Universality and classification of elementary thermal operations

Pedro Hack[1,2,*] and Christian B. Mendl[1,3]

July 23, 2025

[1]Technical University of Munich, School of CIT, Department of Computer Science, Boltzmannstraße 3, 85748 Garching, Germany
[2]German Aerospace Center, Institute of Communication and Navigation, Müncherstraße 20, 82234 Weßling, Germany
[3]Technical University of Munich, Institute for Advanced Study, Lichtenbergstraße 2a, 85748 Garching, Germany
[*]Corresponding Author: pedro.hack@dlr.de

Abstract

Elementary thermal operations are thermal operations that act non-trivially on at most two energy levels of a system at the same time. They were recently introduced in order to bring thermal operations closer to experimental feasibility. A key question to address is whether any thermal operation could be realized via elementary ones, that is, whether elementary thermal operations are **universal**. This was shown to be false in general, although the extent to which elementary thermal operations are universal remained unknown. Here, we characterize their universality in both the sense described above and a weaker one, where we do not require them to decompose any thermal operation, but to be able to reproduce any input-output pair connected via thermal operations. Moreover, we do so for the two variants of elementary thermal operations that have been proposed, one where only deterministic protocols are allowed and one where protocols can be conditioned via the realization of a random variable, and provide algorithms to emulate thermal operations whenever their elementary counterparts are (weakly or not) universal. Lastly, we show that non-deterministic protocols reproduce thermal operations better than deterministic ones in most scenarios, even when they are not universal. Along the way, we relate elementary thermal operations to random walks on graphs.

1 Introduction

The fundamental tools we are interested in here are the **thermal operations** [1–4]. Thermal operations are quantum channels that, given a system in state

ρ with Hamiltonian H_S that is in contact with a heat bath with Hamiltonian H_B, take the form

$$\mathcal{E}(\rho) = \mathrm{Tr}_B \left[U \left(\rho \otimes \frac{e^{-\beta H_B}}{\mathrm{Tr}\left(e^{-\beta H_B}\right)} \right) U^\dagger \right], \tag{1}$$

where U is an energy-preserving unitary ($[U, H_S + H_B] = 0$), Tr_B is the partial trace over the heat bath and β is the inverse temperature.[1]

The study of thermal operations has been pursued in a field known as **resource theory**. The fundamental ideas in resource theory appeared in the chemical physics literature and are originally due to Ruch and collaborators [7–10], Alberti and Uhlmann [11, 12] and Zylka [13]. Despite this, some of the basic tools were introduced outside the realm of physics and are still used in areas ranging from information theory to economics [14, 15], with standard mathematical references in the field being the books by Marshall et al. [16] and Bhatia [17]. Regarding physics, an overview of key results and areas of application can be found in the recent review by Lostaglio [4] and the book by Sagawa [18]. The fundamental contributions to resource theory include [1,19,20], with some of the recent advances on the topic being [21–26]. Regarding applications, algorithmic cooling should be highlighted [2, 27, 28].

The conditions under which some ρ' can via achieved from ρ via some thermal operation $\mathcal{E}$, $\rho' = \mathcal{E}(\rho)$, constitute a long-standing challenge in general. However, whenever ρ or ρ' are energy incoherent, the challenge simplifies and the transitions between ρ and ρ' are fully characterized in terms of their associated population vectors p and p'. In particular, we can translate the question regarding transitions to the **thermo-majorization** relation on probability distributions. More specifically, ρ' can be achieved from ρ provided p' is thermo-majorized by p [1, 29]. Moreover, the latter condition is known to be equivalent to the existence of some stochastic matrix G such that $p' = Gp$ and $Gg = g$, where

$$g = \frac{1}{Z}\left(e^{-\beta E_1}, \ldots, e^{-\beta E_{|\Omega|}}\right)$$

is the Gibbs distribution associated to H_S and $Z = \sum_{i=1}^{|\Omega|} e^{-\beta E_i}$ its partition function [23]. That is, G is a g-stochastic matrix that maps p to p'.[2] Lastly, thermo-majorization relations provide fundamental constraints for general thermodynamic transitions since they characterize the population dynamics induced by arbitrary thermal operations [1, 2, 30].

[1]The restriction to an energy-preserving unitary is justified via the notion of passivity [2, 5, 6].

[2]In general, for $d \in \mathbb{P}_\Omega$, we denote $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ as d-stochastic provided it is stochastic, i.e.,

$$\sum_{i=1}^{|\Omega|} M_{i,j} = 1 \text{ for } 1 \leqslant j \leqslant |\Omega|,$$

and has d as equilibrium distribution $Md = d$ [16]. Here, $\mathbb{P}_\Omega$ stands for the set of probability distributions over some finite set Ω (also known as **standard simplex** and often denoted by $\Delta^{|\Omega|-1}$) and $\mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ for the set of $|\Omega| \times |\Omega|$ matrices with real entries.

In this context, a natural subset of the g-stochastic matrices considered in resource theory is the one where we only allow G to act non-trivially on at most two elements of the state space, the so-called **elementary thermal operations** [2]. Theoretically, this subset of stochastic matrices has already been considered in the study of the fundamental notion of uncertainty known as majorization for its simplicity [16,17,31]. However, from a resource-theoretic point of view, the recent interest in elementary thermal operations comes from experimental considerations. To draw an analogy, the situation is quite similar to the decomposition of unitary matrices in terms of circuits involving at most two-level gates [32], which was pursued given that experimental implementations of such gates via optical devices are well-known [33].[3] In our resource-theoretic context, the main issue is that universal sets of thermal operations, like the so-called **crude** operations [34], often require control over the global system and heat bath. The fact that this issue is considerably improved upon when using elementary thermal operations, which can be well reproduced via the collision or the Jaynes-Cummings model [2, 35, 36], motivated their recent introduction in resource theory [2].

Following the analogy with [32], a fundamental question regarding elementary thermal operations is to what extent they can be used to implement thermal operations, that is, whether they are universal or not. A question analogous to that in [32] would be whether every g-stochastic matrix can be decomposed as a product of elementary thermal operations. A more general question [2] would be whether the same holds if we take several sets of products of elementary thermal operations and condition which one we use on the realization of some random variable (for instance, a coin toss). This is equivalent to asking whether every g-stochastic matrix can be decomposed as a convex combination of products of elementary thermal operations. (It should be noted that, while the incorporation of such a random variable has the clear drawback that post-processing will be required, its advantages will become clear later on. For the moment, it suffices to say that its incorporation does not seem to add further restrictions regarding experimental applicability.)

The fundamental questions above have not yet been fully answered, although some results are known. In this regard, [2] showed that elementary thermal operations (in the stronger form and, hence, in the weaker one without convex combinations), are not universal in general, that is, for any Gibbs distribution. However, whenever the Gibbs distribution is uniform, it is well-known that they are universal [31,37,38]. Despite these results, no characterization of the universality of elementary thermal operations in terms of $g \in \mathbb{P}_\Omega$ is known. Moreover, there is no characterization of universality in the weaker sense, where we do not necessarily require elementary thermal operations to decompose any thermal operation, but we ask whether we can find some elementary thermal operations connecting each pair $p, p' \in \mathbb{P}_\Omega$ that is connected by thermal operations.

[3] Regarding the formal relation between both questions, it is easy to see that the only stochastic matrices that are unitary are the permutation matrices. ($M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ is a **permutation** matrix if it has a single one in each row and column with the remaining entries being zero [16].)

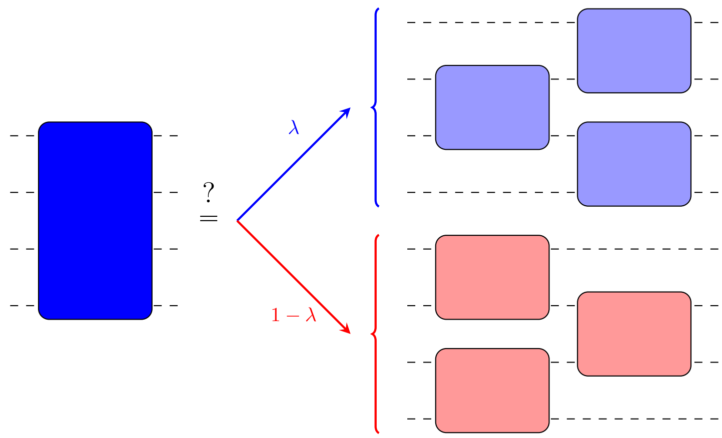


Figure 1: Schematic representation of (Q2): When can thermal operations be decomposed as convex combinations of sequences of two-level thermal operations? In this representation, we ask whether a thermal operation acting on four energy levels (left of the equality) can be decomposed as a convex combination (consisting of two elements with weights λ and $1-\lambda$ in the figure) of two sequences of thermal operations that act only on two energy levels simultaneously.

We will refer to this as **weak universality**. Lastly, there is no characterization of the extent to which the incorporation of random variables (and, hence, post-processing) to elementary thermal operations offers an advantage regarding universality. These constitute our main questions. In summary, we ask:

(Q1) When are elementary thermal operations **weakly** universal?

(Q2) When are elementary thermal operations universal?

(Q3) When does the incorporation of random variables to the elementary thermal operations offer an advantage regarding weak or non-weak universality?

A schematic representation of (Q2) can be found in Figure 1. In particular, we represent the stronger framework where convex combinations are allowed.

1.1 Outline

As our main results, we answer (Q1) and (Q2) fully and make substantial progress on (Q3). We give an overview of our main results in Section 3.

Regarding structure, we first introduce the resource theories and polytopes we consider here in Section 2. We then gather some known results regarding their relationship when the Gibbs distribution is uniform and some basic remarks for the following (Sections 4 and 5). As a first step towards our main questions, we characterize a restricted version of (Q2) in Section 6. Question (Q1) is resolved in Section 7, and (Q2) in Section 8. Our progress regarding (Q3) can be found in Section 9. Lastly, we address the convexity of elementary thermal operations as a consequence of our work regarding (Q3) in Section 10, and relate resource theories to random walks on graphs in Appendix B.

2 Thermal and elementary thermal operations

In this section, we introduce the thermal operations we are concerned with here. Note that, throughout this work, we consider $0 < d \in \mathbb{P}_\Omega$, the only physically relevant situation since, for $1 \leqslant i \leqslant |\Omega|$, $d_i = 0$ implies E_i is infinite. (The reader interested in the case $0 \leqslant d$ may find [39] useful.) Moreover, some of the definitions in this section are usually written for the Gibbs distribution and, hence, do not include the partition function [2, 4]. We simply take some $0 < d \in \mathbb{P}_\Omega$ and note this results in no meaningful difference.

2.1 Thermal operations

We begin by introducing the well-known polytope and resource theory of thermal operations [2, 4].

Definition 1 (Resource theory and polytope of thermal operations (TO))**.** *If $0 < d \in \mathbb{P}_\Omega$ and $\mathcal{P}_{TO}(d)$ is the set of d-stochastic matrices $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$, then the* ***resource theory of thermal operations*** *$\mathcal{R}_{TO}(d)$ is the set*

$$\mathcal{R}_{TO}(d) := \{(p, Mp) : p \in \mathbb{P}_\Omega, M \in \mathcal{P}_{TO}(d)\},$$

that is, $\mathcal{R}_{TO}(d)$ is the set of all possible state transfers $p \to q := Mp$ allowed by d-stochastic matrices. We refer to $\mathcal{P}_{TO}(d)$ as the ***polytope of thermal operations****.*

Note that our definition of thermal operations is somewhat restrictive: It corresponds to the **action of thermal operations on populations** [4, Section 2.1].

In what follows, we will use the term **polytope** to refer to a set of allowed operations and the term **resource theory** to the set of input-output pairs that these operations map.[4]

We say that $q \in \mathbb{P}_\Omega$ can be **achieved** from $p \in \mathbb{P}_\Omega$ via thermal operations (and analogously for the resource theories we will introduce later on) provided $(p, q) \in \mathcal{R}_{\mathrm{TO}}(d)$.

An easy way to determine whether a transition from p to q is possible is to use the ***d*-majorization curve** [4]:

[4] Strictly speaking, a **convex polytope** or **polytope** for simplicity is the convex hull of a nonempty finite set [40]. Although the polytope of thermal operations is indeed a polytope (see Section 8 or [39, 41]), we will sometimes use the term **polytope** loosely. We will return to this point later on.

Definition 2 (d-majorization curve)**.** *If $p, d \in \mathbb{P}_\Omega$ and $0 < d$, then the **d-majorization curve** associated to p can be constructed as follows:*

(a) We consider a permutation $\Pi_p^d \in S_{|\Omega|}$ such that, for $1 \leqslant i < |\Omega|$,

$$\frac{p_{(\Pi_p^d)^{-1}(i)}}{d_{(\Pi_p^d)^{-1}(i)}} \geqslant \frac{p_{(\Pi_p^d)^{-1}(i+1)}}{d_{(\Pi_p^d)^{-1}(i+1)}}. \tag{2}$$

We call Π_p^d the d-permutation of p.[5] *Moreover, we refer to $(\Pi_p^d)^{-1}$ as the d-order of p and define*

$$p^d := \left(p_{(\Pi_p^d)^{-1}(i)}\right)_{i=1}^{|\Omega|}.$$

*Furthermore, for $1 \leqslant i \leqslant |\Omega|$, we say p_i is **associated to** α if $d_i = \alpha$ and extend this definition naturally to p_i^d.*

(b) Plot the pairs of points

$$\left(\sum_{i=1}^{k} d_{(\Pi_p^d)^{-1}(i)}, \sum_{i=1}^{k} p_{(\Pi_p^d)^{-1}(i)}\right)_{k=1}^{|\Omega|}$$

*together with $(0,0)$ and connect them piecewise linearly to form a concave curve. We call this curve the **d-majorization curve** and denote it by c_p^d. It should be noted that the curve is also known as **Lorenz d-curve** or **thermo-majorization curve**.*

We say that p **d-majorizes** or **thermo-majorizes** q, which we denote by $q \preceq_d p$, if c_p^d is never below c_q^d. The reason we include these definitions is that it is well-known [2, 4, 10, 42] that, for any pair $p, q \in \mathbb{P}_\Omega$, there exists a d-stochastic matrix M such that $q = Mp$ if and only if p d-majorizes q, $q \preceq_d p$. This will prove to be useful later on.

2.2 Elementary thermal operations

To introduce the resource theories and polytopes of elementary thermal operations, we need some more definitions. We will denote by $d^\downarrow$ the distribution that results from permuting the components of some $d \in \mathbb{P}_\Omega$ until they are arranged in non-increasing order.

It is not hard to see that, if $Q \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ is a permutation matrix such that $d^\downarrow = Qd$ and $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ is a d-stochastic matrix acting non-trivially on at most two levels, then

$$M^\downarrow := QMQ^T$$

[5] This constitutes an abuse of language, since we should call it **a** d-permutation whenever there is some equality in (2). However, we assume that Π_p^d is unique in general and deal with the problematic scenario when needed, referring to it as **uncertainty** in Π_p^d. Lastly, note that the most pathological case is $p = d$, since any permutation can be the d-permutation of d.

is a $d^{\downarrow}$-stochastic matrix acting non-trivially on at most two levels. (In the following, we assume some fixed Q for each d in order for $M^{\downarrow}$ to be well-defined.) Moreover, by definition, there exist i, j with $1 \leqslant i \leqslant j \leqslant |\Omega|$ and $x, y, z, t \in [0,1]$ such that

$$M^{\downarrow} = \begin{pmatrix} x & y \\ z & t \end{pmatrix} \bigoplus \mathbb{I}_{\setminus(i,j)} = \begin{pmatrix} 1-\gamma_{i,j}y & y \\ \gamma_{i,j}y & 1-y \end{pmatrix} \bigoplus \mathbb{I}_{\setminus(i,j)} = (1-y)\mathbb{I} + yP^{d^{\downarrow}}(i,j), \tag{3}$$

where $\gamma_{i,j} := d_j^{\downarrow}/d_i^{\downarrow}$ and

$$P^{d^{\downarrow}}(i,j) := \begin{pmatrix} 1-\gamma_{i,j} & 1 \\ \gamma_{i,j} & 0 \end{pmatrix} \bigoplus \mathbb{I}_{\setminus(i,j)}.$$

(Note that in (3) we mean that $M^{\downarrow}$ is identity except for $M^{\downarrow}_{k,\ell}$ with $k, \ell \in \{i, j\}$, where it takes the values we specify on the left of $\bigoplus$.)

As a result, we have the following decomposition of M:

$$M = (1-y)\mathbb{I} + yQ^T P^{d^{\downarrow}}(i,j)Q. \tag{4}$$

The decomposition in (4) motivates the following definition, which will be key in order to introduce the elementary resource theories we consider here.

Definition 3 (d-swap and T^d-transform)**.** *If $0 < d \in \mathbb{P}_\Omega$, $Q \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ is a permutation matrix such that $d^{\downarrow} = Qd$ and $1 \leqslant i \leqslant j \leqslant |\Omega|$, then we call $P^d(i,j) \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ a d-**swap** provided*

$$P^d(i,j) = Q^T P^{d^{\downarrow}}(i,j)Q. \tag{5}$$

*In the same scenario, we call $T^d_\lambda(i,j) \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ a T^d-**transform** provided there exists some $0 \leqslant \lambda \leqslant 1$ such that*

$$T^d_\lambda(i,j) = (1-\lambda)\mathbb{I} + \lambda P^d(i,j). \tag{6}$$

As shown in (4), T^d-transforms constitute all d-stochastic matrices that act non-trivially on at most two levels. Moreover, (4) also shows that any such matrix can be decomposed as a convex combination of d-swaps. Moreover, since it uses the Gibbs distribution notation for some inverse temperature β, d-swaps are called β-swaps in [2]. (In fact, in [2], β-swaps were only introduced for $d = d^{\downarrow}$.) T^d transforms are sometimes referred to as **simple** d-stochastic matrices [16] or as **elementary thermal operations** [2]. We follow the notation in [16, 31] when dealing with the uniform case $d = u := (1/|\Omega|, \ldots, 1/|\Omega|)$ and refer to T^u-transforms as T-transforms. Another name for T^u transforms is **elementary doubly stochastic matrices** [43].

Definition 3 allows us to introduce two pairs of elementary thermal resource theories and polytopes, which we call **strong** and **weak**.

Definition 4 (Resource theory and polytope of weak elementary thermal operations (WETO))**.** *If $0 < d \in \mathbb{P}_\Omega$ and $\mathcal{P}_{WETO}(d)$ is the set of finite sequences* *of T^d-transforms, then the* ***resource theory of weak elementary thermal operations*** *$\mathcal{R}_{WETO}(d)$ is the set*

$$\mathcal{R}_{WETO}(d) := \{(p, Mp) : p \in \mathbb{P}_\Omega, M \in \mathcal{P}_{WETO}(d)\}.$$

We refer to $\mathcal{P}_{WETO}(d)$ as the ***polytope of weak elementary thermal operations****.*[6]

Sometimes the literature refers to the fact that some $q \in \mathbb{P}_\Omega$ can be achieved from some $p \in \mathbb{P}_\Omega$ via the resource theory of weak elementary thermal operations by saying that q is **simply d-majorized** by p [16, Definition 14.B.2.a].

Definition 5 (Resource theory and polytope of strong elementary thermal operations (ETO) [2, Definition 1])**.** *If $0 < d \in \mathbb{P}_\Omega$ and $\mathcal{P}_{ETO}(d)$ is the set of convex combinations of finite sequences of d-swaps, then the* ***resource theory of strong elementary thermal operations*** *$\mathcal{R}_{ETO}(d)$ is the set*

$$\mathcal{R}_{ETO}(d) := \{(p, Mp) : p \in \mathbb{P}_\Omega, M \in \mathcal{P}_{ETO}(d)\}.$$

We refer to $\mathcal{P}_{ETO}(d)$ as the ***polytope of weak elementary thermal operations****.*

The ETO resource theory and polytope are the convex hull of the WETO resource theory and polytope, respectively. Hence, we can think of them as the incorporation of random variables to WETO protocols.

Definition 5 and [2, Definition 1] are equivalent. In [2], what we call the **resource theory of strong elementary thermal operations** is referred to as the **resource theory of elementary thermal operations** and, moreover, d-swaps are substituted by T^d transforms (which they call ETOs). In any case, it is easy to see that they are equivalent [2].[7]

If we only consider the sequences of d-swaps with a single d-swap, then we refer to Definition 5 as the **length one** resource theory (and polytope) of strong elementary thermal operations. We analyze this simplified scenario in Section 6 and connect it with random walks on graphs in Appendix B.

The following section provides an overview of the main results in this work.

3 Main results

Our first main result is the following:

Theorem. *If $0 < d \in \mathbb{P}_\Omega$, then the following statements are equivalent:*

(a) $\mathcal{R}_{TO}(d) = \mathcal{R}_{ETO}(d)$.

[6] This constitutes an abuse of language since the WETO polytope is not convex in general. We give an example after the proof of Theorem 9 and characterize its convexity in Corollary 3.

[7] The fact that our definition is included in theirs holds directly and, for the converse, one simply ought to develop the products of T^d-transforms for each sequence in the convex combination and note that the result is indeed a convex combination of products of d-swaps.

