Entanglement and the foundations of statistical mechanics

SANDU POPESCU[1,2], ANTHONY J. SHORT[1]* AND ANDREAS WINTER[3]

[1]H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK
[2]Hewlett-Packard Laboratories, Stoke Gifford, Bristol BS12 6QZ, UK
[3]Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK
*e-mail: tony.short@bristol.ac.uk

Published online: 29 October 2006; doi:10.1038/nphys444

Statistical mechanics is one of the most successful areas of physics. Yet, almost 150 years since its inception, its foundations and basic postulates are still the subject of debate. Here we suggest that the main postulate of statistical mechanics, the equal *a priori* probability postulate, should be abandoned as misleading and unnecessary. We argue that it should be replaced by a general canonical principle, whose physical content is fundamentally different from the postulate it replaces: it refers to individual states, rather than to ensemble or time averages. Furthermore, whereas the original postulate is an unprovable assumption, the principle we propose is mathematically proven. The key element in this proof is the quantum entanglement between the system and its environment. Our approach separates the issue of finding the canonical state from finding out how close a system is to it, allowing us to go even beyond the usual boltzmannian situation.

The great conceptual puzzle of statistical mechanics is how a physical system, despite always being in some definite state, and evolving deterministically, can exhibit thermodynamical properties pertinent to statistical averages, such as the entropy[1].

Here we consider an alternative approach to the foundations of statistical mechanics, suggested to one of us by Yakir Aharonov about twenty years ago. In this approach the usual devices of subjective randomness, ensemble-averaging or time-averaging[2], are not required. We show that, although the universe (that is, the system together with a sufficiently large environment) is in a quantum pure state subject to a global constraint, thermalization results from entanglement between the system and the environment. This leads to a finite entropy of the system, despite the universe itself having zero entropy. Significant results along similar lines have been obtained by Bocchieri and Loinger[3], Lloyd[4] and Gemmer *et al.*[5]; see also very recent work by Goldstein *et al*[6].

We formulate and prove a 'general canonical principle', which states that the system will be thermalized (that is, in the canonical state) for almost all pure states of the universe, and provide rigorous quantitative bounds. In fact, we actually go beyond ordinary thermalization: in the standard statistical setting, energy constraints are imposed on the state of the universe, which determine a corresponding temperature and thermal canonical state for the system. In contrast, we allow completely arbitrary constraints, which leads to the system being in a corresponding generalized canonical state.

Our results are kinematic, rather than dynamical, as we do not consider any particular evolution of the state. However, because almost all states of the universe are such that the system is thermalized, we anticipate that most evolutions will quickly carry any initial state to a thermal state. Furthermore, as information about the system will tend to leak into the environment over time, we might expect that their entanglement, and hence entropy, will increase. It is conceivable that this is the mechanism behind the second law of thermodynamics.

Consider a large isolated quantum mechanical system, 'the universe', that we decompose into two parts, the 'system' S and the 'environment' E. We will assume that the dimension of the environment is much larger than that of the system. In addition, let the state of the universe obey some global constraint R. We can represent this quantum mechanically by restricting the allowed states of the system and environment to a subspace $\mathcal{H}_R$ of the total Hilbert space:

$$\mathcal{H}_R \subseteq \mathcal{H}_S \otimes \mathcal{H}_E,$$

where $\mathcal{H}_S$ and $\mathcal{H}_E$ are the Hilbert spaces of the system and environment, with dimensions d_S and d_E respectively. In standard statistical mechanics, R would typically be a restriction on the total energy of the universe, which then determines a corresponding temperature for the system, but here we leave R completely general.

