

Inadequacy of von Neumann entropy for characterizing extractable work

Oscar C O Dahlsten^{1,2,3,5}, Renato Renner¹, Elisabeth Rieper³
and Vlatko Vedral^{2,3,4}

¹ Institute for Theoretical Physics, ETH Zürich, 8093 Zürich, Switzerland

² Clarendon Laboratory, University of Oxford, Parks Road,
Oxford OX1 3PU, UK

³ Centre for Quantum Technologies, National University of Singapore,
3 Science Drive 2, Singapore 117543, Singapore

⁴ Physics Department, National University of Singapore, 2 Science Drive 3,
Singapore 117543, Singapore
E-mail: cqtcod@nus.edu.sg

New Journal of Physics **13** (2011) 053015 (10pp)

Received 18 December 2010

Published 12 May 2011

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/13/5/053015

Abstract. The lack of knowledge that an observer has about a system limits the amount of work it can extract. This lack of knowledge is normally quantified using the Gibbs/von Neumann entropy. We show that this standard approach is, surprisingly, only correct in very specific circumstances. In general, one should use the recently developed smooth entropy approach. For many common physical situations, including large but internally correlated systems, the resulting values for the extractable work can deviate arbitrarily from those suggested by the standard approach.

⁵ Author to whom any correspondence should be addressed.

Contents

1. Introduction	2
2. Szilard's engine and Bennett's development	3
3. Smooth entropies	4
4. A model for work extraction	6
5. The work value of information	7
6. Standard heat engines and thermodynamical limit	8
7. Experiments	9
8. Subjective versus objective	9
9. Discussion	9
Acknowledgments	9
References	10

1. Introduction

The relation between work and information has been the cause of great debate since the birth of statistical mechanics. Focal points of the debate include Maxwell's demon, Szilard's engine, Landauer's erasure principle and Bennett's reversible measurements [1]–[4]. That there should be such a relation can be understood intuitively. Harnessing motion, e.g. wind, for one's benefit requires knowing its directionality. With regard to thermodynamical work extraction from the pressure of a gas one uses the knowledge that the particles are confined and will only push the piston from one known direction. One of the simplest examples of such extraction is Szilard's engine, described in figure 1.

Previous efforts to quantify the relation between work and information, e.g. for an n -particle Szilard-type engine coupled to a thermal bath at temperature T , yielded expressions of the type $W = (n - S)kT \ln 2$. Here, W is the extractable work, $S = S(\rho)$ the lack of information (entropy) about the system's initial state ρ (the position of the particles in the Szilard engine) and $nkT \ln 2$ the amount of work that would be gained if there were no uncertainty [4, 5]. Feynman argued this expression *defines* entropy S [5].

In this paper, we revisit this relation. We observe that, in the generic case of possibly internally correlated and finite-sized systems, there is a *risk-reward trade-off*; the more work one aims at extracting, the higher the risk that the work extraction process fails. As a main result, we find that if one defines entropy S via the extractable work W as Feynman suggested, then S is in general *not* the entropy suggested by standard textbook thermodynamics, namely the Shannon entropy ($S_{\text{Sh}} = -\sum_i p_i \log p_i$), which equals the Gibbs entropy up to Boltzmann's constant k , and its quantum equivalent, the von Neumann entropy ($S_{\text{vN}}(\rho) = -\text{tr}(\rho \log \rho)$). Instead we find that it is the *smooth entropy* introduced in [6, 7] that characterizes the extractable work in general. Smooth entropy generalizes von Neumann entropy in the sense that the latter is obtained in an appropriate limit. In particular, for systems with no internal correlations our statement reduces, asymptotically with increasing system size, to the standard statement, $W = (n - S_{\text{vN}})kT \ln 2$. However, in general, e.g. for large but internally correlated systems, the respective entropies can deviate arbitrarily.