(b) $\mathcal{R}_{TO}(d) = \mathcal{R}_{WETO}(d)$*.*

(c) $\mathcal{P}_{TO}(d) = \mathcal{P}_{ETO}(d)$*.*

(d) $d^{\downarrow} = (d_0, \dots, d_0, d_1)$*.*

The equivalence between (a) and (d) is proven in Theorem 5. This equivalence characterizes the weak universality of strong elementary thermal operations, and is part of our answer to (Q1). The equivalence between (b) and (d) is shown in Theorem 6. This equivalence characterizes the weak universality of weak elementary thermal operations, and is part of our answer to (Q1). The equivalence between (c) and (d) is proven in Theorem 8. This equivalence characterizes the universality of strong elementary thermal operations, and is part of our answer to (Q2).

Our second main result is the following:

Theorem. *If* $0 < d \in \mathbb{P}_\Omega$*, then the following statements are equivalent:*

(a) $\mathcal{P}_{TO}(d) = \mathcal{P}_{WETO}(d)$*.*

(b) $\mathcal{P}_{ETO}(d) = \mathcal{P}_{WETO}(d)$*.*

(c) $|\Omega| = 2$*.*

The equivalence between (a) and (c) is proven in Theorem 9. This equivalence characterizes the universality of weak elementary thermal operations, and is part of our answer to (Q2). We show the equivalence between (b) and (c) in Theorem 11. This equivalence characterizes the advantage regarding universality that one can achieve by incorporating random variables to the elementary thermal operations, and is part of our answer to (Q3).

Before we state the last main result, we need to introduce **quasi-uniform** distributions.

Definition 6 (Quasi-uniform distribution)**.** $0 < d \in \mathbb{P}_\Omega$ *is a quasi-uniform distribution provided it has, at most, two different entries* $d_i \in \{x, y\}$ *for all* $i \in \Omega$*.*[8]

Using the definition above, we can now state our last main theorem.

Theorem. *If* $0 < d \in \mathbb{P}_\Omega$*, then* $\mathcal{R}_{ETO}(d) = \mathcal{R}_{WETO}(d)$ *only if* d *is quasi-uniform.*

The implication is proven in Theorem 10. This result shows that, in most situations, one can achieve an advantage regarding weak universality by incorporating random variables to the elementary thermal operations, and is part of our answer to (Q3). Provided we are restricted to $|\Omega| = 3$, the converse also holds, as proven in Proposition 4.

[8] For instance, for $|\Omega| = 4$, the only (non-increasingly ordered) quasi-uniform distributions are the uniform distribution u and, taking $d_0 > d_1$, $p_1 = (d_0, d_1, d_1, d_1)$, $p_2 = (d_0, d_0, d_1, d_1)$ and $p_3 = (d_0, d_0, d_0, d_1)$.

Since the main aim of this work is to study the relation between the different resource theories and polytopes we have introduced, we begin by considering the known results for the case where $d = u$ is uniform in the following section. (See also [2, Section 2.3].)

4 Universality and weak universality for uniform equilibrium distributions

The first thing we ought to remark is that all resource theories coincide provided $d = u$. This follows directly from the following theorem, which was originnally developed by Muirhead [37] and was generalized by Hardy et al. [31] to encompass probability distributions.

Theorem 1 (Equivalence TO and WETO resource theories for uniform d [31, 37])**.** *If $d \in \mathbb{P}_\Omega$ is the uniform distribution u, then the thermal operations resource theory is equal to the weak elementary thermal operations resource theory*

$$\mathcal{R}_{TO}(u) = \mathcal{R}_{WETO}(u).$$

Moreover, since

$$\mathcal{R}_{\mathrm{WETO}}(d) \subseteq \mathcal{R}_{\mathrm{ETO}}(d) \subseteq \mathcal{R}_{\mathrm{TO}}(d)$$

for all $d \in \mathbb{P}_\Omega$ by definition, all thermal resource theories considered here are equivalent provided the equilibrium distribution is uniform.

When dealing with thermal operations and (strong) elementary thermal operations, the equivalence can be generalized to polytopes as a direct consequence of Birkhoff's theorem [38] (also sometimes attributed to von Neumann [44]) together with the fact that permutation matrices can be decomposed as products of transpositions.

Theorem 2 (Equivalence TO and ETO polytopes for uniform d [38, 44])**.** *If $d \in \mathbb{P}_\Omega$ is the uniform distribution u, then the polytope of thermal operations is equivalent to the polytope of elementary thermal operations*

$$\mathcal{P}_{TO}(u) = \mathcal{P}_{ETO}(u).$$

The equivalence at the polytope level, however, does not include weak elementary thermal operations, as shown by Marcus et al. [43]. (See also [16].)

Theorem 3 (Equivalence TO and WETO polytopes for uniform d [43, Theorem 1])**.** *If $d \in \mathbb{P}_\Omega$ is the uniform distribution u, then the following statements are equivalent:*

(a) The weak elementary thermal operations polytope is equivalent to the thermal operations polytope

$$\mathcal{P}_{TO}(u) = \mathcal{P}_{WETO}(u).$$

(b) $|\Omega| = 2$.

In the following, we address the generalization of Theorems 1, 2 and 3 to arbitrary equilibrium distributions $0 < d \in \mathbb{P}_\Omega$. Before we start showing our main results, we make some basic remarks that will be useful in the future.

5 Basic observations for the general case

Recall that we only deal with the case where $0 < d \in \mathbb{P}_\Omega$. As a first remark, we note that, given an arbitrary $0 < d \in \mathbb{P}_\Omega$, it is equivalent to show the relation between the different resource theories and polytopes for $d^\downarrow$ than to do so for d. In order to show this, we use the following terminology

$$\mathcal{P}(p) := \{\mathcal{P}_{\text{TO}}(p), \mathcal{P}_{\text{ETO}}(p), \mathcal{P}_{\text{WETO}}(p)\},$$
$$\mathcal{R}(p) := \{\mathcal{R}_{\text{TO}}(p), \mathcal{R}_{\text{ETO}}(p), \mathcal{R}_{\text{WETO}}(p)\}$$

for all $p \in \mathbb{P}_\Omega$.

Lemma 1 (Equivalence under permutation of d)**.** *Consider $0 < d \in \mathbb{P}_\Omega$. If we take a pair of polytopes $A(d), B(d) \in \mathcal{P}(d)$ or resource theories $A(d), B(d) \in \mathcal{R}(d)$ of d and the corresponding polytopes $A(d^\downarrow), B(d^\downarrow) \in \mathcal{P}(d^\downarrow)$ or resource theories $A(d^\downarrow), B(d^\downarrow) \in \mathcal{R}(d^\downarrow)$ of $d^\downarrow$, then the following statements are equivalent:*

(a) $A(d) \subseteq B(d)$.

(b) $A(d^\downarrow) \subseteq B(d^\downarrow)$.

Note that Lemma 1 works more in general for any permutation of d, but we focus on $d^\downarrow$ since it is the one we will use later on.

As a second remark, it is straightforward to relate the different resource theories and polytopes for the smallest dimension, i.e. for $|\Omega| = 2$. In particular, as a consequence of (4), we have the following lemma.

Lemma 2 (Equivalence TO, ETO and WETO polytopes for $|\Omega| = 2$)**.** *If $0 < d \in \mathbb{P}_\Omega$ and $|\Omega| = 2$, then the polytopes of thermal and (weak and strong) elementary thermal operations are equivalent*

$$\mathcal{P}_{TO}(d) = \mathcal{P}_{ETO}(d) = \mathcal{P}_{WETO}(d).$$

As a direct consequence of Lemma 2, all resource theories considered here coincide $\mathcal{R}_{\text{TO}}(d) = \mathcal{R}_{\text{ETO}}(d) = \mathcal{R}_{\text{WETO}}(d)$ provided $|\Omega| = 2$.

As a last remark, the following straightforward property of elementary thermal operations will prove to be useful in the following, specially when justifying no-go results.

Lemma 3 (Monotonicity of support under ETO [2, Lemma 4])**.** *If $0 < d \in \mathbb{P}_\Omega$ and $p, q \in \mathbb{P}_\Omega$ such that $q \in \mathcal{C}_d^{ETO}(p)$, then $|supp(p)| \leqslant |supp(q)|$, where $\mathcal{C}_d^{ETO}(p) := \{q \in \mathbb{P}_\Omega | q = Mp, M \in \mathcal{P}_{ETO}(d)\}$ and $supp(p) := \{i \in \Omega | p_i > 0\}$ for all $p \in \mathbb{P}_\Omega$.*

Before addressing our main questions, in the next section, we characterize the simplest scenario where ETO protocols incorporating random variables appear: the length one ETO polytope.

6 Characterizing the length one ETO polytope

As a first result, in the following theorem, we characterize the d-stochastic matrices M that belong to the length one ETO polytope. This result is closely related random walks on graphs, as we show in Appendix B. Note that we say a d-stochastic matrix $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ satisfies **detailed balance** provided $M_{i,j}d_j = M_{j,i}d_i$ for $1 \leqslant i,j \leqslant |\Omega|$ [45].

Theorem 4 (Characterization of the length one ETO polytope)**.** *If $0 < d \in \mathbb{P}_\Omega$ and $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$, then the following statements are equivalent:*

(a) M can be decomposed as a convex combination of d-swaps.

(b) $M^\downarrow$ is stochastic, it satisfies detailed balance and there exists some $\lambda \in [0,1]$ such that

$$M^\downarrow_{i,i} = \lambda + \sum_{i<j\leqslant|\Omega|} \left(1 - \frac{d^\downarrow_j}{d^\downarrow_i}\right) M^\downarrow_{i,j} + \sum_{\substack{1\leqslant k<j\leqslant|\Omega| \\ k,j\neq i}} M^\downarrow_{k,j} \tag{7}$$

for $1 \leqslant i \leqslant |\Omega|$.

Remark 1 (Algorithm)**.** *Theorem 4 provides a simple algorithm to determine whether some matrix belongs to the length one ETO polytope and, in case it does, it also returns the weights to decompose such a matrix in terms of d-swaps. Up to permutations in d, the algorithm can be summarized as follows:*

(a) Input $0 < d \in \mathbb{P}_\Omega$ and $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$.

(b) Check M is stochastic and satisfies detailed balance.

(c) Calculate λ_i for $1 \leqslant i \leqslant |\Omega|$ following (7) and check $\lambda_i = \lambda$ for $1 \leqslant i \leqslant |\Omega|$ and $0 \leqslant \lambda \leqslant 1$.

(d) Output the decomposition of M in terms of d-swaps: λ times the identity plus $M_{i,j}$ times the d-swap acting non-trivially on the components i and j for all $i < j$.

We prove Theorem 4 in Appendix A. If $|\Omega| = 2$, note that any d-stochastic matrix M satisfies detailed balance and (7) always holds with $\lambda = 1 - M_{1,2}$. Moreover, in case $d \in \mathbb{P}_\Omega$ is the uniform distribution and for arbitrary Ω, then detailed balance reduces to M being symmetric and (7) to

$$M_{i,i} = \lambda + \sum_{\substack{1\leqslant k<j\leqslant|\Omega| \\ k,j\neq i}} M_{k,j} \tag{8}$$

for $1 \leqslant i \leqslant |\Omega|$ and $\lambda \in [0,1]$.

7 Weak universality of elementary thermal operations

In this section, we answer (Q1) by considering first non-determinisitic protocols in Section 7.1 and then determinisitc ones in Section 7.2.

7.1 Weak universality of strong elementary thermal operations

The first question we address is the relation between the thermal and elementary thermal resource theories. We begin by extending [2, Corollary 5], where it was shown that they do not always coincide.[9]

Proposition 1 (Difference TO and ETO resource theories)**.** *If $0 < d \in \mathbb{P}_\Omega$, then the thermal operations resource theory is equal to the elementary thermal operations resource theory*

$$\mathcal{R}_{TO}(d) = \mathcal{R}_{ETO}(d)$$

only if $d^\downarrow = (d_0, \dots, d_0, d_1)$.

As a result of Proposition 1, even if we condition our experimental protocols using random variables, elementary thermal operations are not weakly universal (hence not universal) provided $d^\downarrow \neq (d_0, \dots, d_0, d_1)$. We prove Proposition 1 in Appendix D. Given that $\mathcal{C}_d^{ETO}(p)$ is a closed set for all $p, d \in \mathcal{P}_\Omega$ with $0 < d$ [2, Theorem 6], Proposition 1 actually shows that there exists some $q \in \mathbb{P}_\Omega$ such that there is no sequence of distributions $(q_\varepsilon)_\varepsilon \subseteq \mathbb{P}_\Omega$ which are ε-close to q and achievable from p via elementary thermal operations $q_\varepsilon \in \mathcal{C}_d^{ETO}(p)$.

Since it seems we cannot extend Proposition 1 further, let us establish the equivalence between both resource theories for a low dimensional Ω, where we can calculate everything explicitly. In fact, we can directly show the equivalence at the polytope level, as we do in the following proposition.

Proposition 2 (Equivalence TO and ETO polytopes for $|\Omega| = 3$)**.** *If $0 < d \in \mathbb{P}_\Omega$ and $|\Omega| = 3$, then the following statements are equivalent:*

(a) The thermal operations polytope is equal to the elementary thermal operations polytope
$$\mathcal{P}_{TO}(d) = \mathcal{P}_{ETO}(d).$$

(b) $d^\downarrow = (d_0, d_0, d_1)$.

As a result of Proposition 2, if $|\Omega| = 3$, elementary thermal operations with conditional protocols are universal (and, as one can easily see, also weakly universal) if and only if $d^\downarrow = (d_0, \dots, d_0, d_1)$. We prove Proposition 2 in Appendix

[9] The impossibility of extending Theorem 2 in general was already mentioned by Veinott in [46, p. 2], which points to a manuscript called *On d-majorization and d-Schur convexity* by the same author. The latter was not published according to [16].

E. The proof is straightforward since we can explicitly calculate the extreme points in the TO polytope. As we will show in Theorem 8, we can use the tools in [39,41] and follow a similar strategy for arbitrary dimensions. However, we can use a simpler approach provided we are only interested in the relation between resource theories. To do so, we rely on the following lemma, where $S_{|\Omega|}$ denotes the set of permutations over Ω.

Lemma 4 (Extreme points thermal cone [2, Lemma 12]). *If $0 < d \in \mathbb{P}_\Omega$ and $p, q \in \mathbb{P}_\Omega$ with $q \preceq_d p$, then q can be written as a convex combination of elements in the set $\{p^\Pi\}_{\Pi \in S_{|\Omega|}}$, where, recalling the notation in Definition 2,*

(a) $x_i^\Pi := \sum_{j=1}^{i} d_{\Pi^{-1}(j)}$, and $y_i^\Pi := c_p^d(x_i^\Pi)$, and

(b) $p_i^\Pi := y_{\Pi(i)}^\Pi - y_{\Pi(i)-1}^\Pi$, with $y_0 := 0$,

for $\Pi \in S_{|\Omega|}$ and $1 \leqslant i \leqslant |\Omega|$.

We are now in position to characterize the equivalence between TO and ETO resource theories in general. To show this, in the spirit of the definition of $\mathcal{C}_d^{ETO}(p)$ in Lemma 3, we use the notation

$$\mathcal{C}_d(p) := \{q \in \mathbb{P}_\Omega | q \preceq_d p\} \tag{9}$$

for any $p, d \in \mathbb{P}_\Omega$, $0 < d$. (Later on, we will use the notation $\mathcal{C}_d^{WETO}(p)$ for the natural extensions of these definitions to WETO.)

Theorem 5 (Relation TO and ETO resource theories). *If $0 < d \in \mathbb{P}_\Omega$, then the following statements are equivalent:*

(a) The thermal operations resource theory is equal to the elementary thermal operations resource theory

$$\mathcal{R}_{TO}(d) = \mathcal{R}_{ETO}(d).$$

(b) $d^\downarrow = (d_0, \ldots, d_0, d_1)$.

As a result of Theorem 5, experimental protocols conditioned by a random variable are weakly universal if and only if $d^\downarrow = (d_0, \ldots, d_0, d_1)$. We prove Theorem 5 in Appendix F. At a high level, in order to prove that (b) implies (a), we use the characterization of the extreme points of the thermal resource theory in Lemma 4 to construct an explicit sequence of elementary thermal operations for each extreme point. In order to do so, we exploit the simple structure of the equilibrium distributions d considered in the statement and divide the argument in terms of the d-permutation each extreme point may have.

Remark 2 (Algorithm). *Theorem 5 together with linear programming (which allows us to find the weights decomposing a point that belongs to the convex hull of some finite set [47]) gives us an algorithm to achieve q via (strong) elementary thermal operations on p provided $d^\downarrow = (d_0, \ldots, d_0, d_1)$. Up to permutations in d, the algorithm can be summarized as follows:*

(a) *Input* $0 < d = (d_0, \ldots, d_0, d_1) \in \mathbb{P}_\Omega$ *and* $p, q \in \mathbb{P}_\Omega$ *such that* $q \preceq_d p$.

(b) *Calculate the decomposition of* q *in terms of the extreme points of* $\mathcal{C}_d(p)$ *using linear programming.*

(c) *Calculate* m *and* s *as in the proof of Theorem 5 for each extreme point obtained in (b) and use them, following again Theorem 5, to obtain a sequence of* d*-swaps that (up to permutations) yield the extreme point when applied to* p.

(d) *Find the sequence of swaps that are lacking in (c) in order to achieve each extreme point via* d*-swaps.*

(e) *Output a convex combination of products of* d*-swaps that yield* q *when applied to* p.

Theorem 5 exemplifies the impact that degeneracy can have on a thermodynamic system: Different realizations of **effective two-dimensional** Hamiltonians, that is, those that have (up to degeneracy) two energy levels, behave differently in terms of weak universality. For instance, $d = (d_0, \ldots, d_0, d_1)$ is universal and $d' = (d_0, d_1 \ldots, d_1)$ is not universal provided $d_0 > d_1$. Intuitively, this follows from the asymmetry in the d-swaps whenever the two non-trivially involved energy levels differ.

Theorem 5 illustrates that the general bounds on the number of d-swaps required to reach the extreme distributions in the ETO resource theory can be substantially improved whenever $d^\downarrow = (d_0, \ldots, d_0, d_1)$. In particular, we have shown that any extreme point of $\mathcal{C}_d(p)$ can be obtained using $|\Omega|$ d-swaps. This contrasts with the known general bounds, which escalate like $|\Omega|!$ [2, Theorem 6].

7.2 Weak universality of weak elementary thermal operations

In order to study the relation between the the thermal and weak elementary thermal resource theories, we start again by characterizing the low-dimensional case, where we can calculate everything explicitly.

Proposition 3 (Equivalence TO and WETO resource theories for $|\Omega| = 3$)**.** *If* $0 < d \in \mathbb{P}_\Omega$ *and* $|\Omega| = 3$*, then the following statements are equivalent:*

(a) *The thermal operations resource theory is equal to the weak elementary thermal operations resource theory*

$$\mathcal{R}_{TO}(d) = \mathcal{R}_{WETO}(d).$$

(b) $d^\downarrow = (d_0, d_0, d_1)$.

As a result of Proposition 3, if $|\Omega| = 3$, elementary thermal operations with deterministic protocols are weakly universal if and only if $d^{\downarrow} = (d_0, d_0, d_1)$. We prove Proposition 3 in Appendix G. Note that we deal with resource theories in Proposition 3 instead of polytopes as in Proposition 2. The reason for this will become clear later on.

The characterization in Proposition 3 holds in general. In order to show this, we require a couple more definitions that will help us quantify the difference between two probability distributions.

Definition 7 (Difference maps h_0 and h_1)**.** *Given $p, q, d \in \mathbb{P}_\Omega$ with $0 < d$, the h_0* ***difference map*** *(with respect to d) is defined as follows:*

$$\begin{aligned} h_0 : \ \mathbb{P}_\Omega \times \mathbb{P}_\Omega &\to \{0, 1, \ldots, |\Omega| - 1\} \\ (p,q) &\mapsto |\Pi_p^d(|\Omega|) - \Pi_q^d(|\Omega|)|, \end{aligned}$$

where, whenever there is uncertainty in either $\Pi_p^d(|\Omega|)$ or $\Pi_q^d(|\Omega|)$, we assume for simplicity that they take the values that minimize $h_0(p,q)$. Analogously, the h_1 ***difference map*** *[31] is defined as follows:*

$$\begin{aligned} h_1 : \ \mathbb{P}_\Omega \times \mathbb{P}_\Omega &\to \{0, 1, \ldots, |\Omega|\} \\ (p,q) &\mapsto |\{1 \leqslant i \leqslant n | p_i \neq q_i\}|. \end{aligned}$$

In the particular case where $d = (d_0, \ldots, d_0, d_1)$, $h_0(p,q)$ establishes whether the components of p and q associated to the non-degenerate energy level d_1 are mapped to the same number by their respective d-permutations and, as such, it is a first quantification of the difference between p and q. $h_1(p,q)$ is an actual quantification of their difference, since it outputs the number of positions in which they differ.

We generalize Proposition 3 in the following theorem, which we prove in Appendix H. At a high level, we show that (b) implies (a) by induction on h_0. Moreover, we show the base case $h_0 = 0$ by induction on h_1.

Theorem 6 (Equivalence TO and WETO resource theories)**.** *If $0 < d \in \mathbb{P}_\Omega$, then the following statements are equivalent:*

(a) The thermal operations resource theory is equal to the weak elementary thermal operations resource theory

$$\mathcal{R}_{TO}(d) = \mathcal{R}_{WETO}(d).$$

(b) $d^{\downarrow} = (d_0, \ldots, d_0, d_1)$.

As a result of Theorem 6, elementary thermal operations with deterministic protocols are weakly universal if and only if $d^{\downarrow} = (d_0, \ldots, d_0, d_1)$. Hence, given the result in Theorem 5, Theorem 6 shows that conditioning our experimental protocols via random variables does not augment the cases where elementary thermal operations are weakly universal.