We define $\mathcal{E}_R$, the equiprobable state of the universe corresponding to the restriction R, by

$$\mathcal{E}_R = \frac{1}{d_R} \mathbb{1}_R,$$

where $\mathbb{1}_R$ is the identity (projection) operator on $\mathcal{H}_R$ and d_R is the dimension of $\mathcal{H}_R$. $\mathcal{E}_R$ is the maximally mixed state in $\mathcal{H}_R$, in which each pure state has equal probability. This corresponds to the standard intuition of assigning equal *a priori* probabilities to all states of the universe consistent with the constraints. The equal *a priori* probability postulate is the assumption that the equilibrium thermodynamics of the universe under the restriction R is entirely described by $\mathcal{E}_R$.

We define Ω_S, the canonical state of the system corresponding to the restriction R, as the quantum state of the system when the universe is in the equiprobable state $\mathcal{E}_R$. The canonical state of the system, Ω_S, is therefore obtained by tracing out the environment in the equiprobable state of the universe:

$$\Omega_S = \mathrm{Tr}_E \mathcal{E}_R. \quad (1)$$

We now take the crucial conceptual step of our approach, and consider that the universe is in a pure state $|\phi\rangle$, and not in the mixed state $\mathcal{E}_R$ (which represents a subjective lack of knowledge about its state). We prove that despite this, the reduced state of the system,

$$\rho_S = \mathrm{Tr}_E |\phi\rangle\langle\phi|,$$

is very close to the canonical state Ω_S in almost all cases:

$$\rho_S \approx \Omega_S.$$

That is, for almost every pure state $|\phi\rangle \in \mathcal{H}_R$ of the universe, the system behaves as if the universe were actually in the equiprobable mixed state $\mathcal{E}_R$.

This qualitative result can be stated as a ‘general canonical principle’ that will subsequently be refined to a quantitative theorem: Given a sufficiently small subsystem of the universe, almost every pure state of the universe is such that the subsystem is approximately in the canonical state Ω_S.

Recalling that the canonical state of the system Ω_S is, by definition, the state of the system when the universe is in the equiprobable state $\mathcal{E}_R$, we can interpret this principle as a ‘principle of apparently equal *a priori* probability’: For almost every pure state of the universe, the state of a sufficiently small subsystem is approximately the same as if the universe were in the equiprobable state $\mathcal{E}_R$. In other words, almost every pure state of the universe is locally (that is, on the system) indistinguishable from $\mathcal{E}_R$.

We emphasize that the above are generalized principles, in the sense that the restriction R imposed on the states of the universe is completely arbitrary (and is not necessarily the usual constraint on energy or other conserved quantities). Similarly, the canonical state Ω_S is not necessarily the usual thermal canonical state, but is defined relative to the arbitrary restriction R by equation (1). This generalization incorporates the grand canonical case as well as the standard canonical case, and may lead to many other interesting insights. For example, it may be useful in situations where the usual weak-interaction approximation does not hold.

