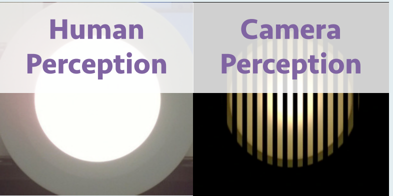
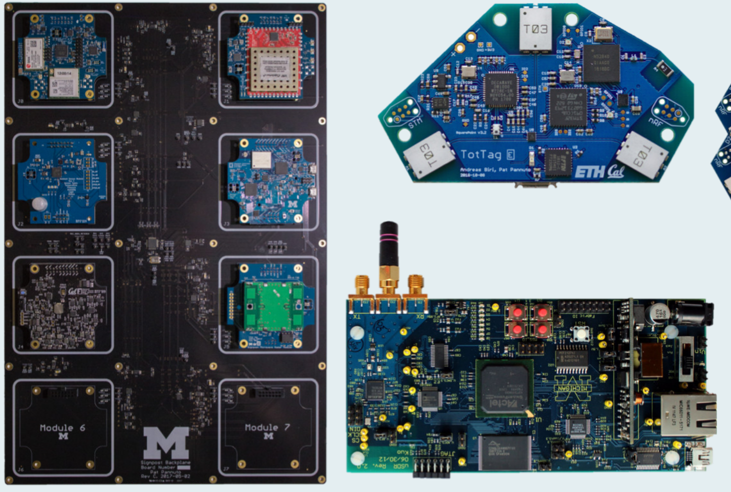
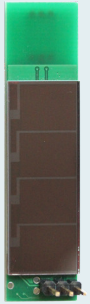
A Modular Platform for Nanopower Computing

Pat Pannuto, UC Berkeley
https://patpannuto.com ppannuto@berkeley.edu

University of California, San Diego
April 24, 2019







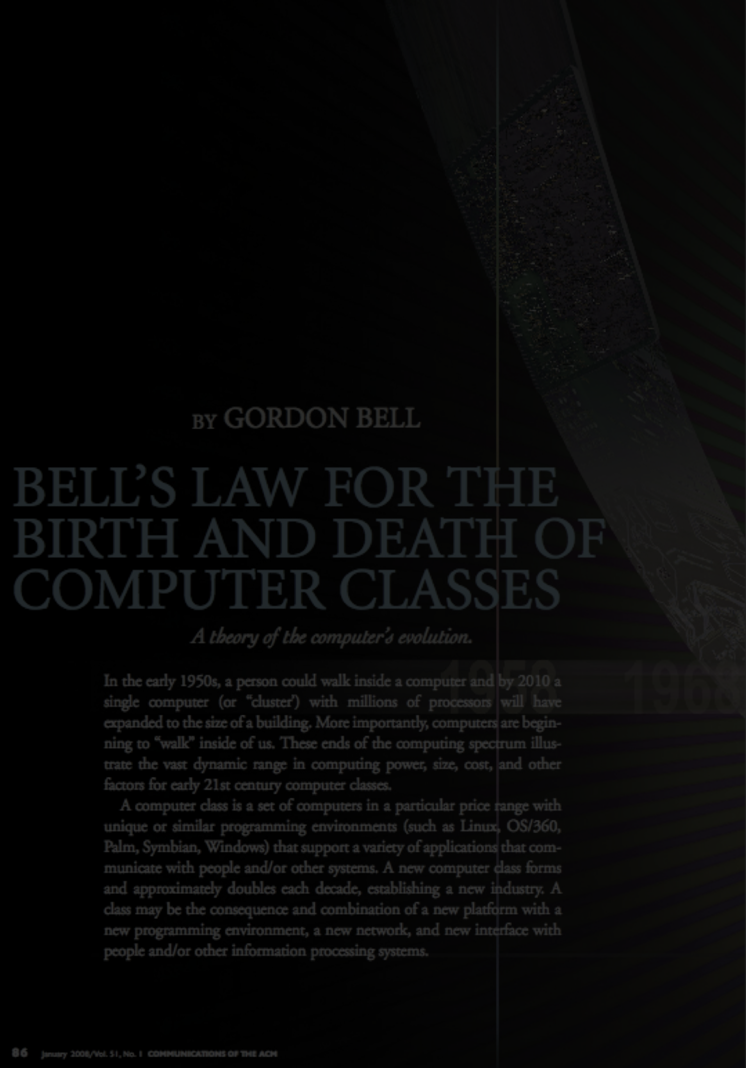
“A computer class is…a

new <u>platform</u> with a

new <u>programming environment</u>, a

new <u>network</u>, and

new <u>interface</u>”



BY GORDON BELL

BELL’S LAW FOR THE BIRTH AND DEATH OF COMPUTER CLASSES

A theory of the computer’s evolution.

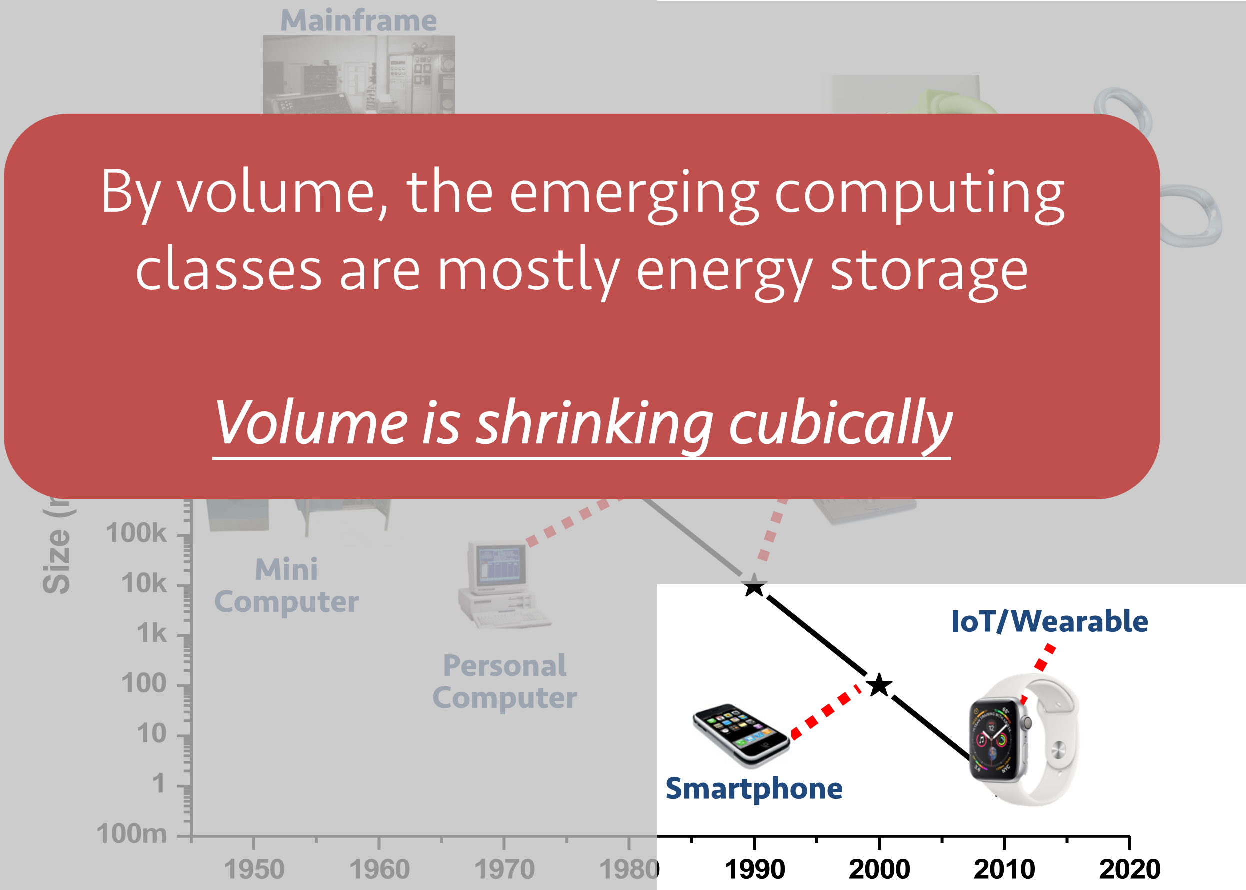
In the early 1950s, a person could walk inside a computer and by 2010 a single computer (or “cluster”) with millions of processors will have expanded to the size of a building. More importantly, computers are beginning to “walk” inside of us. These ends of the computing spectrum illustrate the vast dynamic range in computing power, size, cost, and other factors for early 21st century computer classes.

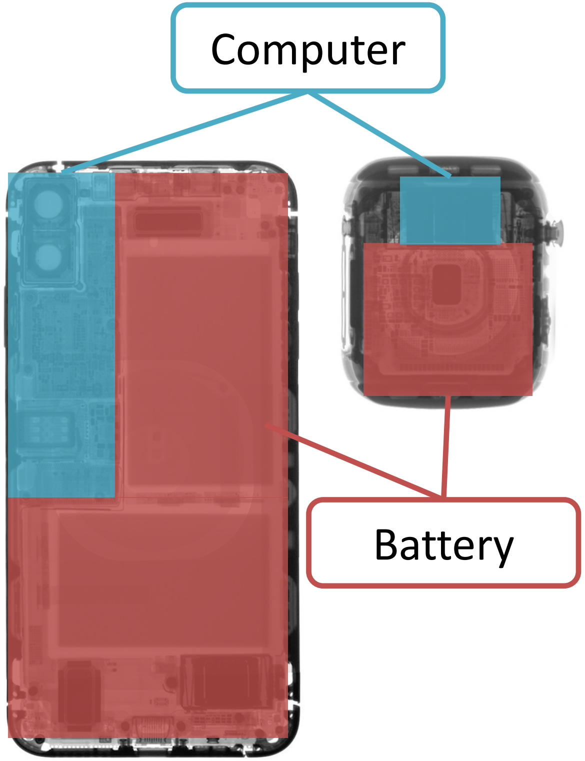
A computer class is a set of computers in a particular price range with unique or similar programming environments (such as Linux, OS/360, Palm, Symbian, Windows) that support a variety of applications that communicate with people and/or other systems. A new computer class forms and approximately doubles each decade, establishing a new industry. A class may be the consequence and combination of a new platform with a new programming environment, a new network, and new interface with people and/or other information processing systems.

Bell's Law captures the evolution of computing platforms

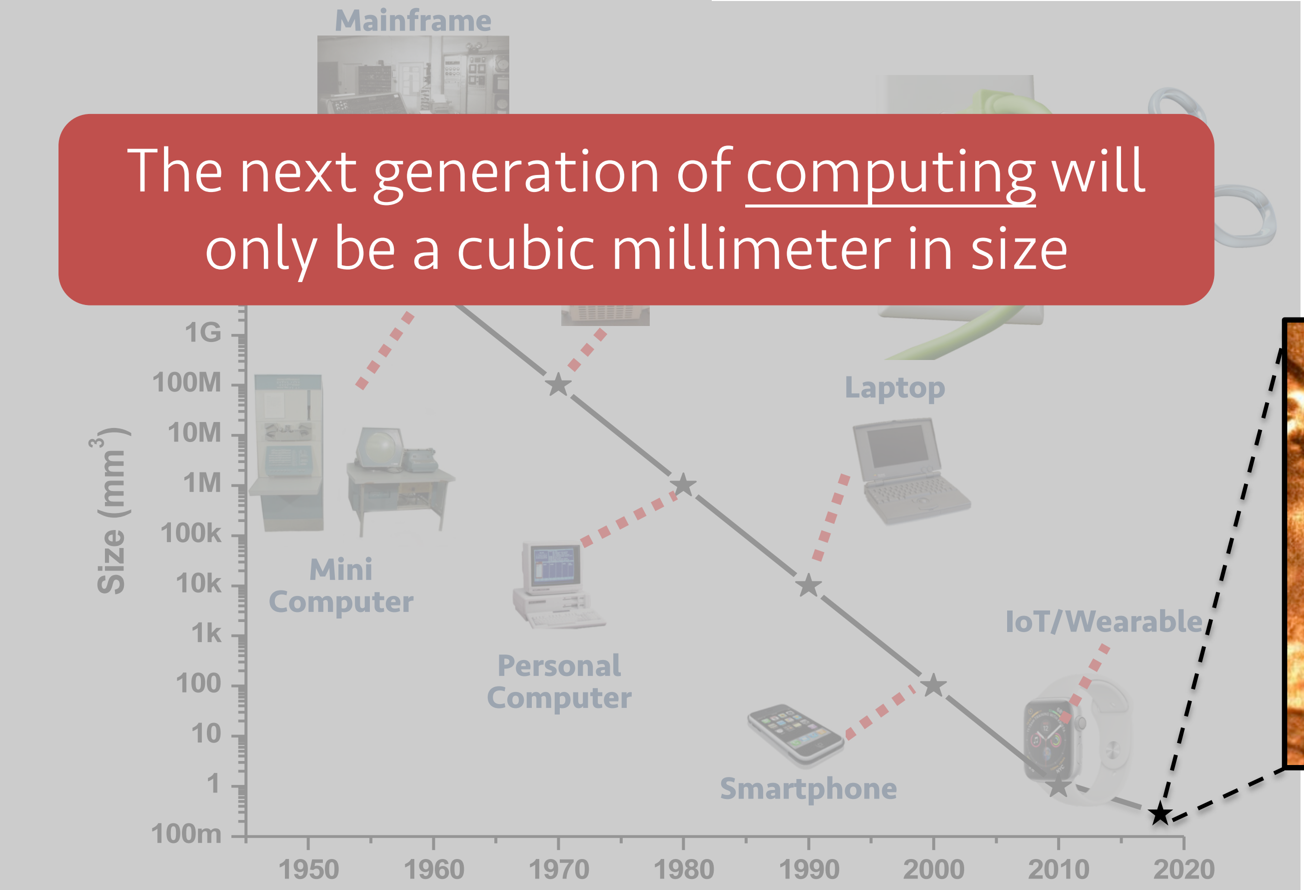
By volume, the emerging computing classes are mostly energy storage