Remark 3 (Algorithm)**.** *If $p, q \in \mathbb{P}_\Omega$ and $q \preceq_d p$, then Theorem 6 provides an algorithm to achieve q via weak elementary thermal operations on p provided $d^\downarrow = (d_0, \ldots, d_0, d_1)$. Up to permutations in d, the algorithm can be summarized as follows:*

(a) Input $0 < d = (d_0, \ldots, d_0, d_1) \in \mathbb{P}_\Omega$ and $p, q \in \mathbb{P}_\Omega$ such that $q \preceq_d p$.

(b) While $h_0(p, q) > 0$ and following Theorem 6, calculate a recursive sequence of T^d-transforms that sequentially reduce h_0 and update p by applying the sequence to it.

(c) While $h_1(p, q) > 0$ and following Theorem 6, calculate a recursive sequence of T^d-transforms that sequentially reduce h_1 and update p by applying the sequence to it.

(d) Output (in order) the sequence of T^d-transforms generated in (b) and (c). This sequence yields q when applied to p.

Theorem 5 follows as a direct corollary of Theorem 6. As we will see in Section 8, the tight relation between TO and WETO resource theories for $d^\downarrow = (d_0, \ldots, d_0, d_1)$ breaks down at the polytope level. In fact, Theorem 6 cannot be extended to polytopes even for uniform d by Theorem 3. This contrasts with the extension of Theorem 5, which we will prove in Section 8.

8 Universality of elementary thermal operations

As a follow up to the previous section, we turn our attention to thermal polytopes, that is, to the universality of elementary thermal operations. Hence, in this section, we answer (Q2) by considering first non-determinisitic protocols in Section 8.1 and then determinisitc ones in Section 8.2.

8.1 Universality of strong elementary thermal operations

The first question we wish to answer is whether Theorem 5 can be extended to polytopes, that is, we would like to know how does Proposition 2 look when we consider $|\Omega| > 3$. In order to do so, we use the work by Jurkat and Ryser [41] and Hartfiel [39]. We begin recalling an algorithm provided in [41].

Definition 8 (Jurkat-Ryser d-algorithm and d-matrix [41])**.** *If $d \in \mathbb{P}_\Omega$, the Jurkat-Ryser d-algorithm is a procedure to construct a matrix of dimension $|\Omega| \times |\Omega|$ that begins with an empty $|\Omega| \times |\Omega|$ matrix A_1 and a couple of vectors $r^1 = s^1 = d$ and, for each step $m \geqslant 2$, does the following:*

(a) Select a position (i_m, j_m) in A_{m-1} that has not been assigned a value yet. If such a position does not exist, return A_{m-1}.

(b) Define A_m as the matrix equivalent to A_{m-1} with the addition of the (i_m, j_m) entry, which equals $\min(r_i^m, s_j^m)$, and fill the rest of row i (column j) with zeros provided $r_i^m = \min(r_i^m, s_j^m)$ ($s_j^m = \min(r_i^m, s_j^m)$).

(c) *Define*

$$r^{m+1} = (r_1^m, \ldots, r_{i-1}^m, r_i^m - \min(r_i^m, s_j^m), r_{i+1}^m, \ldots, r_{|\Omega|}^m),$$
$$s^{m+1} = (s_1^m, \ldots, s_{j-1}^m, s_j^m - \min(r_i^m, s_j^m), s_{j+1}^m, \ldots, r_{|\Omega|}^m).$$

(d) *Return to step (a).*

We say a matrix A is a Jurkat-Ryser d-matrix if it can be constructed following the Jurkat-Ryser algorithm with $r^1 = s^1 = d$. Moreover, for any $m \geqslant 1$, we call r^m and s^m the m-th Jurkat-Ryser row and column d-vectors, respectively.

In the resource-theoretic context, the Jurkat-Ryser d-algorithm is used to construct the extreme points of the **transportation polytope** [22]. The relevance of Definition 8 for our work here is encapsulated in the following theorem, where we denote by $\mathrm{diag}(x_1, \ldots, x_n)$ a diagonal matrix with entries $x_1, \ldots, x_n$.

Theorem 7 (Extreme points TO polytope [41, Theorem 4.1], [39, Lemma 1.1])**.** *If $d \in \mathbb{P}_\Omega$, then the extreme points of the polytope of thermal operations take the form AD^{-1}, where A is a Jurkat-Ryser d-matrix and $D = diag(d_1, d_2, \ldots, d_{|\Omega|})$.*

We prove Theorem 7 in Appendix I. Note that Theorem 7 proves that the TO polytope is in fact a polytope in the sense of [40], since it is clearly convex. In the resource-theoretic context, in analogy to Theorem 7, the Jurkat-Ryser d-algorithm is used to define **β-permutations** [27].

If A is a Jurkat-Ryser d-matrix such that $M = AD^{-1}$ for some extreme point of the TO polytope, then we say A is **associated** to M. In order to determine the relation between the TO and ETO polytopes, a couple more definitions will prove to be useful. The aim of these definitions is to keep track of the previous choices in the Jurkat-Ryser d-algorithm, which will allow us to determine the future entries of a Jurkat-Ryser d-matrix. Hence, we begin by defining the **history** of a Jurkat-Ryser d-matrix.

Definition 9 (History)**.** *If $d \in \mathbb{P}_\Omega$ and A is a Jurkat-Ryser d-matrix, we call a sequence of ordered pairs $((i_{k_l}^A, j_{k_l}^A))_{l=1}^{l_0}$ through which A was constructed the **history** of A and denote it by $H(A)$. Lastly, if $1 \leqslant m \leqslant l_0$, we call the subsequence $((i_{k_l}^A, j_{k_l}^A))_{l=1}^{m} \subseteq H(A)$ the history of A until step m and denote it by $H(A, m)$.*

Although there are several *histories* for some Jurkat-Ryser d-matrix A, we fix here one instance of the Jurkat-Ryser d-algorithm generating A and, hence, one history associated to A. We can do so w.l.o.g. since we only introduce the concept of history in order to determine what the structure of the Jurkat-Ryser d-matrices is, and such a structure is independent of the specific history one may associate to these matrices.

By Proposition 1, the equilibrium distributions in which we are interested have a simple structure, $d = (d_0, \ldots, d_0, d_1)$. Hence, the most important property to determine the entries in a Jurkat-Ryser d-matrix is which components in the Jurkat-Ryser d-vectors have been affected by d_1. That is, for which components in these vectors there exists a sequence of comparisons in the Jurkat-Ryser d-algorithm such that all of them involve a value different from d_0. We call this a **connection** to d_1, as we formalize in the following definition.

Definition 10 (Row and column connection to d_1)**.** *If $d = (d_0, \ldots, d_0, d_1) \in \mathbb{P}_\Omega$ and A and r^m are the Jurkat-Ryser d-matrix and m-th row d-vector, respectively, then we say r^m_a is **row-connected** to d_1 if either $m = 0$ and $a = |\Omega|$ or $m > 0$, $r^m_a > 0$ and there exists a subsequence of the history of A until step $m-1$, $((i^A_{k_\ell}, j^A_{k_\ell}))_{\ell=1}^{\ell_0} \subseteq H(A, m-1)$, such that*

$$\begin{aligned} i^A_{k_{\ell_0}} &= a && \text{and } i^A_{k_1} = |\Omega|, \\ j^A_{k_{2\ell}} &= j^A_{k_{2\ell-1}} && \text{and } i^A_{k_{2\ell+1}} = i^A_{k_{2\ell}} \text{ for } 1 \leqslant \ell, \text{ and} \\ a &\neq i^A_t && \text{for } k_{\ell_0} < t \leqslant m-1. \end{aligned}$$

*We call ℓ_0 the **row-connection length** to d_1 and we define the row-connection between s^m_a and d_1, the column-connection to d_1 of both r^m_a and s^m_a, and their respective lengths in the same vein. Furthermore, we say r^m_a or s^m_a is **connected** to d_1 if it is either row-connected or column-connected and refer to its connection length to d_1 in an analogous way. Lastly, we say a component $A_{i,j}$ is **connected** to d_1 if it was generated, for some $m \geqslant 1$, using r^m_i and s^m_j with at least one of them connected to d_1, and we naturally extend this definition to $M = AD^{-1}$.*

Note that, as one can easily check by contradiction, we do not need to add to Definition 10 constraints like

$$\begin{aligned} j_t &\neq j_{k_{2\ell-1}} && \text{for } k_{2\ell-1} < t < k_{2\ell}, \text{ or} \\ i_t &\neq i_{k_{2\ell}} && \text{for } k_{2\ell} < t < k_{2\ell+1}. \end{aligned}$$

We are now ready to characterize the equivalence between the TO and ETO polytopes, which we address in the following theorem.

Theorem 8 (Equivalence TO and ETO polytopes)**.** *If $0 < d \in \mathbb{P}_\Omega$, then the following statements are equivalent:*

(a) The thermal operations polytope is equal to the elementary thermal operations polytope

$$\mathcal{P}_{TO}(d) = \mathcal{P}_{ETO}(d).$$

(b) $d^\downarrow = (d_0, \ldots, d_0, d_1)$.

As a result of Theorem 8, elementary thermal operations with protocols that can be conditioned using random variables are universal if and only if $d^\downarrow = (d_0, \ldots, d_0, d_1)$. We prove Theorem 8 in Appendix J.[10] At a high level, we show that (b) implies (a) by characterizing, for the specific sort of d we consider,

[10] Note that Theorem 8 can be used to obtain Theorem 5 directly, although it requires using the tools developed in [39, 41].

the extreme points of $\mathcal{P}_{\text{TO}}(d)$ (relaying on the history and the connection to d_1, we first characterize their entries, then the rows and columns they may have and, lastly, the families of rows and columns that may conform them), and by constructing a decomposition of each of these in terms of d-swaps.

Remark 4 (Algorithm)**.** *Given some thermal operation $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$, then Theorem 8 together with linear programming (see the comment below Theorem 5) gives us an algorithm to realize M via strong elementary thermal operation provided $d^{\downarrow} = (d_0, \ldots, d_0, d_1)$. Up to permutations in d, the algorithm can be summarized as follows:*

(a) Input $0 < d = (d_0, \ldots, d_0, d_1) \in \mathbb{P}_{\Omega}$ and $M \in \mathcal{P}_{ETO}(d)$.

(b) Calculate the decomposition of M in terms of the extreme points of $\mathcal{P}_{ETO}(d)$ using linear programming.

(c) Calculate Q as in the proof of Theorem 8 for each extreme point obtained in (b) and use it, following again Theorem 8, to obtain a sequence of d-swaps that (up to permutations) yield the extreme point.

(d) Find the sequence of swaps that are lacking in (c) in order to achieve each extreme point via d-swaps.

(e) Output a convex combination of products of d-swaps that yield M.

8.2 Universality of weak elementary thermal operations

Although, as we have shown, the relation between the ETO and TO polytopes is analogous to that of their polytopes, the situation is quite different when WETO enters the picture. This was already noticed for the uniform case in Theorem 3, which we generalize in the following theorem.

Theorem 9 (Equivalence TO and WETO polytopes)**.** *If $0 < d \in \mathbb{P}_{\Omega}$, then the following statements are equivalent:*

(a) The thermal operations polytope is equal to the weak elementary thermal operations polytope

$$\mathcal{P}_{TO}(d) = \mathcal{P}_{WETO}(d).$$

(b) $|\Omega| = 2$.

As a result of Theorem 8, elementary thermal operations with deterministic protocols are universal if and only if $|\Omega| = 2$. Hence, by Theorem 8, conditioning the protocols via random variables augments the instances where elementary thermal operations are universal. We prove Theorem 9 in Appendix K.

9 Advantage of strong elementary thermal operations over their weak counterpart

In this section, we make substantial progress regarding (Q3) and weak universality in Section 9.1 and fully answer it for the non-weak case in Section 9.2.

9.1 Weak universality advantage

As a result of Theorems 5 and 6, we have the following relation between strong and weak elementary resource theories.

Corollary 1. *If $0 < d \in \mathbb{P}_\Omega$ and $d^\downarrow = (d_0, \ldots, d_0, d_1)$, then the resource theories of strong and weak elementary thermal operations are equal*

$$\mathcal{R}_{ETO}(d) = \mathcal{R}_{WETO}(d).$$

To deal with the relation between these two theories more in general, and noting that it includes the hypothesis in Corollary 1, we recall the definition of quasi-uniform distribution (see Definition 6). The relevance of Definition 6 lies in the fact that, as we show in the following theorem, ETO and WETO resource theories do not coincide whenever d is not quasi-uniform.

Theorem 10 (Difference ETO and WETO resource theories)**.** *If $0 < d \in \mathbb{P}_\Omega$, then the elementary thermal operations resource theory is equal to the weak elementary thermal operations resource theory*

$$\mathcal{R}_{ETO}(d) = \mathcal{R}_{WETO}(d)$$

only if d is quasi-uniform.

As a result of Theorem 10, conditioning our experimental protocols on the realization of some random variable results in the elementary thermal operations being closer to weak universality provided the system has at least three different energy levels. We prove Theorem 10 in Appendix L. At a high level, we first assume $|\Omega| = 3$ and we construct, for any $0 < d \in \mathbb{P}_\Omega$ with non-repeating components, a pair of distributions $p, q \in \mathbb{P}_\Omega$ such that q can be achieved from p via strong elementary thermal operations but not via weak elementary thermal operations. We conclude by showing that we can extend this construction to an arbitrary Ω by appending zeros to both p and q.

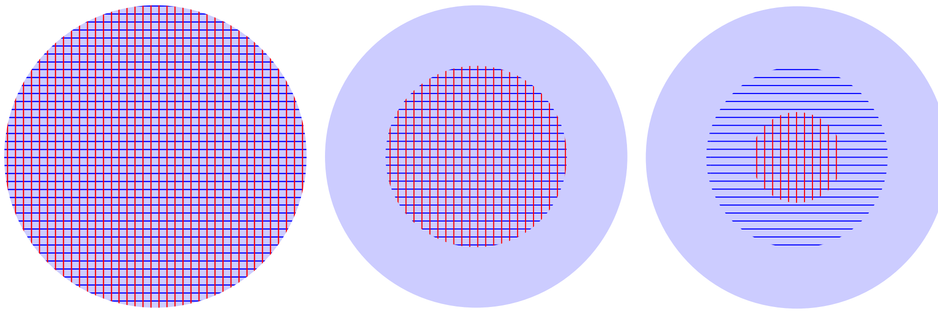
Since it seems we cannot extend the discrepancy between WETO and ETO resource theories in Theorem 10 beyond quasi-uniform distributions, we consider the low-dimensional case $|\Omega| = 3$ in the following proposition. As it turns out, these distributions characterize the equivalence between WETO and ETO resource theories whenever $|\Omega| = 3$.

Proposition 4 (Equivalence ETO and WETO resource theories for $|\Omega| = 3$)**.** *If $0 < d \in \mathbb{P}_\Omega$ and $|\Omega| = 3$, then the following statements are equivalent:*

(a) The strong elementary thermal operations resource theory is equal to the weak elementary thermal operations resource theory

$$\mathcal{R}_{ETO}(d) = \mathcal{R}_{WETO}(d).$$

(b) d is quasi-uniform.



(a) $d^{\downarrow} = (d_0, d_0, d_1)$ with $d_0 \geqslant d_1$.

(b) $d^{\downarrow} = (d_0, d_1, d_1)$ with $d_0 > d_1$.

(c) d is not a quasi-uniform distribution.

Figure 2: Venn diagram representation of the relation among the thermal operations resource theories we have considered in this work depending on $0 < d \in \mathbb{P}_\Omega$ provided $|\Omega| = 3$. In particular, the thermal polytope is filled in light blue and the strong (weak) elementary thermal operations has horizontal (vertical) blue (red) lines. If $|\Omega| > 3$, and we substitute the condition in (a) for $d^{\downarrow} = (d_0, \ldots, d_0, d_1)$ with $d_0 \geqslant d_1$, we replace the one in (b) for d is quasi-uniform and $d^{\downarrow} \neq (d_0, \ldots, d_0, d_1)$ with $d_0 \geqslant d_1$, and we leave condition (c) unaltered, then the diagram looks the same except we do not know whether weak elementary thermal operations are equivalent to strong elementary thermal operations in (b). This figure summarizes Propositions 10 and 4, and Theorems 5 and 6.

As a result of Proposition 4, when $|\Omega| = 3$, conditioning our experimental protocols on the realization of some random variable results in the elementary thermal operations being closer to weak universality provided the system has exactly three different energy levels. We prove Proposition 4 in Appendix M. As a consequence of the results in this section, we pose the following conjecture regarding the general equivalence between ETO and WETO resource theories.

Conjecture 1 (Equivalence ETO and WETO resource theories)**.** *If $0 < d \in \mathbb{P}_\Omega$, then the following statements are equivalent:*

(a) The strong elementary thermal operations resource theory is equal to the weak elementary thermal operations resource theory

$$\mathcal{R}_{ETO}(d) = \mathcal{R}_{WETO}(d).$$

(b) d is quasi-uniform.

9.2 Universality advantage

To conclude this section, we would like to determine the relation between the ETO and WETO polytopes. As we state in the following corollary, we know partly how they are related as a consequence of Theorems 8 and 9.

Corollary 2. *If* $0 < d \in \mathbb{P}_\Omega$ *with* $d^\downarrow = (d_0, \ldots, d_0, d_1)$*, then the following statements are equivalent:*

(a) The (strong) elementary thermal operations polytope is equal to the weak elementary thermal operations polytope

$$\mathcal{P}_{ETO}(d) = \mathcal{P}_{WETO}(d).$$

(b) $|\Omega| = 2$*.*

In the following theorem, we characterize the equivalence between the ETO and WETO polytopes, improving thus on Corollary 2.

Theorem 11 (Equivalence ETO and WETO polytopes)**.** *If* $0 < d \in \mathbb{P}_\Omega$*, then the following statements are equivalent:*

(a) The (strong) elementary thermal operations polytope is equal to the weak elementary thermal operations polytope

$$\mathcal{P}_{ETO}(d) = \mathcal{P}_{WETO}(d).$$

(b) $|\Omega| = 2$*.*

As a result of Theorem 11, conditioning our experimental protocols on the realization of some random variable results in the elementary thermal operations being closer to weak universality provided $|\Omega| \geqslant 3$. We prove Theorem 11 in Appendix N.[11]

10 Convexity of weak elementary thermal operations

To conclude, we use the results in the previous section, in particular Propositions 10 and 4 and Theorem 11, to address the convexity of weak elementary thermal operations in the following corollary.

Corollary 3 (WETO resource theory and polytope convexity)**.** *If* $0 < d \in \mathbb{P}_\Omega$*, then the following statements hold:*

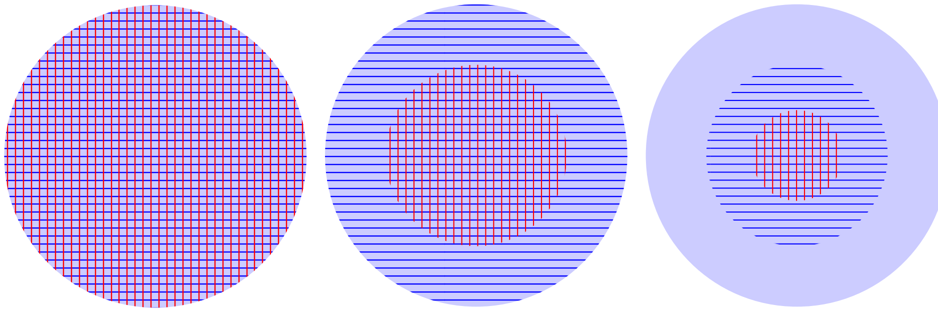
(a) $\mathcal{C}_d^{WETO}(p)$ *is convex for all* $p \in \mathbb{P}_\Omega$ *only if* d *is quasi-uniform.*

(b) If $|\Omega| = 3$*, then* $\mathcal{C}_d^{WETO}(p)$ *is convex for all* $p \in \mathbb{P}_\Omega$ *if and only if* d *is quasi-uniform.*

(c) The polytope of weak elementary thermal operations is convex if and only if $|\Omega| = 2$*.*

We prove Corollary 3 in Appendix O.

[11] Although some instances of the proof of Theorem 11 could be simplified via Theorem 10, we take a simpler approach similar to the one we used when proving Theorem 9.



(a) $|\Omega| = 2$. (b) $d^{\downarrow} = (d_0, \ldots, d_0, d_1)$. (c) $d^{\downarrow} \neq (d_0, \ldots, d_0, d_1)$.

Figure 3: Venn diagram representation of the relation among the thermal polytopes we have considered in this work depending on $0 < d \in \mathbb{P}_{\Omega}$. In particular, the thermal polytope is filled in light blue and the strong (weak) elementary thermal polytope has horizontal (vertical) blue (red) lines. This figure summarizes Theorems 8, 9 and 11.

11 Conclusion

The answers to our main questions are the following:

(Q1) When are elementary thermal operations **weakly** universal?

Both strong and weak elementary thermal operations are weakly universal if and only if $d^{\downarrow} = (d_0, \ldots, d_0, d_1)$. Interestingly, the addition of random variables to our experimental protocols does not offer any advantage regarding the instances where elementary thermal operations are weakly universal.

(Q2) When are elementary thermal operations universal?

Strong elementary thermal operations are universal if and only if $d^{\downarrow} = (d_0, \ldots, d_0, d_1)$. However, weak elementary thermal operations are universal if and only if $|\Omega| = 2$. This illustrates the advantage of adding random variables to our experimental protocols, in contrast with the situation when dealing with weak universality.

