The Szilard Map: Defeating Maxwell's Demon

Hyun Soo Kim*
Reed College, 3203 SE Woodstock Blvd, Portland, OR, USA

The Szilard engine is a hypothetical device that extracts work through the physics of Maxwell's demon. Through the exchange of information and thermodynamic entropy, it presents a functioning system that seemingly violates the second law of thermodynamics. Through an analysis of how the demon's memory plays into the physics of the engine, however, we can see that the second law is not so easily transgressed. This report presents the Szilard map, one of such analyses, as well as my extensions on it. My application of the Szilard map into a multi-particle Szilard engine is demonstrated, and a fractal dimensional analysis of the single-particle Szilard map is discussed.

I. INTRODUCTION

A. Meet the Demon

Consider an almighty, all-knowing demon who knows, at any given time, exactly where a particle is in a closed box. The demon also has an impeccable memory, and can remember where a particle was in the box for as long as it wants to. It can perplex us mortals as much as it would like, but let us assume it was rather merciful today. It gives us a box in thermal equilibrium, with just one gas particle inside it. This particle moves about the box freely and elastically. The demon then claims this box is all we need to create energy–for free! Here is what the demon does:

1. Measure I: The demon fixes a partition in the middle of the box, splitting the box into two equal regions.

2. Measure II: The demon registers which region the single particle is now in, and encodes that into its memory.

3. Feedback I: The demon then lets the partition move freely, so that it can be isothermally pushed by the bouncing particle.

4. Feedback II: The demon takes out the partition.

5. Erase: The demon forgets its encoded memory of the particle's location, and returns to step 1.

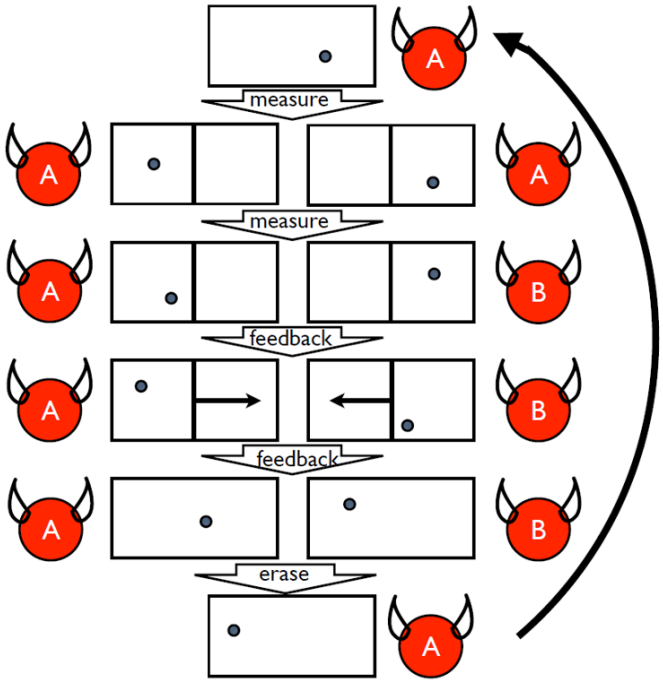

FIG. 1: The Szilard engine protocol. The demon uses two possible memory states of A and B, each corresponding to either location the particle could be relative to the partition. The particle may be either to the left of the partition or to its right.

Figure 1 depicts the demon's cycle.

While the process itself may look simple, the demon points to the physical outcome of the feedback step. By the ideal gas law, we compute the work done by the system as:

$$-\int_{V_0}^{V_1} P dV = -\int_{V_0}^{V_1} \frac{k_B T}{V} dV = k_B T \ln 2, \qquad (1)$$

where k_B is the Boltzmann constant and T is the temperature of the system.

This is actually startling, as the system has done work without receiving any external physical work during the cycle. It seems like the demon was being serious about creating energy for free!

* hskim@reed.edu

B. Important Questions

The above protocol is a type of *information* engine, a hypothetical device that uses abstract information in extracting physical work. To be precise, our protocol specifically is the Szilard engine, named after Leo Szilard. As much fun they are to think about, their implications, at first value, seem to suggest that our understanding of thermodynamics is flawed. After all, the extraction of energy out of null physical input is in clear violation of the second violation of thermodynamics–you cannot create order out of disorder for free.