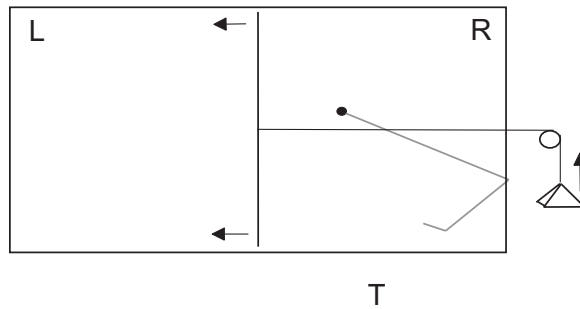


Figure 1. Szilard's paradigmatic example of how work extraction requires information. If one has information that a particle in contact with a thermal reservoir of temperature T is in one given side of a box (L or R), one can insert a divider and trap it there [2]. The divider is attached to a weight and, as the particle bounces around due to its thermal energy, the divider can be pushed in a predictable direction, lifting the weight. The work output is $kT \ln 2$ per such particle (where k is Boltzmann's constant), and this is therefore the work value of one bit of information (L or R) in this context.

Our presentation proceeds as follows. We first describe some basic concepts and methods related to the use of information for work extraction, in particular the idea of employing information compression. We then briefly review the smooth entropy and its use as an uncertainty measure in information theory. Subsequently, we state and explain our main results, relating extractable work and uncertainty, and discuss the implications.

2. Szilard's engine and Bennett's development

In order to motivate the general work extraction scenario that we will use here, it is instructive to recall certain specific examples existing in the literature. Bennett, in particular, considered an agent who controls n Szilard engines (as in figure 1), which it couples to a heat bath at temperature T to extract work [4]. It is assumed that the agent has partial knowledge about the (initial) position of the particle in each of the engines (i.e. whether it is in the left (L) or right (R) compartment). This knowledge is represented by a probability distribution P on particle positions $\{L, R\}^n$. By Szilard's argument, when a particle's position is *known* it can be used to isothermally extract work by letting the 'one-particle gas' expand, so that the work value is given by $\int_{V/2}^V (p) dV = kT \ln 2$ for each such engine. Consequently, if the positions of the particles in each of the boxes are either fully known or completely unknown, the total work value equals $W = (n - n_u)kT \ln 2$, where n_u is the number of completely unknown positions. Bennett notes, crucially, that correlations can be exploited too, even if the marginal distributions on the positions of individual particles are uniformly random. The agent can implement a reversible interaction between the particles to *compress* the total randomness (which is constrained by the correlations) onto the position of as few particles as possible, so that the others can then be used for work extraction. As a simple example, let $n = 2$, $P(LL) = P(RR) = 1/2$. Then performing a reversible interaction satisfying $LL \mapsto LL$ and $RR \mapsto RL$ (e.g. the so-called controlled-not operation) would yield $P(LL) = P(RL) = 1/2$, so that the second particle (whose position is now known to be L) could be used to extract $kT \ln 2$ work [5].

The above considerations, like many other information theoretical arguments, also apply to quantum systems [8, 9]. In fact, the nano and quantum regimes should be the focus for implementations of these ideas, due to the apparently unavoidable presence of friction in macroscopic systems. As shown in [9] (section 3.2), given a two-level system, an agent with perfect knowledge on the system's state can, using an appropriate coupling to a heat bath at temperature T , extract $kT \ln 2$ work, thus reproducing the value of the (classical) Szilard engine.

3. Smooth entropies

In information theory, entropies play a crucial role as measures for *ignorance* and, accordingly, as measures for *information*. (The *information* that a piece of data A gives on B may be defined as the amount by which one's ignorance of B decreases if one learns A .) Already in his seminal 1948 paper [10], where Shannon formulated the basics of the modern mathematical theory of information, he used a quantity that is formally equivalent to the Gibbs entropy to measure the uncertainty of random variables. This quantity is today referred to as the *Shannon entropy* and is ubiquitous in classical information theory. The use of quantities from statistical mechanics in information theory also carries over into the quantum domain. Here, the *von Neumann entropy* S_{vN} is employed, analogously to the Shannon entropy, as an uncertainty measure (and, correspondingly, as an information measure) for quantum states.

