On the physical implementation of logical transformations: Generalized \textit{L}-machines

Neal G. Anderson *

Department of Electrical and Computer Engineering, University of Massachusetts Amherst, Amherst, MA 01003-9292, USA

\section*{A R T I C L E   I N F O}

Article history:
Received 30 April 2009
Received in revised form 30 November 2009
Accepted 11 June 2010
Communicated by M. Hirvensalo

Keywords:
Implementation
\textit{L}-machine
Landauer’s Principle
Physical information theory
Physics of computation
Realism about computation
Representation

\section*{A B S T R A C T}

Any account of computation in a physical system, whether an artificial computing device or a natural system considered from a computational point of view, invokes some notion of the relationship between the abstract-logical and concrete-physical aspects of computation. In a recent paper, James Ladyman explored this relationship using a “hybrid physical–logical entity” – the \textit{L}-machine – and the general account of computation that it supports [J. Ladyman, What does it mean to say that a physical system implements a computation?, Theoretical Computer Science 410 (2009) 376–383]. The underlying \textit{L}-machine of Ladyman’s analysis is, however, classical and highly idealized, and cannot capture essential aspects of computation in important classes of physical systems (e.g. emerging nanocomputing devices) where logical states do not have faithful physical representations and where noise and quantum effects prevail. In this work we generalize the \textit{L}-machine to allow for generally unfaithful and noisy implementations of classical logical transformations in quantum mechanical systems. We provide a formal definition and physical-information-theoretic characterization of generalized quantum \textit{L}-machines (QLMs), identify important classes of QLMs, and introduce new efficacy measures that quantify the faithfulness and fidelity with which logical transformations are implemented by these machines. Two fundamental issues emphasized by Ladyman – realism about computation and the connection between logical and physical irreversibility – are reconsidered within the more comprehensive account of computation that follows from our generalization of the \textit{L}-machine.

© 2010 Elsevier B.V. All rights reserved.

\section*{1. Introduction}

In a recent paper in this journal [1], James Ladyman considered the question of what it means to say that a physical system implements a computation. He discussed two fundamental issues of contemporary interest that require a clear answer to this question for their proper consideration – realism about computation and the relationship between logical and thermodynamic irreversibility – and analyzed these issues using an account of implementation that is embodied in the notion of an \textit{L}-machine. The \textit{L}-machine, introduced in earlier work by Ladyman, Presnell, Short and Groisman [2] (hereafter LPSG), is a “hybrid physical–logical entity that combines a physical device, a specification of which physical states of that device correspond to various logical states, and an evolution of that device which corresponds to the logical transformation \textit{L}” [1].

In defining their \textit{L}-machine, Ladyman and co-workers [1–3] properly associate logical transformations with families of physical processes. Also, they rightly emphasize the significance of faithful representation of logical states in distinguishable physical states. However, because LPSG require faithful representation in distinguishable states, and because they consider...
only classical devices, their $L$-machine can describe only idealized noiseless implementations of logical operations in classical physical systems. An account of implementation based on LPSGs $L$-machine will thus be unable to capture essential features of computing in natural and man-made systems in which noise and quantum effects prevail – including emerging nanocomputing devices – with implications for the generality of foundational analyses that are based on such an account.

In this work, we generalize the $L$-machine to allow for generally unfaithful and noisy physical implementations of classical logical transformations in generally quantum mechanical systems. This requires that we go beyond the question of whether a particular machine implements a specified computation – a question that is predicated on an “all or nothing” view of computation – to ask more generally how well a particular machine implements a specified computation when used in a particular way, and provide meaningful quantitative answers to this question. By allowing for “shades of grey” in computational efficacy, captured by statistical measures in our physical-information-theoretic approach, the notion of an $L$-machine is extended to accommodate a much wider variety of scenarios involving computation in physical systems and thus to provide a more comprehensive basis for the exploration of fundamental issues.

The paper is organized as follows: We begin Section 2 by restating LPSG’s definition of their classical $L$-machine, which we dub the “ideal classical $L$-machine” (ICLM), and discussing its limitations. We then build to our definition of the fully generalized quantum $L$-machine (QLM), discussing the relevant generalizations and introducing associated formal devices along the way, and identify important classes of QLMs. In Section 3, we reconsider the “qualified” computational realism discussed by Ladyman in connection with ideal classical machines, and identify the class of noiseless QLMs for which such a realist conception is tenable. In Section 4, we provide a full physical-information-theoretic characterization of computation in the quantum $L$-machine that draws from and generalizes Winograd and Cowan’s treatment of the noisy classical computation channel [4] and their observations regarding information loss in computation. Here we introduce new quantitative efficacy measures for representational faithfulness and computational fidelity, discuss their interpretation, and consider their connection to information loss in $L$-machines. Next, in Section 5, we consider the connection between logical irreversibility, physical irreversibility and information loss in the implementation of logical transformations by general quantum $L$-machines, and obtain information-theoretic lower bounds on entropy generation and energy flow that are expressed in terms of our efficacy measures. Finally, in Section 6, we summarize our work, discuss limitations of $L$-machines (LPSGs and ours), and briefly consider requirements for more comprehensive physical descriptions of realistic information processing scenarios.

We emphasize that, although the machines we consider in this work are generally quantum mechanical and are analyzed with the help of some ideas familiar from quantum information theory and quantum computation (e.g. [5]), our focus is decidedly on the physical implementation of classical logical transformations\(^1\) in such machines. Our specific concern is the relationship between the structure of the initial to final physical state transformations that characterize a given quantum $L$-machine and the structure of the classical logical transformation that the machine is taken to implement, considered independent of the machine’s internal architecture and workings.

2. Generalizing the $L$-machine

2.1. Ideal classical $L$-machines

We begin with LPSGs definition of an $L$-machine [1–3] – hereafter the “ideal classical $L$-machine” (or ICLM) – which we paraphrase with minor modifications and notational changes for consistency with this work:

**Definition 1.** An $M$-input, $N$-output ideal classical $L$-machine (ICLM) is an ordered set

\[ \{\mathcal{D}, \{D_{i}^{(in)}\}, \{D_{j}^{(out)}\}, \lambda_{i}\} \]

consisting of

- A physical device $\mathcal{D}$, situated in a heat bath at temperature $T$.
- A set \( \{D_{i}^{(in)}\} \) of $M$ macroscopic input states of the device, which are distinct thermodynamic states of the system (i.e. no microstate is a component of more than one thermodynamic state). $D_{i}^{(in)}$ is the representative physical state of the logical input state $x_{i}$.
- A set \( \{D_{j}^{(out)}\} \) of distinct thermodynamic output states of the device. $D_{j}^{(out)}$ is the representative physical state of the logical output state $y_{j}$.
- A time-evolution operator $\lambda_{i}$ for such a device, such that

\[ \lambda_{i}(D_{i}^{(in)}) = D_{j}^{(out)} \quad \forall i \in \{i\} = \{i|L(x_{i}) = y_{j}\} \]

where $L$ is a logical transformation that maps $M$ logical input states $x_{i} \in \{x_{i}\}$ into $N$ logical output states $y_{j} \in \{y_{j}\}$ via $x_{i} \rightarrow L(x_{i}) = y_{j}$.

\(^1\) By “classical logical transformation” we mean any mapping $L$ from a discrete set $\{x_{i}\}$ of “input” elements to a discrete set $\{y_{j}\}$ of “output” elements. For consistency with Ladyman’s terminology, we will generically refer to elements $x_{i} \in \{x_{i}\}$ and $y_{j} \in \{y_{j}\}$ of the domain and range of $L$ as “logical input states” and “logical output states”, respectively.
Paraphrasing Ladyman from Ref. [1], with the same notational changes, such a machine

“...physically implements L in the following sense. If D is prepared in the input state $D_i^{(in)}$ corresponding to the logical input state $x_i \in \{x_i\}$, and is evolved using $A_L$, it will be left in the output state $D_j^{(out)}$ corresponding to the logical output state $y_j = L(x_i) \in \{y_j\}$.”

In ICLM, it is indeed easy to see how a logical transformation $L$ is mirrored in a family of physical processes. The relationship is as natural and transparent as it is precisely because the physical implementation of $L$ supported by ICLM is faithful and noiseless, i.e. unique final states $D_j^{(out)}$ can be associated with specified disjunctions of initial states and the $D_j^{(out)}$ are distinguishable from one another. Evolution operators $A_L$ that do not meet these requirements simply are not admitted in ICLM.

These requirements are, however, highly idealized. Ladyman recognizes that they may be violated even in classical machines, stating that “all that is required in practice is that the appropriate physical [output] state be arrived at with very high probability” [1] and considering “multi-L-machines” that support unfaithful implementations [1,2]. However, the consequences of non-ideal evolutions for the definition of an $L$-machine or for the fundamental issues Ladyman and co-workers have viewed through the prism of ICLM have not been explored. These consequences warrant consideration: If the device output states are unfaithful representations of the logical outputs or if they cannot be perfectly distinguished from one another, then the essential nature of the relationship between the abstract-logical and the concrete-physical is fundamentally altered and can only be captured by generalizing the definition of an $L$-machine. We construct such a generalization in the remainder of this section, and do so in a fully quantum mechanical context.

2.2. Quantum machines and referents

Our objective is to construct a generalized quantum mechanical definition of an $L$-machine that retains the spirit of ICLM but accommodates evolution operators that are not required to evolve distinguishable input states into distinguishable, representationally faithful output states. We begin building toward such a definition by first defining a quantum machine tailored to these purposes:

**Definition 2.** An $M$-input quantum machine (QM) is an ordered set

$$\{D, \{D_i\}, A\}$$

consisting of

- A quantum system (or device) $D$, situated in an environment $E$.
- A set $\{D_i\}$ of $M$ distinguishable initial physical states of $D$.
- A time-evolution operator $A$ for $D$, which maps initial physical states $\hat{D}_i \in \{D_i\}$ of $D$ into final physical states as

$$\hat{D}_i \rightarrow A(\hat{D}_i).$$

QM does not yet involve association of logical states with physical states or logical transformation(s) with the machine. Such associations will, of course, be part of our definition of a quantum $L$-machine, but they will necessarily appear in a modified form to accommodate relaxation of the conditions on the evolution operator that characterize ICLM. The most significant modifications stem from attempts to generalize the output states, for reasons we now discuss.

Consider a partition $\{\{\hat{D}_j\}_i\} = \{\{\hat{D}_i\}_0, \{\hat{D}_i\}_1, \ldots, \{\hat{D}_i\}_{N-1}\}$ of the $M$ initial device states into $N$ subsets $\hat{D}_i = \{\hat{D}_i|_i \in \{i\}_j\}$, where $\{i\}_j$ is the set of indices labeling states that belong to the $j$-th subset. A similar partition is used in ICLM, where each of $N$ final physical device state $D_j^{(out)}$ is uniquely associated with a subset $\{D_i^{(in)}\}_i = \{D_i^{(in)}|i \in \{i\}_j\}$ of initial device states (and a corresponding subset of logical input states). Thus, to generalize ICLM, one would first seek a final quantum machine state that is uniquely associated with each set $\{\hat{D}_i\}_j$ of initial device states. The evolved device states $A(\hat{D}_i) \in \{A(\hat{D}_i)|_i \in \{i\}_j\}$ may, however, all differ from one another for a general $A$, so a unique final machine state generally cannot be associated with the set $\{\hat{D}_i\}_j$ of initial states. The best next thing would be to associate a distinct

---

2 Note that we have relaxed the requirement that the device’s environment is a thermal bath at temperature $T$, and thus relax the requirement that the device states are “thermodynamic states” as well.

3 Here a physical state $D_i$ of $D$ is denoted by its associated quantum mechanical density operator $\hat{D}_i$, indicated by the carat, which is a positive operator with unit trace on the Hilbert (state) space of $D$. Two states $\hat{D}_i$ and $\hat{D}_j$ are distinguishable if and only if their density operators have support on mutually orthogonal subspaces of this state space (so $\text{Tr}[\hat{D}_i\hat{D}_j] = 0$), as this orthogonality ensures the existence of measurements that could distinguish the two states without ambiguity [6].

4 Specifically, $A$ is a completely positive trace-preserving map that transforms density operators on the device’s state space into density operators on the same space (see, for example, [5,6]).

