Irreversibility in unitary quantum homogenisation: Theory and Experiment

Chiara Marletto and Vlatko Vedral
Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom and Fondazione ISI, Via Chisola 5, Torino, Italy and Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543 and Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117542

Laura Knoll, Fabrizio Piacentini, Ettore Bernardi, Alessio Avella, Marco Gramegna, and Ivo Pietro Degiovanni
Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce 91, 10135, Torino, Italy

Enrico Rebufello
Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy and Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce 91, 10135, Torino, Italy

Marco Genovese
Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce 91, 10135, Torino, Italy and, INFN, sezione di Torino, via P. Giuria 1, 10125 Torino, Italy

(Dated: October 1, 2020)

PACS numbers: 03.67.Mn, 03.65.Ud

Is it possible to fit irreversibility exactly in a universe whose laws are time-reversal symmetric? This vexed open problem, dating back to the origin of statistical mechanics [1–3], is crucial for the realisability of logically irreversible computations such as erasure [4–6]. A well-known type of irreversibility is epitomised by Joule's experiment [2], involving a volume of fluid and a stirrer powered by mechanical means, e.g. a suspended weight: while it is *possible* **to construct a cycle heating up the water** *by mechanical means only***, it is** *impossible* **to cool the water down via a cycle that likewise utilises mechanical means only. This irreversibility is of particular interest because, unlike others, it is based on a system that performs some transformation, or task, operating in a cycle. First introduced via a thermodynamic cycle (e.g. Carnot's), this idea was generalised by von Neumann to a** *constructor* **- a system that can perform a given task on another system and crucially retains the ability to perform the task again. On this ground, a task is** *possible* **if there is no limitation to how well it can be performed by some constructor; it is impossible otherwise. Here we shall define a "***constructor-based* **irreversibility", generalising Joule's, requiring that a task T is possible, while the transpose task (where T's input and output are switched) is not. We shall demonstrate the compatibility of this constructor-based irreversibility with quantum theory's time-reversal symmetric laws, confining attention to an example using the universal quantum homogeniser [9, 10]. This result is an essential further step for understanding irreversibility within quantum theory, one of the most fundamental theories of physics we posses at present. We will also simulate the irreversibility experimentally with high-quality photon qubits.**

To define the constructor-based irreversibility, we use *constructor theory*, a recently proposed framework for generalising the quantum theory of computation to cover general tasks, [7, 8]. For our quantum model, we shall consider a universe made of qubits, where all unitary transformations and their transposes are allowed. Also, we shall make the simplifying assumption that qubits are available in unbounded numbers, in any state. Interestingly, we shall see that even in this simplified scenario, constructor-based irreversibility arises. We do not lose generality because this assumption makes our task of proving the claim of irreversibility harder than for the real universe.

A task T is the specification of a physical transformation on qubits. We will confine attention on tasks transforming one quantum state ρ_x to another quantum state ρ_y:

$$T = \{\rho_x \to \rho_y\} \tag{1}$$

whose transpose $T^\sim$ is simply defined as:

$$T^\sim = \{\rho_y \to \rho_x\} \ . \tag{2}$$

We will refer to the *substrate* qubit on which the task T is defined as Q and to the *rest* of the qubits as R. A **constructor** for the task T on a qubit Q is some subsystem of the rest that enables T, *without undergoing any net change in its ability* *to do it again*, i.e., it can be thought of as a cycle. A task is *possible* if the laws of physics do not put any constraint on the accuracy to which it can be performed by a constructor, as defined above. It is *impossible* otherwise. A familiar example of a possible task is the task of cloning any state drawn from a set of mutually orthogonal states. An impossible task is that of cloning a state drawn from a set of two non-orthogonal states, because there is a fundamental limitation to how well such a task can be performed, which is a function of the inner product of the states, [11].

Constructor-based irreversibility will be defined here as the fact that while the task T is possible, its transpose $T^{\sim}$ is not possible. The second law in one of its traditional formulations can also be expressed via a statement of this kind, [12]: this is a long-standing tradition in thermodynamics, initiated by Planck and continuing with axiomatic thermodynamics [13, 14]). Specifically, the second law can be stated as requiring that a cycle (a constructor, in our language) that converts completely work into heat is possible, but a cycle that performs the transpose task is not. Here we provide a quantum model for constructor-based irreversibility and show how it is compatible with quantum theory's time-reversal symmetric laws, but without specialising to thermodynamics: we will consider unitary dynamics with no additional conservation laws; furthermore, there will be no need to assume there being a temperature or thermodynamics equilibrium. Showing compatibility will entail proving that under unitary quantum theory T being possible does not necessarily imply that $T^{\sim}$ is also possible.

We express now the conditions for a *constructor* for the task T to be allowed under our unitary quantum model. To do so, we need to introduce some notation. A unitary transformation acting on both Q and the rest will be denoted by $U_{Q,R}$.

For a fixed task T on Q, a positive ϵ, and a given unitary U representing the dynamics of the whole universe, define the set of quantum states of R that can perform T to accuracy ϵ:

$$V_\epsilon[T] \doteq \{\rho_R \;:\; U(\rho_x \otimes \rho_R)U^\dagger = \rho \;,\; \mathrm{Tr}_R[\rho] \in \epsilon(\rho_y)\} \quad (3)$$

where $\epsilon(\rho_y)$ is the ϵ-ball centered around T's desired output state, ρ_y: $\epsilon(\rho_y) \doteq \{\sigma : F(\rho_y, \sigma) \geq 1-\epsilon\}$, being $F(\cdot,\cdot)$ is the quantum fidelity.