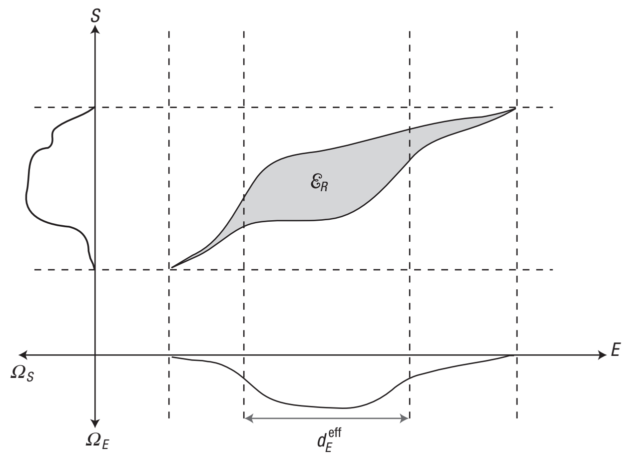

Figure 1 **The equiprobable state of the universe $\mathcal{E}_R$ corresponding to the restriction *R*.** This diagram represents $\mathcal{E}_R$ in a schematic state space of the system and environment, with the configurations of E and S on the horizontal and vertical axes respectively. The central shaded region represents the allowed states of the universe, and the functions plotted along the axes represent the number of configurations in $\mathcal{E}_R$ consistent with a given configuration of E or S. These marginals provide an illustration of the reduced states Ω_E and Ω_S, and show why $d_E^{\mathrm{eff}} = (\mathrm{Tr}\Omega_E^2)^{-1}$ is generally a better measure of the effective environment dimension than the support of Ω_E. Note however, that when all of the non-zero eigenvalues of Ω_E are of equal weight, d_E^{eff} simply corresponds to the dimension of Ω_E's support. Also, when there is no constraint on the accessible states of the universe, (that is, $\mathcal{H}_R = \mathcal{H}_S \otimes \mathcal{H}_E$), then $d_E^{\mathrm{eff}} = d_E$.

To connect the general canonical principle to standard statistical mechanics, all we have to do is to consider the restriction R to be that the total energy of the universe is close to E, which then sets the temperature scale T. The total hamiltonian of the universe H_U is given by

$$H_U = H_S + H_E + H_{\mathrm{int}},$$

where H_S and H_E are the hamiltonians of the system and environment respectively, and H_{int} is the interaction hamiltonian between the system and the environment. In the standard situation, in which H_{int} is small and the density of states of the environment increases approximately exponentially with energy, $\Omega_S^{(E)}$ can be computed using standard techniques[7], and shown to be

$$\Omega_S^{(E)} \propto \exp\left(-\frac{H_S}{k_B T}\right),$$

where k_B is Boltzmann's constant.

This allows us to state the ‘thermal canonical principle’ that establishes the validity (at least kinematically) of the viewpoint expressed in the introduction: Given that the total energy of the universe is approximately E, interactions between the system and the rest of the universe are weak, and that the density of states of the environment increases approximately exponentially with energy, almost every pure state of the universe is such that the state of the system alone is approximately equal to the thermal canonical state $e^{-H_S/k_B T}$, with temperature T (corresponding to the energy E).

We want to stress, however, that our contribution in this paper is to show that $\rho_S \approx \Omega_S$, and has nothing to do with showing that $\Omega_S \propto e^{-H_S/k_B T}$, which is a standard problem in statistical mechanics. In other words, we are interested in the distance between ρ_S and the canonical state, not in finding the canonical state itself. Our approach allows us to separate these two problems.

Finally, we note that the general canonical principle also applies in the case where the interaction between the system and environment is not small. In such situations, the canonical state of the system is no longer $e^{-H_S/k_B T}$, as the behaviour of the system will depend very strongly on H_{int}. Nevertheless, the general principle remains valid for the corresponding generalized canonical state Ω_S.

We now formulate a precise mathematical theorem corresponding to the general canonical principle stated above. The distance between the actual state ρ_S of the system and the canonical state Ω_S is the trace distance[8] $D(\rho_S, \Omega_S) = (1/2)\text{Tr}\sqrt{(\rho_S - \Omega_S)^\dagger(\rho_S - \Omega_S)}$. This distance is equal to the maximal difference between the two states in the probability of obtaining any measurement outcome. The distance $D(\rho_S, \Omega_S)$ therefore quantifies how hard it is to tell ρ_S and Ω_S apart by quantum measurements.

Throughout this paper we denote by $\langle\cdot\rangle$ the average over states $|\phi\rangle \in \mathcal{H}_R$ according to the standard (unitarily invariant) measure. For example, it is easy to see that $\Omega_S = \langle\rho_S\rangle$. The same measure is used to compute volumes of sets of states.