5 One can define (generally nonorthogonal) statistical states representing sets of evolved initial states, provided that the a priori probabilities of the initial states are known (as we have done in [7]), but here we follow LPSG in defining the machine independent of an assumed operating context.
support subspace \( \mathcal{H}^D_j \) of the device's state space with each subset \( \{ \Lambda(\hat{D}_i) \} \) of final device states, where the \( j \)-th support subspace is the union

\[
\mathcal{H}^D_j = \bigcup_{i \in [l_j]} \mathcal{H}^D_i
\]  

of the support subspaces \( \mathcal{H}^D_i \) for the evolved initial states \( \Lambda(\hat{D}_i) \in \{ \Lambda(\hat{D}_i) \} \). The identity operator for the union \( \mathcal{H}^D_j \) is the “shell projector”

\[
\hat{\Pi}^D_j = \hat{D} - \prod_{i \in [l_j]} \left( \hat{D} - \hat{\pi}^D_i \right)
\]  

where the projector \( \hat{\pi}^D_i \) is the identity for \( \mathcal{H}^D_i \) and \( \hat{D} \) is the identity for the full space \( \mathcal{H}^D = \bigcup_{i=1}^M \mathcal{H}^D_i = \bigcup_{j=1}^N \mathcal{H}^D_j \). Yet even this falls short of what we are after, since the support subspaces \( \mathcal{H}^D_j \) need not be disjoint subspaces for a general \( \Lambda \).

Thus, without conditions on \( \Lambda \) analogous to those in ICLM, one cannot associate unique final states of \( \mathcal{D} \) – or even disjoint subspaces of the state space of \( \mathcal{D} \) – with subsets of initial device states. In building from QM to a definition of a fully generalized quantum \( \mathcal{L} \)-machine that is similar in structure to ICLM, something additional will be required if we are to identify appropriate surrogates for ICLM’s output states \( \hat{D}_{\text{out}} \).

We identify and define appropriate surrogates for these states with the help of a referent system, which will be an integral part of our generalized definition an \( \mathcal{L} \)-machine and will play a key role in its analysis. To demonstrate use of a referent system for resolving the problem at hand, and to provide a stepping stone to definition the quantum \( \mathcal{L} \)-machine, we define the following:

**Definition 3.** An \( M \)-input quantum machine with referent (QMR) is an ordered set

\[
\{ \mathcal{D}, \mathcal{R}, \{ \hat{D}_i \}, \{ \hat{\pi}^R_i \}, \{ \hat{\pi}^{RD}_i \}, \Lambda \}
\]

consisting of

- A quantum system (or device) \( \mathcal{D} \), situated in an environment \( \mathcal{E} \).
- A referent system \( \mathcal{R} \).
- A set \( \{ \hat{D}_i \} \) of \( M \) distinguishable initial physical states of \( \mathcal{D} \).
- A set \( \{ \hat{\pi}^R_i \} \) of \( M \) distinguishable pure states of \( \mathcal{R} \).
- A set \( \{ \hat{\pi}^{RD}_i \} \) of \( M \) joint states of \( \mathcal{RD} \), constructed from a one-to-one pairing of \( \hat{\pi}^R_i \in \{ \hat{\pi}^R_i \} \) and \( \hat{D}_i \in \{ \hat{D}_i \} \) into product states \( \hat{\pi}^{RD}_i = \hat{\pi}^R_i \otimes \hat{D}_i \).
- A time-evolution operator \( \hat{\Lambda} \) for \( \mathcal{RD} \), which maps initial physical states \( \hat{\pi}^{RD}_i \in \{ \hat{\pi}^{RD}_i \} \) of \( \mathcal{RD} \) into final physical states as

\[
\hat{\Lambda}(\hat{\pi}^R_i \otimes \hat{D}_i) = \hat{\pi}^R_i \otimes \Lambda(\hat{D}_i).
\]

QMR is just QM together with a stable referent system \( \mathcal{R} \) and a prescription for the joint preparation of the device and referent. By design, the specified joint preparations ensure that the referent system holds an unambiguous and enduring record of the initial state of \( \mathcal{D} \): When the device is initially prepared in the \( i \)-th initial state \( \hat{D}_i \), the referent is initially prepared in the state \( \hat{\pi}^R_i \) and remains in this state as the device state evolves into \( \Lambda(\hat{D}_i) \). The referent system fulfills this “record keeping” role without affecting the behavior of \( \mathcal{DE} \) in any way, as states of the device \( \mathcal{D} \) are evolved locally as \( \hat{D}_i \to \Lambda(\hat{D}_i) \) in QMR exactly as they are in QM. \(^8\)

While benign as far as the device dynamics are concerned, the referent system provides the necessary ingredient for our general quantum \( \mathcal{L} \)-machine. To see this, consider the support subspace \( \mathcal{H}^{RD}_j \) for a set \( \{ \hat{\pi}^R_i \otimes \Lambda(\hat{D}_i) \} \) of final QMR machine states, i.e. the union

\[
\mathcal{H}^{RD}_j = \bigcup_{i \in [l_j]} \mathcal{H}^{RD}_i
\]  

\(^6\) This terminology follows use of the term “core projector” for the identity of the intersection of a set of subspaces in the literature (e.g. [8]).

\(^7\) Note that \( \mathcal{H}^D \) – the support space for the union of evolved initial states – is in general a subspace of the physical device’s Hilbert space.

\(^8\) Note that, while we have taken the referent states \( \hat{\pi}^R \) to be pure and time independent, they can fulfill their record keeping role even if they are mixed and time dependent provided that they are distinguishable and their time evolution is locally unitary. We choose the special case of stationary pure referent states throughout this work for simplicity and without loss of generality: The results and conclusions of this paper are unaffected by this simplification.
where $\mathcal{H}^{RD}_i$ is the support of the $i$-th evolved input state $r^R_i \otimes \Lambda(\hat{D}_i)$ of the device-referent composite $RD$. The identity for this space is

$$\hat{1}^{RD}_j = \sum_{i \in [I]} \hat{1}^{RD}_i$$

(10)

where $\hat{1}^{RD}_i = r^R_i \otimes \hat{1}^D_i$ is the identity for $\mathcal{H}^{RD}_i$. Because of the mutual orthogonality of the referent states, the subspaces $\mathcal{H}^{RD}_j$ are disjoint subspaces – i.e. the associated projectors do satisfy the mutual orthogonality condition

$$\hat{1}^{RD}_j \hat{1}^{RD}_j = \delta_{jj} \hat{1}^{RD}_j$$

(11)

– even when the shell projectors $\hat{1}^{D}_j$ on the device space $\mathcal{H}^D$ are not mutually orthogonal.

Thus, by incorporating a referent system into a quantum machine with a general device evolution operator $\Lambda$, we can identify disjoint subspaces that are uniquely associated with subsets of evolved initial states. The projectors associated with such subspaces will function as surrogates for physical output states in our formal definition of a quantum $L$-machine.

### 2.3. Quantum $L$-machines

Recall that our objective is to describe, through a generalized notion of an $L$-machine, how a logical mapping

$$x_i \rightarrow L(x_i) = y_j \ \forall i \in [I]$$

(12)

from $M$ logical input states $x_i \in \{x_i\}$ to $N$ logical output states $y_j \in \{y_j\}$ is mirrored in a family of quantum-physical processes that generally do not evolve $M$ distinguishable physical input states into $N$ distinguishable physical output states as in ICLM. We showed above how introduction of a referent system – an immutable “identification tag” appended to each initial device state - is useful in this more general context, where one cannot generally associate distinguishable final device states or even disjoint device state (sub)spaces with subsets of initial device states. We formally include such a referent system – called an $L$-referent – within our definition of a quantum $L$-machine, even though its primary role is as a bookkeeping device. The $L$-referent is defined as follows:

**Definition 4.** An $L$-Referent $R_L$ associated with an $M$-input, $N$-output logical transformation $L$ is an ordered set

$$\left\{ R_L, \{\hat{r}^R_{in}\}, \{\hat{r}^R_{out}\}, \{\hat{r}^R_L\}\right\}$$

(13)

consisting of

- A bipartite quantum system $R_L = R_{in}\otimes R_{out}$.
- A set $\{\hat{r}^R_{in}\}$ of $M$ distinguishable pure states of $R_{in}$.
- A set $\{\hat{r}^R_{out}\}$ of $N$ distinguishable pure states of $R_{out}$.
- A set $\{\hat{r}^R_L\}$ of $M$ product states

$$\hat{r}^R_L = \hat{r}^R_{in} \otimes \hat{r}^R_{out} \ \forall i \in [I] = \{i\} \mid L(x_i) = y_j$$

(14)

where $L$, where $L$ is a logical transformation that maps $M$ logical input states $x_i \in \{x_i\}$ into $N$ logical output states $y_j \in \{y_j\}$ via $x_i \rightarrow L(x_i) = y_j$.

The set $\{\hat{r}^R_L\}$ of $M$ physical $L$-referent states embodies the full structure of the abstract-logical transformation $L$. By substituting an $L$-referent for an abstract-logical transformation $L$ in the definition of an $L$-machine, we can thus build the structure of $L$ into the definition in such a way that the abstract-logical states have concrete-physical instantiations. This will give the notion of “representation” of a logical state in a device $D$ an explicit meaning in a quantum $L$-machine that it does not have in LPSG’s classical $L$-machine, where logical states are purely abstract, and will play an important role in our information-theoretic characterization of the quantum $L$-machine.

---

9 The general form for such a projector is

$$\hat{1}^{RD}_j = \hat{1}^{RD}_j - \sum_{i \in [I]} \left( \hat{1}^{RD}_j - \hat{1}^{RD}_i \right)$$

where $\hat{1}^{RD}_j$ is the identity for $\mathcal{H}^{RD}_j$. Since the $\hat{1}^{RD}_i$ are mutually orthogonal here ($\hat{1}^{RD}_i \hat{1}^{RD}_j = \delta_{ij} \hat{1}^{RD}_j$), the product term on the right-hand side simplifies to $\hat{1}^{RD}_j - \sum_{i \in [I]} \hat{1}^{RD}_i$ and the specialized form Eq. (10) results.

10 The referent states $\hat{r}^R_{in}$ and $\hat{r}^R_{out}$ instantiate the logical inputs $x_i$ and $y_j$, respectively.
With this, we finally define the quantum L machine:

**Definition 5.** An M-input, N-output quantum L-machine (QLM) is an ordered set

\[
\{ D, R_L, \{ \hat{D}_i^{(in)} \}, \{ \hat{\rho}_j^{R_LD} \}, \Lambda, \{ \hat{\Pi}_j^{R_LD} \} \}
\]

(15)

consisting of

- A quantum system (or device) \( D \), situated in an environment \( \mathcal{E} \).
- An M-input, N-output L-referent \( R_L \).
- A set \( \{ \hat{D}_i^{(in)} \} \) of M distinguishable physical input states of \( D \).
- A set \( \{ \hat{\rho}_j^{R_LD} \} \) of M joint initial states of \( R_L, D \), constructed from a one-to-one pairing of \( \hat{r}_j^{R_L} \) and \( \hat{D}_i^{(in)} \) into product states \( \hat{\rho}_j^{R_LD} = \hat{r}_j^{R_L} \otimes \hat{D}_i^{(in)} \).
- A time-evolution operator \( \hat{\Lambda} \) for \( R_L, D \), which maps initial states \( \hat{\rho}_j^{R_LD} \) of \( R_L, D \) into final states as

\[
\hat{\Lambda}(\hat{r}_j^{R_L} \otimes \hat{D}_i^{(in)}) = \hat{r}_j^{R_L} \otimes \Lambda(\hat{D}_i^{(in)}).
\]

(16)

- A set \( \{ \hat{\Pi}_j^{R_LD} \} \) of N orthogonal projection operators

\[
\hat{\Pi}_j^{R_LD} = \sum_{i \in \{ j \}} \hat{r}_i^{R_L} \otimes \hat{r}_j^{R_LD}
\]

(17)

on \( \mathcal{H}^{R_L} \otimes \mathcal{H}^{D} \), with

\[
\hat{r}_j^{R_LD} = \hat{r}_j^{R_L} \otimes \hat{r}_j^{R_LD}
\]

(18)

where \( \hat{r}_j^{R_L} \) is the identity for the support of \( \Lambda(\hat{D}_i^{(in)}) \). \( \hat{\Pi}_j^{R_LD} \) is the identity for the support subspace associated with the j-th subset \( \{ \hat{r}_j^{R_L} \otimes \Lambda(\hat{D}_i^{(in)}) \}_{i \in \{ j \}} \) of evolved initial states of \( R_L, D \).

