Electrical theory of thermodynamics and particle scale heat engines

V. Guruprasad

Inspired Research, New York 10509, USA.

Ultimately at single particle scales, the principal forces of interaction are electric and magnetic, and all heat is essentially kinetic energy, hence direct treatment of thermodynamic processes in terms of these forces would be more insightful and appropriate at these scales. The main result is the novel possibility of virtually eliminating the principal cause of inefficiency in current heat engines by raising the efficiency of the Carnot cycle itself to almost unity. The reason this should be possible is that the nascent particle energies in chemical reactions and photovoltaics are equivalent to excess of 10,000 K, and are even higher by seven orders of magnitude in nuclear fission, but it would require directly converting the nascent energies before their dissipation into bulk media. I show that a magnetic or dielectric engine cycle, which in fact involves electromagnetic field interactions with the medium, is indeed capable of such conversion as it can compete against thermal diffusion, instead of waiting for local thermal equilibrium. The resulting scheme also closely relates to regenerative braking. As a necessary foundation for treating the particle scale conversion, general equivalent circuits of heat engines and the circuit equations of state are developed and are shown to even provide analytic refinement over traditional heat engine theory, such as an opportunity for very closely approaching the ideal cycle using phase space velocity control. The difficulties to be addressed for particle scale conversion and virtually unity efficiency are discussed using regenerative hot carrier energy recovery, which could enable nondissipating semiconductor logic at over 100 GHz speeds, as a near-term realizable example.

Nomenclature

α	Specific negative immittance
β	Power gain due to α
χ	Magnetic susceptibility
λ	Label for a virtual photon (Fig. 17)
μ_B	Bohr magneton
ω_0	Angular frequency of cyclic operation
Φ	Total magnetic flux
Φ_0	Reference point flux in phase space
ϕ_0	Phase offset of cyclic operation
Φ_M	Magnetization flux
τ	Engine cycle time
$\widehat{G}_e$	Mean developed engine conductance
$\widehat{R}_e$	Mean developed engine resistance
$\widetilde{N}$	Mean #gates switching in a clock cycle
ξ	Total back-emf ($\equiv \xi_0 + \xi_M$)
ξ_0	Back-emf due to current change
ξ_M	Back-emf due to magnetization change
a_k	Coefficient in system characteristic equation
B_L	Load susceptance $(\equiv X_L^{-1})$
C	Electrical capacitance
$C(T)$	Temperature-dependent capacitance
C_M	Curie constant (of a material)
C_p	Heat capacity at constant pressure
C_x	Heat capacity at constant x
C_x	Heat capacity of engine medium
C_L	Load capacitance from next stage in CMOS
e^*	Label for a hot carrier (Fig. 17)
E_M	Total energy of magnetization
f	Dynamical force (on a medium)
$f(t)$	System response function in time domain
f_c	CMOS clock frequency
$g(t)$	Applied (driving) force function
G_e	Developed engine conductance
G_L	Load conductance ($\equiv R_L^{-1}$)
H	Applied magnetic field intensity
I	Inductive engine current
I_0	Reference point current in phase space
I_b	Driving current source
K	Diffusion constant
k	Summation index
k_B	Boltzmann constant
L	Electrical inductance
$L(T)$	Temperature-dependent inductance
L_0	Vacuum inductance
L_h	Inductance at higher temperature
L_l	Inductance at lower temperature
L_M	Inductance due to a medium
M	Magnetization flux density
N	Number of turns of engine coil
n	Number of moles (in ideal gas law)
n	Occupancy of an energy level
n	Order of the system characteristic equation
p	Label for a phonon (Fig. 17)
p	Pressure (in ideal gas law)
P_d	Dynamic dissipation in CMOS
P_i	Input power, drawn from voltage source V_b
P_i'	Increased power draw due to α
P_L	Power delivered to/dissipated in the load
P_L'	Increased load power due to α
Q	Heat flow or heat content (of a medium)
q	Charge in capacitive engine
q_0	Reference point charge in phase space
R	Electrical resistance
R_e	Developed engine resistance
R_h	Engine resistance R_e at high temperature
R_L	Load resistance
R_l	Engine resistance R_e at low temperature
S	Entropy (of a medium)
s	Laplace operator
s_k	Complex roots of the characteristic equation
T	Temperature (of a medium)
T_c	Curie temperature (of a material)
T_h	High temperature value (Carnot cycle)
T_l	Low temperature value (Carnot cycle)
U_h	Internal energy at higher temperature
U_l	Internal energy at lower temperature
U_M	Potential energy of magnetization
V	CMOS supply voltage
V	Load voltage in capacitive engine
V	Volume (in ideal gas law)
V_0	Reference point voltage in phase space
V_b	Driving voltage source,e.g. battery
V_{in}	Input voltage
V_{out}	Output voltage
W	Net work done by engine
x	Displacement corresponding to force f
$X(t)$	Actuator displacement
x_0	Position of zero displacement
X_e	Engine reactance
Y_L	Load admittance ($\equiv Z_L^{-1}$)
Z_c	Variable control impedance
Z_e	Total engine impedance
Z_L	Load impedance

I. Introduction

On the nano and ultimately single particle scales, the principal forces of interaction would be electric and magnetic, and the particle energies would be kinetic. Photoelectrons and nascent products of chemical reactions typically carry 1-$5\ \mathrm{eV}$, equivalent to $10,000$-$60,000\ \mathrm{K}$. Between such temperatures and a $300\ \mathrm{K}$ ambient, the efficiency of the Carnot cycle would be 0.974-0.994. Nuclear sources would be even more remarkable, since the product nucleii from U^{235} fission have initial energies of 55-$100\ \mathrm{MeV}$, for which the Carnot efficiency would be as high as 0.99999999953, i.e., the minimum theoretic inefficiency would be below 10^{-9}. We do know that the conduction band electron gas in metals and semiconductors generally takes on the order of $100\ \mathrm{fs}$ to reach equilibrium, and that the eventual dissipation into the bulk lattice as heat involves carrier-phonon interactions on the time scale of 1-$10\ \mathrm{ps}$. Utilization of these hot particle temperatures should pose no fundamental issues, as these temperatures are quite real, occurring without, say, intelligent separation from cold particles by a Maxwell demon. These temperatures have been verified by various means including infrared imaging. As the current theoretic limit on the efficiency for mechanical engines is around 60%, and that for third generation photovoltaic power is only 85% even with solar concentrators and full spectrum utilization, particle scale conversion, if achieved, could in effect double our available energy resources virtually overnight.

Moreover, our understanding of heat itself remains incomplete, despite the immense progress in nano and femto physics, without a comparable treatment of conversion using electromagnetic fields, as electromagnetics technology as we know it today had been unavailable for much of the historic development of thermodynamics. Magnetocaloric conversion and magnetohydrodynamics are both essentially mechanical, and are diffusion-favouring processes – there has been no treatment in the literature of *inductive* conversion to avoid mechanical contact and gross mechanical forces altogether, which would thereby enable "conversion at a distance" without a uniform build up of pressure.

One reason why these hot particle temperatures are currently unavailable is therefore simply that we do not as yet have heat engines that could operate at the sub-picosecond speeds to exploit individual particle temperatures. I shall show below that particle scale conversion is achievable with macroscopic engine design, if only we use electromagnetic fields to directly interact with the individual particles. The theory needed for building and operating such engines is developed below, and the associated theoretical and technological difficulties are discussed.

As will be established, *the only difference between known macroscopic engines and the particle scale convertors would be in the preparation of the thermodynamic media and the couplings required between the macroscopic engine load, the hot particles, and the media.* The circuit equations including the thermodynamics would be identical, but were themselves hitherto unknown for want of an electrical equivalent circuit analysis of heat engines, and sufficiently refined analytic treatment of heat engines in general – it will be shown below that analytical refinement leads to the equivalent circuit picture and a general representation of heat engines as *parametric amplifiers* instead of the usual view of *prime movers*, which also dictates thermoelectric and photovoltaic design as emf sources. The complementary question to be instead addressed is what prevented particle scale conversion in prior inductive engines if any.

The second problem has been that traditional thermodynamic theory was concerned with slow, diffusion-limited mechanical or thermoelectric processes. Even in emerging nanoscale power technologies, the focus is on adaptations of mechanical engine principles that would continue to remain similarly slow and diffusion-limited, and thus inherently incapable of *competing against diffusion* to exploit hot particle energies before their dissipation into bulk media and the resulting loss of temperature. Thermodynamic use of electromagnetic fields calls for electromagnetic interactions that similarly go beyond the past direct conversion ideas involving "lumped" electrical components. Again, the applicable direct conversion ideas have themselves also stayed undeveloped, ironically because they were viewed solely in terms of the properties of bulk solid media, which presented dual barriers of very high "thermal mass" and slow diffusion in external combustion designs. These barriers vanish at the single particle scale without any change to the (macroscopic) engine theory, because engine operation in general is concerned solely with dynamical coupling to a medium and its temperature variation, and not with how the heat enters or leaves the medium! In the particle scale engines, the heat must merely occur and vanish as hot particles, and the key problems become electromagnetic field design and inductive coupling. Also, hot particle energy that escapes direct conversion would be still captured by conventional means.

As will be discussed, other basic deficiencies in the prior theory also stand in the way of a requisite understanding of the physics and engineering involved, such as the fallacies of using a permanent magnet bias in the thermomagnetic circuit, of mimicking the flywheel of mechanical engines using sinusoidal excitation, and of relying on mathematical resemblance instead of physical principles in identifying the field variables for the corresponding Carnot cycles. The traditional advocacy of quasi-static operation for the ideal cycle is also revisited in view of our intent to compete with diffusion processes, instead of waiting for them to complete as in equilibrium thermodynamics.

Accordingly, Section II presents the problem of hot carriers in semiconductor logic. Although this problem merely concerns energy recovery and cooling, both phonon interactions and dissipation mechanisms are best understood today in semiconductors, and a successful application of the present ideas to this problem would have immediate utility. The challenges for hot carrier energy recovery and conversion are identified in Section III as those of dynamically coupling the hot particles individually to the load system or circuit, and of effectively "rectifying" the random motions of the hot particles in the process. While the first problem would be addressed by using electromagnetic fields as the coupling mechanism in place of a mechanical piston or turbine, the second problem more particularly concerns entropy and calls for a rigorous thermodynamic solution. The remainder of this paper is organized towards this end as follows.

First, equivalent electrical circuit descriptions of heat engines, along with the corresponding equations of state and description of the Carnot cycle, are obtained in Section IV by applying linear differential analysis and linear control systems principles to heat engine operation. The resulting analytical description of heat engines and thermodynamic conversion as (negative) immittances is a refinement over the traditional perception of heat engines as prime movers, which includes the flywheels mechanisms used for sustaining engine operation, and over parametric amplification, as portrayed in prior direct conversion literature, which would imply strictly sinusoidal but suboptimal excitation. This electrical formalism is necessary for defining and characterizing the inductive circuits needed for the field coupling. A further treatment relating thermal and frictional losses to the engine cycle execution is given in Section V, to complete the engine circuit perspective, and to explain why the quasi-static principle is inapplicable to hot particle conversion.

Second, the thermodynamic theory for the field coupling is developed in Section VI via a first principles analysis of the thermodynamic properties of a medium, to particularly develop a robust insight into the electromagnetic forces and displacements, and to overcome the limitations of the past treatments, which have been purely analytical, in this regard. This first principles analysis explains, for instance, how the gas laws and the Curie-Wiess laws of magnetism are complementary, and are necessary and sufficient to exhaust the space of possible first order equations of state, thus finally assuring us of a complete theory of thermodynamics.

Lastly, Section VII ties these circuit and field perspectives together by considering the detailed interactions of the individual dipole moments in the magnetic medium of an inductive engine with heat and the engine load. The broad principles of realizing particle scale conversion, and the remaining issues for research are also discussed.

II. The opportunity for hot carrier energy recovery

Fig. 1 illustrates the switching contribution to dissipation in discrete CMOS logic.

When the input gate voltage V_{in} is low, the n-channel MOSFET is off, and the p-channel MOSFET conducts, so that the output voltage V_{out} is pulled high by the supply voltage V. When V_{in} is high, the p MOSFET is off, and the n MOSFET conducts, pulling V_{out} low. In either case, as at least one of the MOSFETs is off, there is very little current, and hence little dissipation, in the steady state. The basic CMOS circuit shown is an invertor.

