

The Power of Heterogeneity in Near-Threshold Computing

Neal Barcelo, Michael Nugent, Kirk Pruhs
University of Pittsburgh
E-mail: {ncb30, mnugent, kirk}@cs.pitt.edu

Michele Scquizzato
University of Houston
E-mail: michele@cs.uh.edu

Abstract—One potential method to attain more energy-efficient circuits with the current technology is Near-Threshold Computing. However, this energy savings comes at a cost of increased functional failure, which necessitates that circuits must be more fault-tolerant, and thus contain more gates. Thus, achieving energy savings with Near-Threshold Computing involves properly balancing the energy used per gate with the number of gates used.

We consider both the setting where the supply voltages must be homogeneous and the setting where they may be heterogeneous. We show that, for small circuit error bounds, there are many natural functions that can be computed with a log factor less energy by heterogeneous circuits than is possible with homogeneous circuits, and that this result is tight for many functions. In contrast, we show that there are relations that can be computed with a log-squared factor less energy by heterogeneous circuits than is possible with homogeneous circuits.

I. INTRODUCTION

The threshold voltage of a transistor is the minimum supply voltage at which the transistor starts to conduct current. However, if the designed supply voltage was exactly the ideal threshold voltage, some transistors would likely fail to operate as designed due to manufacturing and environmental variations. In the traditional approach to circuit design the supply voltages for each transistor/gate are set sufficiently high so that with sufficiently high probability no transistor fails, and thus the designed circuits need not be fault-tolerant. One potential method to attain more energy-efficient circuits is *Near-Threshold Computing*, which simply means that the supply voltages are designed to be closer to the threshold voltage. As the power used by a transistor/gate is roughly proportional to the square of the supply voltage [1], Near-Threshold Computing can potentially significantly decrease the energy used per gate. However, this energy savings comes at a cost of a greater probability of functional failure, which necessitates that the circuits must be more fault-tolerant, and thus contain more gates. As the total energy used by a circuit is approximately the energy used per gate times the number of gates, achieving energy savings with Near-Threshold Computing involves properly balancing the energy used per gate with the number of gates used.

Since the relationship between voltage and the log of the failure is approximately linear (see Figure 7 in [2]), we have

K. Pruhs was supported, in part, by NSF grants CCF-1115575, CNS-1253218, CCF-1421508, and an IBM Faculty Award.

that the error as a function of supply voltage v is approximately of the form of $\epsilon(v) = c^{-v}$, for some positive constant c . Using the fact that the energy is proportional to the square of the supply voltage [1], we conclude that the failure-to-energy function for a 65nm SRAM cell is approximately $E(\epsilon) = \Theta(\log^2(1/\epsilon))$. We thus initially adopt this model of the relationship between energy and error. (However, results in this paper can be generalized to a broader class of relationships.)

While it may not currently be practical, in principle the supply voltages need not be homogeneous over all gates, that is, different gates could be supplied with different voltages. Intuitively, heterogeneous voltages should benefit a circuit where certain parts of the computation are more sensitive to failure than others. For example, in order for a circuit to be highly reliable, gates near the output need to be highly reliable. However, it may be acceptable for gates that are far from the output to be less reliable if there is sufficient redundancy in the circuit.

This naturally leads to the question of whether there is any limit to the energy savings possible by allowing heterogeneous supply voltages. Before stating our contributions toward answering this question, it is useful to first consider the related literature.

A. Related Work

The study of the design of fault-tolerant circuits using noisy gates (that is, gates that fail with some known, fixed probability) was initiated with the seminal paper by von Neumann [3]. The general idea of trading accuracy of a hardware circuit and computing architecture for energy savings dates back to at least [4]. A theoretical study of Near-Threshold Computing was initiated in [5]. [5] gave some general upper and lower bounds on the energy required to compute general functions, showed that there are some functions where allowing heterogeneous supply voltages does not give significantly lower energy circuits, and showed that there are some circuits where allowing heterogeneous supply voltages can give significantly lower energy circuits.

Following up on this work, [6] showed that the traditional approach, cranking up the supply voltage sufficiently high so that each of the s gates fails with probability at most δ/s , has approximation ratio $\Theta(\log^2 s)$. The main result of

[6] was that it is NP-hard to achieve a significantly better approximation ratio than the traditional approach to this problem. [7] considered a model where the (heterogeneous) supply voltages specify exactly the failure rate of a gate, rather than just an upper bound on the failure rate. In this model, [7] showed that a single circuit may be able to compute an exponential number of functions. Despite this, [7] showed that with high probability, the optimal energy circuit for a function chosen uniformly at random requires exponential energy (i.e., $\Omega(2^n/n)$ energy).

