Kirk Pruhs

Green Computing Algorithmics

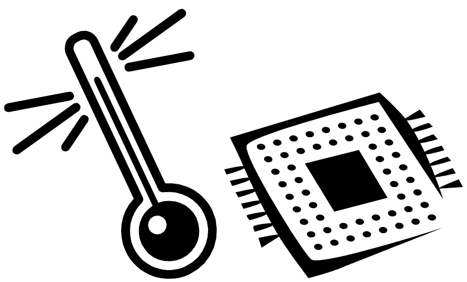
Talk 1: Energy Efficient Circuit Design

ADFOCS 2015

Motivation: Moore’s Gap

- Moore’s Law: Transistor density doubles every 18-24 months.

- Computer performance has not kept pace over the last 10 years due to the prohibitive cost of cooling such a high density of switches. The result is “Moore’s Gap”.



Moore's Law ran smoothly until 2002, when the gap between performance and gate count started to appear.

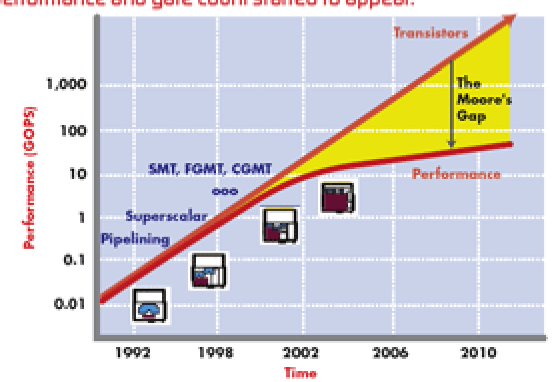


Figure 2

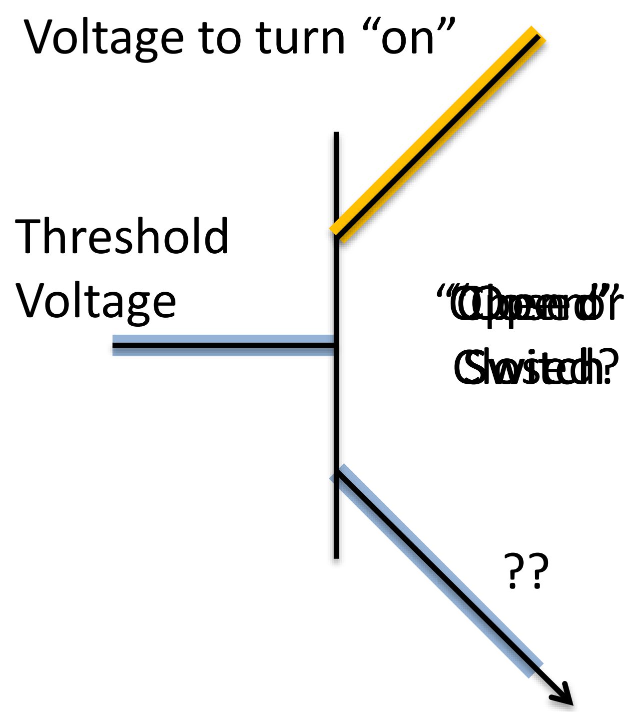
- One possible partial solution: Near Threshold Computing

Transistors 101

High Voltage:

Low Voltage:

- The building blocks of computers
- Acts as a switch when supplied with a high voltage
- Threshold Voltage: The lowest voltage at which the switch works (ideally)
- In reality, the probability that the switch works depends on the difference between the supply and threshold voltages.



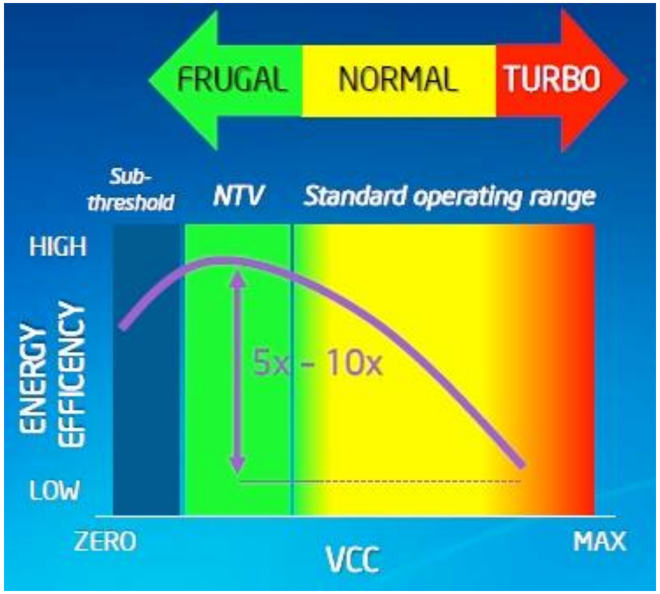
Traditional Approach to Setting the Supply Voltage

- Increase supply voltage so, by the union bound over all transistors, no transistor fails.

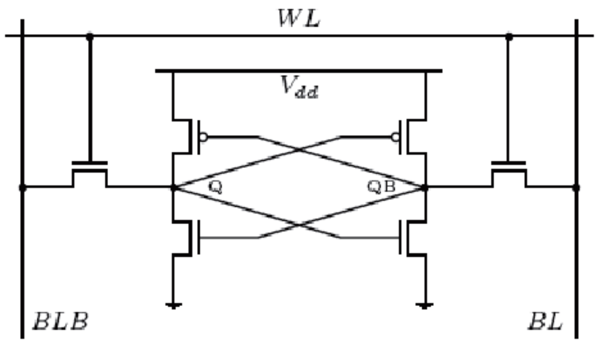
- Benefits:
  - Reliability
  - Also speed

Near-Threshold Computing

- Set the supply voltage close to the threshold voltage
- Advantage: Save energy per transistor
- Disadvantage: Decrease reliability per transistor. Requires fault-tolerant circuits.

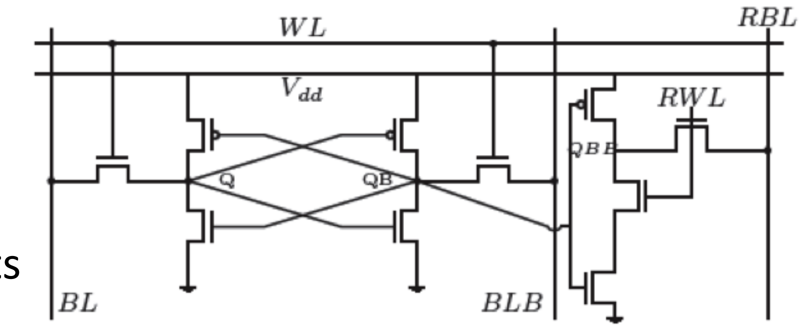


Minimizing Energy is a Balancing Act



(a) Standard 6-transistor design.

SRAM circuits



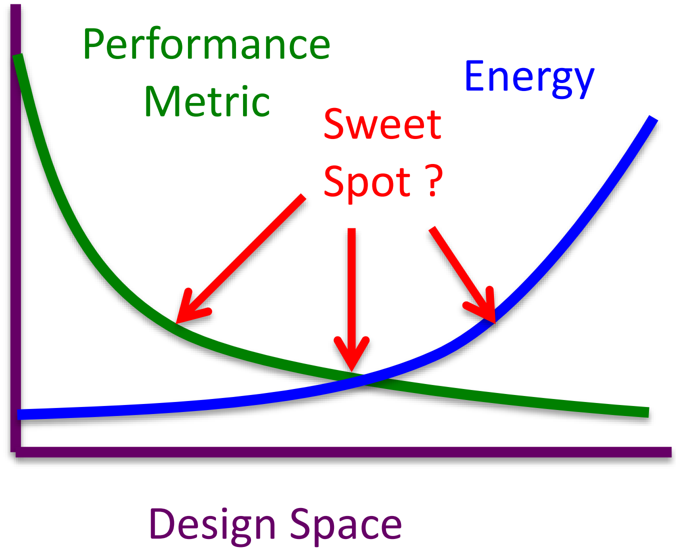
(b) A more fault-tolerant 10-transistor design from [3].

- Traditional:
  - Smaller number of transistors
    - nonfault tolerant circuit
  - Higher energy per transistor

vs.

- Near-threshold:
  - Larger number of transistors in a fault tolerant
  - Lower energy per transistor

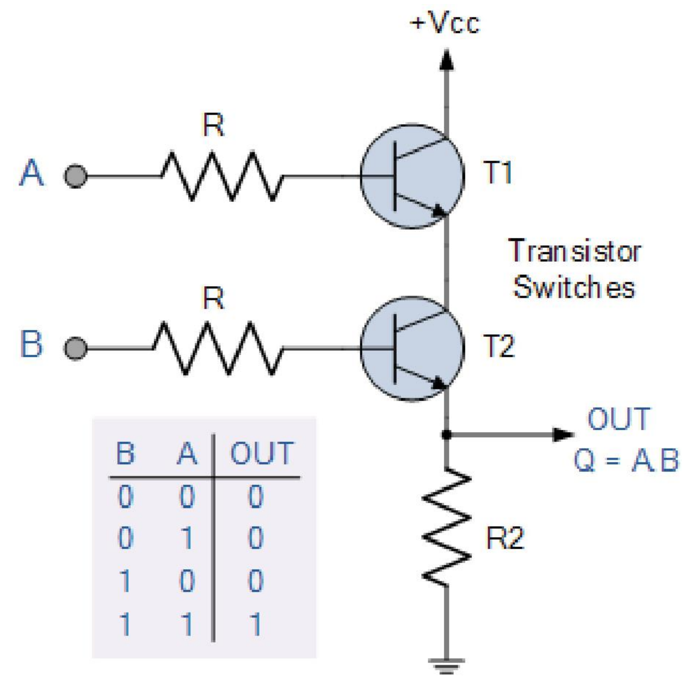
Current State Of Theory of Energy as a Computational Resource: Energy vs. Performance Tradeoffs



From Transistors to Gates

- Gates are composed of resistors and transistors, e.g.,
  - High Voltage: 1
  - Low Voltage: 0

