# Synchronous coherent extraction of heat

*V Guruprasad*
*syncool@worldnet.att.net*

## ABSTRACT

Described are embedded techniques for cooling clocked CMOS circuits by converting the primary phonons excited by the switching currents to electricity. The intent is to extract the energy as coherently as possible before it disperses into the bulk lattice and becomes heat at a lower average temperature. The first method is a heat engine using the bulk lattice as its heat sink and depends on the higher effective temperatures of the primary phonons before their dispersion. The second method uses stimulated emission for cooling, and a close relation is shown to exist between lasers and heat engines. Since coherent removal means performance of work, clock-switched coherent loads may be used to replace ohmic resistance in logic design. This not only overcomes a fundamental difficulty in the design of superconducting logic, but suggests eventual replacement of silicon with copper.

## 1. SUMMARY

It is well known that most of the dissipation in CMOS logic occurs at gate transitions. When a CMOS logic gate switches states, a tiny amount of charge moves through a potential difference and dissipates energy into the lattice, which eventually shows up as heat. Classically, the energy loss manifests initially as local heating within the channel, as the "closest atoms" gain vibrational energy from the colliding charge carriers (Fig. 1). In the more accurate quantum picture, each carrier initially loses its energy to only a few phonon modes that it interacts with; the phonons soon disperse into other lattice modes, eventually becoming heat that can be removed only by conduction to the surface. Since the primary phonons are at least momentarily localised, their energy is relatively coherent and can be removed coherently.

Two such techniques are described, viz. distributed electromagnetic heat engines and sublasing. The engines permit true thermodynamic measurement, for the first time, of transient local temperatures at virtually any speed, and are in a sense d.c. cousins of the laser. The detailed balance between magnetisation states and the engine circuit is particularly relevant to our application domain. Both techniques in effect *simulate* electrical resistance using coherent loads, and can therefore replace ohmic resistance for representing data states in switching logic. The implications are discussed in terms of thermodynamics and superconducting logic.

## 2. APPROACH

There are three aspects of the primary phonon coherence to consider. First, ignoring clock skew for the moment, we note that CMOS gates are designed to transit only at the precisely determined periodic clock transitions. This means the carrier losses have sharp temporal coincidence across the chip, defined by the clock rise and fall times. Second, thermal diffusion is a relatively slow process, taking at least 200 ps to spread beyond 1 μm, the average gate spacing at a density of 100 M gates per cm$^2$. A detection mechanism faster than the clock transitions would see a spatial localisation of the energy at least for times of this order. Third, even the relaxation of the primary phonons within the channel is a finite time process. The low order modes of a rigid periodic lattice are necessarily discrete, and it takes some time for energy concentrated in one mode to get distributed to the others. The local relaxation represents localisation in the momentum (**k**) space. At very low temperatures, the phonon modes are only weakly coupled, resulting in fairly long momentum coherence times. At ordinary temperatures, however, there is sufficient coupling between the lattice modes to reduce the thermal relaxation to the same order as the propagation time for sound.

The periodicity of the clock means that we could apply a heat engine method if it were operable at that speed, ie. if a complete thermodynamic cycle could be executed within one clock period. In fact, we should consider each phase of a clock signal as a period for the thermodynamic period, because transitions can occur in different gates at either edge of the clock. The spatial coherence means of the primary phonons means that the engine cycle must be executed separately by every transiting gate, even though they would be in step because of the common clock. This means the embedded engines are to be operated in parallel using a single macroscopic "piston" in SIMD[1] fashion; we may set the "expansion stroke" to be slightly longer than the clock rise and fall times, in order to "catch" transitions from the entire chip. The "farming out" of the "thermodynamic instruction stream" should be done intelligently in order to handle the clock skew that occurs in many complex chips; this merely entails skewing the "piston" delivery to follow the same path as the clock.

An inherent feature of this approach is that it tends to remove heat preferentially from gates with higher dissipation. Hitherto, the only way to prevent the peak transient temperatures from damaging the gate construction has been to limit the operating voltage and currents. Despite other limitations such as insulation breakdown and material creep, the selective cooling capability means that chips can be operated at higher voltages and currents without a proportionate increase in the degradation. The chips can be alternatively run at higher speeds without shortening the useful life, and the overall utilisation of the gates can be improved. To evaluate the possible impact, consider that at 100 M/cm$^2$ gate densities (corresponding to 1 μm average gate spacing) and 30 W/cm$^2$ dissipation at 100 MHz, we must deal with upto 1 pJ per gate transition, assuming about a third of the gates to transit on the average. Given the heat capacity of silicon to be about 1.57 J/cm$^3$-°K[2], this would cause a 1.6 °K rise in a 1 μm$^3$ channel, and a 103 °K rise in a 0.25 μm$^3$ channel. With the substrate at 30 °C, the maximum efficiency of the electromagnetic engines would be 6% at 50 °C peak transient temperature, 10% at 64 °C, and 23% at 125 °C at which the gates begin to degrade. Such figures would be quite favourable, especially in relation to reversible logic and charge recovery schemes [1] [2], which may require 10 times as many transistors per gate to implement. This decimates the logic density and also raises the per-gate dissipation unless the speed is considerably reduced. The present approach does not significantly affect the density or design of the gates, and may even allow higher operating speeds than currently possible.

