

Thermodynamic uncertainty relations from exchange fluctuation theorems

André M. Timpanaro,^{1,*} Giacomo Guarnieri,^{2,†} John Goold,^{2,‡} and Gabriel T. Landi^{3,§}

¹Universidade Federal do ABC, 09210-580 Santo André, Brazil

²Department of Physics, Trinity College Dublin, Dublin 2, Ireland

³Instituto de Física da Universidade de São Paulo, 05314-970 São Paulo, Brazil.

(Dated: August 15, 2019)

Thermodynamic uncertainty relations (TURs) place strict bounds on the fluctuations of thermodynamic quantities in terms of the associated entropy production. In this work we identify the tightest (and saturable) matrix-valued TUR that can be derived from the exchange fluctuation theorems describing the statistics of heat and particle flow between multiple systems of arbitrary dimensions. Our result holds for both quantum and classical systems, undergoing general finite-time, non-stationary processes. Moreover, it provides bounds not only for the variances, but also for the correlations between thermodynamic quantities. To demonstrate the relevance of TURs to the design of nanoscale machines, we consider the operation of a two-qubit SWAP engine undergoing an Otto cycle and show how our results can be used to place strict bounds on the correlations between heat and work.

Introduction - Over the last decades, technological developments have led to the creation of artificial meso- and nanoscopic heat engines [1, 2], with applications ranging from nano-junction thermoelectrics [3] to quantum dots [4]. Understanding the fundamental principles ruling over the non-equilibrium physics of such devices is therefore one of nowadays most sought-after challenges. One of the key features of these non-equilibrium processes is that they are always accompanied by an irreversible production of entropy. And as the systems become smaller, the fluctuations in the entropy production become significant. This requires one to treat the entropy production Σ as a random variable distributed according to a certain probability distribution $P(\Sigma)$. These distributions satisfy a set of fundamental symmetry relations, known as Fluctuation Theorems (FT) [5–17] which can generally be expressed as $P(\Sigma)/\tilde{P}(-\Sigma) = e^{\Sigma}$, where $\tilde{P}(\Sigma)$ denotes the probability distribution of the time-reversed process. FTs represent a refinement of the second law of thermodynamics, which at the stochastic level is recast in the form $\langle \Sigma \rangle \geq 0$. The additional information they carry, however, can also be used to characterize systems arbitrarily far from equilibrium, which generated enormous interest across many fields of research [3, 4, 18–21].

More recently, another set of powerful results called *Thermodynamic Uncertainty Relations* (TURs) have been discovered [22–25]. TURs impose strict restrictions on the fluctuations of thermodynamic currents (e.g., heat, particles, etc.). Letting Q denote any such *integrated* current (net charge) exchanged during an out-of-equilibrium process over some generic time interval, the TURs bound the signal-to-noise ratio (SNR) of Q according to

$$\frac{\text{Var}(Q)}{\langle Q \rangle^2} \geq \frac{2}{\langle \Sigma \rangle}, \quad (1)$$

where $\text{Var}(Q) = \langle Q^2 \rangle - \langle Q \rangle^2$ denotes the variance, $\langle Q \rangle$ the average charge and $\langle \Sigma \rangle$ the average entropy production. Eq. (1) expresses a tradeoff between process precision, quantified by the SNR, and dissipation, quantified through the entropy production. To reduce fluctuations one must pay the inevitable price

of dissipation. This has important ramifications for the operation of microscopic autonomous engines [24] where fluctuations in the output power may be significant.

TURs were originally discovered in the context of non-equilibrium steady-states of classical time-homogeneous Markov jump-processes satisfying local detailed balance [22–24]. Further refinements and extensions have since then been found for finite-time processes [25–27], periodically driven systems [28–30], quantum systems in linear response [31] and using geometrical arguments based on the manifold of non-equilibrium steady-states [32].

A natural question that emerges is whether TURs, being inequalities, can be viewed as a consequence of FTs, just like the second law $\langle \Sigma \rangle \geq 0$. Explorations in this direction began quite recently, starting with symmetric work protocols [30, 33, 34] and subsequently generalized to include measurement feedback [35, 36]. In this paper we derive a new type of saturable TUR for FTs stemming from heat and particle exchange between multiple systems (Fig. 1(a)). This class of problems is particularly relevant, as it encompasses microscopic autonomous engines, which can be implemented in thermoelectric devices [37, 38] and are now starting to be pursued in controlled quantum platforms [39–45]. A set of charges Q_1, \dots, Q_n (energy, work, heat, particles, etc.) in this case satisfies the so-called Exchange Fluctuation Theorems (EFTs) [11–13] (see also [14, 15])

$$\frac{P(Q_1, \dots, Q_n)}{P(-Q_1, \dots, -Q_n)} = e^{\sum_i A_i Q_i}, \quad (2)$$

where A_i are thermodynamic affinities associated to each charge. The corresponding entropy production is $\Sigma = \sum_i A_i Q_i$.

As our first main result, we show that the EFT (2) implies a generalized TUR for any charge Q_i , of the form

$$\frac{\text{Var}(Q_i)}{\langle Q_i \rangle^2} \geq f(\langle \Sigma \rangle), \quad (3)$$

where $f(x) = \text{csch}^2(g(x)/2)$, $\text{csch}(x)$ is the hyperbolic cosecant and $g(x)$ is the function inverse of $x \tanh(x)$. We prove that this bound represents *the tightest saturable* tradeoff bound

for the SNR of any observable satisfying (2), given $\langle Q_i \rangle$. In fact, we are also able to provide an explicit form for the probability distribution $P(Q_1, \dots, Q_n)$ saturating (3). This is to be compared, for instance, to the bounds derived in [30, 33–35], which are looser but cannot be saturated. On the other hand, a series expansion of $f(x)$ around $x = 0$ yields $f(x) \simeq 2/x - 2/3$, so that for $\langle \Sigma \rangle \ll 1$ one recovers the bound (1). The bound (1), however, does not necessarily apply to all scenarios involving the exchange fluctuation theorem and can be violated. Our bound, on the other hand, is always looser than (1) and always holds in any EFT scenario.

Our framework also allows us to go further and construct a matrix-valued TUR for the covariance matrix $C_{ij} = \text{Cov}(Q_i, Q_j) = \langle Q_i Q_j \rangle - \langle Q_i \rangle \langle Q_j \rangle$ between different charges, similar in spirit to Refs. [27, 31, 32]. In this case, the bound becomes

$$C - f(\langle \Sigma \rangle) \mathbf{q} \mathbf{q}^T \geq 0, \quad (4)$$

where $\mathbf{q} = (\langle Q_1 \rangle, \dots, \langle Q_n \rangle)$ and the inequality is to be interpreted as a condition on the positive semi-definiteness of the matrix on the left-hand side. This bound therefore not only places restrictions on the fluctuations of currents, but also on their correlations.

Eqs. (3) and (4) are the main results of this paper. They hold for (i) quantum and classical systems of arbitrary dimensions and (ii) undergoing arbitrarily finite time processes far from equilibrium. The steady-state scenario is also contemplated as a particular case in which the systems become macroscopically large [12]. Below we start by reviewing the physical scenarios where our results are valid. We then provide the details of the proof and discuss their physical consequences. To illustrate their usefulness, we then apply them to a two-qubit SWAP engine functioning as a nanoscale Otto cycle.

The Exchange Fluctuation Theorem (EFT) scenario -

We consider the scenario depicted in 1(a) and studied in Refs. [11–15]. An arbitrary number M of quantum systems are initially prepared in a factorized grand-canonical state $\rho = \prod_i Z_i^{-1} \exp[-\beta_i (\mathcal{H}_i - \mu_i N_i)]$, where \mathcal{H}_i, N_i are the local Hamiltonians and particle number operators and β_i, μ_i the inverse temperature and chemical potential of the i -th subsystem [46]. The quantum systems are put in contact at time $t = 0$ up to a time τ by means of an arbitrary unitary \hat{U} incorporating the effect of all interactions between the subsystems, as well as any possible external driving. The only assumption is that the external drives are time-symmetric, so that the unitary related to the time-reversed process is simply \hat{U}^\dagger . Classical systems can be treated in a similar way [11].

As a result of this time-dependent protocol, the subsystems exchange both energy and particles with each other; we denote by $Q_{\mathcal{E}_i} = \Delta \mathcal{E}_i$ and $Q_{N_i} \equiv \Delta N_i$ the integrated energy and particle currents during the time window $(0, \tau)$. Following [12, 15], the full statistics of these quantities can be shown to satisfy the FT

$$\frac{P(Q_{\mathcal{E}_1}, \dots, Q_{\mathcal{E}_M}, Q_{N_1}, \dots, Q_{N_M})}{P(-Q_{\mathcal{E}_1}, \dots, -Q_{\mathcal{E}_M}, -Q_{N_1}, \dots, -Q_{N_M})} = e^{\sum_i \beta_i (Q_{\mathcal{E}_i} - \mu_i Q_{N_i})}, \quad (5)$$

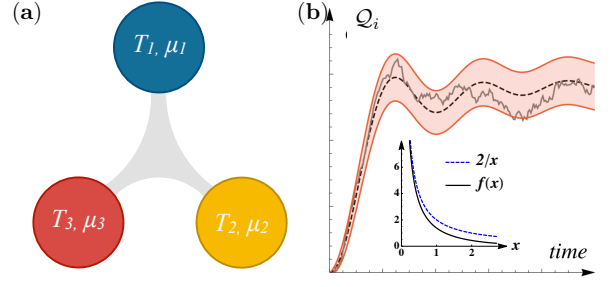


FIG. 1. (a) Exchange Fluctuation Theorem scenario: A system consisting of M (here $M = 3$) subsystems is allowed to interact by means of a unitary U . As a result, the subsystems will exchange energy and particles, amounting to net transferred charges (integrated currents) of energy $Q_{\mathcal{E}_i}$ and particles, Q_{N_i} . (b) For microscopic systems any generic Q_i will be a stochastic variable and fluctuate from one repetition of the experiment to the other, represented pictorially by the jagged gray curve. This is to be contrasted with the average charge $\langle Q_i \rangle$ shown as a dashed line. The fluctuations in Q_i are represented by the variance $\text{Var}(Q_i)$, which we illustrate here by the red interval. Inset: a plot of the function $f(x)$ in the right-hand side of Eq. (3), compared with the traditional bound $2/x$ that appears in Eq. (1).

which is of the form (2).

