nature physics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41567-019-0481-0

In the format provided by the authors and unedited.

Large work extraction and the Landauer limit in a continuous Maxwell demon

M. Ribezzi-Crivellari[1,2] and F. Ritort[1,3]*

[1]Condensed Matter Physics Department, University of Barcelona, Barcelona, Spain. [2]Laboratoire de Biochimie, Institute of Chemistry, Biology and Innovation (CBI), UMR 8231, ESPCI Paris/CNRS, PSL Research University, Paris, France. [3]CIBER-BBN de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Sanidad Carlos III, Madrid, Spain. *e-mail: fritort@gmail.com

Supplementary information of "*Large work extraction and the Landauer limit in the Continuous Maxwell Demon*"

M. Ribezzi-Crivellari[a,b] and F. Ritort[a,c]

[a]Condensed Matter Physics Department, University of Barcelona, C/Marti i Franques s/n, 08028 Barcelona, Spain.

[b]Laboratoire de Biochimie, Institute of Chemistry, Biology and Innovation (CBI), UMR 8231, ESPCI Paris/CNRS, PSL Research University, 10 rue Vauquelin, 75231-Paris Cedex 05, France

[c]CIBER-BBN de Bioingeniería, Biomateriales y Nanomedicin a, Instituto de Sanidad Carlos III, Madrid, Spain.

S1: Information-content calculation for the two-states model

The two-compartment single-particle system and the single DNA pulling experiment can be modeled by a two-states system $\sigma = 0{,}1$ with kinetic rates $k_{\sigma\to\sigma'}$ for the system to change state $\sigma \to \sigma'$. The equilibrium occupancies are given, $P_0 = k_{1\to0}/R$; $P_1 = k_{0\to1}/R$, where $R = k_{1\to0} + k_{0\to1}$ is the full relaxation rate. Rates satisfy the detailed balance condition $k_{0\to1}/k_{1\to0} = P_1/P_0$ with $P_0 = \frac{1}{1+e^{\phi}}$; $P_1 = \frac{1}{1+e^{-\phi}}$. The dimensionless quantity ϕ stands for the free energy difference (in k_BT units) between states 0 and 1,

$$\phi = (G_0 - G_1)/k_BT = \Delta G/k_BT = -\log\left(\frac{P_0}{P_1}\right) \quad . \tag{S1.1}$$

For the two-compartment model $\phi = \log(V_1/V_0)$ whereas in the single DNA pulling experiment ϕ equals the free energy difference between the folded and unfolded states. Let $W_t(\sigma|\sigma')$ be the conditional probability of the system being in state σ at time t if it is in state σ' at time 0. It satisfies the following equation,

$$\frac{\partial W_t(\sigma|\sigma')}{\partial t} = \sum_{\sigma''}[\, k_{\sigma''\to\sigma} W_t(\sigma''|\sigma') - k_{\sigma\to\sigma''} W_t(\sigma|\sigma')] \tag{S1.2}$$

where $\sum_\sigma W_t(\sigma|\sigma') = 1$. The equations are readily solved

$$W_t(1|0) = \frac{k_{0\to1}}{R}(1 - e^{-Rt}) \;;\; W_t(0|0) = 1 - W_t(1|0) \tag{S1.3}$$

$$W_t(0|1) = \frac{k_{1\to0}}{R}(1 - e^{-Rt}) \;;\; W_t(1|1) = 1 - W_t(0|1) \tag{S1.4}$$

The calculation of the information-content I of the stored sequences follows basic steps in information theory. The stored sequences are defined by $\mathbf{0_n} = \left\{\overbrace{\mathbf{0},\ldots\mathbf{0}}^{\boldsymbol{n}}, \mathbf{1}\right\}$

for 0-cycles and $\mathbf{1_n} = \left\{\overbrace{\mathbf{1},\ldots\mathbf{1}}^{\boldsymbol{n}},\mathbf{0}\right\}$ for 1-cycles ($n \geq 1$). The first bit in the cycle specifies the first measurement outcome and the corresponding cycle class, whereas the last bit indicates that the system has changed state and the measurement outcome changes bit (1 for 0-cycles and 0 for 1-cycles). Measurements are made every τ and we assume measurements are free of error. The probability of the sequences $\mathbf{0_n}, \mathbf{1_n}$ are defined as p_n, q_n respectively. These are given by, $p_n = P_0 W_\tau^{n-1}(0|0) W_\tau(1|0)$ and $q_n = P_1 W_\tau^{n-1}(1|1) W_\tau(0|1)$, and satisfy the condition, $\sum_{n=1}^{\infty}(p_n + q_n) = 1$. The information-content of the stored sequences is given by