The approach may have difficulties lurking elsewhere in matters of realisability, for which no claims are made at present. The proposed methods are possibilities arising from purely theoretical developments in heat engines and thermodynamics, and the perceived implications and advantages should hopefully motivate further investigation. The methods depend on physical properties that do not seem powerful on the macroscopic scale, but that should not be a concern. Gas engines obviously have the greatest power densities on the macroscopic scale, but no gas engine can hope to compete with thermal diffusion and relaxation, the very processes responsible for developing the mechanical pressure. A comparable power density is however necessary on the microscopic scale. A typical microprocessor dissipates about 10 W per 1 cm$^2$ of chip surface while operating at about 100 MHz, and has an active thickness of the order of 100 μm, yielding a heat production density of 10 MW/litre. To compare, the power density of a modern automobile engine is about 100 KW/litre of cylinder capacity at 50 Hz (3000 rpm), so our heat source is indeed of a comparable order of magnitude in power density, even though it only amounts to about 1 pJ/gate-event.

We also cannot rely on past studies to determine the feasibility of the approach. Some of these [4] involved a design error costing about one order of inefficiency. The mistake, that of using a permanent magnet, dates back to Edison, whose "thermomagnetic generator" [5] (Fig. 2) was designed to produce only an emf, not power. In a heat engine, the load must compress the medium repeatedly in order to obtain a power gain on the subsequent expansions, the gain resulting solely due to the increased force during the expansion. Edison does not provide for the compression, and even suggests using a commutator, which would have further prevented it. The distinction appears to have generally escaped the physicists' attention, although the merits of varying the field have been noticed recently [6]. Omitting the permanent bias altogether would make the device a true inductive heat engine [7].

---

[1] Single Instruction Multiple Data (streams).

[2] Density = 2.329 gm/cm$^3$ at 25 °C, specific heat = 0.162 cal/gm-°K.

Later developments overcome some other problems in the past studies. Strong fields from superconducting magnets, in excess of 10-15 T, have been shown to yield efficiencies sufficient for power generation in magnetocaloric engines [8]. This translates to 1-3 mT at our operating frequencies, and is still somewhat high. Another factor was the use of sinusoidal load currents. The optimal waveform for executing a magnetic Carnot cycle should be triangular, not sinusoidal [7], and with the correct choice of waveform, the engine cycle can even be modified to approximate to the Carnot form [9]. The latter method is largely constrained by the thermal isolation available for the adiabatic portions of the cycle, but the constraint is not relevant in our application because we are competing with thermal diffusion and relaxation anyway, hence near Carnot efficiencies may be taken for granted.

For the same reason, as already discussed, we are also not constrained by the thermal mass of bulk magnetic or dielectric media, and the application itself overcomes another fundamental difficulty in the past, viz. long diffusion times and the associated temperature drop and cumulative leakage losses. Our engines would operate at such speeds only because the heat is generated and dispersed *in situ* without transportation delay or loss. In any case, we are ideally concerned only with the individual primary phonon modes at near-total coherence, and consequently, at very high effective temperatures, sufficient to cause infrared emissions usable for diagnostic purposes [10].

## 3. ELECTROMAGNETIC HEAT ENGINES

As observed with respect to Edison's design, the principle of a heat engine requires a circulation of energy between the load and the medium, with the net conversion manifesting as an excess of the energy output over that absorbed. There is a corresponding circulation of heat into and out of the medium, with an input excess accounting for the net conversion. A similar circulation also occurs at the primary of an electrical transformer, and the sole difference from a heat engine is that in each cycle, the returned energy, unlike heat, remains usable for the next cycle, and the net conversion represents the power factor in place of the efficiency. These difference fully and correctly account for the special nature of heat and the second law [9]. On the load side, however, a heat engine is no different from a transformer, and conversely, non-thermal Carnot cycles are also easily defined.

The fundamental requirement for an *electrically-operable heat engine* is therefore that its load be electrical and it perform cyclic work on the thermodynamic medium. Since the medium should also absorb and emit work cyclically, the engine itself should be essentially reactive. Naturally, an electrical heat engine, in the above sense, is also the valid circuit model for mechanical heat engines, and a *canonical electrical representation* of a heat engine must involve the fewest possible linear electrical components. We thus arrive at inductive and capacitive devices using temperature-sensitive media for the canonical representations, as shown in Figs. 3 and 4, respectively, and the inductive form indeed resembles a transformer. The *equation of state* for the inductive form is simply

$$i = \frac{N\Phi}{L(T)}, \qquad (1)$$

where $i$ is the coil current acting as the applied force or pressure, $N$ is the number of turns, $\Phi$ is the magnetisation flux excluding contribution from the applied field $H$, representing the "piston displacement", and $L$ is the inductance that varies with the temperature $T$. The adiabatic equation follows from thermodynamic first principles [9]:

$$i = \left[ \frac{\dfrac{c}{L_T}}{1 - \dfrac{\Phi}{\Phi_0}\left(1 - \dfrac{c}{L_T i_0^2}\right)} \right]^{1/2} \qquad (2)$$

where $c$ is the applicable heat capacity of the medium[3], $L$ is the engine inductance, whose temperature dependence is represented by $L_T \equiv dL/dT$. The latter must be computed from the field equation of state in terms of **H**, **M** and $T$, which can be analytically derived from the Curie and Curie-Weiss laws [8] [7] or obtained empirically. The corresponding force and displacement variables for the capacitive model are the voltage $V$ and capacitor charge $Q$, and lead to a similar adiabatic equation.

Fig. 5 illustrates how the inductive form can be used to model a typical mechanical engine. Notice that the thermodynamic component merely amplifies power and cannot develop a motive force on its own. Traditionally, one used flywheel inertia, represented by a capacitor in the figure, to sustain the operation, by using the expansion stroke of one engine cycle to drive the compression for the next, and also to smoothen the output power. The flywheel is really extraneous to the

---

[3] $c_P$ at very low frequencies, $c_V$ otherwise.

thermodynamic conversion, however, and can be eliminated by using an auxiliary driving source as shown. The correct load circuit characterisation of the engine is therefore as a *mean negative resistance* $\overline{R_t}$, producing a current, and power, amplification β given by

$$\beta = \frac{1}{1-\alpha}, \ \alpha = 1 - \frac{1}{\beta} \qquad (3)$$

where α represents the normalised thermodynamic resistance $\overline{R_t}/R_L$, $R_L$ being the load resistance. We must redefine α as the normalised conductance to arrive at the same gain factor β; these factors are particularly useful for controlling the operating point and regulation of engines operated this way [7].

The canonical forms are clearly lumped representations of devices that we may need for our application. Notice, incidentally, that a transmission line heat engine can utilise both magnetic and dielectric properties simultaneously for the power conversion. When the magnetic and electric component waves are in phase, as in a travelling wave, the corresponding thermodynamic cycles are also in step and contribute independently to the travelling wave. If the components go out of phase, the corresponding thermodynamic cycles push and pull each other, amplifying the stationary wave. The thermodynamic conversion is thus available to all kinds of electromagnetic wave energy, and we may use routine electronic design principles to implement the transmission line engines.

We still need to understand the microscopic interactions involved, firstly, because they serve to check the consistency of our macroscopic theories. In the kinetic theory of gases, for instance, one applies the ordinary (macroscopic) dynamical model of rigid bodies to the individual molecules, and derives the gross thermodynamic behaviour by statistical averaging. We need an analogous *kinetic theory of electromagnetism* using atomic polarisations and magnetic moments as the "particle" model. Although such models do exist in the literature, they do not adequately address induction and the transmission of power.

Secondly, details of the microscopic interactions would help in our understanding of electrical resistance and dissipation. Existing theoretical models, such as the Bloch-Sommerfeld theory, generally explain the initial transfer of energy to the lattice phonons, but the mechanisms responsible for making the transfer irreversible remain unclear. The extreme case of reversibility is evidently the domain of the BCS theory of superconductivity, but the latter is not sufficiently intuitive for our application. These shortcomings of prior theory reflect the fact that induction has received only secondary attention in the quantum formulations.

## 4. KINETIC THEORY

Shown in [7] is that the mean engine resistance $\overline{R_t}$ is the time-averaged result of *positive* instantaneous resistance during the magnetisation (compression) stroke, representing absorption of energy from the coil, and *negative* instantaneous resistance during the demagnetisation (expansion). Without the temperature variation, the engine reduces to a simple inductor. The sign of $\overline{R_t}$ depends on the relative phase between the temperature and the load current cycles: if the demagnetisations occur at the higher temperature, the net conversion is from heat to electricity ($\overline{R_t} < 0$), and if they occur at the lower temperature, the result is refrigeration ($\overline{R_t} > 0$). $\overline{R_t}$ is thus the d.c. value of the time-varying $R_t$. $R_t$ represents, of course, the instantaneous induced emf

$$\xi_t = R_t i, \qquad (4)$$

$i$ being the instantaneous engine current, and $\xi_t$ is in turn the statistical sum of the inductions from the innumerable individual dipoles being flipped over by the thermal activity.

It is helpful to develop an intuitive picture of the dipole-coil and dipole-dipole interactions involved. In the first case, Lenz's law dictates that any perturbation of the dipole would induce an emf in the coil, such that the incremental current would tend to maintain the net magnetic moment (Fig. 6). If the dipole be initially aligned along the magnetic moment of the coil, the incremental reaction increases the coil moment, and if the dipole were aligned opposite to the coil moment, the reaction would reduce the coil current. The change in the energy associated with the magnetic moment of the coil is consistent with the change in the total energy of the dipole with respect to the coil field[4]. The coil current also continuously exerts a torque on the dipole, and the incremental reaction changes the torque, again in a manner tending to restore the dipole to its initial orientation. The disturbance thus leads to a restoring force approximately proportional to the angular displacement (at least for small angles), and the result is

---

[4] Recall that the total magnetisation energy is the *negative* of the potential energy.

a harmonic oscillation that continuously exchanges energy back and forth between the coil and the dipole.