The relevance of the Shannon/von Neumann entropy S_{vN} within the theory of information is due to the fact that S_{vN} is related to *operational quantities*, i.e. quantities defined by a specific information processing task. For example, consider a system encoding some information in its state given by a density operator ρ . The maximum space (number of qubits) needed to store this information or, more precisely, the *compression rate*, equals $S_{\text{vN}}(\rho)$. Another example is Shannon's well-known formula for the *capacity of a communication channel*, which is essentially a difference between two Shannon entropies.

It turns out, however, that the use of the Shannon/von Neumann entropy S_{vN} for the characterization of operational quantities only works in an asymptotic sense, and under certain assumptions regarding the structure of the states. For example, the aforementioned correspondence between the space needed to store information from a source ρ and the von Neumann entropy $S_{\text{vN}}(\rho)$ is only valid if the source emits asymptotically many pieces of *independent and identically distributed (i.i.d.)* data, i.e. formally, if $\rho = \sigma^{\otimes n}$ for some large n . Similarly, Shannon's channel capacity formula is based on the assumption that the channel is memoryless and used many times.

Smooth min- and max-entropies were introduced in [6, 7] to overcome these limitations. In contrast to the Shannon/von Neumann entropy, they depend on an additional *smoothness parameter* $\varepsilon \geq 0$. They are defined by

$$H_{\min}^{\varepsilon}(\rho) := \sup_{\bar{\rho}} H_{\infty}(\bar{\rho}), \quad H_{\max}^{\varepsilon}(\rho) := \inf_{\bar{\rho}} H_{1/2}(\bar{\rho}),$$

where the supremum/infimum is taken over all states $\bar{\rho}$ that are ε -close to ρ and where $H_{\infty}(\bar{\rho}) = -\log_2 \|\bar{\rho}\|_{\infty}$ and $H_{1/2}(\bar{\rho}) = 2 \log_2 \text{tr}(\sqrt{\bar{\rho}})$ are the Rényi entropies of order ∞ and $1/2$, respectively. (We note that the extrema in the above definitions can usually be evaluated efficiently.)

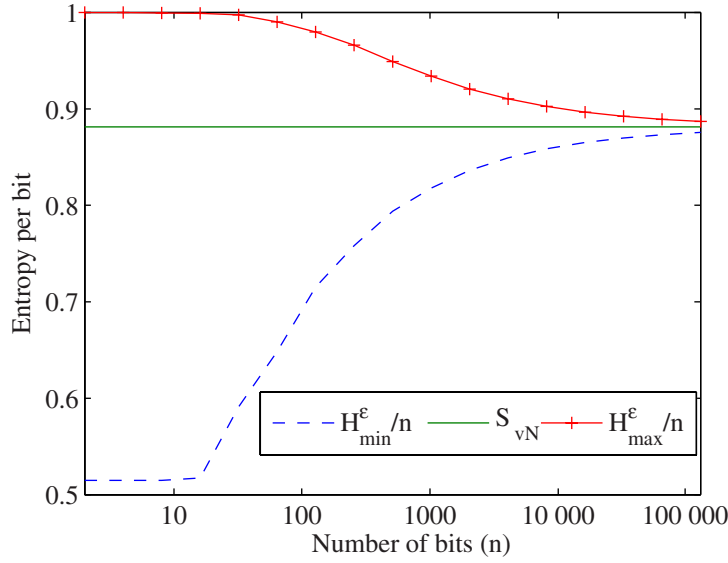


Figure 2. The entropies H_{\min} , $S_{\text{Sh}} = S_{\text{vN}}$ and H_{\max} are evaluated for n qubits, under the assumption that their state is i.i.d., with single-qubit density operator $\rho = \text{diag}(0.7, 0.3)$ (a choice motivated by the experiment [12]). One sees that for this particular case the entropies coincide in the $n \rightarrow \infty$ limit. This figure is relevant for the first two examples in table 1. The first shows that, for $n = 1000$, there is still a significant difference between the different entropies and, accordingly, the different work values. We emphasize that this figure concerns the special case of i.i.d. states—in general, i.e. for states with internal correlations, the entropies can differ arbitrarily, even for $n \rightarrow \infty$. (Technical note: the smoothing parameter ε used implicitly here is 10^{-5} .)

