Low-Energy Computing for Implantable Medical Devices

Paper to Accompany MEDG Seminar of February 21, 1996

Michael P. Frank
http://www.ai.mit.edu/~mpf
MIT AI Lab
545 Technology Sq.
Cambridge, MA 02139

February 20, 1996

1 Description of Talk

As experts in the physics of computation have known for some time, there is no lower bound on the amount of energy dissipation needed to perform computation; in principle, the energy can be made as small as desired [2, 3, 6, 7]. Unfortunately, in traditional computers, CPUs dissipate a substantial amount of energy per computational operation, limiting the effectiveness of computers in severely energy-limited (e.g., battery-powered) contexts.

However, recent developments in digital circuit design have shown that it is possible to create so-called "adiabatic" digital circuits in which the energy per operation can actually be made as low as desired [5, 11, 12, 13, 10]. The main drawback is that individual operations must be performed more slowly. Our group is creating a general-purpose CPU chip based on this technology [9]. What would the applications of such a device be?

One context in which low energy consumption might be particularly desirable is the case of a battery-powered device to be implanted long-term in the human body, since we might imagine that a need for frequent recharging or replacement would be very undesirable to the patient. With the new technology, an implant with a tiny battery could perform essentially as much computation as desired without needing a fresh input of energy. But what sorts of implants might need to do large amounts of computation?

In this talk, I will briefly outline the technological tradeoffs afforded by the new technology, but then I would like to spend most of the time in an open discussion with the audience about what some possible uses of the technology in medical implants might be. I would like audience members to try to think up some possible applications before coming to the talk. To get you started, here are some initial ideas, some of which are pretty far-out. I hope you can help me come up with some more ideas that are closer to reality.

1. Monitoring device to frequently measure e.g. blood-sugar levels and compute & store long-term statistics & patterns, for later communication (via radio or removal).

2. Device as above, but also with a decision-making and planning component to decide, e.g., when to release a stored drug.

3. Artificial retina that performs needed preprocessing of input data before sending to optic nerve. (Run on light energy, without battery?)

4. Other neural-interface implants to replace the computational role of damaged ganglia, etc.

Keep in mind that although we can perform computations with arbitarily little energy, this says nothing about the energy requirements for sensor/effector operations. Therefore, applications that require lots of computation relative to the amount of interaction with their environment are most convincing as roles for the new technology.

I look forward to an interesting discussion.

2 Reversible Adiabatic Circuits

This section outlines some of the principles of operation of the circuit technology upon which our low-energy computer project is based. More details can be found in the papers by Younis and Knight [11, 12, 13] and in Saed Younis's Ph.D. thesis [10].

An *adiabatic* electronic circuit operates dyanamically without exchanging energy with its surroundings. Although perfect adiabaticity is impossible, we can approach it as closely as desired by performing changes *quasistatically*, meaning in a controlled, gradual way where at any moment the system is close to equilibrium.

One of the main restrictions that quasistatic operation places on circuit design is that a transistor must never be switched on while there is a voltage across it. Otherwise, there would be a sudden, non-quasistatic flow of current across it. That sort of indiscriminate switching is the norm in today's digital circuits and is the main source of energy dissipation in modern CPU chips. The dissipation is substantial: the newest processors dissipate tens of watts of power and require substantial cooling to operate properly.

The requirement that transistors never be switched when there is a voltage across them leads to an interesting restriction on the logical operation of the circuit: all operations performed must be *logically reversible* in other words, it must always be possible to reconstruct each operation's inputs from its outputs. Information about the past state of the computation can never be thrown away. Briefly, this is because we need to reconstruct the input so that we can use it to reset input lines, by drawing off charge from them through a transistor without any sudden switching. (If we did not reset the state of wires in the circuit after using them, then each part of the circuit could only be used once.)

Fortunately, the restriction of logical reversibility is not a show-stopper. As was pointed out by Bennett [1], any operation can be made trivially reversible by letting the output include a copy of the input. This naive solution causes the memory requirements of a computation to increase linearly with time, in order to store the inputs of all operations performed so far. But the important point is that arbitary general computations are indeed possible.

Bennett also showed that we can more clever, and avoid the monotonic increase in memory use, by breaking a computation into parts. A part of the computation is performed, the results are copied, and then the part is performed in reverse (which is of course possible since each part is reversible), basically re-absorbing all the intermediate results back into the part's original input. Then we repeat this with the next part of the computation. Only the input to each part needs to be stored, rather than all the intermediate results.

