffe der wahrscheinlichkeitsrechnung”. In: (1933). DOI: 10.1007/978-3-642-49888-6.

[105] J. K. Korbicz, J. I. Cirac, and M. Lewenstein. “Spin Squeezing Inequalities and Entanglement of N Qubit States”. In: *Phys. Rev. Lett.* 95 (12 Sept. 2005), p. 120502. DOI: 10.1103/PhysRevLett.95.120502.

[106] Huan-Yu Ku et al. “Quantifying Quantumness of Channels Without Entanglement”. In: *PRX Quantum* 3 (2 May 2022), p. 020338. DOI: 10.1103/PRXQuantum.3.020338.

[107] Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert. “Numba: a LLVM-based Python JIT compiler”. In: *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*. LLVM ’15. Austin, Texas: Association for Computing Machinery, 2015. ISBN: 9781450340052. DOI: 10.1145/2833157.2833162.

[108] Justo Pastor Lambare. “On the Meaning of Local Realism”. In: *Foundations of Physics* 52.5 (Sept. 2022), p. 98. ISSN: 1572-9516. DOI: 10.1007/s10701-022-00618-1.

[109] Jan-Åke Larsson. “Loopholes in Bell inequality tests of local realism”. In: *Journal of Physics A: Mathematical and Theoretical* 47.42 (Oct. 2014), p. 424003. DOI: 10.1088/1751-8113/47/42/424003.

[110] Jean B. Lasserre. “Global Optimization with Polynomials and the Problem of Moments”. In: *SIAM Journal on Optimization* 11.3 (2001), pp. 796–817. DOI: 10.1137/S1052623400366802.

[111] Jean B. Lasserre. “An Explicit Equivalent Positive Semidefinite Program for Nonlinear 0-1 Programs”. In: *SIAM Journal on Optimization* 12.3 (2002), pp. 756–769. DOI: 10.1137/S1052623400380079.

[112] Jean B. Lasserre. “Convergent SDP-Relaxations in Polynomial Optimization with Sparsity”. In: *SIAM Journal on Optimization* 17.3 (2006), pp. 822–843. DOI: 10.1137/05064504X.

[113] Monique Laurent. “Sums of Squares, Moment Matrices and Optimization Over Polynomials”. In: *Emerging Applications of Algebraic Geometry*. Ed. by Mihai Putinar and Seth Sullivant. New York, NY: Springer New York, 2009, pp. 157–270. ISBN: 978-0-387-09686-5.

[114] Edward Ashford Lee and Sanjit Arunkumar Seshia. *Introduction to embedded systems: A cyber-physical systems approach*. 2nd. MIT press, 2017. ISBN: 9780262533812.

[115] A. J. Leggett. “Realism and the physical world”. In: *Rep. Prog. Phys.* 71.2 (Jan. 2008), p. 022001. ISSN: 0034-4885. DOI: 10.1088/0034-4885/71/2/022001.

[116] A. J. Leggett and Anupam Garg. “Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks?” In: *Phys. Rev. Lett.* 54.9 (9 Mar. 1985), pp. 857–860. DOI: 10.1103/PhysRevLett.54.857.

[117] Paul Vitányi Ming Li. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, 1997. ISBN: 9781489984456. DOI: 10.1007/978-0-387-49820-1.

[118] Göran Lindblad. “Non-Markovian quantum stochastic processes and their entropy”. In: *Comm. Math. Phys.* 65.3 (1979), pp. 281–294. DOI: 10.1007/bf01197883.

[119] Robert Logožar and Alen Lovrenčić. “The modeling and complexity of dynamical systems by means of computation and information theories”. In: *Journal of information and organizational sciences* 35.2 (2011), pp. 173–196.

[120] I. A. Luchnikov, S. V. Vintskevich, and S. N. Filippov. *Dimension truncation for open quantum systems in terms of tensor networks*. arXiv:1801.07418. Jan. 2018. DOI: 10.48550/arXiv.1801.07418.

[121] I. A. Luchnikov et al. “Simulation Complexity of Open Quantum Dynamics: Connection with Tensor Networks”. In: *Phys. Rev. Lett.* 122.16 (Apr. 2019), p. 160401. DOI: 10.1103/PhysRevLett.122.160401.

[122] I. A. Luchnikov et al. “Probing non-Markovian quantum dynamics with data-driven analysis: Beyond “black-box” machine-learning models”. In: *Phys. Rev. Res.* 4.4 (Oct. 2022), p. 043002. DOI: 10.1103/PhysRevResearch.4.043002.

