ASPDAC '2016

Work hard, sleep well - Avoid irreversible IC wearout with proactive rejuvenation

Xinfei Guo, Mircea R. Stan

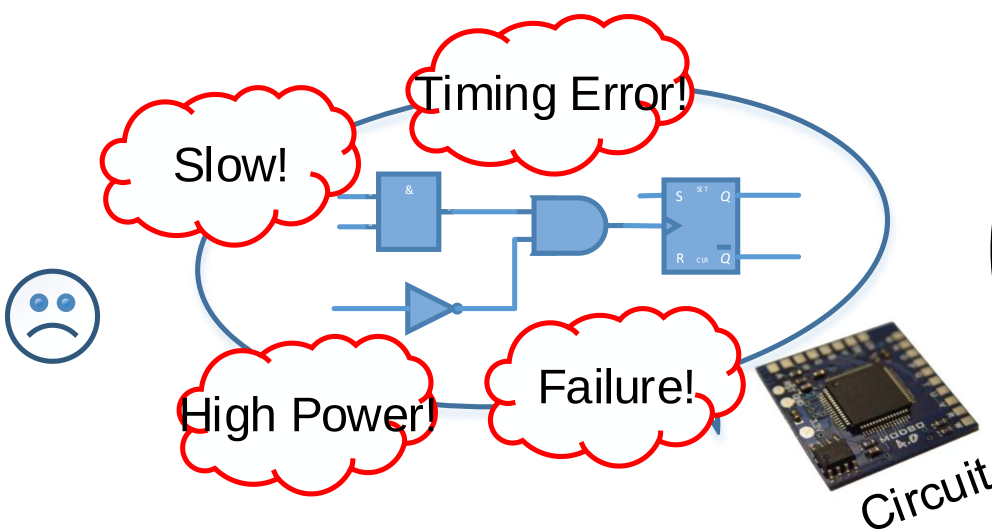
Department of ECE,

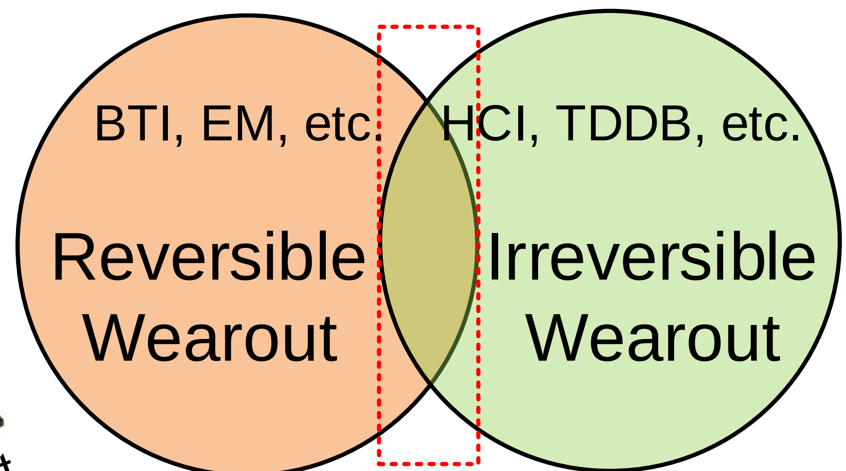
University of Virginia

Jan. 28, 2016

Wearout/Aging

- Front-end of line: BTI, HCI, etc.
- Back-end of line: EM
- A Cross-layer Issue
- Both Reversible and Irreversible part



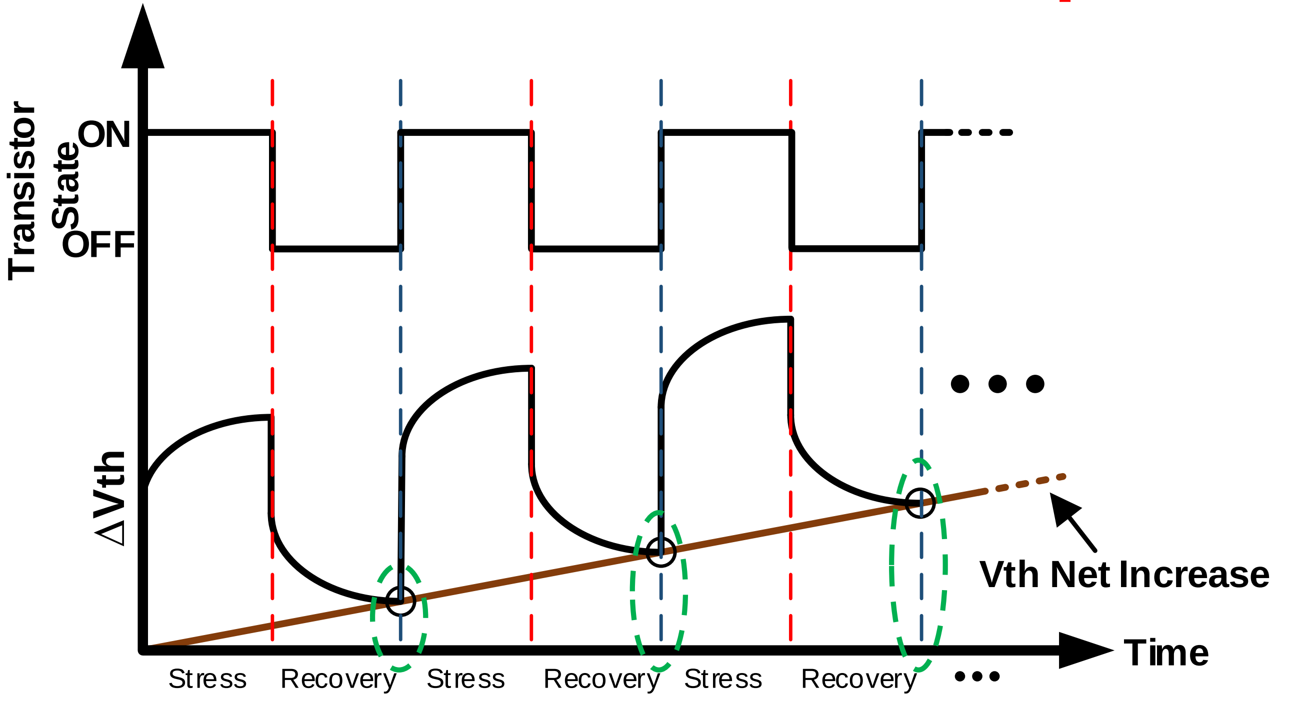


Boundary?

Irreversible Component

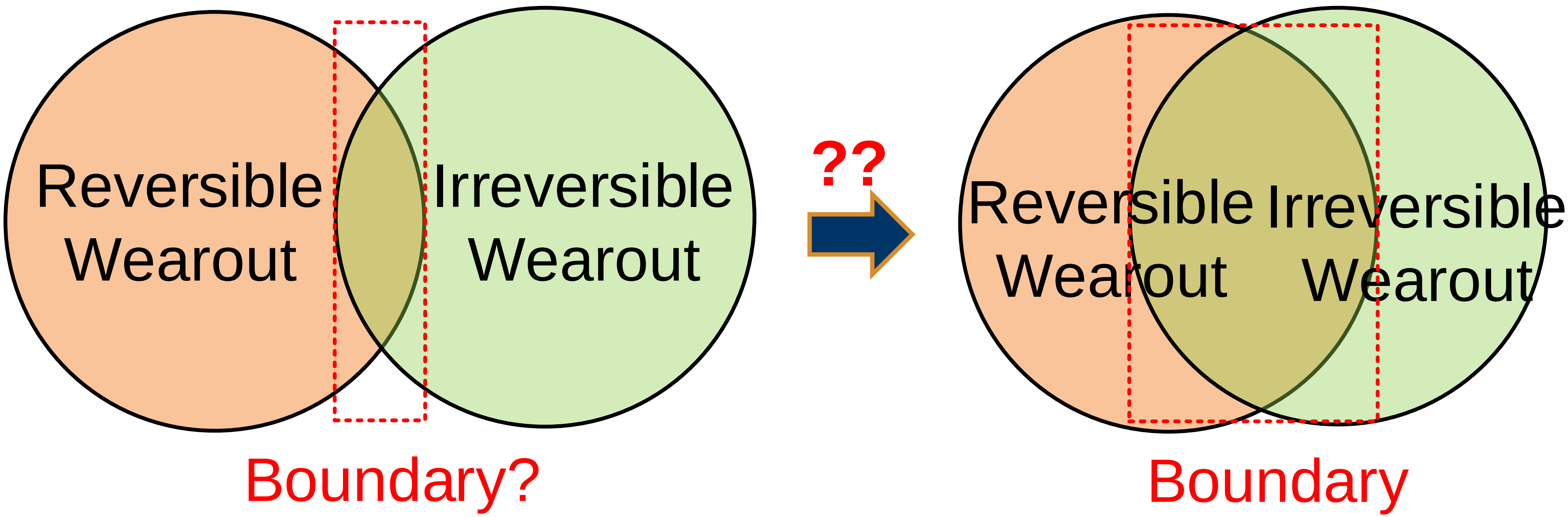
- Biased Temperature Instability (BTI) – Reversible wearout