Volume is shrinking cubically

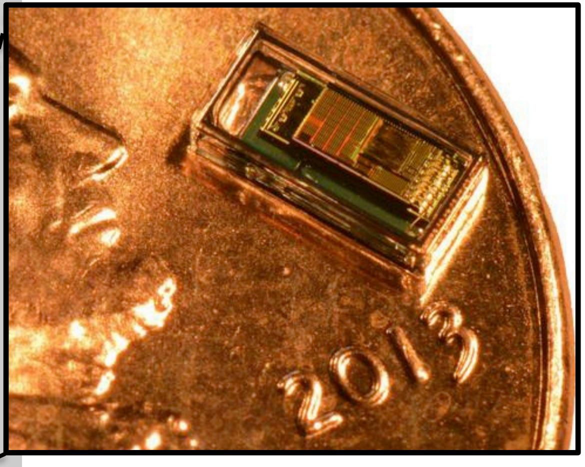




Computational platforms will continue to scale

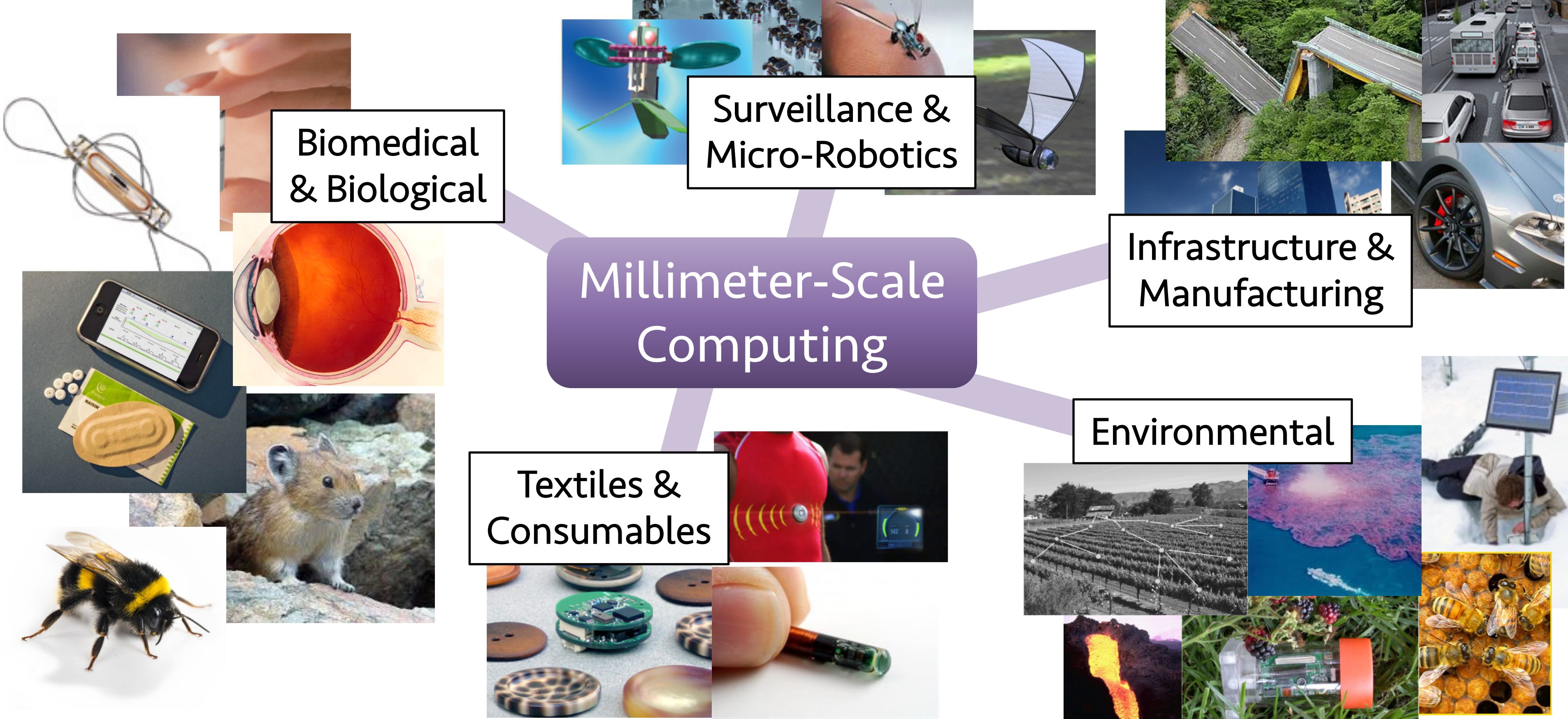


The next generation of computing will only be a cubic millimeter in size

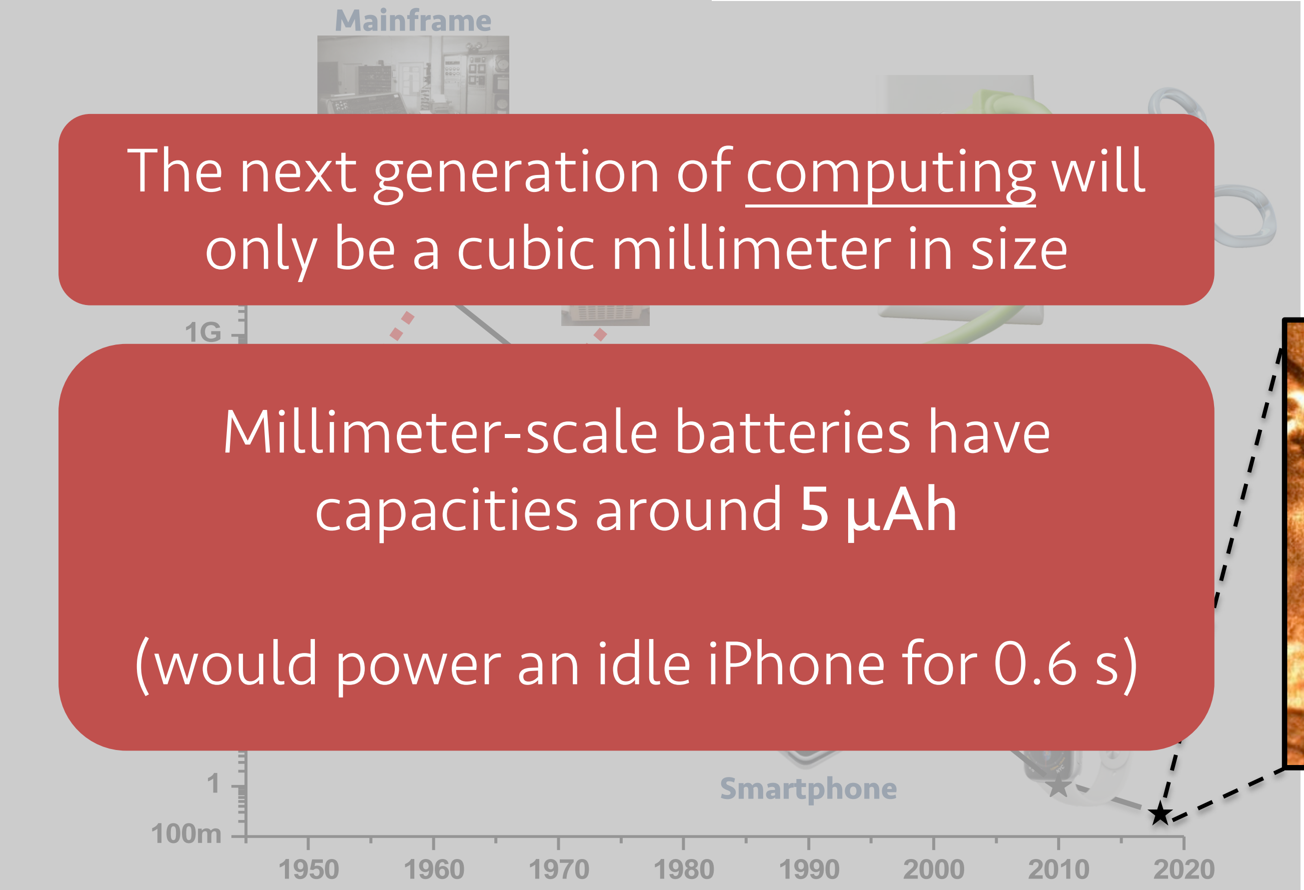


"Smart Dust"

Millimeter-scale form factor is key to opening a wide array of new applications



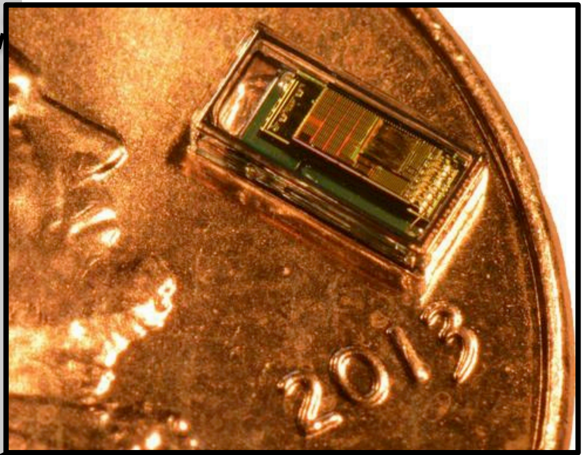
Computational platforms will continue to scale



The next generation of computing will only be a cubic millimeter in size

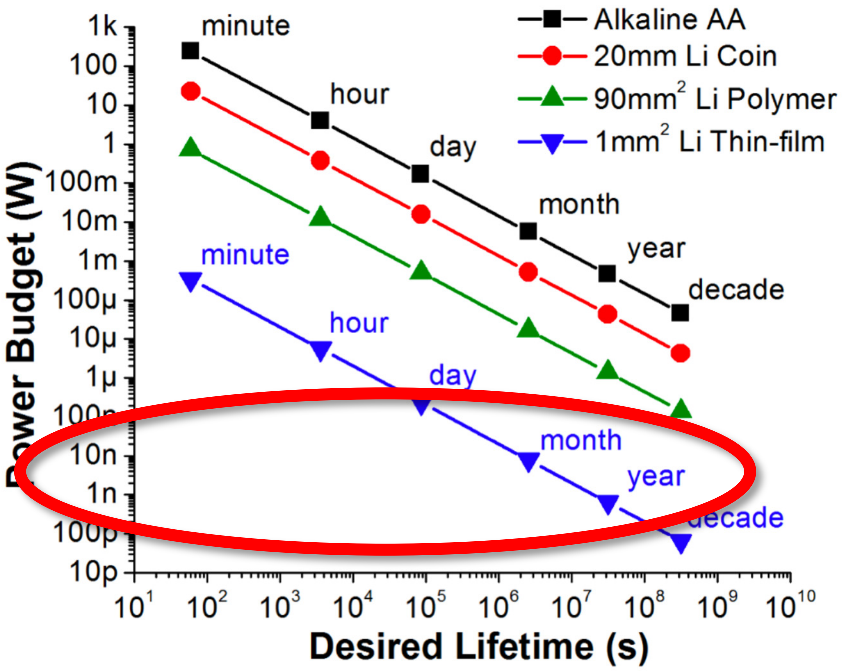
Millimeter-scale batteries have capacities around **5 µAh**

(would power an idle iPhone for 0.6 s)



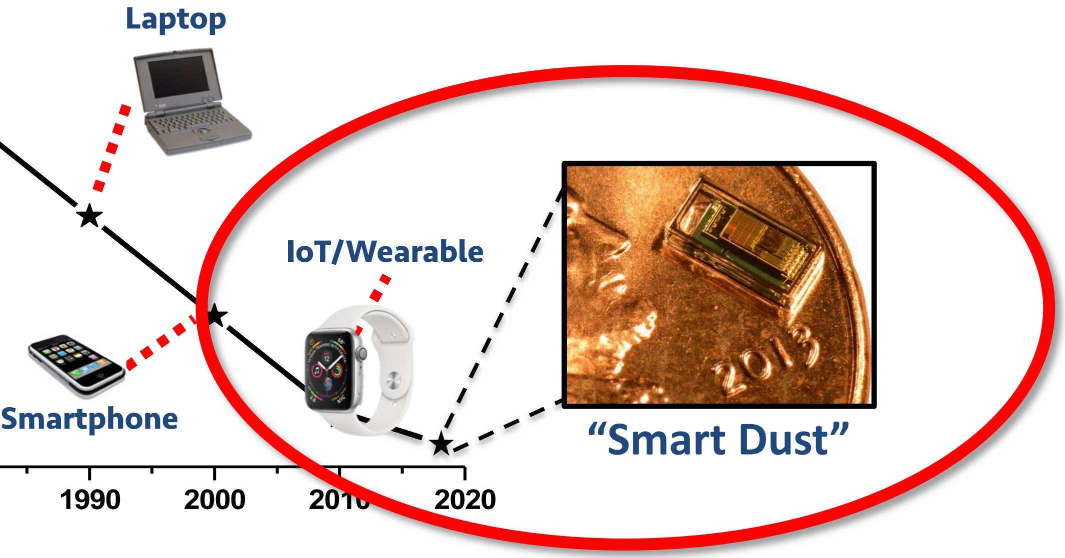
"Smart Dust"

To support their target applications, Smart Dust systems must last longer on less energy



IoT Design Space Challenges: Circuits and Systems
David Blaauw, Dennis Sylvester, Prabal Dutta, Yoonmyung Lee, Inhee Lee, Sechang Bang, Yejoong Kim, Gyouho Kim, **Pat Pannuto**, Ye-Sheng Kuo, Dongmin Yoon, Wanyeong Jung, ZhiYoong Foo, Yen-Po Chen, Jeong Seok-Hyeon, and Myungjoon Choi
Proceedings of the 2014 IEEE Symposium on VLSI Technology (VLSI'14)

Energy constraints will play a central role in the evolution of computing platforms



How must traditional paradigms change, adapt, or re-invent for the new computing classes?

Outline

- Introducing the nanopower computing class

- The development of a modular, nanopower computing platform

  Hardware — Composing micro-scale systems

  Software — Returning modularity to microcontroller software

  Services — Deployment and management of “dust”

- The future of nanopower computing and its broader impact

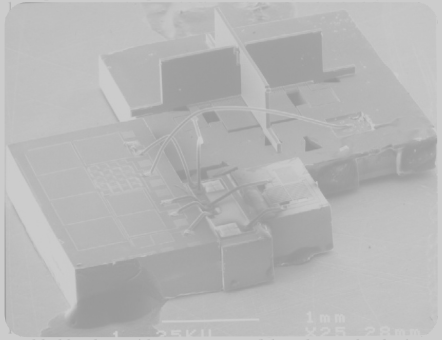
A very brief history of “Smart Dust”

Monolithic: Every sensing system requires a completely new chip

{jmk, randy, pister}@eecs.berkeley.edu

Abstract

Large-scale networks of wireless sensors are becoming an active topic of research. Advances in hardware technology and engineering design have led to dramatic reductions in size, power consumption and cost for digital circuitry, wire-



enabled by the rapid convergence of three key technologies: digital circuitry, wireless communications, and Micro ElectroMechanical Systems (MEMS). In each area, advances in hardware technology and engineering design have led to reductions in size, power consumption, and cost. This has enabled remarkably compact, autonomous nodes, each containing one or more sensors, computation and communication capabilities, and a power supply.

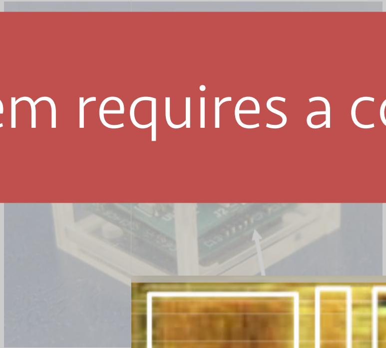
Berkeley’s *Smart Dust* project, led by Professors Pister and Kahn, explores the limits on size and power consumption in

“Smart Dust.” It is certainly within the realm of possibility that future prototypes of Smart Dust could be small enough to remain suspended in air, buoyed by air currents, sensing and communicating for hours or days on end. At least one popular science fiction book has articulated just such a vision [12].

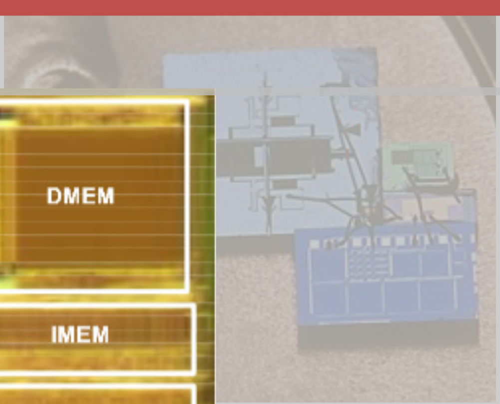
the potential applications of Smart Dust and the challenges they pose. Section 5 discusses related projects from the research community. Section 6 presents our summary and conclusions.

2 Smart Dust Technology

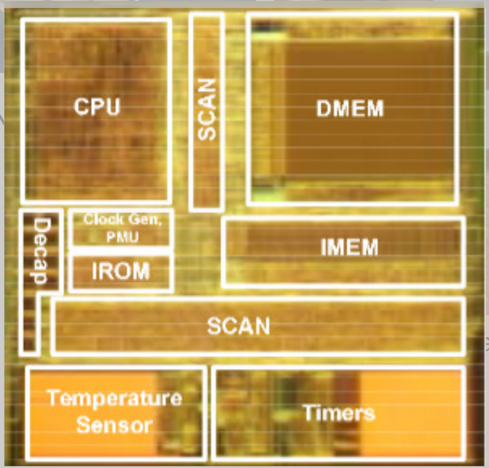
A Smart Dust *mote* is illustrated in Figure 1. Integrated into a single package are MEMS sensors, a semiconductor laser diode and MEMS beam-steering mirror for active optical



T. N

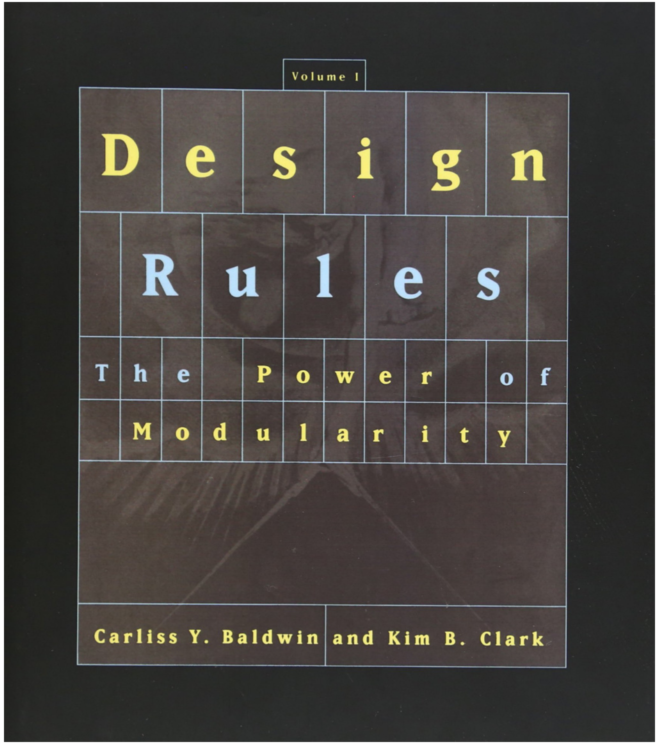


arneke ‘07



Seok ‘08

Claim: Efficient modularity is a necessary condition for the emergence of a computing class

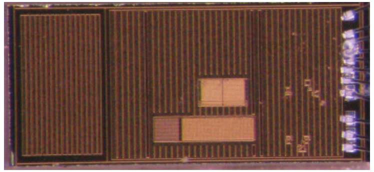


Modularity…

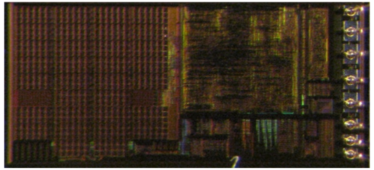
- "increases the range of 'manageable' complexity"
- "allows different parts of a large design to be worked on concurrently"
- "accommodates uncertainty"
- adds costs

Quotes from pages 90-91: What is Modularity?