This is the desired generalization of the definition of an L-machine. Note that, as advertised, the orthogonal projection operators \( \hat{\Pi}_j^{R_LD} \) take the place of ICLMs output states \( D_{\text{out}}^{(i)} \) in this definition, and that the “presence” of the L-referent has absolutely no affect on the dynamics of the QLM’s device and environment (which are just those of the underlying QM).

Several remarks are in order regarding this definition: First, the M input states \( \{ \hat{D}_i^{(in)} \} \) of an M-input, N-output QLM are just the M initial states \( \{ \hat{D}_i \} \) of the “underlying” M-input quantum machine QM on which the QLM has been defined: The change in notation reflects the role these states play as natural representations in \( D \) of the logical input states, or, equivalently, the input referent states.11 This representational role is underwritten by well-defined physical relationship between the states of \( R_L \) and \( D \) in QLM, codified in the set of joint states \( \{ \hat{\rho}_j^{R_LD} \} \). Second, the L-referent is necessarily included as an integral part of the definition of QLM because it is this referent system that “puts the ‘L’ in the QLM”: For a general evolution operator, there is no objective basis for associating even a specific number \( N \) of outputs, much less a particular logical transformation \( L \), with a given quantum machine (see Section 3). At best, one can ask how well the structure of a logical mapping \( L \) from logical input states to logical output states – instantiated in the set \( \{ \hat{D}_i \} \) of physical L-referent states in a QLM – is mirrored in a dynamical mapping \( \Lambda \) from physical input states to physical output states of a quantum machine. We will show in Section 4 how this question can be answered quantitatively using physical-information-theoretic measures for the representational faithfulness and computational fidelity of a quantum L-machine. Third, we emphasize the critical distinction between the referent system in our QLM – which is external to the device \( D \) and remains forever isolated from \( D \) once evolution has commenced – and a memory of the input internal to \( D \) that can be accessed by other components of \( D \) during and after its evolution via \( \Lambda \). Inclusion of such a memory as part of \( D \) would render all logical transformations \( L \) that can be ascribed to the device logically reversible (e.g. [9]), which would in turn allow for physically irreversible implementation of any \( L \) since the input data would remain “known” to the device throughout its dynamical evolution.12 While nothing in our approach prohibits the absence of a memory internal to the device \( D \), the L-Referent cannot and does not play the role of such a memory since interactions between \( R_L \) and \( D \) are disallowed. Fourth and finally, we note that we have not specified whether the input and output referents are intended as actual physical systems or as fictitious systems akin to the ancillas frequently encountered in quantum information theory. While the referent systems need not be “real” to fulfill important bookkeeping roles they will play later in this work, there are arguments for granting at least the input referent the status of a real physical system. These considerations will be discussed in due course.

---

11 Since the logical transformation \( L \) enters QLM through the L-referent, and since all logical states are physically instantiated in the L-referent, the terms “logical states” and “referent states” will be used interchangeably hereafter.

12 One example is the physically reversible implementation of the logically irreversible RESET operation described in section 7 of Ref. [2], which is possible only because of the presence of the memory M internal to \( D \) and its coupling to the remainder of \( D \) (i.e. the register \( R \)) throughout implementation of the logical transformation RESET. Another example is the addition of a “history tape” to a Turing machine, which enables reversible operation of the machine by creating and later accessing a record of the machine states visited during a computation [10–12].
2.4. Classes of quantum L-machines

We conclude this section by identifying important classes of QLMs, including the quantum analog of ICLM. Given the critical relevance of faithful representation and state distinguishability for L-machines, we first define classes of QLMs relevant to these notions. First, we define faithful quantum L-machines:

**Definition 6.** A quantum L-machine is a faithful quantum L-machine if, for every $j$, $\Lambda$ evolves all device input states $\hat{D}_i^{(\text{in})} \in \{\hat{D}_i^{(\text{in})}\} = \{\hat{D}_i^{(\text{in})} | i \in \{i\}_j\}$ into a unique device output state $\hat{D}_j^{(\text{out})}$, i.e.

$$\Lambda(\hat{D}_i^{(\text{in})}) = \hat{D}_j^{(\text{out})} \quad \forall i \in \{i\}_j = \{|i\Lambda(x_i) = y_j\}. \quad (19)$$

This is to say that all $y_j$ have faithful physical representations in $\mathcal{D}$. The resulting simplifications to QLM are replacement of the states $\Lambda(\hat{D}_i^{(\text{in})})$ by $\hat{D}_j^{(\text{out})}$ for all $i \in \{i\}_j$ in Eq. (16) and replacement of the projectors $\hat{P}_j^{\mathcal{D}}$ by $\hat{P}_j^{\mathcal{D}}$ for all $i \in \{i\}_j$ in Eq. (18), where $\hat{P}_j^{\mathcal{D}}$ becomes the identity for the support of $\hat{D}_j^{(\text{out})}$. In a faithful QLM unique device states $\hat{D}_j^{(\text{out})}$ can be associated with each of the logical outputs, rather than just subspaces $\mathcal{H}_j^{\mathcal{D}}$ of $\mathcal{D}$ as in a general QLM, although these output states need not be mutually orthogonal.

Next, we define noiseless quantum L-machines:

**Definition 7.** A quantum L-machine is a noiseless quantum L-machine if the shell projectors

$$\hat{P}_j^{\mathcal{D}} = \hat{I}^{\mathcal{D}} - \prod_{i \in \{i\}_j} \left(\hat{I}^{\mathcal{D}} - \hat{P}_i^{\mathcal{D}}\right) \quad (20)$$

associated with the subsets $\{\Lambda(\hat{D}_i^{(\text{in})})| i \in \{i\}_j\}$ have orthogonal support, i.e. if

$$\hat{P}_j^{\mathcal{D}} \hat{P}_j^{\mathcal{D}}' = \delta_{jj'} \hat{P}_j^{\mathcal{D}} \quad \forall j, j'. \quad (21)$$

The resulting simplification to QLM is replacement of the projection operators $\{\hat{P}_j^{\mathcal{R}_i^{\mathcal{D}}}\}$ by the local projectors $\{\hat{P}_j^{\mathcal{D}}\}$ in Definition 5. In a noiseless QLM, logical output states can be unambiguously associated with orthogonal subspaces of the device’s state space $\mathcal{H}_j^{\mathcal{D}}$ alone, rather than with the full product space $\mathcal{H}_{\mathcal{R}_i^{\mathcal{D}}} \otimes \mathcal{H}_{\mathcal{D}}$. This is not, however, to say that logical outputs can necessarily be associated with physical states of $\mathcal{D}$ in a noiseless QLM as they can in a faithful QLM.

A general quantum L-machine need not be either faithful or noiseless. In addition, as indicated above, a faithful QLM can be noisy and a noiseless QLM can be unfaithful. A QLM can, of course, be both faithful and noiseless, and we call such a machine an ideal quantum L-machine (IQLM). The simplifications to Definition 5 that result from the two conditions together are significant, yielding the definition:

**Definition 8.** An $M$-input, $N$-output ideal quantum L-machine (IQLM) is an ordered set

$$\{\mathcal{D}, R, \{\hat{D}_i^{(\text{in})}\}, \{\hat{D}_j^{(\text{out})}\}, \rho^{\mathcal{R}_i^{\mathcal{D}}}, \Lambda\} \quad (22)$$

consisting of

- A physical device $\mathcal{D}$, situated in an environment $\mathcal{E}$.
- An $M$-input, $N$-output $\mathcal{R}$-referent $R$.
- A set $\{\hat{D}_i^{(\text{in})}\}$ of $M$ distinguishable input states of $\mathcal{D}$.
- A set $\{\hat{D}_j^{(\text{out})}\}$ of $N$ distinguishable output states of $\mathcal{D}$.
- A set $\{\rho^{\mathcal{R}_i^{\mathcal{D}}}\}$ of $M$ joint initial states of $\mathcal{R}_i \mathcal{D}$, constructed from a one-to-one pairing of $\rho^{\mathcal{R}_i^{\mathcal{D}}}$ and $\hat{D}_i^{(\text{in})}$ into product states $\rho_i^{\mathcal{R}_i^{\mathcal{D}}} = \hat{I}^{\mathcal{R}_i^{\mathcal{D}}} \otimes \hat{D}_i^{(\text{in})}$.
- A time-evolution operator $\Lambda$ for $\mathcal{R}_i \mathcal{D}$, which maps initial states $\rho_i^{\mathcal{R}_i^{\mathcal{D}}}$ of $\mathcal{R}_i \mathcal{D}$ into final states as

$$\Lambda(\rho_i^{\mathcal{R}_i^{\mathcal{D}}} \otimes \hat{D}_j^{(\text{in})}) = \rho_i^{\mathcal{R}_i^{\mathcal{D}}} \otimes \hat{D}_j^{(\text{out})} \quad \forall i \in \{|i\rangle = \{|i\Lambda(x_i) = y_j\}. \quad (23)$$

where $L$ is a logical transformation that maps $M$ logical input states $x_i \in \{x_i\}$ into $N$ logical output states $y_j \in \{y_j\}$ via $x_i \rightarrow L(x_i) = y_j$. IQLM is the clearly quantum analog of ICLM, which is also both faithful and noiseless, the most conspicuous difference being the presence of the $\mathcal{L}$-referent in ICLM.

We finally define one additional class of noiseless QLMs.

**Definition 9.** An $N$-output ($N \geq 2$) noiseless quantum L-machine is a logically irreducible quantum L-machine if the underlying quantum machine cannot support any noiseless quantum L-machine with more than $N$ outputs.

Note that noiseless L-machines that are also faithful – i.e. ideal L-machines – are logically irreducible, but noiseless machines that are not faithful need not be logically irreducible. The significance of this class of machines will become apparent in the next section.
Fig. 1. State space structure for evolved input states in various classes of QLMs implementing a 4-input 2-output logical transformation for which \( y = L(x) \) with \( y_0 = L(x_0), \ y_1 = L(x_1) = L(x_2) = L(x_3) \). \( \Lambda(D^\infty) \) is the support of \( \Lambda(D^\infty) \), as indicated in the key. Depicted are (a) a general QLM as well as QLMs that are (b) noisy but faithful, (c) noiseless but unfaithful and logically reducible, (d) noiseless and unfaithful but logically irreducible, and (e) ideal (i.e. faithful, noiseless, and logically irreducible).

To summarize our classification of quantum L-machines, we illustrate key features of the final state space structure of \( D \) for important classes of QLMs in Fig. 1. Specifically, we depict the support subspaces \( \gamma(D) \) for evolved device states \( \Lambda(D^\infty) \) that represent the \( M = 4 \) logical inputs for QLMs implementing a 4-input 2-output logical transformation with \( y_0 = L(x_0), \ y_1 = L(x_1) = L(x_2) = L(x_3) \). The state space structure for a generally noisy and unfaithful QLM is depicted in (a). A faithful but noisy QLM is depicted in Fig. 1(b). The QLM of Fig. 1(c) is noiseless and unfaithful, and is logically reducible since the underlying quantum machine could support a noiseless QLM with four outputs. This is not the case for the QLM of Fig. 1(d), which is also noiseless and faithful but is logically irreducible. Finally, the QLM of Fig. 1(e) is both noiseless and faithful – i.e. ideal – and is thus also logically irreducible.