BUT still with irreversible component



Overview

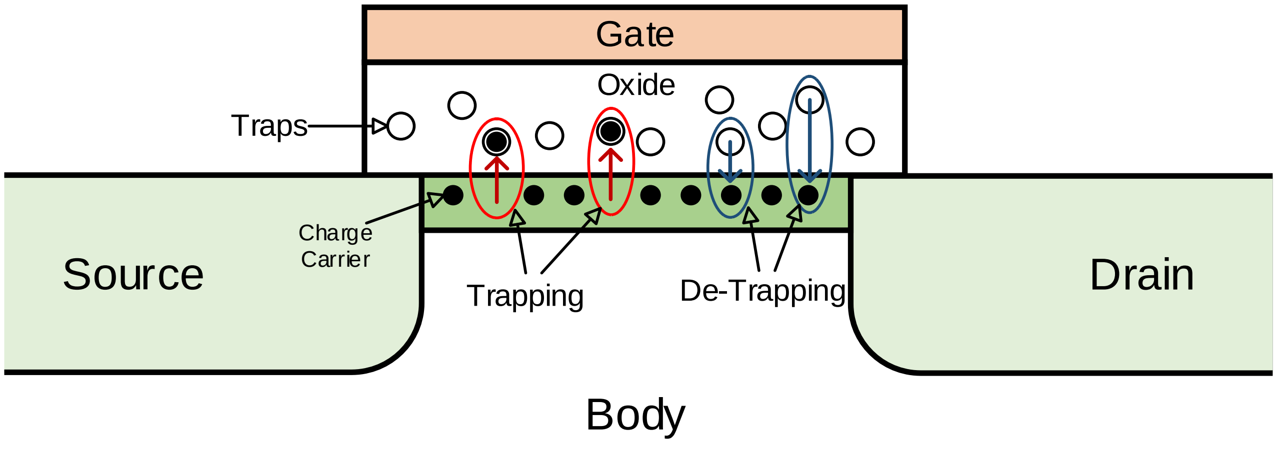
- The boundary is "soft"
- The boundary can be "controlled" & shifted
- The irreversible part can be FULLY avoided



Outline

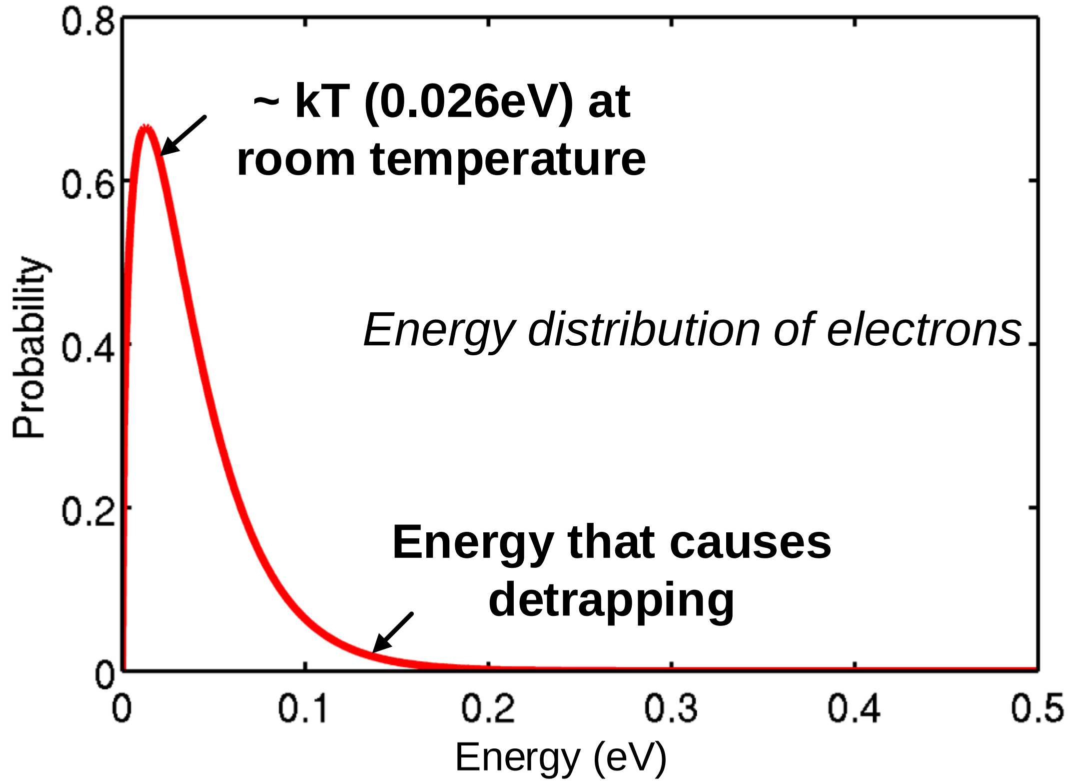
- Overview
- Mechanisms
- Experiments
- Proposed Solution
- Results
- Implementations
- Conclusion

Recovery mechanism (1/2)



- Trapping – Charge carriers overcome a potential barrier
- Detrapping – Trapped charge carriers with a certain probability to escape

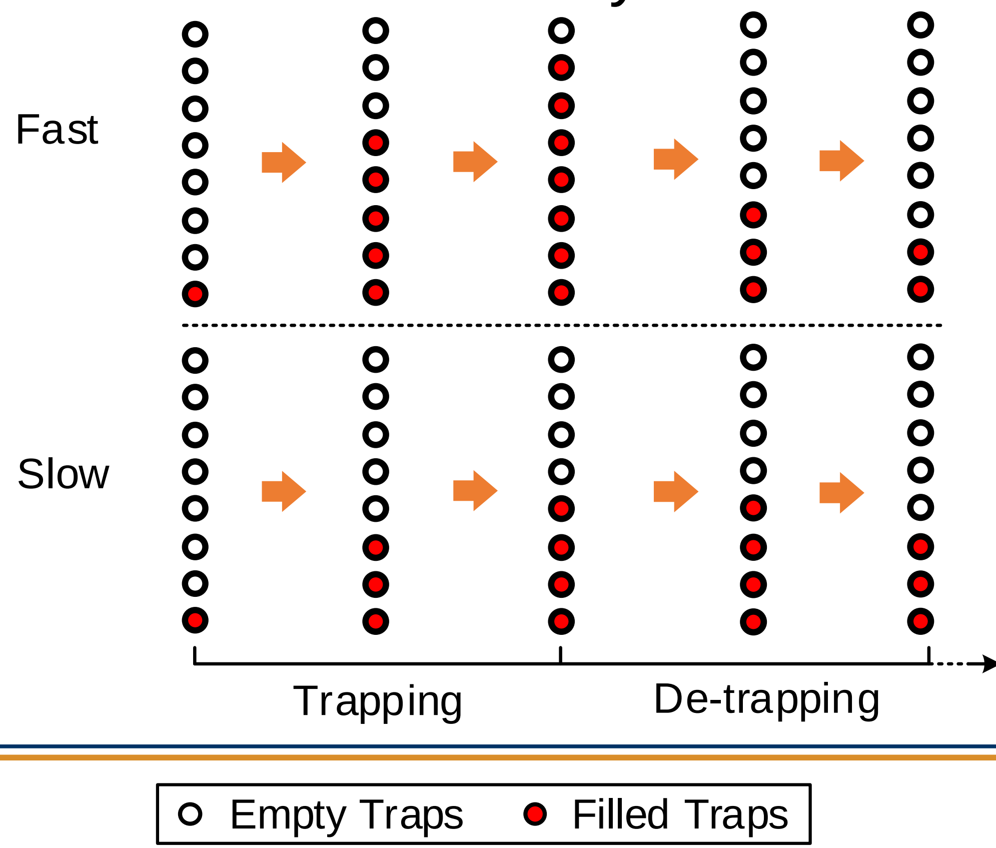
Recovery mechanism (2/2)



