How much is used punched tape worth? A weak and a strong equivalence principle

Tommaso Toffoli (tt@bu.edu)
ECE Department, Boston University, 8 Saint Mary's St., Boston, MA 02215

We study the repeated use of a *monotonic* recording medium—such as punched tape or photographic plate—where marks can be added at any time but never erased. (For practical purposes, also the electromagnetic "ether" falls into this class.) Our emphasis is on the case where the successive users act independently and selfishly, but not maliciously; typically, the "first user" would be a blind natural process tending to degrade the recording medium, and the "second user" a human trying to make the most of whatever capacity is left.

To what extent is a length of used tape "equivalent"—for information transmission purposes—to a shorter length of virgin tape? Can we characterize a piece of used tape by an appropriate "effective length" and forget all other details? We identify two equivalence principles. The *weak* principle is exact, but only holds for a sequence of infinitesimal usage increments. The *strong* principle holds over the entire operational range of the tape, but is only approximate; nonetheless, it is quite accurate even in the worst case and is virtually exact over most of the range—becoming exact in the limit of heavily used tape.

The fact that strong equivalence does not hold *exactly*, but then it does *almost* exactly, comes as a bit of a surprise.

Rummaging through Alice's dump, Bob—a poor computer science student—has found a large amount of used punched tape "in good conditions". He doesn't care for the data that is on the tape: he would like to reuse the tape for storing his own data. He wants to be able to use a standard tape read/punch unit, which can sense holes in the tape and punch new ones but not remove holes that are already there. Of course, the storage density Bob can expect to achieve is less than with virgin tape, and will depend on the actual conditions of the tape.

To what extent is a length of used tape "equivalent"—for information transmission purposes—to a shorter length of virgin tape? Are there any qualitative differences between tapes that have been used to different degrees, or can one characterize a piece of used tape simply by its "effective length" and forget all other details?

The theme we develop is complementary to that of Rivest and Shamir[7] (also cf. [6]). They stress the information-engineering aspects of reusing a tape generated by a cooperative partner in a pre-planned context. On the other hand, we are interested in a situation where the other party, while presumed non-malicious, volunteers no cooperation and pursues independent goals (if any goals can be made out); what we typically have in mind for "the other party" is *natural processes*.[1]

The cumulative channel capacity of randomly-punched used tape was first investigated in [8] (also see references therein), some of whose results we simplify and extend. Coding algorithms that dynamically adjust to "stuck-at-0" faults on the tape (cells that will not punch) sensed during punching and "stuck-at-1" faults sensed during or before punching are discussed in [3] and [5]. A paper related in spirit if not in detailed substance to the present one is "Writing on dirty paper" by Costa[2], whose moral ("Do the best with what you have") we make our own.

If you have no time at all, read just §5—a self-contained, intuitive debriefing.

1 Orientation

Each position on the tape where a hole may appear is called a **cell**; the two possible cell states are **hole** and **blank**. The instructions to the punch unit are **punch** and **spare**, with the following results on the tape

spare	blank	$\mapsto$	blank
spare	hole	$\mapsto$	hole
punch	blank	$\mapsto$	hole
punch	hole	$\mapsto$	hole

A hole (or punch) distribution that factors into identical independent distributions for the individual cells—and is thus characterized by a single number, namely, the hole (or punch) density—will be called **canonical**. We shall assume that on each round of usage or **stage** the tape starts with a canonical hole distribution of density p and comes out with a uniform hole density p'; furthermore, we assume that the intervening punching process packs on the tape the *maximum* amount of new information compatible with the initial density p and the target density p'. According to Shannon's theorem, such maximum efficiency can asymptotically be achieved by means of sufficiently long block codes. From the above assumptions one can prove that both the punch distribution q yielded by an optimal code and the resulting hole distribution p' must be canonical as

[1] As humans become more proficient at exploiting physical mechanisms on a finer and finer scale for computational purposes, computation will look more and more like an attempt to encroach on a turf already jammed near capacity by heavy "native" traffic—the near-equilibrium bustle of microscopic matter (cf. Dyson[4]). The present study is part of a wider program aimed at exploring this kind of computational regime.

well. In what follows, all distributions are canonical unless stated otherwise.

Thus, starting with virgin tape—whose distribution is, of course, canonical with $p = 0$—we conclude that input, punch, and output are canonical at every successive stage. The result of applying a punch density q to a hole density p is a new hole density

$$p' = 1 - (1 - p)(1 - q). \qquad (1)$$

A canonical punch distribution entails that, once the input hole density p is known, there is no further advantage in knowing the position of the individual holes; in other words, punching can be carried out in a *data-blind* fashion.

Let's examine a few distinguished cases.