- For simplicity, consider gates rather than transistors



B	A	OUT
0	0	0
0	1	0
1	0	0
1	1	1

Two transistor AND gate

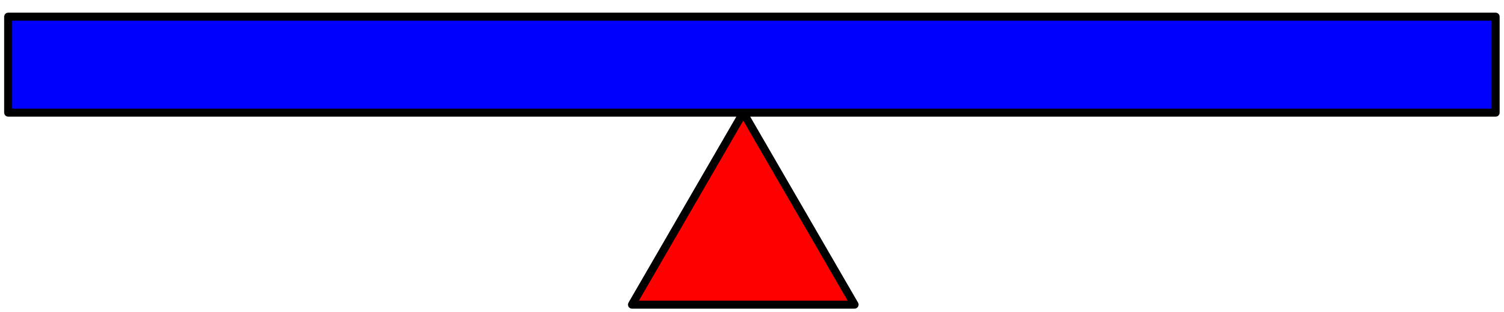
- If you wanted to do algorithmic research on near-threshold computing, what would you have to do first?

Algorithmists'
View of Science/Theory

- Science research tries to model a complex system by something simple, accurate, amenable to math and predictive. Muthu Muthukrishnan's blog

•Accuracy
•Realism
•Predictive

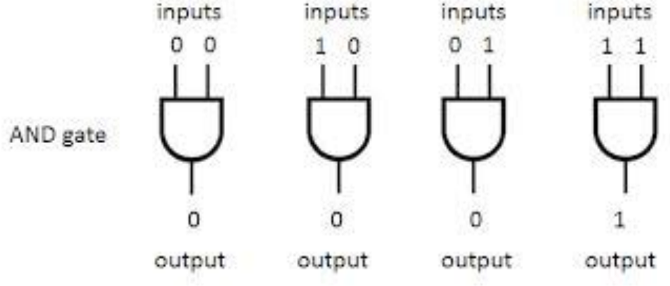
•Simplicity
•Amenable to math



Two Modeling Issues

- Relationship between energy and error at a gate

- What happens when there is a error



Voltage, Energy, and Error

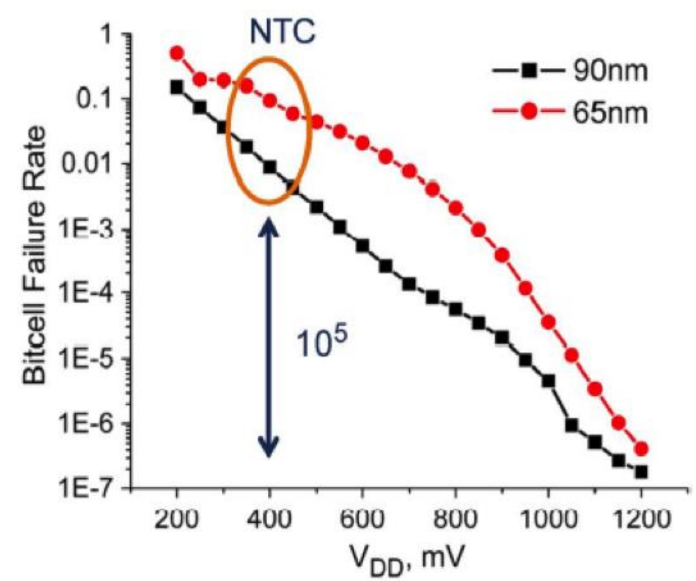
- Let v be the supply voltage
- Power/Energy = v^2
- Error Probability $\varepsilon = 2^{-v}$
- There, error to energy function $E(\varepsilon) = \log^2 1/\varepsilon$
  - Not crucial

INVITED PAPER

Near-Threshold Computing: Reclaiming Moore's Law Through Energy Efficient Integrated Circuits

Future computer systems promise to achieve an energy reduction of 100 or more times with memory design, device structure, device fabrication techniques, and clocking, all optimized for low-voltage operation.

By Ronald G. Dreslinski, Michael Wieckowski, David Blaauw, *Senior Member IEEE*, Dennis Sylvester, *Senior Member IEEE*, and Trevor Mudge, *Fellow IEEE*

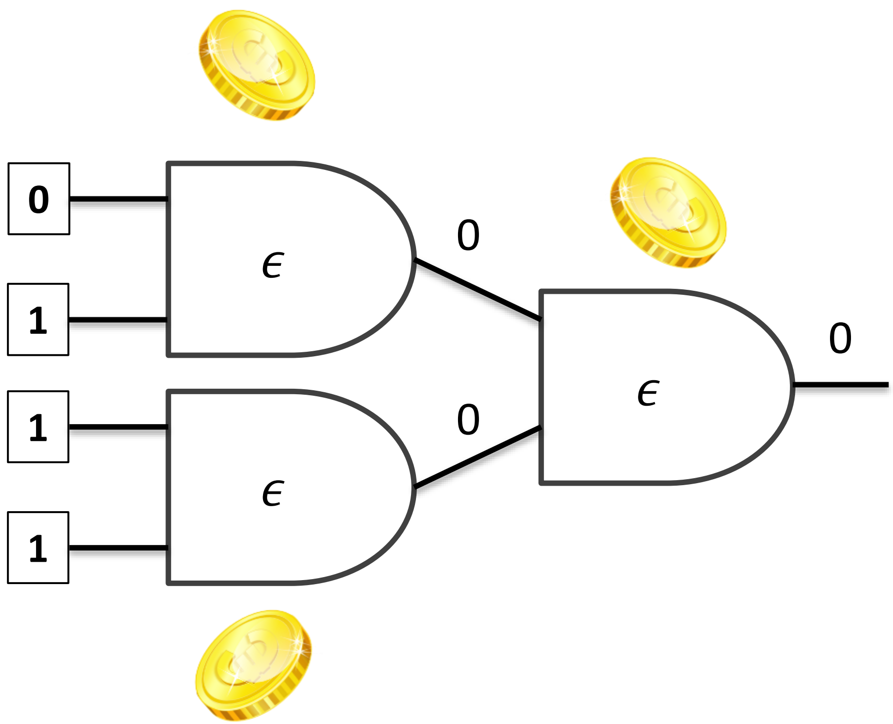


Modeling Errors

- Error type models:
  - Toggle
  - Stuck-at
  - Generally not crucial
- Probability models
  - Exact: probability of error is exactly ε
  - Bounded: probability of error is adversarially set to be in the range $[0, \varepsilon]$
  - Generally not crucial

Example – 3 AND Gates

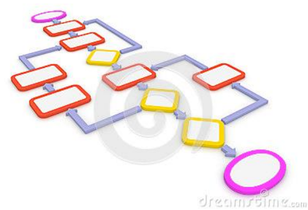
Each gate fails/toggles independently with probability ε



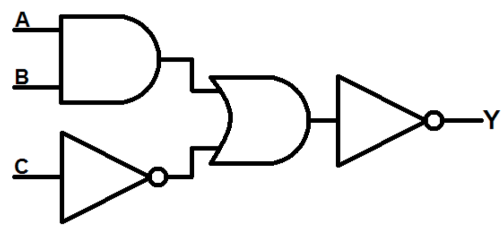
- Now what are some interesting research questions?

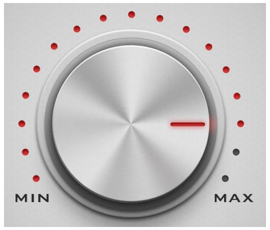
Natural Problem: Finding Minimum Energy Circuit (MEC)