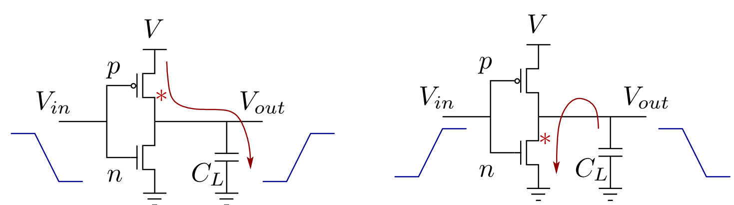

Figure 1. Switching currents in CMOS

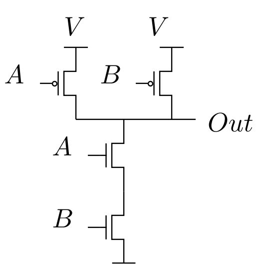

Figure 2. NAND gate

Complex logic is built by combining MOSFET gates in various ways while maintaining the complementary design. Fig. 2 illustrates a CMOS NAND gate. As the n MOSFETs are in series, the output voltage will be pulled down only when both inputs A and B are low, which would also allow the p MOSFETs to be both off.

Almost all of the dissipation in CMOS occurs during switching. The load at the output is essentially the gate capacitance of the next stage of logic, as represented by C_L in Fig. 1, since the gate conductance (at the next stage) would be very small compared to $5\ \mathrm{m}\mho$ $(= 200\ \Omega)$ of the on-state of the p or n MOSFET. When the output is driven high by falling V_{in}, as shown in Fig. 1 left, C_L is charged from V by the p MOSFET. When the output is subsequently driven low by a rising V_{in}, as in Fig. 1 right, this charge drains through the n MOSFET to ground. The rise and fall times are generally kept much shorter than the operating clock period, hence each switching cycle expends a charge of V/C_L, where V is the supply voltage[a], so that the dissipation per gate per switching cycle is V^2/C_L. The switching dissipation, also called dynamic dissipation, of an integrated-circuit (IC) chip is then $P_d = \widetilde{N} V^2 f_c / C_L$, where f_c is the clock frequency, C_L is the average capacitance of the gates, and $\widetilde{N}$ is the mean number of gates actually switching states at each cycle.

[a]The supply voltage is conventionally marked V_{dd} in the literature, although the source and drain are inverted between the p and n MOSFETs. The majority carriers are holes in the p and electrons in the n channels, so the dissipation follows the source-drain inversion, as shown.

Most known ways to reduce the dynamic dissipation either require operating at lower speeds, or involve additional circuit complexity. Another known method is to trade off the dynamic dissipation for steady overall dissipation using current mode logic,[1,2] which is particularly advantageous at high operating speeds of up to $10\ \mathrm{GHz}$.[3] However, we do seem to have hit a wall of late with respect to further increases in processor clock speed because of the dissipation. The present opportunity concerns avoiding dissipation, as in the charge recovery schemes, but without similarly increasing the circuit complexity and without reducing the operating speed. The opportunity particularly concerns the dynamic dissipation in Fig. 1, which can be attributed mostly to hot carriers within the conducting MOSFET channel giving up their energy to the lattice near the drain (marked by asterixes in the figure).

Figs. 3 and 4 illustrate the hot carrier distributions in the conducting MOSFET channels during switching. These figures are graphs from Monte Carlo simulations which have been well verified. In both, the hottest carriers are shown in red, and clearly have energies of $1\ \mathrm{eV} \sim 10^4\ \mathrm{K}$, corresponding to operating voltages of a little over $1\ \mathrm{V}$. Fig. 4 shows that in a silicon-on-insulator (SOI) channel, the hot carriers form near the drain as they gain energy in flight from the potential gradient in the channel and from Coulomb interactions with other carriers.

Low power mobile processors are being operated at $0.3\ \mathrm{V}$ or so to reduce the power, and the hot carrier temperatures would be somewhat lower in these devices. However, if the power could be recovered without dissipation, and if the recovery scheme were more efficient at higher voltages, there would be little reason to reduce the voltage, since a low operating voltage also reduces the noise margin. A similar argument applies to the clock speed in the present method, as we would be competing against the diffusion of heat, and would therefore not require slow operation.

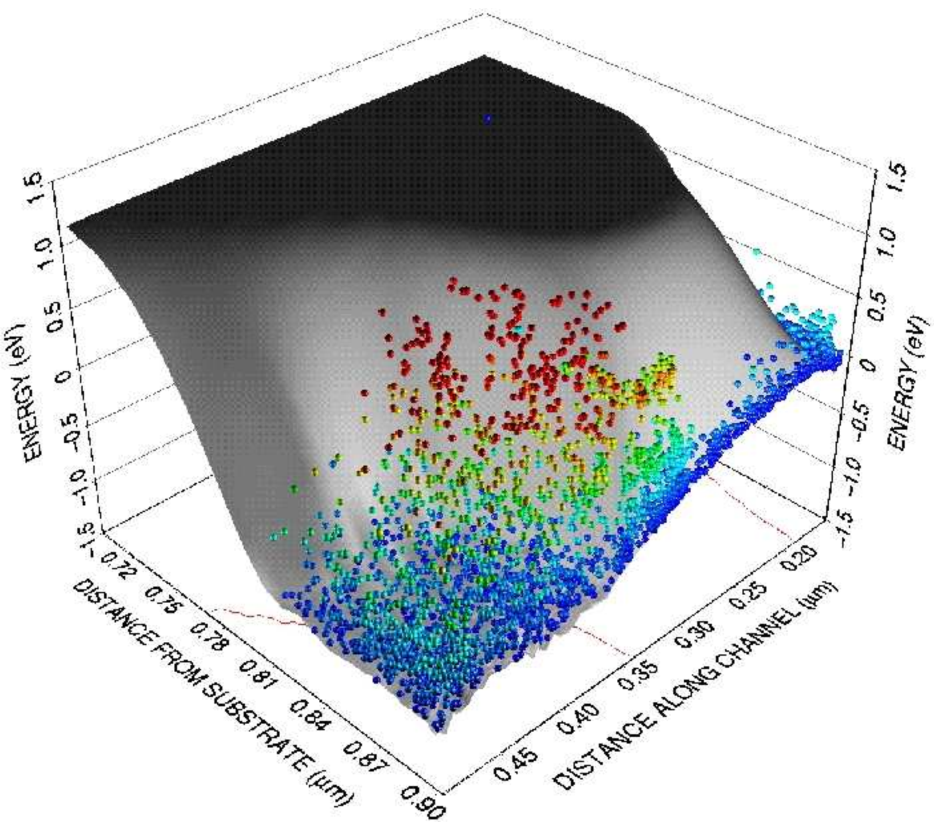

Figure 3. Carrier energy distribution in nMOSFET (Courtesy: IBM DAMOCLES project)

The motivations for considering recovery of the hot carrier energies are as follows. Only a small fraction of the total charge that flows during switching turns into hot carriers, but most of the energy dissipated at each gate switching event can be attributed to the hot carriers. They are also primarily responsible for lattice damage by impact ionization. Yet, as a minority fraction of the charge, they are also less responsible for the overall charge transport necessary for the logic function, hence recovering their energies and slowing them down *before* they arrive at the drain should achieve both power reduction and device longevity at the full operating clock speeds without loss of functionality.

Any mechanism to remove energy from the hot carriers before they reach the drain, without simultaneously slowing the rest of the current, would be thus beneficial for both improving reliability and reducing dissipation. Although this

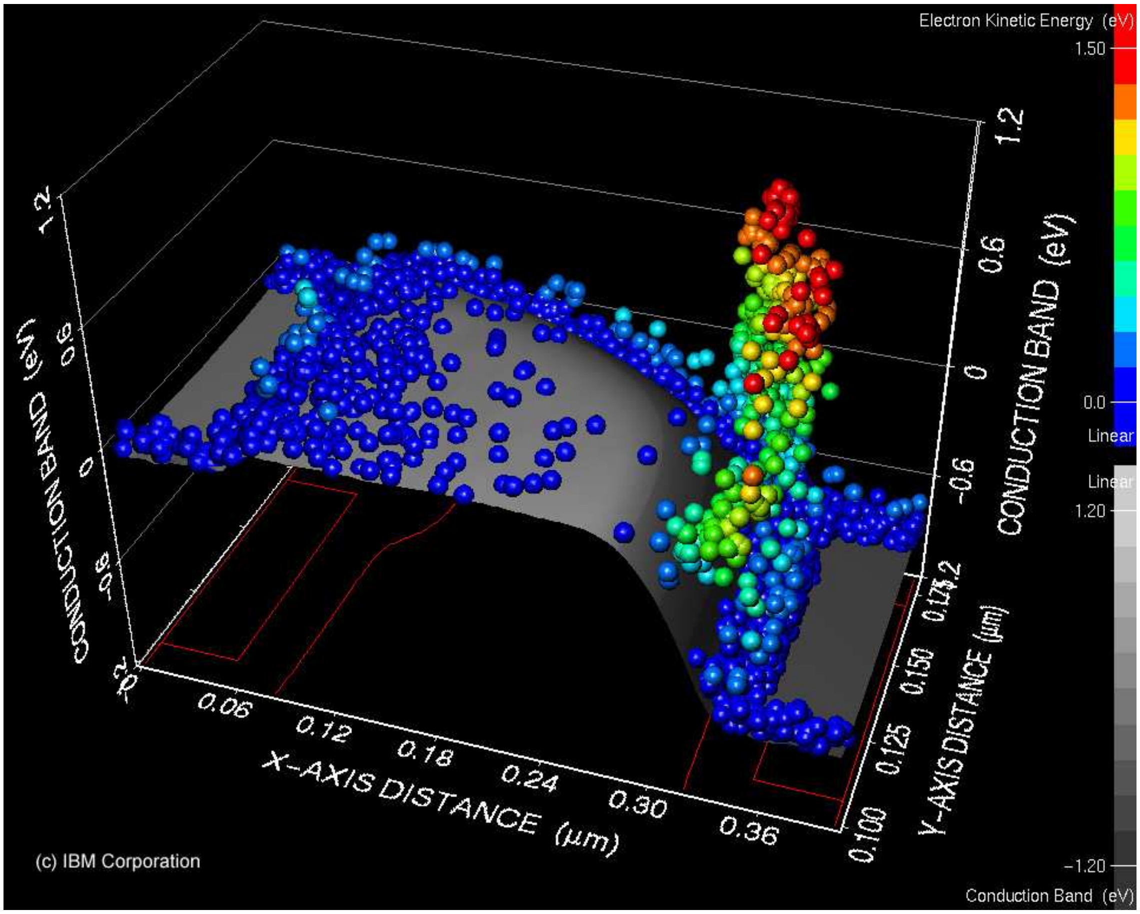

Figure 4. Carrier energy distribution in SOI channel (Courtesy: IBM DAMOCLES project)

concerns energy recovery and cooling, not power generation, the mechanisms of electron-phonon and phonon-phonon interactions involved in dissipation are most well researched in the physics of semiconductors and optoelectronics, and the hot carrier problem represents an opportunity testing the present theory with immediate commercial implications. In addition, similar problems of interaction probabilities and nonuniformity of particle density and energy distributions would be also encountered in the general case, as will be shortly described, and could have similar solutions.

For a sample set of numbers, consider Intel's upcoming quad-core Itanium family processor codenamed Tukwila. Its $170\ \mathrm{W}$ rated dissipation over a die area of $21.5 \times 31.5\ \mathrm{mm}^2$ would amount to a dissipation density of $2.43\ \mathrm{MW\,l}^{-1}$ assuming a substrat of thickness of $100\ \mu\mathrm{m}$. (To compare, the internal combustion engine in a modern sedan develops under $200\ \mathrm{kW}$ per litre of cylinder volume.) Assuming clock distribution similar to that of Itanium 2, the clock power would alone be about 25% of the total power, or 42.5 W. Since standby power has been generally reduced to the order of 4 W, most of the remaining 127.5 W can be attributed to switching dissipation. Given that the chip has 2×10^9 transistors, the mean dissipation would be about $60\ \mathrm{nW}$ per transistor. If all transistors switched at every clock cycle, it would mean 0.025 fJ per switching event. At 1.2 V, this would mean a transfer of $2.5 \times 10^{-17}\ \mathrm{J}/(1.2\ \mathrm{V} \times 1.6 \times 10^{-19}\ \mathrm{C}) \approx 130$ carriers per transistor per transition.