The same exchange oscillation is expected in a dipole-dipole interaction, but the nature of reaction is somewhat different. When one dipole is disturbed from its orientation, both the magnitude and direction of its field change at the location of the other dipole. The latter's response to the magnitude change is a slight expansion or contraction of the orbital "currents", thus changing its magnetic moment. The change in direction also causes a change in torque on the second dipole in addition to its diamagnetic reaction. Both constitute work done on the second dipole by the first, and by Lenz's law, the effects again tend to return the energy and to restore the first dipole to its original orientation. In a system involving large number of dipoles, every dipole influences every other dipole and pairwise oscillations become difficult to isolate. However, but for the occasional randomisation due to thermal activity, the magnetic system remains generally locked in such oscillatory modes, which can be identified with the so-called exchange coupling in quantum theory[5].

The thermal equilibrium is clearly more complex than that of the ideal gas, but the model is in fact sufficient to explain all effects thermal and electrical. We are particularly interested in the equilibrium against coherent and incoherent loads, and in the inductive transfer of energy during magnetisation and demagnetisation. Our system cannot lose energy to a coherent load, such as a motor, because doing work on such a load requires a macroscopically directed force, whereas the forces within the dipole system are strictly oscillatory and closed. Although the system would, and does, perform microscopic work on an incoherent (ohmic) load, the latter also performs similar work on the dipoles in return, equilibrium being reached when the work transfer rates balance. This means the coil also acts as an electrical conduit for heat between the magnetic medium and an ohmic load[6].

Finally, it would be recalled that whenever the coil current rises or falls, macroscopic work is done on or by the coil, respectively. When the current drops, the work done by dipoles on the coil is no longer returned in its entirety, but passed on to the load, leaving the dipoles in a new state of magnetisation with lower energy. Conversely, work must be done on the dipoles during magnetisation, but the dipoles will tend to return the energy immediately, unless thermal activity can carry the energy off and leave the dipole in a new state of mean orientation. In both cases, there are net energy transfers that become observable as electrical pulses in the coil[7].

The steady state oscillations are not directly observable because an observation implies change in the physical state of the observer, representing the incremental knowledge from the observed data, which usually involves exchanging energy with the observed system. The thermal (shot) noise observed in electrical circuits is really due to ohmic resistance that causes random quantum transitions between the stationary states. Large coherent pulses with strongly thermal character have indeed been observed at least in an unusual class of motors involving gigantic air-core coils (100 H - 20 KH) and rotating magnets with large dipole moments [12].

The most basic application of kinetic theory is, of course, in relating the pressure and temperature of a gas. Recall that the pressure is derived as the mean rate of impulses delivered by the molecules on a confining wall, leading to the ideal gas equation of state (Charles' law):

$$p = \frac{rT}{V}. \tag{5}$$

The comparable scenario in magnetism is the Curie law of paramagnetism expressible as

$$H = kTM, \tag{6}$$

where $k = k_B / n\mu^2$, $n$ being the number density of the dipoles and $\mu$ the individual dipole moment. Although we usually think of the inductive reaction as an emf $-Li$, the current is in fact equally representative; the $H$ in eqn. 6 represents the disorienting "pressure" on the coil due to the continuous thermal turnover of the dipoles. The equation differs from the gas law in that $M^{-1}$, rather than $M$, plays the role of the displaced volume, and the incremental work corresponding to $pdV$ is $-\mu_0 HdM$, and is negative. These differences make the energy transfers in a mechanical magnetic engine more difficult to follow [7].

The emf is directly associable with mechanical forces only in electrostatic interactions, as in the capacitive engine. A similar kinetic theory is involved, based on the electrical moments of the molecules, but the forces are considerably simpler to understand. Incidentally,

---

[5] This is simply a classical rendering of the theory.

[6] The thermal conductivity due to the electron gas.

[7] Known as the Barkhausen effect.

although $L(T)$ could be derived from a field description like eqn. 6, it is more appropriate to obtain it empirically because magnetic materials deviate considerably from simple paramagnetism. In our case, the embedded dipoles unlikely to be strongly coupled, but even so, the Curie law only represents equilibrium thermodynamics, and in order to compete with thermal relaxation, we must prepare to deal directly with the quantum interactions and probabilities.