- If $p = 0$ the tape is blank—Bob can resell it as virgin tape.
- If $p = 1-p = 1/2$, the tape has already been utilized by Alice at its maximum information capacity of one bit per cell. That would seem to leave Bob with no room for further information storage. But remember that he doesn't care about the old information: punching new holes will destroy some of it, making room for some of his own! In fact (cf. §3 below), with a punch density $q = 3/5$, Bob can record on the tape as much as about .322 bits per cell.
- If $p = 1$, the tape carries no information for Alice—just as in the case $p = 0$. However, now there is no way Bob can put any information on it. Alice wantonly spoiled the tape.

2 Notation

If p is a probability, it will be convenient to write $\overline{p}$ for $1-p$. Thus, in (1),

$$p' = 1 - \overline{p}\,\overline{q} = \overline{\overline{p}\,\overline{q}}, \quad \text{or} \quad \overline{q} = \overline{p'}/\overline{p}.$$

We shall use natural logarithms throughout. It will be convenient to write $\overline{\ln}\, x = -\ln x$. The **self-information function**, defined as

$$y = x\,\overline{\ln}\, x,$$

will play an important role in the equivalence principles discussed here (cf. §7). The **binary entropy function**, defined by

$$H(p) = p\,\overline{\ln}\, p + \overline{p}\,\overline{\ln}\,\overline{p},$$

is the average of the self-information function over the binary distribution $\{p, \overline{p}\}$.

Both self-information and binary entropy, as defined here, measure information in natural units or **nats**. Conversion of information quantities to binary units or **bits** is achieved by explicitly factoring out the constant

$$\text{bit} = \ln 2 \approx .693;$$

thus, for example, the entropy of four equally probable messages is $\ln 4 = 2\ln 2 = 2$ bit.

If X and Y are random variables, $P(x)$ will denote the probability that $X = x$, and $P(xy)$ the probability that $X = x$ and $Y = y$. The **mutual information** between X and Y is defined as

$$\{X;Y\} = \sum_{x,y} P(xy)\,\overline{\ln}\,\frac{P(x)P(y)}{P(xy)}.$$

For more background on information theory, see the excellent introduction by Abramson[1].

3 Used tape as a monotonic binary channel

Under the above assumptions (§1), used punched tape may be viewed as a communication channel affected by monotonic noise. In the channel diagram of Fig. 1, the input variable X represents the instruction given to the punch unit while scanning a cell, and the output variable Y represents the resulting cell state. An "error" occurs when a cell spared by the punch unit turns out already to contain a hole. The conditional probability $P(\text{hole}|\text{spare})$ associated with this transition equals the current hole density p.

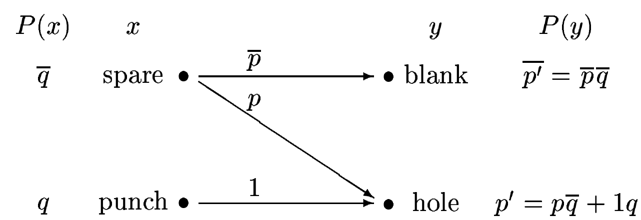

Figure 1: Channel diagram of used punched tape viewed as a monotonic-error binary channel.

From the joint and marginal distributions of X and Y, namely,

x \ y		blank	hole
		$\overline{p}\,\overline{q}$	$1-\overline{p}\,\overline{q}$
spare	$\overline{q}$	$\overline{p}\,\overline{q}$	$p\overline{q}$
punch	q	0	p

(2)

one obtains, for this channel operated at a punch density q, a mutual information

$$\Delta I = H(\overline{p}\,\overline{q}) - \overline{q}H(p) = H(p') - \frac{\overline{p'}}{\overline{p}}H(p). \qquad (3)$$

The quantity ΔI is the amount of *new* information that can be encoded on a tape having a hole density p by punching it with a density q, resulting in a new hole density $p'(p,q) = 1 - \overline{p}\overline{q}$.

The relation expressed by equation (3)—plotted in Fig. 2—completely characterizes the bulk properties of punched tape as a communication channel. The rest of this paper is devoted to extracting some of its implications.

The **capacity** C of the channel is the maximum of ΔI over all possible values of q (or, equivalently, of p'). By equating

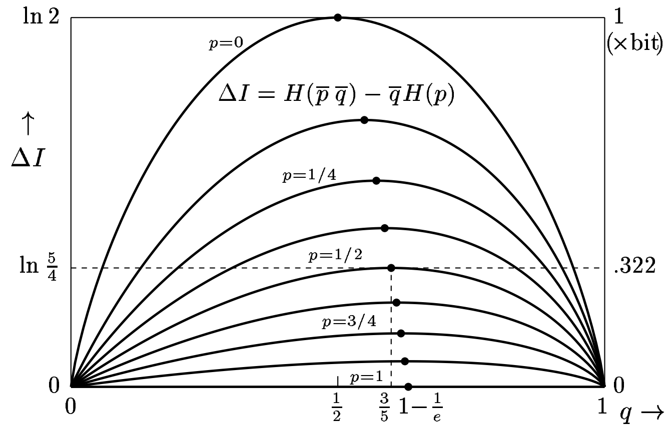

Figure 2: Mutual information ΔI of the "used tape" channel, plotted as a function of the punch density q for various values of the current hole density p treated as a parameter. The maximum of each curve is marked.