We shall denote with $\mathcal{E}[T]$ a set of qubits prepared in a state belonging to V_ϵ: it can be thought of as a machine capable of performing the task T with an error ϵ.

Let us now introduce a measure of how much a given $\mathcal{E}[T]$ stays unchanged in its property of performing T to accuracy ϵ after n repeated usages, on n fresh substrate qubits $Q_1, ..., Q_n$ each in the input state ρ_x. Define, for a given initial state of the rest ρ_R, the reduced operator:

$$\rho_R^{(n)} = \mathrm{Tr}_{Q_n \ldots Q_1}[\mathcal{U}_{Q_n,R} \cdots \mathcal{U}_{Q_1,R}(\bigotimes_{k=1}^{n} \rho_x^{(k)} \otimes \rho_R)\mathcal{U}_{Q_1,R}^\dagger \cdots \mathcal{U}_{Q_n,R}^\dagger] \;,$$

being $\mathcal{U}_{Q_j,R} = U_{Q_j,N} \cdots U_{Q_j,1}$ the operator mediating the interaction between the j-th substrate qubit and the N rest qubits.

We can then define the worst-case deterioration of the machine $\mathcal{E}[T]$ after performing the task n times:

$$\delta_{\mathcal{E}[T]}(n) \doteq \mathrm{Sup}_{\rho_R \in V_\epsilon[T]}\{F(\rho_R, \rho_R^{(n)})\} \;.$$

Most machines lose the ability to perform the task with repeated use. So we expect $\delta_{\mathcal{E}[T]}(n)$ to decrease with n, for a fixed ϵ. How well $\mathcal{E}[T]$ retains its ability to perform the task over and over again, as n increases and ϵ decreases, is crucial to whether it is a constructor for T.

A figure of merit for the resiliency of $\mathcal{E}[T]$ is the *relative deterioration* of $\mathcal{E}[T]$ after having been used n times, defined as:

$$R_{\mathcal{E}[T]}(n) \doteq \frac{\epsilon}{\delta_{\mathcal{E}(T)}(n)} \;.$$

There are two conditions for a constructor for the task T to be allowed under a given unitary law U.

Condition (i). For any positive arbitrarily small ϵ, the set $V_\epsilon[T]$ is non-empty. (This means that the rest can perform the task T to arbitrarily high accuracy, once.)

Condition (ii). The relative deterioration of the rest in its ability to perform the task T with error ϵ goes to zero, as the error ϵ goes to zero and the number n of repeated usages goes to infinity:

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} R_{\mathcal{E}[T]}(n) = 0 \;.$$

If these two conditions are both satisfied, then a sequence of machines $\mathcal{E}[T]$ converges to a limiting machine that perfectly retains the ability to cause the transformation with asymptotically small error, even when used n times, for arbitrarily large n. The limiting point of the sequence of machines $\{\mathcal{E}[T]\}$ is a constructor for the task T [30]

In general, for the task T to be possible, it is not enough that a constructor is allowed (in the above sense) under the given laws: there should be no limitation to how well a constructor could be approximated, *given the supplementary conditions*. Thus that a task is possible is a constraint *on both the supplementary conditions and on the dynamical laws*. Under our simplifying assumptions that a constructor is allowed implies that the task is possible.

A QUANTUM MODEL FOR CONSTRUCTOR-BASED IRREVERSIBILITY: HOMOGENISATION MACHINES

We will now provide a physical model where the dynamical laws are time-reversal symmetric (unitary quantum theory), but the fact that a task $T = \{\rho_x \to \rho_y\}$ is possible does not necessarily imply that its transpose task $T^{\sim}$ is possible. This will show the compatibility of time-reversal symmetric laws with the constructor-based irreversibility.

This model utilises the powerful scheme of quantum homogenisation [9, 10], depicted in Fig. 1 and defined as follows. Consider N qubits, each prepared in the state ρ_y. Let us call it $H_N[\rho_y]$. Suppose the qubit Q is initialised with the state ρ_x and then it interacts with the qubits in $H_N[\rho_y]$ one at a time, via a unitary transformation

$$U_{Q,k} = \cos(\eta)I + i\sin(\eta)\Sigma_{Qk}$$

where Σ_{Qk} is the swap gate on qubit Q and qubit k in the homogenisation machine, and I is the identity (see the appendix for interesting properties of the SWAP). This is a *partial swap*, whose intensity is defined by the real parameter η. The closer η is to $\frac{\pi}{2}$, the closer this transformation is to a standard SWAP. It is a way of slightly modifying the original state of the qubit, to become closer and closer to the desired output state ρ_y. See the figure. The state of the qubit Q after interacting with

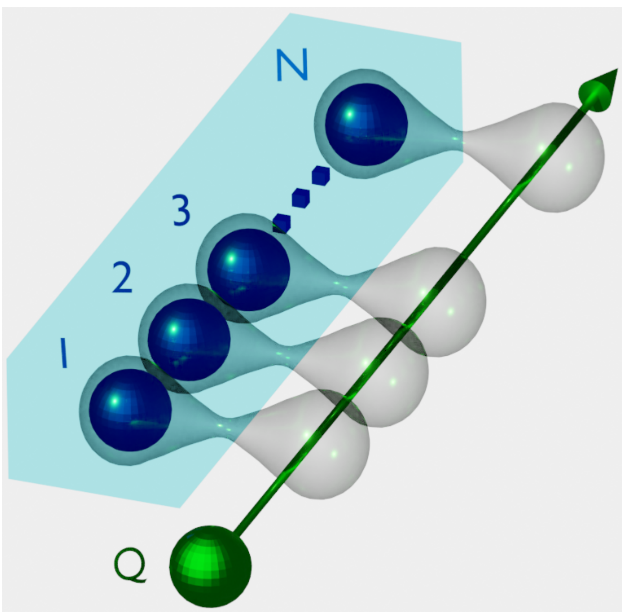

FIG. 1: A qubit Q (in green) interacting with a homogenisation machine (azure shaded box) composed of N qubits (in blue). The grey hourglasses represent each the unitary partial swap $U_{Q,k}$.