From Figs. 3 and 4, we could reasonably expect about 10% to be hot, and to account for say 50% of the switching energy. If we could intercept and convert the energies of say 80% of the hot carriers, we would recover about 50 W, and since this would be returned as electrical power, the dissipation in the chip would reduce from 170 to $120\ \mathrm{W}$. It would still need conventional cooling but for only $120\ \mathrm{W}$. The 80% lower hot carrier dissipation within the individual transistors could be alternatively exploited as an opportunity to clock the chip 5 times faster, i.e., at $12.5\ \mathrm{GHz}$, for the same rate of damage from the hot carriers. At 90% recovery, the same rate of damage could be expected at 10 times faster operation, i.e., at $25\ \mathrm{GHz}$. Though these numbers are conjectural, and constitute goals rather than results, the projected recovery of 50 W out of 170 W is consistent with an alternative approach of clock-power logic,[4] in which dissipation is avoided by preserving charge at high potential, albeit at the cost of additional circuit complexity.

III. Challenges for particle scale conversion

The hot carriers in CMOS are illustrative of the difficulties with converting hot particle energies. The key problems to be addressed, and the nature of possible solutions, are as follows:

Ephemeral state. As explained in the Introduction, the actual lifetimes of particles in the hot states is of the order of femtoseconds before the initial phonon interactions, and the immediate high energy phonon excitations in turn decay within picoseconds. The conversion scheme therefore needs to act effectively within such short lifetimes.

Minority population. Although individual hot particles possess high energies and high effective temperatures, they are constitute only a minority among even the same species of particles. In CMOS, the hot carriers are formed partly by the potential gradient and in part from Coulomb interactions leading to a Maxwellian distribution superposed on the linear motion due to the potential gradient. In the chemical, solar and nuclear energy scenarios mentioned in the Introduction, hot particles are similarly heated or produced *within* a material medium, but loose their energies very quickly to the bulk medium, so that the majority of the species remains relatively cold.

Hence, a method to convert hot particles should be able to either interact selectively with the hot particles, or at least prevent dilution of the conversion efficiency by the cold particles. Selective interaction poses the problem of how the convertor can know which particles are likely to become hot, as their energies must be tapped within the few picoseconds that they will remain hot. This problem would be harder than the Maxwell demon's, since the demon can afford to learn which molecules are more energetic by examining bounced photons.[5,6] Selective interaction is thus unlikely to succeed, but preventing dilution by competing against diffusion is unexplored.

Coupling particle energies. Preventing dilution from the cold interactions will require a non-mechanical approach, because a mechanical piston or turbine can be pushed only by particles in contact, and not by hot particles deep within the medium. The inertia of the coupling mechanism is also in question, as the ratio of its inertia to that of an individual hot particle will determine the fraction of energy transferred – should the coupling not yield at all, or should it yield to the particles but fail to couple the energy onward to an external load, there would be no net transfer. In the latter case, our particle coupling mechanism would be merely execute Brownian motions. The result would be very similar to that historically envisaged by M. Smoluchowski for the Maxwell demon.[6]

Randomness of occurrence and motions. The hot particle energies would kinetic, and their velocities at any instant would be likely along random directions. The CMOS scenario is an exception in this regard, as we can safely assume that at least till the first collision, a hot carrier's velocity vector will be aligned with the voltage gradient. We would also know their approximate locations (MOSFET drains) and occurrence times (clock transitions).

However, these assumption will not be applicable in scenarios where we do not have positional or directional control over the processes producing the hot particles. We could exert some control over the production times, for example, using electro-optic shutters to pulse the exposure to photons, and thereby pulse the generation of hot photoelectrons, but doing so would compromise the continuity of conversion and would be rather difficult in many cases, e.g., with nuclear fission. The hot particle locations would be also in general random.

As will be revisited in Section VII, all four problems are addressable using inductive heat engines with electromagnetic fields. as a key mechanism for coupling the nascent hot particle energies to an external load:

- The heat engine principle eliminates dependence on the locations and directions of motion of the hot particles. As the energies are related to speeds, the result might superficially resemble the Maxwell demon, but the particle energies are utilized as temperature, exactly as in comparable heat engines that use bulk thermodynamic media. The cyclic operation requires pulsed exposure to the hot particle energy source, which is available in the CMOS scenario. Workarounds may be feasible in other scenarios, as suggested for the photovoltaic case above.

  There is no determination of individual particle locations or energies, i.e., all that the engine load sees would be a net emf comprising inductive impulses contributed by the hot particles as and when they do work on the medium and the load. Such impulses are incidentally known in ferromagnetism as the Barkhausen effect.

- The use of fields allow near-instantaneous coupling to the individual particles, and the superposition principle applicable to electromagnetic fields and induction avoids dispersion of the hot particle energies to neighbouring cold particles or to surrounding bulk media in the process of coupling to the load.

  More precisely, *the individual particles interact directly and simultaneously on the external load via the field, instead of having to pass on their energies, by diffusion, to the molecules closest to a piston or turbine blade* as in a traditional (mechanical) heat engine. This is a general property of inductive engine design, to be formally described ahead, whether or not used for particle scale conversion!

- Heat engine principles further suggest the use of thermodynamic media distinct from the heat source. We may thus use magnetic (or dielectric) atoms to mediate dynamical interaction with nonmagnetic or uncharged hot particles, and to enhance the likely low field interaction probabilities of many species of hot particles, such as the neutrons or daughter nuclei in nuclear fission.

The theory of inductive electrical heat engines necessary to explain these various aspects is accordingly developed in the next few sections.

IV. Electrical heat engine circuits and modelling

A. Linear components analysis

Recall from basic system theory that the general integro-differential system equation, is reducible, with suitable choice of variables, to a general differential dynamical equation:

$$\sum_{k=0}^{n} a_k \frac{d^k f(t)}{dt^k} \equiv \left[\sum_{k=0}^{n} a_k \frac{d^k}{dt^k}\right] f(t) = g(t) \tag{1}$$

where $f(t)$ is the system response function, $g(t)$ is the (applied) force function and a_k are real valued coefficients that characterize the system. The corresponding *characteristic equation*

$$\left[\sum_{k=0}^{n} a_k \frac{d^k}{dt^k}\right] f(t) = 0 \tag{2}$$

is solved using a trial solution $f = e^{st}$, to transform Eq. (2) to the polynomial equation

$$\sum_{k=0}^{n} a_k s^k = 0 \tag{3}$$

on dividing further by e^{st}. We know from algebra that Eq. (3) has n complex roots $s_0 \ldots s_{n-1}$, so Eq. (2) is equivalent to the product

$$\prod_{j=0}^{n-1} \left(\frac{d}{dt} - s_j\right) f(t) = 0 \quad , \tag{4}$$

and suggests the form of the particular solution as

$$f(t) = \left\{\prod_{j=0}^{n-1} \left(\frac{d}{dt} - s_j\right)\right\}^{-1} g(t) \quad , \tag{5}$$

which is easy to relate to solution by the Laplace transform, in which the Laplace transform operator $s \sim d/dt$, and the n roots $s_0 \ldots s_{n-1}$ are eigenvalues of the s operator and constitute the poles of the system response. Rigorous treatment of the above steps can be found in basic texts on differential equations and control system theory.

A further result from algebra is that if the coefficients a_k are all real, complex valued s_j can occur only in conjugate pairs. This equivalently means that the polynomial $\sum a_k s^k$ can be factored into at most quadratic factors, i.e., factors of the form $(L\,d^2/dt^2 + R\,d/dt + C^{-1}) \sim (Ls^2 + Rs + C^{-1})$. identifiable with linear electrical circuit elements, viz inductances (L), resistances (R) and capacitances (C). The coefficients in the linear factors of the form $(a\,d/dt + b) \sim (as + b)$ can be interpreted as $a \equiv L$ and $b \equiv R$, or $a \equiv R$ and $b \equiv C^{-1}$, depending on whether f is identified with a current or a charge, respectively. These electrical equivalent representations were once important for the testing of mechanical system designs, but have been replaced by software equation solvers and simulators. In our context, they serve as reminders that any complex dynamical system is ultimately reducible to elements corresponding to L, R and C, and that any subsystem or component may be regarded as reducible to such elements in this analytical sense.

The traditional notion of a gas heat engine operating between two reservoirs at different temperatures and using a flywheel to power the compression strokes, is thus a reducible system equivalent to an oscillatory circuit of at least two elements, one of which is thermodynamically passive and serves as the flywheel. This analytical reduction will be also important because electrical and electronic technologies uniquely allow fine control of the engine cycle execution.

B. Circuit models and equations of state

Fig. 5 illustrates the circuit representation of a heat engine broken down to irreducible electrical elements, and using an electrical driving source in place of the traditional flywheel. The full circuit must include a resistance R_L (more generally an impedance Z_L), or equivalently a conductance G_L (or admittance Y_L), to denote load consumption and ohmic losses, and at least one non-resistive element to represent the thermodynamic conversion.

With hindsight of thermomagnetic and magnetocaloric theory, we may use an inductor L as the representative element, so that its magnetic "core" can also serve as the thermodynamic medium. The result is an L-R circuit, which needs to be driven by a voltage source V_b, in absence of thermodynamic conversion. In the steady state, the load consumes the power $P_L = I^2 R_L$, where I denotes the current in the circuit. The voltage source supplies input power $P_i \equiv IV_b = I^2 R \equiv V_b^2/R$, where R is the total circuit resistance including the ohmic contribution of the inductor L.

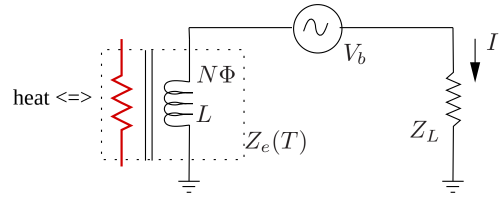

Figure 5. Inductive heat engine circuit

To represent thermodynamic conversion, we need at least one temperature sensitive circuit element. Since R_L (or Z_L) represents load to be driven by the converted power, it is not a desirable candidate. Moreover, Carnot theory requires the thermodynamic medium or device to both produce and consume work in a cycle, whereas a resistance generally lacks the association of a medium and is inherently dissipative by concept. An inductance is a more natural candidate, in this sense, and allows us to attribute the temperature sensitivity needed for thermodynamic conversion, represented by making the inductance a function of temperature $L \equiv L(T)$, to its magnetic core medium. We also represent the heat flow for the thermodynamic conversion by a heat exchanger, resembling a transformer in a generalized sense, and denote the total impedance developed by the inductor as Z_e, including the inductive reactance X_e, the imaginary part of Z_e, and a real part including the inductor resistance.

Fig. 6 illustrates the equivalent capactive heat engine circuit, in which a current source I_b drives load conductance G_L, and the capacitance $C \equiv C(T)$ now needs to temperature dependent, i.e., with a temperature sensitive dielectric medium, to represent thermodynamic conversion. This would be further consistent with the hindsight of dielectric engines. The admittance Y_e developed by the engine would correspondingly include X_C, the capacitive reactance, in its imaginary part. The key result of thermodynamic conversion will be shown to be a negative immittance. A different assignment of the conversion functionality, say to the source or another resistor, cannot yield a fundamentally different model. Our inductive and capacitive circuits are thus quite general and sufficient to represent all heat engines.

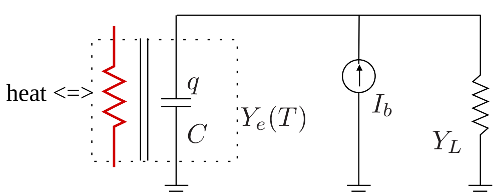

Figure 6. Capacitive heat engine circuit

The thermodynamic variables for the *circuit equations of state* are required to be a temperature T, a force f and a corresponding displacement x. All three must be properties of the magnetic or dielectric medium in the inductor or capacitor, respectively. The distinction between the temperature T and the other two variables is that *the force and dimension variables must define work done on the medium or on the load*. The further distinction between which property constitutes a force and which must be a dimension is also dictated by mechanics: $f.\Delta x$ must constitute work, while $x.\Delta f$ at most concerns an internal change in the material with no work done, where Δx and Δf denote incremental variations in x and f, respectively. These relations hold for a gas engine with the association of f with the pressure p acting on a piston and x with the volume V displaced by the piston in a cylinder containing the working gas.