We must not forget, however, that our laws of thermodynamics are much sturdier than we may sometimes think. If we reexamine what had just happened, we notice that there is a seemingly small but crucial detail that has been overlooked: the demon's memory. Unlike what it may want us to believe, the demon itself is a part of this process through its encoding and erasure of its memory about the particle's location. It seems that this is the point of focus in the demon's deception, and we must find its solution if we are to claim victory against it.

Of course, at first glance the idea of the demon's intangible memory bearing a direct effect on physical work seems unsound. Before skipping to conclusions, however, let us pose two important questions:

- How do we quantify abstract memory into physically applicable quantities?
- How does information manifest in physical dynamics?

If we find appropriate answers to these questions, we may not only find a way to defeat our demon, but even see how we may view our physical reality through tools from information theory. An actual interchange between the abstract notion of information and our physical world sounds, in a way, even more tempting than free energy creation.

C. The Szilard Map

A simple yet effective answer to these questions is the Szilard map. The Szilard map is a mapping of both the particle's location and the demon's memory into a single-two dimensional map. In the Szilard engine, we may encode the state of the engine through two coordinates: (*Position* L *or* R, *Memory State* A *or* B). L means the particle is at the left of the partition, while R means it is at its right. Either A or B can be assigned to each position as the demon's corresponding memory as to whether the particle was L or R. This two-coordinate system is easily mappable into a two-dimensional unit square, with a sub-region in the square depicting one of four possible system states.

The map, however, could also be represented by a two-dimensional gas-filled box in thermal equilibrium. You would initially have a partition through a point in the vertical axis, splitting the box into two regions. You would then uniformly fill one of those two regions with the gas, and enact the following:

1. Measure I: Fix another partition down the middle of the horizontal axis, splitting the gas into two halves.
2. Measure II: Move the walls of one half so that that half is placed above the position of the partition through the vertical.
3. Feedback I: Let the split partition through the horizontal move so that the gas may isothermally expand. Then take out the split partition.
4. Feedback II: Move the partition through the vertical to a certain point.
5. Erase: Move the upper wall down to the very first location of the partition through the vertical, and replace the moved wall with the partition. Return to step 1.

Figure 2 illustrates this process.

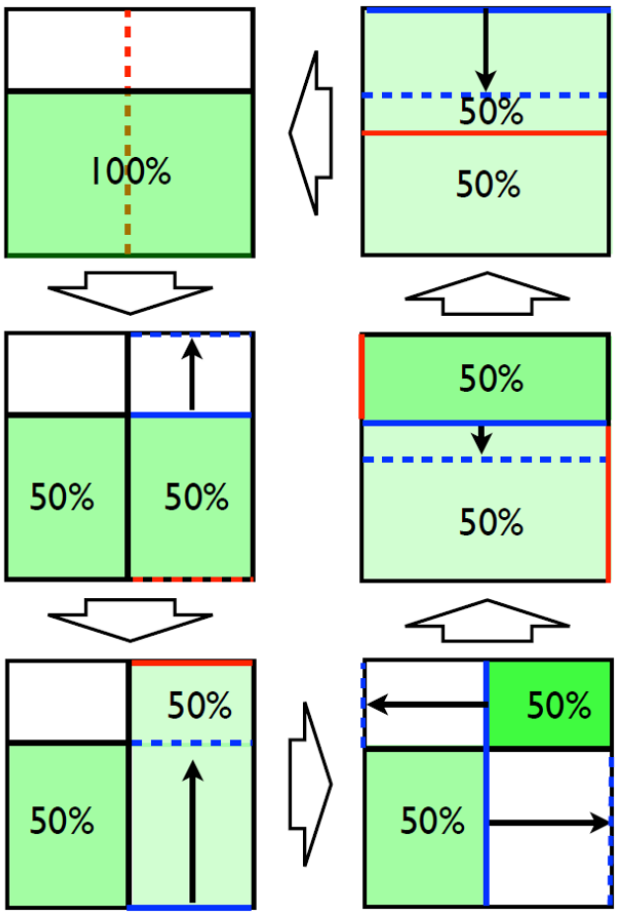

FIG. 2: Representation of the Szilard mapping with a two-dimensional gas-filled box. Filled red indicates removal, dashed red indicates addition. Filled blue indicates a wall to move, dashed blue indicates where to move it to.