3. Realism about computation and L-machines

Computational realism maintains that there is an objective fact of the matter about whether a physical system implements a particular computation. Evaluation of such a claim clearly requires an explicit account of what it means for a physical system to implement a computation. Ladyman has argued that the account of computation embodied in the ideal classical L-machine supports a qualified version of computational realism – a realism about the “transformation structure” implemented by such a machine – since the machine’s ability to implement a given transformation structure objectively “…depends on the structure of the system and the laws governing its time evolution and is independent of exactly what the physical states of the system are taken to represent” [1]. He contrasts this with a stronger realism that would objectively ascribe semantic content to the logical states, noting the “freedom of association” that is inherent in semantic representation in this context.13 In this Section, we argue that even a realism about transformation structure is untenable in the most general quantum L-machines. We then identify the class of quantum L-machines that do support this qualified form of computational realism, which we will hereafter refer to simply as “computational realism”.14

We should first be clear that, in denying computational realism for generalized L-machines, we take it that a realist claim about digital computation amounts to an operational claim that an observer could, at least in principle, establish that a physical system implements a particular transformation structure. For a quantum machine, this entails that an experimenter preparing a device \( D \) in various initial states \( \hat{D}_i \in \{\hat{D}_i\} \), letting the device evolve via \( \Lambda \), and performing a measurement on \( D \), would observe that the preparation and evolution of every input state \( \hat{D}_i \), belonging to the \( j \)-th subset \( \{\hat{D}_i\}_{j} = \{\hat{D}_i | i \in [i]_j\} \) of \( \{\hat{D}_i\} \) is always followed by realization of a unique measurement outcome (call it \( \omega_j \)).15 Only such an unambiguous association

---

13 Consider, for example, the 4-input, 2-output ideal L-machine of Fig. 1(e). By assigning binary strings to the logical inputs and outputs as \([x_0, x_1, x_2, x_3] = [00, 01, 10, 11]\) and \([y_0, y_1] = [0, 1]\), the machine can be regarded as a Boolean AND gate (Ladyman’s L\text{AND}-machine). By instead making the assignments \([x_0, x_1, x_2, x_3] = [11, 10, 01, 00]\) and \([y_0, y_1] = [1, 0]\) and the same underlying machine can be regarded as a Boolean OR gate (Ladyman’s L\text{OR}-machine). Other Boolean operations result from further permutation of the input and output assignments. Because these assignments are arbitrarily, there are no objective grounds for regarding the machine as a physical implementation of one particular 4-input 2-output Boolean operation but not all other structurally equivalent Boolean operations (i.e. operations that are equivalent up to permutation of the binary string assignments to logical states).

14 Recent discussions of computational realism and the role that representation may or may not play in it can be found in Refs [1,13–15].

15 As an aside, we note that this implicitly requires existence of an appropriate referent system. Since the initial device states are generally destroyed during evolution of the device and measurement, an observer can correlate the initial device state with the measurement outcome at the conclusion of any given trial only if a record of the initial state has been maintained. The physical embodiment of this record – e.g. in the observer’s own memory or
of distinct measurement outcomes with disjoint subsets of initial states would allow an observer to objectively ascribe a particular transformation structure to a quantum machine, so the existence of such measurements is necessary for a realist conception of implementation of a logical transformation by the machine.

This essential condition is, however, met in a quantum machine if and only if the support subspaces $\mathcal{H}_i^D$ associated with the subsets $\{\Lambda(\hat{D}_i)\}$ of evolved initial states of $D$ are disjoint, i.e. if the corresponding shell projectors $\hat{P}_i^D$ defined in Eq. (6) are mutually orthogonal.\(^{16}\) Since there are no restrictions on the final states $\Lambda(\hat{D}_i)$ of general quantum machines that would ensure satisfaction of this condition, measurements that would map evolved states belonging to subsets $\{\Lambda(\hat{D}_i)\}$ into distinct outcomes $\omega_D$ generally do not exist for such machines even in principle. Operationally, this would deny any observer the very physical resources that would enable objective association of a unique transformation structure with a quantum machine, thus underlining the basis for a realist conception of computation in general quantum $L$-machines. Realism fails for noisy classical machines for similar reasons.

Two questions remain: First, what requirements must be satisfied for a quantum machine to support a realist conception of computation in the above sense? Second, what can be said about computation in $L$-machines when these conditions are not met? We address the first question below, leaving the second for detailed consideration in Section 4.

The orthogonality condition on output subspaces (i.e. Eq. (21)), shown above to be necessary for a realist conception of a QLM, is necessary and sufficient for defining a noiseless QLM on an underlying quantum machine $QM$: If, for an $M$-input quantum machine, there exists a partition $\{|i\}_{N} = \{|i\}_0, |i\}_1, \ldots |i\}_{N-1}\}$ of the $M$ input indices into any $N \geq 2$ subsets such that the corresponding output support subspaces $\mathcal{H}_i^D$ are disjoint (i.e. the $N$ associated shell projectors are mutually orthogonal), then there exists a nontrivial noiseless QLM for the machine. Thus, satisfaction of the defining conditions for a noiseless QLM alone ensures satisfaction of the necessary conditions for computational realism. The defining conditions for a noiseless QLM are, however, not sufficient for realism, since multiple noiseless QLMs – associated with distinctly different transformation structures – can generally be defined for quantum machines that can support noiseless QLMs (with $M > 3$ and $N \geq 2$). The reason for this is clearly illustrated by the simple example discussed below, which was alluded to by LPSG [2].

Consider an $M$-input quantum machine that noiselessly implements the identity, i.e. that is characterized by the time-evolution operator $\Lambda$ acting as

$$\Lambda(\hat{D}_i) = \hat{D}_i.$$ \(^{24}\)

There is a trivial sense in which this $M$-input "identity" machine can be construed as a noiseless physical implementation of every logical mapping of $M$ inputs into $N \leq M$ outputs. Indeed, any $M$-input, $N$-output noiseless QLM can be defined for the underlying quantum machine simply by appropriate grouping of the input states. To see this, consider such a machine with $M = 4$ inputs. One could specify $N = 2$ logical outputs $\{y\} = \{y_0, y_1\} = \{1, 0\}$ and make the assignments $|i\}_0 = |0\}$ and $|i\}_1 = |1\}$, and the machine is a noiseless 4-input, 2-output QLM with a particular transformation structure. By instead choosing the input assignments $|i\}_0 = |0\}$ and $|i\}_1 = |1\}$, the same machine is a noiseless 4-input, 2-output QLM with a different transformation structure. Specifying logical outputs $\{y\} = \{y_0, y_1\}$ and making the assignments $|i\}_0 = |0\}$, $|i\}_1 = |1\}$, and $|i\}_2 = |2\}$, we have a noiseless 4-input QLM that has yet a different transformation structure and even a different number of outputs (three instead of two). Because the assignment of final physical states to logical outputs can be made arbitrarily, the assignment of even a unique transformation structure to a machine is, in this sense, arbitrary.\(^{17}\) For this reason, the conditions that allow a noiseless QLM to be defined on a given underlying quantum machine are necessary but generally insufficient for realism about computation in quantum $L$-machines.

This inability to objectively associate even a unique transformation structure with a given machine is avoided in LPSG’s ideal classical $L$-machine by the requirement that all logical states must, by definition, have faithful physical representations. This ensures that, if an $M$-input classical machine is to “qualify” as an $M$-input $N$-output ICLM, the underlying classical machine must complete all of the work that the logical transformation $\Lambda$ requires: It must map all physical states $\{D^{(in)}_{ij}\}$ representing inputs $x_i$ that “belong to” the same logical output $y_j$ into a unique final physical state $D^{(out)}_{ij}$. In the ideal classical context, this is to say that no ICLM with more than $N$ outputs could be defined on the underlying classical machine. For example, by this criterion an $M = 4$ input classical “identity” machine could thus support only ICLMs with $N = 4$ logical outputs, since at least one of the output states of any ICLM with $N < 4$ defined on this machine would necessarily be unfaithful.

The analogous requirement in the general quantum context – the requirement that is necessary and sufficient for a realist conception of an $M$-input $N$-output quantum $L$-machine – is that the QLM is logically irreducible. This is to say that

\(^{16}\) This is a straightforward consequence of quantum measurement theory [6].

\(^{17}\) Similar considerations were used by Putnam [16] and Searle [17] to argue that, with complete freedom in the assignment of physical states to logical states, the evolution of any physical system of sufficient microscopic complexity (e.g. a rock or wall) can be interpreted as an implementation of any finite state automata. (See also [18].) For a recent critical discussion of these and other anti-realist views of computation, see [19].
the QLM is an \( M \)-input \( N \)-output noiseless QLM and that the underlying quantum machine can support no noiseless QLMs with more than \( N \) outputs, as discussed in Section 2.4. Unlike ICLM (or IQLM), where logical irreducibility implies faithful representation, realism does not require faithful representation of logical outputs in this more general context. Rather, in addition to the requirement that evolved machine states representing logical inputs belonging to different logical outputs must be perfectly distinguishable from one another, those belonging to the same logical output must not be perfectly distinguishable from one another (as in Fig. 1(d) and (e)).

4. Computation in generalized \( L \)-machines

4.1. General considerations

In the previous section we identified the class of classical and quantum \( L \)-machines – logically irreducible \( L \)-machines – that can be realistically construed as noiseless physical implementations of deterministic logical transformations. In an \( M \)-input \( N \)-output logically irreducible QLM, there must exist precisely \( N \) disjoint subspaces of the underlying quantum machine’s state space such that each of the \( M \) specified initial physical states evolves into a final state with support on no more than one of these subspaces. Such a machine objectively exhibits the “transformation structure” that would qualify it as an unambiguous physical implementation of an \( M \)-input \( N \)-output logical transformation \( L \).

But what about more general (and more realistic) sorts of machines that do not meet these criteria, machines for which the physical states representing the various logical outputs cannot be perfectly distinguished from one another because of classical and/or quantum noise? In the general case, realism about computation is untenable even in the “qualified” form discussed in Section 3. Do such machines implement logical transformations? If so, what exactly does it mean to say that such machines implement logical transformations?

We answer the first question in the affirmative, maintaining that any \( M \)-input machine can be regarded as an \( L \)-machine that implements any \( M \)-input logical operation \( L \) with some degree of efficacy. In addressing the second question, we identify physical-information-theoretic measures that quantify – for a specified \( L \)-machine – how faithfully and with what fidelity the machine implements the logical transformation \( L \). By allowing for shades of grey in computational efficacy, we can relax the realist requirement for objective ascription of a unique transformation structure to an \( L \)-machine without embracing an extreme anti-realism – or “computational nihilism” [19] – that would deny any basis for regarding a physical system as an implementation of any particular logical transformation. The efficacy measures quantify how well a logical transformation is mirrored in a family of physical processes, providing a basis for making such associations according to quantitative criteria and standards.

We emphasize that, in constructing the information-theoretic framework within which these quantitative efficacy measures are defined, we adopt a relational conception of information that regards information as always being “about something else”. On this view, to say that “there is information in \( Y \)” can only mean that “there is information in \( Y \) about \( X \)” where \( X \) is a referent; it is that about which \( Y \) holds information.\(^{18}\) If information is taken to be encoded in the states of physical systems, as it is when physical states of the device \( D \) in a quantum \( L \)-machine are taken to be representations of the physical states of an \( L \)-referent \( R_L \), then to say that “there is information in \( D \) about \( R_L \)” is to say that “there is information in the state of system \( D \) about the state of system \( R_L \)”. The combination of the \( L \)-referent and device in the quantum \( L \)-machine, together with mutual information measures appropriate to the quantum context, provides a natural framework for quantifying and relating faithfulness of representation, fidelity of computation, loss of information, and the physical costs of computation using this relational conception. We now develop these ideas, using a physical-information-theoretic approach that treats the quantum \( L \)-machine as a quantum generalization of Winograd and Cowan’s classical computation channel.

4.2. Classical computation channels

In their classic monograph Reliable Computation in the Presence of Noise [4], Winograd and Cowan adapted Shannon’s conception of the noisy communication channel [20–22] to the information-theoretic characterization of noisy computation. Their “computation channel”, like Shannon’s (discrete) communication channel, consists of an \( M \)-ary input alphabet \( \{x_i\} \), an \( N \)-ary output alphabet \( \{y_j\} \), and a set \( \{q_{ij}\} \) of conditional probabilities that characterize the statistical properties of the channel. (\( q_{ij} \) is the probability that input \( x_i \) is mapped into output \( y_j \).) Given a channel and a distribution \( \{p_i\} \) of input probabilities, random variables \( X = \{ p_i, x_i \} \) and \( Y = \{ q_j, y_j \} \) are defined for the input and output, where \( \{q_j\} \) is the distribution of output probabilities \( q_j = \sum_i p_i q_{ij} \). Information-theoretic entropies are then defined and used to characterize statistical uncertainties in \( X \) and correlations between \( X \) and \( Y \). We now discuss two entropic quantities of particular interest in this work – the conditional entropy and the mutual information – with emphasis on their connection to information loss in computation channels.