This leads us to the main theorem of this paper: The volume of states $V[\{|\phi\rangle \in \mathcal{H}_R | D(\rho_S(\phi), \Omega_S) \geq \eta\}]$ for which the system is further than η from the canonical state Ω_S is related to the total volume of allowed states $V[\{|\phi\rangle \in \mathcal{H}_R\}]$ by

$$\frac{V[\{|\phi\rangle \in \mathcal{H}_R | D(\rho_S(\phi), \Omega_S) \geq \eta\}]}{V[\{|\phi\rangle \in \mathcal{H}_R\}]} \leq \eta',$$

where, for arbitrary $\epsilon > 0$,

$$\eta = \epsilon + \frac{1}{2}\sqrt{\frac{d_S}{d_E^{\text{eff}}}}, \qquad (2)$$

$$\eta' = 4\exp\left(-C d_R \epsilon^2\right).$$

In these expressions, C is a positive constant ($C = (2/9)\pi^{-3}$), d_S and d_R are the dimensions of $\mathcal{H}_S$ and $\mathcal{H}_R$ respectively, and d_E^{eff} is a measure of the effective size of the environment, given by

$$d_E^{\text{eff}} = \frac{1}{\text{Tr}\,\Omega_E^2} \geq \frac{d_R}{d_S},$$

where $\Omega_E = \text{Tr}_S\,\mathcal{E}_R = \langle\rho_E\rangle$. The intuitive meaning of d_E^{eff} is explained in Fig. 1.

Both η and η' will be small quantities, and thus almost all states will be close to the canonical state, whenever $d_E^{\text{eff}} \gg d_S$ (that is, the effective dimension of the environment is much larger than that of the system) and $d_R\epsilon^2 \gg 1 \gg \epsilon$. This latter condition can be ensured when $d_R \gg 1$ (that is, the total accessible space is large), by choosing $\epsilon = d_R^{-1/3}$.

This theorem gives rigorous meaning to our statements about thermalization being achieved for 'almost all' states: we have an exponentially small bound on the relative volume of the exceptional set, that is, on the fraction of states that are far from the canonical state. Interestingly, the exponent scales with the dimension of the constraint space $\mathcal{H}_R$, whereas the deviation from the canonical state is characterized by the ratio between the system size and the effective size of the environment, which makes intuitive sense.

Note that in the special case in which there is no restriction on the allowed states (that is, $\mathcal{H}_R = \mathcal{H}_S \otimes \mathcal{H}_E$), our theorem implies that the reduced state of the system will almost always lie close to the maximally mixed state ($\Omega_S = \mathbb{1}_S/d_S$). This case can be viewed as an infinite-temperature limit of the standard canonical situation, and has previously been studied by a number of authors[9–13].

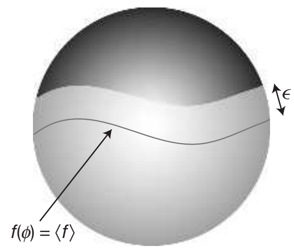

Figure 2 Bounding deviations from the average using Levy's lemma. Levy's lemma[14] is a result in high-dimensional geometry, which states that for almost all points ϕ on a hypersphere of dimension d (where $d \gg 1$) and area $V[\{\phi\}]$, and all functions f that do not vary too rapidly ($|\nabla f| \leq 1$), $f(\phi)$ is approximately equal to its mean value $\langle f \rangle$. The diagram shows the case $d = 2$, in which the hypersphere corresponds to the surface of a normal sphere. The shaded region corresponds to the maximum area $V[\{\phi | f(\phi) - \langle f \rangle \geq \epsilon\}]$ in which f is ϵ greater than average. Although this area is relatively large for $d = 2$, when d becomes large, the relative size of this region compared with the entire hypersphere becomes exponentially small. Specifically, Levy's lemma states that $V[\{\phi | f(\phi) - \langle f \rangle \geq \epsilon\}]/V[\{\phi\}] \leq 4\exp(-(1/9\pi^3)(d+1)\epsilon^2)$.

A major component in the proof of the theorem is a mathematical result known as Levy's lemma[14] (presented in Fig. 2), which plays a similar role to the law of large numbers and governs the properties of typical points on high-dimensional hyperspheres. Owing to normalization, pure quantum states can be represented by points on the surface of a hypersphere, making Levy's lemma a very powerful tool with which to evaluate functions of typical quantum states. It has already been used in quantum information theory to study entanglement and other correlation properties of typical states in large bipartite systems[15].