B. Our Contributions

We reconsider the question how much energy savings is possible by allowing heterogeneous supply voltages under the assumption that one requires that the circuit error bound approaches zero as the size of the input grows (as opposed to being constant, as was assumed in [5]). It is common in situations where randomization is involved to desire/require such “high confidence” bounds. In particular, we assume that the circuit error bound is $\delta = 1/n^c$ for some constant $c > 0$, which is the most common type of high confidence bound. One practical motivation for desiring high confidence bounds is when the function/circuit considered is a component in a larger computation/circuit, and the final error bound is a sum/combination of the error bounds of the various components. In such a setting one might reasonably desire that the error bounds of the components approach zero. Somewhat surprisingly to us, when high confidence error bounds are required, we can resolve positively this open question from [5], for both functions and relations. That is, we show that there are functions and relations that benefit from heterogeneity.

Specifically, in Section II we show that for any Boolean function that is non-degenerate (in the sense that for every input bit affects the output for at least one input), and that can be computed by a circuit with $O(n)$ faultless gates, there is a heterogeneous circuit that consumes $\Omega(\log n)$ less energy than the minimum energy homogeneous circuit. In Section III we show that this log energy savings is tight, as for any such function, a factor of $O(\log n)$ energy savings is the most that can be gained from heterogeneity. Switching to relations, we show in Section IV that for a particular super-majority relation, there is a heterogeneous circuit that uses a factor $\Omega(\log^2 n)$ less energy than any homogeneous circuit. Finally, in Section V we show that, for every relation that has a linear sized circuit, every heterogeneous circuit uses at most $O(\log^2 n)$ less energy than the optimal homogeneous circuit.

Before presenting our formal model we provide some intuition for the high level proof techniques. Turning to the first result, we begin by using the fact that no circuit’s output gate can have failure rate greater than the error bound, while computing correctly with probability at least $1 - \delta$, to lower bound the energy used by any homogeneous circuit. To build our circuit that requires $\Omega(\log n)$ less energy than this lower bound, we borrow a gadget from [8] that allows us to replace each gate in a faultless circuit with $O(\log n)$ gates that fail with constant probability, while ensuring that a constant fraction of

these will compute correctly. Combining this with a majority gadget, also of size $O(\log n)$, placed before the output of the circuit, allows us to recover the correct output with sufficiently high probability.

The second result, showing that this energy savings is tight, follows by extending the techniques from the general energy lower bound of [5]. The idea is first to map a circuit to an equivalent model (in terms of energy and failure probability) where failures occur both at gates and on wires. We then consider the event that for each input bit, all wires emanating from this bit fail, and show that in such a case we can bound the probability that circuit computes incorrectly. We in turn use this to lower bound the total energy used by wires emanating from each input bit, yielding the desired energy lower bound.

Turning to relations, we consider the majority relation which returns a 0 if the input contains at least 75% 0’s, a 1 if the input contains at least 75% 1’s, and either 0 or 1 otherwise. The standard circuit to compute such a relation is a tree of full adders. The approach is similar to our first result in that we add increasing redundancy so that failures become increasingly rare, and near the output, we take the majority using low probabilities of failure. The main complexity comes in accounting for the fact that full adders consist of multiple output bits.

The final result follows by observing that circuits of linear size require $\Omega(n)$ energy, and that, for a relation that can be computed with $O(n)$ faultless gates, setting the voltage such that each gate uses $O(\log^2 n)$ energy is sufficient to ensure that with high enough probability no gate fails, resulting in a homogeneous circuit that uses $O(n \log^2 n)$ energy.

In summary, we essentially show that high confidence computation of functions can be done with logarithmically less energy if heterogeneous supply voltages are allowed (and this is the best possible), and that high confidence computation of relations can be done with log-squared less energy if heterogeneous supply voltages are allowed (and this is best possible).

C. Formal Model

A *Boolean relation* h is a map from $\{0, 1\}^n$ to $\{0, 1\}$, where each input is mapped to 0, 1, or both 0 and 1. If $x \in \{0, 1\}^n$ is mapped to both 0 and 1, this can be thought of as “don’t care” (for example because the input x should not occur in a correctly functioning system). A *Boolean function* f is a Boolean relation where each input is uniquely mapped to either 0 or 1. For any input $x \in \{0, 1\}^n$, denote by x^ℓ the input that has the same bits as x , except for the ℓ -th bit, which is flipped.