\(^{18}\) This is to say that we take correlation entropy (mutual information), rather than self-entropy (self-information), to be the appropriate measure for information content.
The classical conditional entropy

\[ H(X|Y) = \sum_{j=1}^{N} q_j H(x_j|y_j) \]  

(25)

is the average, over all outputs \( y_j \), of the quantity

\[ H(X|y_j) = -\sum_{i=1}^{M} p_{ij} \log_2 p_{ij} \]  

(26)

where \( p_{ij} = q_{ij}/q_j \) is, by Bayes’ Rule, the conditional probability that occurrence of output \( y_j \) resulted from input \( x_i \), \( H(X|y_j) \) serves as a measure of the uncertainty in the input given the \( j \)-th output of a communication or computation channel, with \( H(X|Y) > 0 \) indicating that, for at least one output, information is lost in the mapping from input \( X \) to output \( Y \). This information loss is undesirable in communication channels, where one hopes to infer all \( M \) inputs from the outputs without ambiguity (\( H(X|Y) = 0 \)), but is completely natural (indeed necessary) in noiseless computation channels with \( M > N \); A binary “0” at the output of an ideal Boolean AND gate results from application of either 00, 01, or 10 at the input, but all information about which of these three inputs was applied is “lost” in implementation of the logical transformation.

The classical mutual information is defined as

\[ I(X; Y) = H(X) - H(X|Y) \]  

(27)

where

\[ H(X) = -\sum_{i=1}^{M} p_i \log_2 p_i \]  

(28)

is the classical Shannon self-entropy associated with the input \( X \), \( I(X; Y) \), which is a measure of the correlation between the random variables \( X \) and \( Y \), can be interpreted heuristically as the “information in \( Y \) about \( X \)” and vice versa. In a noiseless communication channel, where the input can be inferred from the output, the mutual information is \( I(X; Y) = H(X) \). In a computation channel, however, we expect less: A noiseless computation channel maps \( N \) subsets \( \{x_i\} = \{x_i | i \in [i]\} \) of the \( M \) inputs \( x_i \) into \( N \) distinct outputs \( y_j \) \( (q_{ij} = 1 \forall i \in [i], q_{ij} = 0 \forall i \notin [i]) \) yielding a mutual information

\[ I(X; Y) = H(Y) - \sum_{j=1}^{N} q_j \log_2 q_j \]  

(29)

where \( H(Y) \) is the Shannon entropy of the output \( Y \). Since \( H(X) > H(Y) \) for noiseless computation channels with \( M > N \),\(^{19} \) i.e. for noiseless computation channels implementing logically irreversible transformations, we have \( H(X) > I(X; Y) \) and thus, by Eq. (27), \( H(X|Y) > 0 \).

These considerations support the quantification of information loss by the conditional entropy and the association of information loss with computation. Using Shannon’s mutual information as an information measure, the information “about \( X \)” that is lost going from input to output is

\[-\Delta I = I(X; X) - I(X; Y) = H(X) - [H(X) - H(X|Y)] = H(X|Y).\]  

(30)

Since \( H(X|Y) > 0 \) for noiseless computation channels with \( M > N \) (see above), it follows that

\[-\Delta I > 0\]  

(31)

for noiseless computation channels implementing logically irreversible transformations.

Wingrad and Cowan recognized the connection between information loss and computation early on, going so far as to identify “the destruction of information” as the defining characteristic of computation [4]:

“We say that computation occurs if \( H(X|Y) > 0 \) i.e., if the output symbols do not completely specify the input configurations; and we say that communication occurs if \( H(X|Y) = 0 \), i.e. if the output symbols completely specify the input configurations...[I]t follows...that computation occurs if \( H(X) > H(Y) \), i.e. if information is lost in going from \( X \) to \( Y \).”

\(^{19}\) This is a consequence of the grouping property of Shannon entropy. See, for example, [22].
While there is more to computation than the selective destruction of information,20 information loss is the defining characteristic of irreversible computation and is effectively quantified by the conditional entropy. We will make use of these considerations, and give them physical interpretations, as we consider the quantum \( L \)-machine from a physical-information-theoretic point of view.

4.3. Quantum \( L \)-machines as computation channels

We now provide a physical-information-theoretic characterization of an \( M \)-input \( N \)-output quantum \( L \)-machine, which is similar to a classical computation channel but with ensembles of quantum states replacing ensembles of classical logical states (i.e. the random variables \( X \) and \( Y \)) and an appropriate quantum correlation measure replacing the Shannon mutual information. We also take a first look at the information loss in implementation of logical transformations by QLMs.

4.2.1. Input and output ensembles

We first define the input ensemble for a quantum \( L \)-machine as

\[
\epsilon_x^{R_LD} = \{p_i, \hat{\rho}_i^{R_LD}\}
\]

where \( p_i \) is the probability that the machine is initially prepared in the state \( \hat{\rho}_i^{R_LD} = \hat{\tau}_i^{R_L} \otimes \hat{\rho}_i^{(m)} \) corresponding to the \( i \)-th logical input \( x \). The density operator describing the statistical state of this ensemble is

\[
\hat{\rho}^{R_LD} = \sum_{i=1}^{M} p_i \hat{\rho}_i^{R_LD}.
\]

To obtain the output ensemble, we first evolve all \( M \) members of the input ensemble via \( \hat{\Lambda} \) to obtain the evolved input ensemble

\[
\epsilon_x^{R_LD'} = \{p_i, \hat{\rho}_i^{R_LD'}\}
\]

where \( \hat{\rho}_i^{R_LD'} = \hat{\tau}_i^{R_L} \otimes \hat{\Lambda}(\hat{\rho}_i^{(m)}) \). The elements of the output ensemble

\[
\epsilon_x^{R_LD'} = \{q_j, \hat{\rho}_j^{R_LD'}\}
\]

can then be “projected out” of the statistical state

\[
\hat{\rho}^{R_LD'} = \sum_{i=1}^{M} p_i \hat{\rho}_i^{R_LD'}
\]

of the evolved input ensemble as

\[
\hat{\rho}_j^{R_LD'} = \frac{1}{q_j} \hat{\Pi}_j^{R_LD} \hat{\rho}^{R_LD'} \hat{\Pi}_j^{R_LD} = \sum_{i \in [i_j]} p_i \hat{\rho}_i^{R_LD'}
\]

where \( \hat{\Pi}_j^{R_LD} \) is the projector associated with the \( j \)-th logical output in the definition of a QLM (Eq. (18)), \( p_i^{(j)} = \frac{p_i}{q_j} \), and

\[
q_j = \text{Tr} \left[ \hat{\Pi}_j^{R_LD} \hat{\rho}^{R_LD'} \right] = \sum_{i \in [i_j]} p_i.
\]

We then define \( N \) output ensembles

\[
\epsilon_j^{R_LD'} = \{p_i^{(j)}, \hat{\rho}_i^{R_LD'} | i \in [i_j]\},
\]

associated with the \( N \) logical outputs, and note that the density operators associated with these ensembles are just the \( \hat{\rho}_j^{R_LD'} \) defined above (Eq. (37)). The \( j \)-th reduced density operator

\[
\hat{\rho}_j^{D'} = \text{Tr}_{R_L} [\hat{\rho}_j^{R_LD'}] = \sum_{i \in [i_j]} p_i^{(j)} \hat{\Lambda}(\hat{\rho}_i^{(m)})
\]

provides a statistical physical representation of the logical output \( y_j \) – for input distribution \( \{p_i\} \) in the device \( D \) alone. Hereafter, when we refer to \( \hat{\rho}_j^{D'} \) as a “physical representation of \( y_j \)” or simply as the \( j \)-th “output state” we will mean it in this statistical sense: As we discussed in Sections 2 and 3, unique final device states generally cannot be associated with logical output states in quantum \( L \)-machines.

\[20\] Reversible logical transformations are certainly “computations” even though they do not entail the destruction of information. The transformation structure associated with a noiseless \( L \)-machine implementing a reversible computation is trivial – it is that of the identity machine (e.g. [23]) – even when the computational task of evolving input states into output states that have the required structure is highly nontrivial (e.g. as it is in prime factorization).
4.3.2. Mutual information

In selecting a “referential” information measure appropriate for this setting, we first note that here “the information about the input that is in the output” is “the information about the identity of the initial quantum state of $D$ that is in the structure of the final quantum state of $D'$”. The obvious choice would seem to be the quantum mutual information, which can be used to quantify correlations between quantum states in much the same way as the classical mutual information quantifies correlations between random variables. There is, however, a hitch: The quantum mutual information quantifies correlations between the states of two systems at a given time, while we seek to quantify correlation between the state of a given system at two times.

Fortunately, a physical record of the initial state of $D$ resides permanently in the state of the input referent $R_{in}$. The correlations of interest are thus captured in the quantum mutual information

$$I(\hat{\rho}^{R_{in}}; \hat{\rho}^{D'}) = S(\hat{\rho}^{R_{in}}) + S(\hat{\rho}^{D'}) - S(\hat{\rho}^{R_{in}D'})$$

(41)
defined for the joint final state $\hat{\rho}^{R_{in}D'}$ of $R_{in}D$, where $S(\hat{\rho}) = -Tr[\hat{\rho} \log_2 \hat{\rho}]$ is the von Neumann entropy. Owing to the form of $\hat{\rho}^{R_{in}D'}$ and the mutual orthogonality of the $\{\hat{r}^{R_{in}}\}$, this is, by the joint entropy theorem, equivalent to

$$I(\hat{\rho}^{R_{in}}; \hat{\rho}^{D'}) = \chi(\epsilon^{D'})$$

(42)
where

$$\chi(\epsilon^{D'}) = S \left( \sum_{i=1}^{M} p_i \Lambda(\hat{D}_i) \right) - \sum_{i=1}^{M} p_i S \left( \Lambda(\hat{D}_i^{in}) \right)$$

(43)
is the Holevo information for the ensemble $\epsilon^{D'} = \{ p_i, \Lambda(\hat{D}_i^{in}) \}$ of the $M$ (generally nonorthogonal) evolved device states.

Support for this choice of information measure is found in the well known Holevo (or Levitin–Holevo) theorem [24–26], which identifies $\chi(\epsilon^{D'})$ as an upper bound on the accessible information for the ensemble $\epsilon^{D'}$. In this context, the accessible information is the maximum classical (Shannon mutual) information about the input $X$ that could be obtained from the outcomes of any physically realizable measurement performed on $D$ in its final quantum state. The Holevo information is bounded as $0 \leq \chi(\epsilon^{D'}) \leq H(X)$, with equality in the upper bound when the final device states $\Lambda(\hat{D}_i^{in})$ are mutually orthogonal and thus distinguishable (and $I(\hat{\rho}^{R_{in}}; \hat{\rho}^{D'}) = H(X)$), and equality in the lower bound when the final device state is completely independent of its initial state (and $I(\hat{\rho}^{R_{in}}; \hat{\rho}^{D'}) = 0$).

4.3.3. Information loss

Using this mutual information measure, the information about the logical input $X$ that is lost as a quantum $L$-machine evolves from its initial to final state is

$$-\Delta I \equiv I(\hat{\rho}^{R_{in}}; \hat{\rho}^{D}) - I(\hat{\rho}^{R_{in}}; \hat{\rho}^{D'})$$

(44)
where $I(\hat{\rho}^{R_{in}}; \hat{\rho}^{D'}) = \chi(\epsilon^{D'})$ is the information about $X$ that is initially in $D$. Noting that $D$ initially holds “all of the information” about $X$, since the $x_i$ are unambiguously encoded in distinguishable input states $\hat{D}_i^{in}$, $I(\hat{\rho}^{R_{in}}; \hat{\rho}^{D}) = H(X)$ and the information loss is

$$-\Delta I = H(X) - \chi(\epsilon^{D'})$$

(45)

for a general quantum $L$-machine. Since $\chi(\epsilon^{D'}) \leq H(X)$, with equality only for mutually orthogonal evolved input states $\{\Lambda(\hat{D}_i^{in})\}$, the information loss $-\Delta I$ is nonnegative and vanishes only for evolution operators $\Lambda$ that completely preserve the distinguishability of the initial device states $\{\hat{D}_i^{in}\}$. We will revisit this result in Section 4.5, following definition of quantitative measures for representational faithfulness and computational fidelity.

4.4. Measures of computational efficacy: faithfulness and fidelity

We now define physical-information-theoretic measures that quantify the faithfulness and fidelity with which a quantum machine implements a logical transformation $L$. We motivate these measures heuristically, and argue that they successfully capture complementary aspects of computational efficacy.