$H_N[\rho_y]$ is

$$\rho_{Q,N} = Tr_{1..N}\{U_{Q,N}...U_{Q,1}\left(\rho_x \otimes \rho_y^{\otimes N}\right)U^\dagger_{Q,1}...U^\dagger_{Q,N}\} .$$

Define now the error in performing the task on Q as

$$\epsilon_N = 1 - F\{\rho_{Q,N}, \rho_y\} .$$

One can show, [9, 10], that ϵ_N tends to zero as the number of qubits N in the machine tends to infinity

$$\lim_{N\to\infty} \epsilon_N = 0 .$$

In other words, the machine $H_N[\rho_y]$ tends to be capable of performing the task T perfectly when N is large, thus satisfying condition (i) for a constructor as defined above; and this is true for any T. In this case, $V_\epsilon[T]$ is a one-dimensional subspace spanned by the state $\rho_y^{\otimes N}$.

This is true for any task T transforming any quantum state ρ_x into any other state ρ_y. However, as we shall see, not all machines using quantum homogenisation satisfy condition (ii) - the resiliency condition: not all of them are constructors; hence, not all such tasks are possible. Specifically, as we shall prove, $T^\sim$ need not be, even if T is.

A special case. Consider the case where ρ_x is a pure state,

$$\rho_x = \frac{I + \mathbf{A}\cdot\sigma}{2}$$

and ρ_y is a maximally mixed state

$$\rho_y = \frac{I}{2} .$$

In this case, the task T goes in the direction of more mixedness, while $T^\sim$ goes in the opposite direction, purifying the state. It is possible to show two facts (see supplementary material).

1) The homogenisation machine $H_N[\rho_y]$ tends, as N increases, to being a constructor for the task T, because the relative deterioration after performing the task once tends to 0:

$$\lim_{N\to\infty}\lim_{n\to\infty} R_{H_{\epsilon_N}[T]}(n) = 0 .$$

2) The optimal candidate to perform $T^\sim$, the homogenisation machine $H_N[\rho_x]$, is *not* a constructor for $T^\sim$. Specifically, one can show that

$$\lim_{N\to\infty}\lim_{n\to\infty} R_{H_{\epsilon_N}[T^\sim]}(n) \not\to 0 .$$

Thus, the possibility of T and the assumption of time-reversal symmetric laws *does not imply* that $T^\sim$ must also be possible. This shows compatibility of time-reversal symmetric laws with constructor-based irreversibility under unitary quantum theory, as promised.

Intuitively, this irreversibility happens because, as N and n go to infinity, $\delta_{\mathcal{E}[T^\sim]}(n)$ increases faster than the error ϵ in performing the task is reduced. An intuitive explanation is that the initial state of the homogenisation machine $H_N[\rho]$, when ρ is pure, belongs to the so-called *symmetric subspace* of N qubits, as defined in [16]. While it is not so when ρ is a maximally mixed state. The symmetric subspace is more fragile to perturbations.

AN EXPERIMENTAL DEMONSTRATION WITH PHOTONS

In order to provide evidence of this mechanism at work, we verified experimentally that the homogenisation machine performing the task $T = \rho_p \to \rho_m$ (being ρ_p a pure state and ρ_m a mixed one) is more efficient with respect to the machine performing the transpose task $T^\sim = \rho_m \to \rho_p$. This represents a clear demonstration of the convergence properties, mentioned above, to show that, even though the underlying dynamics is time-reversal symmetric, $T^\sim$ may not be possible.

In summary, we consider two states: a pure state $\rho_p = |0\rangle\langle 0|$ and a quasi-maximally mixed state $\rho_m = \frac{1+\gamma}{2}|0\rangle\langle 0| + \frac{1-\gamma}{2}|1\rangle\langle 1|$ (with $\gamma \ll 1$, to take into account experimental imperfections in the mixture preparation), comparing the performance of the homogenisation machine $H_N[\rho_m]$ when the incoming substrate qubit is prepared in the state ρ_p (pure to mixed task) with the one of the machine $H_N[\rho_p]$ acting on an incoming qubit prepared in the state ρ_m (mixed to pure task) by measuring the accuracy of each machine in performing their task (i.e. the error ϵ).

This is achieved by implementing a single-photon-based experiment (Fig. 2). Single photons at 1550 nm are generated by a low-noise heralded single photon source [17] and sent to a 1×4 fiber optical switch which addresses them to four different optical paths, one for the substrate qubit and three for the machine ones. Each of the four outputs undergoes a preparation stage in which the photon is prepared in a pure (or mixed) time-bin state.

Then, the substrate and homogenisation machine photons paths are connected to a cascade of three consecutive fiber beam splitters (FBSs, all either 50:50, 76:24 or 90:10, in different configurations) implementing, between the substrate and rest photons, $N = 3$ subsequent partial swaps with swap parameter η ranging from 0.12 (weak interaction) to 1.46 (strong interaction). Finally, the four output channels of the FBSs cascade are sent to two free-running infrared single-photon avalanche diodes (SPADs), whose output signals are addressed to proper time-tagging electronics.