- Input:
  - A function/relation F
  - Desired error bound δ



- Output:
  - Circuit C that computes F
  - Setting of the supply voltage

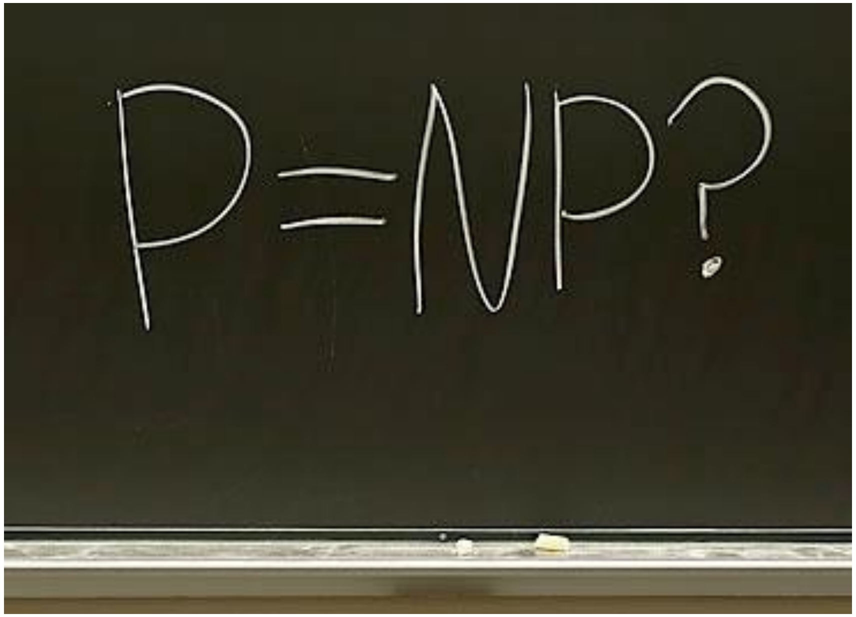


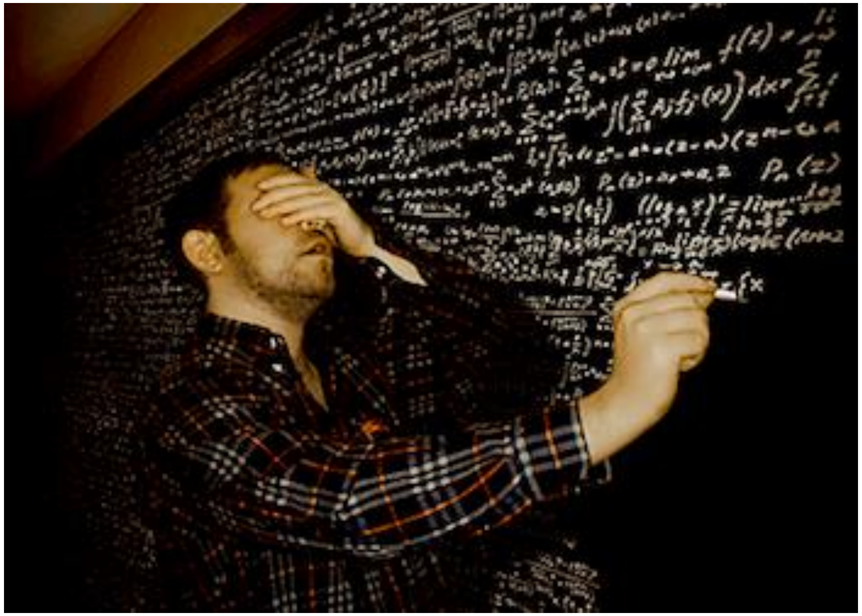


- Constraint: On all inputs the circuit should be correct with probability > 1-δ
- Objective: Minimize power

Unfortunately…

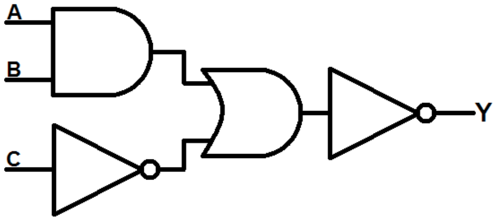
- This problem is too hard. We don't have a clue how to find small circuits. As hard as P vs. NP.



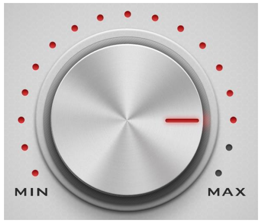


Natural Fallback Problem: Deciding the Right Supply Voltage for a Circuit (MCE)

- Input:
  - Combinatorial circuit
  - Desired error bound δ



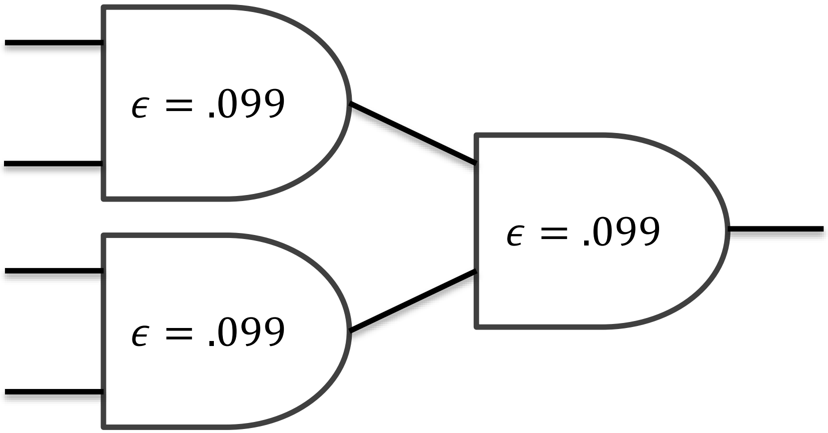
- Output:
  - Setting of the supply voltage



- Constraint: On all inputs the circuit should be correct with probability > 1-δ
- Objective: Minimize power

MCE Problem

Input: Following circuit and δ = .75



- Output:
  - ε ≈ .099
  - Energy = 3 $\log^2$ (1/.099) ≈ 33.39

Obvious Natural Questions

- MEC
  - General lower bounds on energy?
  - General upper bounds on energy?
  - Energy for a random function?
- MCE
  - Complexity?
  - Approximation ratio of traditional approach?
  - Best possible approximation ratio?

Less Obvious Natural Questions

- Can you save energy by making supply voltages heterogeneous?
  - Does it matter whether you computing a function or an injective relation (so you don't care on some inputs)?
  - Does it matter the circuit error bound δ is a constant or if δ goes to 0 as the circuit size increases?

Roadmap

- General bounds
- Approximation for Minimum Circuit Energy (MCE)
- Minimum Energy Circuit (MEC) for random functions
- Energy savings from allowing heterogeneous supply voltages

Previous Work: Fault Tolerant Circuits

- Introduced by **von Neumann** in **1956**: fixed probability of error ϵ
- Upper bound of $O(N \log N)$ given a faultless circuit of size N.
  - Heuristically argued by **von Neumann**, formalized and made more explicit by **Dobrushin and Ortyukov (77)**, **Pippenger (85)**, and **Gács (05).**
- Lower bound of $\Omega(s \log s)$ for functions with sensitivity s.
  - **Dobrushin and Ortyukov (77)** published proof containing errors, correctly proved by **Gács and Gál (94).**

Warmup

- Recall δ is given circuit error
- Recall ε is gate error that needs to be determined
- Recall s = number of gates
- Energy $E(\varepsilon) = \log^2 1/\varepsilon$

- Lemma: $\delta/s \leq \varepsilon \leq \delta$
- Proof?

- Corollary: Energy per gate is $\Omega(1)$.
- Proof?

- Corollary: Energy per gate is $O(\log^2 \delta/s)$, which is $O(\log^2 s)$ if δ is constant.
- Proof?

General Energy Upper Bound

Theorem: Given a circuit C with n gates that computes a relation f (with no failures), and a constant δ, the optimal solution to MEC uses energy O(n log n).

Proof idea:

General Energy Lower Bound

Theorem: Given δ and a relation f with sensitivity s, the optimal solution to MEC requires energy Ω(s log (s/δ)).

- The sensitivity of on an input is the number of bits of that, if flipped, change the output of .
- The sensitivity of is the maximum sensitivity of over all inputs.

Proof idea:

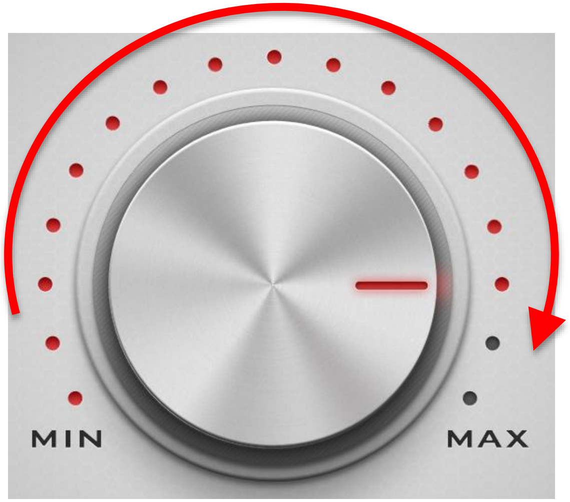
Roadmap

- General bounds
- Approximation for Minimum Circuit Energy (MCE)
- Minimum Energy Circuit (MEC) for random functions
- Energy savings from allowing heterogeneous supply voltages

MCE Problem

- Theorem: For constant δ, the traditional approach is an θ($\log^2 s$) approximation.
  - Proof:
  - Traditional approach sets gate error ε = circuit error bound δ / s
  - So traditional approach E = s $\log^2$ 1/(δ/n)
  - Optimal E ≥ s $\log^2$ 1/δ since ε ≤ δ

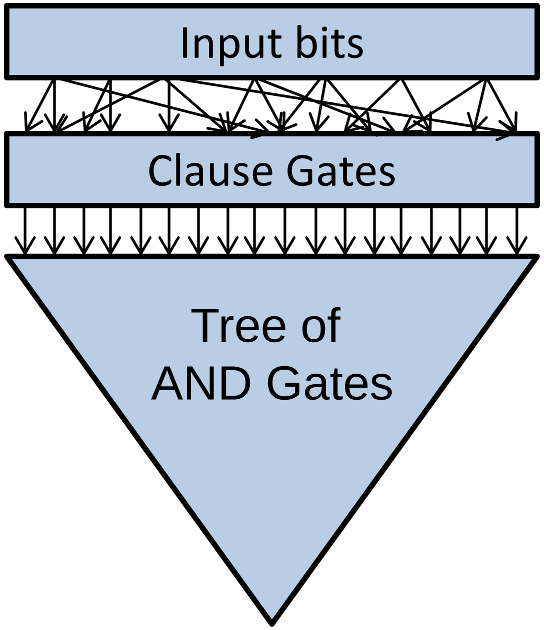
MCE Problem



- Theorem: For constant δ, it is NP-hard to obtain an $o(\log^2 s)$ approximation.
  - Proof:
    - Reduction for gapped version of 3SAT
    - NP-hard Gapped 3SAT Problem:
      - Input: Formula F
      - If F is satisfiable then the output has to be 1
      - If no assignment can not not satisfy more than 15/16 of the clauses of F then the output has to be 0

NP-hardness Hardness via Reduction From 3SAT

- Consider the natural circuit corresponding to formula in conjunctive normal form
  - If no assignment can not not satisfy more than 15/16 of the clauses then AND gates can occasionally fault
    - OK to set $\varepsilon \approx \delta$
    - and hence the power can be a factor of $\log^2 s$ less than traditional amount
  - If the formula is satisfiable,
    - then the AND gates basically can't fault on a satisfying input,
    - So we need $\varepsilon \leq \delta / s$
    - and hence the power is the traditional amount