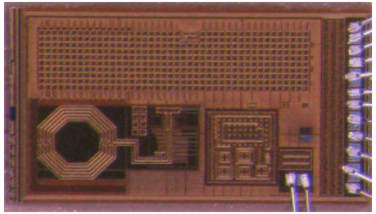
We have modular components…



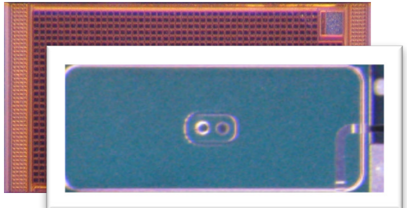
Temperature Sensor
~10 pW standby, < 1 µW active



CPU
~1 nW standby, ~5 µW active

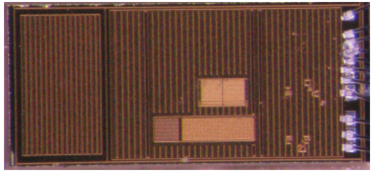


Radio
~10 pW standby, ~10 µW active

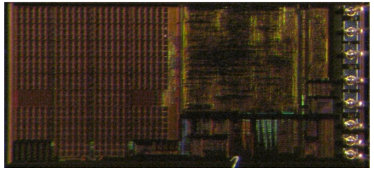


Energy Harvesting & Storage
1~10 nW indoors
2~10 µAh capacity

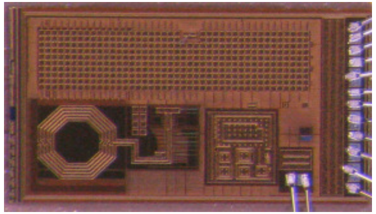
We have modular components… but no suitable, general purpose, common interconnect



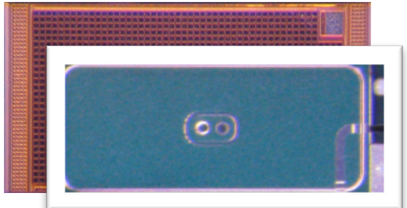
Temperature Sensor
~10 pW standby, < 1 µW active



CPU
~1 nW standby, ~5 µW active



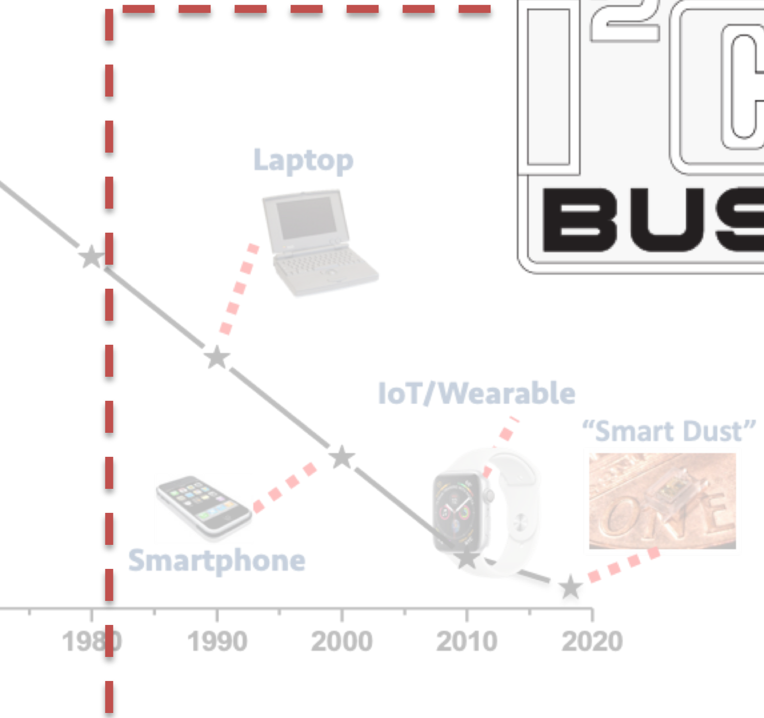
Radio
~10 pW standby, ~10 µW active



Energy Harvesting & Storage
1~10 nW indoors
2~10 µAh capacity

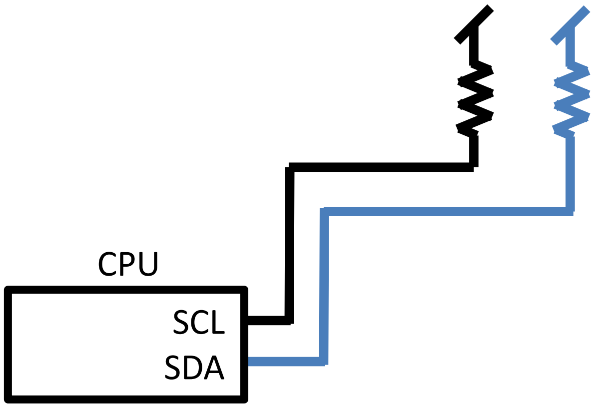
?

What's wrong with existing interchip interconnects? Energy.

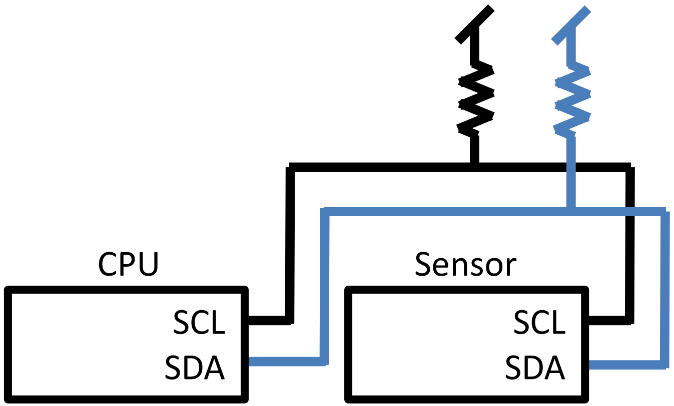


- Inter-Integrated Circuit (I^2C) Bus
  - Fixed wire count
  - Multi-master
  - Compact addressing
  - Hardware acknowledgements
  - Clock stretching
  - Easy voltage level translation
  - …
- Too energy-inefficient for Smart Dust

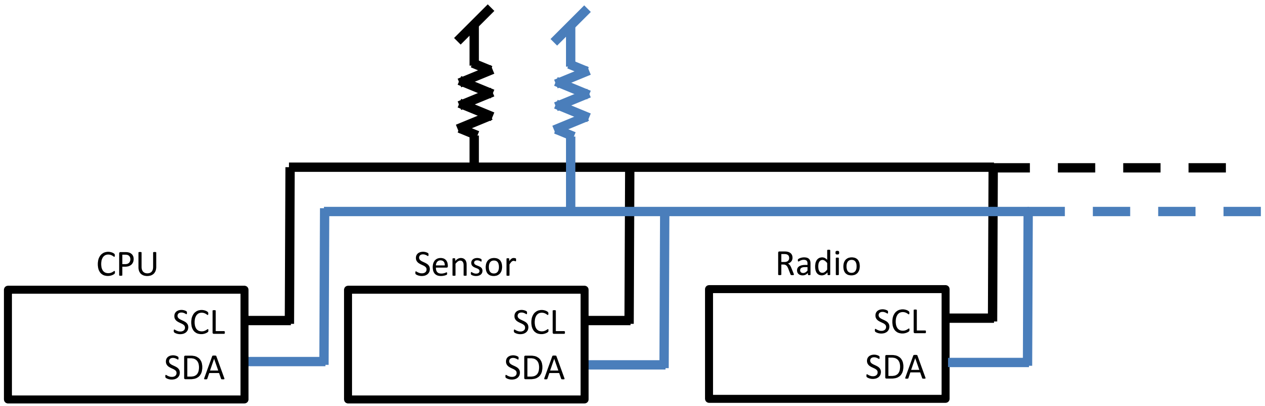
I^{2}C has fixed I/O requirements and a decentralized architecture



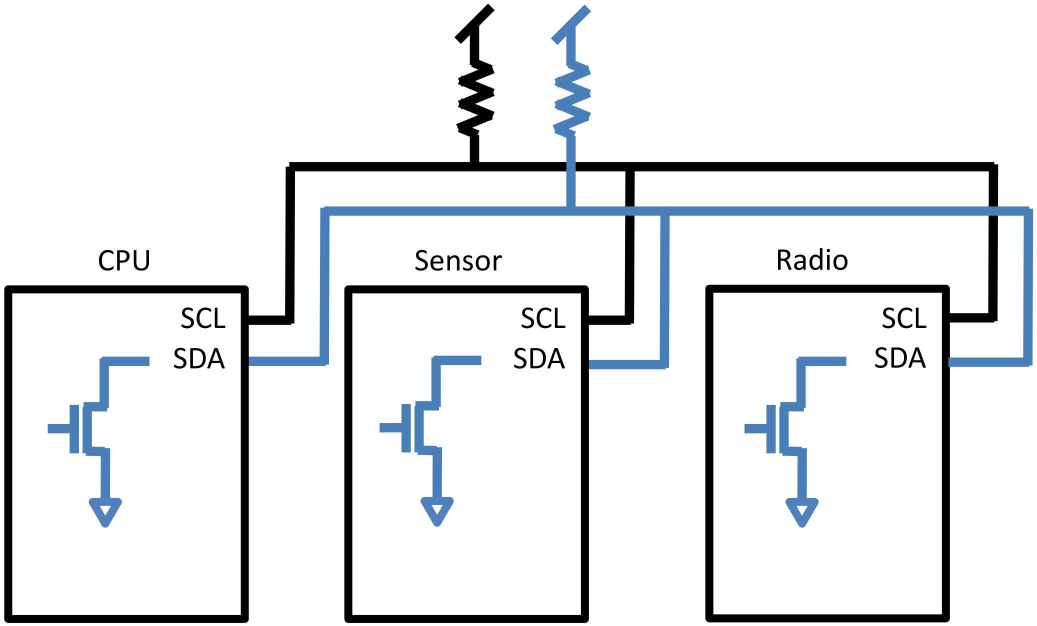
I^{2}C has fixed I/O requirements and a decentralized architecture



I²C has fixed I/O requirements and a decentralized architecture

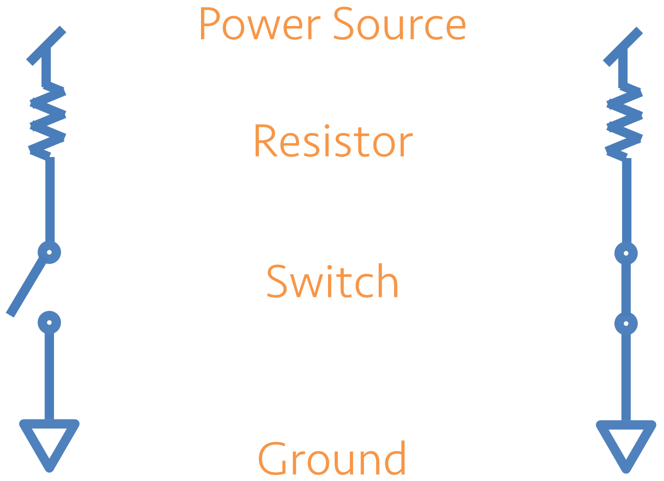


I²C is built on a simple circuit that enables its properties



Open-collector (aka wired-AND)

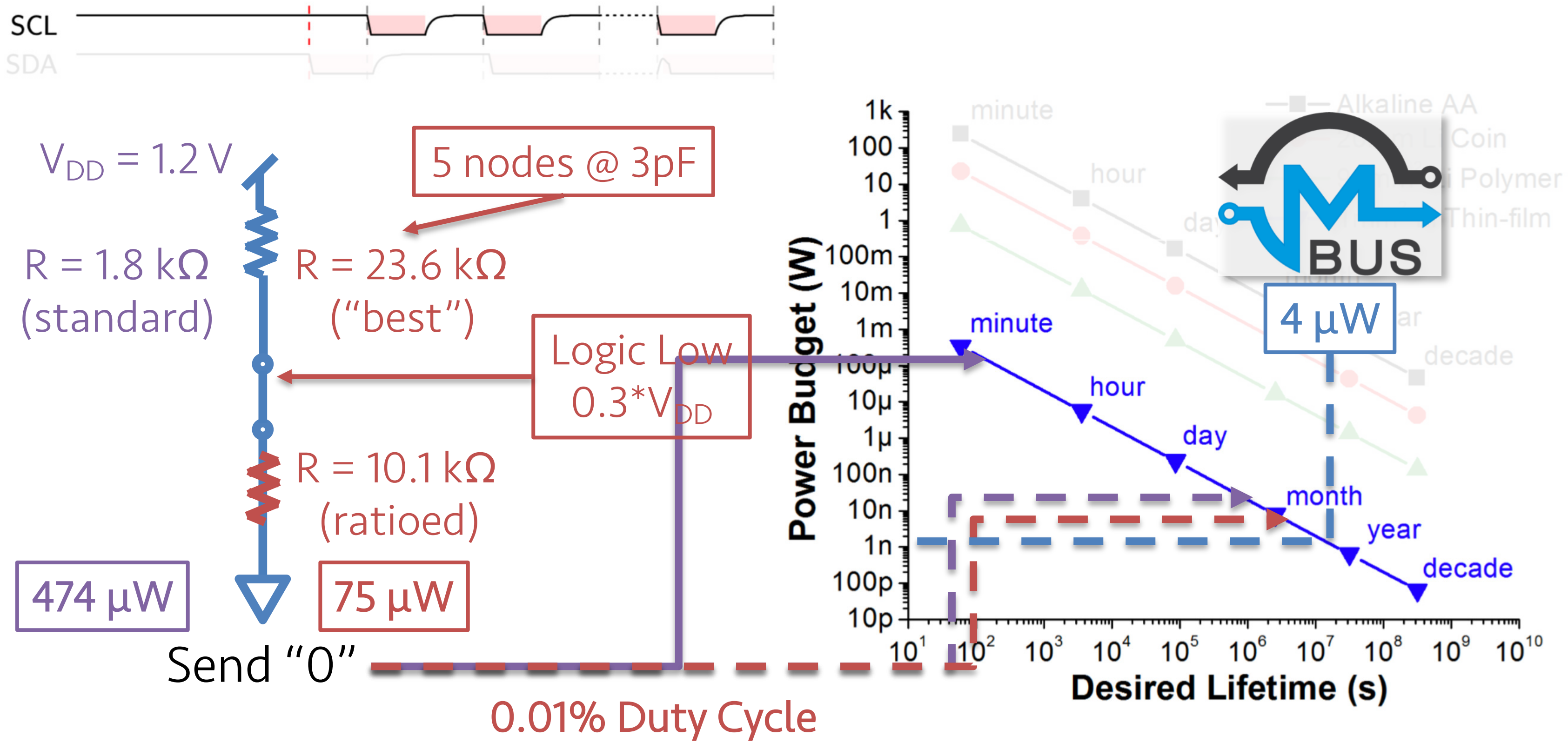
Problem is the energy costs of running an open-collector