We now apply Levy's lemma (as given in Fig. 2) to the $(2d_R - 1)$-dimensional hypersphere of quantum states $|\phi\rangle \in \mathcal{H}_R$ and the function $f(\phi) = D(\rho_S, \Omega_S)$. Rearranging the resulting equation, we obtain our main theorem by proving that

$$\langle D(\rho_S, \Omega_S)\rangle \leq \frac{1}{2}\sqrt{\frac{d_S}{d_E^{\text{eff}}}}. \qquad (3)$$

This proof proceeds in two stages. First, bounding the distance $D(\rho_S, \Omega_S)$ from above by the more convenient quantity $(1/2)\sqrt{d_S}\sqrt{\text{Tr}(\rho_S - \Omega_S)^2}$ using standard techniques, leads to

$$\langle D(\rho_S, \Omega_S)\rangle \leq \frac{1}{2}\sqrt{d_S\left(\left\langle\text{Tr}\rho_S^2\right\rangle - \text{Tr}\langle\rho_S\rangle^2\right)}.$$

Second, we use the key mathematical insight that

$$\left\langle\text{Tr}\rho_S^2\right\rangle \leq \text{Tr}\langle\rho_S\rangle^2 + \text{Tr}\langle\rho_E\rangle^2.$$

The proof of this second inequality proceeds by a representation theoretic argument to evaluate the left-hand side, and uses similar techniques to those in ref. 9, involved in random quantum channel coding[16] and random entanglement distillation (see ref. 17). Combining these two inequalities and substituting $\text{Tr}\langle\rho_E\rangle^2 = 1/d_E^{\text{eff}}$, we obtain equation (3). Full details of this proof, and the others in this paper, can be found in ref. 18.

Note that although the bound (3) being small already suffices to argue that most states have almost canonical reduced state, the bound due to Levy's lemma is exponentially stronger.

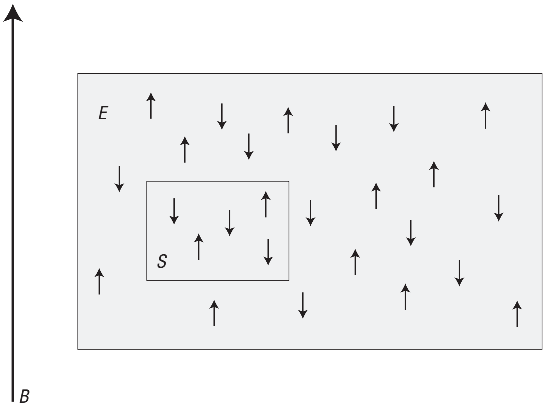

Figure 3 Example: A system of spins. As a concrete application of our theorem, consider a set of n spin-1/2 systems in an external magnetic field B, where a particular subset of k spins form the system S, and the remaining $n-k$ spins form the environment E. We consider a restriction to the energy eigenspace $\mathcal{H}_R$ in which np spins are in the excited state $|1\rangle$ (aligned with the field) and the remaining $n(1-p)$ spins are in the ground state $|0\rangle$ (opposite to the field). With this setup, $d_S = 2^k$ and $d_R \approx 2^{nH(p)}$, where $H(p) = -p\log_2(p) - (1-p)\log_2(1-p)$ is the Shannon entropy of a single spin. Using our theorem, we obtain $D(\rho_S, \Omega_S) \approx 0$ for almost all states whenever $nH(p) \gg 2k$. Projecting onto the typical subspace in which the system contains approximately kp excitations allows us to replace k by $kH(p)$ with very little additional error. This yields $D(\rho_S, \Omega_S) \approx 0$ for almost all states whenever $n \gg 2k$ (that is, whenever the system is fractionally larger than the environment).