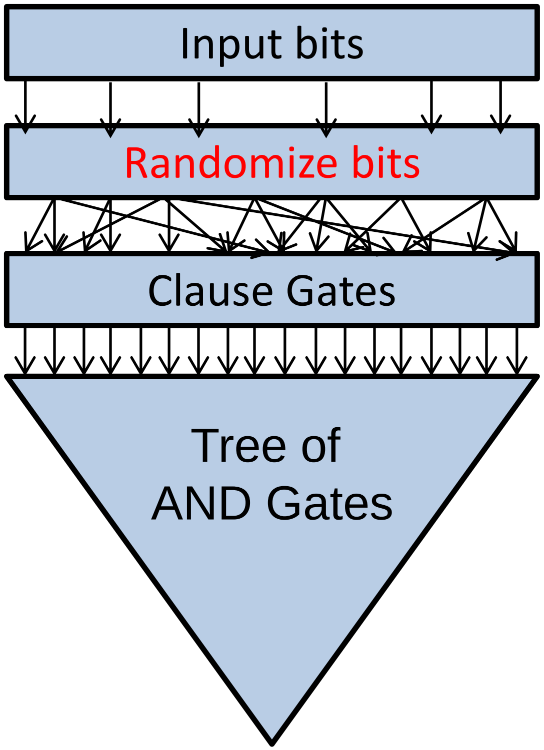


MCE Problem

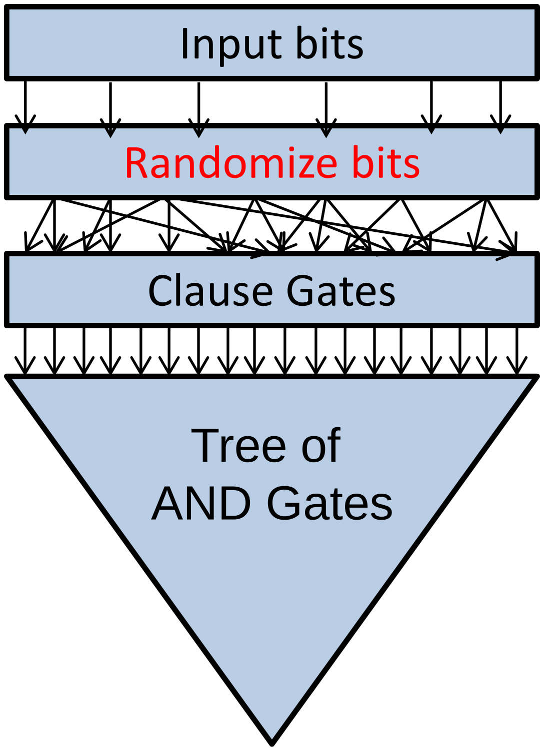
- Theorem: It is NP-hard to obtain an $o(\log^2 n)$ approximation even for a fixed input.
  - Proof:
    - Reduction for gapped version of 3SAT
    - NP-hard Gapped 3SAT Problem:
      - Input: Formula F
      - If F is satisfiable then the output has to be 1
      - If no assignment can not not satisfy more than 15/16 of the clauses of F then the output has to be 0

NP-hardness Hardness via Reduction From 3SAT

- Consider the natural circuit corresponding to formula in conjunctive normal form
  - How do you randomize a bit using faulty gates?

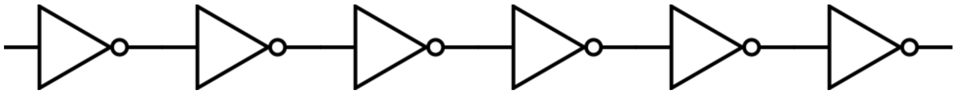


NP-hardness Hardness via Reduction From 3SAT

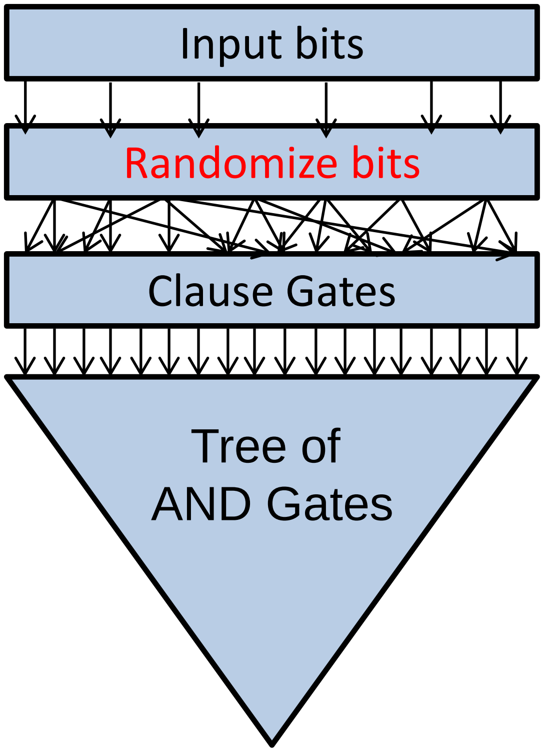


- Consider the natural circuit corresponding to formula in conjunctive normal form

  – How do you randomize a bit using faulty gates?

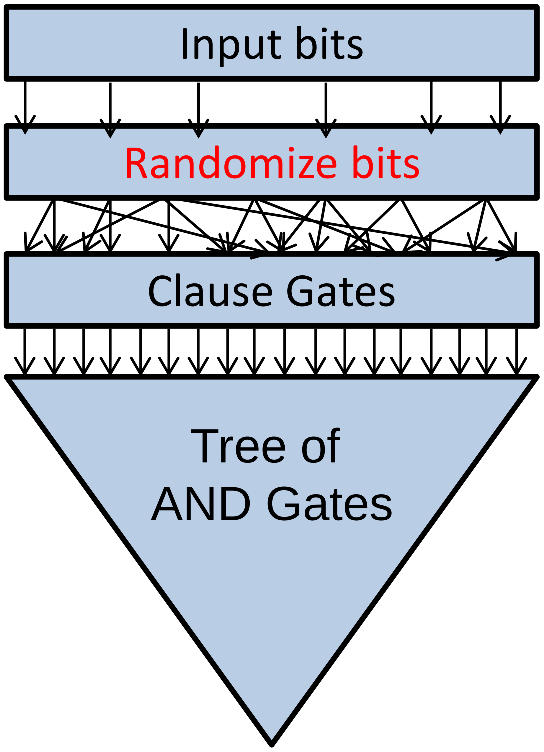


NP-hardness Hardness via Reduction From 3SAT



- Consider the natural circuit corresponding to formula in conjunctive normal form
  - If at most 15/16th of the clauses are satisfiable
    - then many outputs of the clause gates are 0
    - OK if some of the AND gates fail
    - OK to set ε≈ δ
    - and hence the power can be a factor of $\log^2 n$ less than traditional amount

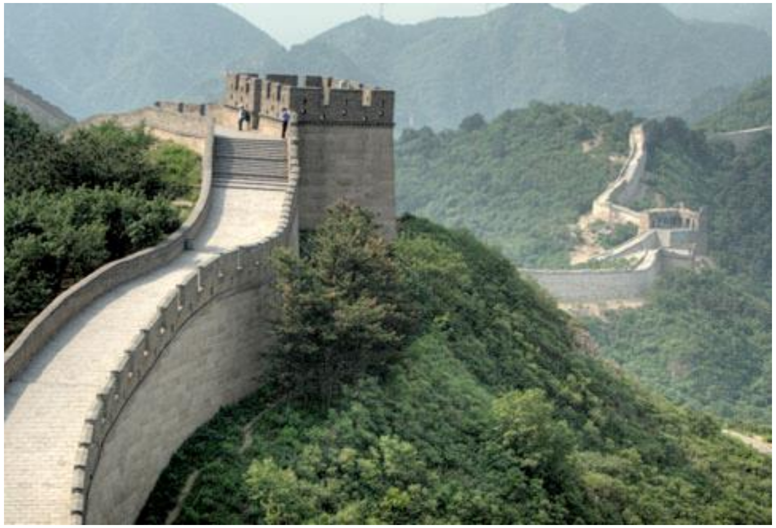
NP-hardness Hardness via Reduction From 3SAT

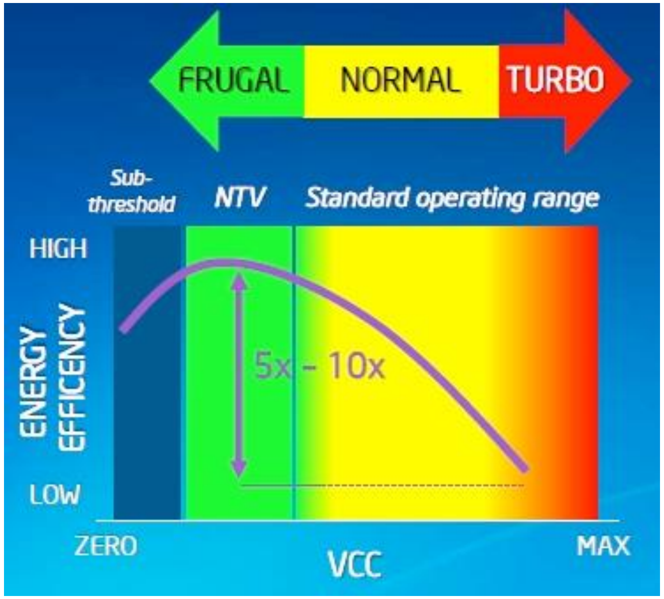


- Consider the natural circuit corresponding to formula in conjunctive normal form
  - If the formula is satisfiable,
    - then the AND gates basically can't fault on the satisfying assignment,
    - So we need $\varepsilon \leq \delta / s$
    - and hence the power is the traditional amount

Take Away Point: Beating the Traditional Approach is NP-hard

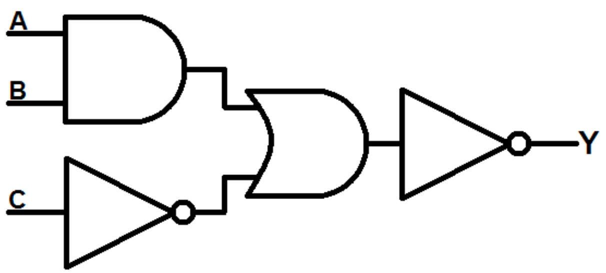
- Therefore, there is a complexity theoretic barrier to <u>systematically</u> achieving energy savings with near-threshold circuits.