A *gate* is a function $g : \{0, 1\}^{n_g} \rightarrow \{0, 1\}$, where n_g is the number of inputs (i.e., the *fan-in*) of the gate. We assume that the maximum fan-in is at most a constant. A *Boolean circuit* C with n inputs is a directed acyclic graph in which each of the n input nodes (i.e., those with no incoming edges) outputs one of the input bits, and where every other node is a gate. The *size* of a circuit, denoted by s , is the number of gates it contains. For any $I \in \{0, 1\}^n$, $C(I)$ denotes the output of the Boolean function computed by Boolean circuit layout C .

In this paper we consider circuits (C, \bar{v}) that consist of both a traditional circuit layout C as well as a vector of supply voltages \bar{v} , one for each gate of C . Every gate g is supplied with a voltage v_g . We say that the supply voltages are *homogeneous* when every gate of the circuit is supplied with the same voltage, and *heterogeneous* otherwise. A circuit is homogeneous when its supply voltages are homogeneous, and heterogeneous otherwise. We say that a gate *fails* when it produces an incorrect output, that is, when given an input x it produces an output other than $g(x)$.

Each non-input gate g fails independently with probability at most $\epsilon(v_g)$, where $\epsilon : \mathbb{R}^+ \rightarrow (0, 1/2)$ is a decreasing function. The voltage supplied to a gate determines both its energy usage and its failure probability, thus we define $\epsilon_g := \epsilon(v_g)$ and drop all future formal reference to supply voltages. Finally we assume there is a decreasing, nonnegative failure-to-energy function $E(\epsilon)$ that maps the failure probability ϵ to the energy used by a gate. Throughout the paper we assume that $E(\epsilon) = \Theta(\log^2(1/\epsilon))$ and that $\lim_{\epsilon \rightarrow 1/2^-} E(\epsilon) > 0$. In the full version of the paper we shall discuss how to generalize our results to other failure-to-energy functions.

A gate that never fails is said to be *faultless*. Given a value $\delta \in (0, 1/2)$ (δ may not be constant), a circuit (C, \bar{v}) that computes a Boolean relation h is said to be $(1 - \delta)$ -*reliable* if for every input I on which $h(I)$ is not both 0 and 1, $C(I)$ equals $h(I)$ with probability at least $1 - \delta$. The *minimal circuit size* for a relation h is the minimum number of faultless gates required by any circuit computing h .

II. OBTAINING LOGARITHMIC ENERGY SAVINGS FOR FUNCTIONS

We show here that, for a wide class of natural functions, allowing heterogeneous supply voltages provides a logarithmic savings in energy. In particular, we show that, when δ is a polynomial function of the minimum circuit size s , it is possible to obtain an $\Omega(\log s)$ energy savings using heterogeneous supply voltages. The result is that many natural Boolean functions can be computed with asymptotically less energy using heterogeneous circuits. Formally, we have the following theorem.

Theorem 1. *For any function f with minimum circuit size s , for any constant $c > 0$, if $\delta = 1/s^c$, the optimal homogeneous circuit requires energy $\Omega(s \log^2 s)$ energy, and the optimal heterogeneous circuit uses $O(s \log s)$ energy.*

Proof. We first provide a lower bound on the energy used by any homogeneous circuit that $(1 - \delta)$ -reliably computes f . Since, by assumption, s gates are required when there are no failures, and because the circuit is homogeneous, gates (in particular, the output gate) can fail with probability at most $1/s^c$. Since it must be that $\epsilon < \delta$, and $E(1/s^c) = \Theta(\log^2 s)$, we have that $\Omega(s \log^2 s)$ energy is required.

The upper bound requires significantly more work, although is still a somewhat straightforward use of techniques from [8], which proves the following as part of the proof of the general

fault-tolerant upper bound in [8] (which is re-stated in this paper as Theorem 6):

Lemma 2 ([8]). *Let the maximum fan-in of any gate be a constant. There is a constant $\epsilon_1 > 0$ and $\theta > 1/2$ such that for any $\epsilon \leq \epsilon_1$, there is a $\rho = \rho(\epsilon) < 1$ such that any gate g of fan-in ℓ can be replaced by a gadget with*

- 1) k input wires for each input to g ,
- 2) k output wires, and
- 3) $\Theta(k)$ gates,

with the property that if, for all i , at least a θ fraction of the i -th set of input wires carries bit b_i , then the probability that fewer than a θ fraction of the output wires carries $g(b_1, \dots, b_\ell)$ is at most ρ^k .