In many cases, it is possible to improve the bounds obtained from the main theorem by projecting the state $|\phi\rangle$ onto a typical subspace of $\mathcal{H}_R$ before proceeding with the analysis. This can allow us to decrease the effective dimension of the system d_S (by eliminating components with negligible amplitude), and increase the effective dimension of the environment d_E^{eff} (by eliminating components of Ω_E with disproportionately high amplitudes), whilst leaving the equiprobable state $\mathcal{E}_R$ largely unchanged.

In addition to altering d_S and d_E^{eff}, this projection will introduce an additional error term in η (equation (2)) given by $4\sqrt{\delta}$, where δ is the relative volume of states in $\mathcal{H}_R$ outside the typical subspace. However, for an appropriately chosen typical subspace, the reduction in $\sqrt{d_S/d_E^{\text{eff}}}$ will often more than compensate for this additional source of error. This is particularly evident for the example of a collection of spin-1/2 particles given in Fig. 3.

Let us look back at what we have done. Concerning the problem of thermalization of a system interacting with an environment in statistical mechanics, there are several standard approaches. One way of looking at it is to say that the only thing we know about the state of the universe is a global constraint such as its total energy. Thus, the way to proceed is to take a bayesian point of view and consider all states consistent with this global constraint to be equally probable. The average over all these states indeed leads to the state of any small subsystem being canonical. But the question then arises: what is the meaning of this average, when we deal with just one state? Also, these probabilities are subjective, and this raises the problem of how to argue for an objective meaning of the entropy. A formal way out is that suggested by Gibbs, to consider an ensemble of systems, but of course this does not solve the puzzle, because there is usually only one actual system. Alternatively, it was suggested that the state of the universe, as it evolves in time, can reach any of the states that are consistent with the global constraint. Thus, if we look at time averages, they are the same as the average that results from considering each state of the universe to be equally probable. To make sense of this image, assumptions of ergodicity are needed to ensure that the universe explores all the available space equally, and of course this does not solve the problem of what the state of the subsystem is at a given time.

What we showed here is that these averages are not necessary. Rather, (almost) any individual state of the universe is such that any sufficiently small subsystem behaves as if the universe was in the equiprobable average state. This is due to massive entanglement between the subsystem and the rest of the universe, which is a generic feature of the vast majority of states. To obtain this result, we have introduced measures of the effective size of the system, d_S, and its environment (that is, the rest of the universe), d_E^{eff}, and showed that the average distance between the individual reduced states and the canonical state is directly related to d_S/d_E^{eff}. Levy's lemma is then invoked to conclude that all but an exponentially small fraction of all states are close to the canonical state.

The main message of our paper is that averages are not needed to justify the canonical state of a system in contact with the rest of the universe—almost any individual state of the universe is enough to lead to the canonical state. In effect, we propose to replace the postulate of equal *a priori* probabilities by the principle of apparently equal *a priori* probabilities, which states that as far as the system is concerned almost every state of the universe seems similar to the average.

We stress once more that we are concerned only with the distance between the state of the system and the canonical state, and not with the precise mathematical form of this canonical state. Indeed, it is an advantage of our method that these two issues are completely separated. For example, our result is independent of the canonical state having boltzmannian form, of degeneracies of energy levels, of interaction strength, or of energy (of the system, the environment or the universe) at all.

In future work, we hope to go beyond the kinematic viewpoint presented here to address the dynamics of thermalization. In particular, we will investigate under what conditions the state of the universe will evolve into (and spend almost all of its later time in) the large region of its Hilbert space in which its subsystems are thermalized. Some results in this direction have already been obtained[19–22], and we hope that the new results and techniques introduced in this paper will lead to further exciting advances in this area.

Received 10 April 2006; accepted 2 October 2006; published 29 October 2006.