[123] Günther Ludwig. *An axiomatic basis for quantum mechanics: Volume 1 derivation of Hilbert space structure*. Springer Science & Business Media, 1985.

[124] Yuanyuan Mao et al. “Structure of dimension-bounded temporal correlations”. In: *Phys. Rev. A* 105 (2 Feb. 2022), p. L020201. DOI: 10.1103/PhysRevA.105.L020201. arXiv: 2005.13964 [quant-ph].

[125] O.J.E. Maroney. “Measurements, disturbances and the quantum three box paradox”. In: *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 58 (May 2017), pp. 41–53. ISSN: 1355-2198. DOI: 10.1016/j.shpsb.2016.12.003.

[126] Owen J. E Maroney and Christopher G Timpson. *Quantum- vs. Macro- Realism: What does the Leggett-Garg Inequality actually test?* 2014. arXiv: 1412.6139 [quant-ph].

[127] George H. Mealy. “A method for synthesizing sequential circuits”. In: *The Bell System Technical Journal* 34.5 (1955), pp. 1045–1079. DOI: 10.1002/j.1538-7305.1955.tb03788.x.

[128] Mohammad Mehboudi, Anna Sanpera, and Luis A Correa. “Thermometry in the quantum regime: recent theoretical progress”. In: *Journal of Physics A: Mathematical and Theoretical* 52.30 (July 2019), p. 303001. DOI: 10.1088/1751-8121/ab2828.

[129] Martin Mevissen et al. “Solving partial differential equations via sparse SDP relaxations”. In: *Pacific Journal of Optimization* 4.2 (2008), pp. 213–241.

[130] Stoyan Mihov and Klaus U. Schulz. *Finite-State Techniques: Automata, Transducers and Bimachines*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2019.

[131] Nikolai Miklin, Jakub J. Borkała, and Marcin Pawłowski. “Semi-device-independent self-testing of unsharp measurements”. In: *Phys. Rev. Research* 2 (3 July 2020), p. 033014. DOI: 10.1103/PhysRevResearch.2.033014.

[132] Simon Milz and Kavan Modi. “Quantum Stochastic Processes and Quantum non-Markovian Phenomena”. In: *PRX Quantum* 2 (3 July 2021), p. 030201. DOI: 10.1103/PRXQuantum.2.030201.

[133] Simon Milz et al. “Kolmogorov extension theorem for (quantum) causal modelling and general probabilistic theories”. In: *Quantum* 4 (Apr. 2020), p. 255. ISSN: 2521-327X. DOI: 10.22331/q-2020-04-20-255.

[134] Simon Milz et al. “When Is a Non-Markovian Quantum Process Classical?” In: *Phys. Rev. X* 10 (4 Dec. 2020), p. 041049. DOI: 10.1103/PhysRevX.10.041049.

[135] Simon Milz et al. “Genuine multipartite entanglement in time”. In: *SciPost Phys.* 10 (2021), p. 141. DOI: 10.21468/SciPostPhys.10.6.141.

[136] Edward F. Moore. “Gedanken-Experiments on Sequential Machines”. In: *Automata Studies. (AM-34), Volume 34*. Ed. by C. E. Shannon and J. McCarthy. Princeton: Princeton University Press, 1956, pp. 129–154. ISBN: 9781400882618. DOI: doi:10.1515/9781400882618-006.

[137] MOSEK ApS. *MOSEK Optimization Suite*. 2024. URL: https://www.mosek.com.

[138] The mpmath development team. *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 1.3.0)*. 2023. URL: http://mpmath.org/.

[139] Kazuhide Nakata et al. “Exploiting sparsity in semidefinite programming via matrix completion II: implementation and numerical results”. In: *Mathematical Programming* 95.2 (Feb. 2003), pp. 303–327. ISSN: 1436-4646. DOI: 10.1007/s10107-002-0351-9.

[140] Miguel Navascués, Masaki Owari, and Martin B. Plenio. "Power of symmetric extensions for entanglement detection". In: *Phys. Rev. A* 80 (5 Nov. 2009), p. 052306. DOI: 10.1103/PhysRevA.80.052306.

[141] Cyril Nicaud. "Average State Complexity of Operations on Unary Automata". In: *Mathematical Foundations of Computer Science 1999*. Ed. by Mirosław Kutyłowski, Leszek Pacholski, and Tomasz Wierzbicki. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 231–240. DOI: 10.1007/3-540-48340-3_21.