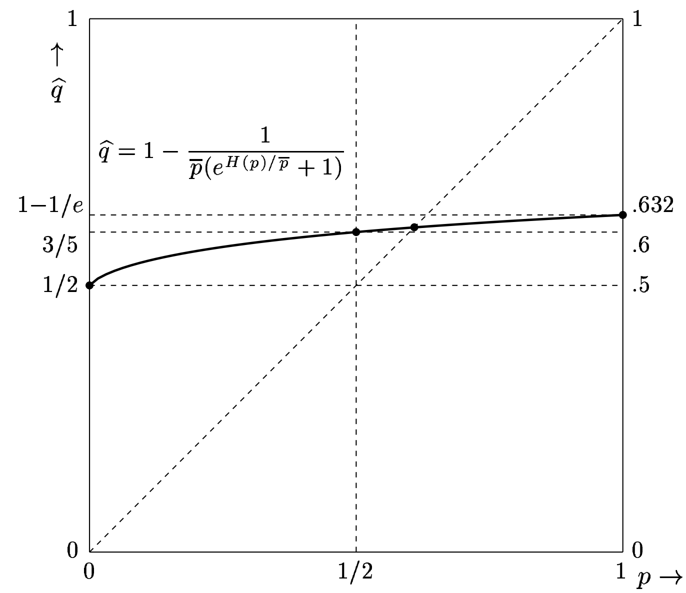

Figure 4: Plot of the optimal punch density $\widehat{q}$ vs current hole density p. The distinguished points are discussed in the text. That lying on the $\widehat{q}=p$ line is discussed in §4.2.

to zero the derivative of ΔI with respect to p',

$$\frac{d\Delta I}{dp'} = \overline{\ln}\,\frac{p'}{\overline{p'}} + \frac{H(p)}{\overline{p}} = 0,$$

one finds that this maximum occurs at

$$\widehat{q} = 1 - \frac{1}{\overline{p}(e^{H(p)/\overline{p}}+1)},\ \text{or}\ \widehat{p'} = \frac{1}{e^{-H(p)/\overline{p}}+1}, \qquad (4)$$

where ΔI attains the value

$$C = \ln(e^{-H(p)/\overline{p}}+1) = \overline{\ln}\,\widehat{p'}, \qquad (5)$$

as plotted in Fig. 3. In particular, for $p = 1/2$,

$$\widehat{q} = \frac{3}{5},\ \widehat{p'} = \frac{4}{5},\ \text{and}\ C = \ln\frac{5}{4} \approx .322\ \text{bit}.$$

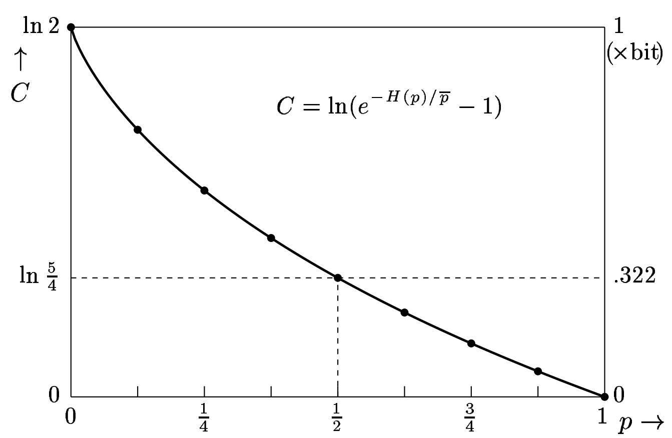

Figure 3: Channel capacity of used tape, C, as a function of the current hole density p; the dots match those of Fig. 2.

4 Cooperation and competition

For sake of contrast with the current context of selfish, independent utilization of the tape by each successive party, in the following two subsections we shall briefly discuss the possibilities of cooperation and competition.

4.1 You shall receive an hundredfold

In the two successive selfish transmission stages starting from virgin tape, Alice got 1 bit's worth of message out of each cell and Bob .322 bit, for a total of 1.322 bit. By collaborating, they could do much better[7, 6]. In fact, if Alice and Bob worked in concert, with a very simple code they could each get two bits' worth of message out of every three cells, for a total of $4/3 \approx 1.333$ bit/cell; with long block codes, they could get up to about 1.55 bit/cell. The advantages of collaboration show up even better when one can plan ahead a long series of transmission stages with a long length of tape: in this situation, the cumulative amount of message worth one can get out of an n-cell length of tape grows as $n \ln n$; therefore, the amount per unit length is unbounded!