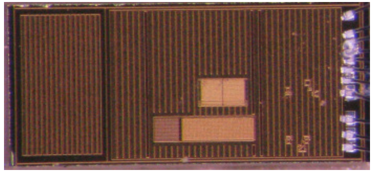
Send “1”: 0 W

Send “0”: …

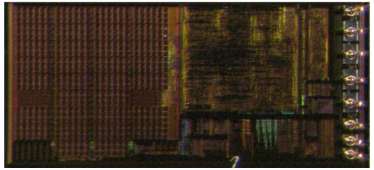
Achieving energy efficiency with I²C sacrifices modularity



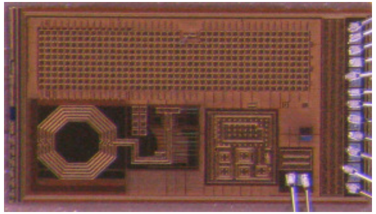
MBus enables Amdahl-balanced modularity for the nanopower computing class



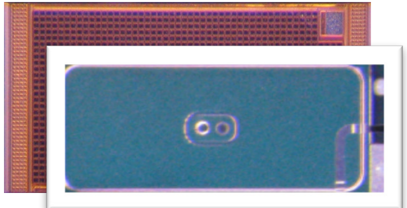
Temperature Sensor
~10 pW standby, < 1 μW active



CPU
~1 nW standby, ~5 μW active



Radio
~10 pW standby, ~10 μW active



Energy Harvesting & Storage
1~10 nW indoors
2~10 μAh capacity

< pW standby
~4 μW active

What makes MBus hard is satisfying <u>all</u> of the design constraints for current and future millimeter-scale systems

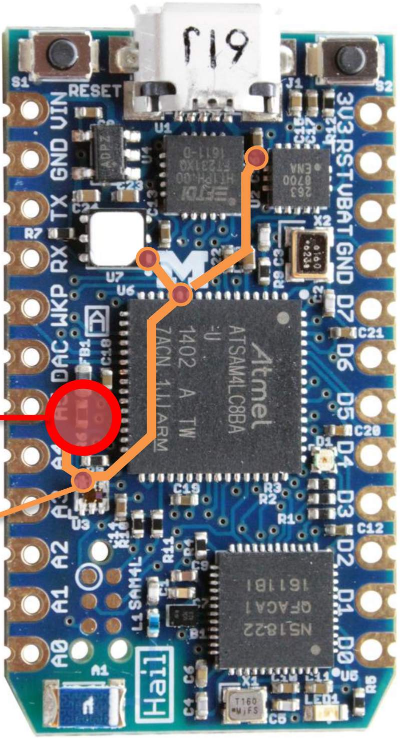
- Low active power
- Fixed pin count (4)
- Minimal standby power
- Multi-master design
- Fully synthesizable
- Robust timing (I/O drive/load)
- Efficient & future-proof addressing
- Data independent behavior
- Hardware acknowledgements
- Power aware design

MBus: An Ultra-Low Power Interconnect Bus for Next Generation Nanopower Systems
Pat Pannuto, Yoonmyung Lee, Ye-Sheng Kuo, ZhiYoong Foo, Benjamin Kempke, Gyouho Kim, Ronald G Dreslinski, David Blaauw, and Prabal Dutta
Proceedings of the 42nd International Symposium on Computer Architecture (ISCA '15). **Top Pick in Computer Architecture.**

Requirement: Low Active Power

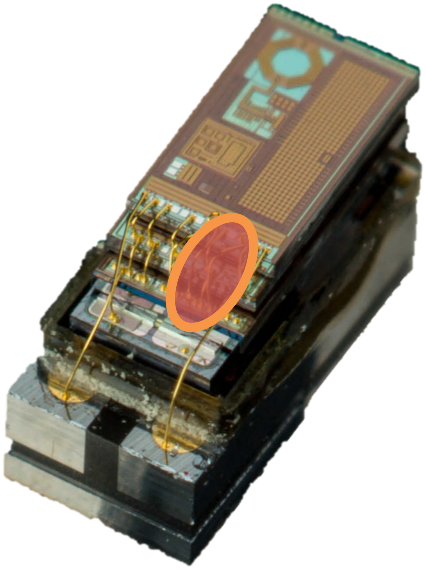
Idea: Eliminate the energy-hungry pullup

- Per clock cycle…



~1,200 pJ lost to pull-up

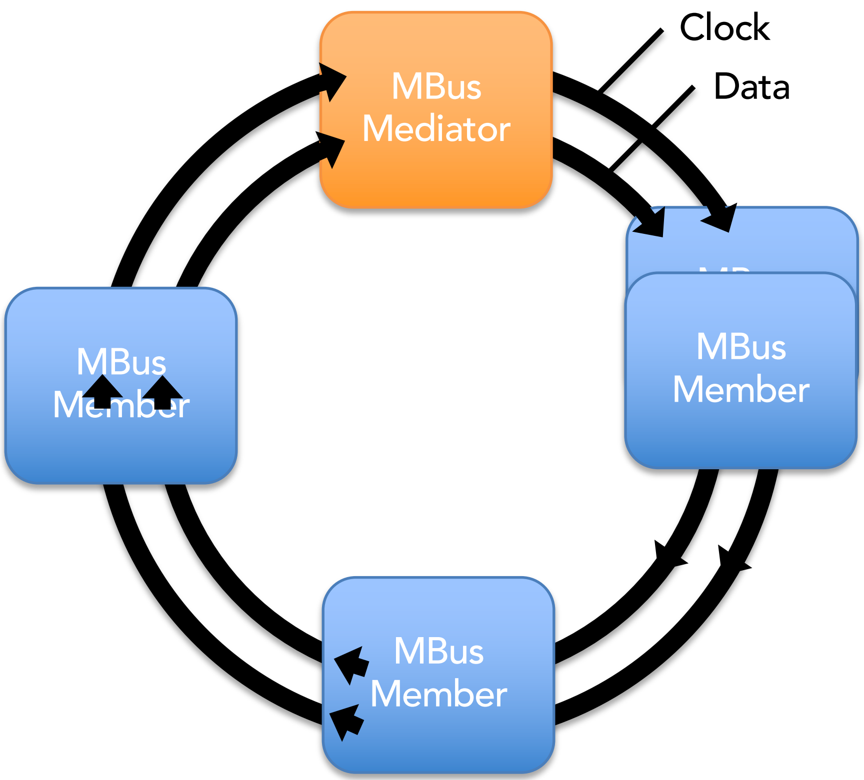
~150 pJ to pads/wires

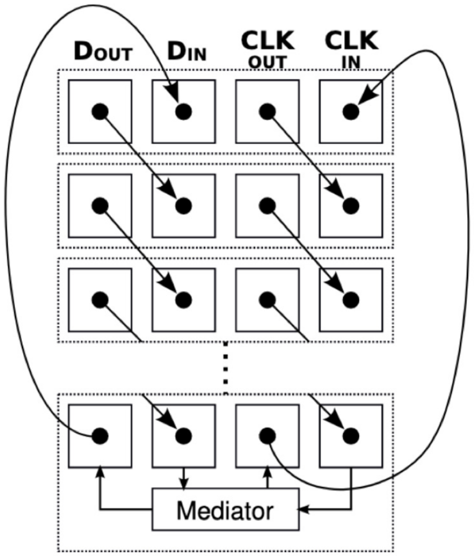


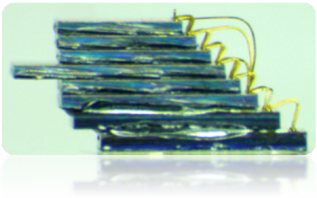
Requirement: Fixed Pin Count

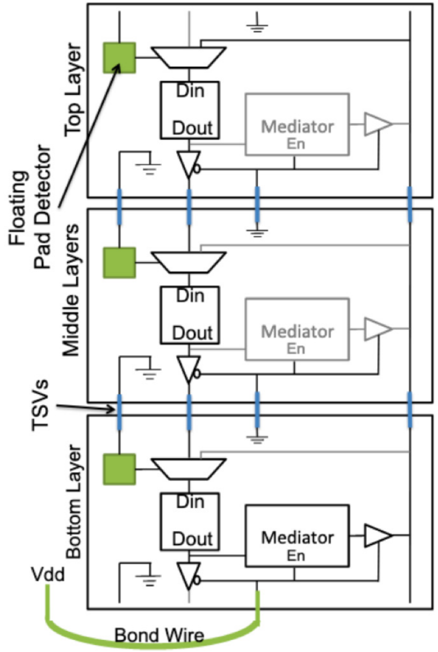
Requirement: Fixed Pin Count

Idea: Ring topology scales and prevents conflicts





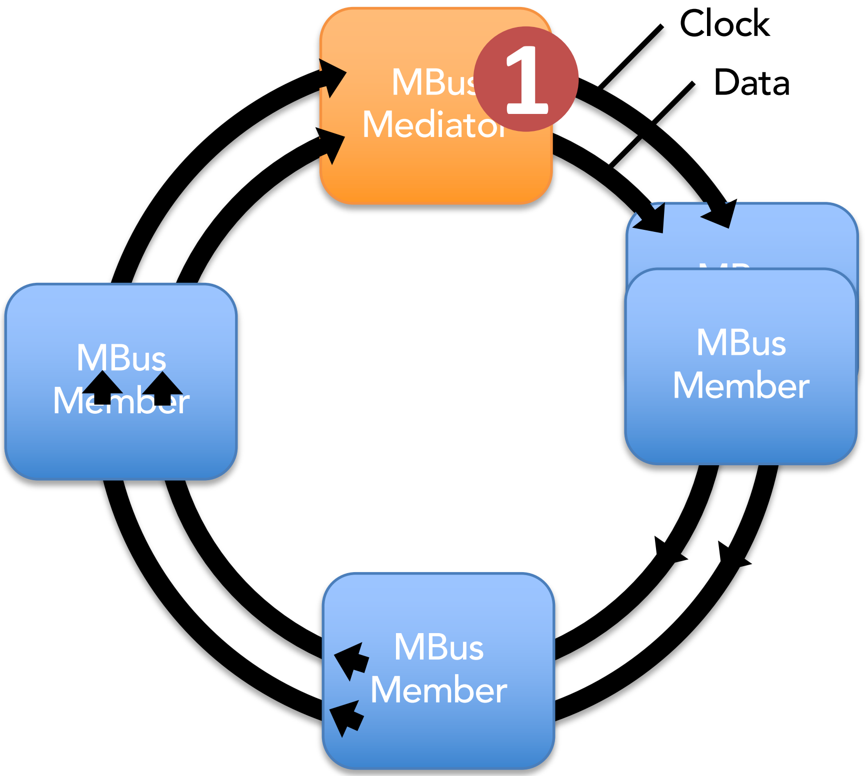




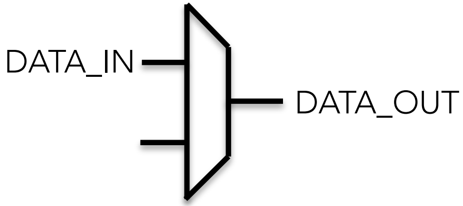
Requirement: Low Active Power & Low Standby Power

Requirement: Low Active Power & Low Standby Power

Idea: Combinational frontend



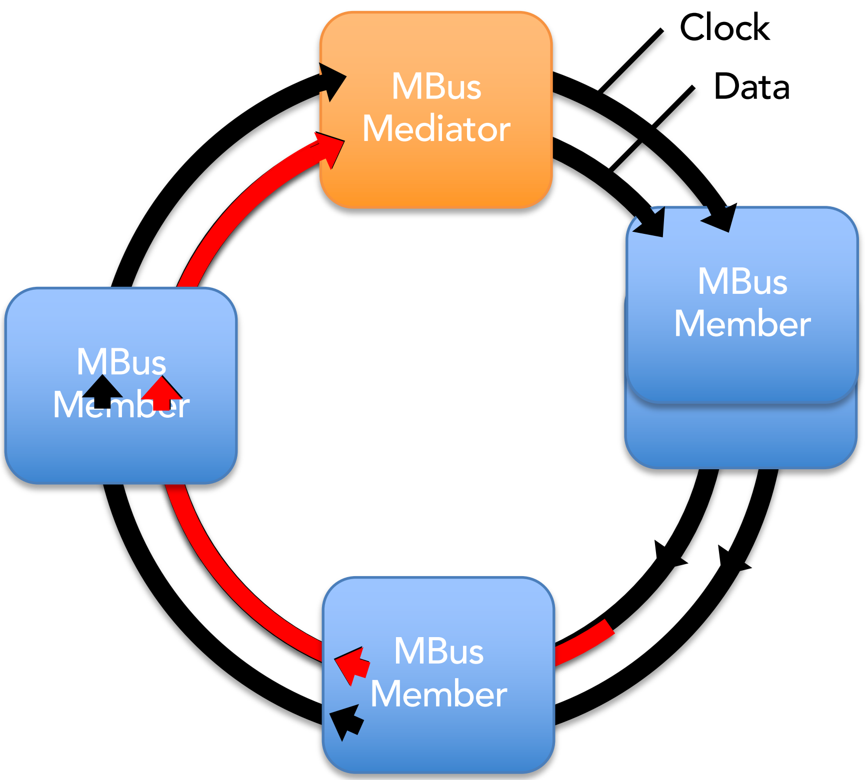
- Clockless "shoot-through" ring



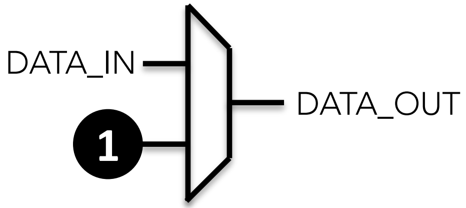
- All bus frontends share one bus clock
  - No local oscillators saves energy
  - Generated by mediator
  - Usually a CPU-like node

Requirement: Multi-Master Design

Requirement: Multi-Master Design
Idea: Topological priority can mitigate rare conflicts



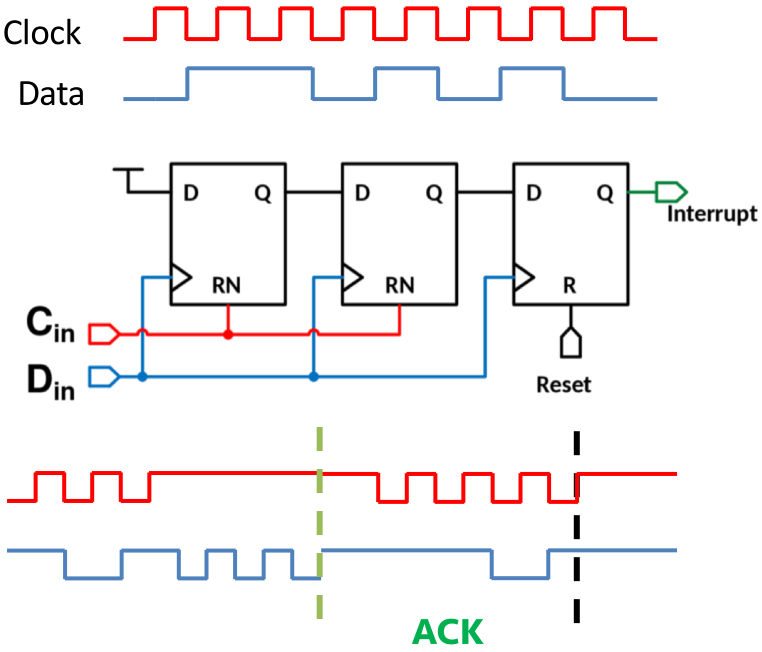
- One node "mediates" arbitration
  - Does not forward during idle



  - Generate bus clock when DATA_IN falls
- Unambiguous winner
- Multiple arbitration rounds allow for non-topological priority schemes

MBus is a clean-slate design, built to satisfy interconnect requirements for this and the next generation of modular systems

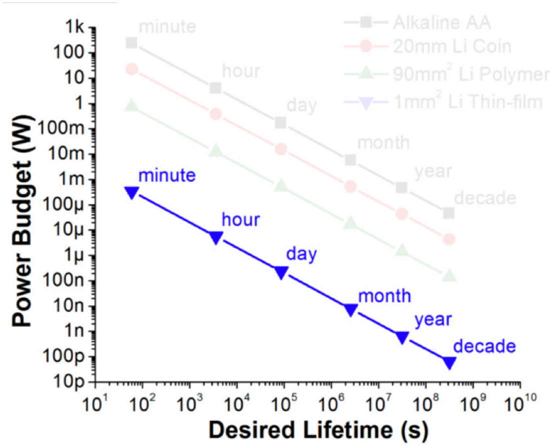
- Low active power
- Fixed pin count (4)
- Minimal standby power
- Multi-master design
- Fully synthesizable
- Robust timing (I/O drive/load)
- Minimal protocol overhead
  - Safe & efficient arbitration
  - Efficient & future-proof addressing
  - Data independent behavior (end of message?)
  - Transaction acknowledgements
- Power aware design