[142] Jiawang Nie and Li Wang. "Regularization Methods for SDP Relaxations in Large-Scale Polynomial Optimization". In: *SIAM Journal on Optimization* 22.2 (2012), pp. 408–428. DOI: 10.1137/110825844.

[143] Jan Nöller et al. *Classical certification of quantum gates under the dimension assumption*. 2024. arXiv: 2401.17006 [quant-ph].

[144] Brendan O'Donoghue et al. "Conic Optimization via Operator Splitting and Homogeneous Self-Dual Embedding". In: *J. Optim. Theory Appl* 169.3 (June 2016), pp. 1042–1068. DOI: 10.1007/s10957-016-0892-3.

[145] Brendan O'Donoghue et al. *SCS: Splitting Conic Solver, version 3.2.2*. https://github.com/cvxgrp/scs. Nov. 2022.

[146] OEIS Foundation Inc. *The On-Line Encyclopedia of Integer Sequences*. Published electronically at http://oeis.org.

[147] Ognyan Oreshkov, Fabio Costa, and Časlav Brukner. "Quantum correlations with no causal order". In: *Nat. Commun.* 3.1 (Oct. 2012), p. 1092. DOI: 10.1038/ncomms2076.

[148] Pablo A Parrilo. "Semidefinite programming relaxations for semialgebraic problems". In: *Mathematical programming* 96 (2003), pp. 293–320. DOI: 10.1007/s10107-003-0387-5.

[149] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035.

[150] Marcin Pawłowski and Časlav Brukner. "Monogamy of Bell's Inequality Violations in Nonsignaling Theories". In: *Phys. Rev. Lett.* 102 (3 Jan. 2009), p. 030403. DOI: 10.1103/PhysRevLett.102.030403.

[151] Azaria Paz. *Introduction to probabilistic automata*. Academic Press, 1971. DOI: 10.1016/C2013-0-11297-4.

[152] Juan Pablo Paz and Günter Mahler. "Proposed test for temporal Bell inequalities". In: *Phys. Rev. Lett.* 71 (20 Nov. 1993), pp. 3235–3239. DOI: 10.1103/PhysRevLett.71.3235.

[153] Asher Peres. "Separability Criterion for Density Matrices". In: *Phys. Rev. Lett.* 77 (8 Aug. 1996), pp. 1413–1415. DOI: 10.1103/PhysRevLett.77.1413.

[154] H. Petersen. "On the language of primitive words". In: *Theoretical Computer Science* 161.1 (1996), pp. 141–156. ISSN: 0304-3975. DOI: 10.1016/0304-3975(95)00098-4.

[155] Felix A. Pollock et al. "Non-Markovian quantum processes: Complete framework and efficient characterization". In: *Phys. Rev. A* 97 (1 Jan. 2018), p. 012127. DOI: 10.1103/PhysRevA.97.012127.

[156] Felix A. Pollock et al. "Operational Markov Condition for Quantum Processes". In: *Phys. Rev. Lett.* 120 (4 Jan. 2018), p. 040405. DOI: 10.1103/PhysRevLett.120.040405.

[157] Michael O Rabin. "Probabilistic automata". In: *Information and Control* 6.3 (1963), pp. 230–245. DOI: 10.1016/S0019-9958(63)90290-0.

[158] L.R. Rabiner. "A tutorial on hidden Markov models and selected applications in speech recognition". In: *Proceedings of the IEEE* 77.2 (1989), pp. 257–286. DOI: 10.1109/5.18626.

[159] Franz Rendl. "Semidefinite Relaxations for Integer Programming". In: *50 Years of Integer Programming 1958-2008: From the Early Years to the State-of-the-Art*. Ed. by Michael Jünger et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 687–726. ISBN: 978-3-540-68279-0. DOI: 10.1007/978-3-540-68279-0_18.

[160] Moritz F. Richter et al. "Classical Invasive Description of Informationally-Complete Quantum Processes". In: *Annalen der Physik* 536.4 (2024), p. 2300304. DOI: 10.1002/andp.202300304.

[161] Martin Ringbauer and Rafael Chaves. "Probing the non-classicality of temporal correlations". In: *Quantum* 1 (Nov. 2017), p. 35. ISSN: 2521-327X. DOI: 10.22331/q-2017-11-25-35.

[162] Ángel Rivas and Susana F Huelga. *Open Quantum Systems: An Introduction*. Springer Berlin, Heidelberg, 2011. ISBN: 978-3-642-23353-1. DOI: 10.1007/978-3-642-23354-8.