The smooth entropies can be seen as a generalization of the von Neumann entropy S_{vN} , in the sense that the latter can be obtained in the i.i.d. limit, i.e. for any $0 < \varepsilon < 1$,

$$S_{\text{vN}}(\rho) = \lim_{n \rightarrow \infty} \frac{1}{n} H_{\min}^{\varepsilon}(\rho^{\otimes n}) = \lim_{n \rightarrow \infty} \frac{1}{n} H_{\max}^{\varepsilon}(\rho^{\otimes n}), \quad (1)$$

(see e.g. [11] for more details).

Smooth min- and max-entropies accurately characterize a number of operational quantities, in the same way as the Shannon/von Neumann entropy does. Crucially, however, nothing needs to be assumed about the structure of the states and no asymptotics are needed. For example, the smooth max-entropy $H_{\max}^{\varepsilon}(\rho)$ quantifies the space that is necessary to store the information encoded in ρ in such a way that the information can be retrieved later, except with a failure probability ε . The standard result mentioned above, which asserts that the compression rate of an i.i.d. source ρ equals $S_{\text{vN}}(\rho)$, is obtained as a special case in the asymptotic limit by virtue of equation (1) (see also figure 2).

As we shall see, the situation is similar in statistical mechanics. We will show that also in the context of work extraction, smooth entropies naturally generalize the von Neumann entropy, in the sense that the latter is retrieved in the special case of large systems consisting of many identical and independent parts.

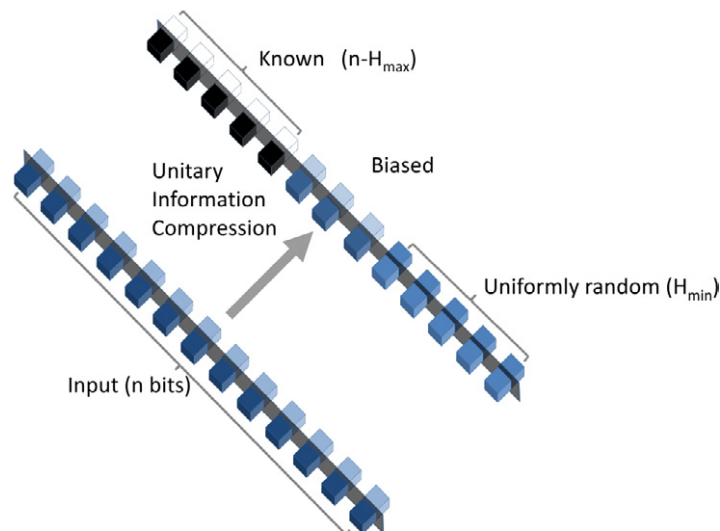


Figure 3. One may think of the state space as that of n (qu)bits, where each bit represents a box, as in figure 1. The reversible compression then results in a state where some bits are fully known, some are biased and some uniformly random. Darker colour indicates higher probability density.

4. A model for work extraction

We consider a work extraction scenario where an *agent* wishes to extract work from a given system in a well-defined setting. This can be termed as a *game* in that the agent (the sole player) has the choice of different strategies to implement. More precisely, the agent controls certain degrees of freedom of the system, described by a state space of size 2^n (corresponding to n qubits). The agent's knowledge is specified by the state ρ of the degrees of freedom under consideration. It is assumed that all states are energetically equivalent and that the agent can make the system undergo any desired reversible evolution. This corresponds to generating a unitary evolution on its state space, as discussed, e.g., in [13]–[15]. (For example, one may think of a scenario as depicted in figure 3, where the agent controls the position, L or R , of particles in n Szilard engines.)