In a manner similar to the proof of the general fault-tolerant upper bound in [8], we use Lemma 2 to replace each gate in the original circuit with a gadget whose input and output is $\Theta(\log s)$ wires, and set the failure probability of this section of the circuit to ϵ_1 , with the result that the probability that less than a θ fraction of the wires carry the correct output (i.e., the output if there were no failures) is at most $1/s^{c+2}$. Since the failure rate is set to be constant, the first part of the circuit uses energy $\Theta(s \log s)$. The probability that any gadget's output does not carry at least a θ fraction of the correct bits is at most $1/s^{c+1}$.

At the end of the circuit, we use the standard majority circuit of size $\Theta(\log s)$ to obtain the output, and set the failure of this section of the circuit to be $1/s^{c+2}$, thus this section of the circuit uses energy $\Theta(\log^3 s)$ and the probability that any gate in this section of the circuit fails is at most $1/s^{c+1}$. \square

III. LOGARITHMIC SAVINGS IS MAXIMAL FOR FUNCTIONS

We now show that, for a large class of natural functions, this $\Theta(\log s)$ savings is the best we can hope to do. In particular, we show a lower bound on the energy used by any heterogeneous circuit that computes a function, in terms of the number of non-degenerate input bits that the function has. For any function f with (1) minimum circuit size that is linear in n , that is $s = \Theta(n)$, and (2) $\Theta(n)$ non-degenerate input bits, when δ is polynomial in s , we can apply this lower bound to show that any heterogeneous circuit computing f must use $\Omega(s \log s) = \Omega(n \log n)$ energy.

We start with the definition of non-degenerate input bits, and then give the main theorem of this section.

Definition 3 (non-degenerate input bit). *The i th input bit to a Boolean function f with n inputs is non-degenerate if there exists some input $I \in [0, 1]^n$ such that $f(I) \neq f(I^i)$.*

Theorem 4. *Let f be a function with b non-degenerate input bits. Then, for any $\delta \in (0, 1/2)$, the optimal heterogeneous circuit requires $\Omega(b \log 1/\delta)$ energy.*

Proof. This proof is quite similar to the proofs of Theorem 1 and Lemma 6 of [5], which in turn use ideas from [9] and [10]. Space constraints force us to defer the full proof to the full version of this paper. \square

IV. OBTAINING LOG-SQUARED ENERGY SAVINGS FOR RELATIONS

In this section we prove that, in contrast with the previous section, there are relations where heterogeneous circuits can obtain a $\omega(\log n)$ energy savings over homogeneous circuits. In fact, we show that a natural supermajority relation obtains a $\Theta(\log^2 n)$ energy savings, which, as we show in Section V, is asymptotically the maximum possible savings for any relation that does not require circuits of superlinear size. Formally, we have the following theorem.

Theorem 5. *Suppose $\delta = 1/n^c$ for some constant $c > 0$. Then there is a relation such that the optimal heterogeneous circuit uses $O(n)$ energy, but the optimal homogeneous circuit requires $\Omega(n \log^2 n)$ energy.*

We cite the following general theorem proved by Pippenger in [11] and formalized by Gacs in [8] that will be useful in our construction in this section of the paper.

Theorem 6 ([11], [8]). *There is an $\epsilon_0 > 0$ such that for any $\epsilon < \epsilon_0$, $\delta \geq 3\epsilon$, and any function f computable by a faultless circuit of size s , there is an $(1 - \delta)$ -reliable circuit computing f of size $O(s \log(s/\delta))$ when gates fail with probability at most ϵ .*

The following relation is quite natural. The relation outputs the majority if at least $3/4$ of the bits are the majority, and otherwise we do not care about the output.

Definition 7. *Let $N_1(x)$ be the number of 1's in the binary string x . The Supermajority Relation (SR) is the following Boolean relation:*

$$SR(x) = \begin{cases} 0, & \text{if } N_1(x) < n/4, \\ 1, & \text{if } N_1(x) > 3n/4, \text{ and} \\ 0 \text{ and } 1 & \text{otherwise,} \end{cases}$$

where x is the input and $|x| = n$.

The proof of the following lemma is deferred to the full version of the paper.