Variations of Eq. (5) may also be naturally constructed. Consider, for instance, the particularly relevant case of $M = 2$ subsystems. Particle conservation implies that it suffices to consider the particle charge $Q_N = \Delta N_2 = -\Delta N_1$ and hence work only with $P(Q_{\mathcal{E}_1}, Q_{\mathcal{E}_2}, Q_N)$. In addition, it may be of interest to change variables and use as thermodynamic quantities a heat charge $Q_H = -\Delta \mathcal{E}_1$ and a work charge $Q_W = \Delta \mathcal{E}_1 + \Delta \mathcal{E}_2$. The EFT for the joint distribution (5) then becomes [47, 48]

$$\frac{P(Q_H, Q_W, Q_N)}{P(-Q_H, -Q_W, -Q_N)} = e^{\delta\beta Q_H + \beta_B Q_W + \delta\beta\mu Q_N}, \quad (6)$$

where $\delta\beta = \beta_B - \beta_A$ and $\delta\beta\mu = \beta_A \mu_A - \beta_B \mu_B$ are the corresponding affinities. This result, as stated, does not assume any form of weak coupling or strict energy conservation (i.e., in general, $\Delta \mathcal{E}_1 \neq -\Delta \mathcal{E}_2$). But if that is the case, then no work is performed and it suffices to deal with $P(Q_H, Q_N)$.

Derivation of the TUR - We now turn to the derivation of our TUR bound. The starting point is a general joint probability distribution $P(Q_1, \dots, Q_n)$ satisfying (2). We first perform a change of variables to $\Sigma = \sum_i A_i Q_i$ and $Z = \sum_i z_i Q_i$, where z_i are a set of auxiliary variables. The corresponding probability distribution $P(\Sigma, Z) = \langle \delta(\Sigma - \sum_i A_i Q_i) \delta(Z - \sum_i z_i Q_i) \rangle$ will then have the same symmetry as Eq. (2) [49]. Namely,

$$\frac{P(\Sigma, Z)}{P(-\Sigma, -Z)} = e^\Sigma. \quad (7)$$

Our bound is now entirely based on the following simple question: *for fixed $\langle \Sigma \rangle$ and $\langle Z \rangle$, what is the probability distribution $P(\Sigma, Z)$, satisfying Eq. (7), which has the smallest possible variance $\text{Var}(Z)$?* We call this the minimal distribution. Our main technical contribution can then be summarized

by the following theorem:

Theorem (“TUR de force”). *For fixed finite $\langle \Sigma \rangle$ and $\langle Z \rangle$, the probability distribution $P(\Sigma, Z)$ satisfying (7), with the smallest possible variance (the minimal distribution) is the distribution*

$$P_{\min}(\Sigma, Z) = \frac{1}{2 \cosh(a/2)} \left\{ e^{a/2} \delta(\Sigma - a) \delta(Z - b) + e^{-a/2} \delta(\Sigma + a) \delta(Z + b) \right\}, \quad (8)$$

where the values of a and b are fixed by $\langle \Sigma \rangle = a \tanh(a/2)$ and $\langle Z \rangle = b \tanh(a/2)$.

The proof is given in the Supplemental Material [50]. We also note that a similar distribution also appears in Ref. [33]. For the minimal distribution (8), the variance of Z is given by

$$\text{Var}(Z)_{\min} = \langle Z \rangle^2 f(\langle \Sigma \rangle), \quad (9)$$

where $f(x)$ is the function discussed below Eq. (3) and $\text{Var}(Z)_{\min}$ is the variance of Z calculated w.r.t. P_{\min} in Eq. (8). Proving that this distribution is minimal hence implies that

$$\text{Var}(Z) \geq f(\langle \Sigma \rangle) \langle Z \rangle^2, \quad (10)$$

for any other probability distribution.

Matrix-valued TUR - We are now in the position to complete the derivation of our TUR. The bound (10) holds for a general combination $Z = \sum_i z_i Q_i$ of the charges, with arbitrary parameters z_i . Let us then write $\langle Z \rangle = \sum_i z_i q_i$, where $q_i = \langle Q_i \rangle$, and $\text{Var}(Z) = \sum_{ij} C_{ij} z_i z_j$, where $C_{ij} = \text{Cov}(Q_i, Q_j)$. Eq. (10) can then also be written as

$$\mathbf{z}^T \left(C - f(\langle \Sigma \rangle) \mathbf{q} \mathbf{q}^T \right) \mathbf{z} \geq 0.$$

But since this must be true for any set of numbers z_i , it follows that the matrix inside the parenthesis must itself be positive semidefinite. We therefore finally arrive at our main result in Eq. (4); viz., $C - f(\langle \Sigma \rangle) \mathbf{q} \mathbf{q}^T \geq 0$. The positive semidefiniteness of this matrix implies that the diagonal entries must also be non-negative. This then leads to Eq. (3).

In addition, a condition on the covariances may be obtained by using the fact that if G is a positive semidefinite matrix, then $-\sqrt{G_{ii}G_{jj}} \leq G_{ij} \leq \sqrt{G_{ii}G_{jj}}$. Applying this to Eq. (4) immediately leads to

$$f(\langle \Sigma \rangle) q_i q_j - M_{ij} \leq C_{ij} \leq f(\langle \Sigma \rangle) q_i q_j + M_{ij}, \quad (11)$$

where $M_{ij}^2 = (\text{Var}(Q_i) - f(\langle \Sigma \rangle) q_i^2)(\text{Var}(Q_j) - f(\langle \Sigma \rangle) q_j^2)$. Beside their magnitude, a particularly relevant information is also contained in the *sign* of the covariances $C_{ij} = \text{Cov}(Q_i, Q_j)$. When C_{ij} is positive (negative), values of Q_i above average imply values of Q_j above (below) average.

It is possible to find a simple criteria determining when $\text{Cov}(Q_i, Q_j)$ will have a well defined sign (namely, the same as that of $q_i q_j$). This will occur whenever the lower and upper bounds in Eq. (11) have the same sign, which amounts to

checking whether $(f q_i q_j)^2 \geq M_{ij}^2$. Using the definition of M_{ij} one then finds

$$\frac{q_i^2}{\text{Var}(Q_i)} + \frac{q_j^2}{\text{Var}(Q_j)} \geq \frac{1}{f(\langle \Sigma \rangle)}. \quad (12)$$

If this inequality is satisfied, then it is guaranteed that $\text{sign Cov}(Q_i, Q_j) = \text{sign } q_i q_j$.

Application to a microscopic engine - To illustrate our results, we consider the application of our bound to an engine composed of two qubits, with energy gaps ϵ_A and ϵ_B , interacting by means of a SWAP unitary $\hat{U} = \frac{1}{2}(1 + \hat{\sigma}_A \cdot \hat{\sigma}_B)$, where $\hat{\sigma}_i$'s are the Pauli matrices [51]. The non-resonant nature of the two qubits means that there will in general be a finite amount of work involved. As shown in Ref. [52], this work can physically be associated with the cost of turning the interaction between A and B on and off. It is not necessary to specify precisely how this takes place, however. All we need is the form of the final unitary \hat{U} .

After the qubits interact, one may reset their states by coupling them individually to two heat baths at different temperatures and allowing them to fully thermalize again (see Fig. 2(a)). Repeating the procedure sequentially then leads to a stroke-based engine operating at the Otto efficiency [53]. We assume A is in contact with the hot bath, so $\beta_A < \beta_B$. The change in energy of qubit A may thus be associated with the heat dumped into the hot reservoir, so we define $Q_H = -\Delta \mathcal{E}_A$. Similarly, the heat dumped to the cold reservoir is $Q_C = -\Delta \mathcal{E}_B$, whereas their mismatch is precisely the work, $W = -Q_H - Q_C = \Delta \mathcal{E}_A + \Delta \mathcal{E}_B$. The engine will thus be characterized by the stochastic variables Q_H and W . The corresponding probability distribution $P(Q_H, W)$, whose calculation details are presented in the Supplemental Material [50], will satisfy the EFT

$$\frac{P(Q_H, W)}{P(-Q_H, -W)} = e^{(\beta_B - \beta_A) Q_H + \beta_B W}, \quad (13)$$

which is clearly of the form (6), so that our basic framework applies.

Fig. 2(b) shows $\langle W \rangle$, $\langle Q_H \rangle$ and $\langle \Sigma \rangle = (\beta_B - \beta_A) \langle Q_H \rangle + \beta_B \langle W \rangle$ as a function of ϵ_B/ϵ_A with fixed $\beta_A/\beta_B = 1/2$. If $(\epsilon_B/\epsilon_A) < (\beta_A/\beta_B)$ the device operates as a refrigerator, consuming work from an external agent to make heat flow from the cold to the hot bath. Instead, if $(\beta_A/\beta_B) < (\epsilon_B/\epsilon_A) < 1$ it operates as a heat engine extracting useful work ($\langle W \rangle < 0$). Finally, if $(\epsilon_B/\epsilon_A) > 1$ the device operates as an accelerator, consuming external work to increase the heat flow from hot to cold.

In Fig. 2(c) and (d) we present results for the fluctuations of Q_H and W respectively. The results are compared with the bound (3) as well as the bound (1), included for comparison. As previously discussed, the bound (1) can be violated depending on the value of ϵ_B/ϵ_A . The bound (3), on the other hand, is minimal and thus can never be violated.