In order to extract work the agent may couple the system or any part of it to a heat bath at temperature T , as well as to a work reservoir. (It thus has the choice of not using any parts of the system deemed to be too uncertain.) For the purpose of illustration, we think of the work reservoir as a suspended weight so that work is stored as gravitational potential energy (as in figure 1). We say that an *amount W of work is successfully extracted* if the weight has been lifted to *at least* an altitude corresponding to a potential energy of W .

We are interested in the dependence of the *amount of work W* the agent can extract from a system on the *information* it has about the system's initial state. Any ignorance about the initial state may, during the work extraction process, translate to ignorance about the direction of certain forces to which the weight is coupled. For example, using a single-particle Szilard engine with some nonzero probability of being L or R necessitates accepting a risk of failure. The agent would need to guess whether it is L or R and the extraction would fail if that guess turns out to be wrong. In general, the agent faces such a trade-off between the risk of failure, in the following denoted p_{fail} , and the work W extracted if successful. Only for specific strategies and states does $p_{\text{fail}} \rightarrow 0$.

Table 1. Dependence of the work value on the initial state. The state in the first row is given by a density operator, e.g. in the L/R -basis for a Szilard engine. The first two examples are discussed further in figure 2. The third example shows that the work value of information to a risk-taking agent can in general deviate arbitrarily from that of an agent taking only very small risks. The fourth example demonstrates the power of information compression, which here enables maximal work extraction even though the subsystems initially all had uniformly random distributions.

Initial state ρ	Min work (theorem 1)	Max work (theorem 2)
$\begin{pmatrix} 0.7 & 0 \\ 0 & 0.3 \end{pmatrix}^{\otimes 1000}, T = T_{\text{room}}$	1.0 eV	3.5 eV
$\begin{pmatrix} 0.7 & 0 \\ 0 & 0.3 \end{pmatrix}^{\otimes n}, n \rightarrow \infty$	$(n - S_N(\rho))kT \ln 2$	
$\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}^{\otimes n} + \frac{1}{2} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}^{\otimes n}$	≈ 0	$\approx nkT \ln 2$
$\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}^{\otimes n} + \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}^{\otimes n}$	$\approx nkT \ln 2$	$\approx nkT \ln 2$

5. The work value of information

In order to recover a simple theory despite this trade-off, we focus on the extreme cases of arbitrarily large- and small-risk tolerances, respectively, as these cases envelope all others. Our main results, theorems 1 and 2, give tight bounds on the two respective work values of information.

The theorems below refer to the scenario described above, where an agent has access to an n -qubit system as well as a thermal bath at temperature T . The system's initial state (conditioned on the agent's knowledge) is given by ρ . The theorems are of most interest when $\varepsilon > 0$ is small, but large enough such that $\ln 1/\varepsilon$ is small compared to H_{\min}^ε .

Theorem 1. *If the agent demands that the failure probability is small, $p_{\text{fail}} \leq 2\varepsilon$, then it can extract an amount of work given by*

$$W \geq (n - H_{\max}^\varepsilon(\rho) - 3 \ln 1/\varepsilon) kT \ln 2,$$

and this inequality is essentially tight in that for $p_{\text{fail}} \leq 2\varepsilon$, $W \leq (n - H_{\max}^{2\varepsilon}(\rho))kT \ln 2$.

Theorem 2. *If the agent chooses a strategy that, upon success, extracts an amount of work satisfying*

$$W \geq (n - H_{\min}^\varepsilon(\rho) + 3 \ln 1/\varepsilon) kT \ln 2,$$

then the success probability $p_{\text{succ}} = 1 - p_{\text{fail}}$ is necessarily small, $p_{\text{succ}} < 2\varepsilon$.

Examples demonstrating the use of these theorems are given in table 1.

The intuition behind the theorems is given by figure 3, which depicts the fact that, after a carefully chosen reversible compression step, roughly $n - H_{\max}^{\epsilon}$ qubits are fully known and H_{\min}^{ϵ} completely unknown. An agent effectively not taking a risk (theorem 1) can only use the fully known qubits. For any amount of risk-taking (theorem 2), the fully unknown qubits are useless for work extraction (they will almost certainly cancel each other out if the agent does try to use them).