References

1. Mandl, F. *Statistical Physics* 2nd edn, Ch. 2, 40–41 (Wiley, New York, 1988).
2. Khinchin, A. I. *Mathematical Foundations of Statistical Mechanics* Ch. 3, 44–47 (Dover, New York, 1949).
3. Bocchieri, P. & Loinger, A. Ergodic foundation of quantum statistical mechanics. *Phys. Rev.* **114,** 948–951 (1959).
4. Lloyd, S. *Black Holes, Demons, and the Loss of Coherence*. Ch. 3, 79–93, Thesis, Rockefeller Univ. (1988).
5. Gemmer, J., Michel, M. & Mahler, G. *Quantum Thermodynamics* Vol. 657 (Lecture Notes in Physics, Springer, Berlin, 2004).
6. Goldstein, S., Lebowitz, J. L., Tumulka, R. & Zanghì, N. Canonical typicality. *Phys. Rev. Lett.* **96,** 050403 (2006).
7. Landau, L. D. & Lifshitz, E. M. *Statistical Physics* Ch. 3, 78–80 (Pergamon, London, 1958).
8. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* Ch. 9, 403–409 (Cambridge Univ. Press, Cambridge, 2000).
9. Lubkin, E. Entropy of an n-system from its correlation with a k-reservoir. *J. Math. Phys.* **19,** 1028–1031 (1978).
10. Lloyd, S. & Pagels, H. Complexity as thermodynamic depth. *Ann. Phys.* **188,** 186–213 (1988).
11. Page, D. N. Average entropy of a subsystem. *Phys. Rev. Lett.* **71,** 1291–1294 (1993).
12. Sen, S. Average entropy of a quantum subsystem. *Phys. Rev. Lett.* **77,** 1–3 (1996).
13. Sommers, H.-J. & Zyczkowski, K. Statistical properties of random density matrices. *J. Phys. A: Math. Gen.* **37,** 8457–8466 (2004).
14. Milman, V. D. & Schechtman, G. *Asymptotic Theory of Finite-Dimensional Normed Spaces* Ch. 2, 5–6 (Lecture Notes in Mathematics, Vol. 1,200, Springer, Berlin, 1986) and 140–141 Appendix V.
15. Hayden, P., Leung, D. W. & Winter, A. Aspects of generic entanglement. *Commun. Math. Phys.* **265,** 95–117 (2006).
16. Lloyd, S. Capacity of the noisy quantum channel. *Phys. Rev. A* **55,** 1613–1622 (1997).

17. Horodecki, M., Oppenheim, J. & Winter, A. Partial quantum information. *Nature* **436,** 673–676 (2005).
18. Popescu, S., Short, A. J. & Winter, A. Entanglement and the foundations of statistical mechanics: Individual states vs averages. Preprint at <http://arxiv.org/abs/quant-ph/0511225> (2005).
19. Tasaki, H. From quantum dynamics to the canonical distribution: General picture and a rigorous example. *Phys. Rev. Lett.* **80,** 1373–1376 (1998).
20. Michel, M., Mahler, G. & Gemmer, J. Fourier's law from Schrödinger dynamics. *Phys. Rev. Lett.* **95,** 180602 (2005).
21. Breuer, H. P., Gemmer, J. & Michel, M. Non-Markovian quantum dynamics: Correlated projection superoperators and Hilbert space averaging. *Phys. Rev. E* **73,** 016139 (2006).
22. Gemmer, J. & Michel, M. Thermalization of quantum systems by finite baths. *Europhys. Lett.* **73,** 1–7 (2006).

Acknowledgements

The authors would like to thank Y. Aharonov and N. Linden for discussions. S.P., A.J.S. and A.W. acknowledge support through the UK EPSRC project 'QIP IRC'. In addition, S.P. also acknowledges support through EPSRC 'Engineering-Physics' grant GR/527405/01 and A.W. acknowledges a University of Bristol Research Fellowship.
Correspondence and requests for materials should be addressed to A.J.S.

Competing financial interests

The authors declare that they have no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/