(Q3) When do the incorporation of random variables to the elementary thermal operations offer an advantage regarding weak or non-weak universality?

Our answers to (Q1) and (Q2) show that the incorporation of random variables is not advantageous regarding the instances where elementary thermal operations are weakly universal, although it does augment the instances where they are universal. Moreover, even if they are not universal, protocols with random variables can realize more thermal operations than those without them in most cases. In particular, they cover more thermal operations if and only if $|\Omega| \geqslant 3$. Lastly, even if they are not weakly universal, protocols with random variables can reproduce more input-output pairs that are connected via thermal operations provided there are at least three different energy levels. In fact, these are the only situations where they offer such advantage provided $|\Omega| \leqslant 3$. However, it is still unknown whether there are instances where the advantage still holds with $|\Omega| > 3$ and only two different energy levels.

To conclude, let us point out a few directions for future research. The most immediate question is whether Conjecture 1 holds. This would conclude the classification of the thermal resource theories we have considered here and, from the technical side, potentially provide an easier proof of the classical result by Hardy et al. in Theorem 1.

The study of thermal operations (1) for general ρ (i.e. not only quasi-classical states) continues to constitute a major challenge in the field, with only a few known results [20]. Closer to our approach, the study of the evolution of coherence under elementary thermal operations also constitutes an area where little is known [2]. Moreover, the extreme points of the ETO polytope and resource theory are still not well-understood, with only some (potentially bad) upper bounds being known [2, 48]. Algorithmic cooling [2, 27] may benefit from our work since the decomposition of TOs into ETOs is key for its experimental realizations. An instance of this is the TOs-relying algorithm in [27, Theorem 1] and its subsequent decomposition into ETOs [27, Section 2.5].

We have characterized the main case of interest where we only allow two levels to act non-trivially. An obvious open question is how this changes whenever we progressively increase the number of levels where we are allowed to simultaneously act non-trivially. A first result showing that one cannot achieve all thermal operations on a system with $|\Omega|$ levels using only operations that act non-trivially on $|\Omega| - 1$ of them was reported in [49, Section III], where a setup analogous to that in Proposition 1 (a) (which we inherit from [2, Corollary 5]) was used. The experimental relevance of doing so is, however, still unclear, since the experimental realization of such models becomes harder as the number of levels where non-trivial action is permitted increases.

An application of our work here could be the establishment of the function characterization of $\preceq_d$ for $d^{\downarrow} = (d_0, \ldots, d_0, d_1)$, that is, the characterization of all the functions that are allowed to be involved in a second law for $\preceq_d$. The key technical tool we have developed that could contribute in doing so is the so-called **path** result in Theorem 6 (see [16, p. 586 and p. 81] and also [17, p. 45]). This becomes more interesting when put together with the question in the previous paragraph, since the establishment of results analogous to the path one but involving larger proper subsets of Ω may lead us to the characterization of any function involved in a second law for arbitrary d. Lastly, the approach in [50, p. 7] may also be useful to study the second law for $\preceq_d$ in general.

Acknowledgments This research is part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus.

References

[1] Michał Horodecki and Jonathan Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature communications*, 4(1):1–6, 2013.

[2] Matteo Lostaglio, Álvaro M. Alhambra, and Christopher Perry. Elementary thermal operations. *Quantum*, 2:52, 2018.

[3] Fernando Brandao, Michał Horodecki, Jonathan Oppenheim, Joseph M Renes, and Robert W Spekkens. Resource theory of quantum states out of thermal equilibrium. *Physical review letters*, 111(25):250404, 2013.

[4] Matteo Lostaglio. An introductory review of the resource theory approach to thermodynamics. *Reports on Progress in Physics*, 82(11):114001, 2019.

[5] Wiesław Pusz and Stanisław L. Woronowicz. Passive states and KMS states for general quantum systems. *Communications in Mathematical Physics*, 58:273–290, 1978.

[6] Andrew Lenard. Thermodynamical proof of the Gibbs formula for elementary quantum systems. *Journal of Statistical Physics*, 19:575–586, 1978.

[7] Ernst Ruch. The diagram lattice as structural principle A. New aspects for representations and group algebra of the symmetric group B. Definition of classification character, mixing character, statistical order, statistical disorder; a general principle for the time evolution of irreversible processes. *Theoretica Chimica Acta*, 38(3):167–183, 1975.

[8] Ernst Ruch and Alden Mead. The principle of increasing mixing character and some of its consequences. *Theoretica chimica acta*, 41(2):95–117, 1976.

[9] Alden Mead. Mixing character and its application to irreversible processes in macroscopic systems. *The Journal of Chemical Physics*, 66(2):459–467, 1977.

[10] Ernst Ruch, Rudolf Schranner, and Thomas H. Seligman. The mixing distance. *The Journal of Chemical Physics*, 69(1):386–392, 1978.

[11] Peter M. Alberti and Armin Uhlmann. *Dissipative motion in state spaces*, volume 33. Teubner, 1981.

[12] Peter M. Alberti and Armin Uhlmann. *Stochasticity and partial order*. Deutscher Verlag der Wissenschaften Berlin, 1982.

[13] Christian Zylka. A note on the attainability of states by equalizing processes. *Theoretica chimica acta*, 68(5):363–377, 1985.

[14] Barry C. Arnold. Majorization: Here, there and everywhere. *Statistical Science*, 2007.

[15] Barry C. Arnold and José María Sarabia. *Majorization and the Lorenz order with applications in applied mathematics and economics*, volume 7. Springer, 2018.

[16] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer: New York, NY, USA, 2 edition, 2011.

[17] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

[18] Takahiro Sagawa. *Entropy, Divergence, and Majorization in Classical and Quantum Thermodynamics*, volume 16. Springer Nature, 2022.

[19] Fernando Brandao, Michał Horodecki, Nelly Ng, Jonathan Oppenheim, and Stephanie Wehner. The second laws of quantum thermodynamics. *Proceedings of the National Academy of Sciences*, 112(11):3275–3279, 2015.

[20] Gilad Gour, David Jennings, Francesco Buscemi, Runyao Duan, and Iman Marvian. Quantum majorization and a complete set of entropic conditions for quantum thermodynamics. *Nature communications*, 9(1):1–9, 2018.

[21] Álvaro M. Alhambra, Jonathan Oppenheim, and Christopher Perry. Fluctuating states: What is the probability of a thermodynamical transition? *Physical Review X*, 6(4):041016, 2016.

[22] Paweł Mazurek. Thermal processes and state achievability. *Physical Review A*, 99(4):042110, 2019.

[23] Naoto Shiraishi. Two constructive proofs on d-majorization and thermomajorization. *Journal of Physics A: Mathematical and Theoretical*, 53(42):425301, 2020.

[24] Matteo Lostaglio and Kamil Korzekwa. Continuous thermomajorization and a complete set of laws for Markovian thermal processes. *Physical Review A*, 106(1):012426, 2022.

[25] Jacopo Surace and Matteo Scandi. State retrieval beyond Bayes' retrodiction. *Quantum*, 7:990, 2023.

[26] Jakub Czartowski, A de Oliveira Junior, and Kamil Korzekwa. Thermal recall: Memory-assisted Markovian thermal processes. *PRX Quantum*, 4(4):040304, 2023.

[27] Álvaro M. Alhambra, Matteo Lostaglio, and Christopher Perry. Heatbath algorithmic cooling with optimal thermalization strategies. *Quantum*, 3:188, 2019.

[28] Kamil Korzekwa and Matteo Lostaglio. Optimizing thermalization. *Physical review letters*, 129(4):040602, 2022.

[29] Matteo Lostaglio, Kamil Korzekwa, David Jennings, and Terry Rudolph. Quantum coherence, time-translation symmetry, and thermodynamics. *Physical review X*, 5(2):021001, 2015.

[30] Kamil Korzekwa. *Coherence, thermodynamics and uncertainty relations.* PhD thesis, Imperial College London, 2016.

[31] Godfrey H. Hardy, John E. Littlewood, George Pólya, György Pólya, et al. *Inequalities.* Cambridge university press, 1952.

[32] Michael Reck, Anton Zeilinger, Herbert J Bernstein, and Philip Bertani. Experimental realization of any discrete unitary operator. *Physical review letters*, 73(1):58, 1994.

[33] Bernard Yurke, Samuel L. McCall, and John R. Klauder. SU (2) and SU (1,1) interferometers. *Physical Review A*, 33(6):4033, 1986.

[34] Christopher Perry, Piotr Ćwikliński, Janet Anders, Michał Horodecki, and Jonathan Oppenheim. A sufficient set of experimentally implementable thermal operations for small systems. *Physical Review X*, 8(4):041049, 2018.

[35] Edwin T Jaynes and Frederick W Cummings. Comparison of quantum and semiclassical radiation theories with application to the beam maser. *Proceedings of the IEEE*, 51(1):89–109, 1963.

[36] Märio Ziman, Peter Štelmachovič, and Vladimír Bužek. Description of quantum dynamics of open systems based on collision-like models. *Open systems & information dynamics*, 12:81–91, 2005.

[37] Robert Franklin Muirhead. Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. *Proceedings of the Edinburgh Mathematical Society*, 21:144–162, 1902.

[38] Garrett Birkhoff. Tres observaciones sobre el algebra lineal. *Univ. Nac. Tucuman, Ser. A*, 5:147–154, 1946.

[39] Darald Hartfiel. A study of convex sets of stochastic matrices induced by probability vectors. *Pacific Journal of Mathematics*, 52(2):405–418, 1974.

[40] Arne Brondsted. *An introduction to convex polytopes*, volume 90. Springer Science & Business Media, 2012.

[41] Wolfgang B. Jurkat and Herbert J. Ryser. Term ranks and permanents of nonnegative matrices. *Journal of algebra*, 5(3):342–357, 1967.

[42] Ernst Ruch, Rudolf Schranner, and Thomas H Seligman. Generalization of a theorem by Hardy, Littlewood, and Pólya. *Journal of Mathematical Analysis and Applications*, 76(1):222–229, 1980.

[43] Marvin Marcus, Kent Kidman, and Markus Sandy. Products of elementary doubly stochastic matrices. *Linear and Multilinear Algebra*, 15(3-4):331–340, 1984.

[44] John Von Neumann. A certain zero-sum two-person game equivalent to the optimal assignment problem. *Contributions to the Theory of Games*, 2(0):5–12, 1953.

[45] David A. Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

[46] Arthur F. Veinott Jr. Least d-majorized network flows with inventory and statistical applications. *Management Science*, 17(9):547–567, 1971.

[47] Stephen P. Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[48] Jeongrak Son and Nelly HY Ng. Catalysis in action via elementary thermal operations. *arXiv preprint arXiv:2209.15213*, 2022.

[49] Paweł Mazurek and Michał Horodecki. Decomposability and convex structure of thermal processes. *New Journal of Physics*, 20(5):053040, 2018.

[50] Alexssandre de Oliveira Junior, Jakub Czartowski, Karol Życzkowski, and Kamil Korzekwa. Geometric structure of thermal cones. *Physical Review E*, 106(6):064109, 2022.

[51] László Lovász. Random walks on graphs. *Combinatorics, Paul Erdös is eighty*, 2(1-46):4, 1993.

[52] Rajeev Motwani and Prabhakar Raghavan. *Randomized algorithms*. Cambridge university press, 1995.

[53] Feng Xia, Jiaying Liu, Hansong Nie, Yonghao Fu, Liangtian Wan, and Xiangjie Kong. Random walks: A review of algorithms and applications. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 4(2):95–107, 2019.

[54] Mario Szegedy. Quantum speed-up of Markov chain based algorithms. In *45th Annual IEEE symposium on foundations of computer science*, pages 32–41. IEEE, 2004.

[55] Andrew M Childs. Universal computation by quantum walk. *Physical review letters*, 102(18):180501, 2009.

[56] Edward Farhi and Sam Gutmann. Quantum computation and decision trees. *Physical Review A*, 58(2):915, 1998.

[57] Pedro Hack, Daniel A Braun, and Sebastian Gottwald. Majorization requires infinitely many second laws. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 480(2296):20230464, 2024.

A Proof of Theorem 4

We assume for simplicity that $d = d^{\downarrow}$ and, hence, $M = M^{\downarrow}$ throughout the proof.

We begin by showing that (a) implies (b). In order to do so, we start by simply fixing a decomposition of $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ is terms of d-swaps

$$M = \lambda \mathbb{I} + \sum_{s=1}^{s_0} \lambda_s P^d(i_s, j_s),$$

where $\lambda + \sum_{s=1}^{s_0} \lambda_s = 1$, and, for $1 \leqslant s \leqslant s_0$, $\lambda, \lambda_s \geqslant 0$ and $1 \leqslant i_s < j_s \leqslant |\Omega|$. In this scenario, we immediately have that M is stochastic and, moreover, it fulfills detailed balance since $P^d(i_s, j_s)$ does and it is the only d-swap that contributes to M_{i_s,j_s} for all $1 \leqslant s \leqslant s_0$. Furthermore, it is easy to see that, for $1 \leqslant s \leqslant s_0$ and $1 \leqslant i \leqslant |\Omega|$, we have that

$$P^d(i_s, j_s)_{i,i} = \sum_{i<j\leqslant|\Omega|} \left(1 - \frac{d_j}{d_i}\right) P^d(i_s, j_s)_{i,j} + \sum_{\substack{1\leqslant k<j\leqslant|\Omega| \\ k,j\neq i}} P^d(i_s, j_s)_{k,j}. \tag{10}$$

To conclude, it is easy to show that (7) holds using (10). We have

$$\begin{aligned}
(M)_{i,i} &= \lambda + \sum_{s=1}^{s_0} \lambda_s \left(\sum_{i<j\leqslant|\Omega|} \left(1 - \frac{d_j}{d_i}\right) P^d(i_s, j_s)_{i,j} + \sum_{\substack{1\leqslant k<j\leqslant|\Omega| \\ k,j\neq i}} P^d(i_s, j_s)_{k,j} \right) \\
&= \lambda + \sum_{i<j\leqslant|\Omega|} \left(1 - \frac{d_j}{d_i}\right) \sum_{s=1}^{s_0} \lambda_s P^d(i_s, j_s)_{i,j} + \sum_{\substack{1\leqslant k<j\leqslant|\Omega| \\ k,j\neq i}} \sum_{s=1}^{s_0} \lambda_s P^d(i_s, j_s)_{k,j} \\
&= \lambda + \sum_{i<j\leqslant|\Omega|} \left(1 - \frac{d_j}{d_i}\right) M_{i,j} + \sum_{\substack{1\leqslant k<j\leqslant|\Omega| \\ k,j\neq i}} M_{k,j},
\end{aligned}$$

for $1 \leqslant i \leqslant |\Omega|$.

We conclude by showing that (b) implies (a). In order to do so, it suffices to notice that any stochastic matrix $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ fulfilling detailed balance and (7) can be decomposed in the following way:

$$M = \lambda \mathbb{I} + \sum_{1\leqslant k<j\leqslant|\Omega|} M_{k,j} P^d(k, j). \tag{11}$$

This constitutes a convex combination of d-swaps since

$$
\begin{aligned}
\lambda + \sum_{1 \leqslant k < j \leqslant |\Omega|} M_{k,j} &= \lambda + \sum_{2 \leqslant k < j \leqslant |\Omega|} M_{k,j} + \sum_{1 < j \leqslant |\Omega|} M_{1,j} \\
&= M_{1,1} - \sum_{1 < j \leqslant |\Omega|} \left(1 - \frac{d_j}{d_1}\right) M_{1,j} + \sum_{1 < j \leqslant |\Omega|} M_{1,j} \\
&= M_{1,1} + \sum_{1 < j \leqslant |\Omega|} \frac{d_j}{d_1} M_{1,j} \\
&= M_{1,1} + \sum_{1 < j \leqslant |\Omega|} M_{j,1} \\
&= 1,
\end{aligned}
$$

where we applied (7) in the second equality, detailed balance in the fourth, and the fact M is stochastic in the last.

To conclude, we verify that (11) holds. This is the case since the following relations hold for $1 \leqslant k < j \leqslant |\Omega|$: (A) $(M_{k,j} P^d(k,j))_{k,j} = M_{k,j}$ by definition, (B) $(M_{k,j} P^d(k,j))_{j,k} = M_{j,k}$ by detailed balance, and (C)

$$
(M_{k,j} P^d(k,j))_{\ell,\ell} = \begin{cases} M_{k,j} & \text{if } \ell \neq k, j, \\ \left(1 - \frac{d_j}{d_k}\right) M_{k,j} & \text{if } \ell = k, \\ 0 & \text{if } \ell = j, \end{cases}
$$

by definition. Hence, applying (7), we have that (11) holds.

B Random walks on complete graphs and the length one ETO polytope

In the simpler scenario where we only allow sequences of d-swaps with length one, the resource theory of strong elementary thermal operations can be interpreted in terms of random walks on complete graphs. In this section, we introduce a new definition of random walk on a complete graph in Section B.1, and use Theorem 4 to relate our definition to the one usually used in the literature in Section B.2.

B.1 ETO random walks on complete graphs

Random walks on graphs have been extensively studied [45, 51–53] and have become of increasing interest given their close connection to quantum walks [54, 55]. Instead of focusing on the general case, we will only consider complete graphs. Take, hence, a complete undirected graph $G = (V_G, E_G)$ without loops, where V_G are the vertices and E_G the edges.[12] Furthermore, assume we have

[12] A graph is **complete** provided there is an edge between any pair of distinct vertices. Moreover, a loop is an edge from a vertex to itself. Lastly, a graph is **undirected** provided its edges have no direction. These definition are taken from [45].

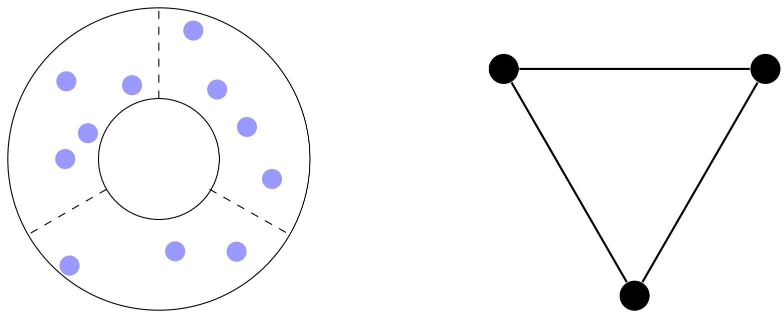


Figure 4: Gas in a toric box (left) and its associated complete graph (right).

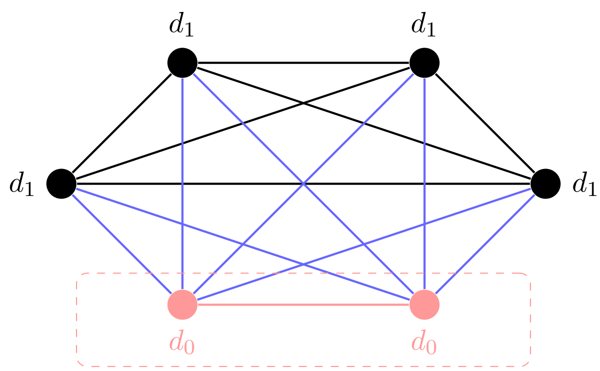


Figure 5: Complete d-graph G with $|V_G| = 6$ and $d^{\downarrow} = (d_0, d_0, d_1, \ldots, d_1) \in \mathbb{P}_V$ with $d_1 < d_0$. The black and red edges represent symmetric channels, and the blue edges asymmetric ones.

some substance that is distributed among the vertices V_G and think of the edges E_G as **channels** through which the substance can be redistributed among the vertices. As an example, when $|V_G| = 3$, we can consider a gas in a **toric** box (see Figure 4). Lastly, we assume that the substance may prefer being in certain vertices over others and model this by incorporating some reference distribution over the vertices $d \in \mathbb{P}_{V_G}$ that may give preference to some subsets of V_G over others. We can think of these preferences among vertices as signaling the existence of some **marked** vertices, as one usually encounters in the study of decision trees [56]. The existence of preferences introduces asymmetries in some channels, as one can see in Figure 5. In the context of the gas in a box, we can think of the introduction of preferences as surrounding it with a heat bath, with the preferences following a Gibbs distribution.

Our aim, and the reason why walks on graphs were introduced [45], is to associate to G a transition matrix M such that, if we assume time to be discrete and consider some initial distribution of the substance over the edges $p_0 \in \mathbb{P}_{V_G}$, then $p_1 = Mp_0 \in \mathbb{P}_{V_G}$ is a distribution that could have been obtained by allowing

Figure 6: Graph associated with the classical thermodynamic example of a gas in a box.

the substance to redistribute itself along the edges of G after one time step. With this in mind, we define the following random walk on complete graphs.

Definition 11 (ETO random walk on a complete graph)**.** *If $G = (V_G, E_G)$ is a complete graph with some preferences among its vertices given by a distribution $d \in \mathbb{P}_{V_G}$, then an* ***ETO random walk*** *on G is a matrix $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ that can be decomposed as a convex combination of d-swaps.*

Note that Definition 11 coincides with the length one ETO polytope and, moreover, instead of assuming that the redistribution of substance takes place among all edges, it allows any distribution over the different edges in E_G (including those in which some of them are not used). Furthermore, the identity can also be given a non-zero weight.