The proofs of the two theorems both employ the same essential ideas. Here we sketch the proof of theorem 1. The general work extraction strategy we consider is as follows. Firstly, the agent acts reversibly (without energy expenditure) on the system, trying to compress the system's state such that it has support on a subspace of size 2^{ℓ} (except with some small failure probability, 2ϵ , and for some ℓ to be determined later). The remaining $n - \ell$ qubits can be set to a predefined known (pure) state. (For example, if the qubits are realized by single-particle Szilard engines, we can imagine that they are all set to the state L .) We call this step the *compression step*. After this step, we consider an *extraction step* where the $n - \ell$ qubits of the system, which are in a known state, are coupled to the heat bath and the work reservoir, in such a way that energy is moved from the former to the latter.

In order to prove the lower bound on W stated by theorem 1, we first look at the compression step. We use a result of [6] which states that any data can be compressed to $\ell = H_{\max}^{\epsilon} + 3 \ln 1/\epsilon$, where H_{\max}^{ϵ} is the smooth max-entropy of the data, and where 2ϵ is the maximum failure probability of this compression (i.e. the probability that the original data cannot be retrieved). This implies that there exists a reversible (unitary) operation U (defined by the compression function), which has the property that $U\rho U^{\dagger}$ has support on ℓ predefined qubits, except with probability 2ϵ . (More precisely, the state $U\rho U^{\dagger}$ is 2ϵ -close to a state of the form $\rho' \otimes \rho''$, where ρ' and ρ'' are states on ℓ and $n - \ell$ qubits, respectively, and ρ'' is a pure state.) This is exactly what needs to be achieved in the compression step and we can thus proceed with the analysis of the extraction step. For this, we apply the result of [9] described above, which asserts that any qubit in a known state can be used to extract an amount of work given by $kT \ln 2$. Consequently, since $n - \ell$ qubits are in the known (pure) state ρ'' , this result implies that the total amount of work that can be extracted is given by $(n - \ell)kT \ln 2$. (One may achieve the same conclusion using, instead of the general process described in [9], the work extraction process proposed by Szilard, under the standard but nontrivial assumption that any thermal fluctuations of the divider position are negligible relative to the total amount of work performed [8, 16].) To show optimality, it suffices to note that the above results concerning both the compressions step [6] and the extraction step [9] are essentially optimal. (More precisely, any scheme that compresses data into fewer than $H_{\max}^{2\epsilon}$ qubits will fail with probability at least 2ϵ .)

6. Standard heat engines and thermodynamical limit

The standard thermodynamical heat-engine scenario for work extraction is included in our general model. It corresponds to a strategy where the system is directly coupled to the heat bath without prior reversible information compression. In general, this is a suboptimal strategy. It could, for example, not extract any work given the fourth example of table 1 (except for large values of the risk tolerance). However, in the thermodynamical limit defined by the second example in the table it can, interestingly, extract an optimal amount of work. In this limit, the extractable work is (combining theorems 1 and 2 with equation (1)) given by

$W = (n - S_{\text{vN}}(\rho))kT \ln 2$, regardless of risk tolerance. The same amount is yielded by the standard strategy.

7. Experiments

While our result is about the fundamental limits of work extraction, it is interesting to ask how well these limits can be approached experimentally. The type of work extraction considered here requires a certain level of control over the system and is experimentally more challenging than standard thermodynamical work extraction, but it is nevertheless inherently easier than full quantum computation. One will in general not require a universal set of unitaries nor a high accuracy to demonstrate nontrivial work extraction via information compression, or to demonstrate the trade-off between risk and possible yield. One could also perform the inverse experiment, which amounts to resetting a system to a predefined state (see e.g. [5]), and here our results correspond to a generalization of Landauer's erasure principle. It seems likely that at least some of the multitude of methods being developed for performing quantum gates will be suitable for such experiments, and that they will be sufficiently good for this application significantly earlier than they can be used for quantum computing [17]. Experiments in this direction have already been performed in the context of nuclear magnetic resonance algorithmic cooling [15, 18].