Fig. 3 shows the results obtained with the 50:50 FBS set, corresponding to $\eta = 0.78$. Plots (a) and (b) report, respectively, the progression of the tasks T (pure-to-mixed) and $T^\sim$ (mixed-to-pure) vs. the interaction time between substrate and homogenisation machine, indicated by the number of partial swaps occurred. The reconstructed diagonal elements ρ_{00} and ρ_{11} of the substrate state density matrix after each PS are reported (we remind that, by construction, in our case $\rho_{01} = \rho_{10} = 0$), considering for task T the substrate in the initial state $\rho_x = \rho_p = |0\rangle\langle 0|$ and the homogenisation machine qubits in the mixed state $\rho_m = 0.55|0\rangle\langle 0| + 0.45|1\rangle\langle 1|$, and viceversa for task $T^\sim$. As evident, the experimentally-reconstructed values of ρ_{00} and ρ_{11} (blue and red dots, respectively) are in very good agreement with the theoretical predictions (azure and yellow bars, respectively) throughout the whole process in both the cases. Plot (c) shows instead the growth of the square fidelity $F^2 = (Tr(\sqrt{\sqrt{\rho_t}\rho\sqrt{\rho_t}}))^2$

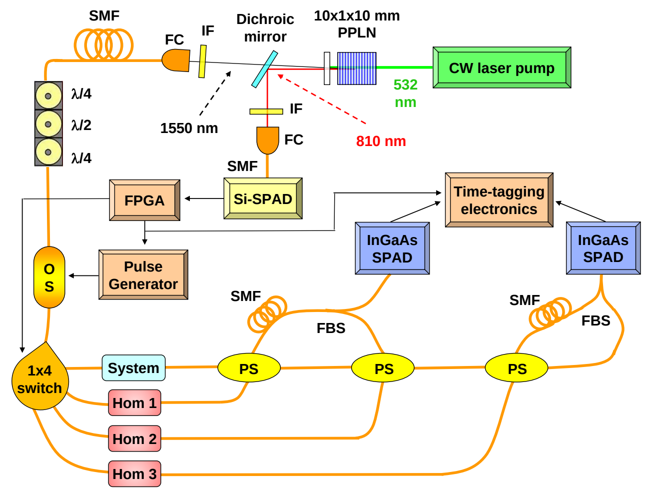

FIG. 2: Experimental setup. Heralded single photons at 1550 nm are produced by spontaneous parametric down conversion in a periodically-poled lithium niobate (PPLN) crystal pumped by a CW laser at 532 nm. The heralding is granted by the detection of a correlated photon at 810 nm by a silicon single-photon avalanche diode (Si-SPAD), triggering a field-programmable gate array (FPGA) operating a fast electro-optical shutter (OS). The OS opens for a 6 ns window only in correspondence of the arrival of a heralded photon, heavily reducing the optical noise in the source output. Afterwards, the heralded photon goes to a 1×4 fiber optical switch, addressing it either to the substrate path (Sub) or to one of the homogenisation machine paths (Hom1, Hom2, Hom3). The heralded single photons then meet a cascade of partial swaps (PS), realised by several sets of fiber beam splitters (each with a different splitting ratio) simulating the interaction between the substrate and the homogenisation machine. Finally, the photons are detected by two free-running indium/gallium-arsenide SPADs (InGaAs-SPADs), whose output is sent to a time-tagging electronics together with the FPGA-validated heralding counts. IF: interference filter. FC: fiber coupler. SMF: single-mode fiber. $\lambda/4$: quarter-wave fiber paddle. $\lambda/2$: half-wave fiber paddle. FBS: 50:50 fiber beam splitter.

between the substrate state ρ after each PS and the task target state ρ_t, being $\rho_t = \rho_m$ for the pure-to-mixed case and $\rho_t = \rho_p$ for the mixed-to-pure one. The experimental points, in good agreement with the theoretical expectations (solid lines) for both tasks, show how the homogenisation machine for the pure-to-mixed task T always outperforms the one for the reverse task $T^\sim$, giving a clear hint on the difficulty gap between the two tasks, at the origin of macroscopic irreversibility. The same holds for the error $\epsilon = 1 - F$ reported in plot (d), indicating the discrepancy level between the substrate state ρ at the end of the process and the target state ρ_t: although the performances difference between the two tasks seems to reduce further and further with time (as expected in this kind of scenario), the pure-to-mixed ϵ is always below the mixed-to-pure one.

By repeating the same analysis for different partial swap probabilities, we obtained the plots of Fig. 4, giving a dynamical picture of the performances of the task being executed by our