The horizontal axis represents the one-dimensional location of the particle in the original engine, with the partition through it representing the demon's partition separating left and right regions. The vertical axis, on the other hand, represents the memory "location". By partitioning the vertical into two as well, we can assign the memory states A and B into each part of the partitioned vertical. This gives us the regions posing as the four possible states from the two-coordinates of the Szilard engine.

Why fill it with gas, then? The purpose of the gas is to use its density in the regions. For the gas in each partitioned region, its relative density there is equivalent to the probability a particle will be in its corresponding state for the Szilard engine. Note how depending on how the two partitions are laid out and how far the partition in step 4 goes down determines the distribution of the gase throughout the protocol.

The Szilard map itself is mathematical, and an analysis of this map as a chaotic system was made by Alec Boyd and Professor Jim Crutchfield. A continuation of the protocol would result in the mapping of Figure 3 [1].

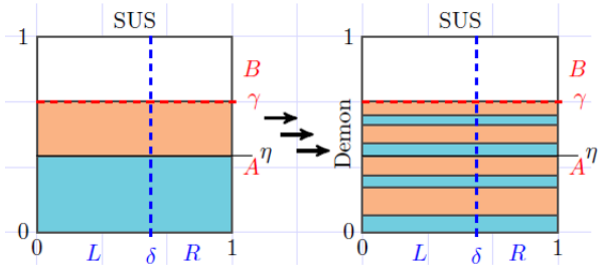

FIG. 3: A continuation of the Szilard map protocol. Note how $\eta = \delta\gamma$ for the distribution to remain uniform.

Note here that for any initial "patch" in the map would gradually be dispersed as the protocol continues over many cycles. This means that for any small uncertainty in the initial conditions of the system, its final outcome could be in rather different regions at the end of many cycles. This leads to the Szilard map being a deterministic chaotic one.

With the interpretation of the Szilard map as a two-dimensional box filled with gas, they could quantify the demon's abstract memory as a physical dimension of the box. This allowed them to calculate the expansions of the gas thermodynamically. Then, with the probabilities of the possible location-memory states, they calculated the expected values of heat released in each phase, reaching the following result [1]:

$$\langle Q_{\text{measure}} \rangle + \langle Q_{\text{erase}} \rangle = k_{\text{B}} T \ln 2 = -\langle Q_{\text{feedback}} \rangle \,, \quad (2)$$

where Q is the heat released during a phase. By Equation 1, we can see that the average sum of heat is 0 throughout the protocol. Since the process is isothermal, this means the average work done is also 0, thus preserving our known laws of thermodynamics.

This result bears significant implications in that we can physically analyze information so that information and physical work are, in effect, interchangeable. There is even a $\ln 2$ term in the work done during feedback and the work required during measurement and erasure, arising from the binary split of the system in the Szilard engine. As the logarithmic base in physics is the natural constant e while in information theory it is the binary 2, this change of base may hint at a handy equivalence between the two.

D. Objectives

The prime objective of this paper was thus to continue investigating the Szilard map. Throughout the UC Davis 2017 REU, I worked with Professor Crutchfield and Alec Boyd and worked on extensions to their Szilard map. This paper presents my extension of their Szilard map to a multi-particle Szilard engine, and my analysis of the

fractal dimension of single-particle Szilard map to further reinforce the suggested interchangeability of abstract information with physical quantities.

II. MULTI-PARTICLE SZILARD MAP

While the Szilard engine may be awesome on its own, as a physical engine it probably would have limitations with just one particle. After all, as it is the case in statistical physics, it gets most interesting when the system involves n-bodies. Thus, a method similar to the Szilard map but applicable to several particles in a system was desired.

This was where we turned to the physical realization of the box described in Section I C. The density of the gas allows us to know the probability a particle has in a state, allowing us to account for all four possibilities simultaneously. With multiple particles, however, there are only so much we can do with this approach.

What we could do with the box, however, was to implement it in computations and directly simulate the multi-particle Szilard engine with it. The correspondence to the axes to the position and memory states still held, so we could simply throw in the many gas particles into this box, rendering it equivalent to just many single particle Szilard engines stacked upon each other. Not only would this keep the advantage of quantifying the demon memory as a workable physical quantity, it also allows us to enact the Szilard map as many times as we'd desire at the same time.