8. Subjective versus objective

Our approach of treating work extraction as a guessing game, apart from generating a well-defined mathematical setting, reconciles the subjectivity of information with the objectivity of extracted work. The extractable work W must be seen as subjective, since different agents may have different knowledge and risk tolerance and accordingly apply different work extraction strategies. The extracted work, obtained using a particular strategy, is, however objective.

9. Discussion

Our results demonstrate that the Gibbs/von Neumann entropy is not generally the correct entropy to use for characterizing extractable work. Instead, the relevant entropy measure(s) for this task are the smooth entropies, which reduce to the Gibbs/von Neumann entropy only under additional assumptions (e.g. that the system under consideration consists of many mutually independent parts). This is important to note as entropy lies at the heart of statistical mechanics and thermodynamics. Additionally, our expressions can differ arbitrarily from the Gibbs/von Neumann expression, e.g. for large but internally correlated systems.

Acknowledgments

We thank J Åberg, R Colbeck, M Gu, A Imamoglu, D Kaszlikowski, M Leifer, S Lloyd, L del Rio and S Winkler for inspiring discussions and comments on earlier versions of this paper. OD and RR acknowledge support from the SNSF (grant no. 200021-119868). The Singapore CQT funded ER and hosted OD for a month during this work. The Pauli Center for Theoretical Studies also supported the collaboration. VV is grateful to EPSRC, the Royal Society, the Wolfson Foundation, NRF (Singapore) and MoE (Singapore) for financial support.

References

- [1] Leff H and Rex A F 2003 *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing* (Bristol: Institute of Physics)
- [2] Szilárd L 1929 Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen *Z. Phys.* **53** 840–56
- [3] Landauer R 1961 *IBM J. Res. Dev.* **5** 183
- [4] Bennett C H 2003 Notes on Landauer's principle, reversible computation and Maxwell's demon *Stud. Hist. Phil. Sci. B* **34** 501–10
- [5] Feynman R P, Hey A J and Allen R W 2000 *Feynman Lectures on Computation* (Cambridge, MA: Perseus)
- [6] Renner R and Wolf S 2004 Smooth Rényi entropy and applications *Proc. IEEE Int. Symp. Information Theory* p 233
- [7] Renner R 2005 Security of quantum key distribution *PhD Thesis* ETH Zürich (arXiv:quant-ph/0512258)
- [8] Zurek W H 2003 Maxwell's demon, Szilard's engine and quantum measurements arXiv:quant-ph/0301076
- [9] Alicki R, Horodecki M, Horodecki P and Horodecki R 2004 Thermodynamics of quantum information systems—Hamiltonian description *Open Syst. Inf. Dyn.* **11** 205–17
- [10] Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* **27** 379–423, 623–56
- [11] Tomamichel M, Colbeck R and Renner R 2009 A fully quantum asymptotic equipartition property *IEEE Trans. Inf. Theor.* **55** 5840–7
- [12] Serreli V, Lee C F, Kay E R and Leigh D A 2007 A molecular information ratchet *Nature* **445** 523–7
- [13] Oppenheim J, Horodecki M, Horodecki P and Horodecki R 2002 Thermodynamical approach to quantifying quantum correlations *Phys. Rev. Lett.* **89** 180402
- [14] Schulman L J and Vazirani U V 1999 Molecular scale heat engines and scalable quantum computation *Proc. STOC '99* pp 322–9
- [15] Fernandez J M, Lloyd S, Mor T and Roychowdhury V 2004 Algorithmic cooling of spins: a practicable method for increasing polarization *Int. J. Quantum Inf.* **2** 461–77
- [16] Berger J 1990 Szilard's demon revisited *Int. J. Theor. Phys.* **29** 985–95
- [17] Everitt H O 2005 *Experimental Aspects of Quantum Computing* (Berlin: Springer)
- [18] Baugh J, Moussa O, Ryan C A, Nayak A and Laflamme R 2005 Experimental implementation of heat-bath algorithmic cooling using solid-state nuclear magnetic resonance *Nature* **438** 470–3