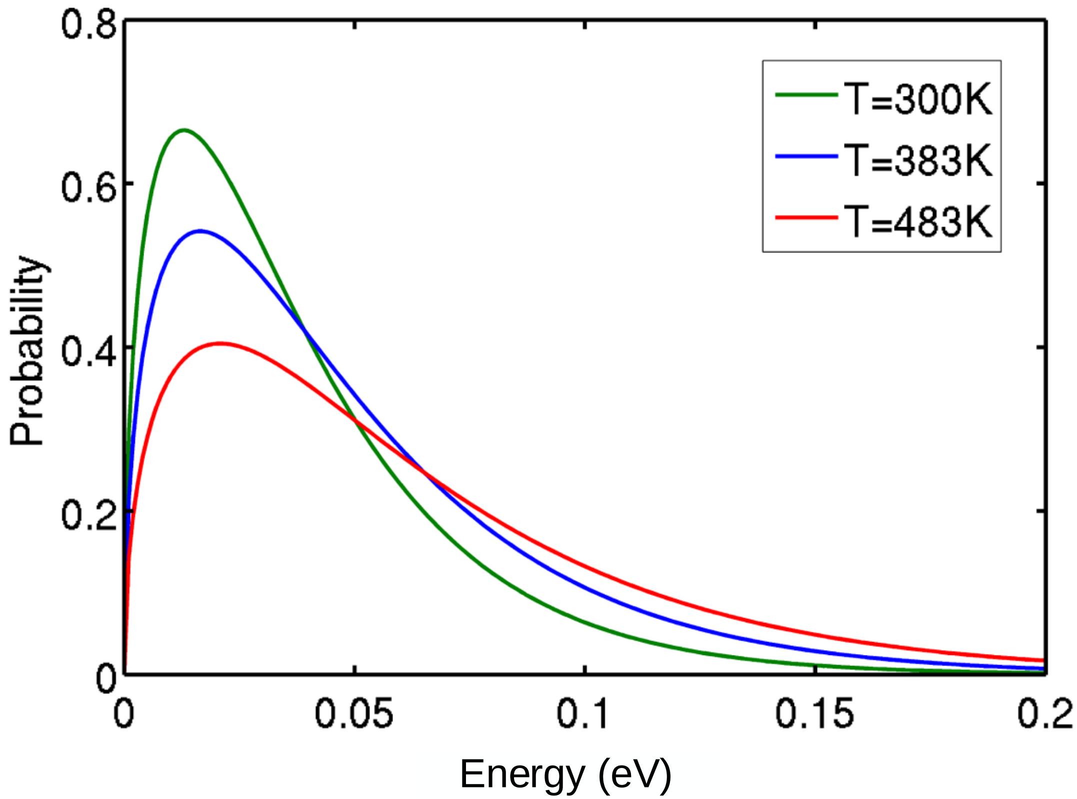
- The probability is high if their energy is higher and the trap energy barrier is lower, and vice-versa.

Fast traps vs. Slow traps

- Fast traps → Lower trap energy barrier → Easier to escape → Fast Recovery → Reversible wearout
- Slow traps → Higher trap energy barrier → Very difficult to escape → Slow/No Recovery → Irreversible wearout



Temperature impact



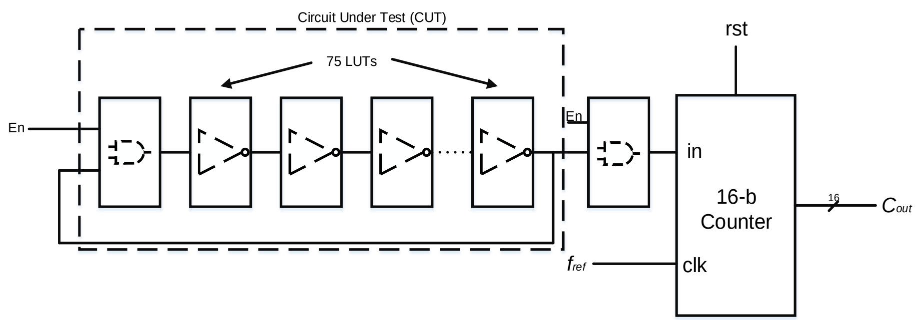
- Temperature can skew the distribution
- Voltage also affects the detrapping via the electrical field

Outline

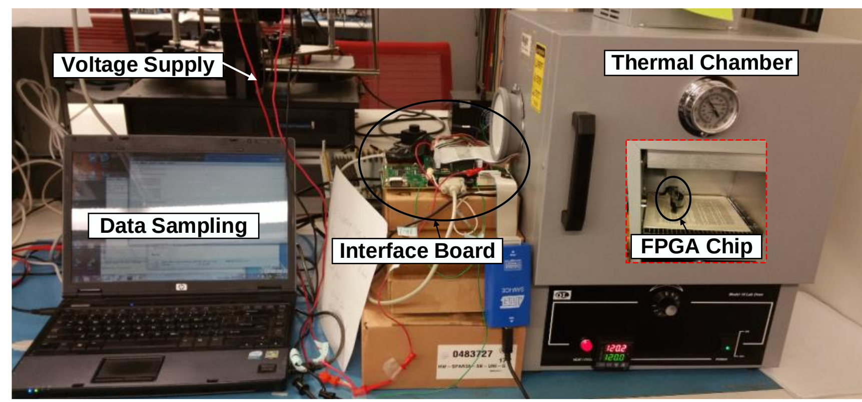
- Motivation
- Mechanisms
- Experiments
- Proposed Solution
- Results
- Implementations
- Conclusion

Experimental Setup

- Accelerated testing methodology
- *40nm* FPGA chips
- Ring Oscillator based test structure
- Measure the oscillation frequency degradation/increase



Test configuration*

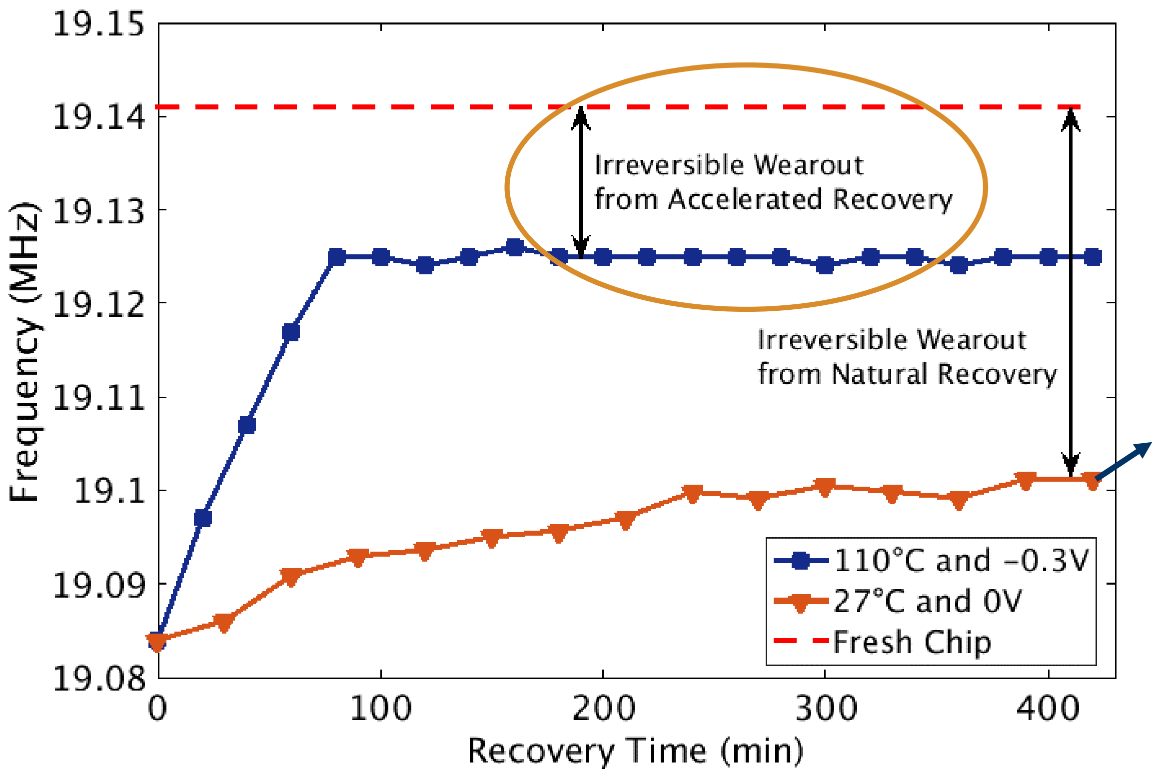


Test setup

* The same test configuration used in [X. Guo et al., DAC ’14]