[163] Ángel Rivas, Susana F Huelga, and Martin B Plenio. "Quantum non-Markovianity: characterization, quantification and detection". In: *Rep. Prog. Phys.* 77.9 (Aug. 2014), p. 094001. DOI: 10.1088/0034-4885/77/9/094001.

[164] Donald J. Rose, R. Endre Tarjan, and George S. Lueker. "Algorithmic Aspects of Vertex Elimination on Graphs". In: *SIAM Journal on Computing* 5.2 (1976), pp. 266–283. DOI: 10.1137/0205021.

[165] Denis Rosset, Francesco Buscemi, and Yeong-Cherng Liang. "Resource Theory of Quantum Memories and Their Faithful Verification with Minimal Assumptions". In: *Phys. Rev. X* 8 (2 May 2018), p. 021033. DOI: 10.1103/PhysRevX.8.021033.

[166] Sebastian Ruder. *An overview of gradient descent optimization algorithms*. 2017. arXiv: 1609.04747 [cs.LG].

[167] Joshua B. Ruebeck et al. "Prediction and generation of binary Markov processes: Can a finite-state fox catch a Markov mouse?" In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.1 (Jan. 2018), p. 013109. ISSN: 1054-1500. DOI: 10.1063/1.5003041.

[168] Carlos Sabín et al. "Impurities as a quantum thermometer for a Bose-Einstein condensate". In: *Sci. Rep.* 4.1 (2014), pp. 1–6. DOI: 10.1038/srep06436.

[169] Fattah Sakuldee, Philip Taranto, and Simon Milz. "Connecting commutativity and classicality for multitime quantum processes". In: *Phys. Rev. A* 106 (2 Aug. 2022), p. 022416. DOI: 10.1103/PhysRevA.106.022416.

[170] Greg Schild and Clive Emary. "Maximum violations of the quantum-witness equality". In: *Phys. Rev. A* 92 (3 Sept. 2015), p. 032101. DOI: 10.1103/PhysRevA.92.032101.

[171] Erwin Schrödinger. "Die gegenwärtige Situation in der Quantenmechanik". In: *Naturwissenschaften* 23.48 (Nov. 1935), pp. 807–812. ISSN: 1432-1904. DOI: 10.1007/BF01491891.

[172] D. Sen. "The uncertainty relations in quantum mechanics". In: *Current Science* 107.2 (2014), pp. 203–218. ISSN: 00113891. URL: http://www.jstor.org/stable/24103129 (visited on 05/15/2024).

[173] Cosma Rohilla Shalizi and James P Crutchfield. "Computational mechanics: Pattern and prediction, structure and simplicity". In: *Journal of statistical physics* 104.3 (2001), pp. 817–879. DOI: 10.1023/A:1010388907793.

[174] Jeffrey Shallit and Ming-Wei Wang. "Automatic complexity of strings". In: *Journal of Automata, Language and Combinatorics* 6 (2001), pp. 537–544. DOI: 10.5555/543313.543324.

[175] E. Specker Simon Kochen. "The Problem of Hidden Variables in Quantum Mechanics". In: *Indiana Univ. Math. J.* 17 (1 1968), pp. 59–87. ISSN: 0022-2518. DOI: 10.1512/iumj.1968.17.17004.

[176] Steven S. Skiena. *The Algorithm Design Manual*. 2nd. Springer Publishing Company, Incorporated, 2008. ISBN: 1848000693.

[177] Paul Skrzypczyk and Daniel Cavalcanti. *Semidefinite Programming in Quantum Information Science*. 2053-2563. IOP Publishing, 2023. ISBN: 978-0-7503-3343-6. DOI: 10.1088/978-0-7503-3343-6.

[178] Morton Slater. "Lagrange Multipliers Revisited". In: *Traces and Emergence of Nonlinear Programming*. Basel: Springer Basel, 1950, pp. 293–306.

[179] Adel Sohbi et al. "Certifying dimension of quantum systems by sequential projective measurements". In: *Quantum* 5 (June 2021), p. 472. ISSN: 2521-327x. DOI: 10.22331/q-2021-06-10-472.

[180] Cornelia Spee. "Certifying the purity of quantum states with temporal correlations". In: *Phys. Rev. A* 102 (1 July 2020), p. 012420. DOI: 10.1103/PhysRevA.102.012420.

[181] Cornelia Spee. "Signaling between time steps does not allow for nonlocality beyond hidden nonlocality". In: *Journal of Physics A: Mathematical and Theoretical* 54.45 (Oct. 2021), p. 455303. DOI: 10.1088/1751-8121/ac2aea.