Positive Result MCE

- Optimal supply voltage can be computed if the circuit is a tree (out-degree of all gates is 1).
  - Suggests tractability for nearly treelike circuits



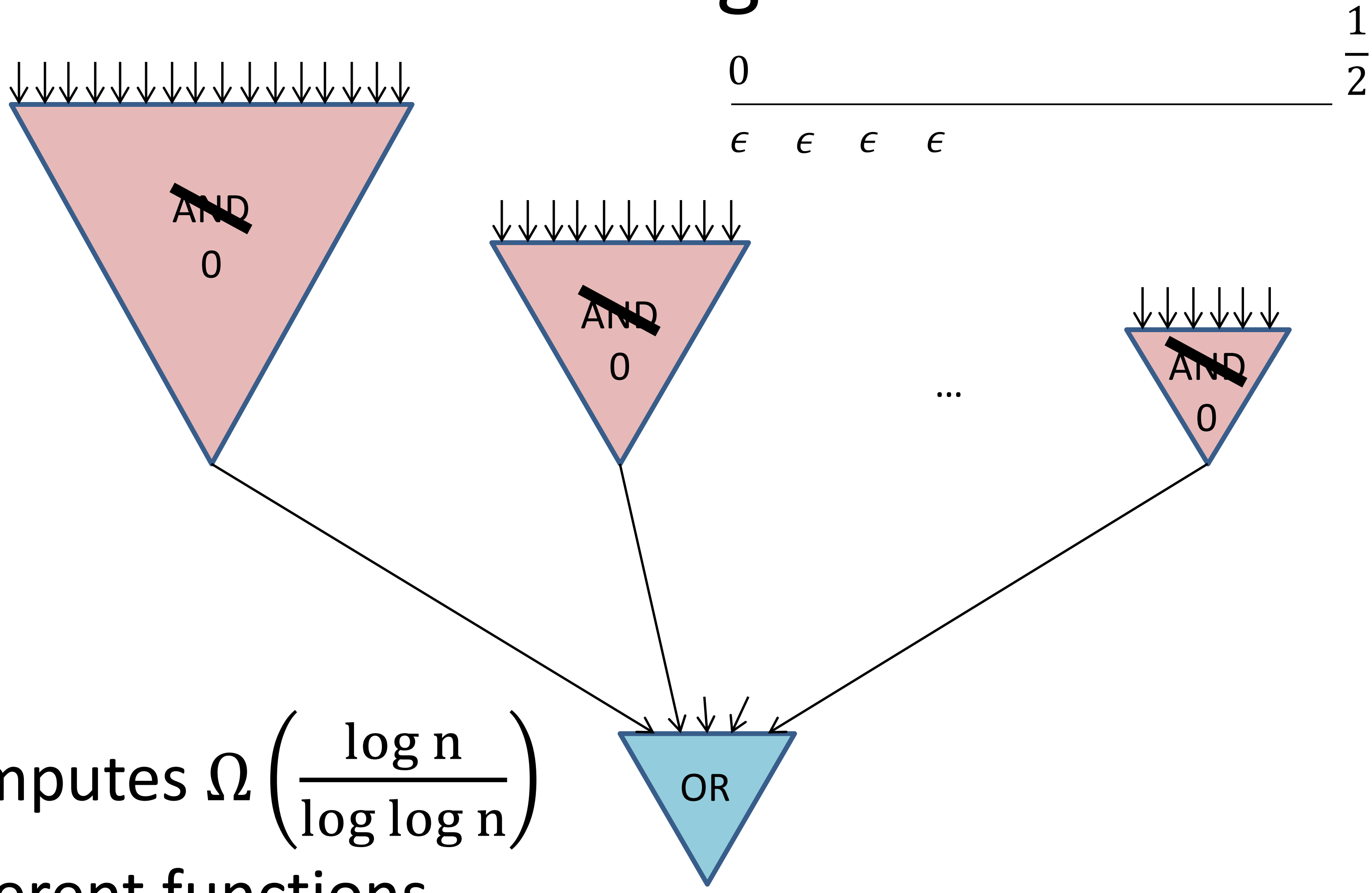
Roadmap

- General bounds
- Approximation for Minimum Circuit Energy (MCE)
- Minimum Energy Circuit (MEC) for random functions
- Energy savings from allowing heterogeneous supply voltages

- Theorem [Shannon 1949]: Almost all Boolean functions on n bits require exponentially sized circuits:
- Proof:
  - There are $2^{\wedge}2^n$ Boolean functions on n bits
  - There are only 2^j circuits described by j bits
    - Thus at most $2^{k \log k}$ circuits of k gates
  - Thus in order for $2^{k \log k} > 2^{\wedge}2^n / 2$ then k needs to at least $2^n/n$

- Theorem: Almost all Boolean functions on n bits require exponential energy circuits
- Proof: Why won't the same proof work as we know you need constant energy per gate?

Why Shannon's Proof Does Not Extend: Homogeneous Case



Computes $\Omega\left(\frac{\log n}{\log \log n}\right)$ different functions

- Theorem: A circuit C with n inputs and s gates can compute at most s 2^n different functions.
  - Proof: Sufficient to show that on each input I (of the 2^n inputs), C can only change its output s times.
  - Let $p_I(\varepsilon)$ be the probability that C outputs a 1 on input I when the gate error is ε
  - In order for the output to change $p_I(\varepsilon)$ has to transition from $<\delta$ to $>(1-\delta)$, or visa versa.

- Theorem: A circuit C with n inputs and s gates can compute at most s 2^n different functions.
- Proof: Sufficient to show that on each input I (of the 2^n inputs), C can only change its output s times.
- Let $p_I(\varepsilon)$ be the probability that C outputs a 1 on input I when the gate error is ε
- In order for the output to change $p_I(\varepsilon)$ has to transition from <δ to >(1-δ), or visa versa.
- $p_I(\varepsilon) = \Sigma_k \Sigma_{G(k, I)} \varepsilon^k (1-\varepsilon)^{s-k}$
  - k in [0, s] is number of gate failures
  - G(k, I) collections of k gates such that when exactly these gates fail, the C on I outputs 1
- How often can $p_I(\varepsilon)$ transition?

- Theorem: Almost all Boolean functions on n bits require exponential energy circuits
- Proof:
  - There are 2^{2^n} Boolean functions on n bits
  - There are only 2^j circuits described by j bits
    - Thus at most $2^{k \log k}$ circuits of k gates
    - Thus there are at most $k\, 2^n$ different functions that each circuit can compute
  - Thus in order for $2^{k \log k}\, k\, 2^n > 2^{2^n} / 2$ then k needs to at least $2^n/n$

Roadmap

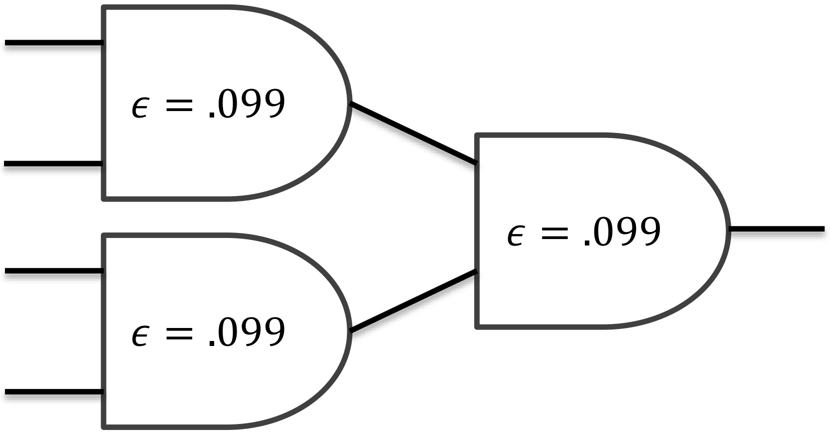
- General bounds
- Approximation for Minimum Circuit Energy (MCE)
- Minimum Energy Circuit (MEC) for random functions
- Energy savings from allowing heterogeneous supply voltages

Homogeneous vs. Heterogeneous Supply Voltages

- Homogeneous supply voltages: the error probability ε_g is the same for each gate g
- Heterogeneous supply voltages: each game g may have a different error probability ε_g
- Essentially all results so far would still hold if heterogeneous voltages were allowed, although sometimes the proof is harder.

MCE Problem: Homogeneous Supply Voltages

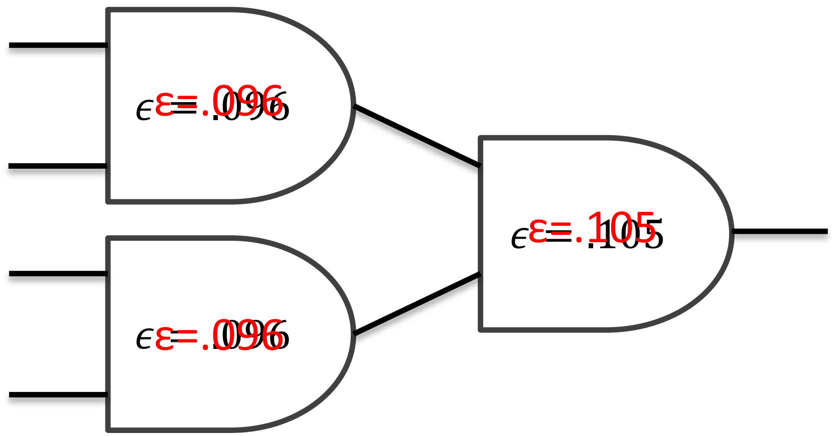
Input: Following circuit and $\delta = .75$



- Output:
  - $\varepsilon \approx .099$
  - Energy = $3 \log^2 (1/.099) \approx 33.39$