MBus: An Ultra-Low Power Interconnect Bus for Next Generation Nanopower Systems
Pat Pannuto, Yoonmyung Lee, Ye-Sheng Kuo, ZhiYoong Foo, Benjamin Kempke, Gyouho Kim, Ronald G Dreslinski, David Blaauw, and Prabal Dutta
Proceedings of the 42nd International Symposium on Computer Architecture (ISCA '15). **Top Pick in Computer Architecture.**

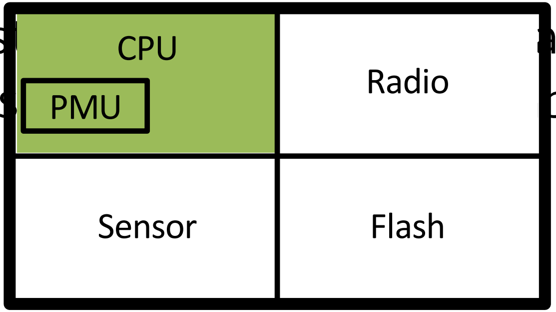
Energy-critical systems push from “dark silicon” to “pitch black” silicon

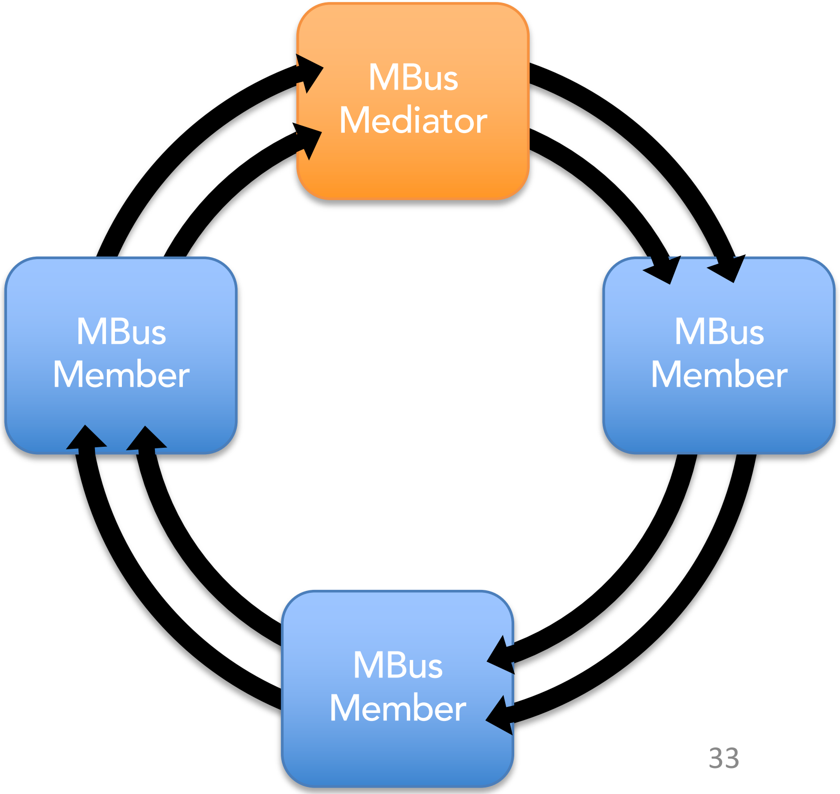
- Clock-gating: “Dark Silicon”
  - Stop driving the clock tree of regions of a chip
  - Eliminates switching power, but not static leakage
- Power-gating: “Pitch Black Silicon”
  - Switch off power to regions of a chip
  - Eliminates (almost) all chip leakage
- Micro-scale systems aggressively power-gate to reach energy budget
  - Power management is a looming systems synthesis problem



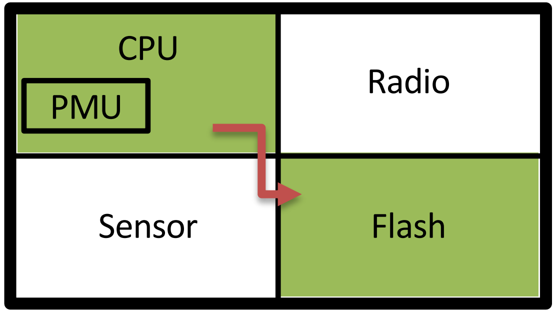
Distributed operation creates challenges for system power management

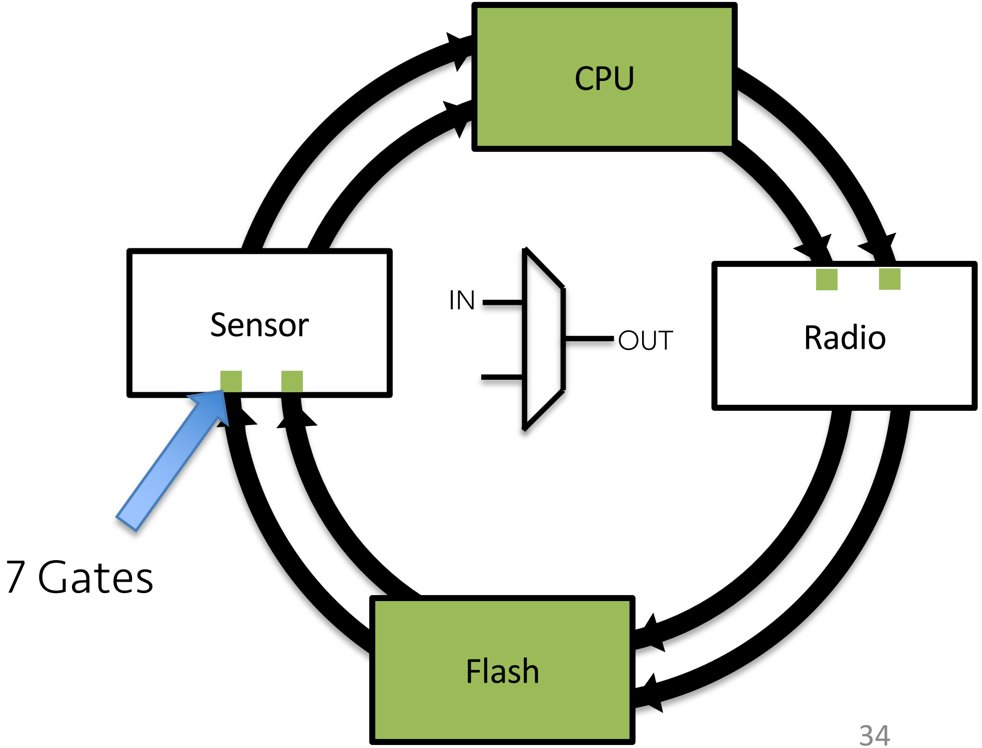
- Monolithic systems can bootstrap, monitor, and manage power state from a centralized local controller
- Dis… ave the bu… ponents



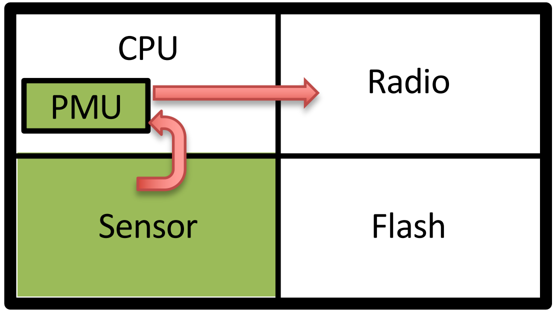


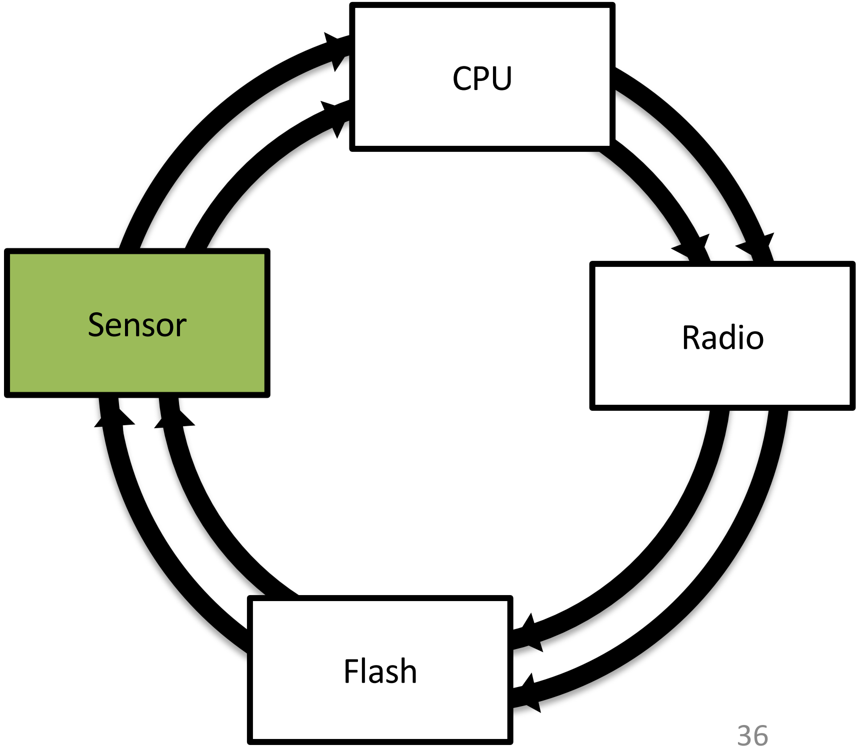
Problem 1: Efficiently "talking through" unpowered chips



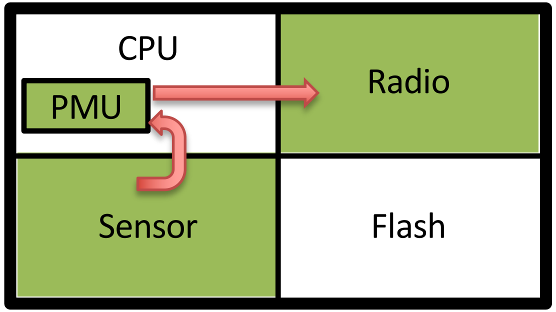


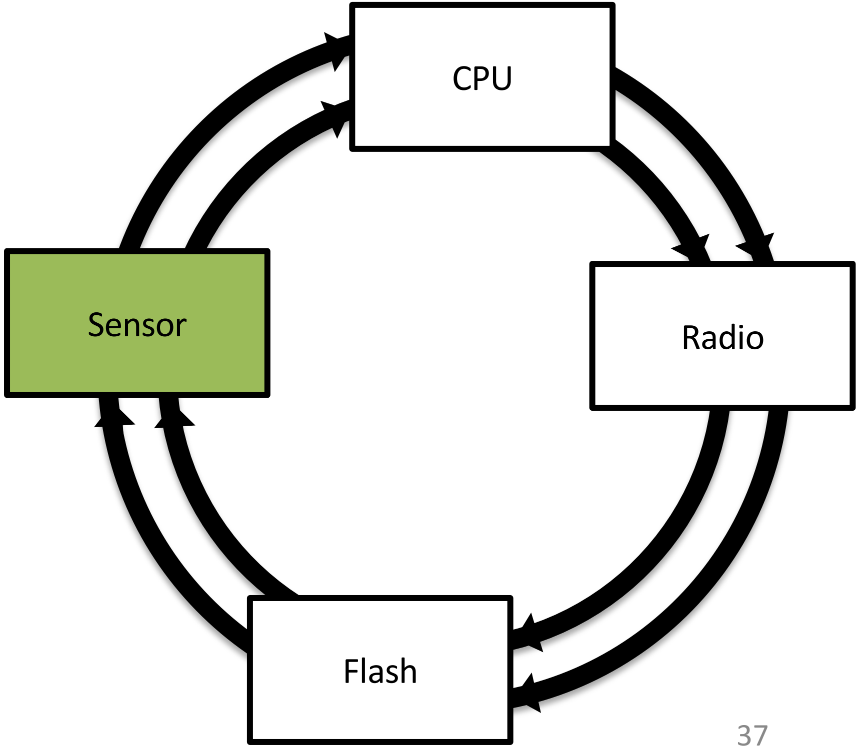
Problem 2: Can I talk to it? Is it on?





Problem 2: Can I talk to it? Is it on?
Problem 3: How do I turn something on?



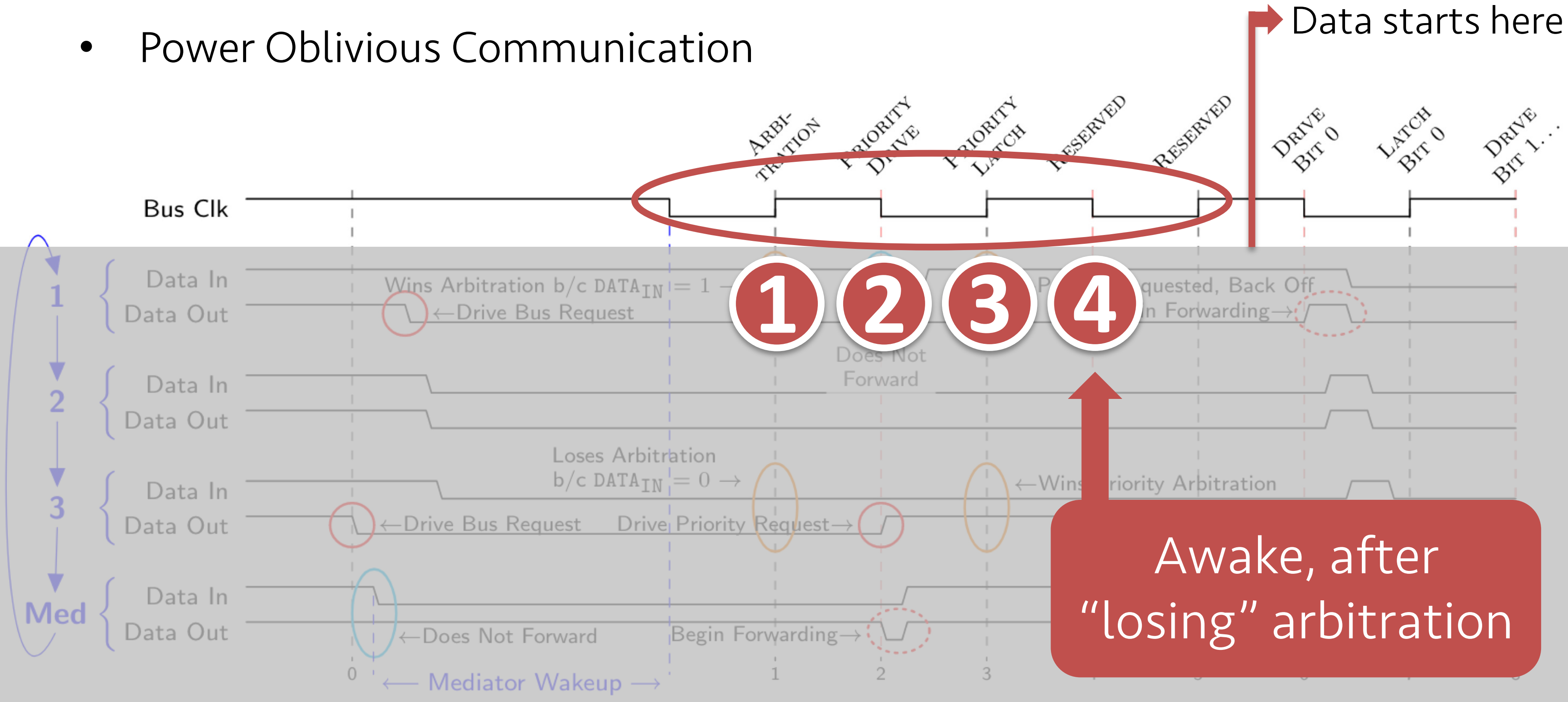


Problem 3: How to turn something on?