$$I = -\sum_{n=1}^{\infty}(p_n \log p_n + q_n \log q_n) \quad . \tag{S1.5}$$

Simple algebra shows that I it can be decomposed into two terms, $I = I(C) + I(S|C)$ where C stands for the cycle class and S stands for the stored sequence. The term $I(C)$ equals the information content related to the cycle class, i.e. the information-content of the first bit in the stored sequences S. Therefore $I(C) = -P_0 \log P_0 - P_1 \log P_1$ which equals the information content in the classical MD. The term $I(S|C)$is the information-content in the ensemble of sequences S conditional to the cycle class or the first bit measurement. It is given by,

$$I(S|C) = -\Big\{P_0\left(\frac{W_\tau(0|0)}{W_\tau(1|0)}\log W_\tau(0|0) + \log W_\tau(1|0)\right) + P_1\left(\frac{W_\tau(1|1)}{W_\tau(0|1)}\log W_\tau(1|1) + \log W_\tau(0|1)\right)\Big\} \quad . \tag{S1.6}$$

Substituting Eqs.(S1.3,S1.4) in the previous expressions gives the result reported in the main text, $I = I(C) + I(S|C) = I_{min} + I_1(\tau)$ with $I_{min}, I_1(\tau)$ given in Eq. (4,5),

$$I_{min} = -\frac{P_0}{P_1}\log(P_0) - \frac{P_1}{P_0}\log(P_1) - P_0 \log P_1 - P_1 \log P_0 \tag{S1.7}$$

$$I_1(\tau) = -\frac{P_0(P_0 + P_1 e^{-R\tau})}{P_1(1-e^{-R\tau})}\log\left(1 + \frac{P_1}{P_0}e^{-R\tau}\right) - \frac{P_1(P_1 + P_0 e^{-R\tau})}{P_0(1-e^{-R\tau})}\log\left(1 + \frac{P_0}{P_1}e^{-R\tau}\right) - \log(1 - e^{-R\tau}) - \left(\frac{P_0\log(P_0)}{P_1} + \frac{P_1\log(P_1)}{P_0}\right)\frac{e^{-R\tau}}{1-e^{-R\tau}} \quad . \tag{S1.8}$$

The τ dependent contribution $I_1(\tau)$ diverges logarithmically in the limit $\tau \to 0$,

$$I_1(\tau \to 0) = -\log(R\tau) + 1 + \frac{P_0^2\log(P_0)}{P_1} + \frac{P_1^2\log(P_1)}{P_0} + O(R\tau) \; . \tag{S1.9}$$

In a similar way one can calculate the average cycle time in the CMD, t_c^{CMD}. This is equal to the average number of steps in one state before the system changes state multiplied by the time interval τ. In other words t_c^{CMD}/τ is equal to the average length of the stored sequences, $\mathbf{0_n}, \mathbf{1_n}$. It is given by,

$$t_c^{CMD}/\tau = \sum_{n=1}^{\infty}(n+1)(p_n + q_n) = 1 + \sum_{n=1}^{\infty} n(p_n + q_n) =$$

$$= 1 + \frac{P_0}{W_\tau(1|0)} + \frac{P_1}{W_\tau(0|1)} = \frac{1}{1-e^{-R\tau}}\left(\frac{1+e^{2\phi}}{e^{\phi}}\right) + 1 \quad , \tag{S1.10}$$

which for finite R, τ is greater than 3 and is given by Eq.(8) in the main text. For the relative power between the CMD and the classical MD we first calculate the average power in the CMD. To optimize the extracted power in the CMD it is natural to take the limit $\tau \to 0$ in Eq.(S1.10). Because W_{max}^{CMD} is independent of τ, maximum power is obtained in the limit $\tau \to 0$. Equation (S1.10) gives the leading behavior,

$$t_c^{CMD} \to \frac{1}{R}\left(\frac{1+e^{2\phi}}{e^{\phi}}\right) \quad . \tag{S1.11}$$