[182] Cornelia Spee, Costantino Budroni, and Otfried Gühne. "Simulating extremal temporal correlations". In: *New J. Phys.* 22.10 (Oct. 2020), p. 103037. DOI: 10.1088/1367-2630/abb899.

[183] Cornelia Spee et al. "Genuine temporal correlations can certify the quantum dimension". In: *New J. Phys.* 22.2 (Feb. 2020), p. 023028. DOI: 10.1088/1367-2630/ab6d42.

[184] John K. Stockton et al. "Characterizing the entanglement of symmetric many-particle spin-$\frac{1}{2}$ systems". In: *Phys. Rev. A* 67 (2 Feb. 2003), p. 022112. DOI: 10.1103/PhysRevA.67.022112.

[185] Philipp Strasberg. "Classicality with(out) decoherence: Concepts, relation to Markovianity, and a random matrix theory approach". In: *SciPost Phys.* 15 (2023), p. 024. DOI: 10.21468/SciPostPhys.15.1.024.

[186] Hemant D. Tagare. "Notes on Optimization on Stiefel Manifolds". In: 2011. URL: https://web.archive.org/web/20240310024321/https://noodle.med.yale.edu/hdtag/notes/steifel_notes.pdf.

[187] D. Tamascelli et al. "Nonperturbative Treatment of non-Markovian Dynamics of Open Quantum Systems". In: *Phys. Rev. Lett.* 120.3 (Jan. 2018), p. 030402. DOI: 10.1103/PhysRevLett.120.030402.

[188] Christino Tamon and Weichen Xie. "A Note on Quantum Markov Models". In: (2019). DOI: 10.48550/arXiv.1911.01953. arXiv: 1911.01953 [quant-ph].

[189] Philip Taranto and Simon Milz. "Hidden Quantum Memory: Is Memory There When Somebody Looks?" In: *arXiv:2204.08298* 7 (Apr. 2022), p. 991. ISSN: 2521-327x. DOI: 10.22331/q-2023-04-27-991.

[190] Philip Taranto et al. "Characterising the Hierarchy of Multi-time Quantum Processes with Classical Memory". In: *arXiv preprint arXiv:2307.11905* (2023). DOI: 10.48550/arXiv.2307.11905.

[191] Armin Tavakoli et al. "Spatial versus sequential correlations for random access coding". In: *Phys. Rev. A* 93 (3 Mar. 2016), p. 032336. DOI: 10.1103/PhysRevA.93.032336.

[192] Armin Tavakoli et al. "Semidefinite programming relaxations for quantum correlations". In: (2023). DOI: 10.48550/arXiv.2307.02551. arXiv: 2307.02551 [quant-ph].

[193] Barbara M. Terhal. "Bell inequalities and the separability criterion". In: *Phys. Lett. A* 271.5 (2000), pp. 319–326. ISSN: 0375-9601. DOI: 10.1016/s0375-9601(00)00401-1.

[194] Jihad Titi. "Matrix methods for the tensorial and simplicial Bernstein forms with application to global optimization". PhD thesis. 2019.

[195] Géza Tóth, Tobias Moroder, and Otfried Gühne. "Evaluating Convex Roof Entanglement Measures". In: *Phys. Rev. Lett.* 114 (16 Apr. 2015), p. 160501. DOI: 10.1103/PhysRevLett.114.160501.

[196] Joel A. Tropp. "Complex equiangular tight frames". In: *Wavelets XI*. Ed. by Manos Papadakis, Andrew F. Laine, and Michael A. Unser. Vol. 5914. International Society for Optics and Photonics. Spie, 2005, p. 591401. DOI: 10.1117/12.618821.

[197] Lieven Vandenberghe and Stephen Boyd. "Semidefinite Programming". In: *SIAM Review* 38.1 (1996), pp. 49–95. DOI: 10.1137/1038003.

[198] Lucas B. Vieira and Costantino Budroni. "Temporal correlations in the simplest measurement sequences". In: *Quantum* 6 (Jan. 2022), p. 623. ISSN: 2521-327x. DOI: 10.22331/q-2022-01-18-623.

[199] Lucas B. Vieira and Costantino Budroni. "Inducing sparsity in semidefinite problems with sparse objectives". (In preparation).

[200] Lucas B. Vieira, Huan-Yu Ku, and Costantino Budroni. *Entanglement-breaking channels are a quantum memory resource*. 2024. arXiv: 2402.16789 [quant-ph].