Before we continue, let us make a remark regarding the classical example of a gas in a box. In particular, we consider d to be the uniform distribution and $|V_G| \geqslant 3$, as illustrated in Figure 6 via its associated graph. Although this is not covered by ETO random walks (since the associated graph is not complete), one could treat this (or any non-complete graph) in an analogous way by taking into account the topology of the specific problem. (The only drawback being the potential lack of symmetry compared to the case of complete graphs.) It should be noted that one cannot take M to be any d-stochastic matrix. In fact, in this scenario, permuting the gas concentration between non-adjacent compartments while leaving the rest unchanged would be allowed by doubly stochastic matrices. Such considerations led to the introduction of the **molecular diffusion ordering** in [57], which corresponds to the transitions associated to non-homogeneous Markov chains whose constituent parts are precisely ETO random walks.

B.2 Random walks on complete graphs

Although **weighted random walks** have also been considered in the literature [45, Chapter 9], we relate here Definition 11 with the usual definition on random walk only in the case when the reference distribution $d \in \mathbb{P}_{V_G}$ is uniform. In this scenario, M_0 is the **simple random walk** on (a complete graph) G provided we have, for all $1 \leqslant i, j \leqslant |V_G|$, that

$$(M_0)_{i,j} = \begin{cases} \frac{1}{|V_G|-1}, & \text{if } i \neq j, \\ 0, & \text{if } i = j. \end{cases} \tag{12}$$

Moreover, M_1 is the **lazy random walk** on G provided

$$M_1 = \frac{1}{2}\left(\mathbb{I} + M_0\right) \tag{13}$$

for some simple random walk M_0.

Hence, the simple random walk assumes that the substance coming from one vertex is redistributed equally among all the vertices it is connected to (in our case, all of them) and the lazy random walk that half of it stays in the original vertex and the rest is redistributed equally among all the vertices it is connected to.

We can use Theorem 4 to directly establish the relation between simple, lazy and ETO random walks on graphs, as we state in the following corollary.

Corollary 4 (Relation between random walks on complete graphs)**.** *If $G = (V_G, E_G)$ is a complete undirected graph without loops, $d \in \mathbb{P}_{V_G}$ is the uniform distribution and $M \in \mathcal{M}_{|V_G|,|V_G|}(\mathbb{R})$, then the following statements hold:*

(a) If M is the simple random walk on G, then M is an ETO random walk on G if and only if $|V_G| = 2$.

(b) If M is the lazy random walk on G, then M is an ETO random walk on G if and only if $|V_G| \leqslant 4$.

Proof. Both statements follow directly from (8). The first follows since we have

$$M_{i,i} = 0 < \frac{1}{2}|V_G| - 1 = \sum_{\substack{1 \leqslant k < j \leqslant |V_G| \\ k,j \neq i}} M_{k,j} \iff 2 < |V_G|$$

for $1 \leqslant i \leqslant |\Omega|$. The second follows since we have

$$M_{i,i} = \frac{1}{2} < \frac{1}{2}\left(\frac{1}{2}|V_G| - 1\right) = \sum_{\substack{1 \leqslant k < j \leqslant |V_G| \\ k,j \neq i}} M_{k,j} \iff 4 < |V_G|$$

for $1 \leqslant i \leqslant |\Omega|$. □

As a last remark, note that weighted random walks consider asymmetries between the different edges in the graph (with the motivation coming from electric networks), while ETO random walks make distinctions at the vertex level.

C Proof of Lemma 1

Take $Q \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ the permutation matrix such that $d^{\downarrow} = Qd$ and note that it suffices to show that (b) implies (a) since the converse follows analogously. Moreover, for simplicity, we only consider two cases:

(A) $A(d^{\downarrow}) = \mathcal{P}_{\mathrm{TO}}(d^{\downarrow})$, $B(d^{\downarrow}) = \mathcal{P}_{\mathrm{ETO}}(d^{\downarrow})$ and we take the corresponding $A(d)$ and $B(d)$. If $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ is a d-stochastic matrix, then $M^{\downarrow}$ is a

$d^{\downarrow}$-stochastic matrix. Hence, by assumption,

$$M^{\downarrow} = QMQ^T = \sum_{k=0}^{k_0} \lambda_k \prod_{\ell=0}^{\ell_0} P^{d^{\downarrow}}(i_{k,\ell}, j_{k,\ell}), \text{ and}$$
$$M = \sum_{k=0}^{k_0} \lambda_k \prod_{\ell=0}^{\ell_0} Q^T P^{d^{\downarrow}}(i_{k,\ell}, j_{k,\ell}) Q,$$

where $\sum_{k=0}^{k_0} \lambda_k = 1$, and $\lambda_k > 0$ and $1 \leqslant i_{k,\ell} < j_{j,\ell} \leqslant |\Omega|$ for $0 \leqslant k \leqslant k_0$ and $0 \leqslant \ell \leqslant \ell_0$.

(B) $A(d^{\downarrow}) = \mathcal{R}_{\mathrm{TO}}(d^{\downarrow})$, $B(d^{\downarrow}) = \mathcal{R}_{\mathrm{ETO}}(d^{\downarrow})$ and we take the corresponding $A(d)$ and $B(d)$. Given a pair $p, q \in \mathbb{P}_\Omega$ such that $q = Mp$ for some d-stochastic matrix $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$, we have that $Qq = QMQ^T Qp$. By assumption, since QMQ^T is a $d^{\downarrow}$-stochastic matrix, we have

$$Qq = \left(\sum_{k=0}^{k_0} \lambda_k \prod_{\ell=0}^{\ell_0} P^{d^{\downarrow}}(i_{k,\ell}, j_{k,\ell}) \right) Qp, \text{ and}$$
$$q = \left(\sum_{k=0}^{k_0} \lambda_k \prod_{\ell=0}^{\ell_0} Q^T P^{d^{\downarrow}}(i_{k,\ell}, j_{k,\ell}) Q \right) p,$$

where $\sum_{k=0}^{k_0} \lambda_k = 1$, and $\lambda_k > 0$ and $1 \leqslant i_{k,\ell} < j_{j,\ell} \leqslant |\Omega|$ for $0 \leqslant k \leqslant k_0$ and $0 \leqslant \ell \leqslant \ell_0$.

Since the other cases follow analogously, this concludes the proof.

D Proof of Proposition 1

By Lemma 1, it suffices to assume that $d = d^{\downarrow}$ throughout the proof.

We show the result by contrapositive, that is, we assume that there is no pair $d_0, d_1 \in \mathbb{R}$ with $0 < d_1 \leqslant d_0$ such that $d = (d_0, \dots, d_0, d_1)$. We split the proof in two cases where, for each of them, we construct a pair of distributions $p, q \in \mathbb{P}_\Omega$ such that q can be attained from p by means of thermal operations but not by elementary thermal operations. We consider d_α the largest component in d and notice that, by assumption, there exist two entries of d, which we name for simplicity d_β and d_γ, such that $d_\gamma \leqslant d_\beta < d_\alpha$.[13] Consider, hence, the following two cases:

(A) $d_\alpha \geqslant d_\beta + d_\gamma$. In this case, we can follow the idea in [2, Corollary 5]. In particular, we can take $q = (1, 0, \dots, 0)$ and $p = (0, a, b, 0, \dots, 0)$ with $d_\gamma / d_\alpha \leqslant b \leqslant (d_\alpha - d_\beta)/d_\alpha$ (this can be done since $d_\alpha \geqslant d_\beta + d_\gamma$ by assumption) and $a = 1 - b$. By Lemma 3, q cannot be achieved from p via

[13]For simplicity, and w.l.o.g., we will assume that d_β and d_γ correspond, respectively, to the second and third entries of d. If that were not the case, we could follow the same argument, taking the appropriate components of both p and q.

elementary thermal operations since it has a smaller support. However, it is easy to check that $q \preceq_d p$.

(B) $d_\alpha < d_\beta + d_\gamma$. In this case, we can take $q = (0, a, b, 0, \dots, 0)$ with $(d_\alpha - d_\beta)/d_\alpha \leqslant b \leqslant d_\gamma/d_\alpha$ (this can be done since $d_\alpha < d_\beta + d_\gamma$ by assumption) and $a = 1 - b$, and $p = (1, 0, \dots, 0)$. q cannot be achieved from p via elementary thermal operations since, given that $d_\alpha > d_i$ for any $i \neq \alpha$ by assumption, the action of any elementary thermal operation on p will result in a non-zero first component. (In case we had multiple entries in d equal to d_α, the action of any elementary thermal operation on p will leave , at least, one non-zero component in the entries associated with those for which d takes the value d_α.) However, it is easy to check that $q \preceq_d p$.

This concludes the proof.

E Proof of Proposition 2

By Lemma 1, it suffices to assume that $d = d^\downarrow$ throughout the proof.

Sufficiency follows from Proposition 1 by contradiction. Assume both polytopes are equivalent for some $d \neq (d_0, d_0, d_1)$. By Proposition 1, there exists a pair $p, q \in \mathbb{P}_\Omega$ such that $q \preceq_d p$ and q is not achievable from p via elementary thermal operations. Since $q \preceq_d p$, there exists some d-stochastic matrix M such that $q = Mp$ and, since the polytopes are equivalent,

$$q = \left(\sum_{k=0}^{k_0} \lambda_k \prod_{\ell=0}^{\ell_0} P^d(i_{k,\ell}, j_{k,\ell}) \right) p,$$

with $\sum_{k=0}^{k_0} \lambda_k = 1$, and $\lambda_k > 0$ and $1 \leqslant i_{k,\ell} < j_{j,\ell} \leqslant |\Omega|$ for $0 \leqslant k \leqslant k_0$ and $0 \leqslant \ell \leqslant \ell_0$. This contradicts Proposition 1.

To prove necessity, since the polytope of elementary thermal operations is always contained inside that of thermal operations by definition, it suffices to note that all the extreme points of the polytope of thermal operations can be obtained as a product of elementary thermal operations. In particular, taking $\gamma = d_1/d_0$, the set of extreme points of the polytope of thermal operations is

$\{\mathbb{I}, A_1, \ldots, A_9\}$, where

$$A_1 = \begin{pmatrix} 1-\gamma & 0 & 1 \\ \gamma & 1-\gamma & 0 \\ 0 & \gamma & 0 \end{pmatrix}, \; A_2 = \begin{pmatrix} 1-\gamma & 0 & 1 \\ 0 & 1 & 0 \\ \gamma & 0 & 0 \end{pmatrix}, \; A_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1-\gamma & 0 & 1 \\ \gamma & 0 & 0 \end{pmatrix},$$

$$A_4 = \begin{pmatrix} 1-\gamma & \gamma & 0 \\ 0 & 1-\gamma & 1 \\ \gamma & 0 & 0 \end{pmatrix}, \; A_5 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-\gamma & 1 \\ 0 & \gamma & 0 \end{pmatrix}, \; A_6 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$A_7 = \begin{pmatrix} \gamma & 1-\gamma & 0 \\ 1-\gamma & 0 & 1 \\ 0 & \gamma & 0 \end{pmatrix}, \; A_8 = \begin{pmatrix} 0 & 1-\gamma & 1 \\ 1 & 0 & 0 \\ 0 & \gamma & 0 \end{pmatrix}, \; A_9 = \begin{pmatrix} 0 & 1-\gamma & 1 \\ 1-\gamma & \gamma & 0 \\ \gamma & 0 & 0 \end{pmatrix}. \tag{14}$$

(This can be calculated using [41] and [39]. The reader interested in how this is done can check Section 8.)

To conclude, we simply notice that, aside from the identity and the elementary thermal operations acting on two levels, we have

$$\begin{aligned} A_1 &= P^d(2,3)P^d(1,3), \\ A_3 &= P^d(2,3)P^d(1,2), \\ A_4 &= P^d(1,3)P^d(2,3), \\ A_7 &= P^d(1,3)P^d(1,2)P^d(1,3), \\ A_8 &= P^d(1,2)P^d(2,3), \\ A_9 &= P^d(2,3)P^d(1,2)P^d(2,3). \end{aligned}$$

This concludes the proof.

F Proof of Theorem 5

By Lemma 1, it suffices to assume that $d = d^{\downarrow}$ throughout the proof. For simplicity, we fix $n = |\Omega|$ and $\gamma = d_1/d_0$ throughout the proof.

Since sufficiency follows from Proposition 1, we only ought to prove necessity. In order to do so, we will show that, for any pair $p, q \in \mathbb{P}_\Omega$, q can be obtained from p by elementary thermal operations provided the same holds for thermal operations. Instead of showing it directly, we can profit from Lemma 4, which provides a finite set of distributions $\{p^\Pi | \Pi \in S_n\}$ such that, if q can be obtained from p by thermal operations, then q can be decomposed as a convex combination of distributions in that set (which we denote by conv),

$$\mathcal{C}_d(p) = \text{conv}\{p^\Pi | \Pi \in S_n\}. \tag{15}$$

Thus, it suffices to show that each distribution in the set is attainable by performing a sequence of elementary thermal operations on p to conclude the proof.

Before we start the argument, note that we can assume w.l.o.g. that $p_i \geqslant p_{i+1}$ for $i = 1, \ldots, n-2$. To conclude the proof, recalling the notation in Definition 2, we take $m = \Pi_p^d(n)$ (in case of ambiguity, take the highest value possible) and we show that, for any permutation $\Pi \in S_n$, p^{Π} is attainable from p by elementary thermal operations. Fix, hence, such a permutation $\Pi_0 \in S_n$, define $s = \Pi_0(n)$, and note that the relation between s and m determines the entries of p^{Π_0}. We distinguish, in particular, three cases:

(A) $m = s$. In this case, we have that

$$y_i^{\Pi_0} = \begin{cases} \sum_{j=1}^{i} p_j & \text{if } 1 \leqslant i < m, \\ p_n + \sum_{j=1}^{i-1} p_j & \text{if } m \leqslant i \leqslant n. \end{cases}$$

By definition of s, we have that $p_n^{\Pi_0} = y_s^{\Pi_0} - y_{s-1}^{\Pi_0} = p_n$. The rest of the components of p^{Π_0} are p_i for some $1 \leqslant i < n$ by definition. The order in which these components are presented is not important, since we can attain any arrangement from another one by elementary thermal operations. Hence, we can assume w.l.o.g. that $p^{\Pi_0} = p$ and, thus, p^{Π_0} is attainable from p via elementary thermal operations.

(B) $m < s$. In this case, we have that

$$y_i^{\Pi_0} = \begin{cases} \sum_{j=1}^{i} p_j & \text{if } 1 \leqslant i < m, \\ p_n + (1-\gamma)p_i + \sum_{j=1}^{i-1} p_j & \text{if } m \leqslant i < s, \\ p_n + \sum_{j=1}^{i-1} p_j & \text{if } s \leqslant i \leqslant n. \end{cases}$$

By definition of s, we have that $p_n^{\Pi_0} = y_s^{\Pi_0} - y_{s-1}^{\Pi_0} = \gamma p_{s-1}$. Following the argument for the case $m = s$, we can assume w.l.o.g. that

$$\begin{aligned} p^{\Pi_0} = (&p_1, \ldots, p_{m-1}, (1-\gamma)p_m + p_n, (1-\gamma)p_{m+1} + \gamma p_m, \ldots, \\ &(1-\gamma)p_{s-1} + \gamma p_{s-2}, p_s, \ldots, p_{n-1}, \gamma p_{s-1}). \end{aligned}$$

We conclude the proof of this case noticing that

$$p^{\Pi_0} = \left(\prod_{k=m}^{s-1} P^d(k,n) \right) p.$$

(C) $s < m$. In this case, we have that

$$y_i^{\Pi_0} = \begin{cases} \sum_{j=1}^{i} p_j & \text{if } 1 \leqslant i < s, \\ \gamma p_i + \sum_{j=1}^{i-1} p_j & \text{if } s \leqslant i < m, \\ p_n + \sum_{j=1}^{i-1} p_j & \text{if } m \leqslant i \leqslant n. \end{cases}$$

By definition of s, we have that $p_n^{\Pi_0} = y_s^{\Pi_0} - y_{s-1}^{\Pi_0} = \gamma p_s$. Following the argument for the case $m = s$, we can assume w.l.o.g. that

$$\begin{aligned} p^{\Pi_0} = (&p_1, \ldots, p_{s-1}, (1-\gamma)p_s + \gamma p_{s+1}, \ldots, (1-\gamma)p_{m-2} + \gamma p_{m-1}, \\ &p_n + (1-\gamma)p_{m-1}, p_m, \ldots, p_{n-1}, \gamma p_s). \end{aligned}$$

We conclude the proof of this case noticing that

$$p^{\Pi_0} = \left(\prod_{k=1}^{m-s} P^d(m-k,n)\right) p.$$

This concludes the proof.

Remark 5. *The decompositions in Theorem 5 coincide with those in Proposition 2 when $|\Omega| = 3$. This is the case since we have that $P^d(1,3)P^d(1,2) = P^d(1,2)P^d(2,3)$ and, hence, we can locate the sequence of swaps at the start.*

G Proof of Proposition 3

By Lemma 1, it suffices to assume that $d = d^{\downarrow}$ throughout the proof.

The fact that (a) implies (b) holds as a direct consequence of Proposition 1 for any finite Ω. We show this by contrapositive: If the implication was false, then we could reach any distribution achievable by TO using finite sequences of T^d-transforms for some $d \neq (d_0, \ldots, d_0, d_1)$. However, this would imply that any process achievable by TO is also achievable by ETO for such d, which contradicts Proposition 1.

We show now that (b) implies (a). We take $\gamma = d_1/d_0$ and $p = (a, b, c)$, noting that we can assume $a \geqslant b$ w.l.o.g. (otherwise, we simply apply $P^d(1,2)$ to p first and then follow the argument below), and $q \in \mathbb{P}_\Omega$ such that $q \preceq_d p$. Since $q \preceq_d p$ and $d = (d_0, d_0, d_1)$ with $0 < d_1 \leqslant d_0$, we know from Proposition 2 that q is contained in the convex hull of $\{A_0 p, \ldots, A_9 p\}$,

$$q \in \mathcal{C}_d(p) = \operatorname{conv}\{A_0 p, \ldots, A_9 p\},$$

where $A_0 = \mathbb{I}$ and the others are defined as in Proposition 2. To conclude the proof, we will give a sequence of T^d-transforms that, when applied to p, yield q. We consider three cases (the cases where some equality holds follow easily from these):

(A) $\gamma a > \gamma b > c$. In this case, $\mathcal{C}_d(p)$ is (roughly) given by Figure 7, with q being achievable by a sequence $T^d(1,2)T^d(2,3)$ if it lies below the dashed line and by $T^d(1,2)T^d(1,3)P^d(2,3)$ if it lies above.

(B) $\gamma a > c > \gamma b$. In this case, $\mathcal{C}_d(p)$ is (roughly) given by Figure 8, with q being achievable by a sequence $T^d(1,2)T^d(1,3)$ if it lies above the dashed line and by $T^d(1,2)T^d(2,3)$ if it lies below.

(C) $c > \gamma a > \gamma b$. In this case, $\mathcal{C}_d(p)$ is (roughly) given by Figure 9, with q being achievable by a sequence $T^d(1,2)T^d(1,3)$ if it lies above the dashed line and by $T^d(1,2)T^d(2,3)P^d(1,3)$ if it lies below.

This concludes the proof.

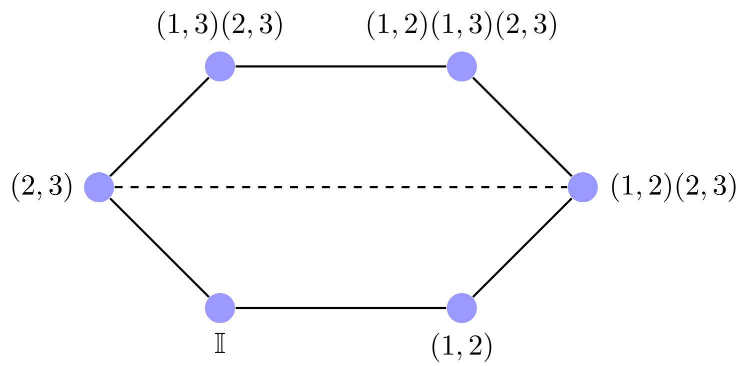


Figure 7: Rough representation of the set $\mathcal{C}_d(p)$ for $d = (d_0, d_0, d_1)$ with $0 < d_1 \leqslant d_0$ and $p = (a, b, c)$ with $\gamma a \geqslant \gamma b \geqslant c$. Note that we label the vertices by the (ordered) sequence of elementary thermal operations that we apply to p to achieve them and that we substitute $P^d(i, j)$ by (i, j).

H Proof of Theorem 6

By Lemma 1, it suffices to assume that $d = d^{\downarrow}$ throughout the proof.

The fact that (a) implies (b) holds analogously to its counterpart in Proposition 3, i.e., as a direct consequence of Proposition 1.

To show that (b) implies (a), we begin taking $p, q \in \mathbb{P}_\Omega$ such that $q \preceq_d p$. We also take $n = |\Omega|$ and $\gamma = d_1/d_0$ for simplicity and note that we can assume w.l.o.g.[14] that

$$p_1 \geqslant p_2 \geqslant \cdots \geqslant p_{n-1} \text{ and } q_1 \geqslant q_2 \geqslant \cdots \geqslant q_{n-1}. \tag{16}$$

Assuming, hence, the desired order for the components of p and q (and recalling the notation Π_p^d from Definition 2), we will prove that q can be achieved from p via T^d-transforms by induction on h_0 (see Definition 7). We begin by dealing with the base case $h_0(p, q) = 0$ in the following lemma. [15]

Lemma 5. *If $d \in \mathbb{P}_\Omega$ with $d^{\downarrow} = (d_0, \ldots, d_0, d_1)$, $p, q \in \mathbb{P}_\Omega$ with $q \preceq_d p$ and $h_0(p, q) = 0$, then there exists a finite sequence of T^d-transforms $(T^d_{\lambda_k}(i_k, j_k))_{k=1}^{k_0}$ such that*

$$q = \left(\prod_{k=1}^{k_0} T^d_{\lambda_k}(i_k, j_k) \right) p,$$

where $0 \leqslant \lambda_k \leqslant 1$ and $1 \leqslant i_k < j_k \leqslant |\Omega|$ for $1 \leqslant k \leqslant k_0$.