4.2 Tape wars averted

We have seen that, if the original tape was punched at a density p by Alice, with an attendant rate $H(p)$ for her message, then Bob can achieve his channel capacity $C(p)$ as in (5) by punching at density $\widehat{q}(p)$ as in (4). In the process, Alice's original message is, of course, degraded. In fact, if Alice tries to read back her message, she will find it contaminated by the same amount of one-way noise as if it had gone through the channel described by exchanging q and p in Fig. 1 and table (2).

Suppose now that Alice, realizing that her tapes are going to be reused (or *concurrently* used—since, as we have seen, the two punching operations commute) by Bob, decides to encode her next batch of tapes so as to make her messages readable even after an anticipated punching by Bob at density q. According to (4) and Fig. 4, as a preventive measure she will have to shift her punch density p to a higher value than $1/2$, thus achieving a smaller rate but greater resistance to Bob's tampering. When Bob realizes that, he will be forced to shift *his* punch density q to a higher value—and so forth.

This is not a zero-sum game: as the arms race unfolds, each party will end up storing progressively less information on the tape. Will the race lead to the mutual destruction of information capacity? Fortunately, the curve of Fig. 4 intersects the line $\widehat{q} = p$ and has a slope there less than 1. Thus, the race converges to a stable point (with $q = p \approx 0.609$), where each party achieves an effective storage capacity of $\approx .240$ bit/cell.

The sum of the two capacities—and these are *coexisting* capacities, with both messages readable at the same time!—is about 0.48 bit/cell, to be compared with the 1 bit/cell Bob and Alice could have achieved by "space-sharing" the tape (e.g., one cell for Alice, one for Bob, and so forth). Thus, the attempt by the two parties to concurrently use the monotonic-write tape, performed in a selfish but rational way, results in an overall loss of storage capacity that is substantial but not crippling.

It must be noted that, even though at equilibrium they are in a symmetrical situation, Alice and Bob cannot use the *very same* block code to encode their messages on the tape. To avoid interference the two codes must be practically uncorrelated or "mutually orthogonal"; this is always possible with long enough block codes.

5 Intercom dialogues

We introduce the issue of *tape equivalence* by means of two dialogues. The **length** ℓ of a piece of tape is the number of cells it contains. Because of the canonical distribution of both holes and punches (§1), the overall capacity of a tape of length ℓ is ℓ times the capacity of a single cell, and similarly for the mutual information.

Dialogue 1

Bob is now an old and stingy facilities officer at Caltech. He can no longer see the individual holes on the tape—his vision is blurred—and he wouldn't any longer know how to start writing a block code. All he cares about is tape as a bulk commodity, and getting the most out of it. He is assisted by Sue, who physically handles the tape and knows how to devise appropriate block codes. Sue has standing instructions to recycle paper tape to the best of her capabilities and not to bother Bob with details.

BOB, *on the intercom*: Sue, we have to send a million-bit message to MIT. Get a piece of tape.

SUE, *from the mail room*: I've got here a reel of tape with a total capacity of one million bits. [*She doesn't tell Bob whether that's a thousand feet of virgin tape, or perhaps ten thousand feet of heavily used tape.*]

BOB: Good! Here is the message. Don't waste any capacity, and make sure you get the tape back from MIT so we can reuse it! By the way, what will be the capacity left on the tape after this message? I want to enter it as an asset in my inventory sheet.

SUE: That will be 333,000 bits.

BOB: So, using the tape at capacity will leave it "shrunk" to .333 of its previous capacity. Well, one third is better than nothing!

BOB, *a week later*: Sue, here is another message for MIT. Since it happens to be 333,000 bits long, let's use the tape you got back from them. [*We assume that, after decoding a message, MIT does not keep a record of the detailed hole pattern received. That might be used for improving transmission efficiency, but at substantial storage cost.*] What will be the capacity left on the tape after this message?

SUE: That will be about 119,000 bits.

BOB, *punching keys on a calculator*: Hey, this time, it will only shrink to 119,000/333,333=.357 of its pre-transmission capacity [*with an accusing tone*] Are you sure you made the best use of my tape last time?

SUE, *chuckling*: Cool off, Bob! Every housewife knows that used tape shrinks less! In fact, really ripe tape only shrinks to $1/e \approx .368$ of its previous capacity upon each usage.

BOB: And brand new tape?