---

21 See, for example, p. 513 in [5].

22 See, for example, p. 513 in [5].
4.4.1. Representational faithfulness

We first generalize the notion of faithful implementation that LPSG defined for ICLM, which is codified in the condition that all device input states “belonging to” the same logical output of $L$ evolve into the same device output state ($\Lambda_t(D^{(in)}_i) = D^{(out)}_j \forall i \in |i_l| = \{i_l(x_i) = y_j\}$). This condition requires that none of the evolved device states $\Lambda(D^{(in)}_i)$ contain any information that could help to identify the state $D^{(in)}_j \epsilon \{D^{(in)}_j\}$ from which it evolved. Generalizing this to QLM, we say that a logical output $y_j$ is represented faithfully in $D$ if the physical representation $\hat{\rho}_j^{D'} = Tr_{R_{in}}(\hat{\rho}_j^{R_{in}D'})$ of $y_j$ in $D$ contains no information about the identity of the logical input $x_i \in |x_i|L(x_i) = y_j$. Quantitatively, this is to say that

$$I(\hat{\rho}_j^{R_{in}}; \hat{\rho}_j^{D'}) = 0$$

(46)

where

$$I(\hat{\rho}_j^{R_{in}}; \hat{\rho}_j^{D'}) = S(\hat{\rho}_j^{R_{in}}) + S(\hat{\rho}_j^{D'}) - S(\hat{\rho}_j^{R_{in}D'}) = \chi(\epsilon_j^{D'})$$

(47)

with $\epsilon_j^{D'} = \{p_i^{(j)}, \hat{\rho}_i^{D'} | i \in |i_l]\}$. A QLM is a faithful physical implementation of $L$ if all logical outputs $y_j$ of $L$ have faithful physical representations in the QLM.

To quantify representational faithfulness, note that the quantity $I(\hat{\rho}_j^{R_{in}}; \hat{\rho}_j^{D'})$ achieves its maximum value of

$$H(X|y_j) = -\sum_{i \in |i_l|} p_i^{(j)} \log_2 p_i^{(j)}$$

(48)

when there exist measurements on the final state of $\mathcal{D}$ that could uniquely identify the logical input, i.e. when the physical output state $\hat{\rho}_j^{D'}$ is a completely unfaithful representation of the logical output $y_j$. This suggests the difference $H(X|y_j) - \chi(\epsilon_j^{D'})$ as the basis for a quantitative measure of how faithfully $y_j$ is physically represented in $\mathcal{D}$. This would further suggest the output average

$$\sum_{j=1}^{N} q_j \left[ H(X|y_j) - \chi(\epsilon_j^{D'}) \right] = \left[ 1 - \frac{1}{H_t(X|Y)} \sum_{j=1}^{N} q_j \chi(\epsilon_j^{D'}) \right] H_t(X|Y)$$

(49)

as a measure of the representational faithfulness of the $L$-machine as a whole, where the right-hand side is obtained by multiplying and dividing the left hand side by its maximum value $\sum_{j=1}^{N} q_j H_t(X|y_j)$ and noting that $\sum_{j=1}^{N} q_j H_t(X|y_j) = H_t(X|Y)$, where $H_t(X|Y)$ is the conditional entropy associated with the logical operation $L$ for input distribution $\{p_i\}$. These considerations suggest the following definition:

**Definition 10.** For a quantum $L$-machine and input distribution $\{p_i\}$, the representational faithfulness is

$$f_t \equiv 1 - \frac{1}{H_t(X|Y)} \sum_{j=1}^{N} q_j \chi(\epsilon_j^{D'})$$

(50)

where $q_j$ and $H_t(X|Y)$ are the $j$-th output probability and the conditional entropy associated with the logical transformation $L$ for input distribution $\{p_i\}$ and $\chi(\epsilon_j^{D'})$ is the Holevo information associated with the ensemble $\epsilon_j^{D'} = \{p_i^{(j)}, \hat{\rho}_i^{D'} | i \in |i_l]\}$ of final (reduced) device states $\hat{\rho}_j^{D'}$ representing the logical output states $y_j$ of $L$.

The product $f_t H_t(X|Y)$ is thus the average, over all logical outputs, of the information “about the logical input” that is lost in producing the physical representations of the logical outputs. The representational faithfulness is bounded as $0 \leq f_t \leq 1$, with equality in the lower bound when all $y_j$ are represented completely unfaithfully – i.e. when no such information is lost – and equality in the upper bound when all $y_j$ are represented faithfully.

4.4.2. Computational fidelity

The representational faithfulness defined above provides an average measure of how faithfully the logical outputs are individually represented in the physical output states of $\mathcal{D}$, independent of how well physical representations of the logical outputs can be distinguished from one another. In considering the computational fidelity, we are instead concerned exclusively with the distinguishability of the representative output states independent of their faithfulness. The fidelity is thus related to the amount of information about the “correct” logical output – encoded in the output referent states – that is reflected in the final physical state of $\mathcal{D}$, i.e. by the quantum mutual information

$$I(\hat{\rho}^{R_{out}}; \hat{\rho}^{D'}) = S(\hat{\rho}^{R_{out}}) + S(\hat{\rho}^{D'}) - S(\hat{\rho}^{R_{out}D'}) = \chi(\epsilon_j^{D'})$$

(51)

This quantity achieves its maximum value of $I(\hat{\rho}^{R_{out}}; \hat{\rho}^{D'}) = H(Y)$ only when the $\hat{\rho}^{D'}$ are mutually orthogonal and thus perfectly distinguishable, and its minimum value of $I(\hat{\rho}^{R_{out}}; \hat{\rho}^{D'}) = 0$ only when the $\hat{\rho}^{D'}$ are identical to one another, i.e. when $\hat{\rho}_j^{D'} = \hat{\rho}_j^{D''}$ for all $j$. This suggests the definition:
Definition 11. For a quantum $L$-machine and input distribution $\{p_i\}$, the computational fidelity is

$$\mathcal{F}_L \equiv \frac{1}{H_L(Y)} \chi(\epsilon_Y^D)$$

(52)

where $H_L(Y)$ is the output entropy associated with the logical transformation $L$ for input distribution $\{p_i\}$, and $\chi(\epsilon_Y^D)$ is the Holevo information associated with the ensemble $\epsilon_Y^D = \{q_j, \rho_j^D\}$ of final device states representing the logical outputs $y_j$ of $L$.

The product $\mathcal{F}_L H_L(Y)$ is thus the amount of information “about the logical output” that is rendered in the final device state of a quantum $L$-machine. The computational fidelity is bounded as $0 \leq \mathcal{F}_L \leq 1$, with equality in the lower bound when the final device state contains no information about the logical output and equality in the upper bound when the output states $\rho_j^D$ have orthogonal support (so measurements exist that could read the logical output from final device state without ambiguity).\(^{23}\)

4.5. Information loss revisited

The representational faithfulness $f_L$ and computational fidelity $\mathcal{F}_L$ quantify complementary aspects of what is required for physical implementation of a logical transformation $L$. Taken together, these two measures are intended to capture the critical similarities between the “transformation structure” of a family of physical processes and an abstract-logical mapping that would qualify the former as an efficacious physical implementation of the latter. That they succeed in this regard is perhaps suggested by their remarkably simple and heuristically suggestive relationship to the information loss, and thus to the “amount of irreversible computation” associated with a family of physical processes. Specifically, as we show in Appendix A, the information lost in evolution of a QLM (Eq. (45)) can be expressed in terms of the efficacy measures as

$$-\Delta I = f_L H_L(X|Y) + (1 - \mathcal{F}_L) H_L(Y).$$

(53)

In this form, the information loss is clearly resolved into components related to faithful representation and computational fidelity. The first term accounts for the desirable information loss that is required to produce faithful representations of logical output states in QLMs with $M > N$, achieving its maximum value of $H_L(X|Y)$ for faithful QLMs (i.e. $f_L = 1$) and vanishing for machines that produce completely unfaithful representations ($f_L = 0$) of all logical outputs. The second term accounts for the undesirable information loss associated with indistinguishability of the output states. This term vanishes for noiseless QLMs (i.e. for $f_L = 1$) that produce perfectly distinguishable representations of logical outputs and achieves its maximum value of $H_L(Y)$ for machines in which the output is completely independent of the input.

With this, it is clear that the total information loss in a QLM with $M > N$ takes its maximum value of $-\Delta I = H_L(X|Y) + H_L(Y) = H(X)$ for QLMs in which all output states are faithful but completely indistinguishable ($f_L = 1, \mathcal{F}_L = 0$), i.e. when all input information is lost. On the other hand, the information loss is minimized at $-\Delta I = 0$ when all output states are completely faithful but are perfectly distinguishable from one another ($f_L = 0, \mathcal{F}_L = 1$), i.e. when no input information is lost. For an ideal $L$-machine ($f_L = 1, \mathcal{F}_L = 1$) with $M > N$, the information loss takes the intermediate value of $-\Delta I = H_L(X|Y)$, precisely that associated with the corresponding noiseless classical computation channel.

We now consider entropy generation and energy flow in quantum $L$-machines and their relationship to information loss, representational faithfulness, and computational fidelity.

5. Logical and physical irreversibility in generalized $L$-machines

The ideal classical $L$-machine, employed by Ladyman in Ref. [1] to help clarify issues surrounding both realism about computation and the connection between logical and thermodynamic irreversibility, was initially devised by LPSG specifically for the latter purpose. In Ref. [2], LPSG presented classical thermodynamic and information-theoretic analyses of ICM to support their qualitative statement of Landauer’s Principle (LP), which is the claim that an $L$-machine implementing a logically irreversible logical transformation is necessarily thermodynamically irreversible (see also [3]). Interest in such claims remains timely nearly half a century after Landauer’s original statement of LP [27,28], both because their validity continues to be questioned on fundamental grounds [29–32] and because the standard “energetic form” of LP implies a lower bound on heat dissipation in computation that is of critical relevance for the future of conventional computing technology: Extrapolation of the historical downward trend in the heat dissipated per logical operation in silicon technology, as studied empirically over several decades by Keyes [33], suggests that Landauer’s fundamental limit will be encountered around the year 2020.

\(^{23}\) Additional discussion of the computational fidelity measure can be found in Ref. [7], which includes consideration of decoherence effects and an example calculation. Note that, because decoherence can never increase the Holevo information associated with an ensemble of quantum states, decoherence can never increase the computational fidelity of a quantum $L$-machine and can never decrease the representational faithfulness or information loss (cf. Eqs. (45), (50) and (51), respectively).
In this section, we obtain bounds on entropy generation and energy flow in general quantum \( L \)-machines. We first show that the total von Neumann entropy generated by implementation of a logical transformation in a QLM is lower bounded by the information loss, preserving the standard “entropic form” of LP in this context. This bound establishes that logical irreversibility indeed requires the generation of entropy in quantum \( L \)-machines, albeit in a sense that is distinctly different from that of LPSG’s statement of LP for reasons that will be discussed. We then obtain an additional lower bound, also dependent on the information loss, on the amount of energy that flows to a quantum \( L \)-machine’s environment during its evolution. The bound is very general in that it assumes that the environment is initially in a thermal state at a well-defined temperature \( T \) without requiring the same of the final environment state or the initial or final device states. The standard “energetic” form of LP is recovered as a corollary of this bound only under the common (but unnecessary) assumption that the \( M \) processes \( \hat{D}^{(m)} \to \Lambda(\hat{D}^{(m)}) \) do not, on average, increase the device entropy. We conclude this section by emphasizing the distinction between “single-shot” implementation of logical operations by \( L \)-machines and full computational cycles, and the implications of this distinction for interpretation of bounds obtained both here and in [2].

5.1. Logical irreversibility, physical irreversibility and entropy generation

For LPSG, a classical \( L \)-machine is thermodynamically irreversible if, for at least one of the logical inputs \( x_i \), the corresponding process \( D^{(m)} \to \Lambda(D^{(m)}) \) increases the total entropy \( S^{\text{DE}}_\text{tot}(x_i) = S^D_i + S^E_i \) of \( \text{DE} \). They note that, in this sense, the thermodynamic irreversibility of an \( L \)-machine implementing a logically irreversible transformation can be established by demonstrating that the average entropy change

\[
\langle \Delta S^{\text{DE}}_\text{tot} \rangle = \sum_{i=1}^{M} p_i \Delta S^{\text{DE}}_\text{tot}(x_i)
\]

is strictly positive for such a machine, since this requires \( \Delta S^{\text{DE}}_\text{tot}(x_i) > 0 \) for at least one of the \( M \) processes. They present a classical thermodynamic analysis – based on a thermodynamic cycle in an ICLM – and an independent information-theoretic analysis of the critical “computational” part of this cycle, both of which aim to establish that \( \langle \Delta S^{\text{DE}}_\text{tot} \rangle > 0 \) in any ICLM implementing a logically irreversible transformation and thus prove their qualitative statement of LP.