The average power is then given by,

$$P_{CMD} = \frac{W_{max}^{CMD}}{t_c^{CMD}} = k_B T R \frac{\left(\frac{\log\left(1+e^{-\phi}\right)}{1+e^{\phi}} + \frac{\log\left(1+e^{\phi}\right)}{1+e^{-\phi}}\right)}{\left(\frac{1+e^{2\phi}}{e^{\phi}}\right)} \quad . \tag{S1.12}$$

In the classical MD work extraction can be applied repeatedly, let us say every relaxation time $1/R$, yielding the optimal power,

$$P_{MD} = \frac{W_{max}^{MD}}{1/R} = k_B T R \left(\frac{\log\left(1+e^{\phi}\right)}{1+e^{\phi}} + \frac{\log\left(1+e^{-\phi}\right)}{1+e^{-\phi}}\right) \tag{S1.13}$$

In the limit $|\phi| \gg 1$ we have, $P_{CMD} = k_B T R|\phi| e^{-|\phi|}\left(1 + O\left(e^{-|\phi|}\right)\right)$ and $P_{MD} = k_B T R|\phi| e^{-|\phi|}\left(1 + O(1/|\phi|)\right)$ giving the result in the text: $P_{CMD}/P_{MD} \to 1 - O(1/|\phi|)$.

S2: Efficiency of uncompressed sequences

In order to evaluate the performance of the CMD with respect to the classical MD one might consider the ratio between the average extracted work and the *uncompressed* information $I_{uncom} = N \log 2$ where N is the average number of measurements performed before the pulling protocol. In the optimal condition $R\tau \gg 1$, N is given by (Eq.8 in the text),

$$N = \frac{1 + e^{2\phi}}{e^{\phi}} + 1 > 3 \qquad (S2.1)$$

In terms of the probability P, this is given by,

$$N = \frac{1}{P(1-P)} - 1 \qquad (S2.2)$$

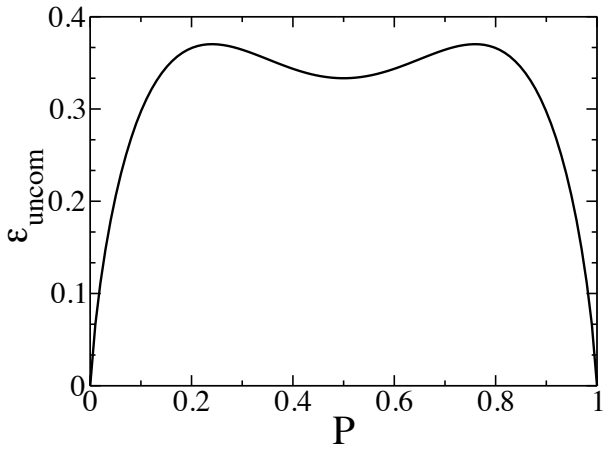

Figure S1. *The efficiency of the CMD considering the information content of uncompressed sequences.*

with $0 \leq P \leq 1$. Note that this expression is symmetric under the transformation $P \to P = 1 - P$ as expected. The *uncompressed* information I_{uncom} as defined above is larger than the minimum information $k_B T I_{min}$ defined in Eq.(4) in the paper because many bits in the stored sequence are redundant. It is important to note that the operation of compression is logically reversible [S1] so no energy is required for the compression $I_{uncom} \to I_{min}$. Therefore the efficiency using the *uncompressed* information I_{uncom} should be lower than the minimum efficiency as given by the information content I_{min} of the compressed sequences, Eq.(7). The efficiency ϵ_{uncom} obtained using the *uncompressed* information I_{uncom} equals,

$$\epsilon_{uncom} = \frac{W_{max}^{CMD}}{k_B T I_{uncom}} = \frac{-P \log(1-P) - (1-P) \log P}{\left(\frac{1}{P(1-P)} - 1\right) \log 2} \qquad (S2.3)$$

This efficiency has a local minimum equal to 1/3 at $P = 1/2$ and two absolute maxima at $P = \frac{1}{4}, \frac{3}{4}$ with $\epsilon_{uncom}^{max} = (2 \log 2 - (\log 3)/4)/(13 \log 2/3) \approx 0.37$. The dependence of ϵ_{uncom} on P is shown in Figure S1. In comparison to the efficiency as

given by the compressed sequences (Eq.7 in the paper) ϵ_{uncom} is always lower than ϵ and vanishes in the limit $P \to 0$,1. It might seem fair to compare the work extracted to the *uncompressed* information. However the true information (equal to the minimum amount of energy required to erase the information content of the stored sequences) equals the information defined by I_{min} rather than I_{uncom}.