Accelerated & Active Recovery

- Natural recovery → Passive recovery
- Negative Voltage → Activate Recovery
- High Temperature → Accelerated Recovery

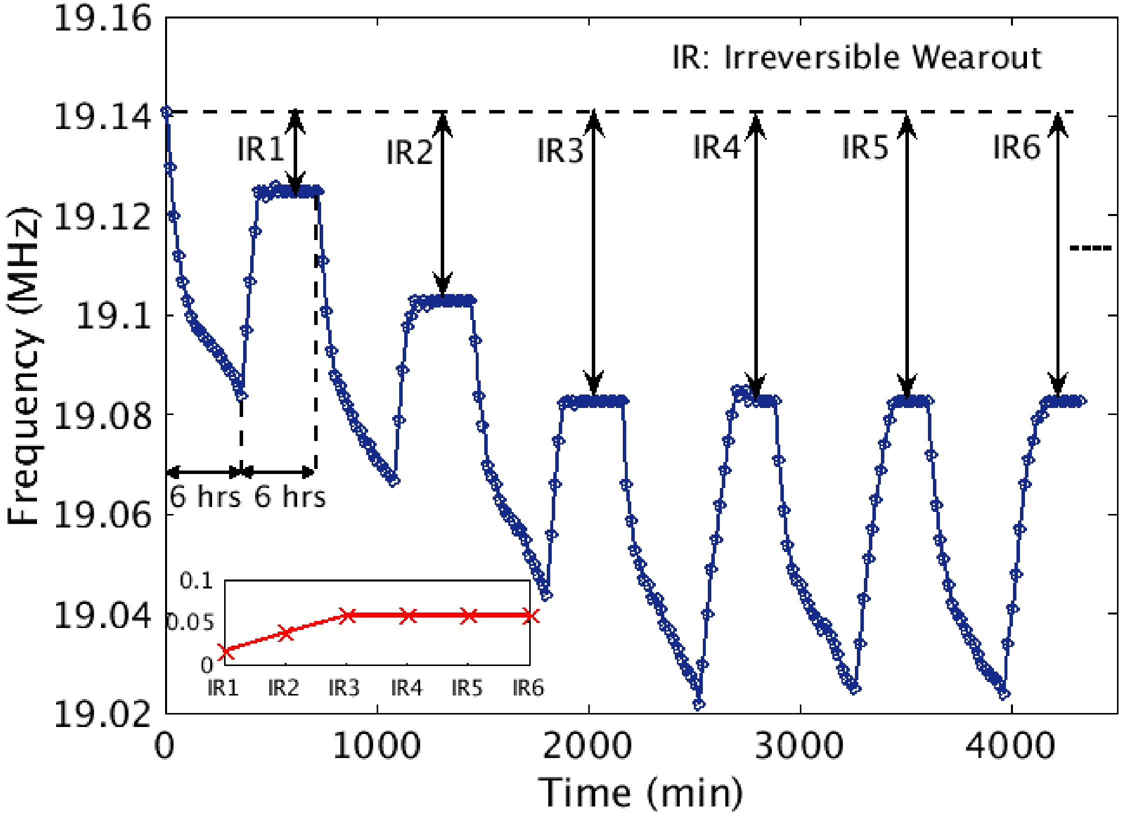


Stay the same for the following days

The boundary is not fixed and is controllable!

Irreversible Wearout During Accelerated & Active Recovery

- Recovery saturates in each cycle
- Irreversible wearout accumulates



IRx: Irreversible Wearout after xth cycle

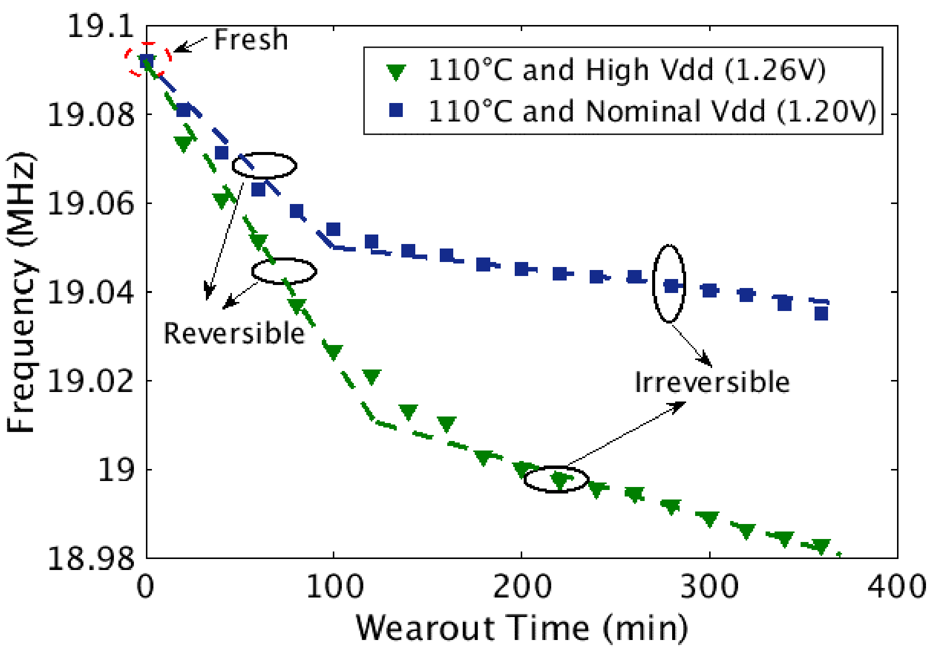
Can we further "remove" or "avoid "all IRs?

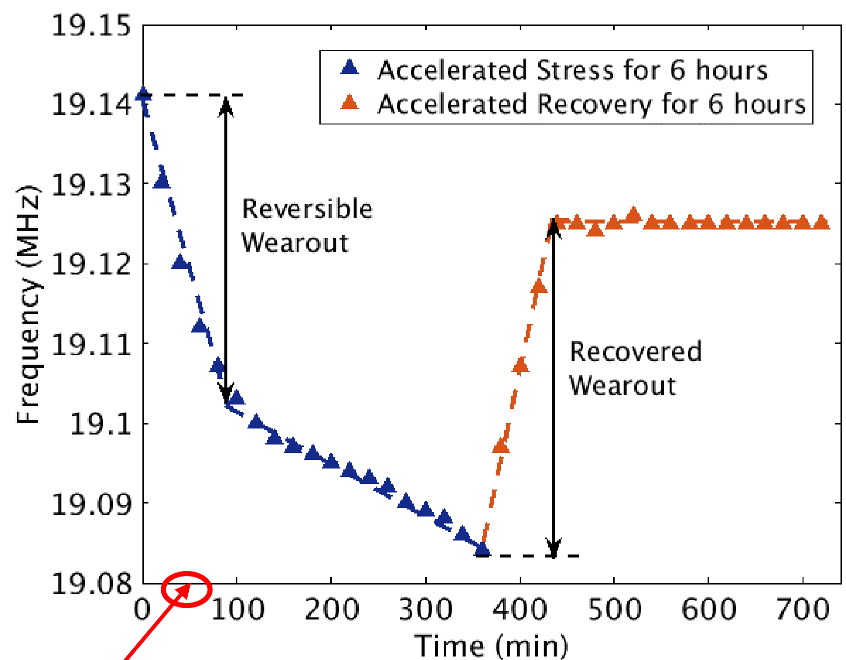
Outline

- Motivation
- Mechanisms
- Experiments
- Proposed Solution
- Results
- Implementations
- Conclusion