Our information-theoretic analysis of the connection between logical and physical irreversibility in general QLMs differs fundamentally from that of LPSG. We identify the aspect of physical irreversibility in QLMs that is most relevant in the present context – namely the impossibility of restoring, by any local operation on \( D \) alone, the initial information-bearing correlations between \( R_{\text{in}} \) and \( D \) that are lost in implementation of a logically irreversible transformation – and consider the physical entropy that is necessarily generated when these correlations are irreversibly lost. Physical irreversibility, so defined, requires an increase in the total entropy \( S^{\text{R}_{\text{in}} \text{DE}}_\text{tot}(x_i) = S^{\text{R}_{\text{in}} \text{D}}+S^E \) of the ensemble of QLM states, as we demonstrate below, but does not require an increase in the average total entropy \( \langle \Delta S^{\text{DE}}_\text{tot} \rangle \) of \( \text{DE} \). For this reason, we regard \( \Delta S^{\text{R}_{\text{in}} \text{DE}}_\text{tot} \) as the entropic quantity relevant to logical irreversibility in QLMs and seek information-theoretic lower bounds on this quantity. In obtaining our bound on entropy generation in a general \( L \)-machine, we define the information-bearing subsystem to be the composite \( R_{\text{in}} \text{D} \) rather than the device \( D \) alone. This implicitly grants the input referent the “status” of a real physical system, thus capturing a realistic feature of computation that is reflected in our referential approach. We will require that \( R_{\text{in}} \) and \( D \) remain isolated from one another throughout evolution of the \( L \)-machine, as noted in Section 2.3, which together with the assumptions of a closed composite \( R_{\text{in}} \text{DE} \) and a stationary referent state, implies global unitary evolution of \( R_{\text{in}} \text{DE} \) from its initial state \( \hat{\rho}^{\text{R}_{\text{in}} \text{DE}} \) to its final state \( \hat{\rho}^{\text{R}_{\text{in}} \text{DE}} \) via

\[
\hat{\rho}^{\text{R}_{\text{in}} \text{DE}'} = \hat{U} \hat{\rho}^{\text{R}_{\text{in}} \text{DE}} \hat{U}^\dagger
\]

with

\[
\hat{U} = \hat{I}^{\text{R}_{\text{in}}} \otimes \hat{U}^{\text{DE}}
\]

where \( \hat{U}^{\text{DE}} \) is a unitary operator governing evolution of \( \text{DE} \). Provided that the initial state is separable, i.e. \( \hat{\rho}^{\text{R}_{\text{in}} \text{DE}} = \hat{\rho}^{\text{R}_{\text{in}} \text{D}} \otimes \hat{\rho}^E \), the increase

\[
\Delta S^{\text{R}_{\text{in}} \text{DE}}_\text{tot} = \left[ S(\hat{\rho}^{\text{R}_{\text{in}} \text{D}'}) + S(\hat{\rho}^E') \right] - \left[ S(\hat{\rho}^{\text{R}_{\text{in}} \text{D}}) + S(\hat{\rho}^E) \right]
\]

24 We note that a nonincreasing \( \langle \Delta S^{\text{DE}}_\text{tot} \rangle \) is not contradicted by quantum generalization of LPSG’s bound, provided in Appendix B, which fails to establish the strict positivity of this quantity in QLMs. It fails for reasons that would also seem to apply to LPSG’s classical information-theoretic proof of their bound (but not their thermodynamic proof, which is obtained independently).

25 The preparation of input states of a real computing device is, after all, conditioned on something external to the device, which is the input referent. The existence of such a referent is required by an observer verifying the operation of the device, as discussed in Footnote 15 of Section 3, and is assumed even by a “trusting user” of a device who does not have access to this referent.
in the total entropy is lower bounded as \( \Delta S_{tot}^{R_{in}D_E} \geq -I \). In thermodynamic units \( ^{(27)} \) (denoted by the tilde), this inequality takes the more familiar form

\[
\Delta S_{tot}^{R_{in}D_E} \geq -k_B \ln(2) \Delta I
\]

where \( k_B \) is Boltzmann’s constant. This bound captures the fact that any loss of correlation between \( R_{in} \) and \( D \) — i.e., any loss \( -\Delta I \) of information about \( R_{in} \), that is held in \( D \) — necessarily creates correlations between \( R_{in}D \) and \( E \) that correspondingly increase the total entropy. This entropy bound, which mirrors the standard “entropic form” of LP, thus reflects the “reallocation” of correlations accompanying information loss in a quantum \( L \)-machine — correlations that cannot be captured by consideration of \( DE \) alone. It is not, however, specifically a bound on thermodynamic entropy, and its connection to thermodynamic bounds requires further investigation.

Writing the information loss in terms of our efficacy measures (Eq. (53)), we finally have

\[
\Delta S_{tot}^{R_{in}D_E} \geq k_B \ln(2) [f_I H_I(X|Y) + (1 - f_I) H_L(Y)].
\]

This expression lower bounds the entropy cost of implementing a logical transformation \( L \) with a specified efficacy, and clearly reveals a positive entropy cost for implementation of a logically irreversible transformation (i.e. \( H_I(X|Y) > 0 \)) with \textit{any} degree of representational faithfulness (\( f_I > 0 \)). This bound simplifies to \( \Delta S_{tot}^{R_{in}D_E} \geq k_B \ln(2)f_I H_I(X|Y) \) for noiseless QLMs and further to \( \Delta S_{tot}^{R_{in}D_E} \geq k_B \ln(2)H_I(X|Y) \) for ideal QLMs.

5.2. Energy flow

We now obtain a lower bound on the average energy that flows to a QLM’s environment as a result of generally noisy and unfaithful implementation of a logical transformation \( L \) by the machine. Recalling that only \( D \) interacts with \( E \), and that the joint evolution of \( DE \) is unitary, we can make use of a result from quantum thermodynamics that is due to Partovi \( ^{(35)} \): In a unitarily evolving composite quantum system \( DE \) in which the environment \( E \) is \textit{initially} in a thermal state at temperature \( T \), any decrease in the von Neumann entropy \( S^D \) of the device results in an increase in the expected energy of the environment that is lower bounded as

\[
\Delta \langle E^E \rangle \geq -k_B T \ln(2) \Delta S^D
\]

with

\[
\langle E^E \rangle = \text{Tr} \left[ \hat{\rho}^E \hat{H}^E \right] = \sum_{i=1}^{M} p_i \text{Tr} \left[ \hat{\rho}_i^E \hat{H}^E \right] = \sum_{i=1}^{M} p_i \langle E^E_i \rangle = \langle \langle E^E_i \rangle \rangle
\]

where \( \hat{H}^E \) is the environment Hamiltonian.

We first consider the entropy reduction

\[
-\Delta S^D = S(\hat{\rho}^D) - S(\hat{\rho}^{D'})
\]

of the device. Writing the initial device entropy as

\[
S(\hat{\rho}^D) = H(X) + \sum_{i=1}^{M} p_i S(\hat{D}_i^{(in)}),
\]

(since the initial device states \( \hat{D}_i^{(in)} \) are mutually orthogonal) and using Eqs. (43) and (45), we can write this as

\[
-\Delta S^D = -\Delta T - \langle \Delta S^D_i \rangle
\]

where \( -\Delta T \) is the information loss and

\[
\langle \Delta S^D_i \rangle = \sum_{i=1}^{M} p_i \left( S(A(\hat{D}_i^{(in)}) - S(D_i^{(in)}) \right)
\]

is the average device entropy reduction. Substitution into Partovi’s inequality (with Eq. (61)) yields

\[
\langle \Delta \langle E^E_i \rangle \rangle \geq k_B T \ln(2) \left( -\Delta T - \langle \Delta S^D_i \rangle \right).
\]

Thus, implementation of logically irreversible operation \( (-\Delta T > 0) \) by a quantum \( L \)-machine necessarily increases the average expected energy of the environment \textit{provided that} the entropy of the representative states, on average, increases by an amount less than the information loss as the device evolves from its initial to final state. This condition is met in the cases

\[^{(26)} \text{This follows from the entropy bound (12) of Ref. [34].} \]

\[^{(27)} \text{The von Neumann entropy of density operator } \hat{\rho} \text{ in thermodynamic units is } \tilde{S}(\hat{\rho}) = -k_B \text{Tr}[\hat{\rho} \ln \hat{\rho}] = k_B \ln(2)S(\hat{\rho}), \text{ which differs from the corresponding “information-theoretic” definition by a factor of } k_B \ln(2).} \]
most commonly studied in the literature, but is by no means necessary in a “single-shot” implementation of $L$ by a QLM. (See discussion in Section 5.3) Note that when the average entropy of $\mathcal{D}$ is nonincreasing, as is most commonly assumed, we have as a corollary

$$\langle \Delta (E_i^L) \rangle \geq -k_B T \ln(2) \Delta \mathcal{D}.$$  

(67)

This result has the standard “energetic form” of LP, which, like its entropic counterpart, is nearly ubiquitous despite the wide variety of definitions for energy, entropy, information, and even temperature that have been used to investigate LP in various contexts.

Our energy bound can finally be written in terms of the computational efficacy as

$$\langle \Delta (E_i^L) \rangle \geq k_B T \ln(2) \left( f_i H_i(X|Y) + (1 - f_i) H_i(Y) - \langle \Delta S_{\rightarrow_1}^L \rangle \right)$$

(68)

which lower bounds the average flow of energy to the environment explicitly in terms of the classical information-theoretic characterization of the desired logical operation $L$, the representational faithfulness and fidelity with which this operation is implemented by the QLM, and the average entropy change of the representative states during evolution. We note that our proof of this bound does not require assignment of thermodynamic entropy to probabilistic mixtures of macrostates, which has been a source of controversy in the recent literature (e.g. [2,31,36]), and is also independent of the entropy bound we obtained in Section 5.1.

5.2. Discussion

The bounds obtained above apply specifically to the physical costs of “single-shot” implementation of a logical transformation $L$ by a QLM, as is the case for LPSGs classical analysis.\(^{28}\) Realistic information processing scenarios typically involve sequences of computational cycles in which the physical process that implements $L$ is preceded by preparation of the appropriate input state and followed by a readout operation. The preparation step requires that the device state is brought into correlation with an appropriate external referent $\mathcal{R}_m$ holding (unknown) input data, overwriting whatever configuration of $\mathcal{D}$ remains from the previous computational cycle and “setting” $\mathcal{D}$ to the appropriate representative state $\hat{\mathcal{D}}^{(\text{in})}_i$ for the new cycle.\(^{29}\) The readout operation is essentially a measurement, which can disturb the final device state in classical or quantum machines and necessarily causes such a disturbance in some noisy quantum machines. These steps generally exact physical costs in addition to those considered here and must be accounted for in comprehensive analysis of machines that perform sequences of computational cycles. We call attention to these issues to emphasize that the bounds obtained here (and in [2]) apply only to physical costs incurred the “computational stroke” of a given cycle – the physical process that implements $L$ – not to complete computational cycles. Evaluation of physical costs associated with full computational cycles will be considered within our referential approach elsewhere (e.g. [37]).

6. Summary and conclusions

In this work, we have introduced and analyzed the quantum $L$-machine (QLM). The QLM represents a full quantum generalization of the ideal classical $L$-machine (ICLM) introduced previously by Ladyman and co-workers [1–3]. Using our quantum generalization, we reconsidered fundamental issues concerning the physical implementation of logical transformations that were recently analyzed by Ladyman [1] using an account of computation rooted in ICLM. Our formal definition of a quantum $L$-machine (QLM) (Definition 5) was deliberately constructed so it retains as much of the structure and spirit of Ladyman’s ideal classical definition (ICLM) as is possible in the more general context. This highlights the fundamental differences between the two types of machines, particularly those differences that admit unfaithful and noisy implementations of logical transformations into QLM, while preserving much of the clarity with which ICLM describes the relationship between the abstract-logical and concrete-physical aspects of computation.