- Cold-booting circuits is much easier if there is a stable clock available
  1. Turn it on
  2. Start your local clock
  3. Connect to the already powered parts
  4. Release power-on reset

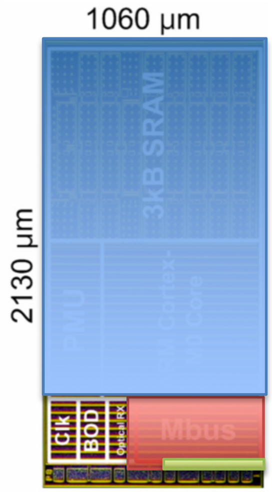
MBus Insight: Problems 2 and 3 can be solved together

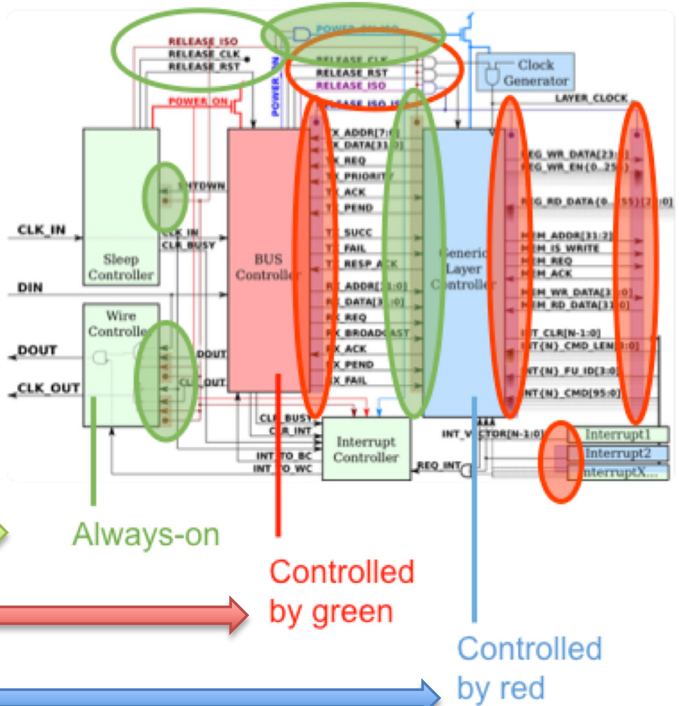
- Power Oblivious Communication



MBus introduces three, transparent, hierarchical power domains to maximize efficiency

- Minimal always powered frontend
  - 32 logic gates, 4 flops
- Small controller active during transmission
  - 27,300 μm^2 in 180 nm process





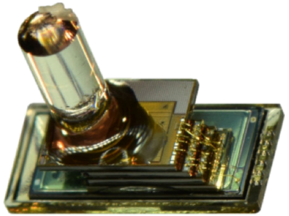
< pW: Wire Controller

~ nW: Bus Controller

~ µW+: Rest of IC

Extending hierarchical power domains one step further
MBus and the next, next, next generation

- MBus abstraction presented to chips:
  - Power control signals
  - Byte-oriented send/receive
  - “Always-on” interrupt



  - 304 nW analog motion detection
  - 20 µW digital image capture mode

Defense Advanced Research Projects Agency › Program Information

Near Zero Power RF and Sensor Operations (N-ZERO)

Dr. Benjamin Griffin

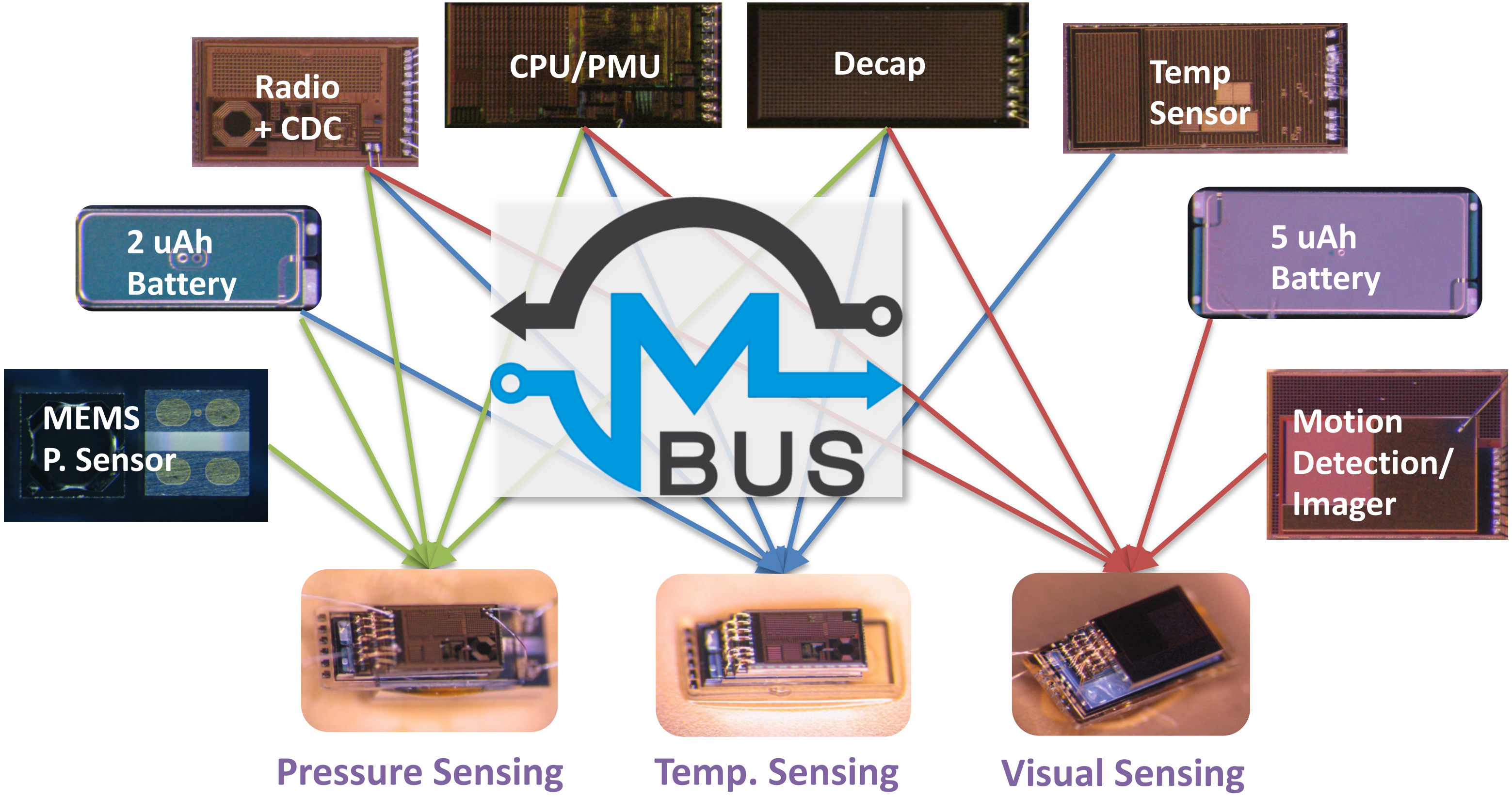
State-of-the-art military sensors rely on “active electronics” to detect vibration, light, sound or other signals for situational awareness and to inform tactical planning and action. That means the sensors constantly consume power, with much of that power spent processing what often turns out to be irrelevant data. This power consumption limits sensors’ useful lifetimes to a few weeks or months with even the best batteries and has slowed the development of new sensor technologies and capabilities. The chronic need to service or redeploy power-depleted sensors is not only costly and time-consuming but also increases warfighter exposure to danger.

The Near Zero Power RF and Sensor Operations (N-ZERO) program has the goal of developing the technological foundation for persistent, event-driven sensing capabilities in which the sensor can remain dormant, with near-zero power consumption, until awakened by an external trigger or stimulus. Examples of relevant stimuli are acoustic signatures of particular vehicle types or radio signatures of specific communications protocols. If successful, the program could extend the lifetime of remotely deployed communications and environmental sensors—also known as unattended ground sensors (UGS)—from weeks or months to years.

N-ZERO will initially focus on two broad areas. One centers on UGS capable of continuous monitoring for infrequent events, with near zero power consumption, but that activate a conventional sensor suite for further sensor data collection and processing when an event of interest is detected and confirmed. Only then would significant power be drawn. The other broad area concentrates on radio receivers that are continuously alert for friendly radio transmissions, but with near zero power consumption when transmissions are not present. A common feature of both focus areas is that the sensing is continuous—no events of interest or communications are missed—and that power-drawing confirmatory sensing and communications functions kick in only when triggered to do so.

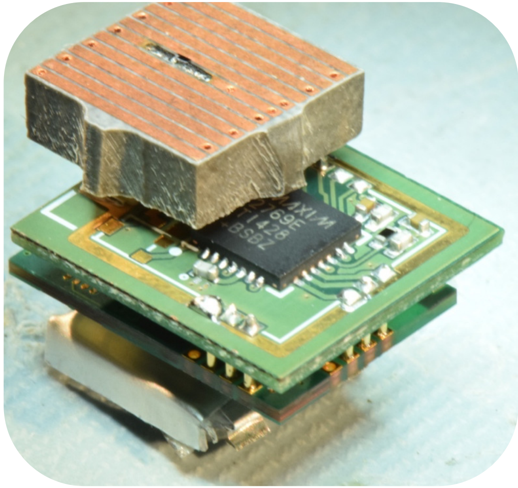
A Millimeter-Scale Wireless Imaging System with Continuous Motion Detection and Energy Harvesting
Gyouho Kim, ZhiYoong Foo, **Pat Pannuto**, Ye-Sheng Kuo, Benjamin Kempke, Mohammad Hassan Ghaed, Suyoung Bang, Inhee Lee, Yejoong Kim, Seokhyeon Jeong, Prabal Dutta, Dennis Sylvester, and David Blaauw
Proceedings of the 2014 IEEE Symposium on VLSI Technology (VLSI’14)

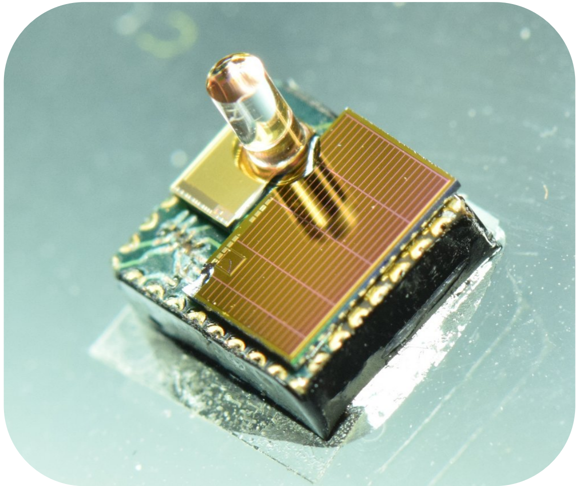
MBus enables an ecosystem of millimeter-scale, nanopower platforms



Seamless and transparent interaction between power-aware and power-oblivious chips

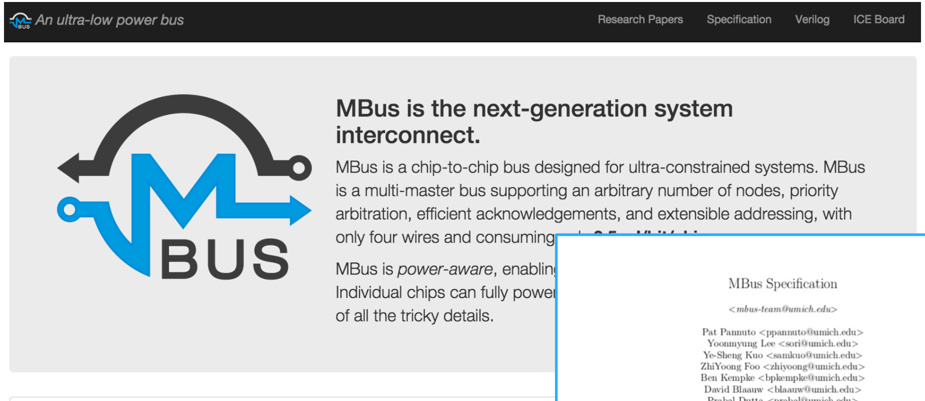
- Facilitates integration with COTS chips

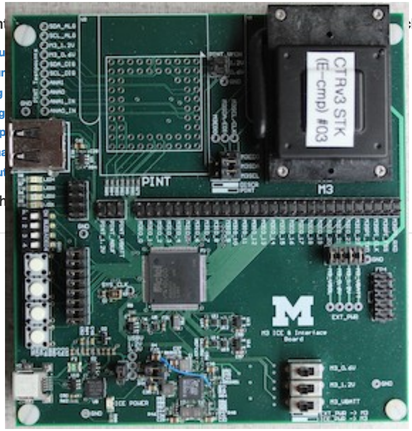




*No current COTS chip support MBus, these integrations leverage more traditional buses or bitbanging

http://mbus.io Get the MBus spec, code, tools, and chips





MBus Specification

<mbus-team@umich.edu>

Pat Pannuto <ppannuto@umich.edu>
Yoonmyung Lee <sori@umich.edu>
Ye-Sheng Kuo <samkuo@umich.edu>
ZhiYoong Foo <zhiyoong@umich.edu>
Ben Kempke <bpkempke@umich.edu>
David Blaauw <blaauw@umich.edu>
Prabal Dutta <prabal@umich.edu>

Revision 0.3 — April 18, 2014

Overview

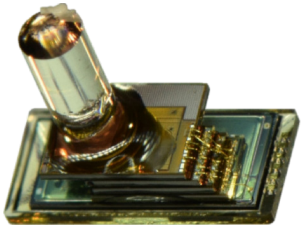
MBus is an ultra-low power system bus. The original design was motivated by the Michigan Micro Mote (M3) project. The goal of MBus, however, is to be a general purpose bus for hyper-constrained systems. MBus requires four pins per node, uses purely digital logic, supports arbitrary length transfers, and features a low-latency priority channel and robust acknowledgments. MBus member nodes do not require a local clock and are capable of completely stateless operation. MBus requires one more powerful node to act as a bus *master* node, whose primary duties are providing the MBus clock and mediating arbitration.

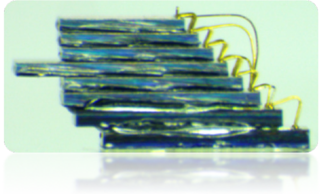
Contents

MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

KU LEUVEN

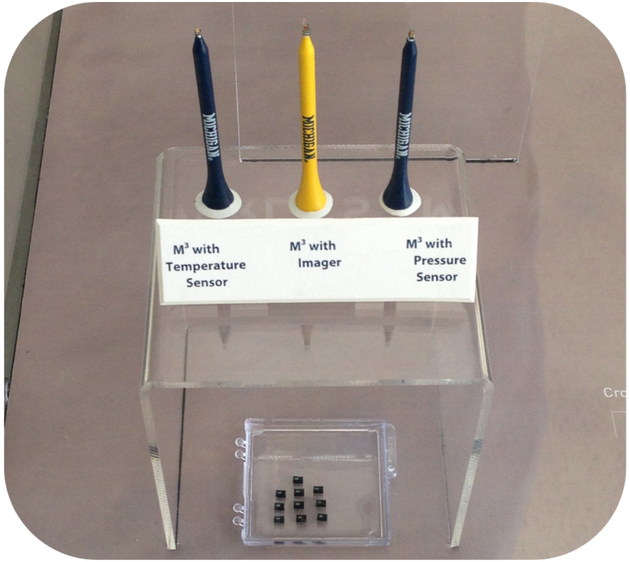
CubeWorks

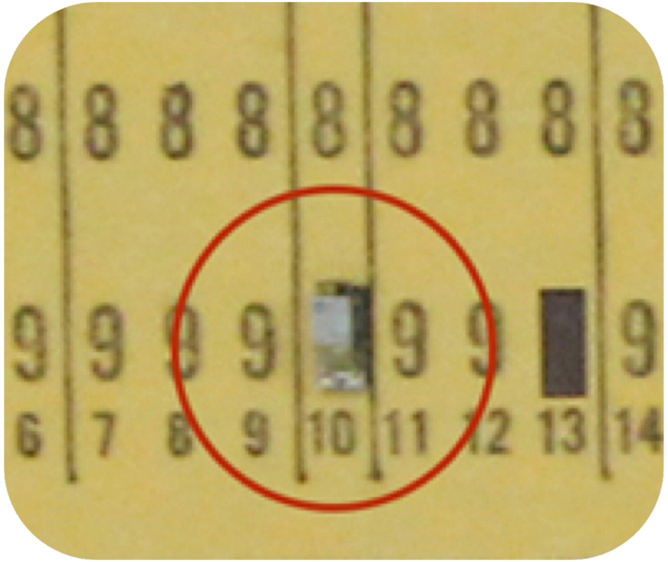




Check out the "World's Smallest Computer" exhibit at Silicon Valley's Computer History Museum!