Sequentiality of reversible and irreversible wearout

- Irreversible wearout follows reversible wearout
- Accelerated Active Recovery saturates

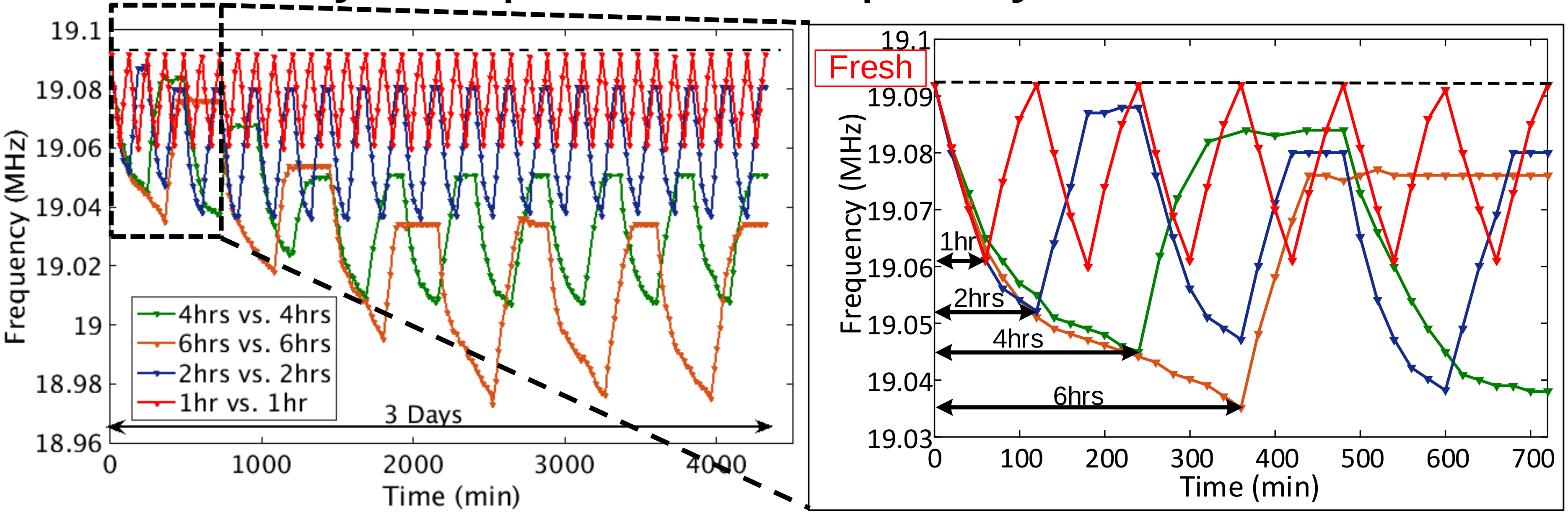




What if we apply the accelerated recovery earlier?

Sleep when getting tired

- Frequency dependency of wearout and recovery
- For *1hr. vs. 1hr.* case, wearout and accelerated recovery compensate completely!

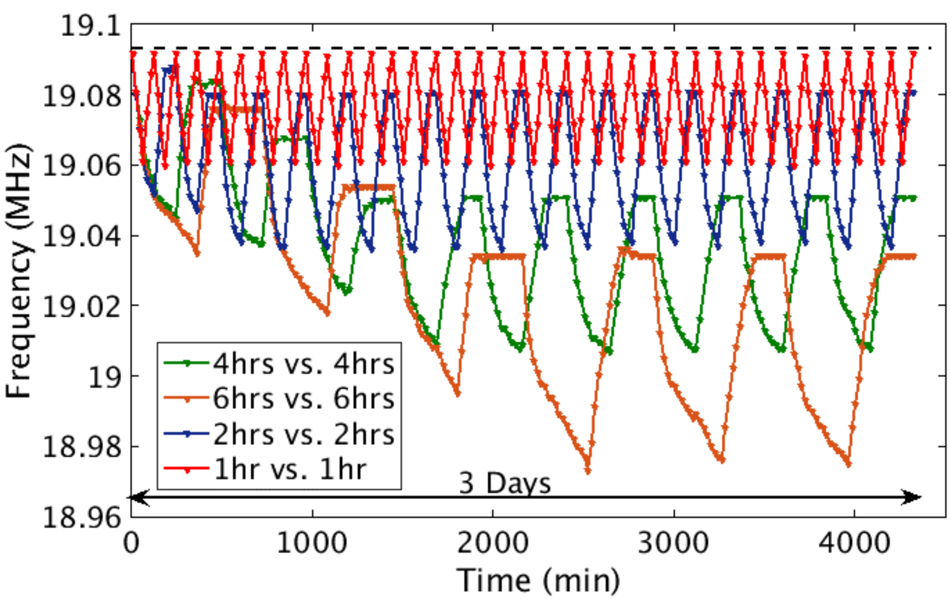


Different “circadian rhythms”

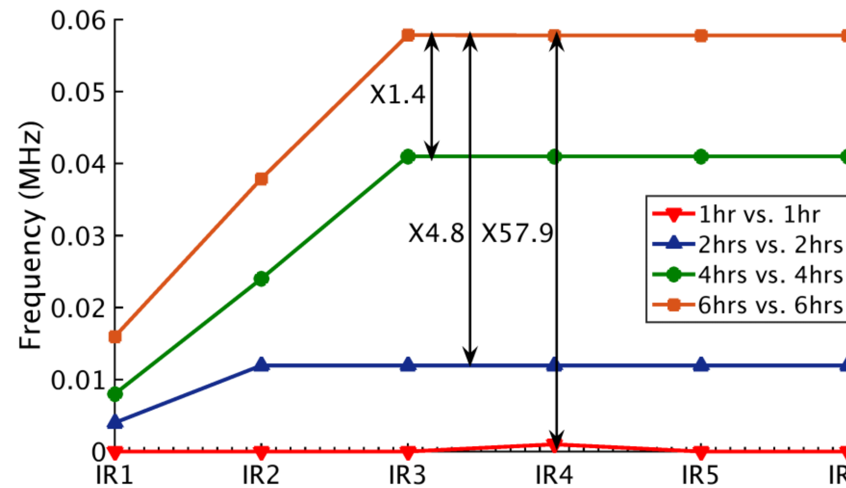
1 hr. Accelerated Wearout ↔ 31 hrs. Normal Operation

Sleep when getting tired

- Frequency dependency of wearout and recovery
- For *1hr. vs. 1hr.* case, wearout and accelerated recovery compensate completely!



Different “circadian rhythms”



Irreversible wearout for the first 6 cycles

1 hr. Accelerated Wearout ↔ 31 hrs. Normal Operation

What does this mean?