## 5. SYNCHRONOUS AND COHERENT COOLING

We construct a conceptual inductive engine using a coil and magnetic dipoles embedded within (or very near) the CMOS channel (Fig. 1). The coil is driven synchronously with the chip clock, so that the current is always at its maximum a little before the clock transitions, and falls during the transitions. The dipoles are thus magnetised and ready to demagnetise when a clock edge arrives. In a gate that does not change state, the embedded dipoles will eventually demagnetise due to thermal relaxation, returning the magnetisation energy without contributing a gain or loss to the coil power. In a gate that does change state, the phonons excited by the transition current directly or indirectly perturb the dipoles independently of the thermal relaxation, thus performing work on the coil.

Note that even in the magnetised state, a fraction of the dipoles are aligned opposite to the applied field. Phonons, both thermal and "ohmic", tend to overturn the dipoles irrespective of their orientation, but because the aligned dipoles are in excess, the overall result is a general shift toward demagnetisation, with corresponding output of energy to the coil. Fig. 7 illustrates the typical cycle in the $M-H$ space.

The Feynman diagram for the overall system is shown in Fig. 8. As per the Bloch-Sommerfeld model, the moving charges (e) pick up energy (A) from the circuit power supply, and lose that energy to lattice phonons (B). These *hot primary phonons* (p*) disperse into other phonon modes (p’, p"), turning into heat (C). When these *warm secondary phonons* strike the dipoles, they induce energy into the coil (D), and thus get coupled into the electrical load (E), where their energy is eventually dissipated (F). If the coil were not energised as described, the dipole transitions would show up as electrical noise. Premagnetising the dipoles just before the clock transition ensures that a majority of the induced impulses come from demagnetisation, and amplify the falling coil current. We thus "partially rectify" the electrical noise, and since the noise is due to an elevated temperature in the channel region, the rectification yields coherent power taken out by the coil current.

There is no thermodynamic reason to wait for the warm phonons. We should aim to intercept the hot primary phonons, which are relatively coherent, consistent with their high effective temperature, and should arrive at a process looking more like Fig. 9, in which the primary phonon (p*) transfers the carrier energy coherently to the coil. The coherence means that the force is carried to the electrical load almost intact. Notice that the dipole itself sheds energy of the order of $2\mu H$ in the process, but this comes from the dipole itself. The work transferred from the hot phonon to the load must be additional to this, and being almost coherent, would be of the order of the carrier loss, corresponding to the high effective temperature.

The phonon dispersion mechanisms are yet to be understood in sufficient detail for us to meaningfully quantify the probabilities and coupling factors. We do know, however, that there are phonons at least hot enough to cause observable infrared emissions [10] corresponding to frequencies in the order of $10^{13}$ Hz, representing a sonic wavelength of only 0.5 nm (5 Å), which is of the order of the lattice constant. Assuming their useful coherence length to be of the same order, this would suggest locating a dipole in every lattice cell, which would cause sufficient exchange couplings to turn the semiconductor into a ferromagnet. On the other hand, we could resign ourselves to catching only cooler phonons with longer wavelengths at a slight distance. Since the infrared emissions are relatively rare, most of the phonons are indeed be cooler and easier to catch, although this also means a lower efficiency of conversion, due to the fact that the warm phonons would exert less force on the dipole-coil load. There is a third possibility that addresses both these difficulties of infrequency of hot phonons and the inefficiency of warm phonon conversion, that is described in the remaining sections.

## 6. COOLING BY SUBLASING EMISSION

The infrared emissions suggest yet another approach for the coherent extraction of the primary phonon energy. Instead of magnetic dipoles, we could impregnate the channel regions with atoms of a suitable lasing material which would be easily excited by the primary phonons, and subsequently discharge these atom by stimulated emission using an infrared illumination. The unexcited atoms simply do not contribute, because they would absorb and emit with equal probability, but those already excited by the gate transitions can but discharge. Even the smallest probability of stimulated emission thus becomes fully productive in terms of removing energy that would otherwise become heat.

The density of implanted lasing atoms is again a limiting issue. An additional difficulty is, of course, the inclusion of an infrared source and the optical paths for the radiation. However, the method has the inherent advantage that it does not require special circuits within the chip and could work with continuous illumination. The relative coherence of the hot phonons between the carriers and the implants means that the energy and momentum must be carried almost unchanged. This means the carrier-phonon and phonon-implant coupling conditions must be closely matched, implying that the implants could also increase the hot phonon production and radiational discharge. Further implications of this are described in the next section.

The quantum picture also reveals a close relationship between heat engines and lasers. Fig. 10 shows the *total* magnetic energy distribution of paramagnetic dipoles with respect to an applied field. Note that a quantum mechanical measurement always yields a discrete answer with respect to the question, which means that in a measurement of the $Z$ magnetisation, the dipoles will always be found pointing in the $+Z$ or $-Z$ direction (up or down, respectively, in the figure), irrespective of the $H$ field. When $H$ is zero, the dipoles are aligned up or down with equal probability and with no difference in their magnetic energy. When a non-zero $H$ field is applied along $+Z$ direction, a majority of the dipoles not only become aligned with $H$, but also gain energy from the field, moving up by $\mu H$[8], while the remaining dipoles lose energy and move down by $\mu H$. The total magnetisation energy of course corresponds to the difference of the populations. This is the equilibrium picture and is precisely what happens during the premagnetisation process. Just around the clock transition, however, the $H$ field begins to drop, but the two sets of dipoles will not move right away toward the demagnetisation line corresponding to $H = 0$: each dipole remains locked in its $\pm\mu H$ state until released by a phonon interaction. Between the dropping of the $H$ field and the phonon event, the dipoles are in an excited state.