Techniques like this can be used at a very low level, so that the traditional basic arithmetic and logical operations of a computer can be implemented by circuits that are not very much more complex than their traditional irreversible counterparts. For example, as part of Younis's work [10] he implemented a multiplication circuit, and demonstrated its energy efficiency.

Bennett also showed [4] (see also [8]) how to use his technique in a sort of recursive way to limit the total memory usage for a reversible implementation of an arbitary computation (whose traditional memory usage is bounded) to only a logarithmic function of running time. So the memory requirements for long-term complete reversibility are really not a serious problem.

In the above outline, we have left out many details. But the important point for our present discussion is that there appear to be no barriers to the construction of a general-purpose CPU chip capable of performing a given computation using arbitarily small amounts of energy. We have a more or less complete architecture for such a processor, and we will fabricate it as soon as we have fleshed out all the low-level details.

3 Technological Tradeoffs

Of course, obtaining an energy savings from these chips demands we trade off energy savings against a number of other variables, of which we must be aware when considering possible applications.

Circuit area. Reversible circuits require an amount of chip area that is larger than the area for a traditional circuit based on the same fabrication technology, by a small constant factor. However, fabrication technology is approaching the point where the density of devices on a chip will be more and more limited by the increased difficulty of cooling the current breed of energy-inefficient circuits, especially as more and more layers of circuitry are added to chips. So at some point, higher circuit densities will probably be possible with reversible circuits than with irreversible circuits anyway.

Clock Speed. An adiabatic circuit that is run at about the same clock speed as a traditional circuit, that is, at the maximum speed that allows sufficient time for signals to propagate from one device to the next, will dissipate *roughly* the same amount of power. As the speed is lowered, energy usage decreases quadratically, so that, for example, a chip running at 1/10th normal speed would use 1/100th the power, or 1/10th the energy per operation.

The speed issue may seem like a major drawback, but it is important to keep in mind that the cooling problem is becoming increasingly important as a limit on the speed of traditional irreversible processors. Once cooling becomes the primary limiting factor, traditional circuits will be unable to operate at the maximum signal propagation speed, and so adiabatic circuits should be able to run faster than the traditional circuits while dissipating less power.

One advantageous feature of the speed-energy tradeoff is that a finite supply of energy can support an unbounded amount of computational work, since we can simply slow down the chip as the energy runs out, and never use more than a certain fraction of the remaining energy for each operation—so that the energy never reaches zero no matter how many operations we do. With a traditional circuit, this mode of operation is impossible because switching a transistor that has a voltage across it will cause a sudden dissipation when the input crosses the threshold, no matter how slowly it does so.

Memory use. As mentioned in the previous section, reversible operation in general requires that extra memory be used to store extra information needed to keep the computation reversible. If we want to continue computing indefinitely, memory needs will in the worst case increase logarithmically with time using Bennett's algorithm [4]. However, this algorithm has its own costs in terms of the number of cycles needed to perform a computation. Namely, it requires polynomially more cycles than the corresponding irreversible computation. The exponent on the polynomial can be made arbitarily close to 1, but only at the cost of an exponentially-larger constant factor on the memory use. This at first appears problematic.

However, our real goal (for practical purposes) is arbitarily low energy usage, rather than perfect reversibility, and this can be achieved without a polynomial run-time cost. Bennett's algorithm is essentially a way to compress the history of a computation by any desired factor, with a constant factor effect on run-time for a given compression factor. In achieving logarithmic memory use over time, the algorithm is essentially compressing the history by greater and greater factors over time, so that the run-time cost is no longer a constant factor.

To achieve a linear run-time cost, we equip our computer with enough memory to execute Bennett's algorithm for an amount of time T that achieves some desired compression factor. After executing for that amount of time, we erase the stored history information and reset the algorithm. There is an energy cost E associated with erasing that information, but the energy per operation E/T can be made arbitrarily small, by making the available memory larger, which increases T exponentially. Using this method, we can run indefinitely without running out of memory (assuming as always that the underlying program that we are running doesn't itself have indefinitely-increasing memory needs).

However, note that making the memory larger to decrease energy per operation is not something we can do at run-time. Therefore, for any particular computer designed in this way, it will have some definite lower bound on energy per operation, and so will not be able to execute an unbounded number of operations on a bounded energy supply. However, for any given energy supply, the bound on the number of operations can be made as large as desired at design time, by giving the machine a memory that is of logarithmic size in the desired number of operations.