[14] If that were not the case, we can first apply a sequence of $P^d(i, j)$ with $1 \leqslant i, j < n$ to reach the desired order for p, follow the argument below to reach q with the desired order, and finally apply another sequence of $P^d(i, j)$ with $1 \leqslant i, j < n$ until we reach q.

[15] It should be noted that a more general version of Lemma 5 was obtained in [34, Theorem 12] for partial level thermalizations, a subset of the weak elementary thermal operations.

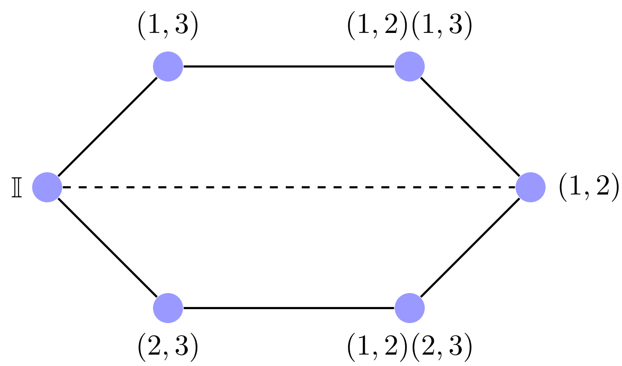


Figure 8: Rough representation of the set $\mathcal{C}_d(p)$ for $d = (d_0, d_0, d_1)$ with $0 < d_1 \leqslant d_0$ and $p = (a, b, c)$ with $\gamma a \geqslant c \geqslant \gamma b$. Note that we use the notation in Figure 7.

Proof. The result follows by extending the argument in Lemma 2 of [31, p. 47]. More specifically, it follows by induction on h_1 (see Definition 7):

If $h_1(p, q) = 0$, then $p = q$ and we have finished.

If $h_1(p, q) = s + 1$ for some $s \geqslant 0$, then, since $\sum_{i=1}^{n}(p_i^d - q_i^d) = 0$ and $\sum_i^\ell q_i^d \leqslant \sum_i^\ell p_i^d$ for $1 \leqslant \ell \leqslant n$ (see Definition 2), there exist some indexes $1 \leqslant k < l \leqslant n$ such that

$$p_k^d > q_k^d,\ p_{k+1}^d = q_{k+1}^d,\ \ \ldots,\ p_{l-1}^d = q_{l-1}^d,\ p_l^d < q_l^d.$$

We distinguish three cases:

(A.1) p_k^d, q_k^d, p_l^d and q_l^d are associated with d_0. In this case, we can follow the proof of Lemma 2 in [31, p. 47] and obtain by T^d-transforms on p some p' such that $h_1(p', q) \leqslant s$ and $q \preceq_d p'$.

(A.2) p_k^d and q_k^d are associated with d_0 and p_l^d and q_l^d are associated with d_1. In this case, we have that $\gamma p_k^d > \gamma q_k^d \geqslant q_l^d > p_l^d$. Hence, there exists some $0 \leqslant \lambda_0 \leqslant 1$ such that $\lambda_0 \gamma p_k^d + (1 - \lambda_0) p_l^d = q_l^d$. If $(1 - \lambda_0 \gamma) p_k^d + \lambda_0 p_l^d \geqslant q_k^d$, then we take $p' = T_{\lambda_0}^d(k, l)p$. Otherwise, we consider some $0 \leqslant \lambda_1 < \lambda_0$ such that $(1 - \lambda_1 \gamma) p_k^d + \lambda_1 p_l^d = q_k^d$ and we take $p' = T_{\lambda_1}^d(k, l)p$. In any case, p' is obtained by applying T^d-transforms on p and $h_1(p', q) \leqslant s$. Moreover, we have

$$\sum_{i=1}^{\ell} (p')_i^d = \sum_{i=1}^{\ell} p_i^d \geqslant \sum_{i=1}^{\ell} q_i^d \text{ for } \ell = 1, \ldots, k-1,$$
$$(p')_k^d \geqslant q_k^d,\ (p')_\ell^d = p_\ell^d = q_\ell^d \text{ for } \ell = k+1, \ldots, l-1,$$
$$\sum_{i=1}^{\ell} (p')_i^d = \sum_{i=1}^{\ell} p_i^d \geqslant \sum_{i=1}^{\ell} q_i^d \text{ for } \ell = l, \ldots, n.$$

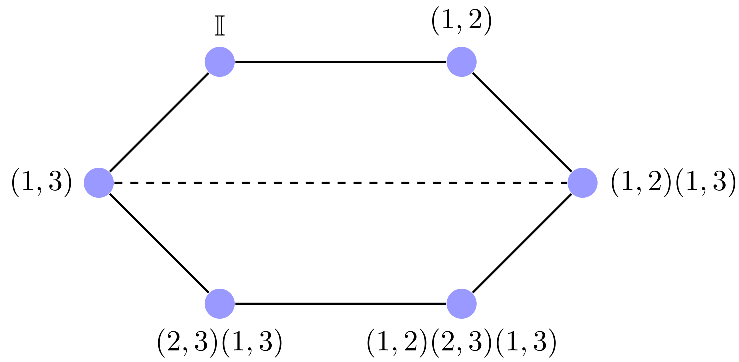


Figure 9: Rough representation of the set $\mathcal{C}_d(p)$ for $d = (d_0, d_0, d_1)$ with $0 < d_1 \leqslant d_0$ and $p = (a, b, c)$ with $c \geqslant \gamma a \geqslant \gamma b$. Note that we use the notation in Figure 7.

Hence, since p' and q have the same d-order by construction, $q \preceq_d p'$.

(A.3) p_l^d and q_l^d are associated with d_0 and p_k^d and q_k^d are associated with d_1. In this case, we have that $p_k^d > q_k^d \geqslant \gamma q_l^d > \gamma p_l^d$ and we can follow an argument analogous to that in $(A.2)$.

This concludes the proof. □

As a result of Lemma 5, to conclude, we only ought to show that, if $h_0(p, q) > 0$, then, applying some T^d-transforms to p, we can obtain some p' such that $q \preceq_d p'$ and $h_0(p', q) < h_0(p, q)$. We take, hence, $h_0(p, q) = m + 1$ for some $m \geqslant 0$ and consider two cases:

(B.1) $\Pi_p^d(n) > \Pi_q^d(n)$. In this scenario, we take i_0 the component for which $i_0 + 1 = \Pi_p^d(i_0) + 1 = \Pi_p^d(n)$, note that $\gamma p_{i_0} > p_n$, and define $p' = T_\lambda^d(i_0, n)p$ for some $0 \leqslant \lambda \leqslant 1$. We distinguish two cases:

(B.1.1) $h_0(p, q) > 1$. In this case, we ought to see that there exists such a λ fulfilling

$$\begin{aligned} &p'_n \geqslant \gamma p'_{i_0} \text{ and} \\ &p'_n + \Delta \geqslant \gamma q_{i_0 - 1}, \end{aligned} \tag{17}$$

where we take $\Delta = \sum_{j=1}^{i_0-1} p_j - (q_n + (1-\gamma) q_{i_0-1} + \sum_{j=1}^{i_0-2} q_j)$ and the first equation assures that $h(p', q) = m$ while the second one assures that $q \preceq_d p'$. (This is the case since the Lorenz d-curve of p' coincides with that of p except for the components that we are modifying, given that $\gamma p_{i_0} > p'_n$ and $\gamma p'_{i_0} > p_n$. Moreover, the conditions in (17) suffice to assure that, in the region where it differs from that of p, the Lorenz d-curve of p' is not below that q.)

Isolating λ, the equations in (17) are equivalent to

$$
\begin{aligned}
\lambda &\geqslant \frac{\gamma p_{i_0} - p_n}{\gamma(1+\gamma)p_{i_0} - (1+\gamma)p_n} \text{ and} \\
\lambda &\geqslant \frac{\gamma q_{i_0-1} - (\Delta + p_n)}{\gamma p_{i_0} - p_n},
\end{aligned} \tag{18}
$$

respectively. It is not difficult to see that the right hand side of both inequalities is bounded by 1: For the first one we use that $\gamma p_{i_0} > p_n$ by assumption and for the second that $\gamma q_{i_0-1} \leqslant \gamma p_{i_0} + \Delta$ since $q \preceq_d p$. Hence, we can find some $0 \leqslant \lambda \leqslant 1$ fulfilling (17).

(B.1.2) $h_0(p,q) = 1$. This case is analogous to $(B.1.1)$, substituting (17) by

$$
\begin{aligned}
&p'_n \geqslant \gamma p'_{i_0} \text{ and} \\
&p'_n + \Delta \geqslant q_n,
\end{aligned} \tag{19}
$$

with $\Delta = \sum_{j=1}^{i_0-1}(p_j - q_j)$, and (18) by

$$
\begin{aligned}
\lambda &\geqslant \frac{\gamma p_{i_0} - p_n}{\gamma(1+\gamma)p_{i_0} - (1+\gamma)p_n} \text{ and} \\
\lambda &\geqslant \frac{q_n - (\Delta + p_n)}{\gamma p_{i_0} - p_n}.
\end{aligned}
$$

It is not difficult to see that the right hand side of both inequalities is bounded by 1: The first follows like (18) and the second since $q \preceq_d p$ and, hence, $q_n \leqslant \gamma p_{i_0} + \Delta$. Thus, there exists some $0 \leqslant \lambda \leqslant 1$ fulfilling (19).

(B.2) $\Pi_p^d(n) < \Pi_q^d(n)$. In this scenario, we take i_0 the component for which $\Pi_p^d(i_0) - 1 = \Pi_p^d(n)$, note that $p_n > \gamma p_{i_0}$, and define $p' = T_\lambda^d(i_0, n)p$ for some $0 \leqslant \lambda \leqslant 1$. We can conclude, analogously to $(B.1.1)$, by finding such a λ fulfilling

$$
\begin{aligned}
&\gamma p'_{i_0} \geqslant p'_n \text{ and} \\
&p'_{i_0} + \Delta \geqslant q_{i_0},
\end{aligned} \tag{20}
$$

where we take $\Delta = \sum_{j=1}^{i_0-1}(p_j - q_j)$ and the first equation assures that $h(p', q) = m$ while the second one assures that $q \preceq_d p'$. These equations are equivalent to

$$
\begin{aligned}
\lambda &\geqslant \frac{p_n - \gamma p_{i_0}}{(1+\gamma)p_n - \gamma(1+\gamma)p_{i_0}} \text{ and} \\
\lambda &\geqslant \frac{q_{i_0} - (\Delta + p_{i_0})}{p_n - \gamma p_{i_0}},
\end{aligned} \tag{21}
$$

respectively. It is not difficult to see that the right hand side of both inequalities is bounded by 1: For the first one we use that $\gamma p_{i_0} < p_n$ by assumption and for the second that $q_{i_0} \leqslant \Delta + p_n + (1-\gamma)p_{i_0}$ since $q \preceq_d p$.

By induction, this concludes the proof.

I Proof of Theorem 7

Most of the theorem is due to Jurkat and Ryser [41, Theorem 4.1] and its final form due to Hartfiel [39, Lemma 1.1]. The only thing we ought to notice is that we have to take the transpose of the result in [39] (since it uses a different convention for stochastic matrices) and that, if a matrix A can be constructed via the Jurkat-Ryser algorithm, then its transpose A^T can as well. (We simply follow the steps in the construction of A transposing the indexes we use in each of them.)

J Proof of Theorem 8

By Lemma 1, it suffices to assume that $d = d^{\downarrow}$ throughout the proof.

Sufficiency follows as in Proposition 2, i.e., as a direct consequence of Proposition 1.

To prove necessity, we will construct a decomposition as a product of d-swaps of an arbitrary extreme point of the thermal operations polytope M (whose form we know from Theorem 7). We take $\gamma = d_1/d_0$ and $n = |\Omega|$ for simplicity, and assume throughout that $d_1 < d_0$. (The case with equality is well known by Theorem 2.)

As a first step, we determine the possible entries of M. In particular, we note that, for $1 \leqslant i, j \leqslant n$, $M_{i,j} \in \{1, \gamma, 1-\gamma, 0\}$. We show this in the following lemma, where we start proving and analogous result for Jurkat-Ryser d-matrices.

Lemma 6. *If $d = (d_0, \ldots, d_0, d_1) \in \mathbb{P}_\Omega$ and $\gamma = d_1/d_0$, then the following statements hold:*

(a) If A is a Jurkat-Ryser d-matrix, then

$$A_{i,j} \in \{d_0, d_1, d_0 - d_1, 0\} \tag{22}$$

for $1 \leqslant i, j \leqslant |\Omega|$.

(b) If M is an extreme point of the thermal operations polytope, then

$$M_{i,j} \in \{1, \gamma, 1-\gamma, 0\} \tag{23}$$

for $1 \leqslant i, j \leqslant |\Omega|$.

Proof. (a) By definition, the only non-zero components of A consist of the minimum of r_i^m and s_j^m for some $m \geqslant 1$ and $1 \leqslant i, j \leqslant |\Omega|$ (see Definition 8). Moreover, aside from the zeros which are generated by the Jurkat-Ryser algorithm (which are not used again for any comparison later on) we have that $r_i^m = s_j^m = d_0$ unless r_i^m or s_j^m is connected to d_1. In particular, one can see that $r_i^m = d_1$ if it is row-connected to d_1 and $r_i^m = d_0 - d_1$ if it is column-connected to d_1. Similarly, $s_j^m = d_1$ if it is column-connected to d_1 and $s_j^m = d_0 - d_1$ if it is row-connected to d_1. To conclude, note that, if r_i^m (s_j^m) is row-connected to d_1, then s_j^m (r_i^m)

is either column-connected to d_1 or not connected to d_1 at all, and vice versa. Hence, in case r_i^m and s_j^m are both connected to d_1, there are only two possible scenarios:

(a.1) r_i^m is row-connected to d_1 and s_j^m is column-connected to d_1. In this case, we have $r_i^m = s_j^m = d_1$.

(a.2) r_i^m is column-connected to d_1 and s_j^m is row-connected to d_1. In this case, we have $r_i^m = s_j^m = d_0 - d_1$.

As a result, we have that

$$(r_i^m, s_j^m) = \begin{cases} (a,b), & \text{or} \\ (b,a), \end{cases}$$

where

$$(a,b) \in \{(d_0,d_0),(d_0,d_1),(d_0,d_0-d_1),(d_1,d_1),(d_0-d_1,d_0-d_1)\}.$$

Hence, (22) holds.

(b) This follows as a direct consequence of (a) since, by Theorem 7, $M = AD^{-1}$, where A is a Jurkat-Ryser d-matrix and $D = diag(d_0, \ldots, d_0, d_1)$. To conclude, it suffices to notice that $A_{i,|\Omega|} \in \{0, d_1\}$ for all $1 \leqslant i \leqslant |\Omega|$, since, provided it is not zero, we have for all $m \geqslant 1$ that $s_{|\Omega|}^m = d_1$ and hence, as we argued in (a), $r_i^m \in \{d_0, d_1\}$ for $1 \leqslant i \leqslant |\Omega|$.

This concludes the proof. □

Now that we know what entries M may have, we show, in the following lemma, how they may be distributed along its lines.[16]

Lemma 7. *If $d = (d_0, \ldots, d_0, d_1) \in \mathbb{P}_\Omega$, $d_1 < d_0$, $\gamma = d_1/d_0$ and M is an extreme point of the thermal operations polytope, then the lines of M fulfill the following properties:*

(a) The last row has either a 1 in the last column or a γ in another column and the rest are zeros.

(b) If a row which is not the last one has a 1 in the last column, then it also has a $1-\gamma$ and the rest are zeros.

(c) If a row which is not the last one has a γ, then it also has a $1-\gamma$ (both not in the last column) and the rest are zeros.

(d) The last column has a single one and the rest are zeros.

(e) If a column has a $1-\gamma$, then it also has a γ and the rest are zeros.

Proof. (a) This follows from Lemma 6 and the fact that $Md = d$.

[16] If $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$, then a **line** of M is either a row or a column [41].

(b) Take R_i the i-th row for $1 \leqslant i < |\Omega|$ and assume it has a one in the last column. Since $Md = d$, there must be another entry in R_i that is neither 0 nor 1. Moreover, there can be no γ in R_i. To show this, we consider two cases:

(b.1) There is one γ in R_i that was generated before the 1 at step m. Then, arguing as in Lemma 6, γ must come from the comparison of $r_i^m = d_0$ and $s_j^m = d_1$ to avoid filling the row with zeros. As a result, we have $r_i^{m+1} = d_0 - d_1$ and, by the argument in Lemma 6, the next non-zero entry in R_i will be $1 - \gamma$. However, this scenario is impossible since the existence of the 1 contradicts the fact that $Md = d$.

(b.2) There is one γ in R_i that was generated after the 1. If we assume the 1 was generated at step m, then, arguing as in Lemma 6, it must come from the comparison of $r_i^m = d_0$ and $s_j^m = d_1$ to avoid filling the row with zeros and, furthermore, the next non-zero entry in the row will be $1 - \gamma$. However, this scenario is impossible since the existence of the γ contradicts the fact that $Md = d$.

In summary, by Lemma 6 (b), a row with the 1 in the last column has a $1 - \gamma$ somewhere and the rest are zeros.

(c) Take R_i the i-th row for $1 \leqslant i < |\Omega|$ and assume it has a γ, which cannot be in the last column by Lemma 6 (b). Since R_i is not the last row and $Md = d$, then there must be another non-zero entry in R_i. Moreover, R_i cannot have another γ. If there were another γ, consider the γ that appeared first at step m. Arguing as in Lemma 6, it must come from the comparison of $r_i^m = d_0$ and $s_j^m = d_1$ to avoid filling the row with zeros. Thus, the next non-zero entry in the row will be a $1 - \gamma$ and the existence of a second γ contradicts the fact that $Md = d$. Lastly, R_i cannot have a 1. This is the case since, if the one was not on the last column, then this would contradict the fact that $Md = d$. Furthermore, if it were on the last column, we can follow the proof of (b). In summary, since $Md = d$, a row which is not the last one and has γ will also have another a $1 - \gamma$ and the rest zeros.

(d) This follows from Lemma 6 (b) plus the fact that M is a stochastic matrix.

(e) Take C_i the i-th column for $1 \leqslant i < |\Omega|$ a column with a $1 - \gamma$. (It cannot be the last column by (b).) Since M is stochastic, there must be another non-zero entry which cannot be a 1. Moreover, there can be no other $1 - \gamma$ in C_i. To show this, let us consider the $1 - \gamma$ that appeared first at step m. Arguing as in Lemma 6, such $1 - \gamma$ must come from the comparison of $r_i^m = d_0 - d_1$ and $s_j^m = d_0$ to avoid filling the column with zeros. Thus, the next non-zero entry in the row will be a γ. However, the existence of a second $1 - \gamma$ in C_i contradicts the fact that M is stochastic. In summary, by Lemma 6 (b), a column with a $1 - \gamma$ entry will have another entry with a γ and the rest zeros.

This concludes the proof. □

Now that we have identified the possible lines M may have, we will establish how they are positioned in M relative to each other in the following lemma.

Lemma 8. *If $d = (d_0, \ldots, d_0, d_1) \in \mathbb{P}_\Omega$, $d_1 < d_0$, $\gamma = d_1/d_0$ and M is an extreme point of the thermal operations polytope, then one of the following holds:*

(a) M is a permutation matrix with $M_{|\Omega|,|\Omega|} = 1$.

(b) There exists a sequence of pairs $Q = ((i_k, j_k))_{k=0}^{t_0}$ such that

$$\begin{aligned} &j_0 = |\Omega|, && M_{i_0,j_0} = 1, \\ &M_{i_{k-1},j_k} = 1-\gamma, && M_{i_k,j_k} = \gamma, \quad \text{if } 1 \leqslant k < t_0, \\ &M_{i_{t_0-1},j_{t_0}} = 1-\gamma, && M_{i_{t_0},j_{t_0}} = \gamma, \end{aligned} \tag{24}$$

*where $i_{t_0} = |\Omega|$, $1 \leqslant j_{t_0} < |\Omega|$, and $0 < t_0 < \infty$. Moreover, the **submatrix**[17] $M_0 = M \backslash M[i_0, \ldots, i_{t_0}; j_0, \ldots, j_{t_0}]$ is a permutation matrix.*

Proof. By Lemma 7 (d), M must have a 1 and the rest zeros in the last column. In case the 1 is in the last row, then, following Lemma 6, it was introduced at some step m_0 comparing $r_n^{m_0} = d_1$ with $s_n^{m_0} = d_1$ and, for all $m \neq m_0$, both r_i^m and s_j^m are not connected to d_1. Hence, M is a permutation matrix with $M_{n,n} = 1$, as stated in (a).