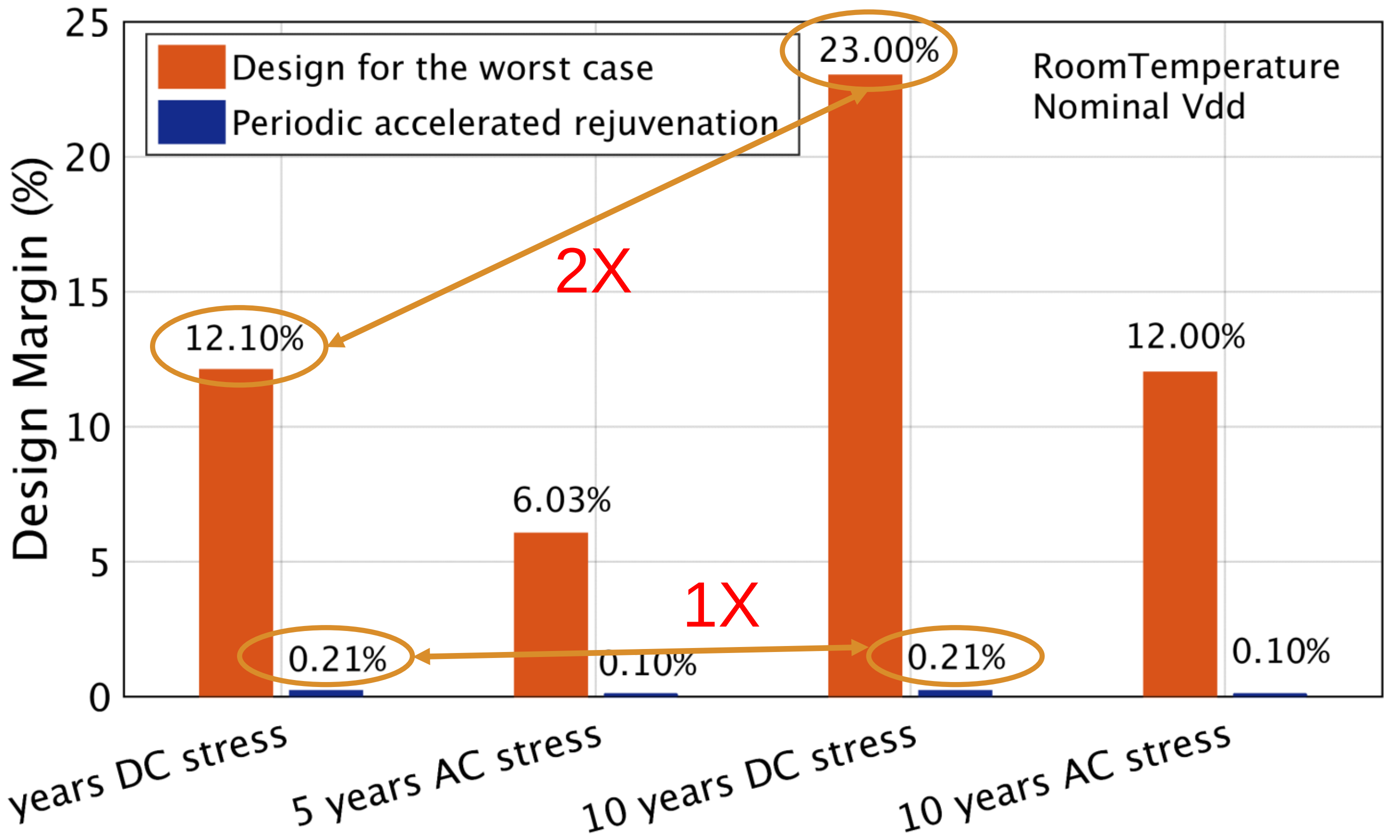
- Irreversible Wearout is completely avoided!
- Operation time $\leq$ 31 hours, and then followed by $\geq$ 1 hour of *Accelerated Active Recovery*
- Reduction of Design Margin (Guardband)
- Higher Average Performance $\rightarrow$ Higher levels of performance and power efficiency most of the lifetime

Outline

- Motivation
- Mechanisms
- Experiments
- Proposed Solution
- Results
- Implementations
- Conclusion

Reduction of Design Margin

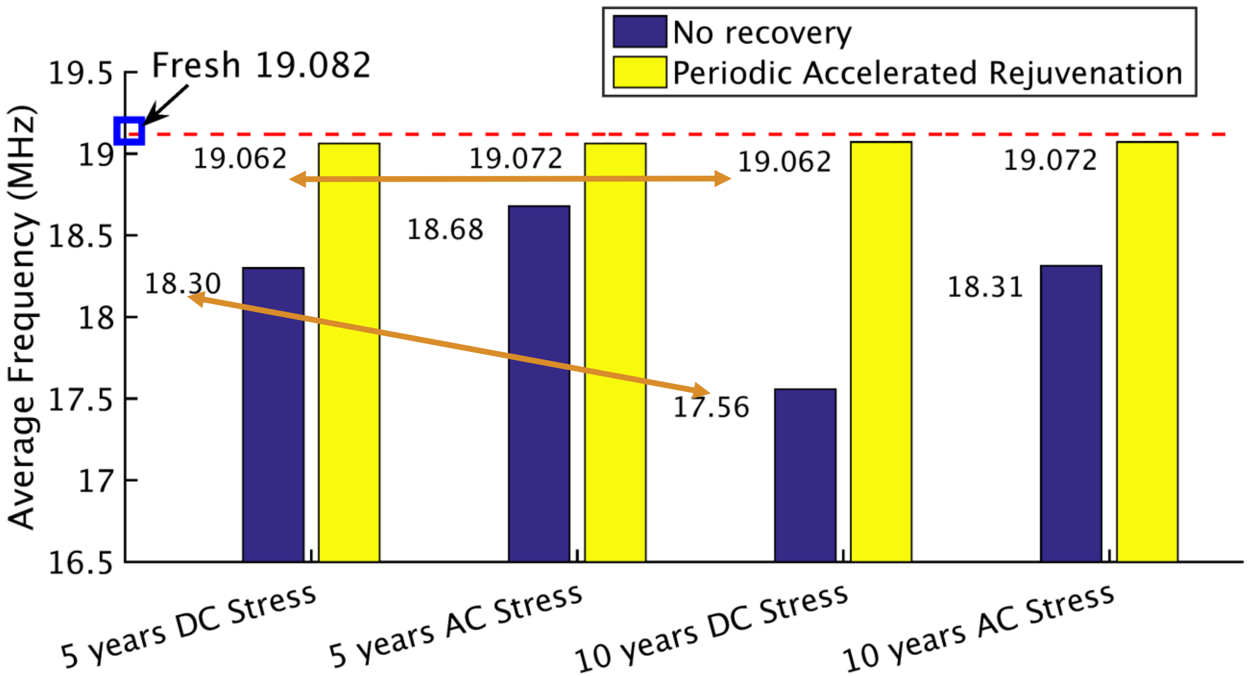
- >60X Reduction for all cases *
- Almost the same margin for any lifetime constraint



* Modeled based on the device wearout model in [Y. Cao, et al. TCAD ’14] and [V. Huard, et al. Springer ’15]

Performance Improvement

- The average performance is close to the fresh during the whole lifetime
- The average performance doesn’t scale with the increase of the lifetime constraint



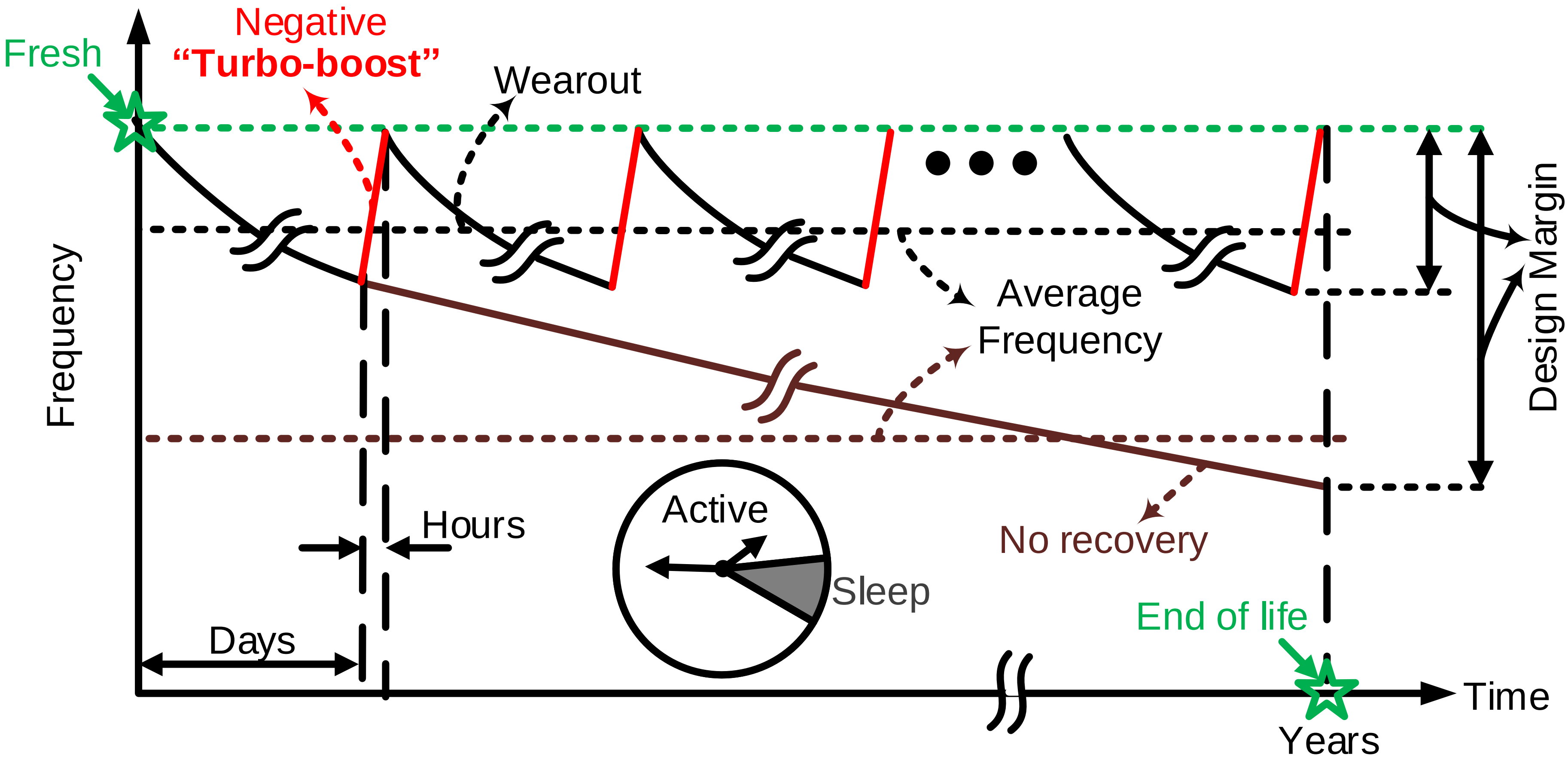
* Modeled based on the device wearout model in [Y. Cao, et al. TCAD ’14] and [V. Huard, et al. Springer ’15]