Summary of Tradeoffs If built today, a reversible computer chip would be larger and slower than its reversible counterpart, but as the underlying fabrication technology improves, and cooling becomes a more limiting concern, we will get to the point where the reversible processors will be able to be both smaller and faster than their irreversible counterparts.

An aside: This is not to say that reversible computing will enevitably win out for computing in general. In contexts where total energy usage of the system is not a concern, it may turn out that intensive active cooling of an irreversible circuit running at top propagation-delay speed allows for faster operation than running an irreversible circuit at less than top speed. But the coolant-based system may be less economical. In any case, in this paper we are concerned with applications where energy is limited, and active cooling is not an option; I believe in those contexts, reversible computing will be very important.

Another major tradeoff for reversible computers is memory use. However, variations of Bennett's algorithm may be used to limit memory needs to a reasonable level, with a constant-factor slowdown, while still making the energy per operation as small as desired.

4 General Applications

Currently, adiabatic reversible computing is a case of a technology looking for an application. The ability to run extremely dense future generations of circuits at reasonable speeds without cooling is one possible application, but we are not yet certain that cooling an irreversible circuit instead would be less economical, in cases where the energy is available for doing active cooling.

Therefore, a more plausible sort of application is one where energy is very limited, but lots of computation is needed. Here is a short list of general areas where this might be the case at some point:

1. Automated deep-space probes travelling far from the Sun (therefore no solar power) for long periods.
2. Personal portable computers. These currently also have major energy needs for displays, disks, and peripherals, so that a low-energy CPU would not drastically lengthen battery life. However, this situation could change.
3. Computers in scientific data-gathering devices located undersea or underground where solar power is not available and transmitted power is inconvenient.
4. Medical implants within the human body. (Subject of the rest of the talk.)

For all of these application areas, it is very important to keep in mind that although we can do arbitary amounts of munging of data inside a computer with as little energy usage as we want, the process of inputting data from sensors, our outputting data to effectors, may not be amenable to such control. In fact, thermodynamics demands that we dissipate at least $kT \ln 2$ energy for each bit of information that we input beyond what we are able to store internally (since those bits must be erased). Output also requires a similar minimum energy loss, unless we output only to systems which are themselves reversible and capable of returning the output energy to us. Moreover, typical sensors and effectors will probably dissipate far more energy than the kT limit, due to engineering inefficiencies beyond the control of us computer scientists.

Therefore, in thinking about low-energy applications, it is helpful to focus on cases where large amounts of computation must be performed in between external operations. For example, a deep-space probe may need to run a long, complex planning program to figure out the optimum action (e.g., navigational maneuver) to perform given a very limited amounts of input data (e.g., positions of a few reference points). In interplanetary space far from the Sun, we have plenty of time to do these computations, but not much energy. This is a promising area of application for reversible computing.

The point of this talk is to get medical technologists thinking about what promising applications there might be for low-energy computing in medicine. The obvious realm of application in medicine is in implanted medical devices, since presumably wires going outside the body to a large power supply would be undesirable, and therefore the preferred mode of power would be a small battery. If there is some class of plausible implanted device that would need to perform large amounts of computation between external operations not requiring much energy, then that would constitute a promising candidate application for our first reversible CPU's, which should exist in a few years.

5 Implantable Devices—Initial Thoughts

Here we will explore some very preliminary ideas for what sorts of implantable devices might benefit from low-energy computing technology.

We will divide our study of possible device functions into two general classes: control and data-gathering. Control is when the result of computation is used directly to take some action that affects the body in some way. Data-gathering is when the purpose of computation is solely to preprocess data gathered from sensors for more efficient long-term storage. The stored data may be later communicated to the outside world, for use in diagnosis or in managing a program of treatment. Of course, the two functions may be combined; some of the unit's computation may be used to preprocess input data for storage and later transmission, and other computations may mull over the stored data in the process of deciding upon an action to perform at some later time. But a conceptual separation of data-gathering from control seems a useful way to break down the issues.

5.1 Data-Gathering Issues

As stated in the previous section, the actual inputting of data has an intrinsic energy cost, connected with the fact that unless we intend to store every single bit of data we gather, we will dissipate energy irretrievably every time we get a new piece of input, in order to throw the previous piece away. At minimum, kt ln 2 per bit, but practically speaking, much more. However, if the amount of computation performed per bit of input is large, than reversible energy-saving technology will have a large impact even if most input bits are thrown away, because we will be saving the energy needed to store all of the many bits of intermediate results involved in the computation.