Finally, in Fig. 2(e) we present results for the covariance $\text{Cov}(W, Q_H)$. Studies on the correlations between thermodynamic quantities are still incipient [31]. As can be seen in the

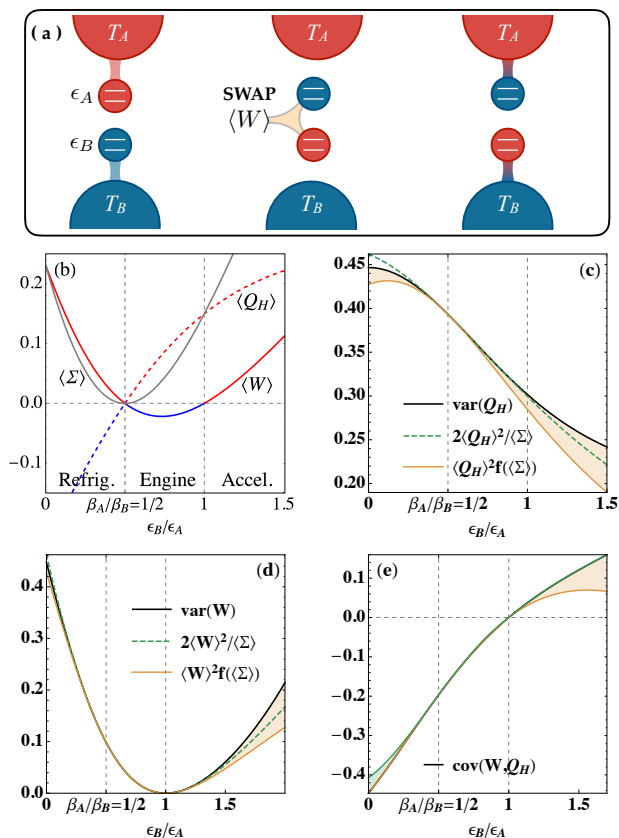


FIG. 2. Fluctuations of heat and work in a two-qubit SWAP engine. (a) Schematic operation of the engine: Two qubits thermalize with two baths at different temperatures. Then they are uncoupled from the baths and allowed to interact with each other by means of a SWAP operation, which produces a certain amount of work W . Repeating this procedure sequentially allows the device to operate as either a refrigerator, an engine or an accelerator. (b) Averages of the work $\langle W \rangle$, heat to the hot bath $\langle Q_H \rangle$ and entropy production $\langle \Sigma \rangle$ as a function of ϵ_B/ϵ_A for $\beta_A/\beta_B = 1/2$. The different regimes of operation of the engine are separated by dashed vertical lines. (c) Fluctuations in the heat to the hot bath $\text{Var}(Q_H)$. The orange line represents our bound (3) and the green-dashed line represents the bound (1), included for comparison. For small values of the detuning ϵ_B/ϵ_A , (i.e. in the refrigerator regime), one can see that the bound (1) is violated (the black line lies below the green dashed one) while (3) is always valid. (d) Same but for the fluctuations in the work $\text{Var}(W)$. (e) The correlations between heat and work, as measured by the covariance $\text{Cov}(W, Q)$. The two orange lines represent the bounds in Eq. (11).

image, in both the heat engine and the refrigerator regimes, the two quantities are negatively correlated, whereas for the accelerator they become positively correlated. The covariance in this case is bounded by the interval in Eq. (11), which is represented by the two orange lines in Fig. 2(e). For all parameters of this model, in the refrigerator regime the two bounds are always negative. Eq. (12), establishing the sign of $\text{Cov}(W, Q_H)$ is always satisfied only in the refrigerator regime. Thus, in this regime $\text{Cov}(W, Q_H) < 0$ and work and heat are always anti-correlated. In the other operation regimes, such a general claim cannot be made.

Comparison with other TURs - As discussed in the introduction, an expansion of Eq. (3) when $\langle \Sigma \rangle \ll 1$ leads to the original TUR Eq. (1). These two bounds, however, must be compared with care, as they are derived for different physical scenarios. The original TUR (1) was obtained for time-homogeneous Markovian jump-processes. Our bound, on the other hand, was derived assuming only the EFT. The two scenarios do not coincide. Indeed, as shown in Fig. 2 (c), the bound (1) can actually be violated in the EFT case. One situation for which the two scenarios could coincide is if the subsystems are macroscopically large. In this case there may exist intermediate time intervals for which the exchange of energy will resemble that of a non-equilibrium steady-state [54].

It is also important to compare our bound with the one derived in Refs. [30, 33] which, translated into our notation, implies replacing the function $f(x)$ in Eq. (3) with

$$\frac{\text{Var}(Q_i)}{\langle Q_i \rangle^2} \geq \frac{2}{e^{\langle \Sigma \rangle} - 1}. \quad (14)$$

This bound is looser than both the original TUR (1) and our generalized TUR (3). Moreover, relevant to the present letter, this bound was obtained by a different route than the one employed here, by means of a chain of inequalities [34]. However, as we have just proved, the bound (3) is the tightest possible bound and can only be saturated for a minimal distribution. As a consequence, the bound (14) can never be saturated. Indeed, in Ref. [35], by the same authors, the bound (14) was replaced by a bound structurally identical to Eq. (3).

Finally, we mention the connection with Ref. [32], where some of us have considered the non-equilibrium steady-state of a system connected to two infinite baths, a scenario where the original TUR (1) can also be violated [55, 56]. As this scenario does not satisfy an EFT, in Ref. [32] we approached the problem using the Zubarev statistical ensemble, which allowed us to show that a TUR of the form (1) also exists, but looser by a numerical factor. The two approaches therefore deal with different scenarios, but are both motivated by the same drive to generalize TURs beyond their original formulation and into the quantum regime.

Conclusions - In this Letter we have rigorously derived a new matrix-valued TUR solely as a consequence of EFT. This new tradeoff represents the tightest bound achievable on both the signal-to-noise ratio of any integrated current and for the covariance matrix between any pair of currents. Our derivation also allowed us to explicitly find the distribution saturating this ultimate bound. This result helps to answer in the affirmative the question of whether TURs, being inequalities, can also be viewed as a consequence of fluctuation theorems, much like the second law is obtained through Jensen's inequality. It hence places an important cornerstone in the direction of understanding and controlling non-equilibrium thermodynamic processes.

Acknowledgements - The authors acknowledge the International Institute of Physics in Natal, where part of this work took place, for both the financial support and the hospitality. The authors acknowledge fruitful discussions with D. Soares-

Pinto, F. B. Brito, C. Boraschi, P.-L. de Assis, F. Rouxinol, A. Amul, M. Terra Cunha and M. Motta. GTL and JG acknowledge the hospitality of the Kauli Seadi school, where part of this work was developed. GTL acknowledges the financial support from the University of São Paulo, the São Paulo Funding Agency FAPESP (grants 2017/50304-7, 2017/07973-5) and the Brazilian funding agency CNPq (Grant No. INCT-IQ 246569/2014-0). This work was supported by an SFI-Royal Society University Research Fellowship (J.G.) This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 758403).