The graph clearly resembles population inversion in a laser, and we can draw an interesting comparison as follows. Unlike ordinary lasers, our sublasing uses hot phonons for pumping, while the engine needs premagnetisation. Spontaneous discharge is not harmful in the engine and is actually productive in sublasing. Conversely, the stimulation for d.c. emission comes from the warm or hot phonons in the engine case.

## 7. SUPERCONDUCTING LOGIC

The fact that the hot phonons are relatively coherent suggests that we could eventually do away with the phonon involvement altogether. However, the ohmic resistance plays an important role in current technology: some form of resistive loss is actually necessary for chip or system level design.

Historically, this issue has been considerably mixed up with notions of logical and thermodynamic reversibility [13], but the prior theory does not distinguish between ohmic and coherent resistance. Recall that the consumption of a given quantity $q$ of heat from an elevated temperature $T$ must contribute a total entropy of $q(T_A - T)^{-1}$, $T_A$ being the ambient temperature, to the environment. This entropy must be produced irrespective of whether the heat were allowed to simply dissipate into the environment, or was used to drive a heat engine. In the latter case, the engine may produce very little entropy, ideally wasting no more than $q^* = qT_A/T$, and the balance of entropy can be accounted for only by the consumption of the output work $w = q - q^*$ leading to its eventual dissipation at $T_A$ (Fig. 11). *The required* $kT\ln 2$ *dissipation per bit of erasure can therefore be met by removing the same amount of energy coherently from the electrical circuit.*

This subtlety of classical thermodynamics has been entirely missed in the chain of reasoning leading to Bennett's proof of the logical reversibility of computation [14] and the thermodynamically reversible models of computation [15]. A computer is essentially a device capable of a finite set of I/O values, having a finite internal memory, and designed to execute sequences of states representing different combinations of its memory and output data values, when driven by a discrete clock signal and a power supply. We begin by observing that a cyclic sequence can be implemented with at most frictional losses at any temperature, ie. without a theoretical minimum dissipation, and that this applies to any finite sequence of such cycles, because the combination would again be periodic. However,

---

[8] Taking $H$ in Telsa to avoid writing $\mu_0$'s.

such a device would be merely a counter, and any such sequence is merely a digital form of a stationary state. *A general computer must be capable of executing not only any of a set of predetermined sequences, but must be switchable between disjoint sequences in response to successive input data.*

In a typical processor chip, the predetermined sequences correspond to its instruction set, while the input is the program. To be useful, a computation must also produce an observable change in its output, but a repetitive output sequence would be predictable and convey no information. In a fundamental sense, *a computer by itself cannot produce information and can only transport information from input to output.* This is because the very notion of information requires the involvement of an input: consider that an input-less device must be either predictable, in which case we do not call its output information, or unpredictable, which we then call noise! The standpoint implies that *the essential dissipation of computation is simply the physical work indirectly performed by the input data source on the final destination.* The ordinary separation of input lines from the power supply indeed reflects the fact that every logic gate is designed with a built-in capability for amplifying its input signal, and an ordinary logic circuit performs amplification many times over before eventually driving its output load. Note also that true noise components in the output do nothing to affect the lower bound on entropy production. These observations also hold for information in the thermodynamic sense, and constitute a statement of conformance to the second law.

Superconductive logic has thus had a more fundamental problem inhibiting its deployment than refrigeration, cost or manufacture. The complete absense of resistance makes it impossible to switch between sequences without doing work on a *coherent* load. The associated difficulty for logic design lies in not being able to involve switching functionality in the intermediate stages of processing. Not surprisingly, most practical applications of superconductive logic have indeed been devices like ring counters and shift registers.

Our coherent cooling techniques, in effect, present embedded charge carriers with a coherent load to lose their energy to, and can thus help distribute switching functionality within superconducting logic. One way to do this is to design logic gates which can switch only when an applied field is relaxed, thus allowing the circuit to do work on the withdrawing field. Magnetic engines are most efficient at lower temperatures anyway and might be well suited for such an application. This principle of "clocked coherent resistance" is not dependent on superconductivity, and could be used at ordinary temperatures to reduce the proportion of ohmic resistances in semiconductor logic, and, in the limit, possibly replace silicon with copper altogether.