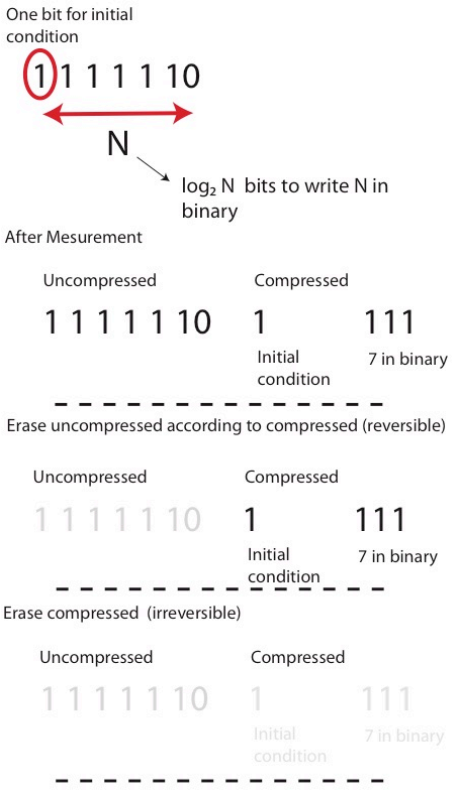

Figure S2: *The total energy expenditure for measurement erasure is set by* I_{min} *rather than* I_{uncom} *(see text for details).*

A related argument could be made to emphasize this. Let us consider the information-content of a typical string produced by the CMD as shown in Figure S2. Clearly the information content of the string can be reduced to 1) the first bit and 2) the length of the string of identical symbols. We refer to this representation as the 'compressed information'. The 'compressed information' can be stored in $1+\log_2 N$ bits : one for the first measurement and $\log_2 N$ to store N in binary format. Here N is taken as the average number of bits. However we should emphasize that N changes from string to string making the argument asymptotically valid only in the limit of large N. This condition is met in the limits $P_0 \to 0$ or 1.

It is important to realize that compression is a logically reversible operation that does not require energy expenditure [Ref S1]. Therefore in a first step we can reversibly compress the multiple-bits stored sequences into just one bit plus the additional $\log_2 N$ bits. To complete the cycle one still needs to erase the $1+\log_2 N$ bits where we have stored the compressed information. This requires a total amount of work equal to:

$$W_{eras} = k_B T \log 2(\log_2 N + 1) \simeq k_B T \log N \quad (N \gg 1) \qquad (S2.4) \quad .$$

This expression coincides with the expression $k_B T I_{min}$ obtained by the CMD in the limits $P_0 \to 0$ and 1. To show this let us consider the case $P_0 \to 0$ (the same argument applies if $P_0 \to 1$). The average number of bits (or length) of the stored sequences is given by the Eq. (S2.2) . This is,

$$N = \frac{1}{P_0(1-P_0)} - 1 \qquad (S2.5)$$

The average number of bits in the limit $P_0 \to 0$ diverges as $N \to 1/P_0$ and therefore $W_{eras} \to -k_B T \log P_0$. This should be compared with the expression of $k_B T I_{min}$ (Eq.4) in the manuscript in the same limit, $k_B T I_{min} \to -k_B T \log P_0$.

Summing up, a minimum of two measurements is necessary for our Szilard motor to operate. Indeed, At $P_0 = P_1 = 1/2$, data compression is not effective, three bits are necessary on average and the CMD is way less efficient than the classical MD in this regime. Efficiency is 33% for the CMD and 100% for the classical MD. Still, as we mentioned above, the CMD attains 100% efficiency asymptotically if $P_0 \to 0$ or 1. This is due to the fact that compression in this regime can reduce the large strings of experimental measurements to a few bits of relevant information. It is in this regime that the CMD can deliver arbitrary large amounts of work. The calculation of the compression of information $I_{uncom} \to I_{min}$ for arbitrary P_0 along the reasoning outlined above requires including the contribution of the variable length N of the stored sequences. In fact, the average value of N given in Eq. (S2.5) does not account alone for the full information-content of the ¨book¨ containing compressed sequences in the above format (one first bit plus ($\log_2$N) bits for the length). In this regard the expression $W_{eras} \to -k_B T \log P_0$ is just a lower bound to the exact expression for the minimum information-content,$k_B T I_{min}$ as given in (Eq.4) in the manuscript.