Assume now that the 1 in the last column of M is in some row R_{i_0} with $1 \leqslant i_0 < n$. We will show that (b) holds. In particular, in this scenario, we define the sequence of pairs $Q = ((i_k, j_k))_{k=0}^{t_0}$, where we take i_0 as in the previous line and $j_0 = n$. Moreover, we define, for all $k \geqslant 1$, j_k such that the column C_{j_k} has a $1-\gamma$ in row i_{k-1}, and i_k such that the row R_{i_k} has a γ in column C_{j_k}. We follow this procedure until we reach some $t_0 > 0$ for which j_{t_0+1} it is no longer defined. To conclude that (24) holds, we check the following properties:

(b.1) There exists some $t_0 > 0$ for which Q is well-defined. To show this, we rely on Lemma 7. In particular, there exists a single $1-\gamma$ in row R_{i_0} by Lemma 7 (b). Moreover, for all $k \geqslant 1$, there exists a single γ in column C_{j_k} by Lemma 7 (e). Furthermore, while $i_k < n$, there exists a single $1-\gamma$ in row R_{i_k} by Lemma 7 (c).

(b.2) $t_0 < \infty$. To show this, we first note that we never repeat a pair $(i_{k_0}, j_{k_0}) = (i_{k_1}, j_{k_1})$ for $0 \leqslant k_0 < k_1$. If that were the case, then we would have $(i_{k_0-v}, j_{k_0-v}) = (i_{k_1-v}, j_{k_1-v})$ for all $0 \leqslant v \leqslant k_0$ by the uniqueness properties in Lemma 7. However, $i_0 \neq i_k$ for all $k > 0$, since there is no γ in the i_0 row by Lemma 7 (b). Hence, since there are no repeated pairs in

[17] If $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$, then M_0 is a **submatrix** of M, denoted

$$M_0 = M \backslash M[a_1, \ldots, a_n; b_1, \ldots, b_m],$$

if it is equal to M after eliminating rows $\{a_1, \ldots, a_n\}$ and columns $\{b_1, \ldots, b_m\}$ with $1 \leqslant a_i, b_j \leqslant |\Omega|$ for $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$ [41].

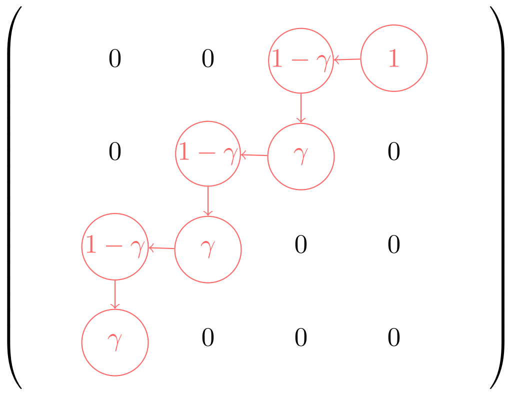


Figure 10: Extreme point of the thermal polytope that fulfills Lemma 8 (b) for $|\Omega| = 4$. We highlight in red the components that correspond to the Q associated to M and connect them with arrows that join one step in the recursion (24) with the following one. Several examples with $|\Omega| = 3$ can be found in (14).

Q and the number of γ in M is finite, there exists some $t_0 > 0$ such that $i_{t_0} = n$. (This is the case since we are not in scenario (a) and Lemma 7 (a) holds.) Thus, since there is no $1 - \gamma$ in row R_n by Lemma 7 (a), j_{t_0+1} is not defined.

To conclude the proof, we only ought to show that the submatrix M_0 is a permutation matrix. To show this, we first notice that $M_{i,j}$ is connected to d_1 if and only if $(i, j) \in Q$. Hence, $(M_0)_{i,j} \in \{0, 1\}$. Moreover, whenever we have $i = i_k$ and $j \neq j_k$ (or vice versa) for some $0 \leqslant k \leqslant t_0$, then $M_{i,j} = 0$. Hence, given that M is a stochastic matrix, we have that M_0 is a permutation matrix. □

(We include an extreme point of the thermal polytope that fulfills Lemma 8 (b) in Figure 10.)

Now that we know the structure of M, we conclude the proof defining a product of ETOs N_2 such that $M = N_2$. We consider two cases:

(a) If we are in the scenario of Lemma 8 (a), then M is a permutation matrix that acts as the identity on the last column. Hence, by Theorem 2, there exists a sequence of pairs $((x_k, y_k))_{k=0}^{t_2}$ with $x_k, y_k < n$ for $0 \leqslant k \leqslant t_2$, such that

$$M = N_2 := \prod_{k=0}^{t_2} P^d(x_k, y_k).$$

(b) If we are in the scenario of Lemma 8 (b), then we start by defining

$$N_0 := \prod_{k=0}^{t_0} P^d(j_k, n),$$

where $(j_k)_{k=0}^{t_0}$ are the second components of the pairs in Q from Lemma 8.

It is not difficult to see by induction that any product

$$\prod_{q=1}^{q_0} P^d(u_q, n),$$

where $u_q < n$ for $1 \leqslant q \leqslant q_0$ and $u_q \neq u_{q'}$ whenever $q \neq q'$, fulfills the following relation between rows: Row R_{u_1} has a 1 in column C_n and a $1-\gamma$ in column C_{u_1}. Moreover, row R_{u_k} has a γ in column $C_{u_{k-1}}$ and a $1-\gamma$ in column C_{u_k} for $1 < k \leqslant q_0$. Finally, row R_n has a γ in column u_{q_0}. (Furthermore, the entries that are not mentioned for each row are zero.)

Hence, although potentially in different rows, N_0 possesses one horizontal line equivalent to each R_{i_k} for $1 \leqslant k \leqslant t_0$, where $(i_k)_{k=0}^{t_0}$ are the first components of the pairs in Q from Lemma 8. Thus, there exists a sequence $(o_k)_{k=0}^{t_0-1}$ with $o_k < n$ for $0 \leqslant k \leqslant t_0 - 1$ such that

$$N_1 := \left(\prod_{k=0}^{t_0-1} P^d(j_k, o_k) \right) N_0$$

coincides with M inside $[i_0, \ldots, i_{t_0}; j_0, \ldots, j_{t_0}]$.

To conclude, note that, since $N_1 \backslash N_1[i_0, \ldots, i_{t_0}; j_0, \ldots, j_{t_0}]$ and M_0 are permutation matrices, there exists a sequence of pairs $((a_k, b_k))_{k=0}^{t_1}$ with $a_k, b_k < n$ for $0 \leqslant k \leqslant t_1$, such that

$$M = N_2 := \left(\prod_{k=0}^{t_1} P^d(a_k, b_k) \right) N_1.$$

This concludes the proof.

K Proof of Theorem 9

By Lemma 1, it suffices to assume that $d = d^{\downarrow}$ throughout the proof.

(b) implies (a) by Lemma 2. We prove the converse by contrapositive. In particular, if $d \neq (d_0, \ldots, d_0, d_1)$ up to permutations, then the result holds already at the resource theory level by Proposition 1. We take, hence, $d = (d_0, \ldots, d_0, d_1)$ with $d_1 < d_0$. (The case where $d_1 = d_0$ is known by Theorem 3.) To deal with this situation, we distinguish two cases and provide, for each of them, a d-stochastic matrix $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ that cannot be decomposed as a product of T^d-transforms:

(A) $|\Omega| \geqslant 4$. In this scenario, the result follows as a consequence of the proof of Theorem 3 (see [43]). In particular, we can take

$$M := M_0 \bigoplus \mathbb{I}_{\backslash(1,\ldots,|\Omega|-1|)},$$

with $M_0 \in \mathcal{M}_{|\Omega|-1,|\Omega|-1}(\mathbb{R})$ a doubly stochastic matrix such that $(M_0)_{i,j} > 0$ if $i \neq j$ and $(M_0)_{i,j} = 0$ if $i = j$. (For instance, we could define M_0 as in (12).) It is clear that M is a d-stochastic matrix. To conclude, assume that M can be decomposed as a sequence of T^d-transforms

$$M = \prod_{k=1}^{N} T^d_{\lambda_k}(i_k, j_k) \tag{25}$$

and consider N_0 the first index such that $\lambda_{N_0} > 0$ and $j_{N_0} = |\Omega|$. In this scenario, we have that

$$\left(\prod_{k=1}^{N_0} T^d_{\lambda_k}(i_k, j_k) \right)_{|\Omega|,|\Omega|} < 1.$$

It is then easy to see that this implies $(\prod_{k=1}^{N_1} T^d_{\lambda_k}(i_k, j_k))_{|\Omega|,|\Omega|} < 1$ for $N_0 \leqslant N_1 \leqslant N$, which contradicts the definition of M. Hence, for $1 \leqslant k \leqslant N$, $j_k = |\Omega|$ implies $\lambda_k = 0$, thus, $T^d_{\lambda_k}(i_k, j_k) = \mathbb{I}$. As a result, M can be decomposed as a product of T^d transforms if and only if M_0 can be decomposed as a product of T-transforms. The latter is, however, false due to the proof of Theorem 3.

(B) $|\Omega| = 3$. In this scenario, the argument above does not hold since Theorem 3 requires M_0 to be at least a 3×3 matrix. In the spirit of the proof of Theorem 3, however, we take

$$M := \begin{pmatrix} 1-\phi & \phi & 0 \\ \phi & 0 & \frac{1-\phi}{\gamma} \\ 0 & 1-\phi & 1-\frac{1-\phi}{\gamma} \end{pmatrix} \tag{26}$$

with $1 - \gamma < \phi < 1$. Note that M is clearly a d-stochastic matrix.

Since there exists some $0 \leqslant \lambda' \leqslant 1$ such that

$$\prod_{k=1}^{k_0} T^d_{\lambda_k}(i, j) = T^d_{\lambda'}(i, j), \tag{27}$$

where $1 \leqslant i, j \leqslant 3$ and $0 \leqslant \lambda_k \leqslant 1$ for $1 \leqslant k \leqslant k_0$, and, moreover,

$$T^d_\lambda(1, 3) = P^d(1, 2) T^d_\lambda(2, 3) P^d(1, 2) \tag{28}$$

for any $0 \leqslant \lambda \leqslant 1$, then M can be decomposed as product of T^d-transforms if and only if there exist some $\ell, m \in \{0, 1\}$ and $N \geqslant 0$ such that

$$M = \left(T^d_{\lambda_A}(2,3)\right)^\ell \left(\prod_{k=1}^{N} T^d_{\lambda_k}(1,2) T^d_{\beta_k}(2,3) \right) \left(T^d_{\lambda_B}(1,2)\right)^m, \tag{29}$$

where $0 \leqslant \lambda_A, \lambda_B, \lambda_k, \beta_k \leqslant 1$ for $1 \leqslant k \leqslant N$. To conclude the argument, it is straightforward to check that this is not the case. (For instance, one can argue by direct calculation using the zeros in M.)

As a last remark, note that, in case $\phi = 1$, then M equals $P^d(1,2)$ and, hence, it belongs to the WETO polytope. Moreover, if $\phi = 1 - \gamma$, then M also belongs to the WETO polytope. In particular, $M = P^d(1,2)P^d(2,3)P^d(1,3)$.

This concludes the proof.

Remark 6. *Note that, for each* $1-\gamma < \phi < 1$*,* M *in* (26) *exemplifies that the WETO polytope is not convex in general since, aside from its extreme points, the segment joining* $P^d(1,2)$ *and* $P^d(1,2)P^d(2,3)P^d(1,3)$ *does not belong to it by Theorem 9.*

L Proof of Theorem 10

By Lemma 1, it suffices to assume that $d = d^{\downarrow}$ throughout the proof.

We argue by contrapositive, that is, we assume we have some $d \in \mathbb{P}_\Omega$ with three different entries and will show that, in this scenario, there exists some pair $p, q \in \mathbb{P}_\Omega$ such that q is achievable from q via (strong) elementary thermal operations but not by its weaker counterpart.

We first note that we can restrict ourselves to the case where $|\Omega| = 3$, as we show in the following lemma.

Lemma 9. *Take* $0 < d \in \mathbb{P}_\Omega$ *and* $0 < d' \in \mathbb{P}_{\Omega'}$ *such that* $\Omega \subseteq \Omega'$ *and* $d'_i = d_i$ *for all* $i \in \Omega$*. If* $p, q \in \mathbb{P}_\Omega$*, then the following statements hold:*

(a) If $q \in C_d^{ETO}(p)$*, then* $(q,0) \in C_{d'}^{ETO}((p,0))$*, where* $(r,0) \in \mathbb{P}_{\Omega'}$ *and*

$$(r,0)_i := \begin{cases} r_i, & \text{if } i \in \Omega \\ 0, & \text{if } i \in \Omega' \backslash \Omega. \end{cases}$$

(b) If $supp(p) = supp(q)$ *and* $q \notin C_d^{WETO}(p)$*, then* $(q,0) \notin C_{d'}^{WETO}((p,0))$*.*

Proof. (a) Straightforward.

(b) For simplicity, we assume throughout the proof that $\text{supp}(p) = \text{supp}(q) = \Omega$. The same method can be followed whenever $\text{supp}(p) = \text{supp}(q) < \Omega$. Moreover, by Lemma 1, we assume $d' = (d')^{\downarrow}$ and $d = d^{\downarrow}$.

We begin assuming $(q,0) \in C_{d'}^{WETO}((p,0))$ and argue by contradiction. By assumption, we have that

$$(q,0) = \left(\prod_{k=1}^{\ell} T_{\lambda_k}^d(i_k, j_k)\right)(p,0) \tag{30}$$

for some $\ell \geqslant 0$ and $1 \leqslant i_k < j_k \leqslant |\Omega'|$ for $1 \leqslant k \leqslant \ell$. Note that, in the remainder of the proof, we use the notation

$$t^m := \begin{cases} (p,0), & \text{if } m = 0 \\ \left(\prod_{k=1}^{m} T_{\lambda_k}^d(i_k, j_k)\right)(p,0), & \text{if } 1 \leqslant m \leqslant \ell. \end{cases} \tag{31}$$

Since $q \notin C_d^{WETO}(p)$, there must be some $1 \leqslant k_0 \leqslant \ell$ such that $i_{k_0} \notin \Omega$ (or $j_{k_0} \notin \Omega$). We fix k_0 the first index with this property such that $d'_{i_{k_0}} \neq d_m$ ($d'_{j_{k_0}} \neq d_m$) for all $m \in \Omega$. (We deal with the remaining cases later on.) Note we can assume w.l.o.g. that

$$t_s^{k_0} > 0 \text{ for } s = i_{k_0} \ (s = j_{k_0}). \tag{32}$$

In particular, since k_0 is minimal by construction, we can assume that $d'_{j_{k_0}} = d_m$ ($d'_{i_{k_0}} = d_m$) for some $m \in \Omega$.

The existence of such a k_0 leads to contradiction. To show this, we consider the following two cases:

(A) $j_{k_0} \notin \Omega$. By assumption, we have that $d'_{i_{k_0}} > d'_{j_{k_0}}$. In particular, by (32), we obtain that

$$|\mathrm{supp}(t^{k_0})| > |\mathrm{supp}((p,0))| = |\mathrm{supp}((q,0))|.$$

This, together with (30), contradicts Lemma 3.

(B) $i_{k_0} \notin \Omega$. In this case, we have that $T^d_{\lambda_k}(i_k, j_k)$ is a d-swap to avoid contradicting Lemma 3. Moreover, to avoid contradicting (30), there must be some $k_0 < m_0 \leqslant \ell$ such that

$$t^m_{i_{k_0}} = 0 \text{ for all } m \geqslant m_0. \tag{33}$$

Hence, there must be some $i_{k_0} > w_0 \in \Omega$ and two sequences $(n_p)_{p=1}^N$, with $k_0 < n_p < n_{p+1} \leqslant \ell$ for all p, and $(v_p)_{p=1}^N \subseteq \Omega'$ such that $t^{n_{p+1}}_{v_{p+1}} > 0$ and $t^{n_{p+1}}_{v_p} = 0$, where $v_0 = k_0$, $d'_{v_{p+1}} \geqslant d'_{v_p}$ by (A) and $v_N = w_0$. However, to avoid contradicting (33), we must have $t^{n_{N-1}}_{v_N} = 0$. Despite this, $t^0_{v_N} > 0$ since $\mathrm{supp}(p) = \Omega$. Thus, we can argue analogously that there must be some $w_0 > w_1 \in \Omega$ with similar properties to those of w_0 and $d_{w_1} > d_{w_0}$. (Note that we could have $d_{w_1} = d_{w_0}$. However, we could argue in a similar fashion that this would imply the existence of some $w_1 > w'_1 \in \Omega$ with equivalent properties and such that $d_{w'_1} > d_{w_1} = d_{w_0}$.) Following the argument recursively, we obtain an unbounded strictly decreasing sequence $(w_n)_{n \geqslant 0} \subseteq \Omega$. This contradicts the finiteness of Ω.

By (A) and (B), if (30) holds, then, for all $1 \leqslant k \leqslant \ell$,

$$d'_{i_k}, d'_{j_k} \in \{d_i | i \in \Omega\}.$$

We conclude noting two properties. First, whenever $d_{i_k} = d_{j_k}$ with $t^{k-1}_{i_k} > 0$ and $t^{k-1}_{j_k} = 0$ or vice versa, then, for all $1 \leqslant k \leqslant \ell$, $T^d_{\lambda_k}(i_k, j_k)$ in (30) must be a swap to avoid contradicting Lemma 3. Second, whenever $d_{i_k} > d_{j_k}$, we must have $t^{k-1}_{i_k}, t^{k-1}_{j_k} > 0$ by the same reason. (If $t^{k-1}_{j_k} = 0$ we can argue as in (A) and, if $t^{k-1}_{i_k} = 0$, as in (B).) Using these properties, it is easy to see that (30) implies $q \in C_d^{WETO}(p)$. (In particular, for all $0 \leqslant m \leqslant \ell$, there exists some $r^m \in \mathbb{P}_\Omega$ such that $t^m = (r^m, 0)$ up to permutations and $r^m \in \mathcal{C}_d^{WETO}(p)$.) This yields the desired contradiction and concludes the proof of (b).

This concludes the proof. □

(Note that one may not need the assumption on the support in Lemma 9. However, this is not important for our purposes here.)

As a result of Lemma 9, it suffices to fix $|\Omega| = 3$ and find a pair $p, q \in \mathbb{P}_\Omega$ with $\text{supp}(p) = \text{supp}(q) = \Omega$ such that $q \in C_d^{ETO}(p) \backslash C_d^{WETO}(p)$. We conclude the proof showing such a pair exists.

In the following, we fix $|\Omega| = 3$ and take $d = (d_0, d_1, d_2)$ with $d_0 > d_1 > d_2$ (we can do so by assumption). Moreover, we fix $\alpha = d_1/d_0$, $\beta = d_2/d_0$ and $\gamma = d_2/d_1$, and take some $p = (a, b, c) \in \mathbb{P}_\Omega$ such that

$$\begin{aligned} &\gamma b < c < \beta a, \text{ and} \\ &\beta((1-\alpha)a + (1-\gamma)b + c) < c. \end{aligned} \tag{34}$$

(Note that both conditions in (34) can be simultaneously satisfied. This is easy to see by assuming $b = 0$ and noticing that, in this scenario, the conditions are fulfilled provided $1/(1+\beta) < a < 1/(1+\tau\beta)$, where $\tau = (1-\alpha)/(1-\beta)$. It is then easy to get some p fulfilling (34) with $0 < b$, that is, with $\text{supp}(p) = \Omega$.)

To conclude, we argue by contradiction that there exists some $q \in C_d^{ETO}(p)$ that cannot be achieved via weak elementary thermal operations. In order to do so, we first note that $C_d^{ETO}(p)$ is given by Figure 11. (Note that this can be achieved via direct calculation using [2, Theorem 6], or [48, Theorem 5] for simplicity, and eliminating the non-extreme points later on.)

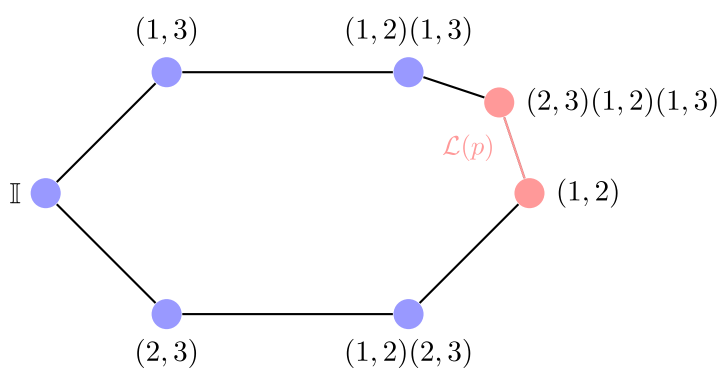


Figure 11: Rough representation of the set $C_d^{ETO}(p)$ for $d = (d_0, d_1, d_2)$ with $d_1 > d_1 > d_2$ and $p = (a, b, c)$ fulfilling (34). We include in red $\mathcal{L}(p)$ as defined in (35). Note that we use the notation in Figure 7.

We will show there exists some

$$q \in \mathcal{L}(p) := \left[P^d(2,3)P^d(1,2)P^d(1,3)p, P^d(1,2)p\right] \subseteq C_d^{ETO}(p) \tag{35}$$

such that q cannot be achieved from p via weak elementary thermal operations, where $[A, B]$ stands for the segment joining points A and B. (Note that, for any $q \in \mathcal{L}(p)$, we have $\text{supp}(q) = \Omega = \text{supp}(p)$. Moreover, $\mathcal{L}(p) \backslash \text{ext}(\mathcal{L}(p)) \neq \emptyset$, where $\text{ext}(\cdot)$ denotes the set of extreme points.)