MCE Problem: Heterogeneous Supply Voltages

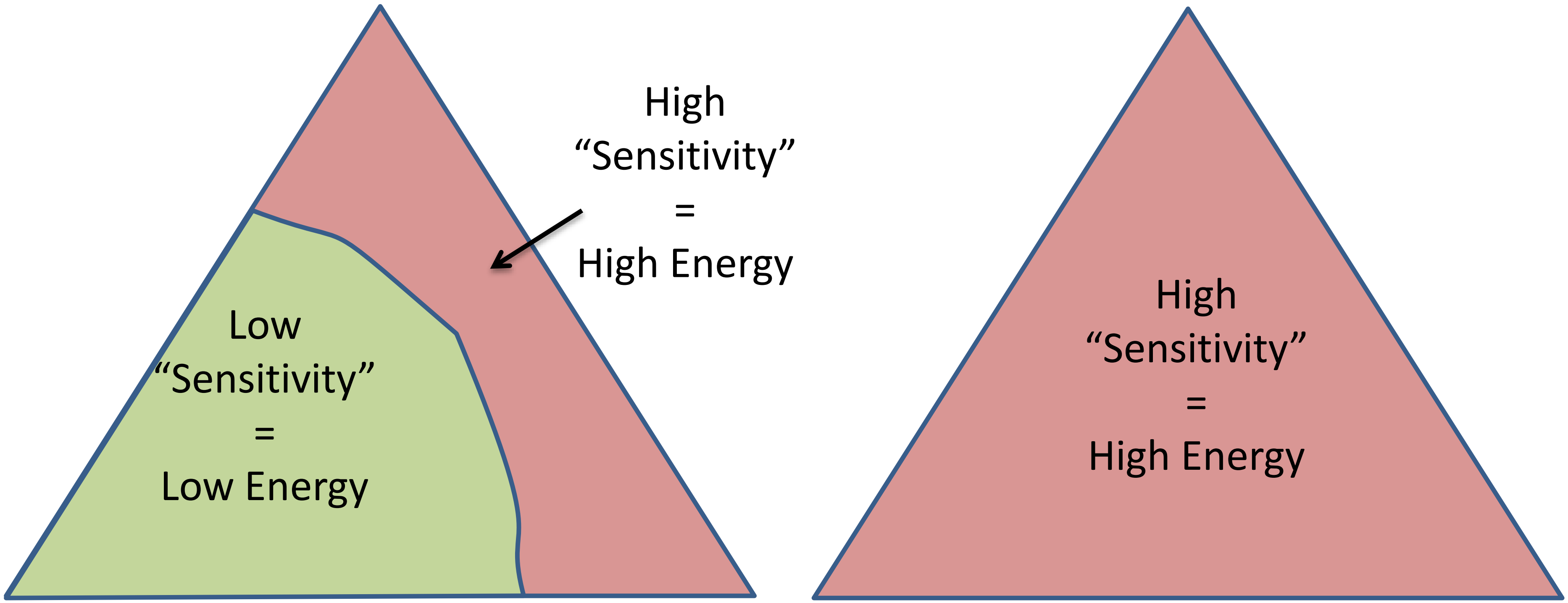
Input: Following circuit and δ = .75



Output:

- ε's as shown above
- Energy = $2 \log^2 (1/.096) + \log^2 (1/.105) \approx 33.39$
- Better by .1 % compared to optimal homogeneous

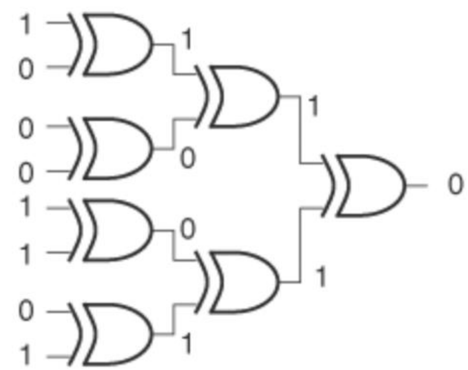
- Can you think of a circuit where heterogeneity would help save significant energy, and a circuit where it wouldn't?



A Case Where Heterogeneity Does Not Help

Theorem: For constant $\delta \in \left(0, \frac{1}{2}\right)$, the solutions to MEC and HMEC for parity are within a constant.

- Parity has sensitivity n so the lower bound says that $\Omega(n \log n)$ energy is required.
- This circuit computes the parity function:



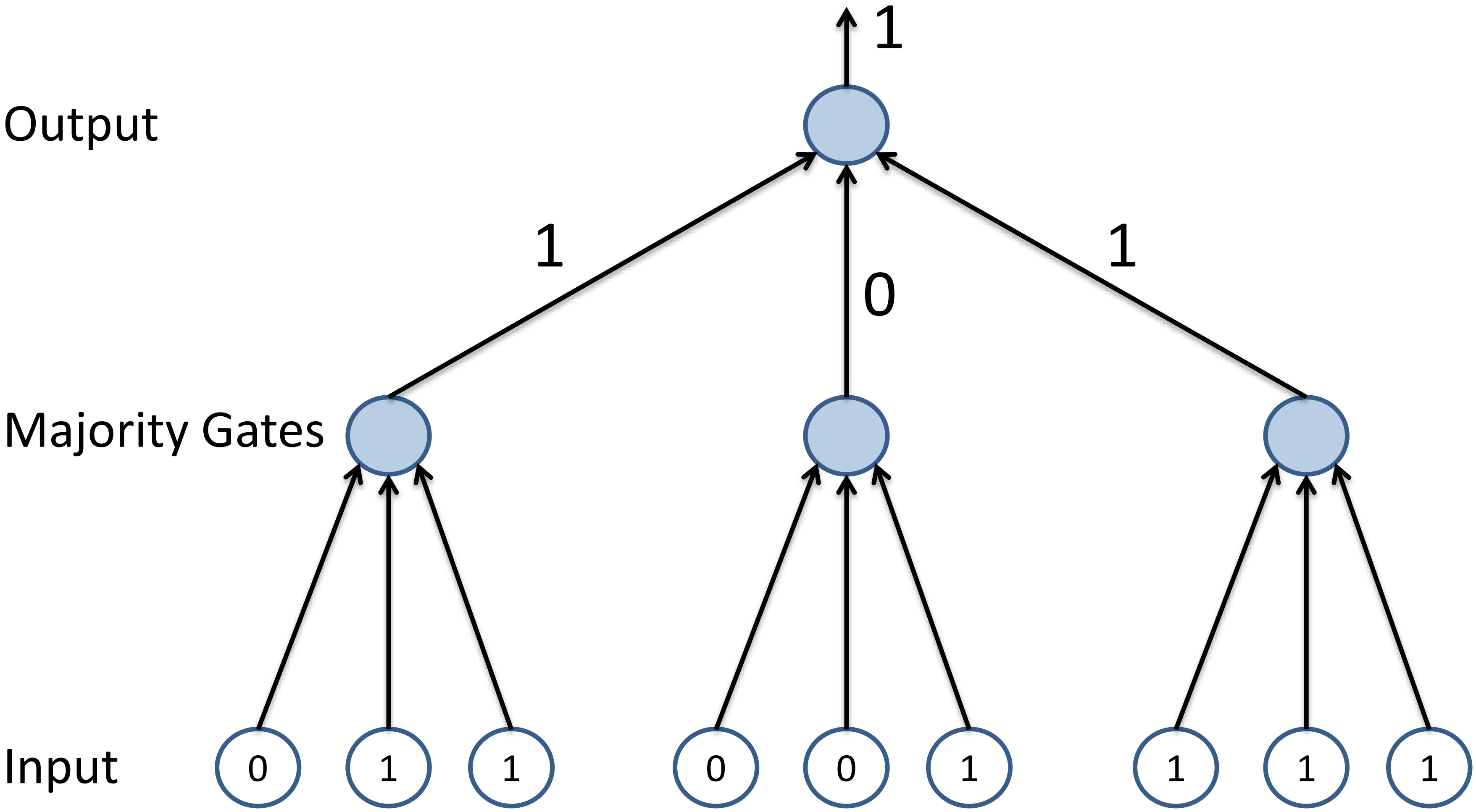
- It has size $O(n)$ so the upper bound says $O(n \log n)$ energy sufficient (and it uses homogeneous failure rates).

A Case Where Heterogeneity Does Help

- Relation: The logarithmic supermajority relation, LSR(x) =
  - 1 if > n – (log3 n)/2 of the input bits are 1
  - 0 if > n – (log3 n)/2 of the input bits are 0
  - Don't care otherwise

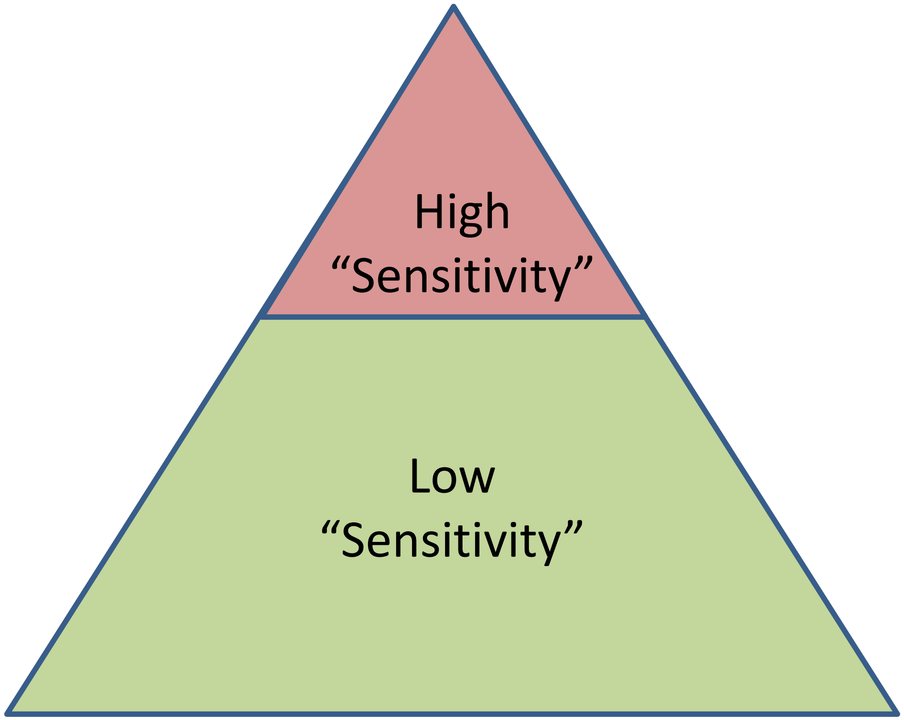
- Circuit Computing This Relation: Ternary tree of majority gates