Lemma 8. *When $\delta = 1/n^c$, for some constant $c > 0$, SR can be computed by a circuit with heterogeneous voltages using $O(n)$ energy.*

We can now prove our main theorem, which is straightforward given the previous lemma.

Proof of Theorem 5. By Lemma 8, SR can be computed by a heterogeneous circuit that uses $O(n)$ energy. It remains to show that any homogeneous circuit computing SR uses $\Omega(n \log^2 n)$ energy. Note that since gates in any homogeneous circuit computing SR cannot fail with probability more than δ , and since $\delta = 1/n^c$, the energy used by each gate must be at least $\Omega(\log^2 n)$. Additionally, it is obvious that any circuit correctly computing SR must have gates connected to at least half the inputs, and so any circuit computing SR using gates of constant fan-in must have $\Omega(n)$ gates. Therefore,

any homogeneous circuit computing SR must use $\Omega(n \log^2 n)$ energy. \square

V. LIMITATIONS ON SURPASSING LOG-SQUARED SAVINGS FOR RELATIONS

In this section we observe that, for relations with faultless circuits of linear size, heterogeneous supply voltages can yield at most $O(\log^2 n)$ energy savings.

Theorem 9. *Let h be any relation with minimum circuit size $s = O(n)$, and let $\delta = 1/s^c$ for some $c > 0$. Then the optimal heterogeneous circuit uses $\Omega(s)$ energy, and the optimal homogeneous circuit uses $O(s \log^2 s)$ energy.*

Proof. Consider any relation h that can be computed by a faultless circuit of size $s = O(n)$. Any heterogeneous circuit computing h must have size at least s (since the circuit must compute correctly even if no gate fails), and thus must use $\Omega(s)$ energy. On the other hand, if $\delta = 1/s^c$ for some $c > 0$, then a homogeneous circuit computing h can be constructed by setting the failure rate to $1/s^{c+1}$. By the union bound, the probability that even a single gate fails in this homogeneous circuit is at most $1/s^c$. Additionally, this circuit uses $O(s \log^2 s)$ energy. \square

VI. OPEN PROBLEMS

The main question left open by this paper is to determine if supply voltage heterogeneity allows asymptotic energy savings when the circuit error bound is constant.

REFERENCES

- [1] J. Butts and G. Sohi, "A static power model for architects," in *Proceedings of the 33rd annual ACM/IEEE International Symposium on Microarchitecture (MICRO)*, 2000, pp. 191–201.
- [2] R. G. Dreslinski, M. Wiczkowski, D. Blaauw, D. Sylvester, and T. N. Mudge, "Near-threshold computing: Reclaiming Moore's law through energy efficient integrated circuits," *Proceedings of the IEEE*, vol. 98, no. 2, pp. 253–266, 2010.
- [3] J. von Neumann, "Probabilistic logics and the synthesis of reliable organisms from unreliable components," in *Automata Studies*, C. E. Shannon and J. McCarthy, Eds. Princeton University Press, 1956, pp. 329–378.
- [4] K. V. Palem, "Energy aware computing through probabilistic switching: A study of limits," *IEEE Trans. Computers*, vol. 54, no. 9, pp. 1123–1137, 2005.
- [5] A. Antoniadis, N. Barcelo, M. Nugent, K. Pruhs, and M. Scquizzato, "Energy-efficient circuit design," in *Proceedings of the 5th conference on Innovations in Theoretical Computer Science (ITCS)*, 2014, pp. 303–312.
- [6] —, "Complexity-theoretic obstacles to achieving energy savings with Near-Threshold Computing," in *Proceedings of the 5th International Green Computing Conference (IGCC)*, 2014, pp. 1–8.
- [7] N. Barcelo, M. Nugent, K. Pruhs, and M. Scquizzato, "Almost all functions require exponential energy," in *Proceedings of the 40th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, 2015, pp. 90–101.
- [8] P. Gács, *Algorithms in Informatics*. Budapest: ELTE Eötvös Kiadó, 2005, vol. 2, ch. Reliable Computation.
- [9] P. Gács and A. Gál, "Lower bounds for the complexity of reliable Boolean circuits with noisy gates," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 579–583, 1994.
- [10] R. L. Dobrushin and S. I. Ortyukov, "Lower bound for the redundancy of self-correcting arrangements of unreliable functional elements," *Problems of Information Transmission*, vol. 13, pp. 59–65, 1977.
- [11] N. Pippenger, "On networks of noisy gates," in *Proceedings of the 26th Symposium on Foundations of Computer Science (FOCS)*, 1985, pp. 30–38.