SUE: That's the worst! It shrinks to $\ln 5/\ln 2 - 2 \approx .322$ of its previous capacity. Here's the whole picture! (Fig. 5)

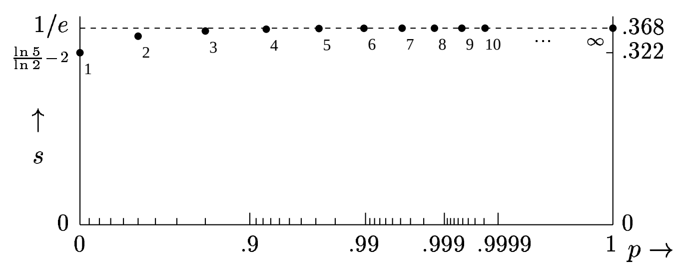

Figure 5: "Shrinkage coefficient" $s = C(\widehat{p'})/C(p)$ for successive full-capacity usages (labeled 1, 2, ..., ∞) of an initially virgin tape. As the tape gets more thoroughly used, the shrinkage coefficient rapidly converges to $1/e$.

The two physical parameters of a piece of tape, namely, its length ℓ (in cells) and its current hole density p, completely characterize its "response"—in terms of amount of information transmitted and capacity left—upon each successive usage, including usages with a punch density $q < \widehat{q}$ (where some of the capacity is saved for later) or $q > \widehat{q}$ (where some capacity is wasted), according to equations (1), (3), and (5).

In particular, the capacity of a piece of tape of length ℓ is $\ell C(p)$ (cf. (5)). If we measure capacity in bits (cf. §2), this can be thought of as the **reduced length** of the tape—i.e., the number of cells of virgin tape having the same overall capacity. Bob would have been delighted to find that two pieces of tape having the same reduced length are completely equivalent for information-transmission purposes.

Such an equivalence principle would allow him to characterize a piece of tape by means of a single information-theoretical parameter—the reduced length—rather than the two physical parameters ℓ and p, and greatly simplify his inventory bookkeeping.

If such an equivalence held, then, as a specific consequence, the shrinkage coefficient of Dialogue 1, defined as

$$s(p) = \frac{C(\widehat{p'})}{C(p)},$$

would be independent of p. Unfortunately, as we have seen in the dialogue, this is only approximately true (Fig. 5). We'll return to this problem, with better tools, in §7.

Dialogue 2

Sue is on vacation. Her temporary replacement, Willie, is being indoctrinated by Bob about the need to conserve tape. To test his coding capabilities, Bob chooses a spool of tape just like the one he gave Sue the first time.

BOB: Here is a length of used tape, Willie, and a million-bit message to be sent to MIT. Please transmit the message as efficiently as you can.

WILLIE: Is it urgent?

BOB: Not, really. Take your time, but do a good job!

BOB, *a month later*: Well, did you get the tape back from MIT?

WILLIE: Here it is!

BOB: What's its capacity now?

WILLIE: 580,000 bits, more or less.

BOB: What? It only shrank to .580 of its original length? How did you manage that?

WILLIE: You know, haste makes waste. So I first encoded only a small fraction of the message on the tape, using a very low punch density. MIT decoded that, wrote it down, and sent back the tape. Then I encoded on the same tape another increment of the message, sent it to MIT, and so on. The tape must have gone back and forth twenty times!

BOB: In the limit of an infinite number of infinitesimal increments, how much information could you transmit in this way?

WILLIE: Starting from virgin tape, about 2.37 bits/cell (precisely, $\frac{\pi^2}{\ln 2}$).

BOB: That's amazing!

WILLIE: And, of course, at any intermediate moment the transmission "mileage" already used plus that which is still left on the tape equals a constant—provided you always travel very slowly.

BOB: I got it! Your "mileage left" is the **effective length** I was looking for. No matter how different they look physically, two pieces of tape (say, one short and fresh and the other long and stale) having the same effective length are equivalent for information transmission purposes.

WILLIE: Slow down, Bob! That is true only as long as you use them up *slowly*. By comparing Sue's performance with mine, you realize that, when one tries to cram onto a tape a substantial fraction of its channel capacity at once, there are losses by "friction", as it were.[2] Well, one can tell the difference between fresh tape and well-worn tape by the fact that the former exhibits *just a little more friction* than the latter.

6 Weak equivalence

Let us explore in more detail what Bob discovered with Willie's help.

Suppose that we start with virgin tape and record on it a small amount dI of information by punching it at a very low density. We ship the tape but ask the recipient to send it back to us after transcribing the message. We then record on this "slightly used" tape an additional small amount of information,[3] further increasing the hole density. We continue in this way, sending one after the other a large number of messages each having a small information contents, until the tape is completely filled with holes. If at each stage the encoding is done optimally, what is the cumulative information $\int dI$ of the messages we sent?