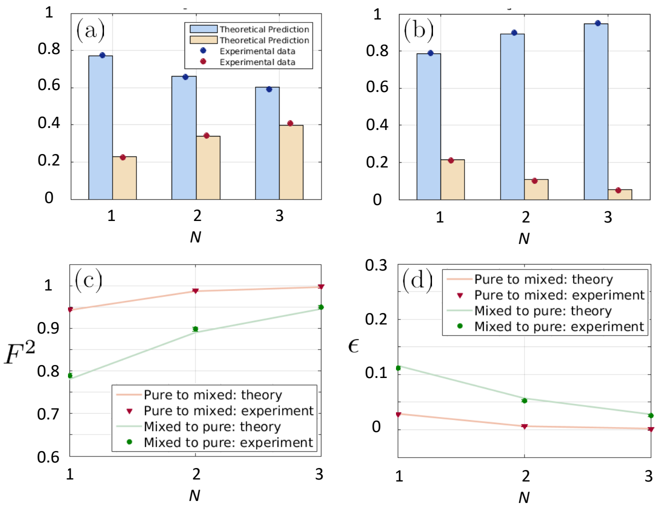

FIG. 3: Results obtained for PS parameter $\eta = 0.78$, realised by means of a set of three 50:50 FBSs. Panel (a): pure-to-mixed task T. The plot hosts the ρ_{00} (left side, in azure/blue) and ρ_{11} (right side, in yellow/red) elements of the density matrix of the substrate state, initially in the pure state $\rho_p = |0\rangle\langle 0|$, after each PS with one of the homogenisation machine qubits in the mixed state $\rho_m = 0.55|0\rangle\langle 0| + 0.45|1\rangle\langle 1|$ (experimentally measured), showing the progression of the substrate state evolution induced by the machine. The bars represent the theoretical predictions, while the dots indicate the experimentally-reconstructed values, reported with the associated experimental uncertainties. Panel (b): mixed-to-pure task $T^{\sim}$. The plot structure is the same as the one in panel (a), but here the substrate initial state is $\rho_m = 0.55|0\rangle\langle 0| + 0.45|1\rangle\langle 1|$ (as from experimental verification), while the homogenisation machine qubits are in the pure state $\rho_p = |0\rangle\langle 0|$. Panels (c) and (d), respectively: square fidelity F^2 and error ϵ between the task target state and the substrate state while the task is being executed (i.e. after each interaction with one of the homogenisation machine qubits), both in the pure-to-mixed case (in orange) and in the mixed-to-pure one (in green). The lines indicate the theoretical expectations, while the dots represent the experimental values.

3-qubit homogenisation machine: by looking at panels (a)-(c), one can see how, for each value of the swap parameter η, the error ϵ decreases as N increases, as expected. The three plots give a clear indication on how the homogenisation machine for task T performs better than its counterpart for $T^{\sim}$, reaching smaller error for every η value, especially in the cases of weak interaction ($\eta < 0.5$).

Concerning the relative deterioration $R_{H_{\epsilon_N}[\rho_t]}(n) \doteq \frac{\epsilon_N}{\delta_{H_{\epsilon_N}(\rho_t)}(n)}$ of the homogenisation machines implemented in our experiment, we evaluated the deterioration $\delta_{H_{\epsilon_N}(\rho_t)}(n)$ with a recursive method considering, for the j-th substrate state ($j = 1, ..., n$), the machine initialized in a state as close as possible to the one of the machine outgoing the $(j-1)$-th substrate-machine interaction. Although we were not able to directly measure the entanglement induced in the homogenisation machine qubits after their interaction with the substrate, we could nevertheless observe correlations among them. For each j, then, we reconstructed these correlations (in the computational basis $\{|0\rangle, |1\rangle\}$) and reproduced them while initializing the machine state for the $(j+1)$-th usage. The results for both tasks T and $T^{\sim}$ are reported in Fig. 5. The figure illustrates, for the two experimental η values corresponding to the weak interaction limit, the dependance of the relative deterioration on $n = N$, giving a hint of the asymptotic limit of $R_{H_{\epsilon_N}[\rho_t]}(n)$. In both plots, one can appreciate how the relative deterioration decreases, asymptotically reaching 0 in the limit $n, N \to \infty$, only for the pure-to-mixed task T, while for the reverse task $T^{\sim}$ it seems to be steadily growing. Our experimental results, obtained evaluating the fidelities of the detected events in the computational basis, are in good agreement with the theoretical predictions. Although the theoretical curves with our approximation of the homogenisation machine state deviate, for big n, N, from the ones obtained with the homogeniser entangled state of the complete theory (especially in the mixed-to-pure task), in our experimental scenario ($N \leq 3$) the solid and dashed curves are superimposed, demonstrating the reliability of our approximation. Furthermore, our numerical simulations show that the relative deterioration $R_{H_{\epsilon_N}[\rho_m]}(n)$ goes to 0 for $n, N \to \infty$, while $R_{H_{\epsilon_N}[\rho_p]}(n)$ reaches a non-zero numerical value: by definition, this excludes the possibility of $H_{\epsilon_N}[\rho_p]$ being a proper constructor, unequivocally demonstrating the irreversibility of the pure-to-mixed task T as performed by the homogenisation machine.

CONCLUSIONS

This paper has addressed the old-age problem of reconciling irreversibility with reversible unitary dynamics with a radically different approach, considering the irreversibility based on possibility/impossibility of tasks rather than on dynamical trajectories being permitted or disallowed. This notion of irreversibility extends the irreversibility of classical axiomatic thermodynamics to a general information-theoretic scenario. The problem is solved by using the conceptual framework of constructor theory, where the fact that a certain task T is possible does not imply that the transpose task $T^{\sim}$ is possible too. Here we address a specific example (with homogenisation machines), also providing an experimental demonstration of this mechanism at work. As shown by the results obtained in our single-photon-based experiment, the homogenisation machine implementing the pure-to-mixed task T always outperforms its counterpart for the reverse task $T^{\sim}$. Furthermore, by looking at the relative deterioration of both machines it is evident how the one for $T^{\sim}$ suffers much higher degradation than the one realising T, ultimately not satisfying condition ii and thus failing to qualify as a proper constructor: this gives a clear proof of the compatibility of the constructor-based irreversibility with unitary quantum theory, providing also a hint about the emergence of a thermodinamical arrow of time in quantum mechanics [20–25]. Another significant application of our results concerns quantum computational processes and