The particles in our simulations were classical, elastic, and non-interacting, with the square box enclosed by hard walls. Their velocities were distributed across the Boltzmann distribution, while their positions were distributed randomly in the box. We then implemented Langevin dynamics for the box, dynamics approximating the thermal fluctuations for particles in a thermal medium. We chose Langevin dynamics as it was more practical to simulate Langevin motion than to simulate each gas molecule in the box. Under Langevin dynamics, each particle would obey the following equation of motion per time step:

$$\Delta v = -\beta v \Delta t + R(t), \tag{3}$$

where $R(t)$ is a normal distribution with $\langle R(t)\rangle = 0$ and $\langle R(t)R(t')\rangle = 2\beta k_{\mathrm{B}} T \delta(t-t')\sqrt{\Delta t}$. β is a damping constant. The $\sqrt{\Delta t}$ is from the normal distribution's properties in a finite-sized step stochastic differential equation.

Before moving on to implement the moving walls and partitions, we confirmed that our simulations were working as expected. Namely, we could check the time-average total kinetic energy of the particles, which would be predicted by thermodynamics as the gas is isothermal and the particles obey Langevin dynamics. In two dimensions, the expected average energy would be $Nk_{\mathrm{B}}T$, where N is the number of particles. Trials at various values confirmed that our simulations behaved as expected, with the time-average temperature converging to the expected value. Figure 4 depicts one of our confirmations.

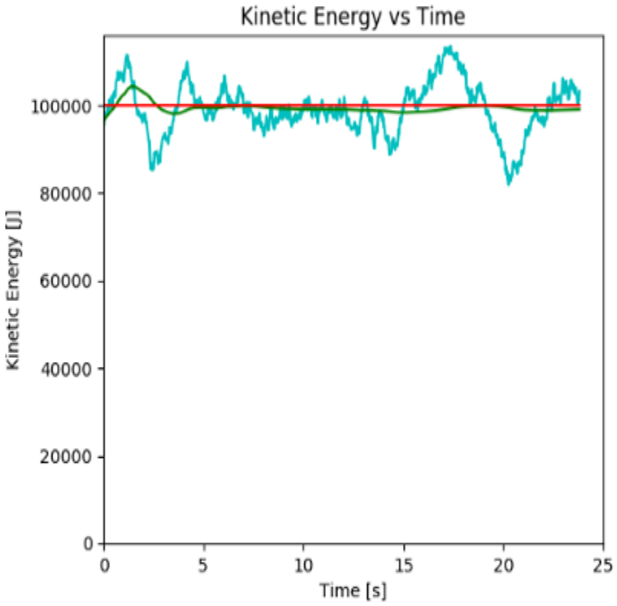

FIG. 4: An instance of testing our Langevin particles. Red is the expected value of the time-average total kinetic energy, cyan is the computed total kinetic energy at a given time, and green is its time-average.

Next, we implemented the moving partitions. The particles collided elastically with the partitions and the surrounding unmoving walls. One point to note here was that although the walls may be physical and thus exchanging momentum with the particles, it needed to look as if did not give momentum to the particles. This was because otherwise the system would be thrown off-equilibrium due to the additional kinetic energy given by the walls' motion. Thus, the box protocol needed to be quasi-static, and the walls needed to move at a speed significantly slower compared to the particles' speeds.

The partitions of the box moved according to the protocol described in Section I C. The placement of the partitions, δ for the location axis and γ for the memory axis, were adjustable parameters, as well as η, the destination of the moving axis in step 4. These parameters, on average, determine the distribution of the particles in the regions throughout the cycle.

Figure 5 is an instance of our simulations, and Figure 6 is a test of whether the quasi-static motion solves our concern of excess momentum transfer. Figure 7 is a contrasting test with non-quasi-static partitions, where we can see the effects of the additional momentum transfer between the particles and the partitions.