For the inductive heat engine, the natural candidates for f and x would be the current I and the effective magnetic flux $N\Phi$, respectively, where Φ denotes the actual total flux and N, the number of turns of the inductor coil. To check, consider that $I.\Delta(N\Phi)$ represents incremental change in magnetic energy, and this would indeed translate to work on the load when the flux decreases, since in the circuit representation, the voltage source cannot sink power. Likewise, no power flow between the inductor and load is contributed by an incremental change in the current at a constant flux represented by $(N\Phi).\Delta I$; the continuous $I^2 R_L$ power dissipated in the load would come entirely from the driving voltage source V_b in this case. Also, although the force variable is I, it is in effect driven by the applied voltage V_b, so V_b needs to be variable, at least in magnitude, for executing thermodynamic cycles.

Current source representation is usually less familiar, but it should be analogously clear that $V.\Delta q$ would involve electrical work done on or by the capacitor, and the capacitor can only do work on the load, whereas $q.\Delta V$ would not entail such work and can only arise from an intrinsic change in capacitance due to heat flow. Hence, in the capacitive heat engine, V constitutes the force, and q, as the corresponding charge displacement.

The known formulae $LI = N\Phi$ for the total flux in an inductor, and $q = CV$ for the charge in a capacitor, lead to the corresponding circuit equations of state

$$\frac{I}{N\Phi} = \frac{1}{L(T)} \quad , \tag{6}$$

for the inductive heat engine

$$\frac{V}{q} = \frac{1}{C(T)} \quad , \tag{7}$$

and for the capacitive heat engine, respectively, to compare to the gas equation of state $pV = nRT$. As in the gas heat engine theory, setting T constant in Eqs. (6) and (7) determines isotherms over the respective thermodynamic spaces.

To determine the adiabats of the inductive engine, we need to set $I.\Delta(N\Phi) + C_x \Delta T = 0$, where C_x denotes the heat capacity of the magnetic medium and would be C_p for a bulk medium. This yields the derivative $d\Phi/dT$ in terms of I and C_x, defining the slope of the adiabat at any given point in the $\langle I, N\Phi, T\rangle$ phase space. The adiabatic curve is then obtained by integrating the slope from a starting point $(I_0, N\Phi_0)$. The integration is tedious but straightforward, and leads to the adiabatic equation

$$I = \left(\frac{C_x}{L_T}\right)^{1/2} . \left(1 - \frac{\Phi}{\Phi_0}\left[1 - \frac{C_x}{L_T I_0^2}\right]\right)^{-1/2} \quad \text{where} \quad L_T \equiv \frac{dL(T)}{dT} \quad . \tag{8}$$

The corresponding capacitive adiabatic condition is $V.\Delta q + C_x \Delta T = 0$, and yields the capacitive adiabatic equation

$$V = \left(\frac{C_x}{L_T}\right)^{1/2} . \left(1 - \frac{q}{q_0}\left[1 - \frac{C_x}{C_T V_0^2}\right]\right)^{-1/2} \quad \text{where} \quad C_T \equiv \frac{dC(T)}{dT} \quad , \tag{9}$$

where V_0 and q_0 similarly identify a starting point in the $\langle V, q, T\rangle$ space.

C. Generalized operating principles

It would be straightforward to verify that the isotherms defined by Eqs. (6) and (7) do intersect the adiabats of Eqs. (8) and (9), so that Carnot cycles can be defined just like in gas heat engine theory, and further that these cycles would be consistent with the known cycles of heat engines using magnetic and dielectric media. The thermodynamics would not be new, but the principles of operation are unobvious, have not been adequately treated in the past, and are critical for our vision of near-unity heat convertors using electromagnetic fields instead of pistons or acoustic waves. They are also important for completion of our analytical refinement of mechanical engine theory by circuit analysis, and lead to new insight that could be applied to existing heat engines and their applications.

To begin with, as first noted by Solomon[7,8] using traditional analysis, the inductive engine circuit of Fig. 5 requires no permanent magnet or static magnetic bias for its operation, which was once thought necessary. The circuit current I itself provides the needed magnetization, via the flux relation $LI = N\Phi$, and power would be generated as back-emf induced by demagnetization. More particularly, we have in standard electrical engineering theory the relation

$$\xi_M = -N\dot{\Phi}_M \equiv -N\frac{d\Phi_M}{dt} = -L_M\frac{dI}{dt} \equiv -\chi L_0 \frac{dI}{dt} \tag{10}$$

for the instantaneous back-emf due to any flux change, where the subscript M identifies the magnetization (material) contribution, the subscript 0 refers to the vacuum component, and χ denotes the (temperature-dependent) susceptibility of the magnetic material. The total instantaneous back-emf is then

$$\xi \equiv \xi_0 + \xi_M = -N(\dot{\Phi}_0 + \dot{\Phi}_M) = -(L_0 + L_M)\frac{dI}{dt} \equiv -(1+\chi)\, L_0 \frac{dI}{dt} \equiv -L\frac{dI}{dt} \quad . \tag{11}$$

The difference from standard electrical engineering is that ξ_M will now have a contribution from both changing of the current I and variation of L_M due to change of susceptibility χ_M over the thermodynamic cycle, i.e., as

$$\xi_M = -\left(\frac{dL_M}{dt} I + L_M \frac{dI}{dt}\right) \equiv -L_0\left(\frac{d\chi_M}{dt} I + \chi_M \frac{dI}{dt}\right) \quad . \tag{12}$$

We should expect to see ξ_M vary in magnitude as well as in sign over a complete cycle, since during magnetization, ξ_M would oppose V_b, and during demagnetization, ξ_M should add to V_b in order to contribute energy converted from heat. The total energy changes in the isothermal magnetization and demagnetization processes would be

$$\Delta U_l = \left.\frac{L_l I^2}{2}\right|_a^b = L_l\,(I_b^2 - I_a^2)/2 \quad \text{and} \quad \Delta U_h = \left.\frac{L_h I^2}{2}\right|_c^d = L_h\,(I_d^2 - I_c^2)/2 \quad , \tag{13}$$

respectively, where the subscripts l and h refer to the low (T_l) and high (T_h) temperature limits of the cycle, and a, b, c and d denote end points of the T_l and T_h isothermal processes. Since adiabatic processes by definition involve no heat transfer, they cannot contribute to the total work ΔW done by the cycle. We thus have $\Delta W = \Delta U_l - \Delta U_h$.

Further, these energy changes signify power flow in or out of the magnetic medium during the respective processes. The power flows are equivalent, in the circuit perspective, to instantaneous equivalent resistances

$$R_e \equiv \frac{\xi_M}{I} = -\left(\frac{dL_M}{dt} + L_M . \frac{1}{I}\frac{dI}{dt}\right) \equiv -L_0\left(\frac{d\chi_M}{dt} + \chi_M . \frac{1}{I}\frac{dI}{dt}\right) \quad . \tag{14}$$

In particular, ξ_M only changes during the adiabatic processes, so for the isothermal equivalent resistances, we have

$$R_l = -L_l\, d(\ln I)/dt \quad \text{and} \quad R_h = -L_h\, d(\ln I)/dt \quad . \tag{15}$$

As the current would be changing in opposite ways during these processes, the mean resistance over a cycle would be

$$\widehat{R}_e = \tau^{-1}\left(\int_c^d R_h\, dt - \int_a^b R_l\, dt\right) \qquad \text{where } \tau \equiv \oint dt \text{ is the total cycle time.} \tag{16}$$

We would expect $\widehat{R}_e$ to be negative when there is conversion of heat, and negative if the engine is operated in reverse as a heat pump or refrigerator. The negative resistance $\widehat{R}_e < 0$ of an inductive engine implies *parametric amplification* of the power drawn from source V_b. We would expect an analogous equivalent conductance $\widehat{G}_e$ from a capacitive engine of the same sign and inverse magnitude as $\widehat{R}_e$. We shall revisit some of these aspects in more detail in Section VI.

Fig. 7 is a graph illustrating how the *power gain* (β) varies with the *specific negative resistance* developed by the inductive engine, which we define as

$$\alpha \equiv -\widehat{R}_e/R_L \quad . \tag{17}$$

For the equivalent capacitive engine, α is likewise defined as *the specific negative conductance* $-\widehat{G}_e/G_L$, G_e denoting the mean engine conductance. Note that α is dimensionless. In accordance with electrical engineering terminology, we may refer to α more generally as *specific negative immittance*, and say that the engine immittance $\widehat{R}_e$ or $\widehat{G}_e$ changes the ideal reactance X_L or Y_L into an impedance $Z_L \equiv \widehat{R}_e + iX_L$ or admittance $Y_L \equiv \widehat{G}_e - iB_L$, where B_L is called susceptance $\equiv X_L^{-1}$. (We retain $i \equiv \sqrt{-1}$, instead of j, for consistency with physics and mechanical engineering.)

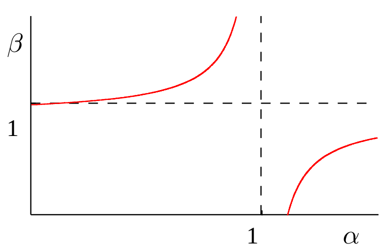

Figure 7. Operating point and gain

Without thermodynamic conversion, the rms (root-mean-square) current in the inductive engine circuit would be $I_{b(rms)} \equiv V_{b(rms)}/R_L$. The presence of conversion would raise it to

$$I_{rms} = \frac{V_{b(rms)}}{R_L + \widehat{R}_e} = \frac{V_{b(rms)}/R_L}{1 + \widehat{R}_e/R_L} \equiv \frac{V_{b(rms)}/R_L}{1-\alpha} \equiv \frac{I_{b(rms)}}{1-\alpha} \quad , \tag{18}$$

so the *current gain* is $1/(1-\alpha)$. This further means that the voltage developed across the load would rise from V_b to

$$V_{rms} = I_{rms}R_L = I_{b(rms)}R_L/(1-\alpha) = V_{b(rms)}/(1-\alpha) \tag{19}$$

in the presence of conversion. The load power must rise from $P_L \equiv I_{b(rms)}^2 R_L = V_{b(rms)}^2/R_L \equiv V_{b(rms)} I_{b(rms)}$ to

$$P'_L \equiv V_{rms} I_{rms} = P_i/(1-\alpha)^2 \quad . \tag{20}$$

The power drawn from the source also rises from P_i to $P'_i \equiv V_{b(rms)} I_{rms} = P_i/(1-\alpha)$, with help of Eq. (18). Hence the actual power gain is

$$\beta \equiv \frac{P'_L}{P'_i} = \frac{P_i/(1-\alpha)^2}{P_i/(1-\alpha)} = \frac{1}{1-\alpha} \quad , \tag{21}$$

which is identical to the current gain and also to the rise in the power drawn from the source, consistent with the total load power growth by the factor $\beta^2 \equiv 1/(1-\alpha)^2$ from Eq. (20). The inverse relation

$$\alpha = 1 - 1/\beta \tag{22}$$

allows us to determine $\widehat{R}_e$ by measuring the power gain, as $\widehat{R}_e = -\alpha R_L$. Treatment for the capacitive engine would be analogous. The above equations lead to the following observations:

- The parametric amplification is only defined for the interval $0 \le \widehat{R}_e < R_L$ corresponding to $\alpha \in [0,1)$. At $\alpha \ge 1$, the gain goes through infinity and becomes negative, so $\tau^{-1}\int R_h\,dt \ge R_L$, or equivalently, $\xi_M \ge V_b$. In other words, the circuit would become overly reactive and impossible to operate.

- The increase in the power drawn from the driving source by the same factor β limits how close α can be taken to unity with a finite driving source. A variable resistance in series with the load can help regulate α, and hence the engine gain β, the power drawn, and the output power to the load.