Assume that at a generic stage of this process we start with a hole density p and increase it to $p' = p + dp$ by issuing punch commands with a probability dq per cell. The channel diagram is the same as Fig. 1, but with input and output probabilities as in Fig. 6.

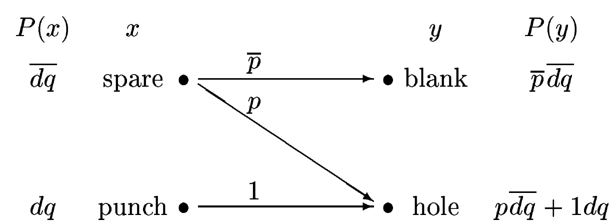

Figure 6: Channel diagram for used tape, with input and output probabilities corresponding to incremental use of the channel capacity.

The hole density increment is $dp = \overline{p}dq$, as the blanks, which appear with density $\overline{p}$, are turned into holes with probability dq, while the holes, with density p, remain unaffected. The mutual information of this infinitesimal punching operation, calculated from (3) using dq in place of q,

[2]This behavior is qualitatively similar to that of mechanical systems. Consider a battery of internal resistance R connected to a load of impedance r. It will be convenient to use the normalized variable $p = 1/(1 + r/R)$, which goes from 0 to 1 as r/R goes from ∞ to 0. The maximum *power transfer* occurs when $p = 1/2$ (i.e., $r = R$); in this case, half of the energy is dissipated by friction in R. As $p \to 0$, energy is transfered to the load more slowly but less of it is wasted by friction. As $p \to 1$, one gets less power out *and* wastes a greater fraction of the energy. Indeed, to an untrained eye the power transfer curve $2p(1-p)$—an inverted parabola—is hard to tell apart from the binary entropy $H(p)$.

[3]For this, we need, of course, a very long block code.

is

$$dI = H(p+dp) - \frac{\overline{p+dp}}{\overline{p}} H(p) \tag{6}$$

$$= \left(\overline{\ln}\,\frac{p}{\overline{p}} + \frac{H(p)}{\overline{p}}\right) dp = \frac{\overline{\ln}\,p}{\overline{p}} dp. \tag{7}$$

The indefinite integral of the integrand in the last expression is

$$\int \frac{\overline{\ln}\,p}{1-p} dp = -\mathrm{Li}_2(1-p),$$

where Li_2 is the *dilogarithm* function.[4] Thus, the **effective capacity** of a tape of hole density p, i.e., the total amount of information that can be transmitted via it in successive small increments until all holes have been punched up, is

$$Q(p) = \int_p^1 dI = -\mathrm{Li}_2(1-x)\Big|_p^1 = \mathrm{Li}_2(\overline{p}), \tag{8}$$

as plotted in Fig. 7 (compare with the qualitatively similar behavior of C, in Fig. 3); for virgin tape ($p=0$), the effective capacity is $\mathrm{Li}_2(1) = \pi^2/6$ (cf. [8]). Note that, by (8),

$$dI = -dQ.$$

Since Q is a *function of state* of the tape (i.e., it depends only on its state and not on the specific sequence of operation that led to that state), dI is an *exact* differential.

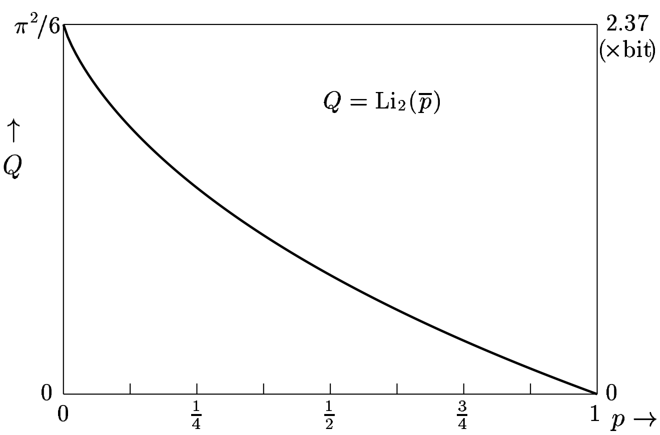

Figure 7: Effective capacity Q as a function of the current hole density p.

The above quantities are on a per-cell basis. Let us define the **effective length** (cf. Dialog 2) of a piece of tape of length ℓ and hole density p as $\lambda = \ell Q(p)$. If by a sequence of small incremental messages we transmit an amount of information I per cell, and thus a total amount $S = I\ell$ for the entire piece of tape, the new effective length will be $\lambda' = \ell(Q - I)$. The corresponding shrinkage coefficient[5] will be

$$\frac{\lambda'}{\lambda} = 1 - \frac{I}{Q} = 1 - \frac{S}{\lambda},$$

which is *independent* of the physical parameters ℓ and p and depends only on the ratio between two information-theoretical quantities, i.e., the total amount S of information transmitted and the effective length λ of the tape. We shall call this the **weak equivalence principle** for monotonic-write media.