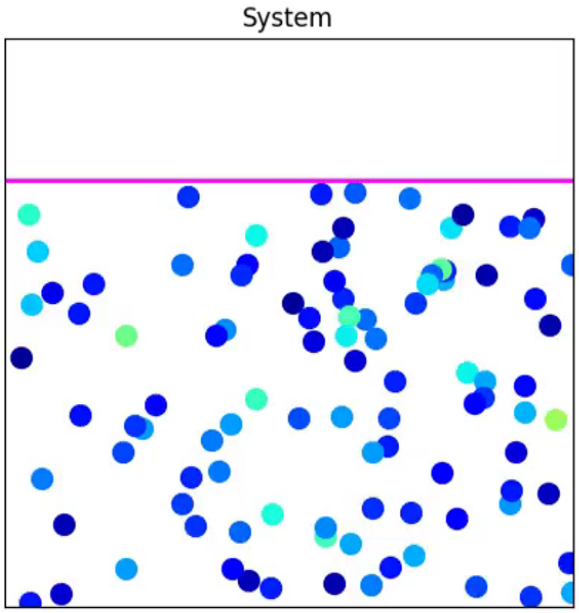

FIG. 5: An instance of our final multi-particle box.

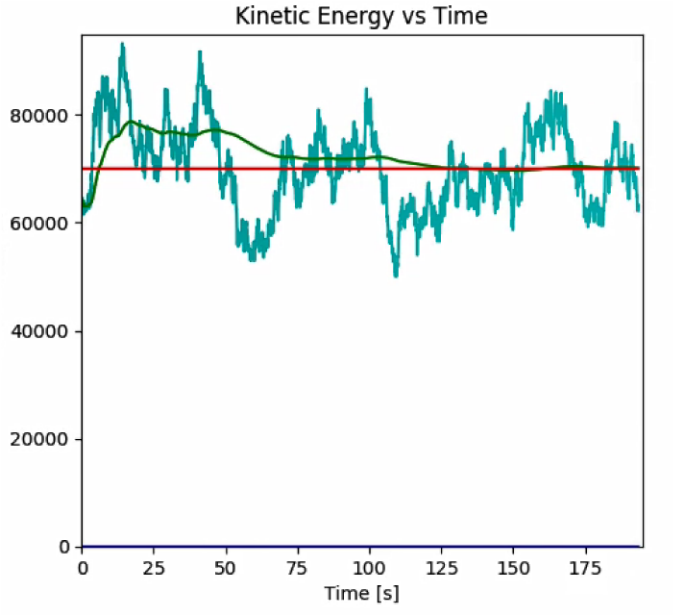

FIG. 6: An test of the quasi-static partitions. Red is the expected value of the time-average total kinetic energy, cyan is the computed total kinetic energy at a given time, and green is its time-average.

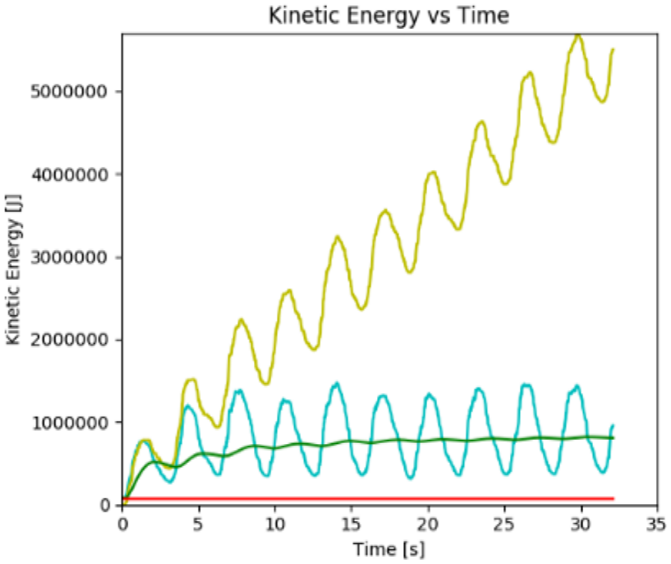

FIG. 7: An test with non-quasi-static partitions. Red is the expected value of the time-average total kinetic energy, cyan is the computed total kinetic energy at a given time, and green is its time-average. Yellow is the total work done on the particles at a given time.