V. Friction, leakage and phase space motion

A general barrier in working with heat engines is the large set of partial differential relations that need to be worked out for each engine cycle. We mostly confine to Carnot cycles to avoid such complication. The preceding analysis presents an illustrative opportunity for more closely analyzing motion in the thermodynamic phase space, i.e., the execution of thermodynamic processes, as follows. The generalized equation of state $f = f(x,T)$ further implies

$$\Delta f = \left.\frac{\partial f}{\partial x}\right|_T \Delta x + \left.\frac{\partial f}{\partial T}\right|_x \Delta T \equiv f_{,x}\,\Delta x + f_{,T}\,\Delta T \quad , \tag{23}$$

using the comma suffix to denote partial differentiation keeping the other thermodynamic variable constant. Clearly, $f_{,x} \equiv \partial f/\partial x|_T$ represents isothermal elasticity of the medium, and $f_{,T} \equiv \partial f/\partial T|_x$ corresponds to the variation of pressure in a gas with temperature at constant volume. From the conservation of energy, Eq. (33), we also have

$$\frac{dU}{dt} + \frac{dQ}{dt} + \frac{dW}{dt} = 0 \quad . \tag{24}$$

We already have $dW/dt = f\,dx/dt$, from the definition of incremental work, $\Delta W = f\,\Delta x$. For the internal energy changes, we may write

$$\Delta U = \left.\frac{\partial U}{\partial T}\right|_x \Delta T + \left.\frac{\partial U}{\partial x}\right|_T \Delta x \equiv C_x \Delta T + U_{,x}\,\Delta x \quad , \tag{25}$$

where C_x denotes the heat capacity of the medium at constant x following the notational convention of C_V and C_P in gas theory. Using this to evaluate dU/dt and replacing dW/dt by the load power $f\,dx/dt$ transforms Eq. (24) to

$$(f + U_{,x})\,\frac{dx}{dt} + \frac{dQ}{dt} + C_x \frac{dT}{dt} = 0 \quad , \tag{26}$$

or

$$\frac{dx}{dt} = -\frac{dQ/dt + C_x\,dT/dt}{f + U_{,x}} \quad . \tag{27}$$

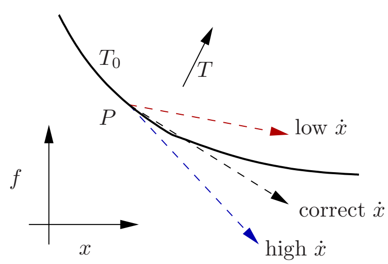

Figure 8. Phase space motion control

Equation (27) determines the displacement velocity necessary to achieve a desired amount of force f at a given heat flow rate dQ/dt and rate of change of temperature dT/dt, or equivalently, the velocity necessary to ensure a desired rate of change of temperature under a given amount of applied force f and heat flow rate dQ/dt, and so on. Using the known thermodynamic identity

$$U_{,x} \equiv \left.\frac{\partial U}{\partial x}\right|_T = \left(T \left.\frac{\partial f}{\partial T}\right|_x - f\right) \quad , \tag{28}$$

which is independent of the state function $f(x,T)$, we may eliminate the dependence on f in Eq. (29) and obtain

$$\frac{dx}{dt} = -\frac{dQ/dt + C_x\,dT/dt}{T(\partial f/\partial T)|_x} \quad . \tag{29}$$

Equation (29) determines the displacement velocity necessary to achieve desired heat flow dQ/dt and rate of change of temperature dT/dt, at any given point in the thermodynamic phase space. Fig. 8 illustrates the concept.

Say our object is to move in the f-x phase space along the isotherm $T = T_0$ (solid curve in the figure). From Eq. (29), we obtain $\dot{x} = -\dot{Q}/(T_0\,\partial f/\partial T)$ for the required speed, where we evaluate $\partial f/\partial T$ for the instantaneous point P. If our actual speed is too low, the heat inflow will cause the temperature to rise, and our motion will deviate upward from the tangent to the isotherm at P in the direction of increasing T. Conversely, if our speed is too high, we will find ourselves moving away from the tangent in the direction of decreasing T.

To ensure motion along the isotherm, we must ensure that the actual speed exactly matches the right side of Eq. (29) at each successive point on the path. In the denominator, $(\partial f/\partial T)|_x$ is the projection of the normal to the isotherm, i.e., the increasing T arrow, on the f (upward) direction. Likewise, for adiabatic motion, we may set $dQ/dt = 0$ in Eq. (29), and obtain $\dot{x} = -C_x\,(dT/dt)/(T_0\,\partial f/\partial T)$. The rate of change of temperature is ordinarily not a constraint for the adiabatic operations, i.e., we may raise dx/dt to the maximum possible, to shorten the time for leakage and any residual heat flow, and instead compute $\dot{x}.dT/dt = -(T_0\,\partial f/\partial T)/C_x$ as the rate of change of temperature to expect.

Equations (23)-(29) are all quite general and not particularly dependent on the form of the state function f, or even the usual interpretation of Q as heat, T as the temperature, and U as internal (mechanical) energy, so long as the work premise $\Delta W = f\,\Delta x$ and Eq. (24) hold. The only additional constraint required by traditional thermodynamic theory is that all of the changes must be performed *reversibly*, because we have not yet accounted for frictional losses and ensured the validity of the observable f, x and T as thermodynamically representative of the whole medium.

Traditional theory mandates quasistatic operation for the validity of the thermodynamic variables, on grounds that operating slowly would ensure that the equilibrium conditions are attained at each step so that intrinsic quantities like pressure and temperature become homogeneous within the medium. Frictional losses would be also minimized.

The traditional advice overlooks the fact that leakages of heat, and often also of pressure, are just as unavoidable as friction. Friction even disappears totally in superfluids, but not radiative and conductive leakage. The per-cycle loss due to leakage would be proportional to the time taken for the cycle, meaning that an ideal quasistatic engine would have absolutely zero efficiency regardless of reversibility with respect to friction and internal states. Leakage has not received the same attention, however, because we are unlikely to ever want to run engines so slowly as to make leakage a concern of theory. In the exceptionally few cases of slow engine design, such as cryogenic magnetic refrigerators, the occurrence of leakage becomes obvious and mundane. In practical heat engines, therefore, reversibility is limited principally by frictional losses, temperature drops due to thermal diffusion especially with condensed thermodynamic media and external high or low temperature baths, and inhomogeneity of thermal conditions, from either operating too fast for diffusion processes within the medium, or from turbulence. In internal combustion engines, the inhomogeneity issue is mostly avoided by keeping the piston speed subsonic!

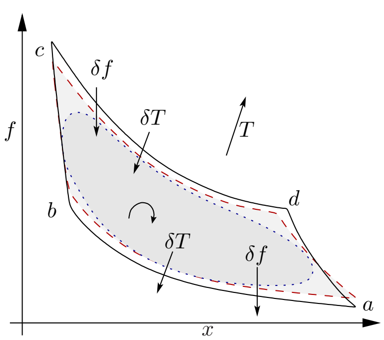

Figure 9. Effect of frictional losses

Fig. 9 illustrates the general impact of frictional losses on a heat engine cycle (solid lines) in the phase space. These losses are of two kinds:

- Dynamical loss $\delta f \propto R\,dx/dt$ due to mechanical friction in mechanical engines, and ohmic and hysteretic losses in electrical engines. It reduces the force available to compress the medium, when work is being done on the medium during compression strokes, and reduces the force available from the medium to work on the load during expansion strokes.
- Thermal loss $\delta T \propto K\,dQ/dt$ due to the heat flow, where K denotes the applicable diffusion resistance. This reduces the temperature available to heat the medium during the high temperature processes, and also raises the exhaust temperature (during the low temperature processes).

The losses thus squeeze the engine cycle by shifting it downward during the high temperature processes, and upward during the low temperature processes. A smaller engine cycle results, yielding less work, as shown by the broken lines. Both losses are frictional since they are proportional to the speed of operation, represented by the heat flow rate in the thermal case. They would be certainly eliminated by quasistatic operation, but it would both reduce the throughput to zero and do nothing to reduce leakages, so the theoretic reversibility would be achieved but at zero actual efficiency!

On the other hand, sinusoidal motion of the piston due to the flywheel in mechanical engines, which was replicated in prior direct conversion studies employing magnetic and dielectric media, would likely miss executing the corners of an ideal cycle. For an ideal cycle with sinusoidal motion given by $x = x_0 \sin(\omega_0 t + \phi_0)$, where ω_0 is the angular frequency of the motion and ϕ_0 is an initial phase offset, Eq. (29) would require

$$\begin{aligned} \frac{dQ/dt + C_x\,dT/dt}{T(\partial f/\partial T)|_x} &= -\frac{dx}{dt} = -x_0 \cos(\omega_0 t + \phi_0) \\ \text{whence} \quad dQ/dt + C_x\,dT/dt &= -x_0\;T(\partial f/\partial T)|_x \cos(\omega_0 t + \phi_0) \quad . \end{aligned} \tag{30}$$

Equation (30) says that the sum of the heat flow rate dQ/dt and the rate of change of temperature dT/dt must exactly fit the sinusoidal profile dictated by the right side throughout the cycle. Ignoring variations due to the $T(\partial f/\partial T)$ factor for the moment, we need to ensure $dQ/dt = 0$ along the adiabats, so the temperature must rise and fall sinusoidally with the motion. This happens to be easy as the adiabatic temperature change is dictated by the motion itself. However, we would also need to ensure $dT/dt = 0$ along the isotherms ($a \to b$ and $c \to d$ in the figure), so dQ/dt would have to also rise and fall sinusoidally during these times, which is not easy when working with piston engines.

We would want the peak heat transfer rates well within the isothermal processes, preferably near c, where the difference in f between the two isotherms is largest, so as to maximize the net work per cycle. Automotive variable fuel injection and ignition timing technologies are based on similar reasoning.

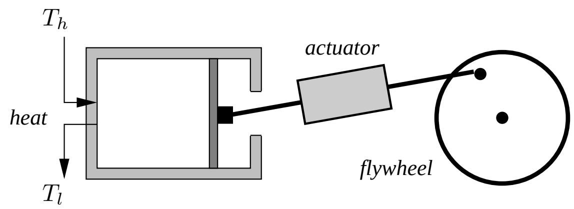

Figure 10. Varying the crankshaft

However, as the sinusoidal profile would reduce the speed at the ends of each stroke, the leakages will inevitably further diminish the available cycle area, particularly compromising the adiabatic portions, as indicated by the dotted line in Fig. 9. We therefore need to be slower during the isothermal motions, to minimize the frictional losses, and fast during the adiabats, to minimize the leakage. The optimal velocity profile therefore cannot be sinusoidal and cannot be realized using just the traditional flywheel.

One approach is to modulate the velocity profile, say by coupling the crankshaft to the flywheel via a cam. A more flexible approach is to build an actuator into the crankshaft, as shown in Fig. 10, such that the length of the crankshaft can be varied during the cycle by an electronic control system. From Eq. (30), we get

$$-\frac{dx}{dt} = -x_0 \cos(\omega_0 t + \phi_0) + X(t) = \frac{dQ/dt + C_x\, dT/dt}{T(\partial f/\partial T)|_x}$$

where $X(t)$ is the cam or actuator control contribution. We thus get

$$X(t) = \frac{dQ/dt + C_x\, dT/dt}{T(\partial f/\partial T)|_x} + x_0 \cos(\omega_0 t + \phi_0) \quad . \tag{31}$$

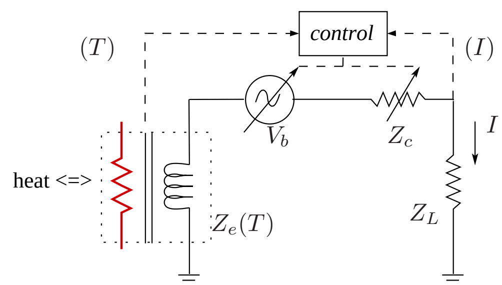

Figure 11. Effect of frictional losses

$X(t)$ would vanish, obviating the velocity modulation, if and only if the first term on the right exactly cancelled the sinusoidal contribution of the flywheel represented by the second term.

Fig. 11 shows the corresponding technique for an inductive engine, where we may modulate one or both of the source V_b and an optional control impedance Z_c based on feedback from sensors monitoring the temperature T and the load current I during the cycle. We must have $dx/dt \equiv N d\Phi/dt \equiv \xi$ of Eq. (11). Equation (29) then yields

$$V_b + IZ_c = \frac{dQ/dt + C_x\, dT/dt}{T(\partial f/\partial T)|_x} - IZ_L \quad , \tag{32}$$

as the condition to be met by varying V_b and Z_c for the optimal velocity profile, as it would be generally more difficult to control the heat flow dQ/dt and the load Z_L.