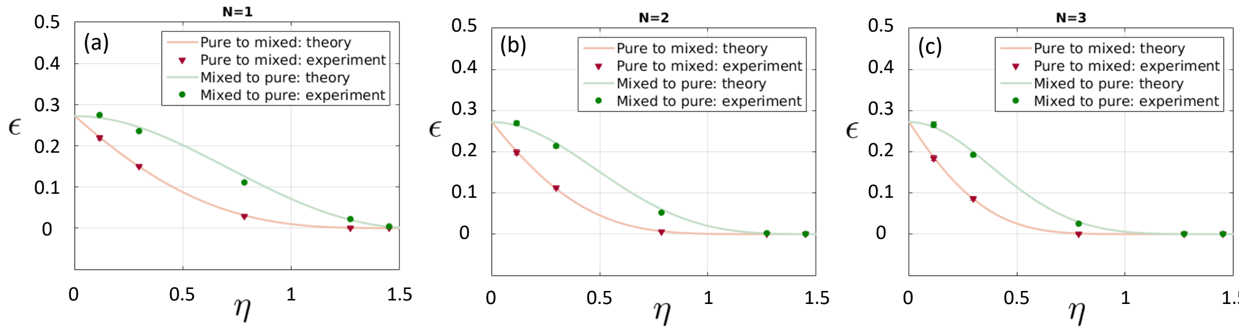

FIG. 4: Dynamics of the performances of our 3-qubit homogenisation machines for the pure-to-mixed task T (in orange) and its reverse, i.e. the mixed-to-pure one $T^{\sim}$ (in green). Panels (a), (b) and (c) show, respectively, the error ϵ vs. the PS parameter η at different moments of the tasks execution, i.e. for $N = 1, 2, 3$. The solid curves represent the theoretical predictions, while the dots indicate the experimental data, together with the associated uncertainties (usually too small to be appreciated).

their reversibility. Finally, there are intriguing further questions that this work opens up. First, it would be interesting to generalise this result in the context of the more general framework of constructor theory, without assuming a specific quantum theory model. Then, one could also recast this formal analysis via the resource theory of catalysis and of purity, [26], a powerful tool to analyse constructor-based irreversibility in a more general way. For example, we conjecture that the irreversibility would be present in any task sending a quantum state to a more mixed version of it. In the long term, it would be essential to explore the connection between this irreversibility and other kinds of irreversibility in physics: the statistical-mechanics irreversibility; and the irreversibility of quantum thermodynamics's second laws [27–29]. In order to assess this, one would have to find and entropy function (or a family of entropies, such as Rényi entropies) to quantify the constructor-based irreversibility, and then connect this with existing generalisations of the second laws in quantum thermodynamics. We leave all this to future work.

SUPPLEMENTARY MATERIAL

The swap is defined as the gate Σ_{12} with the property that

$$\Sigma_{12}\left|\psi\right\rangle\left|\phi\right\rangle = \left|\phi\right\rangle\left|\psi\right\rangle, \forall \left|\psi\right\rangle\left\langle\phi\right| .$$

It has (up to a phase) the following representation in terms of Pauli operators:

$$\Sigma_{12} = \frac{1}{2}(I_{12} + X_1X_2 + Y_1Y_2 + Z_1Z_2)$$

with $\Sigma_{12}^2 = I_{12}$.

Provided that the two subsystems have the same information-theoretic capacity (e.g. they are two qubits) the swap can also be realised by physically swapping the substrates or by performing three C-Not gates, as follows:

$$\Sigma_{12} = CNOT_{1\rightarrow 2}CNOT_{2\rightarrow 1}CNOT_{1\rightarrow 2}$$

where $CNOT_{1\rightarrow 2} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X$.

It also has the property that it leaves invariant any state ψ belonging to the symmetric supspace, and it adds a -1 phase to those belonging to the antisymmetric subspace. Specifically, $\Sigma_{12}\left|S_{12}\right\rangle = -\left|S_{12}\right\rangle$ and $\Sigma_{12}\left|T_{12}\right\rangle = \left|T_{12}\right\rangle$ where $\left|S_{12}\right\rangle$ is the singlet state and $\left|T_{12}\right\rangle$ is any one of the triplet states.

The partial swap is the gate:

$$U = \cos(\eta)I_{12} + i\sin(\eta)\Sigma_{12}$$

It corresponds to applying coherently (reversibly) the swap with probability s^2 and leave the state alone with probability c^2 where $c = \cos(\eta)$ and $s = \sin(\eta)$.

Let us compute the action of the partial swap on two generic initial states of two qubits A and B:

$\rho_a = \frac{I+\mathbf{A}\cdot\sigma}{2}$ and $\rho_b = \frac{I+\mathbf{B}\cdot\sigma}{2}$, where A and B are two real-valued 3-vectors, and $\sigma = (\sigma_x, \sigma_y, \sigma_z)$, a 3-vector of Pauli matrices. We compute:

$$\begin{aligned} S(\rho_a \otimes \rho_b)S &= c^2\rho_a \otimes \rho_b + s^2\rho_b \otimes \rho_a && (4)\\ &- \frac{cs}{8}(\mathbf{A} - \mathbf{B}) \cdot (\sigma \otimes I \wedge I \otimes \sigma) && (5)\\ &- \frac{cs}{8}(\mathbf{A} \wedge \mathbf{B}) \cdot (\sigma \otimes I - I \otimes \sigma) . && (6)\end{aligned}$$