This simulation provides us a method to compute the exchange of information and work numerically. Rather than attempt to calculate the average work as it has been for a single particle [1], we can run this box several times and directly compute the average total work in the box throughout the many trials. Through our trials, we confirmed that the box continues to exert, on average, 0 additional work onto the particles, as found for the single-particle case in [1]. This gives even more evidence that the Szilard engine, even with multiple particles, does not violate the second law of thermodynamics as claimed by the demon. Even with more than one particle, the demon itself uses energy in changing its memory about where the particles are.

III. FRACTAL DIMENSION

The first question of how to quantify information as a physical quantity is now an answered one. Does the Szilard map, then, also answer the second question of how information manifests in physics?

Let us return to the single-particle Szilard map. The map allows us to directly quantify the probability a particle would be in a position-memory state through the density of the gas. The density of the gas, in turn, is determined by the placement of the partitions, δ for the location axis and γ for the memory axis, and η, the destination of the moving axis in step 4 in Section I C. Since these are all geometric quantities in terms of the shape of the box as well, we may argue that the relative volume of each region in the box is also its probability.

From this view, we may take the volumes in the Szilard map as the physical representations of the information of a possible state. To be precise, we may apply the probabilistic view of the Szilard map's volumes and compute the Shannon entropy of the system. The Shannon entropy is a measure of the uncertainty of a system, which is useful in probabilistic systems such as our Szilard map. For a random variable X with probability p_i for the i-th possibility out of k total possibilities, its Shannon entropy $H(X)$ defined as:

$$H(X) = -\sum_{i}^{k} p_i \log_2 p_i. \tag{4}$$

As the protocol runs through several cycles, the distribution of the possible states changes, so the Shannon entropy will change alongside the volume.

One point of curiosity here is just how correlated information, defined by the Shannon entropy, is to the physical volumes of the system. In fact, let us return to the continuation of the Szilard map through many complete cycles, as depicted in Figure 3. One key observation of this continued mapping is that the mapping is self-similar. A repetition of the protocol results in a self-similar structure of the partitions. Thus, it becomes possible to characterise the Szilard map as a fractal system with a fractal dimension. Since the mapping is complete in 2-D, the fractal dimension d_f of the Szilard map is simply 2.

So then the question here would be then to compare the geometric fractal nature of the Szilard map, shown by the information dimension of 2, and the self-similarity in the information of the system due to its volume change. To go one step further, since we have already argued that the information manifests in the volume itself, how could we compare the propagation of Shannon entropy to the rate of the discrete partitioning in the self-similar map?

To quantify this comparison, we define the *information dimension estimate*:

$$\hat{d}_i = \frac{H(X)}{\log_2 b}, \tag{5}$$

where $H(X)$ is the Shannon entropy and b is the number of discretizations of the system. For our estimate, we assume that the system can divided up into b discrete sections, and each location-memory state region may occupy a certain number of the discrete sections. Since the information of the system is in the demon memory, we split the memory axis into b segments, and used it at the completion of each cycle, just at the end of the erasure step. Theoretically, the true information dimension of the system would be when $b \to \infty$.

To compute the information dimension, we first need to find what the probability distribution of the map is. Since the probabilities are essentially the volumes occupied by each region, we simply need to know how that changes with each cycle. At the end of a cycle, the map is "squeezed" back down below γ, while the division between the two regions is at η. After another cycle, the entire map is again pushed back down below γ, but this time with four regions, with their vertical lengths as η^2, $\eta(\delta-\eta)$, $\eta(\delta-\eta)$, and $(\delta-\eta)^2$ respectively. Continuing on this protocol, we realize that the vertical lengths of the distinct regions at the end of a protocol, are in fact, in the form of a binomial expansion, which is evident from the structure of the map with its binary regions and self-replication.

With this knowledge, we may compute $\hat{d}_i$ at a finite but large value for b.

One point we may see here is that $\hat{d}_i$ will change depending on what η is. To be precise, its relation to the partition parameters δ and γ are crucial, since that will determine how the initial partitioning is preserved by the end of the cycle. If $\eta = \delta\gamma$, in particular, the proportions of the initial partitioning will be also in the final partitioning, allowing a uniform dispersion of the mapping and the gas in its physical equivalent (see Section I C). η is therefore the key variable in our estimate of the information dimension.

We first computed the particular case of when $\eta = \delta\gamma$. Regardless of what b was, $\hat{d}_i = 2$ in this case, which quite fittingly, is also the fractal dimension of the map. This shows that when $\eta = \delta\gamma$, there is no information lost in the system compared to the volumetric self-similarity of the map.