We began by defining quantum $L$-machines and identifying relevant classes of these machines, paying particular attention to the physical representation of logical output states in these machines and discussing the fundamental problems that indistinguishability of these states pose for computational realism. We then identified the class of machines – logically irreducible QLMs (Definition 9) – that objectively implement particular “transformation structures” and can thus be unambiguously regarded as physical implementations of particular logical transformations. Next, we provided a full physical-information-theoretic characterization of the implementation of logical transformations in QLMs and discussed the underlying “referential” approach on which it is based. We identified two quantitative efficacy measures – the representational faithfulness $f_i$ (Definition 10) and the computational fidelity $f_{\mathcal{L}}$ (Definition 11) – that together capture how well a quantum system, used in a particular way, implements a classical logical transformation $L$. These measures depend

\(^{28}\) Note that the classical thermodynamic cycle considered by LPSG was used to explore the irreversibility of the specific step that implements $L$, i.e. the irreversibility of “single-shot” implementation of $L$. Their thermodynamic cycle was not presented as a computational cycle, and does not generally correspond to such a cycle (except, perhaps, for ideal $L$-machines implementing the single-output transformation RESET).

\(^{29}\) Note that, for simplicity, analyses of computational cycles often assume decomposition of the overwriting step into a RESET operation – which erases the results of the previous computation from $\mathcal{D}$ – and a subsequent WRITE operation that prepares $\mathcal{D}$ in a new representative input state.
only on classical information-theoretic entropies associated with the logical transformation \( L \) and entropic measures that quantify various correlations between the initial and final quantum states of the computing device and physical records of the corresponding logical input and output states for \( L \) instantiated in the state of a referent system. The information loss \(-\Delta \mathcal{F}\) in a QLM, and thus the physical irreversibility of a QLM implementing logically irreversible operations, is shown to depend on the efficacy measures in a simple and heuristically satisfying way (Eq. (53)). The connection between logical and physical irreversibility was considered within our approach, and a lower bound on entropy generation was obtained for a general QLM that is expressed in terms of the faithfulness and fidelity with which the quantum machine implements a specified \( L \) (Eq. (59)). The entropy generation is lower bounded by the locally irreversible loss of correlations between the state of the computing device and a physical record of the input data encoded in a referent system – the signature of implementation of a logically irreversible transformation in our approach – upholding the qualitative statement of Landauer’s principle and taking its standard quantitative form as well. We also obtained a lower bound on the flow of energy into a QLM’s environment that is expressed in terms of the efficacy with which a logical transformation \( L \) is implemented and the average change in the device entropy resulting from evolution of machine (Eq. (68)). These bounds were compared and contrasted with standard statements of Landauer’s Principle and with analyses of logical and physical irreversibility in ICLMs by Ladyman and co-workers.

The quantum \( L \)-machine introduced in this work represents a nontrivial generalization of its ideal classical counterpart. This generalization extends the notion of an \( L \)-machine to a much wider variety of realistic scenarios involving computation in physical systems, providing a correspondingly more comprehensive basis for the exploration of fundamental issues related to the physical implementation of logical operations in natural and artificial systems. The associated computational efficacy measures together provide a quantitative answer to the question of how well a given QLM implements a specified logical transformation \( L \), even in machines for which a realist conception of computation is not possible, by appropriately capturing “shades of grey” in representational faithfulness and computational fidelity associated with physical implementation of logical transformations. Even with these generalizations, however, the \( L \)-machine lacks features that will ultimately be required for comprehensive description of realistic information processing scenarios: It can describe only “single-shot” implementation of logical transformations, or, equivalently, only the “computational stroke” of a cycle that must be repeated over multiple uses if the machine is to operate on a random sequence of logical inputs. Comprehensive treatment of such scenarios requires a physical description of information processing that goes well beyond that offered by classical or quantum \( L \)-machines: It must be sufficiently rich in structure to include input and output systems among the information-bearing degrees of freedom, interactions between these systems and the computing device and environment in various steps (i.e. the interactions associated with LOAD and READ operations), and multi-component environments if it is to capture all relevant correlations and track all relevant entropy, energy and information flows. Such a description, based on a referential approach and quantitative efficacy measures similar to those used in this work, is currently under development.

**Acknowledgements**

The author gratefully acknowledges the National Science Foundation for partial support of this work (CCF-0649691) and Gualtiero Piccinini for valuable comments on the manuscript and discussions of related topics.

**Appendix A. Information loss and computational efficacy**

In this Appendix, we show that the information loss (Eq. (45)) can be written in terms of the computational efficacy measures \( f_l \) and \( f_r \) as in Eq. (53). By the grouping property of the Shannon entropy, the first term on the right-hand side of Eq. (45) can be rewritten as

\[
H(X) = H_j(Y) + \sum_{j=1}^{N} q_j H_j(X | y_j). \tag{A.1}
\]

The sum in the second term

\[
\chi(\mathcal{F}_j) = \mathcal{S} \left( \hat{\rho}_j^{\text{pr}} \right) - \sum_{i=1}^{M} p_i \mathcal{S} \left( \Lambda(\hat{D}_i^{\text{in}}) \right) \tag{A.2}
\]

on the right-hand side of Eq. (45) can be rewritten as

\[
\sum_{i=1}^{M} p_i \mathcal{S} \left( \Lambda(\hat{D}_i^{\text{in}}) \right) = \sum_{j=1}^{N} \left[ \sum_{i \in [j]} p_i \mathcal{S} \left( \Lambda(\hat{D}_i^{\text{in}}) \right) \right] = \sum_{j=1}^{N} q_j \left[ \sum_{i \in [j]} p_i^{(j)} \mathcal{S} \left( \Lambda(\hat{D}_i^{\text{in}}) \right) \right]. \tag{A.3}
\]
With this, and the addition and subtraction of the quantity \( \sum_{j=1}^{N} q_j \chi(\varepsilon_j^{D'}) \), we have

\[
\chi(\varepsilon_x^{D'}) = \chi(\varepsilon_x^{D'}) + \sum_{j=1}^{N} q_j \chi(\varepsilon_j^{D'})
\]

\hspace{\text{where}}

\[
\chi(\varepsilon_i^{D'}) = S(\hat{\rho}_i^{D'}) - \sum_{j=1}^{N} q_j S(\hat{\rho}_j^{D'})
\]

\[
\chi(\varepsilon_j^{D'}) = S(\hat{\rho}_j^{D'}) - \sum_{i \in |\mathcal{N}_j|} p_i^{(j)} S(\hat{\rho}_i^{D'})
\]

Substitution into Eq. (45) yields

\[
-\Delta I = H_I(Y) + \sum_{j=1}^{N} q_j H_I(X|y_j) - \chi(\varepsilon_Y^{D'}) - \sum_{j=1}^{N} q_j \chi(\varepsilon_j^{D'})
\]

\[
= H_I(Y) - \chi(\varepsilon_Y^{D'}) + H_I(X|Y) - \sum_{j=1}^{N} q_j \chi(\varepsilon_j^{D'})
\]

or

\[
-\Delta I = \left(1 - \frac{\chi(\varepsilon_Y^{D'})}{H_I(Y)}\right) H_I(Y) + \left[1 - \frac{1}{H_I(X|Y)} \sum_{j} q_j \chi(\varepsilon_j^{D'})\right] H_I(X|Y)
\]

which, by Eqs. (50) and (52), is Eq. (53).

**Appendix B. Generalized LPSG entropy bound**

In this Appendix, we generalize the information-theoretic approach used by LPSG [2] to lower bound the average increase \(\langle \Delta S_{i}^{D,E} \rangle = \sum_{i=1}^{M} p_i \Delta S_{i}^{D,E}(x_i)\) in the total (device + environment) thermodynamic entropy for the \(M\) processes \(D_i^{(in)} \rightarrow A_1(D_i^{(in)})\) in ICLM. The analogous average von Neumann entropy increase for a QLM is

\[
\langle \Delta S_{i}^{D,E} \rangle = \sum_{i=1}^{M} p_i \left( S(A(D_i^{(in)})) - S(D_i^{(in)}) + \Delta S_i^{E} \right)
\]

where \(\Delta S_i^{E} = S(\hat{\rho}_i^{E'}) - S(\hat{\rho}_i^{D'})\) is the entropy increase of the environment for the \(i\)-th process. We can write the increase

\[
\Delta S_{i}^{D,E} = S(\hat{\rho}_i^{D'}) + S(\hat{\rho}_i^{E'}) - S(\hat{\rho}_i^{D}) - S(\hat{\rho}_i^{E})
\]

in the total entropy \(S_{i}^{D,E} = S(\hat{\rho}_i^{D}) + S(\hat{\rho}_i^{E})\) of the ensemble of processes as

\[
\Delta S_{i}^{D,E} = S \left( \sum_{i=1}^{M} p_i A(D_i^{(in)}) \right) - H(X) - \sum_{i=1}^{M} p_i S(D_i^{(in)}) + \Delta S_i^{E}
\]

where \(\Delta S_i^{E} = S(\hat{\rho}_i^{E'}) - S(\hat{\rho}_i^{D'})\) and the mutual orthogonality of the input states has been used. Rearranging Eq. (B.3), substituting into (B.1), and simplifying via Eqs. (43) and (45), we have

\[
(\Delta S_{i}^{D,E}) = -\Delta I + \Delta S_{i}^{D,E} - \left(\Delta S_i^{E} - \sum_{i=1}^{M} p_i \Delta S_i^{E}\right)
\]

The total entropy change \(\Delta S_{i}^{D,E}\) is nonnegative: Rewriting the final ensemble entropy as

\[
S(\hat{\rho}_i^{D'}) + S(\hat{\rho}_i^{E'}) = S(\hat{\rho}_i^{D,E'}) + I(\hat{\rho}_i^{D'}; \hat{\rho}_i^{E'}) = S(\hat{\rho}_i^{D,E}) + I(\hat{\rho}_i^{D'}; \hat{\rho}_i^{E'})
\]

\[
= S(\hat{\rho}_i^{D}) + S(\hat{\rho}_i^{E}) + I(\hat{\rho}_i^{D'}; \hat{\rho}_i^{E'}),
\]

\(\text{The first, second and third equalities follow from the definition of the quantum mutual information} I(\hat{\rho}_i^{D'}; \hat{\rho}_i^{E'}), \text{the invariance of entropy under unitary evolutions, and the initial separability of} D,E, \text{respectively.}\)
and substituting into (B.2) to obtain $\Delta S_{\text{tot}}^{DE} = I(\hat{\rho}^D; \hat{\rho}^E)$, $\Delta S_{\text{tot}}^{DE} \geq 0$ follows from the nonnegativity of the quantum mutual information. (This is the “Second Law” for von Neumann entropies in this context.) We thus have

$$\langle \Delta S_{\text{tot}}^{DE} \rangle \geq -\Delta I \equiv \left( \Delta S^E - \sum_{i=1}^{M} p_i \Delta S_i^E \right). \tag{B.6}$$

The strict positivity of $\langle \Delta S_{\text{tot}}^{DE} \rangle$ forQLMs with $-\Delta I > 0$ would clearly follow from the inequality (B.6) if $\Delta S^E = \sum_{i=1}^{M} p_i \Delta S_i^E$ for $-\Delta I > 0$ or equivalently if $S(\rho^E) = \sum_{i=1}^{M} p_i S(\rho_i^E)$ for $-\Delta I > 0$ – but this does not hold in general: The quantity $\sum_{i=1}^{M} p_i S(\rho_i^E)$ is, by the convexity of the von Neumann entropy, a lower bound for $S(\rho^E)$, and is equal to $S(\hat{\rho}^E)$ only if the environment state does not gain information about the initial device state as the device–environment composite evolves in time.

However, LPSG assumed equality in the analogous bound in their classical information-theoretic analysis (cf. Eq. (23) of [2]), taking the entropy of the mixture of final environment states to be the average of the component entropies. The validity of LPSG’s information-theoretic proof of LP hinges on this unexplained assignment of entropy to the mixture of environment states, which is inconsistent with their assignment of entropy to the mixture of initial device states in the very same proof (cf. Eqs. (1) and (21) of [2]).

References


[27] Here $S(\hat{\rho}^D) = S(\hat{\rho}^E)$ ∀i, and thus $S(\hat{\rho}^E) = \sum_{i=1}^{M} p_i S(\hat{\rho}_i^E)$, since the initial environment state is independent of the initial device state.