7 Strong equivalence

Let us now explore in more detail what Bob discovered with Sue's help.

Whether we intend to utilize a piece of tape incrementally, as in Dialogue 2, or in discrete installements, as in Dialogue 1, the *effective length* λ defined above provides a more natural measure of a tape's information capacity than the *reduced length* introduced in Dialogue 1.

Armed with this measure, let us now turn our attention from the special case of the limit of an infinite sequence of infinitesimal messages to the general case of *discrete* messages, where the weak principle is not applicable.

Our goal is to eliminate the physical parameters p and q between equations (3) and (8), and thus write a relation directly between (a) the effective length λ of a piece of tape before the transmission of a message, (b) the effective length λ' after the transmission, and (c) the amount S of information conveyed by the message. If such a relation exists, it may be assumed to be of the form

$$f(\lambda, \lambda', S) = 0$$

and, since we are assuming a canonical hole distribution before and after punching, it must satisfy the scaling property

$$f(a\lambda, a\lambda', aS) = 0 \quad \text{for any } a.$$

Setting $a = 1/\lambda'$, we obtain a relation between two variables

$$g(\sigma, \mu) = f(\sigma, 1, \mu) = 0,$$

where

$$\sigma = \frac{\lambda'}{\lambda} = \frac{Q(p')}{Q(p)} \quad \text{and} \quad \mu = \frac{S}{\lambda} = \frac{I(p,q)}{Q(p)}.$$

The variable μ—which is the mutual information for a given stage of utilization of the tape—can be thought of as the information rate per unit of effective length of the tape, and σ as the shrinkage coefficient attendant to that stage.

Since the variables σ and μ depend on two parameters, p and q, we cannot *a priori* expect to eliminate *both* parameters when solving for μ with respect to σ. However, for a given initial hole density p treated as a fixed parameter, we can eliminate just q and write

$$\mu = \mu_p(\sigma).$$

The result of this elimination, performed numerically for different values of p, are shown in Fig. 8, which also shows the values of the eliminated parameter q on the $\mu(\sigma)$ curves.

Paralleling the weak equivalence principle of the previous section—which states that tapes having the same effective

[4] This is one of the *polylogarithm* functions, defined by $\mathrm{Li}_n(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^n}$.

[5] This quantity is analogous to but distinct from the shrinkage coefficient of Dialogue 1, which is a ratio of channel capacities.

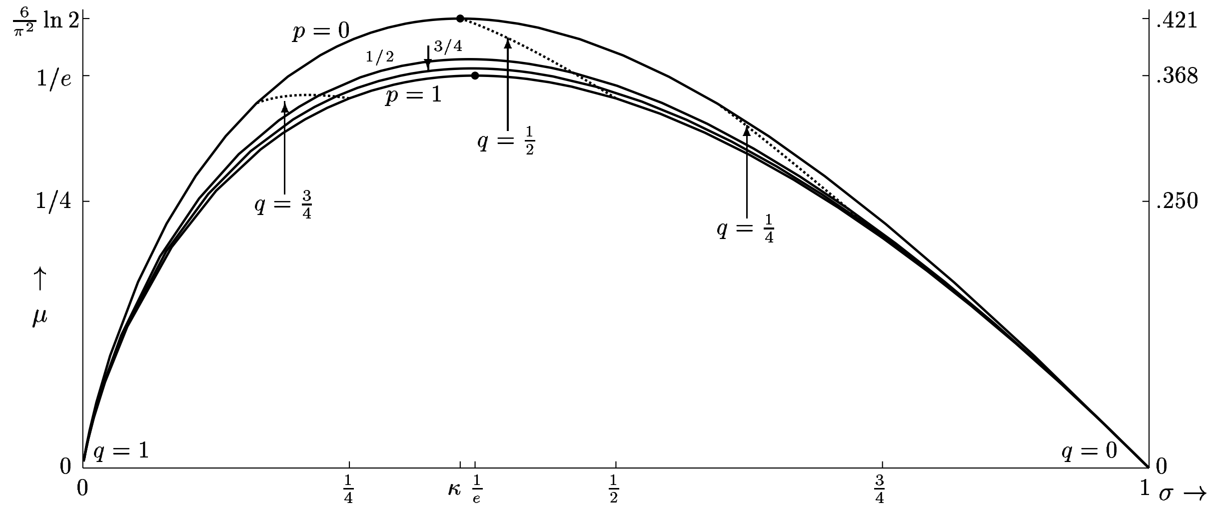

Figure 8: Mutual information per unit of effective length, μ, as a function of the shrinkage coefficient σ, for different values of the initial hole density p. The dotted lines represent loci of equal values for the running parameter q. For any value of p, the maximum of μ represents the corresponding channel capacity given in units of effective length. The maximum of μ_0 occurs at $\kappa = \frac{1}{2}(1 - \frac{6}{\pi^2}\ln^2 2)$. For utilization rate below capacity, the same mutual information can be obtained with two different values of shrinkage, one corresponding to rational usage of the tape and the other to needlessly wasteful usage.