The reduced density operators are, for the first and second qubits:

$$\mathrm{Tr}_B[\rho_a \otimes \rho_b] = \frac{1}{2}[I + c^2\mathbf{A} + s^2\mathbf{B} + \frac{cs}{4}(\mathbf{A} \wedge \mathbf{B})] \cdot \sigma \quad (7)$$

and

$$\mathrm{Tr}_A[\rho_a \otimes \rho_b] = \frac{1}{2}[I + c^2\mathbf{B} + s^2\mathbf{A} - \frac{cs}{4}(\mathbf{A} \wedge \mathbf{B})] \cdot \sigma \quad (8)$$

Special case. In the notation of our paper, let us assume the qubit Q has the initial state $\rho_x = \frac{1}{2}(I + \mathbf{A} \cdot \sigma)$ and the homogenisation machine is initialised in N identical copies of

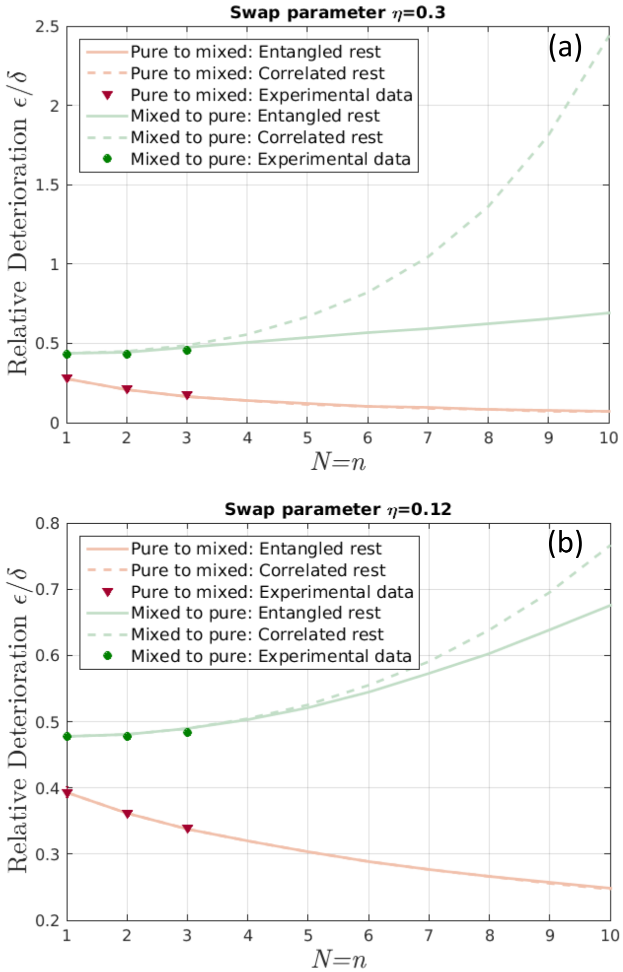

FIG. 5: Relative deterioration of the homogenisation machines for the pure-to-mixed tasks T and its counterpart $T^{\sim}$ for $\eta = 0.3$ (plot (a)) and $\eta = 0.12$ (plot (b)), as a function of the number of machine qubits N and machine usages n. The solid curves represent the theoretically-expected behavior considering the entangled rest state of the complete theory for $n > 1$, while the dashed curves show the predictions given by our correlated (but not entangled) rest approximation. The triangles (dots) indicate, instead, the experimental values obtained for the homogenisation machine performing the task T ($T^{\sim}$).

the state $\rho_y = \frac{I}{2}$. For the purpose of going through an analytical calculation, we will make three simplifying assumptions:

1. We are working in the limit of weak swaps, i.e., $\sin(\eta)$ is small compared to one;

2. We approximate the error after n usages of the same machine on n different input qubits as the n-th power of the deterioration in one usage

$$\delta_{\mathcal{E}[T]}(n) \approx \delta_{\mathcal{E}[T]}(1)^n$$

3. We assume that the state of the N qubits in the homogenisation machine after the interaction with the qubit Q can be approximated as the tensor product of the reduced density operators of each qubits.

We have also performed extensive simulations for the case where assumptions (ii) and (iii) are relaxed, and the result still holds, in the form that the relative deterioration does not go to zero for the "mixed-to-pure" case, but it does for the "pure-to-mixed" case. This indicates that the approximations (ii) and (iii) introduce an error in the analytical calculation which does not affect the proof of the asymmetry, as promised.

1) PURE TO MIXED CASE

The deterioration of the machine after one usage is:

$$\delta_{\mathcal{E}[T]}(1) = \Pi_{k=1}^{N} \frac{1}{2}(1 + \sqrt{(1 - s^4 c^{4(k-1)})})$$

The error in delivering the task is:

$$\epsilon_N(T) = \frac{1}{2}\left(1 - \sqrt{1 - c^{4N}}\right)$$

Approximating the above functions to first order in s^2, we have that the deterioration of the machine is constant and very small, so that $\delta_{\mathcal{E}[T]}(n) \approx 1$.

Hence in this limit, one has:

$$\lim_{N\to\infty} \lim_{n\to\infty} R_{\mathcal{E}[T]}(n) \approx 0$$

Thus the homogenisation machine in this case tends to a constructor.

2) MIXED TO PURE CASE

In this case, the deterioration of the machine after one usage is:

$$\delta_{\mathcal{E}[T^{\sim}]}(1) = \Pi_{k=1}^{N} \frac{1}{2}(1 + c^2 + s^4 \sum_{l=0}^{k-2} c^{2k})$$

where we define $\sum_{l=0}^{k-2} c^{2k} \doteq 0$ for $k = 1$.