S3 : Mean work per cycle in a Szilard engine, the adiabatic case

In order to understand work extraction from the Szilard Engine we shall compute the work extracted per cycle in a limit case i.e. when the forward (FW) part of the protocol is instantaneous while the reverse part of the protocol (RV) is adiabatic. Likewise in the main text, 0 (1) denotes de folded (unfolded) state respectively. The initial value of the control parameter, λ_0, will determine the probability of finding the molecule in the folded state, P_0, or in the unfolded state, P_1. The logarithm of the ratio of these two probabilities is denoted by ϕ and equals the free energy difference between the folded and the unfolded state,

$$\phi = -\log\left(\frac{P_0}{P_1}\right) = -\log\left(\frac{e^{-\beta G_0}}{e^{-\beta G_1}}\right) = \beta(G_0 - G_1) = \beta\Delta G \quad . \tag{S3.1}$$

Let us assume the molecule is initially in the unfolded state 1. The control parameter λ will be raised instantaneously to the value $\lambda_H > \lambda_0$. During this transition the system will be confined to the unfolded free energy branch G_1 and the total work performed *on* the system will be:

$$W_1^{FW} = G_1(\lambda_H) - G_1(\lambda_0) \tag{S3.2}$$

After reaching λ_H the protocol will be reversed and the control parameter will be returned adiabatically to λ_0. Along this part of the protocol the work performed *on* the system will be :

$$W_1^{RV} = G(\lambda_0) - G(\lambda_1) \quad , \tag{S3.3}$$

where the full equilibrium free energy $G(\lambda)$ and the partial free energies $G_1(\lambda), G_0(\lambda)$ are related by :

$$G(\lambda) = -k_B T \log\left(e^{-\beta G_0(\lambda)} + e^{-\beta G_1(\lambda)}\right) \quad . \tag{S3.4}$$

Now if $\lambda_H \gg \lambda_0$ thermodynamic stability implies $G_0(\lambda_H) \gg G_1(\lambda_H)$ and as a consequence we can approximate:

$$G(\lambda_H) = -k_B T \log\left(e^{-\beta G_0(\lambda_H)} + e^{-\beta G_1(\lambda_H)}\right) \approx G_1(\lambda_H) \tag{S3.5}$$

and

$$W_1^{RV} \approx G(\lambda_0) - G_1(\lambda_H) \quad . \tag{S3.6}$$

This allows us to compute the maximum average work per cycle as a function of ϕ :

$$W_1 = W_1^{RV} + W_1^{FW} \approx$$

$$\approx G(\lambda_0) - G_1(\lambda_0) = -k_B T \log\left(e^{-\beta G_0(\lambda_0)} + e^{-\beta G_1(\lambda_0)}\right) - G_U(\lambda_0) =$$

$$= -k_B T \log\left(e^{-\beta G_1(\lambda_0)}(1 + e^{-\beta \Delta G(\lambda_0)})\right) - G_1(\lambda_0) = -k_B T \log\left(1 + e^{-\phi}\right) \quad . \tag{S3.7}$$

A similar computation can be performed if the molecule is initially found in the folded state, replacing λ_H with $\lambda_L < \lambda_0$. In this case the maximum average work per cycle will be given by:

$$W_0 \approx -k_B T \log\left(1 + e^{\phi}\right) \tag{S3.8}$$

Equations (S3.7) and (S3.8) are the equivalent of the two-compartment ideal gas model given in the text with the transformation $P_0 = V_0/V = 1/(1 + e^{\phi})$. Summing up, the maximum average work per cycle in the classical MD will be given by,

$$W_{max}^{MD} = P_0 W_0 + P_1 W_1 = k_B T \left(\frac{\log\left(1+e^{\phi}\right)}{1+e^{\phi}} + \frac{\log\left(1+e^{-\phi}\right)}{1+e^{-\phi}}\right) \tag{S3.9}$$

whereas for the CMD,

$$W_{max}^{CMD} = P_0 W_1 + P_1 W_0 = k_B T \left(\frac{\log\left(1+e^{-\phi}\right)}{1+e^{\phi}} + \frac{\log\left(1+e^{\phi}\right)}{1+e^{-\phi}}\right) \tag{S3.10}$$

which are Eq.(1,3) in the main text.