Outline

- Motivation
- Mechanisms
- Experiments
- Proposed Solution
- Results
- Implementations
- Conclusion

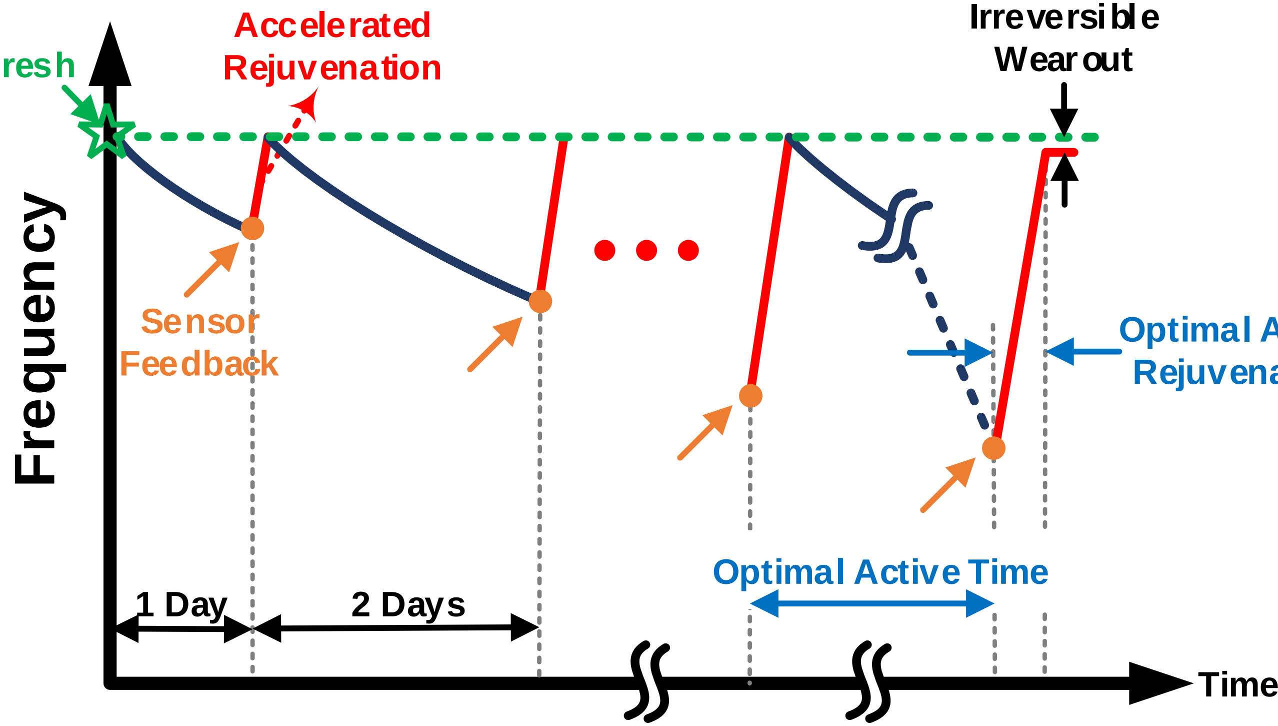
Negative "Turbo Boost"

- Schedule Accelerated Recovery Proactively

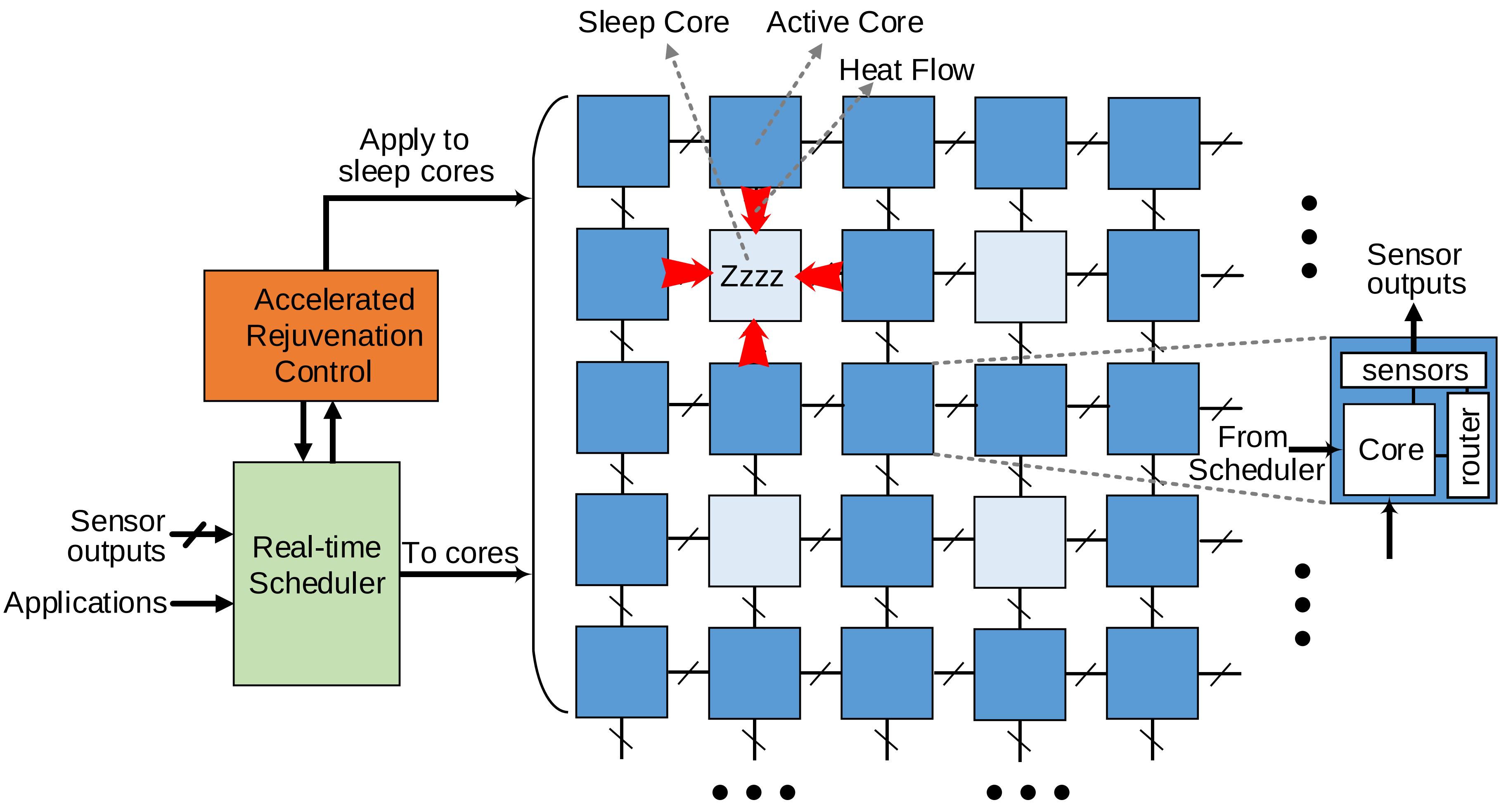


Right balance

- Mobile devices: Human Circadian Rhythms
- Server applications: Utilize core redundancy and employ novel scheduling