## 8. CONCLUSION

Although the above methods are at best conjectural, they introduce significant opportunity for theoretical and technological development. They describe an important application domain for distributed realisations of electrical heat engines and lasers in step with the frontline of electronics technology. They illustrate how parallelising techniques, such as SIMD, hitherto only employed in computing, can be applied to physical engineering as well. It has also been argued that for correctly scaling thermodynamics to molecular dimensions, one must use electric and magnetic properties, which are the dominant forces at that scale. While microscopic steam engines would have their uses, they simply cannot compete with the speed of electronics, nor lead to a zero-incoherence future. Moreover, the thermodynamic applications of electromagnetism are themselves important, opening new opportunities such as the ones just described. Like all new solutions to old problems, these methods cannot right away replace existing practices, but the promise of quantum improvements will hopefully motivate further investigation.

## About..

The author is a software engineer, presently employed as such at the IBM T.J. Watson Research Center, NY 10598. The work presented here is the result of independent, private investigation in physics over the past two decades, mostly while living in India. The author owns the patents cited in this paper and no endorsement by IBM is implied.

## Acknowledgements

To Jeff Kash and J C Tsang, for being there with a (dis)proof of concept when I was groping for one; to Vijay R Kirloskar, Chairman and M.D. of Kirloskar

Electric Company, India, for encouraging my private quests; to Paramahansa Tewari, then Chief Project Engr, Kaiga Project, for first hand introduction to challenges; and to Rolf Landauer and C J Tan at IBM, for their tolerance and general encouragement.

## REFERENCES

[1] E Fredkin and T Toffoli, *Conservative logic,* Intl. Jour. Theor. Phys., Vol 21. Nos. 3/4, 1982, pp. 219-253.

[2] Y Moon and D K Jeong, *Efficient Charge Recovery Logic,* Symposium on VLSI circuits: Dig. of Tech. Papers, Jun 1995.

[3] L Brillouin and H P Iskenderian, Elect. Commun. 25, 300 (1948).

[4] K H Spring, *Direct generation of electricity,* Acad Press, 1965.

[5] T A Edison, *Pyromagnetic Generator,* US Patent 476983, 14 June 1892.

[6] Dan Solomon, *Improving the performance of a thermomagnetic generator by cycling the magnetic field,* J Appl Phy, 63(3), 1988.

[7] V Guruprasad, *Electromagnetic heat engines and method for cooling a system with predictable bursts of heat,* US Patent about to issue, 1997.

[8] R Rosensweig, *Ferrohydrodynamics,* 1965.

[9] V Guruprasad, *Control method for Carnotizing heat engines,* Provisional specification 1997.

[10] J A Kash and J C Tsang, *Hot luminescence from CMOS circuits: A picosecond probe of internal timing,* Phys. Stat. Sol. (b) 204, 507 (1997).

[12] Excerpts of testimony before the US Senate by Dr. Roger Hastings, *The energy machine of Joseph Newman,* The Joseph Newman Publ Co, 1987.

[13] R Landauer, *Irreversibility and Heat Generation in the Computing Process,* IBM Journal, July 1961.

[14] C H Bennett, *Logical Reversibility of Computation,* IBM Jour. of Res. Dev., Nov 1993.

[15] C H Bennett and R Landauer, *The Fundamental Physical Limits of Computation,* Scientific American, July 1985.

[16] S Carnot, *Reflections on the Motive Power of Heat...* 1824, Dover reprint.

[17] E Fermi, *Thermodynamics,* Dover, 1936.

[18] C H Bennett, *Demons, Engines and the Second Law,* Scientific American, Nov 1987.

[19] D G Jablonski, *A Heat Engine Model of a Reversible Computation,* Proc of the IEEE, Vol 78, No 5, 1990.

[20] Dan Solomon, *Thermomagnetic mechanical heat engines,* J Appl Phy, 65 (9), 1989.

[21] G L Roth, *Thermoelectromagnetic energy conversion system,* US Patent 3790829, 1974.

[22] R P Feynman, Leighton and M Sands, *Lectures on Physics,* 1964.

[23] M Toda, R Kubo, N Saito, *Statistical Physics II, Inequilibrium Statistical Mechanics,* Springer-Verlag, 1992.

[24] S G Younis and T F Knight, *Asymptotically Zero Energy Split-Level Charge Recovery Logic,* Intl. Workshop on Low Power Design, 1994. Also http://www.ai.mit.edu/people/tk/lowpower/low94.ps.

[25] C H Bennett, *Time/space Tradeoffs for Reversible Computation,* Soc. for Ind. and Appl. Math. Jour. on Computing, 18(4):766-776, 1989.

[26] R P Feynman, *Quantum mechanical computers,* Foundations of Physics, 16(6), 1986.