A sinusoidal load current is generally preferred for a.c. electrical power, whereas Eq. (30) has shown that strictly sinusoidal load operation tends to be suboptimal because of the difficulty of controlling heat flow. High capacity power conditioning technology does exist and continues to evolve, however, so the above approach may be usable and could help improve efficiency in general.

VI. Generalized thermodynamic principle

In Section A, we introduced the generalized notions of a force f, a corresponding displacement x in a medium and a temperature T, as thermodynamic variables. The stated test for identifying which property constitutes force and which the displacement is that $f.\Delta x$ should constitute work done on or by the medium, whereas $x.\Delta f$ should not. In the case of a volume of gas in a cylinder with a piston, f may be identified with either the pressure of the gas or the total force on the piston, and x, with the volume of the gas or the total displacement of the piston, respectively.

A corresponding constraint exists for the third variable T that it should be associated with a non-work energy transfer ΔQ via a fourth variable S, such that $\Delta Q = T.\Delta S$, consistent with the traditional definition of (incremental) entropy as $\Delta S = \Delta Q/T$. The only analytical constraint between Q and the dynamical pair (f, x) is the conservation of energy, requiring the existence of a third energy variable U such that

$$\Delta U + \Delta Q + \Delta W = 0 \quad , \tag{33}$$

where $\Delta W \equiv f.\Delta x$. When T is interpreted as temperature, Q can be readily identified as heat, S, as the entropy, and U, as the internal energy, but the equations developed so far are not dependent on this interpretation. For example, we could use them identically for the mechanical transformer shown in Fig. 12, in which we may identify f and x with the force on the larger piston at right and its displacement, respectively, and use the pressure p and the displacement volume V of the smaller piston at left in the place of T and S, respectively. The law of conservation of energy for the transformer requires $p.\Delta v + f.\Delta x + \Delta U = 0$, where ΔU must to include any actual heat transfers and temperature changes during operation. Such interpretations also carry over to the notion of engine cycles and energy conversion, and permit us to treat quasi-static execution of Carnot cycles in electrical power transformers, in place of the strictly sinusoidal analysis currently used to represent distortion and hysteretic losses, for example. Similar issues of needing slow operation to maintain the homogeneity of the variables and to minimize frictional losses would be encountered.

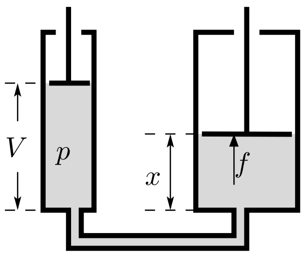

Figure 12. Fluidic transformer

The resulting generality is especially important in our context. In particular, it would be assuring to note that the electrical equations of state (6) and (7) are the only fundamental physical forms of the thermodynamic equation of state besides the gas law. This follows from noting that if f, x and T refer to concurrent properties of the same physical medium, it should be possible to relate each as a function of the other two, i.e., as $f = f(x, T)$. The notion of a state particularly requires that f be single-valued at each point over the domain of engine operation in the x-T space.

We may additionally require that f be analytic over this domain, so that first order partial derivatives can exist; relations between these partial derivatives would be trivial generalizations of known thermodynamic identities like Maxwell's and Bridgeman's relations. The adiabatic condition for obtaining adiabats from the equation of state is

$$f.\Delta x + C_x.\Delta T = 0 \quad , \tag{34}$$

since $f.\Delta x \equiv \Delta W$ and $C_x.\Delta T \equiv \Delta U$, so that by Eq. (33), there can be no heat transfer ΔQ; we already employed this to derive the adiabats, Eqs. (8) and (9), for the inductive and capacitive engines, respectively.

Given the continuity, it would be further possible to construct a Taylor expansion to compute f at any point in the neighbourhood of a reference location in the x-T space, and a polynomial expansion of f in x and T should suffice for a large class of applications. The lowest order polynomial forms for $f(x, T)$ would be the linear equations

$$fx = rT \quad \text{and} \quad f/x = rT \quad , \tag{35}$$

where r denotes a constant of proportionality. Dimensional arguments help in ruling out simple additive combinations like $f + x = rT$. The only other basic combinations, $fx = r/T$, $x/f = rT$ and $1/fx = rT$, differ merely in inverting the definition of T. Equations (35) are thus the simplest irreducible set of forms for equations of state. The first was associated with the gas law by identifying f with pressure p and x with volume V in a gas heat engine.

The analogous associations, as explained in Section A, of f and x with the current I and the effective flux $N\Phi$ in the inductive engine circuit, and with voltage V and the charge q in the capacitive engine circuit, respectively, identify the circuit equations of state (6) and (7) with the second of Eqs. (35). The Curie-Weiss law

$$\frac{H}{M} \equiv \frac{1}{\chi} = -\frac{T - T_c}{C_M} \tag{36}$$

for a paramagnetic or ferroelectric material, where C_M denotes its Curie constant and T_c, its Curie temperature, is an extension of this form. The Curie-Weiss isotherms would be parallel to the Curie isotherms, which clearly radiate linearly through the origin. As the adiabatic condition concerns only the incremental changes in the temperature, the Curie-Weiss adiabats would be similar.

The circuit equations look similar, but are not identical, as the form of $L(T)$ and $C(T)$ are yet to be determined. In an actual magnetic engine, $L(T)$ would depend on both the coil geometry and the magnetic core, and we would also need to account for variations in current and field distributions at high operating frequencies. For equivalent circuits of mechanical heat engines, $L(T)$ and $C(T)$ must be computed from the mechanical model.

Nevertheless, the overall similarity of form to the Curie law suggests a similar phase space geometry, so the engine cycles would be similar as well, but for the difference that unlike most mechanical engines, the electrical engines can operate with both positive and negative displacement values, as shown by the ideal inductive engine cycle depicted in Fig. 13 using an a.c. excitation current. Since a mechanical engine would be unlikely to accommodate a negative force and displacement, the equivalent circuit engine cycle must be generally also considered to operate within one quadrant, unlike a real electrical heat engine.

The above considerations look mundane, but there has been so much confusion in the past that the field itself has remained undeveloped and poorly understood.

For example, the patent for an electromagnetic heat engine[9] was immediately reported by PhysicsWeb as a (purely) magnetic engine,[10] missing both the novelty of converting hot carrier energies, and the century of history of magnetic engine design starting with Tesla (1889) and Edison (1992),[11] cited as prior art by the patent examiner! The oldest error is the presence of a permanent magnet in the magnetic circuit by Edison [*op. cit.*], which has been retained in the studies of Brillouin and Iskenderian.[12] This has been now corrected by Solomon,[7,8] but with focus still on the magnetic (mechanical) cycle and missing the full perspective, as follows.

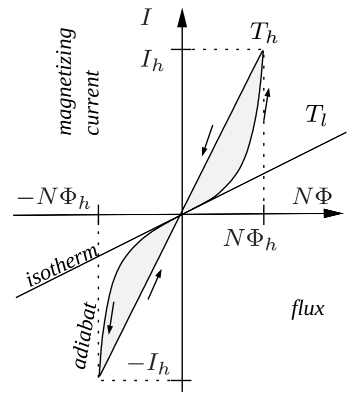

Figure 13. Magnetic Carnot cycle

The physical essence of thermodynamic conversion is that the same range of x is traversed in the successive compression and expansion strokes, but with different forces f depending on the temperature. Net work results corresponding to $x.\Delta f$, where Δf is the difference in f between the strokes. This key principle implies that thermodynamic gain can only result from dynamical work input in the compression strokes. Edison's bias magnet violates this principle by seeking to obtain an emf from the temperature change alone, without requiring a magnetic compression by the load current. This error in Edison's design has remained unnoticed.

A further problem in Edison's design is that by restricting full use of the magnetization cycle, and thereby limiting the converted energy per cycle, the permanent magnet bias made the thermomagnetic engine more of a sensor than a convertor of heat, since the defining object of a heat engine is the generation of power, and not merely the emf. Sensors typically have a high output impedance, whereas a power source must sustain substantial currents, which calls for a low output resistance. By the theory in Section C, a low R_e would mean high α, assuming that we would still have $R_L > R_e$. This could be viewed as requiring a high induced emf ξ_M, based on Eq. (14), but a lower R_L would achieve the same, and low load resistance is in fact the norm for generator equipment, so we do not automatically need very large ξ_M to develop power. The impedance considerations are crucially important for hot carrier conversion.

The dynamical constraint that $f.\Delta x$ define work ΔW, would seem obvious and intuitive, but have evidently posed difficulties in the past. For example, at least one old textbook, *Solid state physics* by A.J. Dekker, even insists that the H field is the analogue of volume and the magnetization M, that of pressure, for all thermodynamic purposes (page 447, 1981 reprint). Subsequent work on magnetocaloric conversion appears to have corrected this confusion. Another frequent source of error is the multitude of CGS units traditionally used for magnetism, and still present in standard reference material, which make it especially difficult to relate the magnetic quantities with electrical power. A further source of confusion is the universal use of potential instead of total energy for representing quantum levels – which causes the magnetization energy to be exactly negative of the potential energy indicated by the level populations. An associated fallacy in much of past work is the notion that ferromagnetism is necessary for high conversion densities

Lastly, as the Curie equation of state (second of Eqs. 35) involves x in the denominator, the displacement actually relates inversely to work as $f.\Delta x = -\Delta W$, instead of $+\Delta W$ as in gas theory. In the inductive engine, this means that work is done *on* the medium to magnetize it when $x \sim N\Phi$ increases, and conversely, work is done *by* the medium *on* the load during demagnetization when $x \sim N\Phi$ decreases. This difference affects the design of mechanical magnetic engines like Tesla's thermo-magnetic motor and the more recent magnetocaloric engines, but its main impact on the electrical engine circuit design and operation is to merely reverse the phase of the work flow to and from the medium with respect to the driving voltage or current.

A related common confusion concerns the ease of magnetizing a ferromagnetic medium, as this actually leads to early onset of saturation at a low energy density. As with gases, it is the linear equation of state of paramagnetism that allows for indefinitely greater energy densities, and even magnetocaloric engines are only able to provide substantial conversion when the applied field H is raised well past saturation to 15-20 T where the susceptibility is again linear.[13] We need the medium to provide a high density of dipole, but without saturating, so the ideal medium for an inductive heat engine would be superparamagnetic. Actual power densities and throughputs using bulk solid or fluid media have been treated in the literature.[13,14] The past studies have established that a heat engine using a solid medium heated and cooled only by contact with hot and cold reservoirs would be severely limited in its performance by two main factors – the large "thermal mass" of solid media and the poor conductivity of heat in solids as compared to gases. These limitations hold identically for an inductive or capacitive heat engine under the same conditions.

What is different for particle scale conversion, as in the CMOS hot carrier conversion scenario described in Section II, is that hot particle production processes are necessarily *in situ*, and that their energy conversion must compete with diffusion to avoid dilution of the hot particle energies in any case. Both problems associated with condensed matter, of high thermal mass and slow thermal diffusion, would be thus avoided.

VII. Implementing particle scale heat engines

All of the preceding theory of inductive electrical heat engines was developed with the intent of application to the conversion of hot particle energies, so as to raise the Carnot efficiency to near unity. As explained in Section III, the four key problems to be addressed for this purpose are

i. obtaining energy from the hot particles within their femtosecond lifetimes or within the picosecond lifetimes of the associated phonon excitations, so as to minimize the dilution of their effective temperatures;

ii. obtaining energy from the hot particles even though the high energy state will be likely confined to a minority in the particle species even within the femtosecond lifetimes, i.e., converting the hot particle energies independently of the presence of, and likely unavoidable interaction with, similar and more numerous cold particles;

iii. effective dynamical coupling all the way from the hot particles to an external load system or circuit, since a poor coupling would not only limit the conversion rate but would in effect reflect the energy back into the medium, resulting in bulk heat at a lower temperature that would be inefficient to convert; and

iv. converting without dependence on the occurrence times, locations, or directions of motion of the hot particles.