The big picture



Conclusion

- Irreversible vs. Reversible Wearout
- Frequency dependency
- *Sleep-when-getting-tired* Strategy
- Reduce guardband & Maintain high performance
- Negative “Turbo-boost”
- Future Work: Optimized scheduling method that considers power, thermal and wearout budgets together

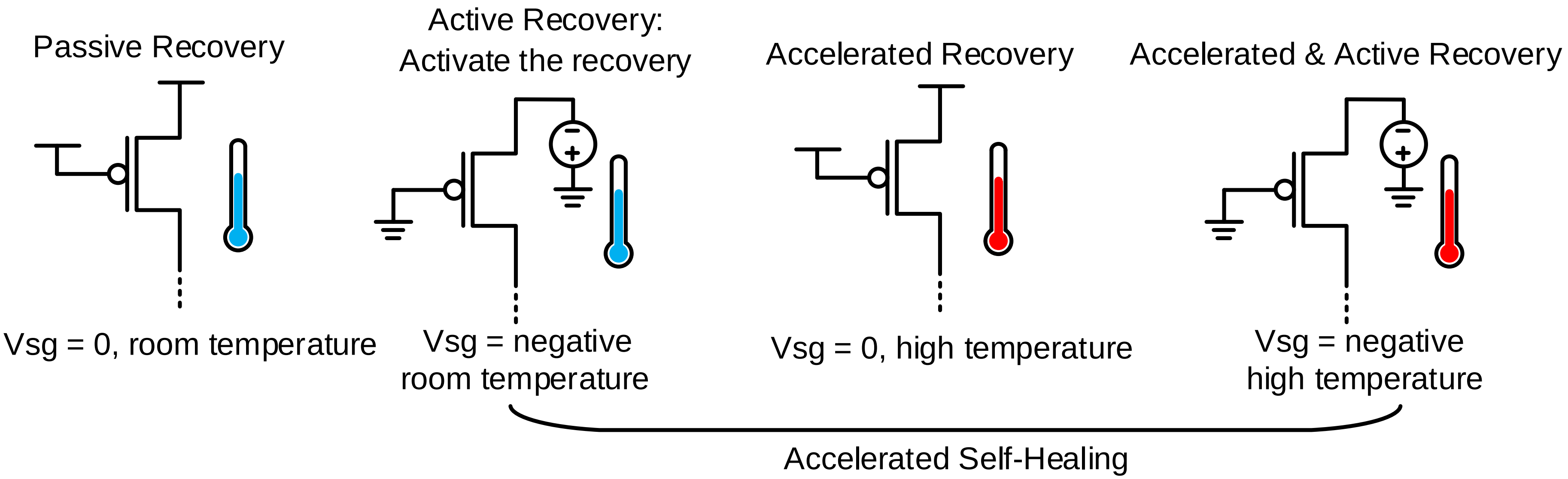
Thank you!

Q & A

This work is funded by NSF, SRC and C-FAR.

Backup Slides

Accelerated Self-Healing

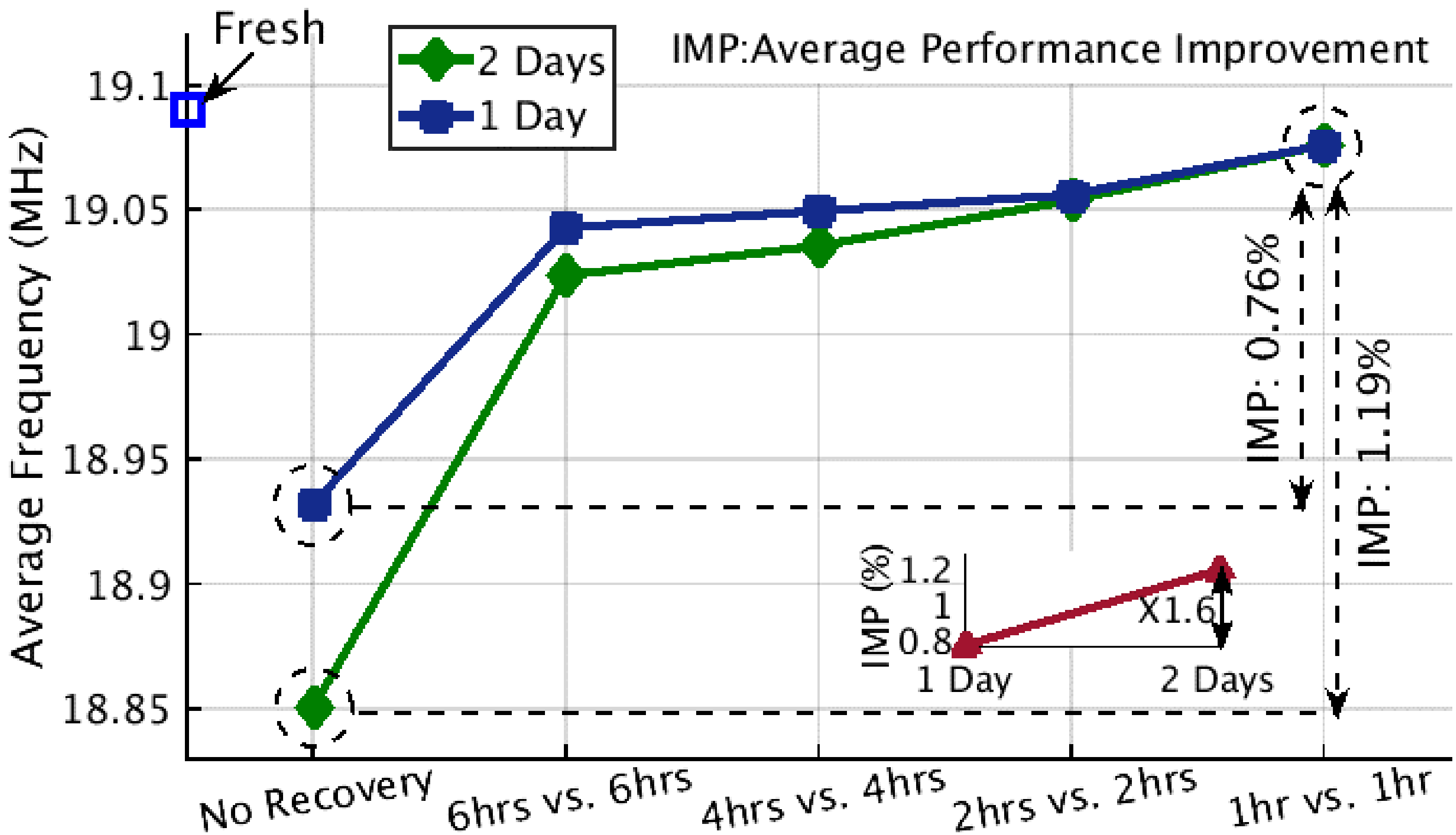


The distribution of kinetic energies

$$f_E(E) = A \times (\frac{1}{kT})^{3/2} \times \sqrt{E} \times \exp(-\frac{E}{kT})$$

- Majority of the electrons are at low energy in *meV* range
- The center energy of even the lowest energy of the trap is in order of several *kT*

Measured Average performance improvement (IMP) for 1 day and 2 days



Test cases

- All start from fresh
- Total test time: 3 days

TABLE I Summary of periodic accelerated rejuvenation test cases

Case Name	Chip No.	Cycle stress time	Cycle accelerated recovery time	# of cycles
6 hrs vs. 6 hrs	1	6 hours	6 hours	6
4 hrs vs. 4 hrs	2	4 hours	4 hours	9
2 hrs vs. 2 hrs	3	2 hours	2 hours	18
1 hr vs. 1 hr	4	1 hours	1 hours	32