On the other hand, when computed with cases for $\eta \neq \delta\gamma$, $\hat{d}_i < 2$, with a dependence on b. This indicates that when the map does not preserve the initial proportions of the regions, some degree of information is lost throughout the cycles. This also seems in line with the conclusion from [1] that the Szilard map is chaotic, as any slight deviation in η from $\delta\gamma$ would result in some loss of information.

Interestingly, this is also directly connected to the average work done by the Szilard engine at the end of a cycle. If $\eta = \delta\gamma$, there is no change in the proportions of the system from its initial to its final state, so expected value of the total work done by the map remains 0. If $\eta \neq \delta\gamma$, however, the system requires more work to compensate for this inefficiency. Thus we suspected a correlation between the two, and produced the result in Figure 8.

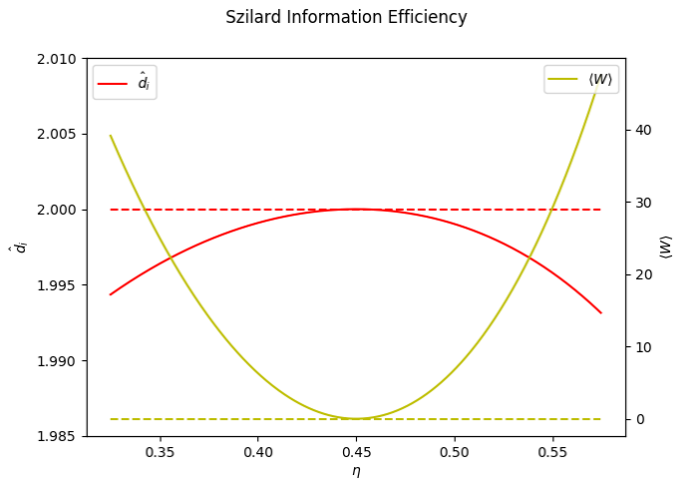

FIG. 8: A plot of $\hat{d}_i$ (in red, read with the left axis), and the expected work $\langle W \rangle$, (in yellow, read with the right axis). For this plot, $\delta = 0.6$, $\gamma = 0.75$, $b = 20000$.

The extrema of both values are at $\eta = \delta\gamma$. In fact, after some inspection, we found that the exact relation between the Shannon entropy of the system and the average work after a cycle was:

$$\langle W \rangle = k_{\mathrm{B}} T \ln 2 \left| H(\eta) - H(\delta\gamma) \right|, \tag{6}$$

where $H(\eta)$ is the Shannon entropy at a given η, and $H(\eta)$ is the Shannon entropy when $\eta = \delta\gamma$. This result is highly reminiscent of the result in not only [1], but also of Landauer's principle, which states that the information required to erase a bit of information is $k_{\mathrm{B}} T \ln 2$. This seems appropriate, as the effect of η occurs in the erasure phase. This, we believe, is a point of future development, as we extend the computation of the information dimension to multiple particles. This result seemingly reinforces a direct interchangeability between physical quantities and their respective information, bringing more light to the question of how information manifests in physical reality.

IV. CONCLUSION

Throughout my time at the UC Davis summer REU, I looked into a multi-particle extension of the Szilard map, as well as a fractal dimensional analysis of the system. Both investigations suggest a strong connection between abstract information and physical quantities, and hopefully it points to a much richer field of exploration. One point of interest would be in whether there is a minimum for the information dimension for the Szilard engine and other alternative types of information engines. Considering the significance of the structure of the system, some ideas from other fields such as combinatorial geometry and network theory may be useful, especially when we possibly turn to the quantum versions of our results or information engines with interacting particles.

I would like to thank Professor Jim Crutchfield and Alec Boyd for their work and their invaluable support during the project. I would also like to thank Ryan James, Greg Wimsatt, and the rest of the Complex Structures group at UC Davis for their advice and feedback, as well as assistance in coding issues. Finally, I am thankful to Professor Rena Zieve and the UC Davis REU, for giving me the opportunity to have my own go at defeating Maxwell's demon.

[1] A. B. Boyd and J. P. Crutchfield, Phys. Rev. Lett. **116**, 190601 (2016).