* a.timpanaro@ufabc.edu.br

† guarnieg@tcd.ie

‡ gooldj@tcd.ie

§ gtlandi@if.usp.br

- [1] G. Benenti, G. Casati, K. Saito, and R. S. Whitney, *Physics Reports* **694**, 1 (2017).
- [2] N. Li, J. Ren, L. Wang, G. Zhang, P. Hänggi, and B. Li, *Rev. Mod. Phys.* **84**, 1045 (2012).
- [3] Y. Dubi and M. Di Ventra, *Reviews of Modern Physics* **83**, 131 (2011).
- [4] M. Josefsson, A. Svilans, A. M. Burke, E. A. Hoffmann, S. Fahlvik, C. Thelander, M. Leijnse, and H. Linke, *Nature nanotechnology* **13**, 920 (2018).
- [5] G. Gallavotti and E. G. D. Cohen, *Physical Review Letters* **74**, 2694 (1995).
- [6] C. Jarzynski, *Physical Review E* **56**, 5018 (1997).
- [7] G. E. Crooks, *Journal of Statistical Physics* **90**, 1481 (1998).
- [8] B. Piechocinska, *Physical Review A* **61**, 062314 (2000).
- [9] H. Tasaki, (2000), arXiv:0009244 [cond-mat].
- [10] J. Kurchan, (2000), arXiv:0007360 [cond-mat].
- [11] C. Jarzynski and D. K. Wójcik, *Physical Review Letters* **92**, 230602 (2004).
- [12] D. Andrieux, P. Gaspard, T. Monnai, and S. Tasaki, *New Journal of Physics* **11**, 043014 (2009).
- [13] K. Saito and Y. Utsumi, *Physical Review B - Condensed Matter and Materials Physics* **78**, 115429 (2008).
- [14] M. Esposito, U. Harbola, and S. Mukamel, *Reviews of Modern Physics* **81**, 1665 (2009).
- [15] M. Campisi, P. Hänggi, and P. Talkner, *Reviews of Modern Physics* **83**, 771 (2011).
- [16] C. Jarzynski, *Annual Review of Condensed Matter Physics* **2**, 329 (2011).
- [17] P. Hänggi and P. Talkner, *Nature Physics* **11**, 108 (2015), arXiv:1311.0275.
- [18] A. Nitzan and M. A. Ratner, *Science* **300**, 1384 (2003).
- [19] M. Galperin, M. A. Ratner, and A. Nitzan, *Journal of Physics: Condensed Matter* **19**, 103201 (2007).
- [20] J. P. Pekola and I. M. Khaymovich, *Annual Review of Condensed Matter Physics* (2018).
- [21] S. Krinner, T. Esslinger, and J.-P. Brantut, *Journal of Physics: Condensed Matter* **29**, 343003 (2017).
- [22] A. C. Barato and U. Seifert, *Physical Review Letters* **114**, 158101 (2015).
- [23] P. Pietzonka, A. C. Barato, and U. Seifert, *Physical Review E* **93**, 052145 (2016), arXiv:1512.01221.
- [24] P. Pietzonka and U. Seifert, *Physical Review Letters* **120**, 190602 (2017), arXiv:1705.05817.
- [25] P. Pietzonka, F. Ritort, and U. Seifert, *Physical Review E* **96**, 012101 (2017), arXiv:1702.07699.
- [26] T. R. Gingrich, J. M. Horowitz, N. Perunov, and J. L. England, *Physical Review Letters* **116**, 120601 (2016).
- [27] A. Dechant, *Journal of Physics A: Mathematical and General* **52**, 035001 (2018), arXiv:1809.10414.
- [28] A. C. Barato, R. Chetrite, A. Faggionato, and D. Gabrielli, *New Journal of Physics* **20**, 103023 (2018).
- [29] V. Holubec and A. Ryabov, *Physical review letters* **121**, 120601 (2018).
- [30] K. Proesmans and C. Van den Broeck, *EPL (Europhysics Letters)* **119**, 20001 (2017).
- [31] K. Macieszczak, K. Brandner, and J. P. Garrahan, *Physical Review Letters* **121**, 130601 (2018), arXiv:arXiv:1803.01904v3.
- [32] G. Guarnieri, G. T. Landi, S. R. Clark, and J. Goold, (2019), arXiv:1901.10428v2.
- [33] N. Merhav and Y. Kafri, *Journal of Statistical Mechanics: Theory and Experiment* **2010** (2010), 10.1088/1742-5468/2010/12/P12022.
- [34] Y. Hasegawa and T. Van Vu, arXiv preprint arXiv:1902.06376 (2019).
- [35] T. Van Vu and Y. Hasegawa, arXiv preprint arXiv:1904.04111 (2019).
- [36] P. P. Potts and P. Samuelsson, (2019), arXiv:1904.04913.
- [37] F. Hartmann, P. Pfeffer, S. Höfling, M. Kamp, and L. Worschech, *Physical Review Letters* **114**, 146805 (2015).
- [38] H. Thierschmann, R. Sánchez, B. Sothmann, F. Arnold, C. Heyn, W. Hansen, H. Buhmann, and L. W. Molenkamp, *Nature Nanotechnology* **10**, 854 (2015).
- [39] J. Roßnagel, S. T. Dawkins, K. N. Tolazzi, O. Abah, E. Lutz, F. Schmidt-Kaler, and K. Singer, *Science* **352**, 325 (2016), arXiv:1510.03681.
- [40] J. Klaers, S. Faelt, A. Imamoglu, and E. Togan, *Physical Review X* **7**, 031044 (2017), arXiv:1703.10024.
- [41] J. P. S. Peterson, T. B. Batalhão, M. Herrera, A. M. Souza, R. S. Sarthour, I. S. Oliveira, and R. M. Serra, , 1 (2018), arXiv:1803.06021.
- [42] C. T. Schmiegelow, V. Kaushal, J. Schulz, and U. G. Poschinger, 2018 arXiv:1808.02390.
- [43] D. J. Saunders, J. Nunn, I. A. Walmsley, R. Uzdin, and E. Poem, , 1 (2018), arXiv:1710.08716v2.
- [44] J. Gong, D. Poletti, and M. Mukherjee, , 1 (2018), arXiv:arXiv:1812.01303v1.
- [45] G. Maslennikov, S. Ding, R. Hablutzel, J. Gan, A. Roulet, S. Nimmrichter, J. Dai, V. Scarani, and D. Matuskevich, *Nature Communications* (2017), 10.1016/j.nucengdes.2009.07.007, arXiv:1702.08672.
- [46] This also contemplates the more typical scenario of a system coupled to multiple baths, but only provided the system starts in a thermal state. Otherwise, the currents will generally not obey a FT. For a more detailed discussion, c.f. Ref. [15].
- [47] N. A. Sinitsyn, *Journal of Physics A: Mathematical and Theoretical* **44**, 405001 (2011).
- [48] M. Campisi, *Journal of Physics A: Mathematical and Theoretical* **47**, 245001 (2014).
- [49] S. Lahiri and A. M. Jayannavar, *Indian Journal of Physics* **89**, 515 (2015).
- [50] *See supplemental material.*
- [51] M. Campisi, J. Pekola, and R. Fazio, *New Journal of Physics* **17**, 035012 (2015).
- [52] F. Barra, *Scientific Reports* **5**, 14873 (2015), arXiv:1509.04223.
- [53] G. De Chiara, G. Landi, A. Hewgill, B. Reid, A. Ferraro, A. J. Roncaglia, and M. Antezza, *New Journal of Physics* **20**,

113024 (2018), arXiv:1808.10450.

[54] G. T. Landi and D. Karevski, *Physical Review E* **93**, 032122 (2016), arXiv:1412.4230.

[55] B. K. Agarwalla and D. Segal, *Physical Review B* **98**, 155438

(2018).

[56] K. Ptaszyński, *Physical Review B* **98**, 085425 (2018), arXiv:1805.11301v2.

Supplemental Material

This Supplementary Material is divided in two sections. In Sec. I we provide details on the proof of Theorem 1 on the main text, which is the theorem required to demonstrate our generalized TUR. Then, in Sec. II we provide some details on how to compute the thermodynamic quantities for the 2-qubit engine example summarized in Fig. 2 of the main text. An appendix is also included in the end, where we include more technical calculations postponed from Sec. I.

I. PROOF OF THEOREM 1 OF THE MAIN TEXT.

Here we prove the following result: if we have a distribution $P(x, y)$ satisfying the symmetry relation

$$P(x, y) = P(-x, -y)e^x, \quad (\text{S1})$$

and with $\langle x \rangle$ and $\langle y \rangle$ finite, then

$$\text{Var}(y) \geq \langle y \rangle^2 f(\langle x \rangle), \quad (\text{S2})$$

with $f(x) = \text{csch}^2(g(x)/2)$, where $g^{-1}(x) = x \tanh x$.

Throughout the whole proof we call distributions obeying (S1), FT distributions. Also note that eq (S1) implies that P can be written as

$$P(x, y) = S(x, y)e^{x/2}, \quad (\text{S3})$$

where S is symmetric, which is a property we will use throughout the proof to build the distributions.

The proof is organized in four parts, as follows:

- I. A. We first show that, if we restrict ourselves to FT distributions with at most 2 points in their support and find the minimum of $\text{Var}(y)$ given the values of $\langle x \rangle$ and $\langle y \rangle$, then this minimum is reached by exactly one distribution, which we will call the minimal distribution for these values. Furthermore, the equality in (S2) holds for the minimal distribution.
- I. B. We then show (by brute force) that if P is a FT distribution with either 3 or 4 points in its support and P' is the associated minimal distribution, then $\text{Var}(y)_{P'} \leq \text{Var}(y)_P$.
- I. C. With this result, we use the law of total variance to show that this implies the bound for the case of discrete FT distributions, with finite support.
- I. D. Finally, we show that given any FT distribution ρ , with finite $\langle x \rangle$, $\langle y \rangle$ and $\langle y^2 \rangle$, we can build a sequence of discrete FT distributions with finite support, whose averages $\langle x \rangle$, $\langle y \rangle$ and $\langle y^2 \rangle$ converge to the averages of ρ .

I. A. The minimal distribution

We first recall that Eq. (S1) implies $\langle e^{-x} \rangle = 1$, which by Jensen's inequality implies that $\langle x \rangle \geq 0$, with the equality ($\langle x \rangle = 0$) implying that $\text{Var}(x) = 0$. We want to find the minimal distribution, that is the distribution with at most 2 points in its support such that $\text{Var}(y)$ is minimum given $\langle x \rangle$ and $\langle y \rangle$. We also want to show that the minimal distributions obey the equality in (S2). Since the case $\langle x \rangle = 0$ works differently, we will start with it.

The $\langle x \rangle = 0$ case

If $\langle x \rangle = 0$, then $x = 0$ everywhere in the support and Eq. (S3) implies that P is symmetric. As a consequence, every FT distribution with $\langle x \rangle = 0$ also has $\langle y \rangle = 0$. The minimum variance is trivially achieved by the distribution where $(0, 0)$ is the only point in the support, as $\text{Var}(y) = 0$ in this case. Also, it's trivial that this distribution obeys the equality in (S2).

The $\langle x \rangle > 0$ case

Since the only possible FT distribution with a 1 point support has $\langle x \rangle = 0$, then we only need to consider the case with 2 points. Eq (S3) means that a FT distribution with support $\{(-a, -b), (a, b)\}$ must be of the form

$$P(x, y) = \begin{cases} pe^{a/2} & \text{for } (x, y) = (a, b), \\ pe^{-a/2} & \text{for } (x, y) = (-a, -b), \end{cases} \quad (\text{S4})$$

where $p = 1/(2 \cosh(a/2))$. The values of a and b are then fixed by the first moments according to:

$$\begin{aligned} \langle x \rangle &= a \tanh(a/2), \\ \langle y \rangle &= b \tanh(a/2). \end{aligned} \quad (\text{S5})$$

Using the previously defined function g , defined such that $g(x)^{-1} = x \tanh(x)$, we have

$$\begin{aligned} a &= 2g(\langle x \rangle/2), \\ b &= \frac{a\langle y \rangle}{\langle x \rangle}. \end{aligned} \quad (\text{S6})$$

Since $y^2 = b^2$ for both points in the support, $\langle y^2 \rangle = b^2$, and we immediately arrive at

$$\text{Var}(y) = \langle y \rangle^2 f(\langle x \rangle) \quad (\text{S7})$$

So in this case there is exactly one distribution with the prescribed first moments and it obeys the equality in (S2).