Supermajority Circuit



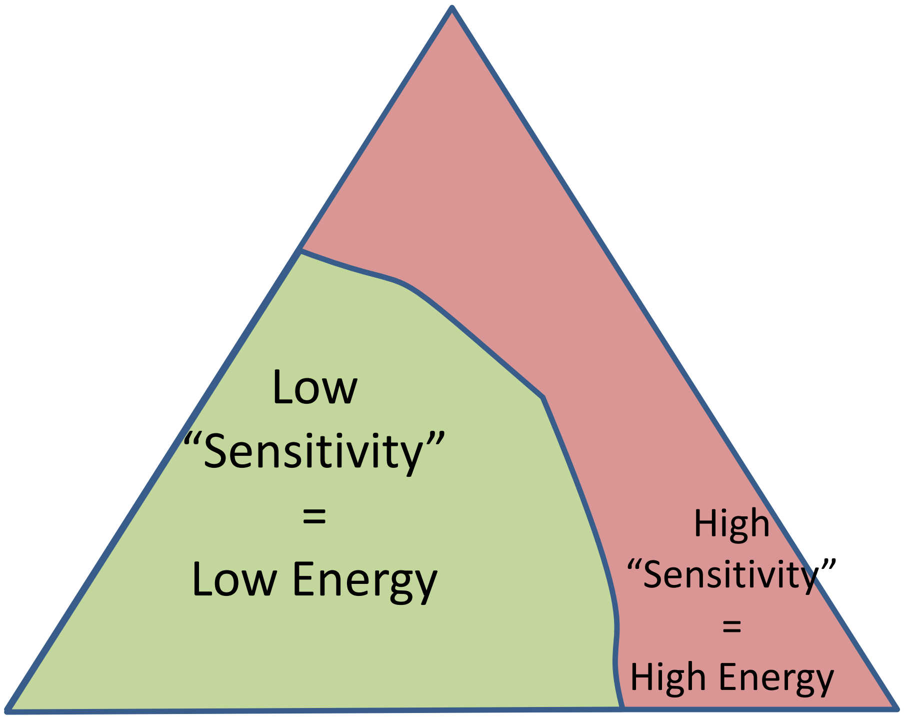
Supermajority Circuit

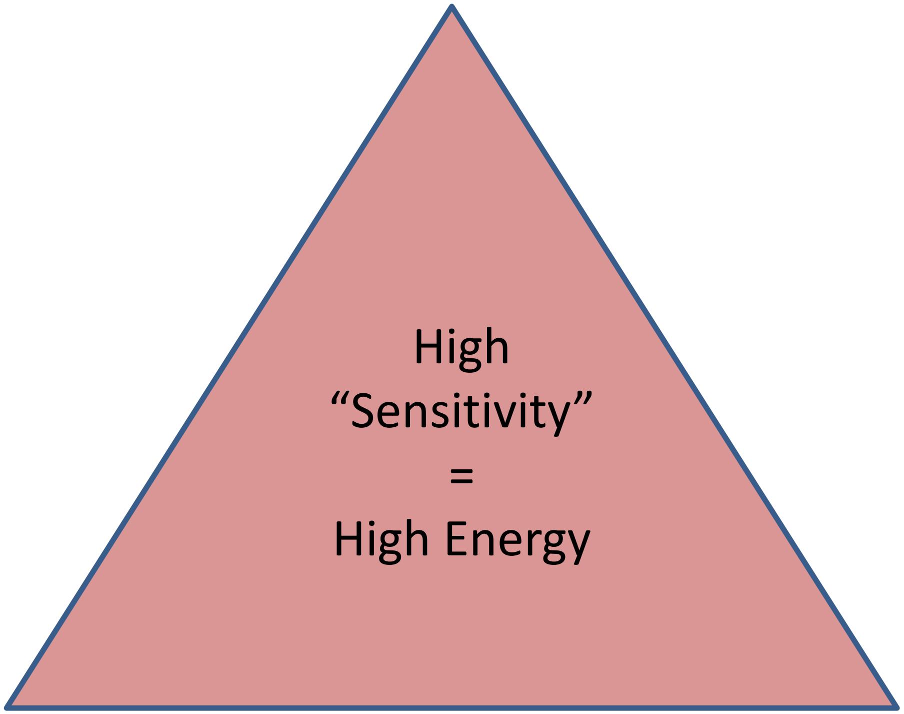
- Root is more “sensitive” to failures.



Big Question

- Are there functions/relations where there is an almost minimum energy circuit that is homogeneous?
- Are there functions/relations where there is no almost minimum energy circuit that is homogeneous?





Energy Savings From Heterogeneity

	δ = constant	δ = 1/poly
Functions	Θ(1) for some $O(\log^2 n)$ for all	• Θ(log n) for all functions with linear sized circuits
Injective Relations	Θ(1) for some $O(\log^2 n)$ for all	• $\Omega(\log^2 n)$ for some • $O(\log^2 n)$ for relations with linear sized circuits

Energy savings for function/relation F = ratio

$$\frac{\text{Minimum energy used by a homogeneous circuit for F}}{\text{Minimum energy used by a heterogeneous circuit for F}}$$

Energy Savings From Heterogeneity

	δ = constant	δ = 1/poly
Functions	Θ(1) for some $O(\log^2 n)$ for all	• Θ(log n) for all functions with linear sized circuits
Injective Relations	Θ(1) for some $O(\log^2 n)$ for all	• $\Omega(\log^2 n)$ for some • $O(\log^2 n)$ for relations with linear sized circuits

Proofs?

$\Omega(\log n)$ Energy Savings For Functions when $\delta = 1/n$

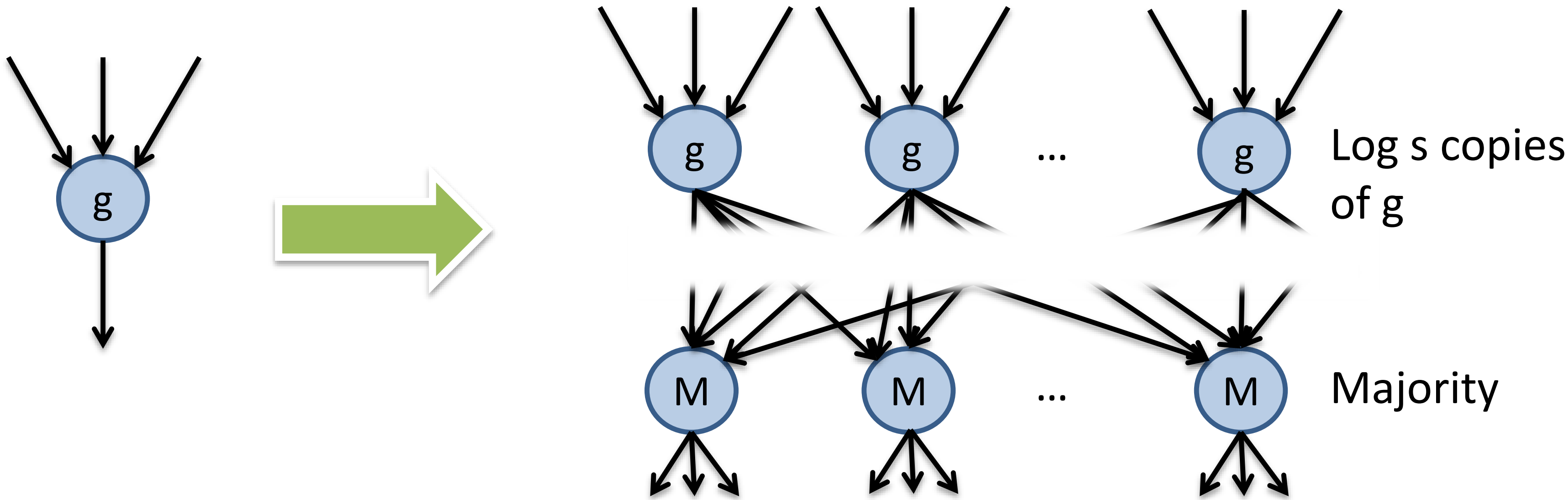
- **Theorem**: For $\delta = 1/n$, and for every function F on n input bits with a s gate circuit:
  - Every circuit with homogeneous supply voltages that computes F uses $\Omega(n \log^2 n)$ energy.
  - There is a circuit with heterogeneous supply voltages that computes F using $O(s \log s)$ energy

- Note: $\Omega(\log n)$ energy savings if $s=\Theta(n)$

- Proof: First statement is trivial

Proof: Heterogeneous Circuit Using O(s log s) Energy

- Each of s gates g in error-free circuit is replaced by a gadget:

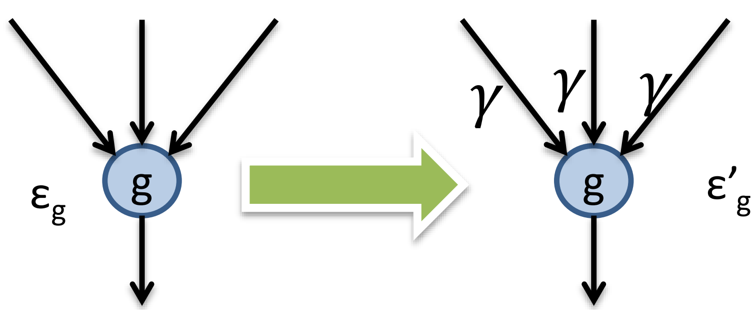


- ε = constant for these gadgets, so energy Θ(s log s)
- If most of the inputs into the gadget are correct, then most of the outputs are correct
- Final majority circuit of size log n with $\varepsilon = 1/s^2$, so energy $\Theta(\log^3 s)$

O(log n) Limit on Energy Savings for Functions when δ=1/n

Theorem: For any function F whose output depends on all n inputs, every heterogeneous circuit that computes F requires Ω(n log n) energy

Proof Idea: log n copies with constant error per copy is the most energy efficient way to get aggregate error 1/n