S4 : Dissipation losses during the work extraction cycle.

In our experimental protocol, the 'fast change' is performed at a pulling speed of 2 nm/ms and takes approximately 10 ms. On such timescales the molecule remains in a specific folding state during the transition. As a consequence the main source of dissipation is due to the friction affecting the plastic bead used in the manipulation, an effect that is intrinsically captured in our measurements. In Figure S3 we demonstrate that this effect is negligible. Figure S3A shows a typical realization of our pulling protocol as a function of time during a work extraction cycle. The force is directly measured by detecting changes in the linear momentum of the trapping beam. Fig S3B shows the same data in terms of the control parameter λ. Here the orange points correspond to the first and fast part of the pulling (i.e. when the position of the trap is rapidly changed) whereas the red line corresponds to the second and slow part of the pulling cycle (i.e. when the original position of the trap is adiabatically recovered during the work extraction cycle). Finally the green data correspond to the partial equilibrium in the folded state as estimated from the second slow part of the pulling work extracting cycle.

The partial equilibrium corresponds to the state of the system when configurations are constrained to the folded basin but the mechanical and thermodynamic variables are equilibrated. Partially equilibrated states are described by Boltzmann-Gibbs distributions restricted to the set of configurations belonging to specific state (in this case the folded state).

In presence of strong friction effects we would expect the green and orange (and the corresponding work estimations) to differ. The total work along the cycle can be measured taking the difference between the force along the slow and fast part of the protocol. These quantities are shown in Fig S3C (green line shows the estimation based on the partial equilibrium force and orange line shows the direct measurement). In Fig S3D we show the estimated work obtained by numerical integration of the curves in Fig. S3C. for different values of the upper integration limit λ:

$$W = \int_0^{\lambda} (f_{RE} - f_{FW}) d\lambda \ . \qquad \text{(S4.1)}$$

The estimation shows that the two estimates are consistent within the experimental error, showing that friction effects are negligible. This is further illustrated in S3E, that shows how for the condition $P_0 = P_1 = 1/2$, the Landauer limit is reached by increasing the pulling speed (Red data low speed, green data medium speed, blue data fast speed, dashed line Landauer limit).

The extent of friction effects can also be estimated by introducing a simple model. In the absence of hopping the system can be effectively approximated as a series of two springs (see Fig. S4). When the trap-pipette distance is increased at constant speed v the equation of motion for the bead position, x, reads:

$$\gamma \dot{x} = -k_M x - k_T (x - \lambda) + \eta \qquad \text{(S4.2)}$$

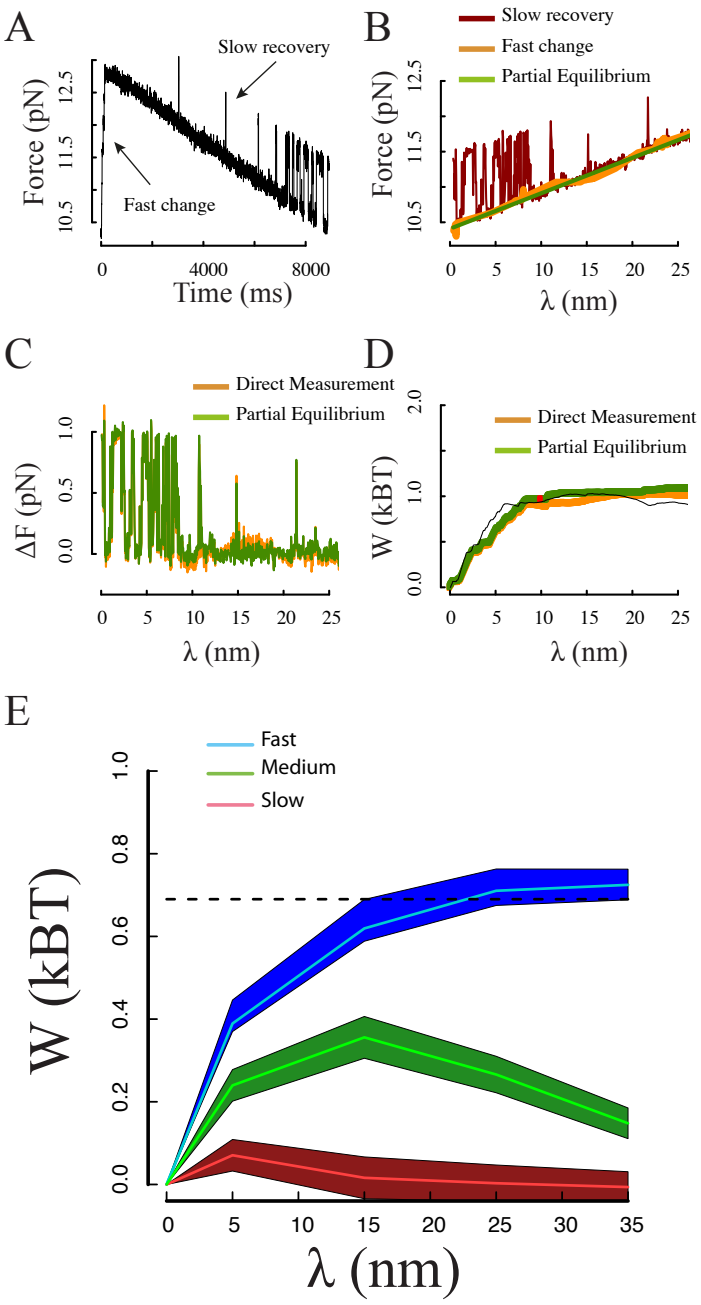

Figure S3: *A) Pulling protocol during work extraction showing the first fast change followed by the slow recovery. B) Pulling trajectory as a function of the control parameter λ. Orange line shows the fast change, red lines shows the slow recovery. Green line shows the estimated equilibrium force in the unfolded state. C) Force difference between the fast and slow part of the trajectory. The integral of this quantity gives the work extracted in the cycle. In orange we show the estimate based on the measured force and in green the estimate based on the partially equilibrated force in the unfolded state. D) Work estimates based on direct measurement (orange) and estimate of the partially equilibrated force (green). The two estimates are compatible within the experimental error. E) Average work extracted for different pulling speeds. High speed (blue data), medium speed (green data), low speed (red data).*

where η is a Gaussian white noise with correlation $< \eta_t \eta_s >= k_B T \gamma \delta(t-s)$. Moreover k_M is the stiffness of the molecular tether and k_T is the stiffness of the optical trap. Taking the average on both sides gives,

$$\gamma < \dot{x} >= -k_M < x > -k_T(< x > -\lambda). \qquad \text{(S4.3)}$$

We consider the case of constant pulling speed i.e. $\lambda = \lambda_0 + vt$, which can be solved as :

$$< x_t >= [R\lambda_0 + Rvt] + e^{-\omega t}(x_0 - R\lambda_0) + \frac{vk_T}{\omega(k_M + k_T)}(1 - e^{-\omega t}), \quad \text{(S4.4)}$$

where $\omega = (k_M + k_T)/\gamma, R = \frac{k_T}{(k_M+k_T)}$ and x_0 is the initial condition. The first term between brackets corresponds to the equilibrium solution, $x_{EQ} = [R\lambda_0 + Rvt]$ .The position of the bead relates to the force measured in the optical trap as:

$$< f >= k_T(< x > -\lambda), \qquad \text{(S4.5)}$$

and the total contribution of friction to the overall dissipation along the extracting cycle can be estimated by integrating the difference between f above ad the equilibrium force $f_{EQ} = k_T\left(x_{EQ} - \lambda\right)$:

$$W_{DISS} = \int_0^{\Delta\lambda} (< f > -f_{EQ})\, d\lambda \approx \frac{k_T v R}{\omega}\Delta\lambda \qquad \text{(S4.6)}$$

where transients have been neglected. In our system $\omega \sim 3\ \mathrm{ms}^{-1}, R = \frac{1}{20}, v = 2\frac{\mathrm{nm}}{\mathrm{ms}}, k_T = 0.06\frac{\mathrm{pn}}{\mathrm{nm}}, k_M = 1\ \mathrm{pn/nm}$ and $\gamma = 3\ \mathrm{pN\ ms/nm}$ so:

$$W_{DISS} = 0.04\frac{pn}{nm} \approx 0.01\ k_B T, \qquad \text{(S4.7)}\ ,$$

well within our experimental error.