I. B. Supports with 3 or 4 points

We now show that if a FT distribution has either 3 or 4 points in its support, then the associated minimal distribution (that is, the one with the same values of $\langle x \rangle$ and $\langle y \rangle$) has a smaller or equal $\text{Var}(y)$. By the same argument we gave in the last section, the case $\langle x \rangle = 0$ is trivial, as it implies $\langle y \rangle = 0$ and the corresponding minimal distribution has $\text{Var}(y) = 0$. As such we assume $\langle x \rangle > 0$. We also don't need to consider the case $\langle y \rangle = 0$, because in this case the minimal distribution trivially has $\text{Var}(y) = 0$, so we will also assume $\langle y \rangle \neq 0$

The 3 point case

In this case, the support is $\{(-a, -b), (0, 0), (a, b)\}$, with $a > 0, b \neq 0$, and Eq. (S3) implies P is of the form

$$P(x, y) = \begin{cases} pe^{a/2} & \text{for } (x, y) = (a, b), \\ pe^{-a/2} & \text{for } (x, y) = (-a, -b), \\ q & \text{for } (x, y) = (0, 0) \end{cases} \quad (\text{S8})$$

Imposing normalization and the values of the first moments gives

$$\begin{aligned} q + 2p \cosh(a/2) &= 1, \\ 2ap \sinh(a/2) &= \langle x \rangle, \\ 2bp \sinh(a/2) &= \langle y \rangle, \end{aligned} \quad (\text{S9})$$

whose solution is

$$\begin{aligned} p &= \frac{\langle x \rangle}{2a \sinh(a/2)}, \\ q &= 1 - \frac{\langle x \rangle}{a \tanh(a/2)}, \\ b &= \frac{a\langle y \rangle}{\langle x \rangle}. \end{aligned} \quad (\text{S10})$$

Furthermore, imposing $q \geq 0$, we have that a is a free variable that obeys the bound

$$a \tanh(a/2) \geq \langle x \rangle. \quad (\text{S11})$$

To prove that the minimal distribution has a smaller or equal $\text{Var}(y)$, we minimize

$$\langle y^2 \rangle = \frac{\langle y \rangle^2}{\langle x \rangle} \frac{a}{\tanh(a/2)},$$

over a subject to the bound (S11). This function is monotonic increasing in a , meaning that the global minimum of $\langle y^2 \rangle$ happens when a is the smallest possible. On the other hand, $a \tanh(a/2)$ is monotonic, which implies that the smallest value of a allowed by the bound (S11) is such that $a \tanh(a/2) = \langle x \rangle$ and hence $q = 0$. The distribution thus has only 2 points in the support).

The 4 point case

In this case, the support is $\{(-x_a, -y_a), (-x_b, -y_b), (x_a, y_a), (x_b, y_b)\}$ and the distribution is of the form

$$P(x, y) = \begin{cases} p_a e^{x_a/2} & \text{for } (x, y) = (x_a, y_a), \\ p_a e^{-x_a/2} & \text{for } (x, y) = (-x_a, -y_a), \\ p_b e^{x_b/2} & \text{for } (x, y) = (x_b, y_b), \\ p_b e^{-x_b/2} & \text{for } (x, y) = (-x_b, -y_b). \end{cases} \quad (\text{S12})$$

The main idea is similar to the 3 point case. One minimizes $\langle y^2 \rangle$ constrained by the normalization, the values of the first moments and the fact that $p_a, p_b \geq 0$, to find that the minima are such that either one of p_a or p_b is 0 and hence there are actually only 2 points in the support. Since one now has more variables, however, the algebra becomes substantially more complicated. In order not to hamper the readability of this Supplemental Material, we postpone this demonstration to Sec. III below.

I.C. Finite Supports

Having proven the results for 3 and 4 points in the support, it is now straightforward to generalize the argument to distributions having finite support. We make the following definitions:

Definition .1 (Symmetric Sets). A set S in \mathbb{R}^2 is said to be symmetric iff $(a, b) \in S \Leftrightarrow (-a, -b) \in S$.

Definition .2 (Reducible Sets). A set S is said to be reducible iff S is finite, symmetric and such that for every FT distribution $P(x, y)$ that has S as support, there exists another FT distribution $P'(x, y)$ with support S' such that

$$\langle x \rangle_{P'} = \langle x \rangle_P \quad \langle y \rangle_{P'} = \langle y \rangle_P \quad \text{Var}(y)_{P'} \leq \text{Var}(y)_P \quad \text{and} \quad |S'| < |S|.$$

We now consider the following lemma:

Lemma .1. If S is reducible, then so is any finite symmetric set that contains it.

Before proving the lemma, we note that what we proved in the last section was that symmetric sets with 3 or 4 points are reducible (P' in the definition we just gave would be the minimal distribution associated with $\langle x \rangle_P$ and $\langle y \rangle_P$). The reason why these are important is that any symmetric set with an even (odd) number of points has a symmetric subset with 4 (3) points in it, so proving the lemma implies that all finite symmetric sets with more than 2 points are reducible, from which it follows by induction that the minimal distribution associated with given values of $\langle x \rangle$ and $\langle y \rangle$ has the smallest possible value of $\text{Var}(y)$ out of all the FT distributions with finite support (as you can step by step remove points from the support, until you get to a support with at most 2 points, without ever increasing the variance), proving the bound for this class of distributions.

Proof of the Lemma: This lemma follows from the laws of total expectation and variance. Suppose that S is reducible and let $P(x, y)$ be a FT distribution with finite support $S \cup T$, where $S \cap T = \emptyset$. Note that this implies that T is finite and symmetric. Using the law of total variance, we have

$$\text{Var}(y) = \mathbb{E}(\text{Var}(y | I_S)) + \text{Var}(\mathbb{E}(y | I_S)), \quad (\text{S13})$$

where I_S is the random variable that indicates S (that is $I_S = 1$ if $(x, y) \in S$ and $I_S = 0$ otherwise).

Let A and B be the distributions conditioned to $I_S = 1$ and $I_S = 0$ respectively and let $p > 0$ be the probability of S as an event. Since both S and T are symmetric, this means that A (support S) and B (support T) have the correct symmetry and hence they are both FT distributions. It follows that

$$\mathbb{E}(y|I_S) = \begin{cases} \mathbb{E}(y)_A & \text{with probability } p, \\ \mathbb{E}(y)_B & \text{with probability } 1 - p, \end{cases} \quad (\text{S14})$$

and

$$\text{Var}(y|I_S) = \begin{cases} \text{Var}(y)_A & \text{with probability } p, \\ \text{Var}(y)_B & \text{with probability } 1 - p. \end{cases} \quad (\text{S15})$$

By the law of total expectation, $\mathbb{E}(\mathbb{E}(y|I_S)) = \mathbb{E}(y)$, while Eqs. (S14) and (S15) imply that

$$\text{Var}(\mathbb{E}(y|I_S)) = p\mathbb{E}(y)_A^2 + (1-p)\mathbb{E}(y)_B^2 - \mathbb{E}(y)^2. \quad (\text{S16})$$

In addition,

$$\mathbb{E}(\text{Var}(y|I_S)) = p\text{Var}(y)_A + (1-p)\text{Var}(y)_B \quad (\text{S17})$$

Combining these results then leads to,

$$\text{Var}(y)_P = p(\text{Var}(y)_A + \mathbb{E}(y)_A^2) + (1-p)(\text{Var}(y)_B + \mathbb{E}(y)_B^2) - \mathbb{E}(y)_P^2. \quad (\text{S18})$$

Finally we use that S is reducible. Since A is a FT distribution with support S , there exists another FT distribution C , with support R , satisfying $|R| < |S|$ and such that

$$\mathbb{E}(x)_C = \mathbb{E}(x)_A \quad \mathbb{E}(y)_C = \mathbb{E}(y)_A \quad \text{and} \quad \text{Var}(y)_C \leq \text{Var}(y)_A.$$

Consider then the distribution P' , describing the following random process:

- Draw a point (x, y) from the distribution P .
- If $(x, y) \notin S$, then (x, y) is the result of the draw.
- If $(x, y) \in S$, then draw (x', y') from C . (x', y') is the result of the draw.

This implies that $P' = pC + (1-p)B$, which is a FT distribution with support $R \cup T$. Furthermore B is P' conditioned to $I_S = 0$ and C is P' conditioned to $I_S = 1$. If $z = x$ or y , as drawn from P' , then

$$\mathbb{E}(z|I_S)_{P'} = \begin{cases} \mathbb{E}(z)_C & \text{with probability } p, \\ \mathbb{E}(z)_B & \text{with probability } 1 - p, \end{cases} \quad (\text{S19})$$

and from the law of total expectation one may immediately show that $\mathbb{E}(z)_{P'} = \mathbb{E}(z)_P$, implying that P and P' have the same first moments. Doing the same reasoning we did to get to Eq. (S18) for P' leads us to

$$\begin{aligned} \text{Var}(y)_{P'} &= p(\text{Var}(y)_C + \mathbb{E}(y)_C^2) + (1-p)(\text{Var}(y)_B + \mathbb{E}(y)_B^2) - \mathbb{E}(y)_{P'}^2 \\ &= p(\text{Var}(y)_C + \mathbb{E}(y)_A^2) + (1-p)(\text{Var}(y)_B + \mathbb{E}(y)_B^2) - \mathbb{E}(y)_P^2 \\ &= \text{Var}(y)_P + p(\text{Var}(y)_C - \text{Var}(y)_A) \\ &\leq \text{Var}(y)_P, \end{aligned} \quad (\text{S20})$$

Thus, P' has the same first moments as P but a smaller or equal variance. The support of P' is also smaller, since $|T \cup R| \leq |T| + |R| < |T| + |S| = |S \cup T|$. In other words, no matter what FT distribution P is, we can build another FT distribution P' with a smaller support and smaller or equal variance, so $S \cup T$ is reducible, finishing the proof.