As stated in Section III, problems (i) through (iii) can be addressed by using electromagnetic induction as the coupling mechanism, as the induction field can interact instantly and independently with all particles in its spatial distribution. More particularly, problem (ii) concerns the fact that in a traditional mechanical heat engine, the effective pressure on the piston or turbine blade corresponds to the bulk temperature of the medium, and not the hot particle temperatures that would be momentarily reached in internal combustion. A similar premature dilution of the hot particle energies to the colder particles or to the bulk medium before conversion can be avoided with an inductive coupling, because all of the particles would be able to interact directly with the load by electromagnetic induction throughout the engine cycle. In fact, as will be explained shortly, *the direct parallel interaction with all regions of the medium obviates dependence on the uniformity of temperature altogether.* Thus, the capability for competing against diffusion has been present all along but had remained unrecognized, thanks to the prevailing mechanical perception of thermodynamics and for want of the dynamical perspective developed for the first time above in Sections V and VI.

As further stated in Section III, problem (iv) would be adequately addressed by exploiting the hot particle energies purely as heat and using the same principles of heat engine operation that have been proven to work with bulk magnetic and dielectric media, the only differences being in how heat occurs within the particulate media in the first place and how adequate coupling of the particle energies to the engine load circuit is obtained. The first aspect, the occurrence of high particle temperatures, does not *per se* concern heat engine theory, although it was pointed out in the Introduction that photovoltaic, chemical and nuclear energy sources all entail hot particle production. The second aspect, the form of coupling, is a matter of electromagnetic field and circuit design, and is independent of thermodynamics altogether. We still need some form of pulsing of the hot particle energy source. This is available in the form of clock transitions which drive the switching in synchronous CMOS logic, and we would need analogous workarounds in other cases.

Below, we review the operation of an inductive heat engine from a particle perspective, and discuss the constraints due to the operating principles of heat engines, the feasibility of achieving the suitable conditions, and the practicality of particle scale conversion. Independence of the precise times and velocities of the hot particles is proved in subsection B, and of their precise locations, in subsection C below, thereby fully resolving problem (iv).

A. Particle scale perspective of the magnetic cycle

Recall from Section IV-C (page 10) that in an inductive engine, we would have power flow into and out of the medium during magnetization and demagnetization, respectively, and further that the direction of the power flow is reflected in the instantaneous sign of induced emf ξ_M (Eq. 12) and the equivalent resistance R_e (Eq. 14).

As remarked in Section VI, paramagnetism is to be generally preferred for thermodynamic purposes, since the linearity assures high conversion power densities similar to gas engines. Practical power generation designs, including the magnetocaloric scheme described by Rosensweig,[13] call for superparamagnetic media, whose higher susceptibility is needed more for electrical engineering reasons very similar to those governing transformer design, rather than just conversion density. These include presenting high enough primary impedance to avoid a short circuit, regulation of the secondary voltage with varying loads, and confining the flux, all three of which would be equally applicable to an inductive engine used for power generation.

The subpicosecond response necessary for particle scale conversion makes paramagnetism mandatory. While flux confinement would be still an operational constraint, the greater inertia of superparamagnetic particles, which are really ferromagnetic grains too small to form complete domains, could limit their speed of response. As will be discussed in the next subsection, this may not rule out their utility altogether. What we need is a distribution of atomic dipoles that is thin enough to remain paramagnetic and to permit the hot particle production, but dense enough to ensure interaction with a large fraction of the hot particles. As remarked in the Introduction, the energy of any hot particles that escape conversion by the (super)paramagnetic dipoles would remain within the material environment of the hot particles, and simply degenerate into bulk heat at a lower temperature, which can continue to be converted by conventional means, in reverse to cogeneration, which concerns utilization of waste heat from conventional power generation.

Fig. 14 illustrates a further detail concerning the dipoles. At left is the usual Boltzmann distribution of the occupation number n of dipoles against the magnetic potential energy $U_M = -\mu \cdot \mathbf{H}$, with the axes reversed. Such a diagram is commonly used in texts to explain the quantum picture of magnetization. It also closely relates to the energy level representation used in laser theory. The potential energy is however misleading in the heat engine context, as it might erroneously suggest that magnetization should release energy, as a steady state of nonzero magnetization requires that more dipoles be aligned along an applied H field than in the other direction, which would seem to require less overall energy than the unmagnetized equilibrium in which there would be equal populations at both levels. However, we do know that it takes an electric current to cause a net magnetization, and that energy is definitely absorbed from the current, which means that the *total* energy of magnetization E_M must vary inversely as U_M. In fact, we know from electrical theory that $E_M = -U_M$, so for purposes of reasoning about power conversion, the figure at the right, showing the variation of n against E_M, would be more helpful.

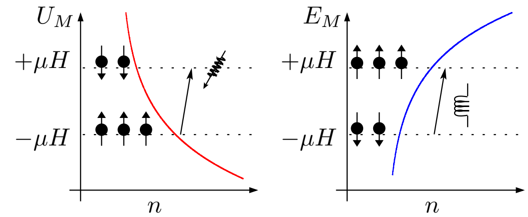

Figure 14. Dipole potential and total energies

Fig. 14 (right) also brings out the right intuition with respect to heat, because we tend to think of a higher energy level as inherently less stable, and therefore that dipoles at the higher level would have a tendency to be pushed down to the lower level with the release of energy. This rule of stability is properly associated with the potential and not the total energy in the context of conservative fields, but in the case of magnetization and heat, we do know that raising the temperature generally causes demagnetization, meaning that heat indeed tends to expel dipoles from the higher total energy state of magnetization. In fact, we know from quantum theory that demagnetization cannot occur without heat. Landauer's theory concerning dissipation in computation suggests that both state transitions are impossible without performing work and incurring dissipation.[15] The ratio of the strengths of the thermal and magnetization forces, given by $k_B T / 2\mu_B H$ where $\mu_B \equiv 9.274 \times 10^{-24}$ J T^{-1} denotes the Bohr magneton, is over 200 at ordinary temperatures, meaning that the equilibrium dipole distribution is strongly, if stochastically, controlled by thermal activity. The actual speed of change, including for flipping a dipole into alignment with an applied H field, however depends on the strength of the dipole-phonon interactions, which are characterized by relaxation times that might be distinct from, and possibly longer than, the hot particle relaxation times. In particular, nuclear magnetic relaxation times are much longer, as much as several seconds, because of the extremely weak coupling of nuclear moments to the phonons. The lattice relaxation times would constrain the rate of "priming" of the hot particle convertor, but not its conversion ability, as will become clear from considering the actual operation of a hot particle energy convertor, discussed next.

B. Thermodynamic principle for particle scale conversion

As stated in Section III, the rationale for using a heat engine is that we cannot use any mechanism that would depend on the instantaneous locations or directions of motion of the hot particles. A heat engine would still require a pulsed supply of hot particles, but it would be a first step.

To appreciate how an engine can operate under these conditions, consider first how the hot particle energies might be hypothetically harnessed without using the inductor of Fig. 5 as a heat engine. If the hot particles possess magnetic moments of their own, we could position the inductor near their expected region of occurrence and examine the total induced emf $\xi \equiv -N(\dot{\Phi}_0 + \dot{\Phi}_M)$ (Eq. 11). As Φ_0 would be not be changing in absence of the cyclic engine operation, and would be likely just zero, all of the induced emf ξ must come from change in the magnetization Φ_M, but this too would be likely zero in absence of engine operation. The individual dipoles would be flipped in random ways by the random occurrences and collisions in random directions by the hot particles, the mean induced emf $\langle \xi \rangle$ would be zero, and the instantaneous emf would be noise corresponding to the equivalent temperature of the impinging hot particles. The Boltzmann distribution is a statistical equilibrium, in which dipoles are constantly flipping in and out of alignment with the instantaneous H. The incidence of hot particles would tend to demagnetize by raising the energies of some of the dipoles, such that these dipoles will then behave as if they are at a higher temperature ($\sim T_h$ in the Carnot cycle treatment of Section IV-C).

Fig. 14 helps explain how the inductive engine can convert the energies of hot particles despite their randomness of locations and motions. Assume that we are given an inductive engine with a paramagnetic medium in which hot particles are injected in periodic bursts. The engine can convert heat to work only during demagnetization, i.e., when the engine current I is dropping in *magnitude*. Whenever the current is building up in magnitude the medium would be getting magnetized, drawing energy from the driving source V_b. It follows that we must synchronize the engine cycle to the hot particle pulses, such that the hot particles occur mostly, if not always, during the demagnetization phase.

Fig. 15 is a timing diagram illustrating this notion. At time $t = a$, the engine current I begins to diminish, crossing 0 at time $t = b$, and peaking in the reverse direction at $t = c$. The time interval (a, b) is the demagnetization phase of the engine cycle, during which incident hot particles would be converted. Any hot particles arriving during the magnetizing phase (b, c) would cause a loss in the overall magnetization, and hence in the capacity to convert hot particles in the next demagnetization phase (c, d). (Notice that the inductive engine can execute two thermodynamic cycles a-c and c-e in each cycle a-e of the engine current, and that the terminology of magnetization and demagnetization phases has exact meaning with respect to these thermodynamic cycles.)

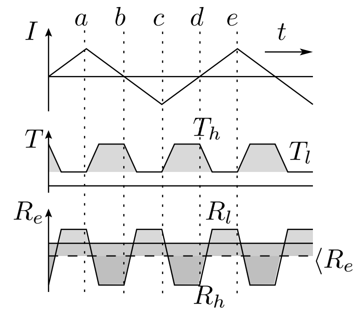

Figure 15. Engine timing

The figure also illustrates the positive and negative resistances R_e developed by the magnetization and demagnetization, respectively. If enough hot particle excitation occurs during each demagnetization phase, the negative resistance R_h, corresponding to the isothermal demagnetization in the traditional Carnot cycle, will exceed the positive resistance R_l developed during magnetization, in the sense of making the mean developed resistance $\langle R_e \rangle$ negative (Eq. 16).

The timing diagram Fig. 15 and this explanation are idealized, since we cannot be *a priori* certain of the uniformity of the hot particle bursts, nor of their sufficiency to produce a uniformly raised temperature T_h throughout the medium. The uniformity is shown, in the next subsection, to be not an issue for inductive conversion, however.

C. Transparency of cold parts

A remarkable property of inductive heat engines is that portions of the medium that do not undergo the same temperature variation merely do not make a net contribution to $\langle R_e \rangle$. This differs from the case of mechanical engines somewhat, as illustrated in Fig. 16, which depicts the cylinder of a gas heat engine and the magnetic medium of an inductive engine side by side. The thermodynamic media are shown shaded, to indicate high temperature state, with two cold regions in each case, as shown in dark blue.

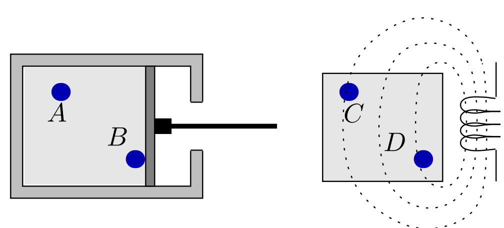

Figure 16. Superposition of inductive conversion

In the gas engine, when the piston encounters cold region B, there would be momentarily loss of pressure on the piston, but no immediate effect from a more distant cold region A, resulting in a "knocking"-like behaviour. In practice, knocking in internal combustion engines is due to explosive detonations of the fuel-air mix rather than cold spots, since the gases would have ample time to attain uniform temperature and pressure at subsonic piston speeds. So the net effect of cold regions would be to lower the overall temperature, reducing the conversion efficiency for the entire medium.

In the inductive engine, all portions of the medium are subject to the H field from the engine current I at the same time. Consequently, cold regions C and D would have immediate, concurrent effect on the throughput, but the engine current profile would not be otherwise impacted. More importantly, if the heating and cooling processes occur within the medium, instead of by conduction from the surface, as assumed in Edison's thermomagnetic engine[11] and in direct conversion studies by Brillouin and Iskenderian,[12] there would be no reason from engine design perspective to provide any means to improve diffusion across the medium, and the slower diffusion of heat in solids would even help to keep the cold regions isolated, so that the temperature in other regions of the medium would not be diminished. Thus, even outside the context of hot particle conversion, inductive engines tend to conserve the peak conversion efficiencies.