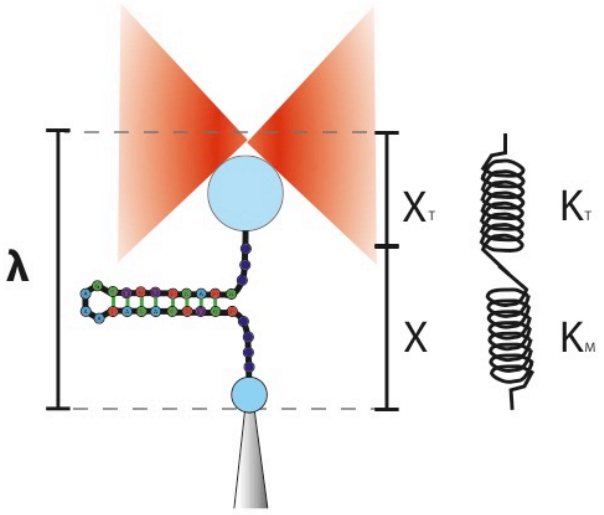

Figure S4: *Linear modelling of the experimental system. In the absence of folding/unfolding transitions the system can be modelled as a series of two springs. Here k_M and k_T are the stiffness of the molecular tether and of the optical trap respectively.*

S5 : Key qualitative features of the CMD as a work-extraction machine

The remarkable aspects of the CMD are the following:

1) The CMD can extract arbitrary large amounts of work per cycle. In contrast the Szilard engine is limited to an average of $k_B T \log 2$ per cycle.

This is an essential point, especially at the nanoscale. In a macroscopic setting we can always imagine performing a given transformation (e.g. lifting a weight, Fig. S5 upper panel) either by delivering small amounts of work several times or by one single application of a large amount of work. At the microscopic scale the situation is quite different. For example, the enzymatic breaking of a single chemical bond in a single enzymatic cycle often requires the expenditure of large amount of energy, that might well exceed $k_B T \log 2$. As a consequence, if Maxwell-Demon-like devices are to be found in nature or used in technology, they must be able to deliver a specific amount of work per cycle. This can be achieved with information-to-energy conversion devices which store multiple bit sequences, rather than just one-bit devices such as the standard Szilard engine.

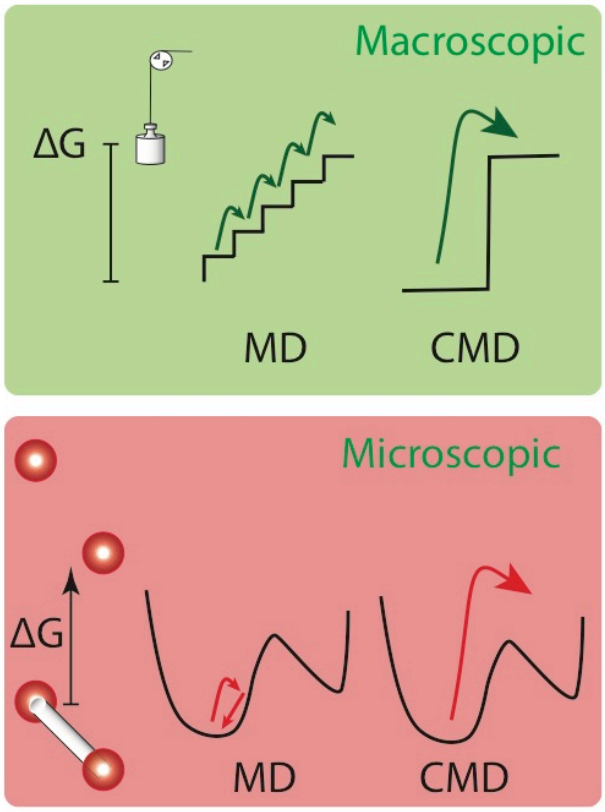

Figure S5: Macroscale v.s. microscale. *At the macroscale thermodynamic transformations can often be performed using multiple cycles each delivering a small amount of work. At the microscale, e.g. in enzymatic reactions, a fixed amount of work must be delivered in a single cycle which is possible using a CMD rather than a standard MD*

2) The efficiency of our scheme increases with the amount of work delivered per cycle. Notably efficiency does asymptotically reach 100% in the limit $P_0 \to 0$ or 1, a regime dominated by rare events. This limit matches the efficiency of the standard Szilard engine while delivering more work per cycle. It is in this sense that we consider the CMD being able to exploit rare events better than the MD does.

References

[S1] C. H. Bennett, *The thermodynamics of computation: a review,* Int. J. Theor. Phys. **21**, 905 (1983)