I.D. Extension to general support

We now finally bridge the gap between finite support and the general case using the following convergence theorem:

Theorem .1. *For every FT distribution ρ , with support $\mathcal{S} \subseteq \mathbb{R}^2$ and a finite family of differentiable functions $\{\phi_j\}$, such that $\langle |\phi_j| \rangle < \infty \forall j$, there exists a sequence $\{D_n\}_{n=1}^\infty$ of discrete FT distributions with finite support, such that*

$$\lim_{n \rightarrow \infty} \langle \phi_j \rangle_{D_n} = \langle \phi_j \rangle_\rho.$$

This theorem completes the proof of our bound, because if there existed a distribution with infinite support, violating the bound

$$\text{Var}(y) \geq \langle y \rangle^2 f(\langle x \rangle) \Leftrightarrow \langle y^2 \rangle \geq \langle y \rangle^2 (f(\langle x \rangle) + 1),$$

then since f is continuous and x, y, y^2 are differentiable then the sequence $\{D_n\}_{n=1}^\infty$ associated with the family $\{x, y, y^2\}$ that the theorem guarantees exists would eventually violate the bound too, leading to a contradiction because all D_n would have finite support. An important detail is that this sequence is not necessarily weakly convergent to ρ (which is not useful for us anyway, as weak convergence guarantees only convergence of the expectations of bounded continuous functions). Instead all this sequence guarantees is convergence of $\langle x \rangle, \langle y \rangle$ and $\langle y^2 \rangle$ (or any other finite family of differentiable functions).

Proof. We begin by noticing that if P is a FT distribution

$$\begin{aligned} \int_{\mathbb{R}^2} P(x, y) \phi(x, y) dx dy &= \frac{1}{2} \int_{\mathbb{R}^2} (P(x, y) \phi(x, y) + P(-x, -y) \phi(-x, -y)) dx dy \\ &= \frac{1}{2} \int_{\mathbb{R}^2} S(x, y) (\phi(x, y) e^{y/2} + \phi(-x, -y) e^{-y/2}) dx dy \\ &= \int_{\mathbb{R}^2} T(x, y) \left(\frac{\phi(x, y) e^{y/2} + \phi(-x, -y) e^{-y/2}}{e^{y/2} + e^{-y/2}} \right) dx dy, \end{aligned} \quad (\text{S21})$$

where $T(x, y) = S(x, y) \cosh(y/2)$. Substituting $\phi = 1$ in (S21) shows us that T is a normalized distribution. Furthermore, T is symmetric, there is a one to one correspondence between T distributions and P distributions and we can write any expected value of P as an expected value of T :

$$\langle \phi \rangle_P = \left\langle \frac{\phi(x, y) e^{y/2} + \phi(-x, -y) e^{-y/2}}{e^{y/2} + e^{-y/2}} \right\rangle_T.$$

Since $\phi(x, y)$ being differentiable implies $\frac{\phi(x, y) e^{y/2} + \phi(-x, -y) e^{-y/2}}{e^{y/2} + e^{-y/2}}$ is also differentiable, then to prove the convergence of the expected values in the FT case we just need to prove the convergence of these deformed expected values in the symmetric case. In other words, proving the theorem for symmetric distributions (which will be much less cumbersome) automatically proves it for FT distributions.

Let \mathcal{S}_n denote the square $[-n, n]^2$. We can break \mathcal{S}_n into a symmetrical partition $\mathcal{S}_{n,m}$ with $(2m+1)^2$ disjoint squares. More precisely, $\mathcal{S}_{n,m} = \{I_{n,m}(k) \times I_{n,m}(q) \mid k, q \in \{-n, 1-n, \dots, n\}\}$ with

$$I_{n,m}(s) \equiv \begin{cases} \left[\frac{2ns-n}{2m+1}, \frac{2ns+n}{2m+1} \right] & \text{if } s > 0, \\ \left[\frac{2ns-n}{2m+1}, \frac{2ns+n}{2m+1} \right] & \text{if } s < 0, \\ \left[\frac{-n}{2m+1}, \frac{n}{2m+1} \right] & \text{if } s = 0. \end{cases}$$

The elements of the partition can be identified by the integer vectors \vec{k} that index them. For a given pair (n, m) we can define $s_{\vec{k}}^{n,m}$ as the square defined by \vec{k} in $\mathcal{S}_{n,m}$ and $c_{\vec{k}}^{n,m}$ as its center. Let $\rho(x, y)$ be a symmetric distribution, \vec{x} be the random vector described by it and $\phi(x, y)$ be a differentiable function. We consider the non-normalized discrete distributions $\bar{D}_{n,m}$ defined over the centers of the squares in $\mathcal{S}_{n,m}$:

$$\bar{D}_{n,m} \left(c_{\vec{k}}^{n,m} \right) = \Pr \left(\vec{x} \in s_{\vec{k}}^{n,m} \right),$$

and 0 elsewhere. Since $\mathcal{S}_{n,m}$ has $(2m+1)^2$ elements, then all $\tilde{D}_{n,m}$ have finite support. Furthermore, since the partition $\mathcal{S}_{n,m}$ is symmetric, $\tilde{D}_{n,m}$ is also symmetric. We now define

$$\mathcal{E}_{n,m}[\phi] = \sum_{\vec{k}} \phi\left(c_{\vec{k}}^{n,m}\right) \tilde{D}_{n,m}\left(c_{\vec{k}}^{n,m}\right),$$

and

$$\mathbb{E}_n[\phi] = \int_{\mathcal{S}_n} \phi(\vec{x}) \rho(\vec{x}) \, d\vec{x},$$

which is to be understood as a Riemann-Lebesgue integral. It follows that for every partition $\mathcal{S}_{n,m}$, we can use the mean value theorem to get

$$\begin{aligned} \mathbb{E}_n[\phi] &= \sum_{\vec{k}} \int_{s_{\vec{k}}^{n,m}} \phi(\vec{x}) \rho(\vec{x}) \, d\vec{x} \\ &= \sum_{\vec{k}} \phi\left(\tilde{\mu}_{\vec{k}}^{n,m}\right) \int_{s_{\vec{k}}^{n,m}} \rho(\vec{x}) \, d\vec{x} \\ &= \sum_{\vec{k}} \phi\left(\tilde{\mu}_{\vec{k}}^{n,m}\right) \tilde{D}_{n,m}\left(c_{\vec{k}}^{n,m}\right), \end{aligned} \tag{S22}$$

where $\tilde{\mu}_{\vec{k}}^{n,m} \in s_{\vec{k}}^{n,m}$, so the difference $\Delta_{n,m}[\phi] = \mathcal{E}_{n,m}[\phi] - \mathbb{E}_n[\phi]$ is

$$\Delta_{n,m}[\phi] = \sum_{\vec{k}} \tilde{D}_{n,m}\left(c_{\vec{k}}^{n,m}\right) \left(\phi\left(c_{\vec{k}}^{n,m}\right) - \phi\left(\tilde{\mu}_{\vec{k}}^{n,m}\right) \right).$$

This implies that

$$|\Delta_{n,m}[\phi]| \leq \sum_{\vec{k}} \tilde{D}_{n,m}\left(c_{\vec{k}}^{n,m}\right) \left| \phi\left(c_{\vec{k}}^{n,m}\right) - \phi\left(\tilde{\mu}_{\vec{k}}^{n,m}\right) \right|.$$

Finally, ϕ is differentiable and \mathcal{S}_n is compact, so ϕ is Lipschitz-continuous in \mathcal{S}_n , with associated constant κ_n (that is $|f\phi(\vec{x}) - \phi(\vec{y})| \leq \kappa_n |\vec{x} - \vec{y}| \, \forall \vec{x}, \vec{y} \in \mathcal{S}_n$), implying

$$|\Delta_{n,m}[\phi]| \leq \kappa_n \sum_{\vec{k}} \tilde{D}_{n,m}\left(c_{\vec{k}}^{n,m}\right) \frac{2n\sqrt{N}}{m} = \frac{2n\kappa_n \mathcal{P}_n \sqrt{2}}{m}.$$

where $\mathcal{P}_n = \Pr(\vec{x} \in \mathcal{S}_n)$. Here, the factor $2n\sqrt{N}/m$ represents the largest distance between points in $s_{\vec{k}}^{n,m}$

We now turn to the family $\{\phi_j\}$ of functions whose expectations we want to converge to the ones in ρ . For all of them $|\Delta_{n,m}[\phi_j]|$ is bounded according to (S23) (with each function having its Lipschitz constant $\kappa_{j,n}$). Furthermore, since the family is finite there exists a Lipschitz constant $\mathcal{K}_n = \max_j \{\kappa_{j,n}\}$ that works for all of them, implying

$$|\Delta_{n,m}[\phi_j]| \leq \frac{2n\mathcal{K}_n \mathcal{P}_n \sqrt{2}}{m} \, \forall j.$$

Whence,

$$|\Delta_{n,m}[\phi_j]| \leq \epsilon \quad \text{if} \quad m > \frac{2n\mathcal{K}_n \mathcal{P}_n \sqrt{2}}{\epsilon}.$$

Let $\{\epsilon_n\}$ be a sequence of positive numbers such that $\epsilon_n \rightarrow 0$ when $n \rightarrow \infty$ and let m_n be an integer such that $m_n > 2n\mathcal{K}_n \mathcal{P}_n \sqrt{2}/\epsilon_n$. It follows that $|\Delta_{n,m_n}[\phi_j]| \leq \epsilon_n \, \forall j$. Let n_0 be the smallest n such that $\mathcal{P}_n > 0$. Finally, we will prove that the sequence $\{D_n\}_{n=1}^{\infty}$ we are looking for can be built using $D_n = \tilde{D}_{n,m_n}/\mathcal{P}_n$ for $n > n_0$ (since we only care for the limits, $n \leq n_0$ is irrelevant). Since $\langle |\phi_j| \rangle < \infty$, then the Lebesgue's Dominated Convergence Theorem implies that

$$\langle \phi_j \rangle_{\rho} = \int_{\mathbb{R}^2} \phi_j(\vec{x}) \rho(\vec{x}) \, d\vec{x} = \lim_{n \rightarrow \infty} \int_{\mathcal{S}_n} \phi_j(\vec{x}) \rho(\vec{x}) \, d\vec{x}, \tag{S23}$$