[201] Lucas B. Vieira et al. "Witnessing environment dimension through temporal correlations". In: *Quantum* 8 (Jan. 2024), p. 1224. ISSN: 2521-327X. DOI: 10.22331/q-2024-01-10-1224.

[202] Pauli Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[203] Giuseppe Vitagliano and Costantino Budroni. "Leggett-Garg macrorealism and temporal correlations". In: *Phys. Rev. A* 107 (4 Apr. 2023), p. 040101. DOI: 10.1103/PhysRevA.107.040101.

[204] Hayato Waki et al. "Sums of Squares and Semidefinite Program Relaxations for Polynomial Optimization Problems with Structured Sparsity". In: *SIAM Journal on Optimization* 1.1 (2006), pp. 218–242. DOI: 10.1137/050623802.

[205] Amos Waterland et al. "ASC: automatically scalable computation". In: *Proceedings of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems*. ASPLOS '14. Salt Lake City, Utah, USA: Association for Computing Machinery, 2014, pp. 575–590. ISBN: 9781450323055. DOI: 10.1145/2541940.2541985.

[206] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. DOI: 10.1017/9781316848142.

[207] Mirjam Weilenmann, Costantino Budroni, and Miguel Navascues. "Optimisation of time-ordered processes in the finite and asymptotic regime". In: (2023). DOI: 10.48550/arXiv.2302.02918. arXiv: 2302.02918 [quant-ph].

[208] L. Welch. "Lower bounds on the maximum cross correlation of signals (Corresp.)" In: *IEEE Transactions on Information Theory* 20.3 (1974), pp. 397–399. DOI: 10.1109/tit.1974.1055219.

[209] Stephen Wiesner. "Conjugate coding". In: *ACM Sigact News* 15.1 (1983), pp. 78–88. DOI: 10.1145/1008908.1008920.

[210] Mark M. Wilde and Ari Mizel. "Addressing the clumsiness loophole in a Leggett-Garg test of macrorealism". In: *Found. Phys.* 42.2 (Feb. 2012), pp. 256–265. ISSN: 1572-9516. DOI: 10.1007/s10701-011-9598-4.

[211] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe. *Handbook of semidefinite programming: theory, algorithms, and applications*. Vol. 27. Springer Science & Business Media, 2012. DOI: 10.1007/978-1-4615-4381-7.

[212] Kang-Da Wu et al. "Implementing quantum dimensionality reduction for non-Markovian stochastic simulation". In: *Nature Communications* 14.1 (May 2023), p. 2624. ISSN: 2041-1723. DOI: 10.1038/s41467-023-37555-0.

[213] Shibei Xue et al. "Quantum filter for a class of non-Markovian quantum systems". In: *54th IEEE Conference on Decision and Control (CDC)*. Dec. 2015, pp. 7096–7100. DOI: 10.1109/cdc.2015.7403338.

[214] Shibei Xue et al. "Modeling for Non-Markovian Quantum Systems". In: *IEEE Trans. Control Syst. Technol.* 28.6 (Nov. 2020), pp. 2564–2571. ISSN: 1558-0865. DOI: 10.1109/tcst.2019.2935421.

[215] Mihalis Yannakakis. “Computing the Minimum Fill-In is NP-Complete”. In: *SIAM Journal on Algebraic Discrete Methods* 2.1 (1981), pp. 77–79. DOI: 10.1137/0602010.

[216] Xiao-Dong Yu et al. “Quantum-Inspired Hierarchy for Rank-Constrained Optimization”. In: *PRX Quantum* 3 (1 Mar. 2022), p. 010340. DOI: 10.1103/PRXQuantum.3.010340.

[217] Xiao Yuan et al. “Universal and operational benchmarking of quantum memories”. In: *npj Quantum Information* 7.1 (July 2021), p. 108. ISSN: 2056-6387. DOI: 10.1038/s41534-021-00444-9.

[218] Yang Zheng, Giovanni Fantuzzi, and Antonis Papachristodoulou. “Chordal and factor-width decompositions for scalable semidefinite and polynomial optimization”. In: *Annu. Rev. Control* 52 (2021), pp. 243–279. ISSN: 1367-5788. DOI: 10.1016/j.arcontrol.2021.09.001.

[219] Marek Żukowski. “Temporal inequalities for sequential multi-time actions in quantum information processing”. In: *Front. Phys.* 9.5 (Oct. 2014), pp. 629–633. ISSN: 2095-0470. DOI: 10.1007/s11467-013-0400-2.