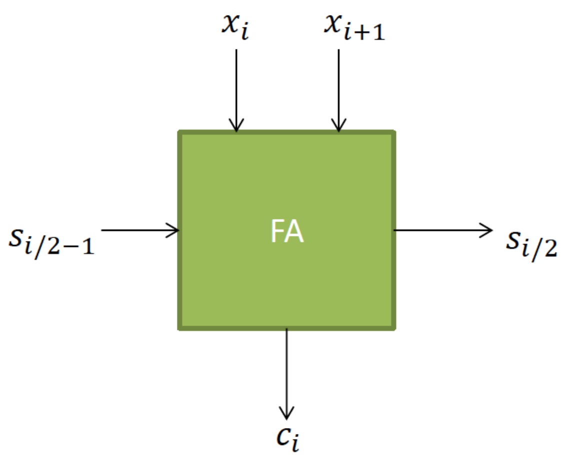


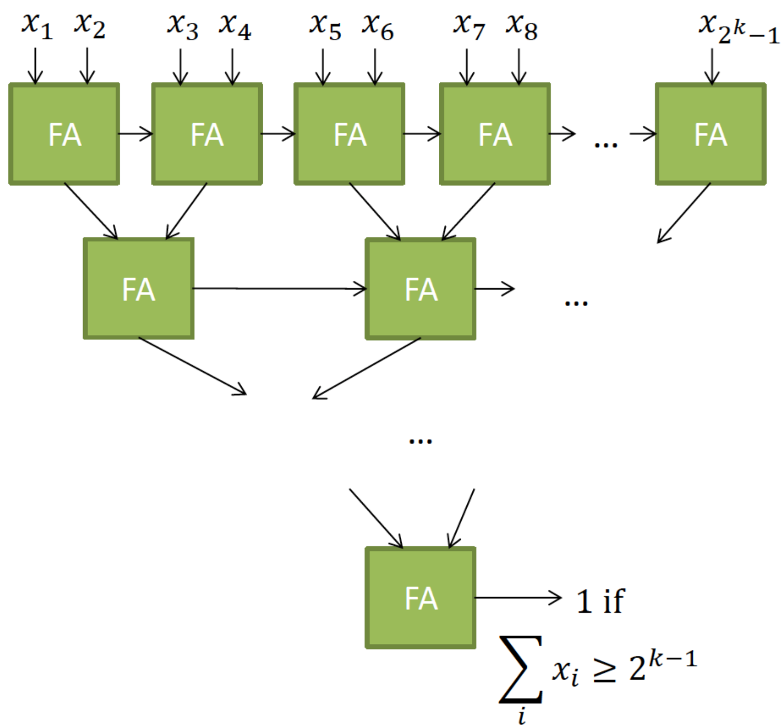
$\Omega(\log^2 n)$ Energy Savings For Supermajority Relation when $\delta = 1/n$

- Supermajority Relation: Only needs to be correct if ¾ majority.

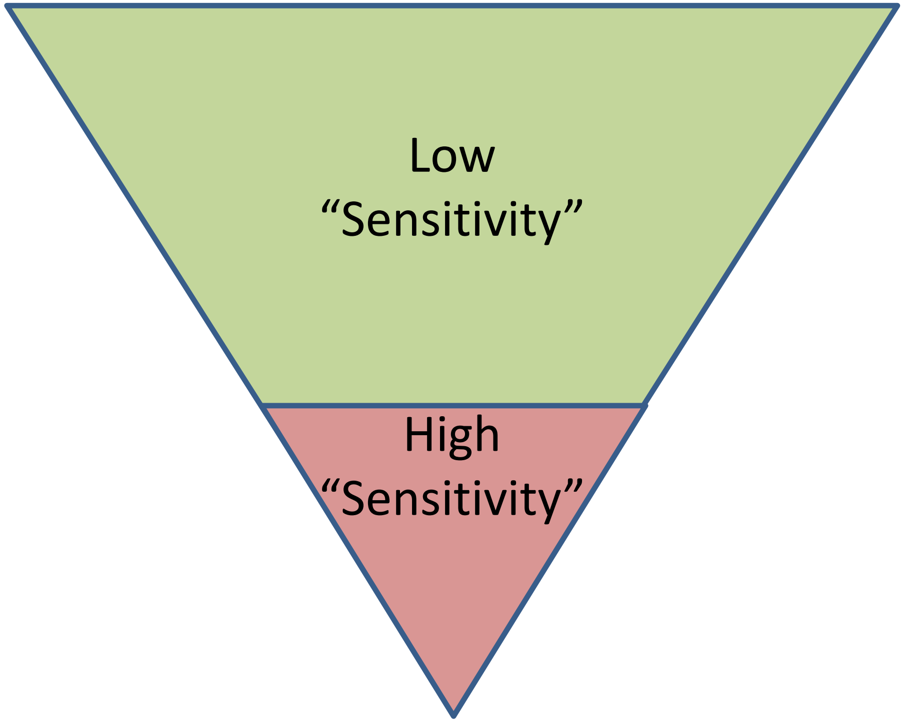
- **Theorem**: For $\delta = 1/n$,
  - Every circuit with homogeneous supply voltages that computes the supermajority relation and that uses $\Omega(n \log^2 n)$ energy.
  - There is a circuit with heterogeneous supply voltages that computes the supermajority relation and uses $O(n)$ energy
- Proof: First statement is trivial

Proof: Low Energy Heterogeneous Circuit

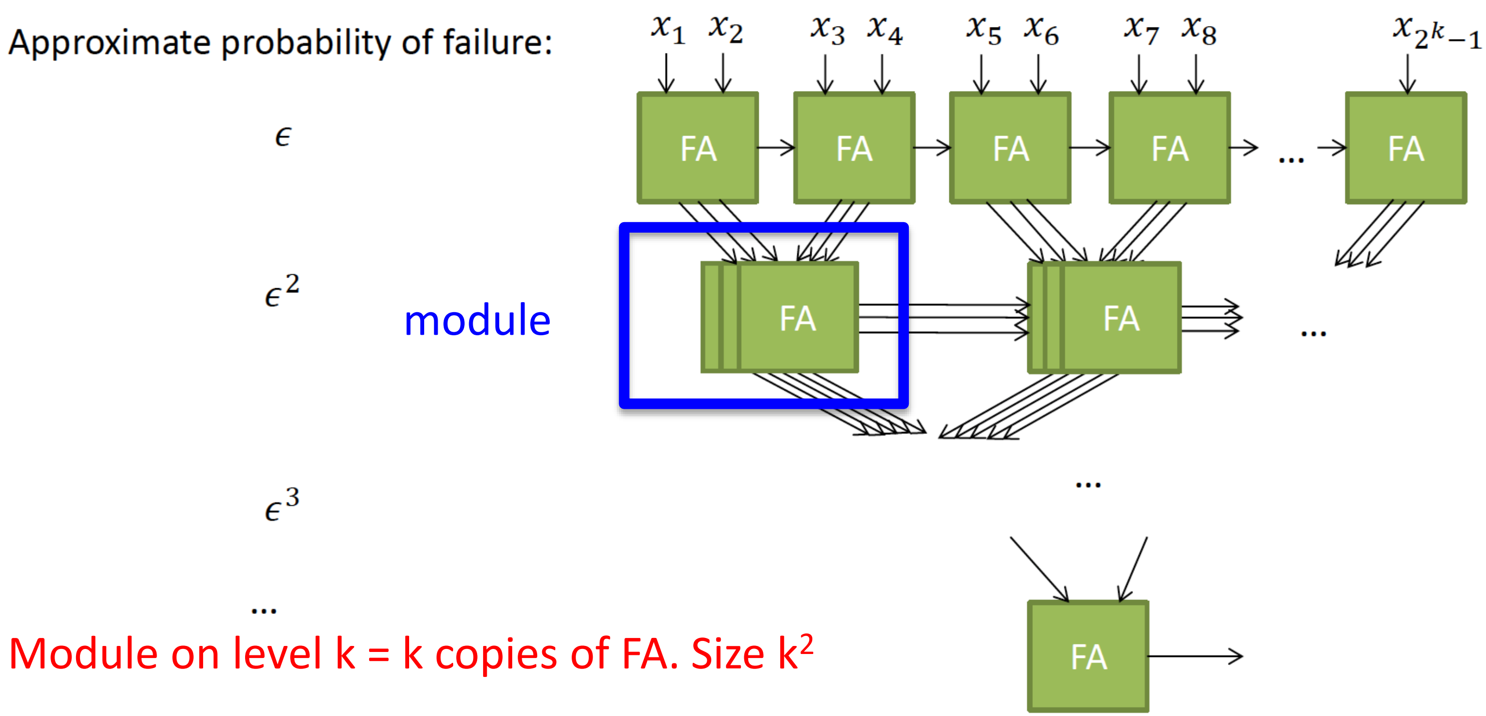




Intuition



Proof: Low Energy Heterogeneous Circuit



- Module on level k = k copies of FA. Size k^2
- ε is small constant
- Number of gates = energy = $\Sigma_{k=1 \dots \log n} (n/2^k)(k^2) \log 1/\varepsilon^k$
- Final level, high energy

Future

- Establish a more complete understanding of the power of heterogeneous supply voltages

- Find good algorithmic problems involving tradeoffs between energy and fault-tolerance in other settings.

Covered Papers: Thanks to my collaborators

- A. Antoniadis, N. Barcelo, M. Nugent, K. Pruhs, and M. Scquizzato. *Energy-efficient circuit design*. 5th conference on Innovations in Theoretical Computer Science (ITCS 2014).

- A. Antoniadis, N. Barcelo, M. Nugent, K. Pruhs, and M. Scquizzato. *Complexity-theoretic obstacles to achieving energy savings with Near-Threshold Computing.* 5th International Green Computing Conference (IGCC 2014).

- N. Barcelo, M. Nugent, K. Pruhs, and M. Scquizzato. *Almost all functions require exponential-energy circuits.* 40th International Symposium on Mathematical Foundations of Computer Science (MFCS 2015).

- N. Barcelo, M. Nugent, K. Pruhs, and M. Scquizzato. *The power of heterogeneity in Near-Threshold Computing.* 6th International Green and Sustainable Computing Conference (IGSC 2015).

Thanks for listening!