which is finite. On the other hand,

$$\begin{aligned}
\left| \langle \phi_j \rangle_\rho - \langle \phi_j \rangle_{D_n} \right| &= \left| \int_{\mathbb{R}^2} \phi_j(\vec{x}) \rho(\vec{x}) \, d\vec{x} - \sum_{\vec{k}} \phi_j(c_{\vec{k}}^{n,m_n}) D_n(c_{\vec{k}}^{n,m_n}) \right| \\
&= \left| \int_{S_n^c} \phi_j(\vec{x}) \rho(\vec{x}) \, d\vec{x} + \int_{S_n} \phi_j(\vec{x}) \rho(\vec{x}) \, d\vec{x} - \frac{1}{\mathcal{P}_n} \sum_{\vec{k}} \phi_j(c_{\vec{k}}^{n,m_n}) \tilde{D}_{n,m_n}(c_{\vec{k}}^{n,m_n}) \right| \\
&= \left| \int_{S_n^c} \phi_j(\vec{x}) \rho(\vec{x}) \, d\vec{x} + \left(1 - \frac{1}{\mathcal{P}_n}\right) \int_{S_n} \phi_j(\vec{x}) \rho(\vec{x}) \, d\vec{x} - \frac{\Delta_{n,m_n}[\phi_j]}{\mathcal{P}_n} \right| \\
&\leq \left| \int_{S_n^c} \phi_j(\vec{x}) \rho(\vec{x}) \, d\vec{x} \right| + \left(\frac{1 - \mathcal{P}_n}{\mathcal{P}_n} \right) \int_{S_n} |\phi_j(\vec{x})| \rho(\vec{x}) \, d\vec{x} + \frac{|\Delta_{n,m_n}[\phi_j]|}{\mathcal{P}_n} \\
&\leq \left| \langle \phi_j \rangle_\rho - \int_{S_n} \phi_j(\vec{x}) \rho(\vec{x}) \, d\vec{x} \right| + \left(\frac{1 - \mathcal{P}_n}{\mathcal{P}_n} \right) \langle |\phi_j(\vec{x})| \rangle + \frac{\epsilon_n}{\mathcal{P}_n}.
\end{aligned} \tag{S24}$$

However, each of these three terms go to zero separately as $n \rightarrow \infty$: first, due to Eq. (S23),

$$\lim_{n \rightarrow \infty} \left| \langle \phi_j \rangle_\rho - \int_{S_n} \phi_j(\vec{x}) \rho(\vec{x}) \, d\vec{x} \right| = 0.$$

Second, since $\mathcal{P}_n \rightarrow 1$ as $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \left(\frac{1 - \mathcal{P}_n}{\mathcal{P}_n} \right) \langle |\phi_j(\vec{x})| \rangle = 0.$$

And finally, by construction,

$$\lim_{n \rightarrow \infty} \frac{\epsilon_n}{\mathcal{P}_n} = 0$$

This therefore implies that

$$\lim_{n \rightarrow \infty} \left| \langle \phi_j \rangle_\rho - \langle \phi_j \rangle_{D_n} \right| = 0,$$

or, what is equivalent,

$$\lim_{n \rightarrow \infty} \langle \phi_j \rangle_{D_n} = \langle \phi_j \rangle_\rho, \tag{S25}$$

which concludes our proof. □

II. THERMODYNAMICS OF THE TWO-QUBIT SWAP ENGINE

In this section we provide additional details on the two-qubit engine discussed in Fig. 2 of the main text. We consider two qubits with Hamiltonians $H_i = -\epsilon_i \sigma_+^i \sigma_-^i$, where $i \in \{A, B\}$, prepared in thermal states $\rho_i = e^{-\beta_i H_i} / \text{tr}(e^{-\beta_i H_i})$. The two qubits are then put to interact by means of SWAP interaction:

$$U = \frac{1}{2}(1 + \sigma^A \cdot \sigma^B) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{S26}$$

The interesting aspect of using a SWAP unitary is that it does not require one to describe the detailed microscopic (and possibly time-dependent) interaction operator $V_{AB}(t)$. It suffices to understand the overall unitary generated by this interaction.

Since the two qubits are in general not resonant, the SWAP usually performs a finite amount of work. For concreteness we assume $T_A > T_B$. The state of the two qubits are measured before and after the interaction, from which one can reconstruct $\Delta\mathcal{E}_A$ and $\Delta\mathcal{E}_B$. After this process, the two qubits are then allowed to thermalize once again with their respective baths. Hence, one can define the heat flowing to the hot bath as $Q_H = -\Delta\mathcal{E}_A$. Similarly, we define the heat flowing to the cold bath as $Q_C = -\Delta\mathcal{E}_B$. Their mismatch is the overall work that had to be performed, $W = \Delta\mathcal{E}_A + \Delta\mathcal{E}_B = -Q_H - Q_C$.

The cumulant generating function (CGF) of Q_H and W is then given by

$$C(\lambda_H, \lambda_W) = \ln \langle e^{\lambda_H Q_H + \lambda_W W} \rangle = \ln \text{tr} \left\{ U^\dagger e^{-\lambda_H H_A} e^{\lambda_W (H_A + H_B)} U e^{\lambda_H H_A} e^{-\lambda_W (H_A + H_B)} (\rho_1 \otimes \rho_2) \right\}. \quad (\text{S27})$$

From the characteristic function one can compute all cumulants of Q_H and W . In particular:

$$\langle Q_H \rangle = \frac{\partial C}{\partial \lambda_H}, \quad (\text{S28})$$

$$\langle W \rangle = \frac{\partial C}{\partial \lambda_W}, \quad (\text{S29})$$

$$\text{Var}(Q_H) = 2 \frac{\partial^2 C}{\partial \lambda_H^2}, \quad (\text{S30})$$

$$\text{Var}(W) = 2 \frac{\partial^2 C}{\partial \lambda_W^2}, \quad (\text{S31})$$

$$\text{Cov}(Q_H, W) = \frac{\partial^2 C}{\partial \lambda_W \partial \lambda_H}. \quad (\text{S32})$$

The entropy production in this case is

$$\sigma = (\beta_B - \beta_A) Q_H + \beta_B W. \quad (\text{S33})$$

Carrying out the explicit computation of the CGF (S27) for our problem leads to

$$C(\lambda_H, \lambda_W) = \ln \left\{ \frac{e^{-\epsilon_A(\lambda_H + \lambda_W) - \epsilon_B \lambda_W} \left(e^{\epsilon_A \lambda_H + \epsilon_B \lambda_W} + e^{\epsilon_A(\beta_A + \lambda_W)} \right) \left(e^{\epsilon_A \lambda_H + \beta_B \epsilon_B + \epsilon_B \lambda_W} + e^{\epsilon_A \lambda_W} \right)}{(e^{\beta_A \epsilon_A} + 1)(e^{\beta_B \epsilon_B} + 1)} \right\}. \quad (\text{S34})$$

The formulas for the first moments read

$$\langle Q_H \rangle = \epsilon_A \frac{e^{\beta_B \epsilon_B} - e^{\beta_A \epsilon_A}}{(e^{\beta_A \epsilon_A} + 1)(e^{\beta_B \epsilon_B} + 1)}, \quad (\text{S35})$$

$$\langle W \rangle = (\epsilon_B - \epsilon_A) \frac{e^{\beta_B \epsilon_B} - e^{\beta_A \epsilon_A}}{(e^{\beta_A \epsilon_A} + 1)(e^{\beta_B \epsilon_B} + 1)}. \quad (\text{S36})$$

From this one can directly verify that this system operators at the Otto efficiency:

$$\eta = -\frac{\langle W \rangle}{\langle Q_H \rangle} = 1 - \frac{\epsilon_B}{\epsilon_A}. \quad (\text{S37})$$

The expressions for the cumulants can be computed in a similar way, but are somewhat cumbersome and will thus not be displayed.

The standard fluctuation theorem is contained in the identity

$$C(\lambda_H = \beta_A - \beta_B, \lambda_W = -\beta_B) = 0, \quad (\text{S38})$$

which can be verified from Eq. (S34). This, however, does not imply that the forward and backward distributions are the same, which is a stronger condition appearing in the exchange fluctuation theorems and which formed the basis for the present work (cf Eq. (2)) of the main text.

In addition to Eq. (S38), distributions for which the forward and backward process are identical also satisfy the stronger condition,

$$C(\beta_A - \beta_B - \lambda_H, -\beta_B - \lambda_W) = C(\lambda_H, \lambda_W), \quad (\text{S39})$$

which, again, one may verify for Eq. (S34). Eq. (S39) immediately implies the detailed exchange fluctuation theorem

$$\frac{P(Q_H, W)}{P(-Q_H, -W)} = e^{(\beta_B - \beta_A)Q_H + \beta_B W}, \quad (\text{S40})$$

which is Eq. (12) of the main text.

APPENDIX: PROOF OF MINIMALITY FOR A 4-POINT DISTRIBUTION

Here we prove the minimality argument for the 4-point distribution (S12). We once again impose our constraints

$$\begin{aligned} 2p_a \cosh(x_a/2) + 2p_b \cosh(x_b/2) &= 1, \\ 2p_a x_a \sinh(x_a/2) + 2p_b x_b \sinh(x_b/2) &= \langle x \rangle, \\ 2p_a y_a \sinh(x_a/2) + 2p_b y_b \sinh(x_b/2) &= \langle y \rangle. \end{aligned} \quad (\text{S41})$$

The system (S41) can be slightly simplified if we define $q_a = 2p_a \cosh(x_a/2)$ and $q_b = 2p_b \cosh(x_b/2)$, as well as \bar{x} and \bar{y} , such that $\bar{x} \tanh(\bar{x}/2) = \langle x \rangle$ and $\bar{y} \tanh(\bar{y}/2) = \langle y \rangle$:

$$\begin{aligned} q_a + q_b &= 1, \\ q_a x_a \tanh(x_a/2) + q_b x_b \tanh(x_b/2) &= \bar{x} \tanh(\bar{x}/2), \\ q_a y_a \tanh(x_a/2) + q_b y_b \tanh(x_b/2) &= \bar{y} \tanh(\bar{y}/2). \end{aligned} \quad (\text{S42})$$

Without loss of generality we have $0 \leq x_a \leq \bar{x} \leq x_b$ (because of the mean value theorem) and $\bar{x}, \bar{y} \neq 0$ (because we are assuming $\langle x \rangle, \langle y \rangle \neq 0$). In terms of these new variables the second moment becomes

$$\begin{aligned} \langle y^2 \rangle &= 2p_a y_a^2 \cosh(x_a/2) + 2p_b y_b^2 \cosh(x_b/2) \\ &= q_a y_a^2 + q_b y_b^2. \end{aligned} \quad (\text{S43})$$

Finally, the minimal distribution associated with \bar{x} and \bar{y} is the one with support $\{(-\bar{x}, -\bar{y}), (\bar{x}, \bar{y})\}$, so we can reword what we want to prove as showing that any solution to (S42) with $q_a, q_b \geq 0$ is such that $\langle y^2 \rangle \geq \bar{y}^2$.