We begin fixing some $q \in \mathcal{L}(p)$ such that

$$q \neq \left(\prod_{k=1}^{m} P^d(i_k, j_k) \right) p \tag{36}$$

for all $m \geqslant 1$ and $1 \leqslant i_k < j_k \leqslant 3$ for $1 \leqslant k \leqslant m$. (This can be done since there is a continuum number of points points in $\mathcal{L}(p)$.) Hence, if we assume that $q \in C_d^{WETO}(p)$, then there exists at least one T^d-transform that is not a d-swap in the sequence of T^d-transforms that yield q when applied to p. That is, we have

$$\begin{aligned} q &= \left(\prod_{k=m+2}^{\ell} P^d(i_k, j_k) \right) \left((1-\lambda)\mathbb{I} + \lambda P^d(i_{m+1}, j_{m+1}) \right) p_0, \text{ with} \\ p_0 &:= \left(\prod_{k=1}^{m} T^d_{\lambda_k}(i_k, j_k) \right) p, \end{aligned} \tag{37}$$

for some $\ell \geqslant m+1$, $m \geqslant 0$, $0 < \lambda < 1$ and $1 \leqslant i_k < j_k \leqslant 3$ for $1 \leqslant k \leqslant \ell$.

We will show (37) leads to a contradiction, implying $q \notin C_d^{WETO}(p)$. In order to do so, we prove a stronger result in the following lemma.

Lemma 10. *If we define, for all $r, s \in \mathbb{P}_\Omega$ and $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$, $M[r, s] := [Mr, Ms]$, and, for any segment segment $L \subseteq \mathbb{R}^3$, $E(L) := L' \bigcap \mathcal{C}_d^{ETO}(p)$ with $L' \subseteq \mathbb{R}^3$ the line such that $L \subseteq L'$, then*

$$\mathcal{L}(p) \notin \bigcup_{\ell \geqslant 0} B_\ell, \tag{38}$$

where

$$B_\ell := \left\{ E\left(\left(\prod_{k=1}^{\ell} P^d(i_k, j_k) \right) \left[p_0, P^d(i_0, j_0) p_0 \right] \right) : 1 \leqslant i_k < j_k \leqslant 3 \right.$$
$$\left. \text{for } 1 \leqslant k \leqslant \ell \text{ and } p_0 \text{ is given by (37)} \right\}.$$

Proof. We will prove that $\mathcal{L}(p) \notin B_\ell$ for all $\ell \geqslant 0$ by induction.

If $\ell = 0$, then each element in B_ℓ has one constant component by definition. However, it is not difficult to see that the extreme points of $\mathcal{L}(p)$ differ in all their components. Hence, $\mathcal{L}(p) \notin B_0$.

Assume now that $\mathcal{L}(p) \notin B_\ell$ for some $\ell \geqslant 0$. To show this implies $L(p) \notin B_{\ell+1}$, we note that

$$B_{\ell+1} = \left\{ E\left(P^d(i_{\ell+1}, j_{\ell+1}) L \right) : L \in B_\ell \text{ and } 1 \leqslant i_{\ell+1} < j_{\ell+1} \leqslant 3 \right\}, \tag{39}$$

and prove separately the following three claims:

(A) $\mathcal{L}(p) \neq E(P^d(1,2)L)$ for all $L \in B_\ell$. Since $P^d(1,2)$ leaves the third component of every probability distribution unchanged, it is clear that we only need to consider the cases where L belongs to the area with horizontal blue lines in Figure 12. Given that $\mathcal{L}(p) \neq L$ for all $L \in B_\ell$ by assumption, it is then easy to show that $\mathcal{L}(p) \neq E(P^d(1,2)L)$ using that $P^d(1,2)P^d(1,3)p, p \notin \mathcal{L}(p)$ and that $P^d(1,2)P^d(1,2) = \alpha\mathbb{I} + (1-\alpha)P^d(1,2)$.

(B) $\mathcal{L}(p) \neq E(P^d(1,3)L)$ for all $L \in B_\ell$. Since $P^d(1,3)$ leaves the second component of every probability distribution unchanged, it is clear that we only need to consider the cases where L belongs to the shaded blue area in Figure 12. Given that $\mathcal{L}(p) \neq L$ for all $L \in B_\ell$ by assumption, it is then easy to show that $\mathcal{L}(p) \neq E(P^d(1,2)L)$ using that

$$P^d(1,3)p, P^d(1,3)P^d(2,3)p, P^d(1,3)P^d(1,2)P^d(2,3)p,$$
$$P^d(1,3)P^d(1,2)p \notin \mathcal{L}(p)$$

and that $P^d(1,3)P^d(1,3) = \beta\mathbb{I} + (1-\beta)P^d(1,2)$.

(C) $\mathcal{L}(p) \neq E(P^d(2,3)L)$ for all $L \in B_\ell$. Since $P^d(2,3)$ leaves the first component of every probability distribution unchanged, it is clear that we only need to consider the cases where L belongs to the area with vertical red lines in Figure 12. Given that $\mathcal{L}(p) \neq L$ for all $L \in B_\ell$ by assumption, it is then easy to show that $\mathcal{L}(p) \neq E(P^d(2,3)L)$ using that $P^d(2,3)P^d(1,2)p, P^d(1,2)P^d(1,3)p \notin \mathcal{L}(p)$ and that $P^d(2,3)P^d(2,3) = \gamma\mathbb{I} + (1-\gamma)P^d(2,3)$.

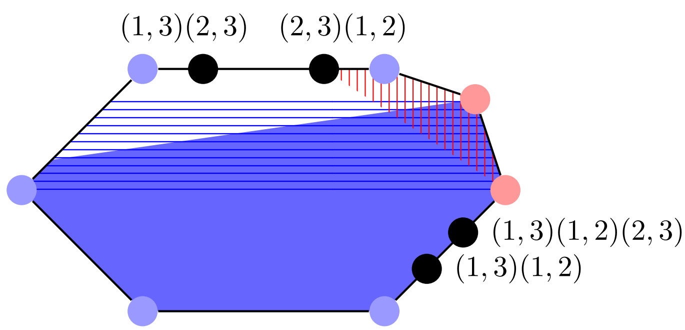


Figure 12: Rough representation of the set $C_d^{ETO}(p)$ for $d = (d_0, d_1, d_2)$ with $d_1 > d_1 > d_2$ and $p = (a, b, c)$ fulfilling (34). We include labels for the black points which, although not plotted in Figure 11 since they are not extremal, are useful when proving Lemma 10. Moreover, we include the areas which are mentioned in (A)-(C) in that lemma. Note that we use the notation in Figure 7.

By the recursive relation in (39), we can conclude from (A)-(C) that $L(p) \notin B_{\ell+1}$. By induction, (38) holds. This concludes the proof. □

By Lemma 10, we have that, if $q \in E(L)$ for some $E(L) \in B_\ell$ and $\ell \geqslant 0$, then $q \in \text{ext}(E(L))$, where $\text{ext}(\cdot)$ denotes the set of extreme points. This contradicts the fact that $0 < \lambda < 1$ in (37). In particular, (37) implies that $q \in L_0 \backslash \text{ext}(L_0)$, where $L_0 \in B_\ell$ for some $\ell \geqslant 0$. This yields the desired contradiction and concludes the proof.

M Proof of Proposition 4

By Lemma 1, it suffices to assume that $d = d^{\downarrow}$ throughout the proof.

We only prove necessity since sufficiency follows from Theorem 10. Since d is quasi-uniform, then we either have $d = (d_0, d_0, d_1)$ with $d_0 \geqslant d_1$ or $d = (d_0, d_1, d_1)$ with $d_0 > d_1$. Since the statement is true for the first instance (by putting together Propositions 2 and 3), we fix $d = (d_0, d_1, d_1)$ with $d_0 > d_1$ and $\gamma = d_1/d_0$ in the following. We will show that, for any $p = (a, b, c) \in \mathbb{P}_\Omega$ (which we can assume fulfills $b \geqslant c$ w.l.o.g.), we have $\mathcal{C}_d^{ETO}(p) \subseteq \mathcal{C}_d^{WETO}(p)$. In order to do so, we distinguish the following cases (the instances where some equality holds follow easily from these):

(A) $\gamma a > b > c$. In this case, $\mathcal{C}_d^{ETO}(p)$ is (roughly) given by Figure 13, with $q \in C_d^{ETO}(p)$ being achievable by a sequence $T^d(2,3)T^d(1,2)$ if it lies to the left of the dashed line and by $T^d(2,3)T^d(1,3)P^d(1,2)$ if it lies to the right.

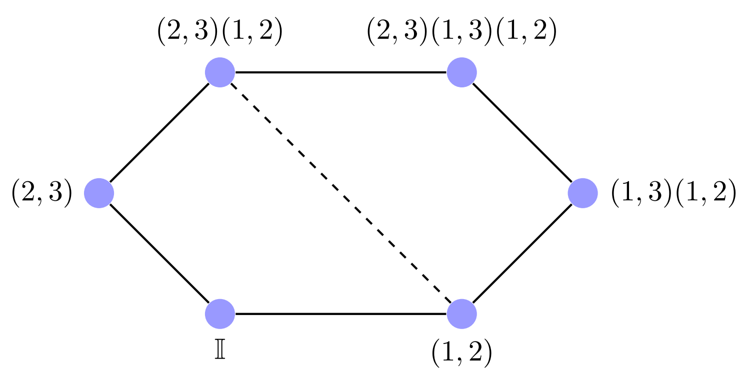


Figure 13: Rough representation of $\mathcal{C}_d^{ETO}(p)$ for $d = (d_0, d_1, d_1)$ with $0 < d_1 < d_0$ and $p = (a, b, c)$ with $\gamma a \geqslant b \geqslant c$. Note that we use the notation in Figure 7.

(B) $b > \gamma a > c$. In this case, $\mathcal{C}_d^{ETO}(p)$ is (roughly) given by Figure 14, with $q \in C_d^{ETO}(p)$ being achievable by a sequence $T^d(2,3)T^d(1,3)$ if it lies to the right of the dashed line and by $T^d(2,3)T^d(1,2)$ if it lies to the left.

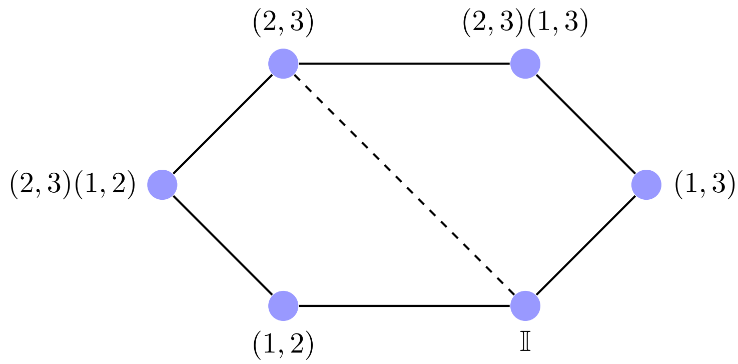


Figure 14: Rough representation of $\mathcal{C}_d^{ETO}(p)$ for $d = (d_0, d_1, d_1)$ with $0 < d_1 < d_0$ and $p = (a, b, c)$ with $b \geqslant \gamma a \geqslant c$. Note that we use the notation in Figure 7.

(C) $b > c > \gamma a$. In this case, $\mathcal{C}_d^{ETO}(p)$ is (roughly) given by Figure 15, with $q \in C_d^{ETO}(p)$ being achievable by a sequence $T^d(2,3)T^d(1,3)$ if it lies to the right of the dashed line and by $T^d(2,3)T^d(1,2)P^d(1,3)$ if it lies to the left.

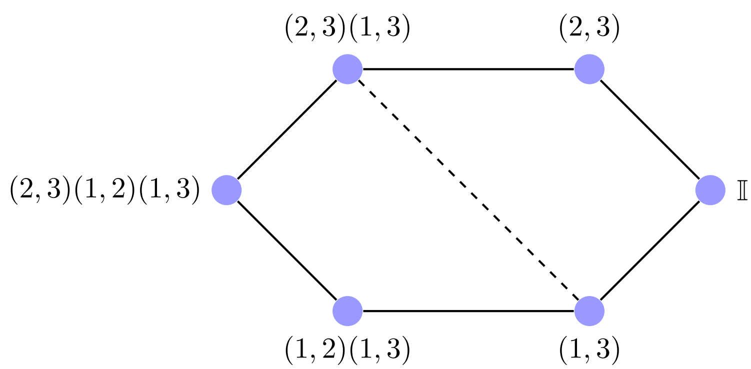


Figure 15: Rough representation of $\mathcal{C}_d^{ETO}(p)$ for $d = (d_0, d_1, d_1)$ with $0 < d_1 < d_0$ and $p = (a, b, c)$ with $b \geqslant c \geqslant \gamma a$. Note that we use the notation in Figure 7.

This concludes the proof.

N Proof of Theorem 11

By Lemma 1, it suffices to assume that $d = d^{\downarrow}$ throughout the proof. Furthermore, we take $d_0 := 1$ and $d_{|\Omega|+1} := 0$.

Necessity is straightforward by Lemma 2. To prove sufficiency, we argue by contrapositive. Hence, we take $|\Omega| \geqslant 3$ and argue on the number of jumps in d.

In particular, we consider the following cases:

(A) There exist some $1 \leqslant \alpha < |\Omega|$ and $2 \leqslant k \leqslant |\Omega| - \alpha$ such that $d_{\alpha-1} > d_\alpha = d_{\alpha+1} = \cdots = d_{\alpha+k} > d_{\alpha+(k+1)}$. In this scenario, following the proof of Theorem 9 for $|\Omega| \geqslant 4$, we have that

$$M := M_0 \bigoplus \mathbb{I}_{\backslash(\alpha,\ldots,\alpha+k)},$$

with $M_0 \in \mathcal{M}_{k+1,k+1}(\mathbb{R})$ a doubly stochastic matrix such that $(M_0)_{i,j} > 0$ if $i \neq j$ and $(M_0)_{i,j} = 0$ if $i = j$, does not belong to the WETO polytope. (To show this, aside from Theorem 9, one can consider the number of ones along the main diagonal to conclude that, whenever we have some $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ acting as the identity on some subset $\Omega' \subseteq \Omega$ and the entries of d in Ω' differ from those in $\Omega\backslash\Omega'$, then we can restrict ourselves to the weak elementary thermal operations on $\Omega\backslash\Omega'$ to prove this result. This is also useful for proving the other cases.) However, by Theorem 2, M does belong to the ETO polytope.

(B) There exist some $1 \leqslant \alpha \leqslant |\Omega| - 2$ such that $d_{\alpha-1} > d_\alpha = d_{\alpha+1} > d_{\alpha+2} > d_{\alpha+3}$ or $d_{\alpha-1} > d_\alpha > d_{\alpha+1} = d_{\alpha+2} > d_{\alpha+3}$. In this scenario, the first instance follows by Theorems 9 (in the case $|\Omega| = 3$) and 1. For the second instance, we take $0 < \phi_0 < 1$ (sometimes a more restrictive choice of ϕ_0 can make the argument easier) and consider

$$M := (1-\phi_0)\, P^d(\alpha, \alpha{+}1) \bigoplus \mathbb{I}_{\backslash(\alpha,\alpha+1)} + \phi_0 P^d(\alpha, \alpha{+}2) \bigoplus \mathbb{I}_{\backslash(\alpha,\alpha+2)}. \quad (40)$$

It is clear, by definition, that M belongs to the ETO polytope. To conclude this case, we can profit from (27) and, analogously to (28), from the fact that, whenever $d_{\alpha+1} = d_{\alpha+2}$ and $0 \leqslant \lambda \leqslant 1$, then

$$T^d_\lambda(\alpha, \alpha+2) = P^d(\alpha+1, \alpha+2) T^d_\lambda(\alpha, \alpha+1) P^d(\alpha+1, \alpha+2).$$

This allows us to conclude that M is part of the WETO polytope if and only if there exist some $\ell, m \in \{0, 1\}$ and $N \geqslant 0$ such that

$$M = \left(T^d_{\lambda_A}(\alpha+1, \alpha+2)\right)^\ell \left(\prod_{k=1}^{N} T^d_{\lambda_k}(\alpha, \alpha+1) T^d_{\beta_k}(\alpha+1, \alpha+2)\right) \times \left(T^d_{\lambda_B}(\alpha, \alpha+1)\right)^m,$$

where $0 \leqslant \lambda_A, \lambda_B, \lambda_k, \beta_k \leqslant 1$ for $1 \leqslant k \leqslant N$. It is then straightforward to check that the last equation is never satisfied. (For instance, one can argue by direct calculation using the zeros in M.)

(C) There exist some $1 \leqslant \alpha \leqslant |\Omega| - 2$ such that $d_{\alpha-1} > d_\alpha > d_{\alpha+1} > d_{\alpha+2} > d_{\alpha+3}$. In this case, we can also consider M as in (40) and argue similarly as in (b) for most of the instances. (This is the case since in (b) we usually argue on the number of zeros and, if certain components of some product of T^d-transforms are non-zero when $d_{\alpha+1} = d_{\alpha+2}$, then they will also be non-zero whenever we consider the analogous product $d_{\alpha+1} > d_{\alpha+2}$.) For the rest of the instances, one can directly check that the result holds.

(D) $|\Omega| = 2m$ for some $m \geqslant 2$ and $d_{2p-2} > d_{2p-1} = d_{2p} > d_{2p+1}$ for $1 \leqslant p \leqslant m$. This case can be reduced to case (b) since it only differs in the introduction of copies of a permutation matrix to (29). In particular, we take $1 \leqslant \alpha < m$ and

$$M := M_0 \bigoplus \mathbb{I}_{\setminus(2\alpha-1,2\alpha,2\alpha+1)},$$

with $M_0 \in \mathcal{M}_{3,3}(\mathbb{R})$ defined as in (26). Note that M belongs to the ETO polytope by Theorem 8. To conclude, assume it belongs to the WETO polytope. In this case, it is easy to see that M must have a decomposition like (29) with the possible inclusion of permutations $P := P^d(2\alpha{+}1, 2\alpha{+}2)$ along the sequence. We conclude showing that any such sequence reduces to (29) and, hence, M is not in the WETO polytope by Theorem 9. Consider, thus, the first P appearing in the sequence. Since P commutes with $T^d_\lambda(2\alpha-1, 2\alpha)$, then it is followed either by some $T^d_\lambda(2\alpha, 2\alpha+1)$ or it is the leftmost matrix in the product. In any case, we can consider separately the case where the matrices to the right of P acted trivially and non-trivially on the $2\alpha + 1$ component and, using that $M_{2\alpha+2,2\alpha+2} = 1$, conclude that either the sequence does not yield M or it is equivalent to a sequence with a fewer number of P matrices. Following this argument recursively we reach a sequence like (29) and obtain the desired conclusion.

This concludes the proof.

O Proof of Corollary 3

(a) We can argue by contrapositive, assuming we have a non quasi-uniform distribution d such that $d_1 > d_2 > d_3$ w.l.o.g. and noting that, if we take $(p, 0), (q, 0) \in \mathbb{P}_\Omega$ as in Theorem 10, then

$$(q, 0) \in (\mathcal{L}(p), 0) := \left[\left(P^d(2, 3)P^d(1, 2)P^d(1, 3)p, 0\right), \left(P^d(1, 2)p, 0\right)\right]$$

with $\mathrm{ext}((\mathcal{L}(p), 0)) \subseteq \mathcal{C}^{WETO}_d((p, 0))$ and $(q, 0) \notin \mathcal{C}^{WETO}_d((p, 0))$.

(b) Necessity follows by (b), while sufficiency follows by Proposition 4 since $\mathcal{C}^{WETO}_d(p) = \mathcal{C}^{ETO}_d(p)$ and the latter is convex by definition.

(c) By Lemma 1, it suffices to assume that $d = d^{\downarrow}$. Necessity follows from Lemma 2 since the TO polytope is convex and, in this instance, equal to the WETO polytope. To prove sufficiency, note that, provided $|\Omega| > 2$, there exists some $M \in \mathcal{M}_{|\Omega|,|\Omega|}(\mathbb{R})$ that belongs to the ETO polytope and not to the WETO polytope by Theorem 11. Moreover, by definition,

$$M = \sum_{k=1}^{k_0} \lambda_k \prod_{\ell=1}^{\ell_0} P^d(i_{k,\ell}, j_{k,\ell}),$$

where $\sum_{k=1}^{k_0} \lambda_k = 1$ and $\lambda_k \geqslant 0$ and $1 \leqslant i_{k,\ell} < j_{k,\ell} \leqslant |\Omega|$ for $1 \leqslant k \leqslant k_0$ and $1 \leqslant \ell \leqslant \ell_0$. Lastly, since $\prod_{\ell=1}^{\ell_0} P^d(i_{k,\ell}, j_{k,\ell})$ belongs to it for $1 \leqslant k \leqslant$

k_0, the WETO polytope does not contain a convex combination of points in it. Hence, it is not convex.

Alternatively, provided d is not quasi-uniform, we can show this using (a). This is the case since, for any such d, we can use (a) to find some $p \in \mathbb{P}_\Omega$ such that $\mathcal{C}_d^{WETO}(p)$ is not convex, which is impossible provided $\mathcal{P}_{\text{ETO}}(d)$ is convex. (If $r, q \in \mathcal{C}_d^{WETO}(p)$ and $\mathcal{P}_{\text{ETO}}(d)$ is convex, then $(1-\lambda)q + \lambda r = (1-\lambda)M_0 p + \lambda M_1 p = M_2 p \in \mathcal{C}_d^{WETO}(p)$ for all $0 \leqslant \lambda \leqslant 1$, where $M_0, M_1, M_2 \in \mathcal{P}_{\text{ETO}}(d)$.)