capacity are indistinguishable at slow utilization rates—a **strong equivalence principle** would be one that is valid for any rate of utilization of the tape at any transmission stage, from an infinitesimal hole-density increment (q close to 0) to gross overpunching (q close to 1). I don't know whether it is more surprising that, strictly speaking, punched tape does *not* obey a strong equivalence principle, or that, after all, it turns out to do so *to a very good approximation*. In fact, as is clear from Fig. 8, after eliminating q between μ and σ some dependence on p remains, but this dependence is slight in any case and rapidly vanishes as p approaches 1. Intuitively, when expressed in terms of a more natural set of variables, the one-parameter family of curves of Fig. 2 nearly collapses onto a *single curve* (Fig. 8).

The curves $\mu_p(\sigma)$ all have slope -1 at $\sigma = 1$; this is an expression of the weak equivalence principle (i.e., for small q, the effective length decreases by an amount equal to the amount of information transmitted). They all have slope ∞ at $\sigma = 0$, signifying that the waste of effective capacity increases precipitously when one punches at a density much greater than that needed for transmitting at channel capacity.

The worst-case departure of the μ_p curves from the limiting curve $\mu_1 = \lim_{p\to 1}\mu_p$ occurs near the maximum bulge of the curves, and is substantially the same as the departure of s from its $1/e$ limit as plotted in Fig. 5. The curves μ_p are not likely to be expressible in closed form; however, as is easy to prove, the limiting curve μ_1 is nothing but the familiar *self-information* function $\mu = \sigma\,\overline{\ln}\,\sigma$. To the same approximation as the strong equivalence principle holds, *this function gives the information-transfer characteristics of punched tape (i.e., for any message, the capacity used by it, that wasted, and that left after the message) over the tape's entire utilization range.*

Let us remark that the self-information function appears in the limit also in Fig. 2. In fact, one can show that

$$\lim_{p\to 1}\frac{I(p,q)}{C(p)} = e\overline{q}\,\overline{\ln}\,\overline{q}.$$

8 Conclusions

A piece of randomly punched tape is described by two physical parameters—its length ℓ and its hole density p. We have raised the question of whether the tape's behavior as an information transmission commodity can be usefully characterized by a single information-theoretical parameter—its *effective length* λ. We have concluded that this is the case

- in the "quasi-static" limit of slow utilization rate (*weak equivalence principle*);
- for any utilization rate (*strong equivalence principle*)
  - exactly, in the limit of already heavily used tape, and
  - approximately—but with good accuracy even in the worst case—over the whole range of previous and future uses of the tape.

9 Acknowledgements

This research was funded in part by NSF (9305227-DMS) and in part by ARPA through the Ultra Program (ONR N00014-93-1-0660) and the CAM-8 project (ONR N00014-94-1-0662). I am indebted to Peter Elias, Matteo Frigo, and Mark Smith for useful discussions, and to Ronald Rivest for some references. I found Wolfram Research's *Mathematica* invaluable in this exploration.

10 References

[1] Abramson, Norman, *Information Theory and Coding*, McGraw-Hill (1963).

[2] Costa, Max, "Writing on dirty paper", *IEEE Trans. Info. Theory* **IT-29** (1983), 439–441.

[3] Dolev, Danny, David Maier, Harry Marson, and Jeffrey Ullman, "Correcting faults in write-once memory", *Proc. 16th Anuual ACM Symp. on Theory of Computing*, ACM (1984), 225–229.

[4] Dyson, Freeman, "Time without end: Physics and biology in an open universe", *Rev. Mod. Phys.* **51** (1979), 447–460.

[5] Heegard, Chris, and Abbas El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Info. Theory* **IT-29** (1983), 731–739.

[6] Maier, David, "Using write-once memory for database storage", *1982 ACM Symposium on Principles of Database Systems*, ACM (1982), 239–246.

[7] Rivest, Ronald, and Adi Shamir, "How to reuse a 'Write-Once' memory", *Information and Control* **55** (1982), 1–19.

[8] Wolf, Jack, Aaron Wyner, Jacob Ziv, and János Körner, "Coding for a Write-Once Memory", *AT&T Bell Lab. Tech. J.* **63** (1984), 1089–1112.