The first complication is that the nature of the solutions of the system (S42) change on certain cases, that we'll consider separately.

The case $x_a = x_b = \bar{x} \neq 0$: The system (S42) can be simplified to

$$\begin{aligned} q_a + q_b &= 1, \\ q_a y_a + q_b y_b &= \bar{y}, \\ q_a y_a^2 + q_b y_b^2 &= \langle y^2 \rangle, \end{aligned} \quad (\text{S44})$$

and $\langle y^2 \rangle \geq \bar{y}^2$ follows from Jensen's inequality, finishing the proof in this case.

The case $y_a = c x_a$ and $y_b = c x_b$ ($c \neq 0$): The third equation in the system (S42) is superfluous (note that this case can only happen if $\langle y \rangle = c \langle x \rangle$, which implies $\bar{y} = c \bar{x}$), leading to

$$q_a = \left(\frac{x_b \tanh(x_b/2) - \bar{x} \tanh(\bar{x}/2)}{x_b \tanh(x_b/2) - x_a \tanh(x_a/2)} \right), \quad (\text{S45})$$

$$q_b = \left(\frac{\bar{x} \tanh(\bar{x}/2) - x_a \tanh(x_a/2)}{x_b \tanh(x_b/2) - x_a \tanh(x_a/2)} \right), \quad (\text{S46})$$

and

$$\frac{\langle y^2 \rangle}{c^2} = x_a^2 + (\bar{x} \tanh(\bar{x}/2) - x_a \tanh(x_a/2)) \left(\frac{x_b^2 - x_a^2}{x_b \tanh(x_b/2) - x_a \tanh(x_a/2)} \right).$$

The point is that this means $\langle y^2 \rangle$ is monotonic increasing with x_b , for $x_b > 0$. To see this, it suffices to show that the function

$$\frac{x_b \tanh(x_b/2) - x_a \tanh(x_a/2)}{x_b^2 - x_a^2}, \quad (\text{S47})$$

is decreasing in x_b . Since x^2 is increasing and invertible for $x > 0$, if we define $z_{a(b)} = x_{a(b)}^2$, then saying (S47) is decreasing with x_b , for $x_b \geq 0$, is the same as saying that

$$\frac{\sqrt{z_b} \tanh(\sqrt{z_b}/2) - \sqrt{z_a} \tanh(\sqrt{z_a}/2)}{z_b - z_a}, \quad (\text{S48})$$

is decreasing with z_b , for $z_a, z_b \geq 0$, which follows from $\sqrt{x} \tanh(\sqrt{x}/2)$ being concave for $x > 0$. If $\langle y^2 \rangle$ is increasing with x_b , then its global minimum happens when x_b is the smallest possible. From Eqs. (S45) and (S46), we see that this means $x_b = \bar{x}$. But then $q_a = 0$ and the minimal distribution is the one that attains the minimum.

All other cases: The general argument is at heart similar to the one used in the last case, but with a more complex algebra. Let $z = \bar{x}y/\bar{y}$. The system (S42) can be slightly rewritten as

$$\begin{aligned} q_a + q_b &= 1, \\ q_a x_a \tanh(x_a/2) + q_b x_b \tanh(x_b/2) &= \bar{x} \tanh(\bar{x}/2), \\ q_a z_a \tanh(x_a/2) + q_b z_b \tanh(x_b/2) &= \bar{x} \tanh(\bar{x}/2), \\ q_a z_a^2 + q_b z_b^2 &= \langle z^2 \rangle. \end{aligned} \quad (\text{S49})$$

Solving the first 2 equations in the system (S49) we have again

$$\begin{aligned} q_a &= \frac{x_b \tanh(x_b/2) - \bar{x} \tanh(\bar{x}/2)}{x_b \tanh(x_b/2) - x_a \tanh(x_a/2)}, \\ q_b &= \frac{\bar{x} \tanh(\bar{x}/2) - x_a \tanh(x_a/2)}{x_b \tanh(x_b/2) - x_a \tanh(x_a/2)}, \end{aligned} \quad (\text{S50})$$

so $q_a, q_b > 0$ iff $x_a < \bar{x} < x_b$.

Combining the second and third equations from the system (S49) leads to:

$$q_a(z_a - x_a) \tanh(x_a/2) + q_b(z_b - x_b) \tanh(x_b/2) = 0, \quad (\text{S51})$$

which thus implies that

$$\frac{z_b - x_b}{z_a - x_a} = \frac{-q_a \tanh(x_a/2)}{q_b \tanh(x_b/2)} := \lambda, \quad (\text{S52})$$

which can equivalently be written as

$$z_b = x_b + \lambda(z_a - x_a). \quad (\text{S53})$$

This result means one can specify x_a , x_b and z_a . However, the constraints of normalization and the prescribed moments define q_a , q_b and z_b . What we want is then to find the values of x_a , x_b and z_a which yield the smallest $\langle z^2 \rangle$. Substituting Eq. (S53) we have

$$\begin{aligned} \langle z^2 \rangle &= q_a z_a^2 + q_b z_b^2 \\ &= z_a^2 (q_a + q_b \lambda^2) + 2z_a \lambda q_b (x_b - \lambda x_a) + q_b (x_b - \lambda x_a)^2 \\ &= A z_a^2 + 2B z_a + C. \end{aligned} \quad (\text{S54})$$

$A > 0$, so we can readily minimize over z_a . The minimum happens at $z_a = -B/A$, where we get $\langle z^2 \rangle = E(x_a, x_b)$, with

$$E(x_a, x_b) = C - \frac{B^2}{A} = \frac{q_a q_b (x_b - \lambda x_a)^2}{q_a + \lambda^2 q_b}. \quad (\text{S55})$$

Next we multiply by $q_b^2 \tanh^2(x_b/2)$ in the numerator and denominator, then substitute λ , leading to

$$E(x_a, x_b) = \frac{q_a q_b (x_b q_b \tanh(x_b/2) + x_a q_a \tanh(x_a/2))^2}{q_a q_b^2 \tanh^2(x_b/2) + q_b q_a^2 \tanh^2(x_a/2)}. \quad (\text{S56})$$

We now note that the expected value $\langle x \rangle$ shows up in the numerator and cancel some terms out

$$\begin{aligned} E(x_a, x_b) &= \frac{\bar{x}^2 \tanh(\bar{x}/2)^2}{q_b \tanh^2(x_b/2) + q_a \tanh^2(x_a/2)} \\ &= \frac{\bar{x}^2 \tanh(\bar{x}/2)^2}{\tanh^2(x_a/2) + q_b(\tanh^2(x_b/2) - \tanh^2(x_a/2))}. \end{aligned} \quad (\text{S57})$$

Since \bar{x} is fixed, then minimizing E over x_b (for a given x_a) is equivalent to maximizing

$$\tilde{E}(x_b; x_a) = q_b(\tanh^2(x_b/2) - \tanh^2(x_a/2)), \quad (\text{S58})$$

over x_b for a given x_a . We recall that $x_a < \bar{x} < x_b$. What we want to show is that the x_b that maximizes Eq. (S58) is \bar{x} and for that it suffices to show that $\tilde{E}(x_b; x_a)$ is strictly decreasing in x_b .

We substitute Eq. (S50) to get \tilde{E} explicitly as a function of x_a and x_b :

$$\frac{\tilde{E}(x_a; x_b)}{\bar{x} \tanh(\bar{x}/2) - x_a \tanh(x_a/2)} = \frac{\tanh^2(x_b/2) - \tanh^2(x_a/2)}{x_b \tanh(x_b/2) - x_a \tanh(x_a/2)}. \quad (\text{S59})$$

Since \bar{x} and x_a are fixed and $\bar{x} > x_a$, then we just have to prove the right hand side is decreasing in x_b for $x_b > x_a$. It suffices to prove that

$$\frac{x_b \tanh(x_b) - x_a \tanh(x_a)}{\tanh^2(x_b) - \tanh^2(x_a)}, \quad (\text{S60})$$

is increasing. Since $\tanh^2(x)$ is monotonic increasing and invertible for $x > 0$, if we define $w_{a(b)} = \tanh^2(x_{a(b)})$ then saying (S60) is increasing is tantamount to saying that

$$\frac{\operatorname{arctanh}(\sqrt{w_b}) \tanh(\operatorname{arctanh}(\sqrt{w_b})) - \operatorname{arctanh}(\sqrt{w_a}) \tanh(\operatorname{arctanh}(\sqrt{w_a}))}{w_b - w_a} = \frac{\sqrt{w_b} \operatorname{arctanh}(\sqrt{w_b}) - \sqrt{w_a} \operatorname{arctanh}(\sqrt{w_a})}{w_b - w_a}, \quad (\text{S61})$$

is increasing in w_b , for $0 < w_a, w_b < 1$, which follows from $\sqrt{x} \operatorname{arctanh}(\sqrt{x})$ being convex in this interval. This implies that for $\langle z^2 \rangle$ (and hence $\langle y^2 \rangle$) to be minimum, then x_b must be the smallest possible, implying $x_b = \bar{x}$, but then by Eq. (S50) we have $q_a = 0$ meaning the minimum is attained by the minimal distribution.