Computer History Museum





Outline

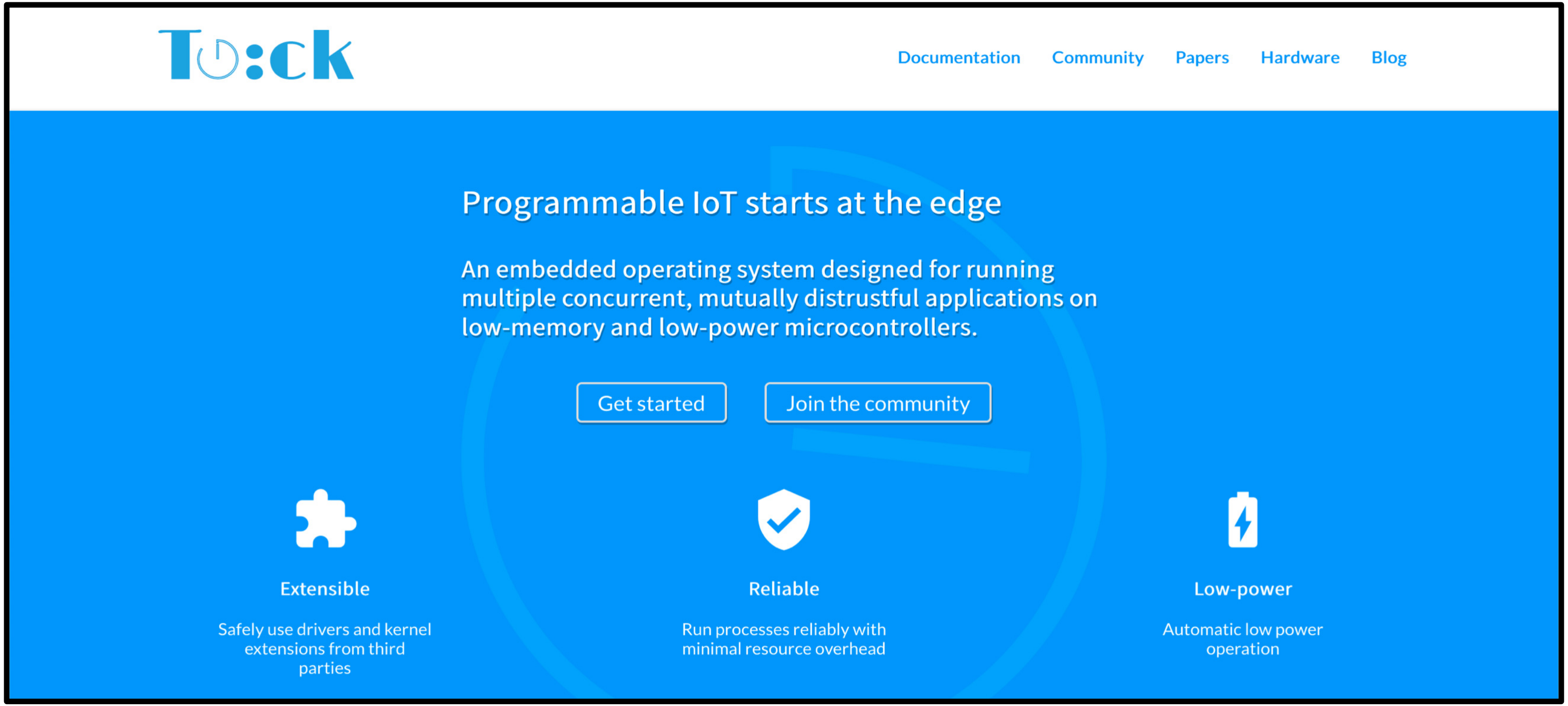
- Introducing the nanopower computing class
- The development of a modular, nanopower computing platform

  Hardware — Composing micro-scale systems

  Software — Returning modularity to microcontroller software

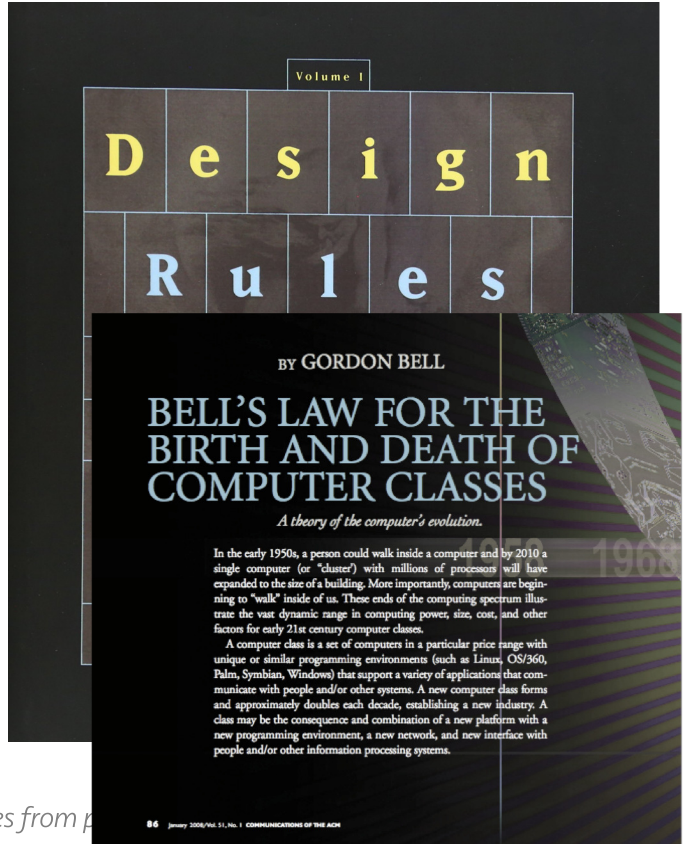
  Services — Deployment and management of “dust”
- The future of nanopower computing and its broader impact

Tock is a new embedded operating system designed for safe, robust multitenancy on microcontrollers



Multiprogramming a 64kB Computer Safely and Efficiently
Amit Levy, Bradford Campbell, Branden Ghena, Daniel B. Giffin, **Pat Pannuto**, Prabal Dutta, and Philip Levis
Proceedings of the 26th Symposium on Operating Systems Principles (SOSP'17)

The case for multitenancy on a millimeter-scale computer



BY GORDON BELL

BELL'S LAW FOR THE BIRTH AND DEATH OF COMPUTER CLASSES

A theory of the computer's evolution.

In the early 1950s, a person could walk inside a computer and by 2010 a single computer (or "cluster") with millions of processors will have expanded to the size of a building. More importantly, computers are beginning to "walk" inside of us. These ends of the computing spectrum illustrate the vast dynamic range in computing power, size, cost, and other factors for early 21st century computer classes.

A computer class is a set of computers in a particular price range with unique or similar programming environments (such as Linux, OS/360, Palm, Symbian, Windows) that support a variety of applications that communicate with people and/or other systems. A new computer class forms and approximately doubles each decade, establishing a new industry. A class may be the consequence and combination of a new platform with a new programming environment, a new network, and new interface with people and/or other information processing systems.

86 January 2008/Vol. 51, No. 1 COMMUNICATIONS OF THE ACM

Modularity…

- "increases the range of 'manageable' complexity"
- "allows different parts of a large design to be worked on concurrently"
- "accommodates uncertainty"

"And generality always wins"

- *Universality* – any computer can take over the function of another given sufficient memory and interfaces

Quotes from p

Networking often drives multitenancy, but comes with additional risk for unattended devices

- Multitenancy != multi-app
  - Black box vendor libraries (e.g. BLE stacks)
  - FCC compliance
- Modular computing means arbitrary compute
  - Phones, watches… fixed-function to platform
- Multiprogrammability amortizes costs
  - Siloed sensor infrastructure is expensive
  - Future-proof, short & long-run adaptive

LILY HAY NEWMAN SECURITY 08.09.18 12:30 PM

A NEW PACEMAKER HACK PUTS MALWARE DIRECTLY ON THE DEVICE

Researchers Hack Into Michigan's Traffic Lights

Security flaws in a system of networked stoplights point to looming problems with an increasingly connected infrastructure.

09 FBI: Smart Meter Hacks Likely to Spread

APR 12

A series of hacks perpetrated against so-called "smart meter" installations over the past several years may have cost a single U.S. electric utility hundreds of millions of dollars annually, the **FBI** said in a cyber intelligence bulletin obtained by KrebsOnSecurity. The law

Microcontrollers rather than microprocessors dominate the emerging compute class, and must learn to become a platform

- Why is the software so different? The hardware is different…
  - MPU (Memory Protection Unit) rather than MMU (Memory Management Unit)
  - Transient storage (SRAM) is limited, and not growing (because limited energy!)
    - From 2004 to 2017… 10 kB → 64 kB

	Min CPU	Min RAM	Min Disk
Windows 10 (32-bit)	1 GHz	1 GB	16 GB

An embedded OS must balance stability and flexibility with fewer resources on a less flexible hardware platform

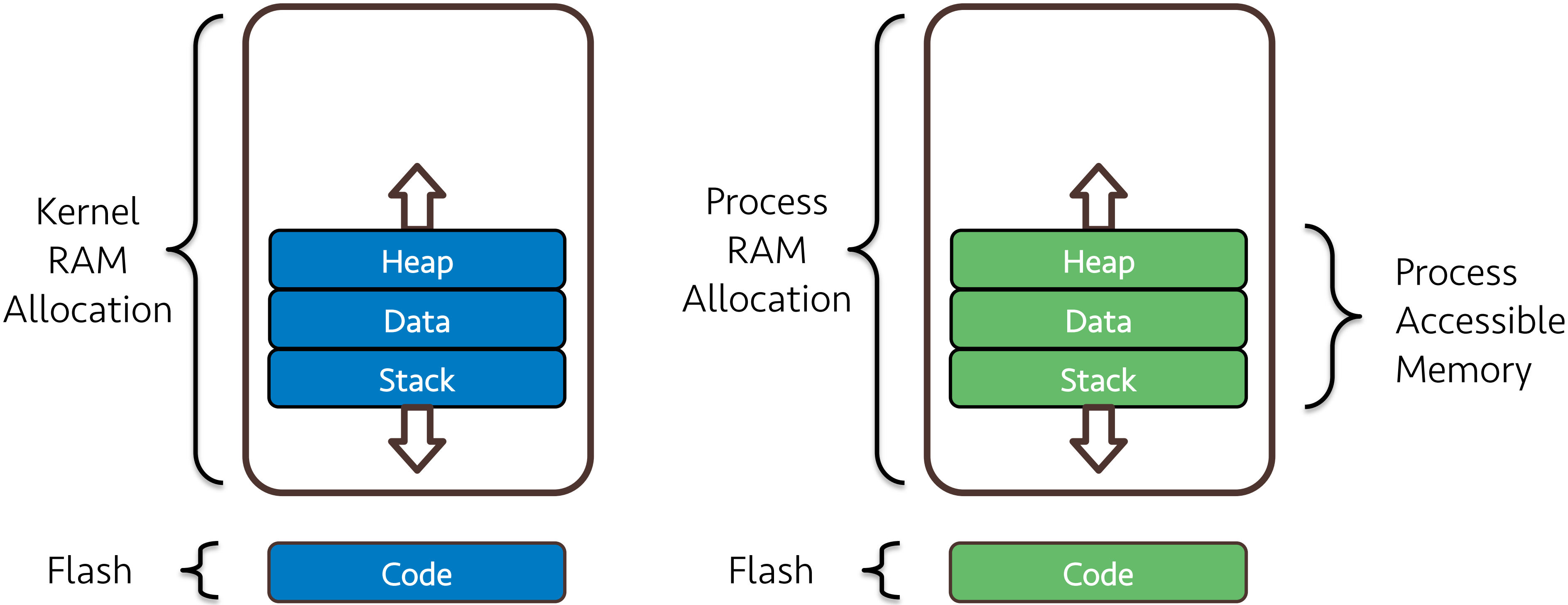
- Challenge: How to be **robust** and **adaptive** at the same time?
  - Robust: Memory exhaustion is common, and no virtual memory out
    - Kernel must be statically allocated
    - (co
  - Adapt
    - Ke
  - These

The benefits and costs of writing a POSIX kernel in a high-level language

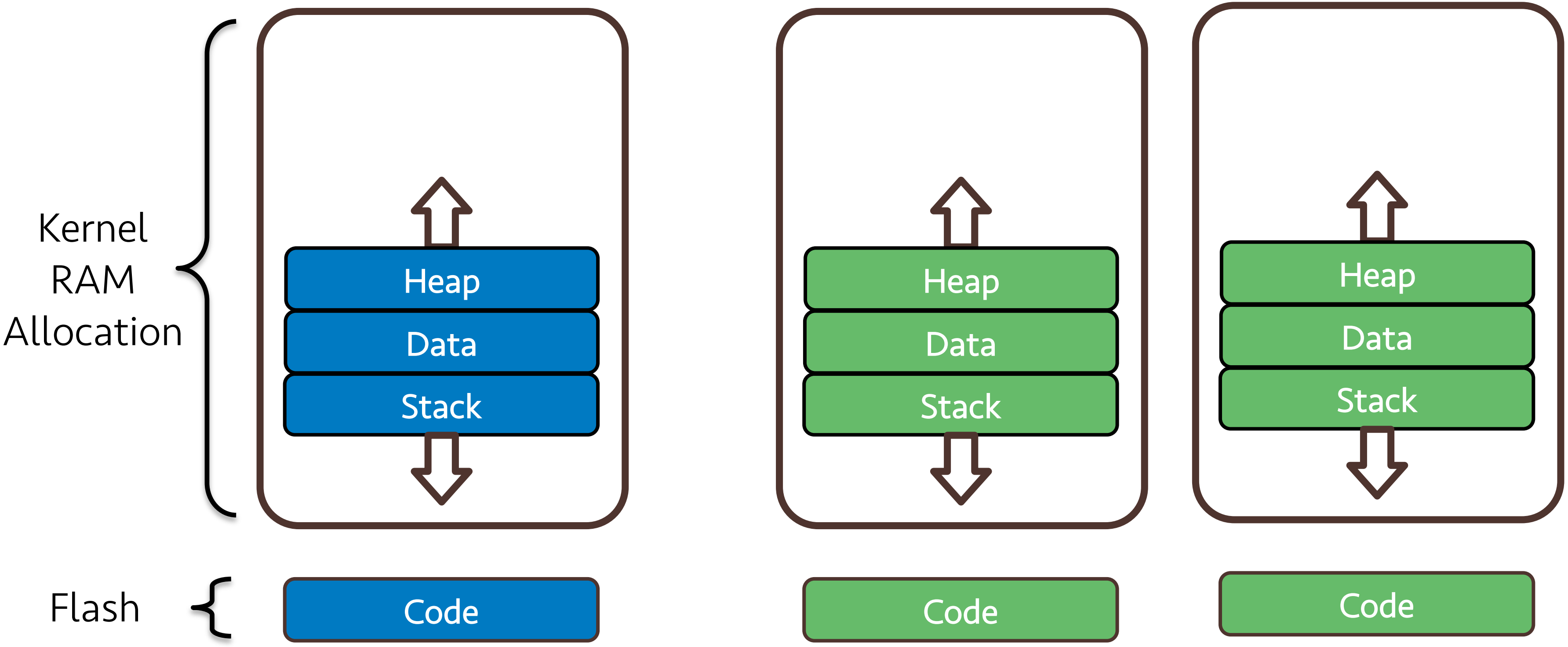
Cody Cutler, M. Frans Kaashoek, and Robert T. Morris, *MIT CSAIL*

"The most challenging puzzle was handling the possibility of running out of kernel heap memory"