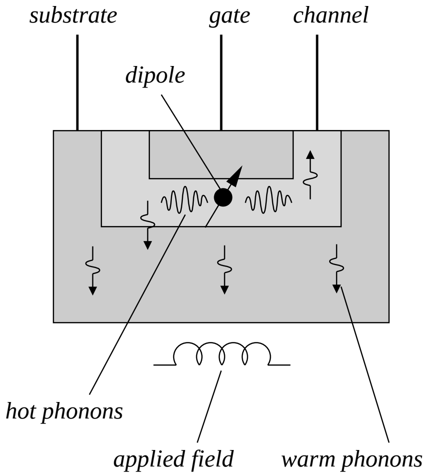


**Fig. 1. Synchronous cooling principle**

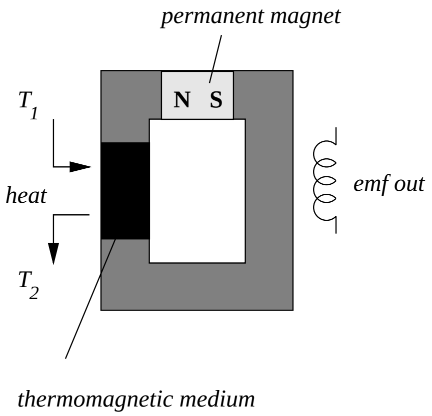
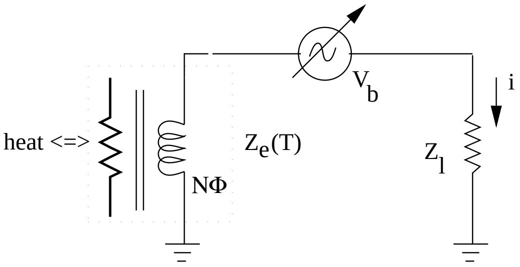
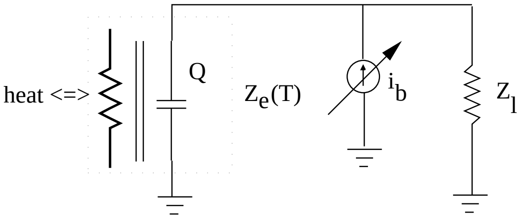


**Fig. 2. Edison’s thermomagnetic sensor**



Equation of state:
$\iota = \iota\ (N\Phi, T)$

**Fig. 3. Inductive heat engine**



Equation of state:
$V = V\ (Q, T)$

**Fig. 4. Capacitive heat engine**

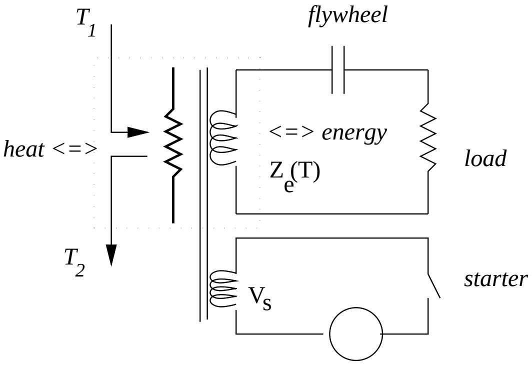


**Fig. 5. Equiv. ckt. of a mechanical engine**

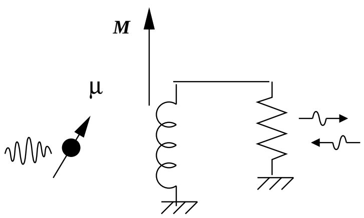


**Fig. 6. Dipole-coil exchange interaction**

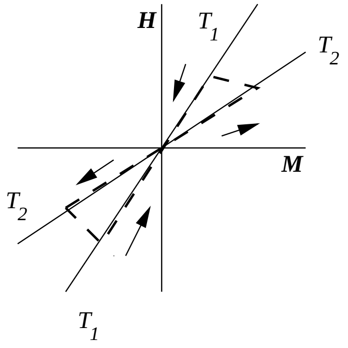


Fig. 7. Synchronous cooling cycle

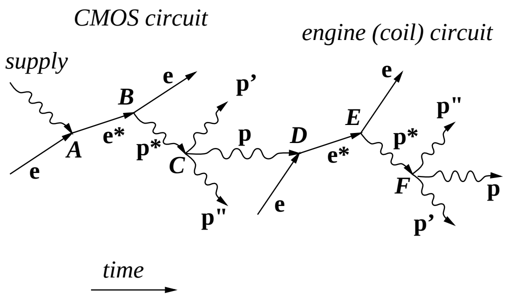


Fig. 8. Synchronous cooling

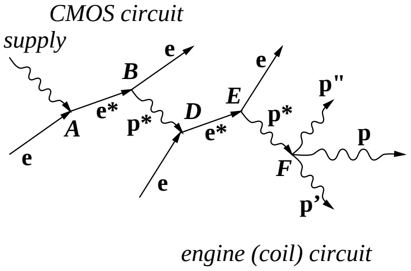


Fig. 9. Coherent cooling

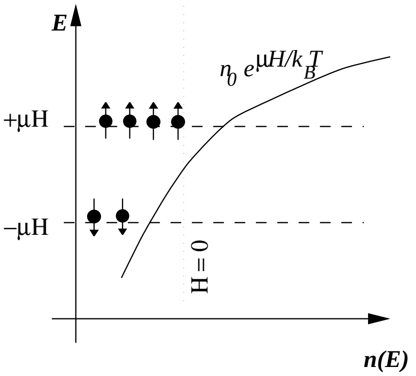


Fig. 10. D.c. population inversion

Fig. 11. Entropy of consumption