A larger concern is the energy usage of the sensors that actually gather the input bits. This will of course vary among applications. Sometimes, the physical entity that the sensor is measuring may actually be able to provide energy to power our device's measurement of that entity. For example, if the implant is an artificial retina, then the measured light could perhaps provide the energy to power its own measurement, for example via the photoelectric effect. If there is extra energy left over, this could conceivably power the computational operation of our device, and we may not need any long-term energy storage at all. (Of course, this is just a thought, not a design!) Other possible energy-providing sensors might include sound sensors, motion sensors, and perhaps some types of chemical sensors.

As for the purpose of data-gathering, if the device has no control component, then data-gathering can only be for the purpose of transmission to the outside world. If raw data is just stored verbatim until memory fills up, there is no need for computation. But raw data is almost always not what you finally want; generally you want some pre-processed summary of the data. This summary could be computed within the implanted device, and stored for later retrieval.

How are results from the device's measurements to be transmitted to the outside world? Earlier I mentioned that wires leading to outside the body might be considered undesirable; otherwise we would not really need a low-energy computer because those same wires could be used to recharge power frequently. However, having to do frequent rechargings, or carry around a large external battery, might be considered more undesirable than the wires themselves. A wire only used for occasional downloading of data might have only an unobtrusive jack visible on the surface, and the patient might not notice it at all between the occasional visits with his doctor or sessions with his Guardian Angel (tm) in which the data is downloaded from the implant.

During those same occasional download sessions, we could also take the opportunity to recharge the implant's internal battery until the next session, so that the device could continue functioning indefinitely, with no effect on the life of patient other than the occasional download/recharging sessions. We can make these redundant from the point of view of power needs, and therefore the sessions will only be needed as often as data is actually needed for diagnosis or treatment.

Another possibility is to do both the data transmission and power-recharging via RF signals. Instead of the wire, we have an internal antenna; a wire that that does not come to the skin surface, and can be just threaded along some convenient space in between tissues when the device is first implanted. During the recharge/download sessions, an external device could transmit RF power which the implant uses to simultaneously recharge its battery and transmit its data. This would be the ideal from the point of view of patient comfort; it might be completely unnoticeable by the patient. If the device were sufficiently non-toxic, it could perhaps be left in place even after it is no longer needed, thus cutting in half the number of implant-related surgical procedures that are required.

5.2 Control Issues

Control is of course a major concern from the point of view of power requirements. It is hard to see how any major mechanical effects on the body could be performed without significant power consumption. One exception is perhaps pacemaker devices implanted near the heart, which might actually be able to recharge themselves using the electrical or mechanical energy of the beating of the heart, and use some of their stored energy to send small pacemaking pulses.

However, there a couple of other obvious classes of effects that might be performed with very little power:

1. Releasing very tiny amounts of a powerful drug from an internal supply in the implant.
2. Transmitting small electrical signals to individual neurons.

The first item is of course only practical if the drug in question is powerful enough in small enough quantities, or needed infrequently enough, so that the implant can store enough to last the patient for a long time. Obviously, if frequent procedures were needed to replace the supply of drug, those same procedures could also be used to recharge battery power, obviating the need for low-power computing. One possibility is that the drug in the implant could be reserved for use only in emergencies, when the patient forgets or is for some reason unable to take his normal oral or injected dose of the drug.

The second item, of course, requires a lot of improvement of technologies for connecting computer chips to neurons. I don't know if there is any technology capable of this yet. I also don't know the magnitude of the electrical pulses required to stimulate a neuron; however, I suspect it is not enormously larger than the energy involved in switching today's transistors. And of course a major problem for any sophisticated nervous-system interaction is knowing which neurons to connect to, and what signals to send. These issues I know nothing about, and leave to our friends over in BCS to ponder.

Control applications might turn out to be even better than data-gathering applications as candidates for low-power computing. If the purpose is the careful timing of a powerful drug, we might easily imagine that lots of computation would be required to carefully choose the optimum time for delivery of the drug based on scanty information garnered from sensors. If the purpose is neural interactions, then clearly we might want to do something pretty sophisticated to send the right pattern of signals to have the desired effect. But of course, any time complex software is involved, you have to be very, very careful about putting it in control!! We wouldn't want an obscure bug in the program to suddenly pop up, and poison the patient, or give him a seizure. Therefore we will have to be extremely confident about the algorithms used in the control implants before they can ever be safely used in actual trials. The opportunities for lawsuits are tremendous! This is not unusual for medical technology. But the programs for control implants will be especially scary to write.