Once more, let us approximate the error after n uses as the product of the fidelities:

$$\delta_{\mathcal{E}[T^{\sim}]}(n) = \delta_{\mathcal{E}[T]}(1)^n$$

The error in delivering the task is:

$$\epsilon_N(T^{\sim}) = \frac{1}{2}(1 - s^2 \sum_{k=1}^{N-1} c^{2k})$$

Let us approximate the above functions to first order in s^2:

$$\delta_{\mathcal{E}[T^{\sim}]}(n) \approx (1 - \frac{s^2}{2})^{Nn}$$

Then we see that:

$$\lim_{N\to\infty} \lim_{n\to\infty} R_{\mathcal{E}[T^{\sim}]}(n) \approx \infty\ .$$

A numerical simulation shows that when relaxing assumptions (ii) and (iii) the relative deterioration is non-zero, but finite; while in the case of the pure to mixed task it is zero. Also in this case, therefore, the homogenisation machine is a constructor for the task T (pure to mixed), but it is not a constructor for the reversed task $T^{\sim}$, displaying the promised asymmetry.

[1] Baylin, M. A survey of thermodynamics, New York, 1994.
[2] Buchdahl, H.A., The Concepts of Classical Thermodynamics, Cambridge University Press (1966).
[3] Wallace, D. The Quantitative Content of Statistical Mechanics, Studies in the History and Philosophy of Modern Physics 52 (2015) pp.285–293.
[4] Landauer, R. The physical nature of information, Phys. Lett. A **217** ,1996.
[5] Bennett, C. H. Demons, engines and the second law. Scientific American, 257, 108-116, 1987.
[6] Earman, J., and Norton, J. D. (1999). Exorcist XIV: The wrath of Maxwell's demon. Part II. Studies in History and Philosophy of Modern Physics, 30, 1-40.
[7] Deutsch, D. Constructor Theory, *Synthese* **190**, 18, (2013).
[8] Deutsch, D., Marletto, C., Constructor theory of information, Proceedings of the Royal Society A, **471**, 2174, (2015).
[9] Ziman, M., et al., Quantum homogenization, Phys. Rev. A 65, 042105 (2002).
[10] Scarani V. et al, Phys. Rev. Lett., 88, 097905, (2002).
[11] Wootters, W.K. and Zurek, W.H. Nature 299 (1982), pp. 802-803
[12] J. Uffink, Stud. Hist. Phil. Mod. Phys., B 32 (3), 2001.
[13] C. Carathéodory, Mathematische Annalen, 67, 355-386, 1909.
[14] E. Lieb, J. Yngvason, Phys. Rept. 310, 1-96, 1999.
[15] B. Coecke, T. Fritz; R. Spekkens, Information and Computation 250, 59–86, 2016.
[16] Barenco, A., et al. Stabilisation of quantum computations by symmetrisation, arXiv: quant-ph/9604028
[17] Brida, G., et al. Applied Phys. Lett. 101 (2012) 221112.
[18] Genovese, M., et al. Physics Reports 413/6 (2005) 319.
[19] Bogdanov, Yu. I., *et al. Phys. Rev. Lett.* **105**, 010404 (2010).
[20] Ghosh, A., Niedenzu, W., Mukherjee, V., Kurizki, G. in *Thermodynamics in the quantum regime - Recent Progress and Outlook* (Springer International Publishing, 2019).
[21] Uzdin, R. *Phys. Rev. Applied* **6**, 024004 (2016).
[22] Lostaglio, M., Jennings, D., and Rudolph, T. *Nat. Commun.* **6**, 6383 (2015).
[23] Riechers, P., and Gu, M. , arxiv: 2002.11425 (2020).
[24] Scarani, V. *EPJ Spec. Top.* **151**, 41-49 (2007).
[25] Manzano, G., Subero, D., Maillet, O., Fazio, R., Pekola, J.P., and Roldán, É., arxiv: 2008.01630 (2020).
[26] A. Anshu, et al. Phys. Rev. Lett. 121, 190504, 2018.
[27] J. Goold, M. Huber, A. Riera, L. del Rio, P. and Skrzypczyk, J. Phys. A: Math. Theor. 49, 143001, 2016.
[28] R. Alicki, R. Kosloff, Introduction to Quantum Thermodynamics: History and Prospects. In: Binder F., Correa L., Gogolin C., Anders J., Adesso G. (eds) Thermodynamics in the Quantum Regime. Fundamental Theories of Physics, vol 195. Springer, Cham., (2018).
[29] J. Gemmer, M. Michel and G. Mahler, Quantum Thermodynamics (Berlin: Springer), (2009)
[30] A constructor generalises the notion of a catalyst in resource theory [15], relaxing the requirement that it stay in exactly the same state, to the requirement that it stay exactly within the same set of states.

Acknowledgments: We are grateful to David Deutsch, Benjamin Yadin, Paul Raymond-Robichaud and Maria Violaris for useful discussions. This research was supported by the grant number (FQXi FFF Grant number FQXi-RFP-1812) from the Foundational Questions Institute and Fetzer Franklin Fund, a donor advised fund of Silicon Valley Community Foundation. CM's research was also supported by the Templeton World Charity Foundation and by the Eutopia Foundation. VV thanks the Oxford Martin School, the John Templeton Foundation, the EPSRC (UK). This work has received funding from the European Union's Horizon 2020 and the EMPIR Participating States in the context of the project EMPIR-17FUN01 "BeCOMe", and by the European Union's Horizon 2020 project "Pathos". This research is also supported by the National Research Foundation, Prime Ministers Office, Singapore, under its Competitive Research Programme (CRP Award No. NRF- CRP14-2014-02) and administered by Centre for Quantum Technologies, National University of Singapore.