In the CMOS scenario, nonswitching CMOS gates would be entirely transparent to the inductive engine cycle. The dipoles located near their drain regions would not result in net loss or gain of energy as they would be paramagnetic. The analysis has also established that knowledge of which CMOS gates switch in a given clock transition is irrelevant to the scheme, and that the nonswitching of some gates can have no impact on the conversion efficiency at the switching gates in the same engine cycle. Analogous reasoning can be applied to the other scenarios, to justify the expectation of the near-unity efficiencies using inductive heat engines. Nonetheless, the very use of a paramagnetic medium adds thermal mass to the hot particles, so there would be some degree of compromise, as discussed next.

D. Coupling through direct impacts and phonons

As explained in subsection B, it is impossible to exploit direct electromagnetic coupling with the hot particles without precise *a priori* knowledge of their locations and velocities. We are therefore generally limited to treating hot particle energies as heat, albeit in an inhomogeneous form, and using thermodynamic conversion. Doing so would have been impossible with a mechanical heat engine or using thermoelectricity, which are both limited by thermal diffusion. We established in subsection C that electromagnetic induction provides an alternative coupling that is not affected by thermal inhomogeneity, and provides the sum of the local contributions without diminishing the local efficiencies to a common bulk temperature. This is necessary both to be able to ignore the nonswitching gates in the CMOS scenario, and for focusing the conversion to hot particles in the unavoidable presence of a bulk cold medium in the first place.

A further problem to be addressed is the general likelihood of a low probability of interaction with the hot particles, for the following reason. Most of the hot particle energies would be in linear, if random, motions. Collision with a hot particle would more be likely to transfer linear than angular momentum to the colliding atom or particle. Flipping an electric dipole moment can be accomplished without angular momentum, but a changing electric field is necessary for capacitive engine operation, which would be problematic in most power recovery or generation scenarios, including CMOS, chemical and nuclear. An inductive engine would be generally preferred, but inductive coupling by a magnetic dipole requires flipping of its orientation, which entails angular momentum transfer.

However, as the very possibility of harnessing hot particle energies and the heat engine principles underlying this possibility are both new, we cannot rule out future solutions to the coupling problem from current knowledge of dipole interactions with hot particles in any of the scenarios envisioned above.

Rotational phonon modes and spin-phonon relaxation, albeit slow, are known in solids, raising the possibility of coupling mediated by the energetic initial phonons resulting from the hot particle interactions within the lattice of a solid medium. In semiconductors, carrier dissipation is indeed known to involve the initial production of longitudinal optical (LO) phonons, which then decay into acoustic longitudinal or transverse (LA/TA) phonons more slowly. Since the terminology of phonon modes as optical or acoustic refers to their energies, the probability of LO-dipole interaction would determine the effectiveness of the hot particle conversion. Fig. 17 summarizes these possibilities graphically.

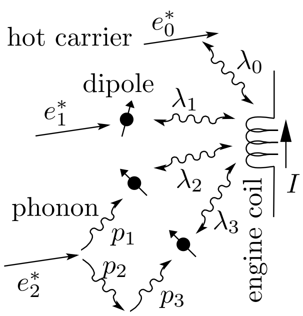

Figure 17. Energy flows

Virtual photon labelled λ_0 represents the direct induction from hot carrier e_0^* to the engine coil. This direct induction path would be generally unusable because it would require *a priori* knowledge of the hot carrier velocities, but in the CMOS scenario, they would be localized to the MOSFET channels and heading towards the drain ends.

Thermodynamic conversion calls for interaction with dipoles that are magnetically cycled. A first hot carrier, marked e_1^* is shown impacting and flipping a dipole, which results in inducing an emf into the engine coil, as represented by virtual photon λ_1. A second hot carrier e_2^* is shown impacting a lattice atom and producing two phonons p_1 and p_2 as a result. The first phonon p_1 then encounters a second dipole, making it flip and thereby inducing an emf and transferring some of its energy to the engine coil, via the induction interaction labelled λ_2.

In CMOS, the initial phonons p_1 and p_2 would be both LO, and cannot carry angular momentum. We would need a secondary TA phonon p_3 from the relaxation of p_2 to flip a (third) dipole and induce energy (λ_3) into the engine coil. Tertiary phonon couplings can also occur.

Clearly, the direct and first order thermodynamic paths, $e_0^* \to \lambda_0$ and $e_1^* \to \lambda_1$ respectively, are the most efficient, allowing the full hot particle momentum to bear on the coil induction. With each degree of indirectness, through one ($p_1 \to \lambda_2$), two ($p_2 \to p_3 \to \lambda_3$), or more phonon stages, we would have additional likely contributions, but at longer relaxation times, requiring slower operation of the inductive engine, and declining energies. The interaction energies correspond to equivalent temperatures of the initial hot particle and of the phonons, respectively, all of which can be measured by nonthermodynamic methods today.

It might be thought, based on the potential energy picture in Fig. 14, that the energy transfer to the coil would be limited to $\mu_B H$, or less than $1/200$ of phonon energies at 300 K, let alone the hot particle temperatures mentioned. In classical studies with bulk magnetic and dielectric engines,[12] as well as in electrical transformers, this potential energy cycle does account for all of the energy throughput, but only because the magnetization and demagnetization processes in the traditional scenarios are driven by thermal phonons which are themselves in equilibrium at all times. The dipole population distribution in Fig. 14 is dictated by the phonon equilibrium and induction into engine load or transformer secondary circuits is a side-effect of the temperature variation on this distribution in the classical engines, causing the inclination of the exponential profile to vary, and of the field variation in the transformers, changing the gap between the energy levels. All of the interactions portrayed in Fig. 17 are, however, dynamical and nonthermal, all hot particle and phonon temperatures mentioned refer to kinetic energies against the backdrop of the thermal phonon equilibrium. Thus, in the $e_1^* \rightarrow \lambda_1$ path, the hot carrier delivers work to both the dipole and the coil, and the dipole acts as a catalyst. Similar reasoning would apply to the indirect phonon coupling paths.

The near unity Carnot efficiency of conversion anticipated in the Introduction clearly correspond only to the direct and at most the first order paths. These would appear to have very low probabilities in current materials, but in material science studies, a macroscopic load circuit impedance matched with the dipoles is generally absent[b], and the presence of the dipole implants itself would also likely change the available phonon modes. Nanoscale structures could be used to enhance phonon couplings, and ways to employ capacitive engines also remain to be explored.

E. Flux distribution and synchronization of heat

All of the preceding subsections assumed that an adequate magnetic field density H (or the electric field density E if using a capacitive engine) can be made available to each participating dipole. A fast changing field is difficult to make uniform over a large volume, but the field variation only needs to be locally synchronized with hot particle production. A more serious limitation concerns the density of the dipole distribution – a high superparamagnetic density can tend to both push out the total flux B, as in the Meissner effect in superconductivity, and potentially obstruct hot particle production, say by blocking light (superparamagnetic ferrofluids, for example, are black liquids).

The related problem, already mentioned, is the pulsing of the hot particle (heat) source, which is necessary for heat engine cycles. While pulsing is readily available in the CMOS energy recovery problem from the synchronous logic, it would have to be additionally implemented other scenarios. For solar power or optical energy recovery, the answer may lie in LCD or electro-optic shutters, trading off continuity of conversion for simplicity. Pulsed combustion is an option with chemical energy, as in internal combustion engines, with engine cycle speeds of $100\ \mathrm{Hz}$ or so. High cycle speeds do not necessarily mean greater efficiency, since the field strength variation would tend to become sinusoidal at higher speeds because of rise and fall times. This would be both less efficient thermodynamically, as explained in Section V, because the diminished field strength during part of demagnetization phase would also reduce the strength of the inductive coupling between the dipoles and the engine coil. A precise estimate of the impact on the hot particle interactions would be difficult without further study.

Pulsing is unlikely to be an option with nuclear power of any kind, including in pacemaker batteries where higher conversion efficiency would be useful, for example, to reduce the radioactive element size for the required power level. One alternative might be to use a moving fluid or solid medium, so that hot particles get created periodically in each region of the medium on exposure to a radiant energy source. Pulsing the energy release producing the hot particles or the phonon channels mediating their dipole interactions using a triggered control mechanism is another possibility.

Conclusion

An electrical circuit analysis of heat engines and thermodynamic conversion has been presented (Section IV) based on dynamical first-principles reasoning for electrical quantities analogous to Carnot's treatment of the gas engine. The correctness of the approach is qualitatively demonstrated by reproducing the characteristic parametric amplification of direct converting dielectric and magnetic engines, and by leading to identical considerations of friction and leakage. (Quantitative considerations have been given in a patent,[9] and have been verified for consistency with prior treatments and empirical results.) The approach yields new analytical insight with some potential for improving the performance of existing heat engines and extending to nonthermal conversions (Sections V, VI).

The approach enables us to take thermodynamic reasoning beyond the traditional realm of macroscopic treatments of heat, to the possibility of capturing the intense nascent energies released even in ordinary combustion, let alone the hot neutrons and nuclear fragments in fission, for thermodynamic conversion at the extreme temperatures represented by these hot particles against the backdrop of the bulk medium (Sections II, III). These nascent energies ordinarily turn quickly to heat in the bulk medium through subpicosecond relaxation processes, meaning that these nascent energies are indeed distinct from the mere high energy end of the equilibrium Maxwellian distribution of the bulk medium, and thus a legitimate high temperature source for operating a heat engine. Detailed reasoning in Section VII establishes that conversion of these nascent energies was hitherto impossible solely because only the molecules coming in contact with the mechanical piston, turbine blade or acoustic wave can perform work in existing heat engines, and that using electromagnetic fields, in place of the mechanical piston, blade or wave, would enable instant, direct performance of work on a macroscopic load by the nascent hot particles. The thermodynamic problem concerns our ignorance of the precise times and locations of occurrence of the hot particles in a bulk medium and of their speeds and directions of motion. The inductive and capacitive engine formalisms resulting in the electrical approach not only enable this use of

[b]Detection by RF coils in NMR is consistent with inductive work, although the method of excitation is different and nonthermodynamic.

electromagnetic fields, but lead to a precise thermodynamic solution via an ensemble of magnetic or dielectric dipoles to mediate and catalyze the performance of work on the electrical load by the nascent hot particles independently of their location and velocity distributions. Such an engine would be suitable for use as a “front-end” convertor in most power plants as all of the nascent hot particle energies that escape conversion would be captured into the bulk medium as heat anyway by the relaxation processes, and would thus remain suitable for conventional power generation.

Detailed consideration of the interactions between the hot particles, dipoles and the engine load circuit, including zero, one or more orders of phonon interactions, have been given in order to bring out the likely problems and practical limitations that may arise. Given the novelty of the electrical approach and more so of the possibility of hot particle conversion, it is likely that additional issues will surface in the course of validation and testing of these ideas. However, the insight and clarity of reasoning that have been achieved make it seem fair to conclude that the basic thermodynamic issues have been addressed, and that the additional issues will be mostly likely of engineering and material science.

References

[1] D J Allstot, G Liang, and H C Yang. Current-mode logic techniques for CMOS mixed-mode ASICs. In *Proc IEEE Integ Ckts Conf*, volume 49, pages 25.2/1–25.2/4, May 1991.

[2] A H Ismail and M I Elmasry. A low power design approach for MOS current mode logic. In *Proc IEEE Intl SOC Conf*, pages 143–146, Sep 2003.

[3] S Khabiri and M Shams. Implementation of MCML universal logic gate for 10 GHz-range in 0.13 μm CMOS technology. In *Proc IEEE Intl Symp Circuits and Systems*, volume 2, pages 654–656, May 2004.

[4] W Athas et al. Ac1 - a clock-powered microprocessor. *IEEE Proc Low Power Electronics*, 1997.

[5] L Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Zeitschrift für Physik*, 53:840–856, 1929.

[6] C H Bennett. Demons, Engines and the Second Law. *Sc Am*, Nov 1987.

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

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

[9] V Guruprasad. Electromagnetic heat engines and method for cooling a system with predictable bursts of heat. *US Patent*, 5714829, 1998.

[10] Magnetic heat engines. Feb 1998. `http://physicsweb.org/articles/news/2/6/21`.

[11] T A Edison. Pyromagnetic generator. *US Patent*, (476983), Jun 1892.

[12] L Brillouin and H P Iskenderian. *Elect Comm*, 25(300), 1948.

[13] R Rosensweig. *Ferrohydrodynamics*. 1965.

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

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