6 Discussion Topic: Particular Implant Applications

At this point I would like to solicit open discussion about particular medical tasks that could be appropriately performed by implants equipped with low-energy CPU's. I don't know much about medicine, so I am hoping the audience can help come up with some realistic scenarios for how this technology might be used. Here are some initial ideas; I don't really know enough yet to evaluate them properly.

1. An implant for diabetics that automatically measures bloor sugar and insulin levels several times a day, and compiles statistics (such as an average daily profile) on this data which it stores for later radio transmission.

2. In a riskier application, the above implant could also be programmed to deliver an emergency supply of insulin if blood sugar levels fluctuate by a dangerous amount.

3. Another risky application is a smart pacemaker device, which could possibly get its energy from the heart itself. The ability to do lots of computation might enable some sophistication in the device's choice of heart rate, which could be varied according to some relevant factor in the blood (e.g., CO2 levels).

4. A very futuristic application might be the artificial retina mentioned earlier, that perhaps obtains its energy from the light itself. This would require more sophistication in connecting circuits to neurons than is currently available (as far as I know).

5. A wide range of other neural-implant applications is conceivable, such as replacing damaged nerves or ganglia, if enough about their function was known so that we could imitate it. Brain implants could perhaps suppress seizures in epileptics. In the more distant future, implants could perhaps help replace brain regions damaged by stroke or tumor, although recovering any lost memories might be impossible. (Anybody read Minsky's "The Turing Option?")

Again, these suggestions are from someone who is naive about medicine. It is hoped that the audience will be able to revise these suggestions and help come up with some much more realistic ones.

7 Conclusion

In a few years, reversible adiabatic computing technology will likely provide us with the ability to compute with essentially as little energy as desired. Small batteries will no longer limit the amount of computation that a self-contained device can perform.

This advancement in fundamental technology will enable us to consider new applications for computer technology, applications that were previously thought to be unworkable due to the inconveniences caused by power requirements.

One promising area of application is the creation of smart implantable medical devices that contain powerful on-board computers but require only a tiny battery to operate for long periods with no energy input. Such devices might gather data, release drugs when needed, or assist damaged nerves or organs. What are some diseases for which a computationally powerful implant might be particularly useful? What other medical applications might there be? These are the questions that we will hopefully begin to answer in today's talk.

References

[1] C. H. Bennett. Logical reversibility of computation. *IBM Journal of Research and Development*, 6:525–532, 1973.

[2] C. H. Bennett. The thermodynamics of computation, a review. *International Journal of Theoretical Physics*, 21(12):905–940, 1982.

[3] C. H. Bennett. Notes on the history of reversible computation. *IBM Journal of Research and Development*, 32(1):281–288, 1988.

[4] C. H. Bennett. Time/space trade-offs for reversible computation. *Society for Industrial and Applied Mathematics Journal on Computing*, 18(4):766–776, 1989.

[5] J. G. Koller and W. C. Athas. Adiabatic switching, low energy computing, and the physics of storing and erasing information. In *Physics of Computation Workshop*, 1992.

[6] R. Landauer. Uncertainty principle and minimal energy dissipation in the computer. *International Journal of Theoretical Physics*, 21(3/4):283–297, 1982.

[7] R. Landauer. Computation: A fundamental physical view. *Found. Physics*, 16:260–266, 1986.

[8] Robert Y. Levine and Alan T. Sherman. A note on bennett's time-space tradeoff for reversible computation. *Society for Industrial and Applied Mathematics Journal on Computing*, 19(4):673–677, 1990.

[9] Carlin J. Vieri. Pendulum: A reversible computer architecture. Master's thesis, MIT Artificial Intelligence Laboratory, 1995.

[10] S. G. Younis. *Asymptotically Zero Energy Computing Using Split-Level Charge Recovery Logic.* PhD thesis, MIT Artificial Intelligence Laboratory, 1994.

[11] S. G. Younis and T. F. Knight, Jr. Practical implementation of charge recovering asymptotically zero power CMOS. In *Proceedings of the 1993 Symposium in Integrated Systems*, pages 234–250. MIT Press, 1993.

[12] S. G. Younis and T. F. Knight, Jr. Asymptotically zero energy split-level charge recovery logic. In *International Workshop on Low Power Design*, pages 177–182, 1994.

[13] S. G. Younis and T. F. Knight, Jr. Harmonic resonant rail drivers for adiabatic logic. In *Proceedings of the 1995 Symposium on Advanced Research in VLSI.* MIT Press, 1995.