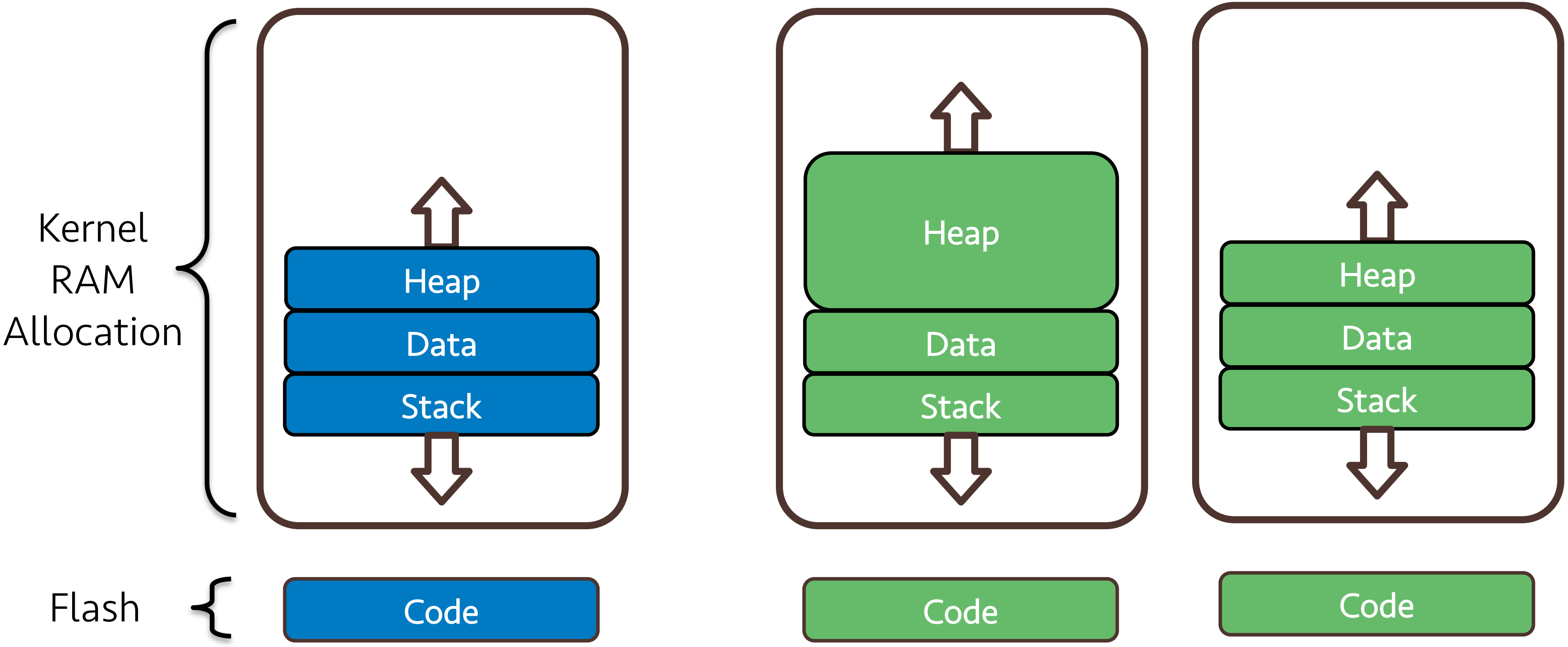
Traditional kernel heap allocation results in shared fate across all processes



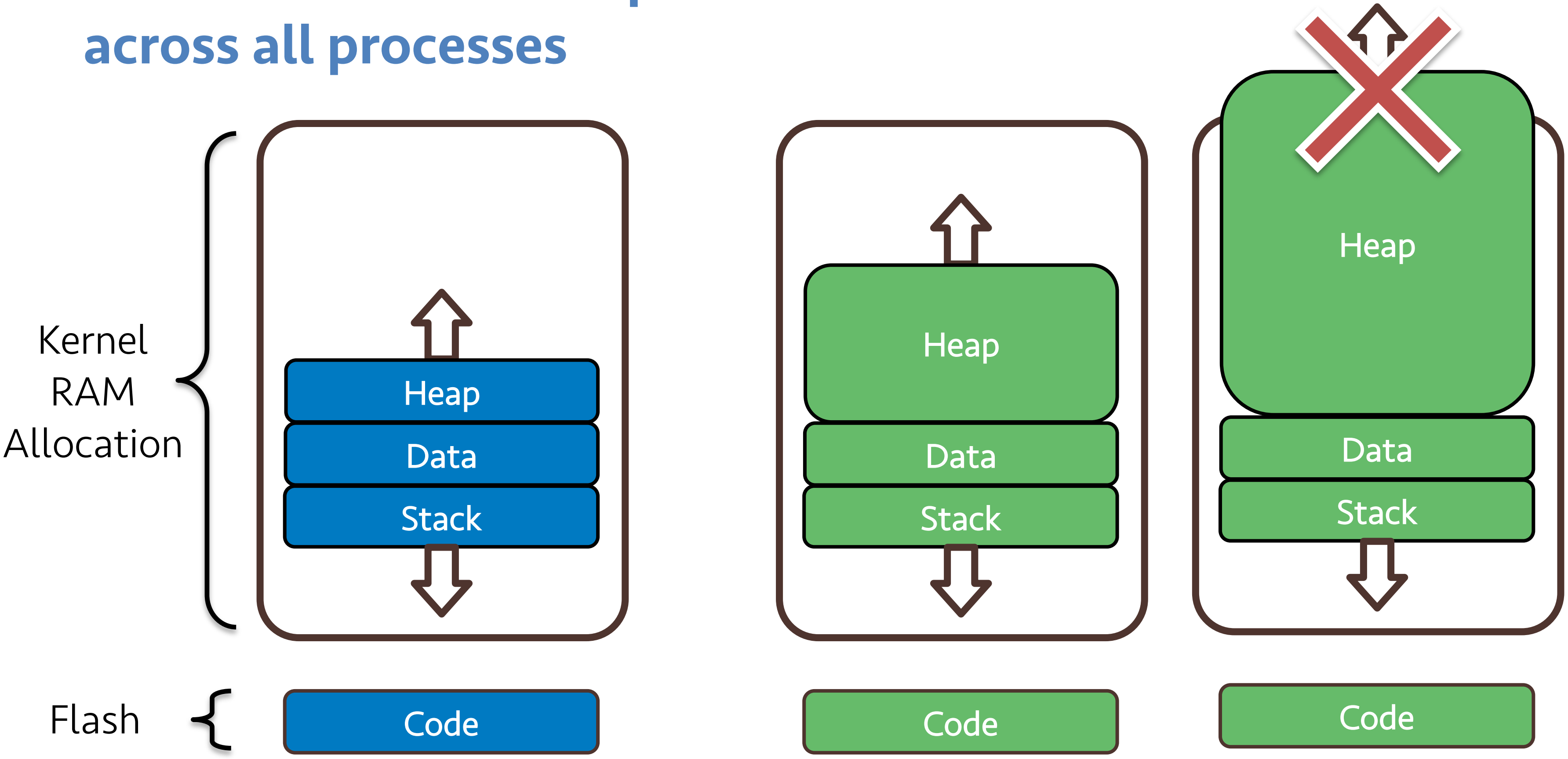
Traditional kernel heap allocation results in shared fate across all processes



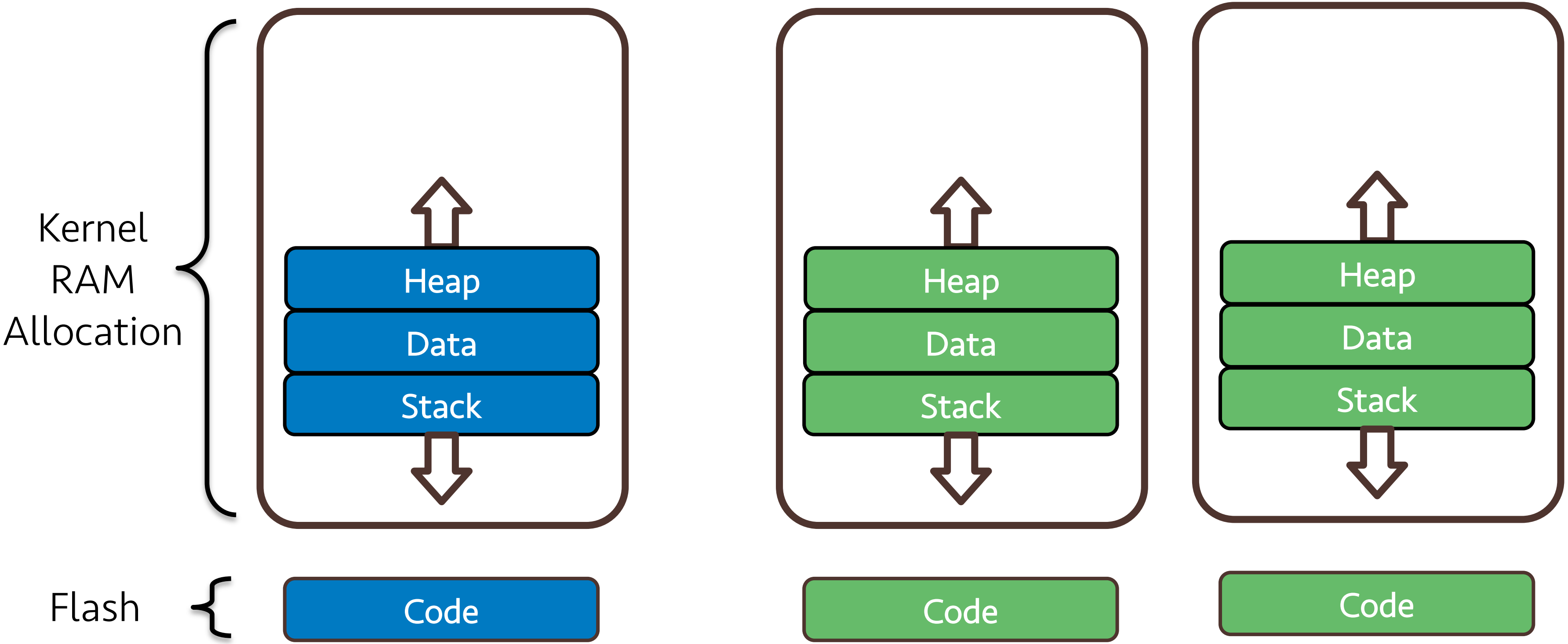
Traditional kernel heap allocation results in shared fate across all processes



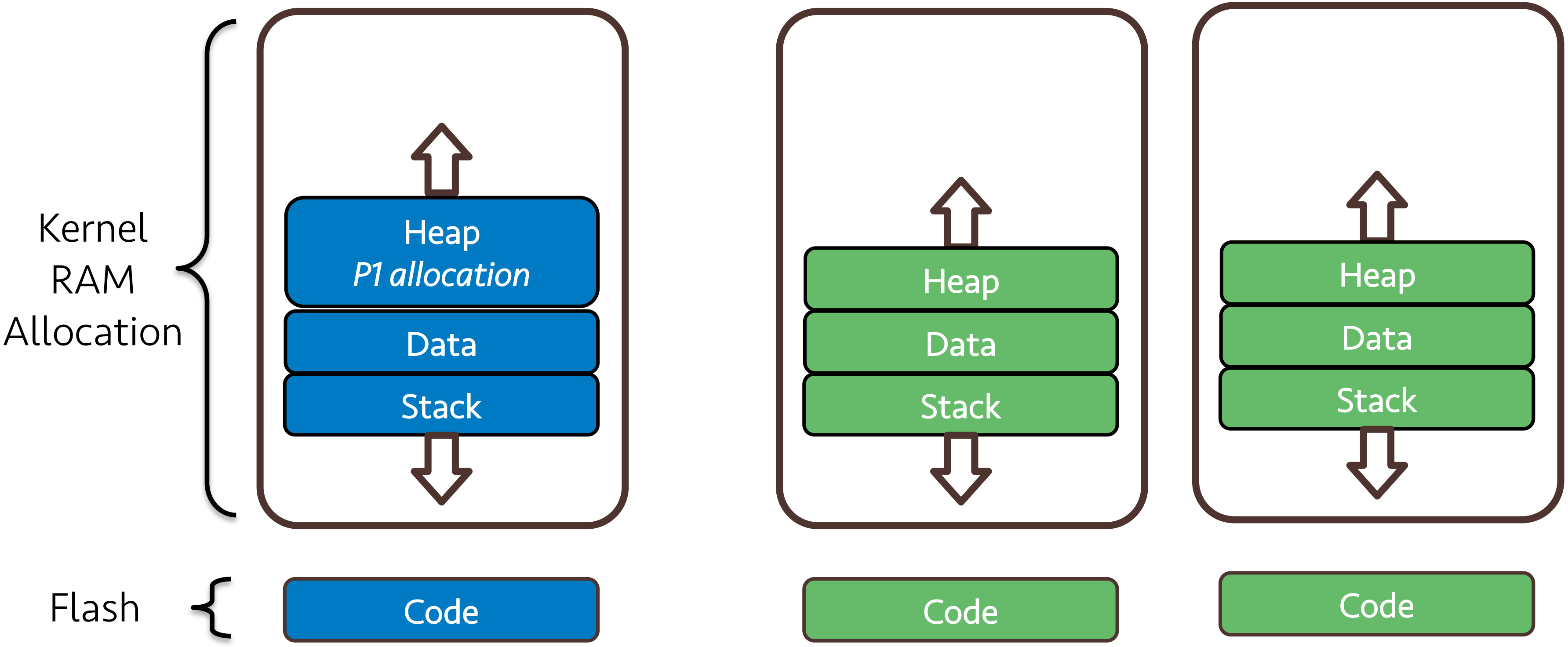
Traditional kernel heap allocation results in shared fate across all processes



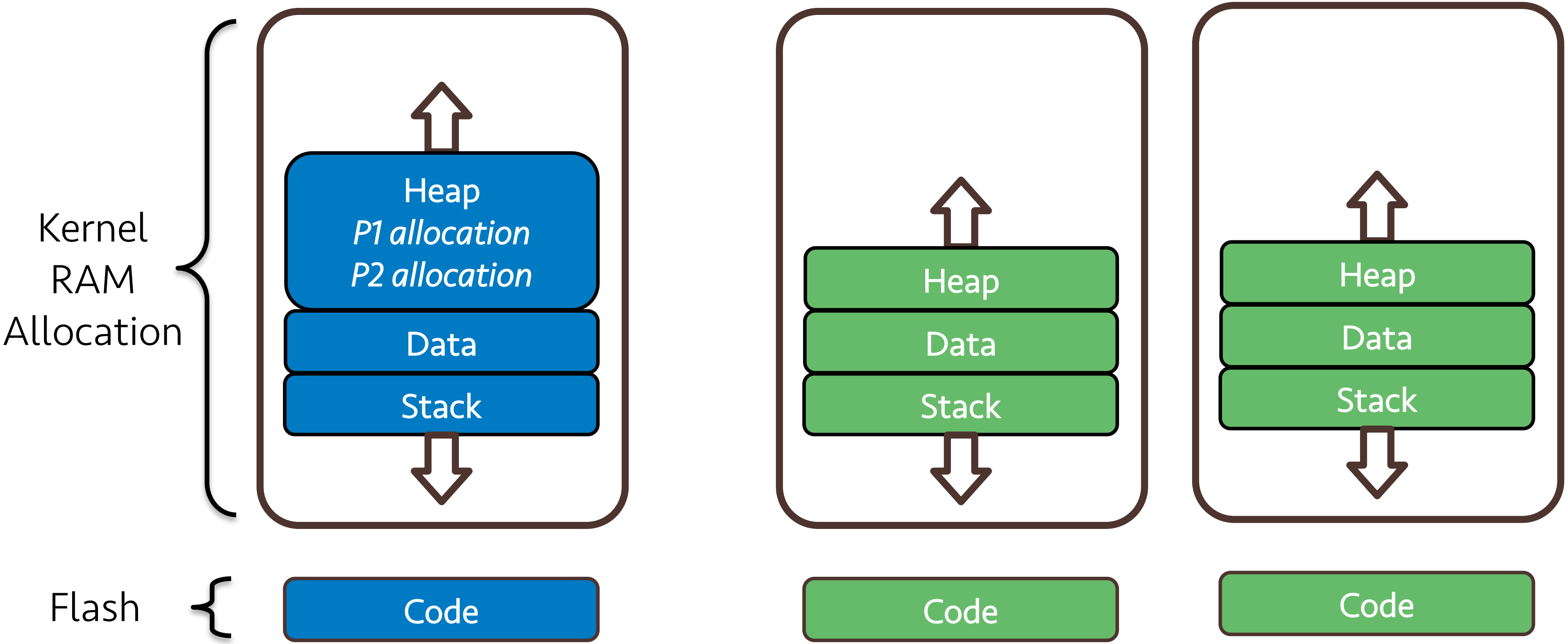
Traditional kernel heap allocation results in shared fate across all processes



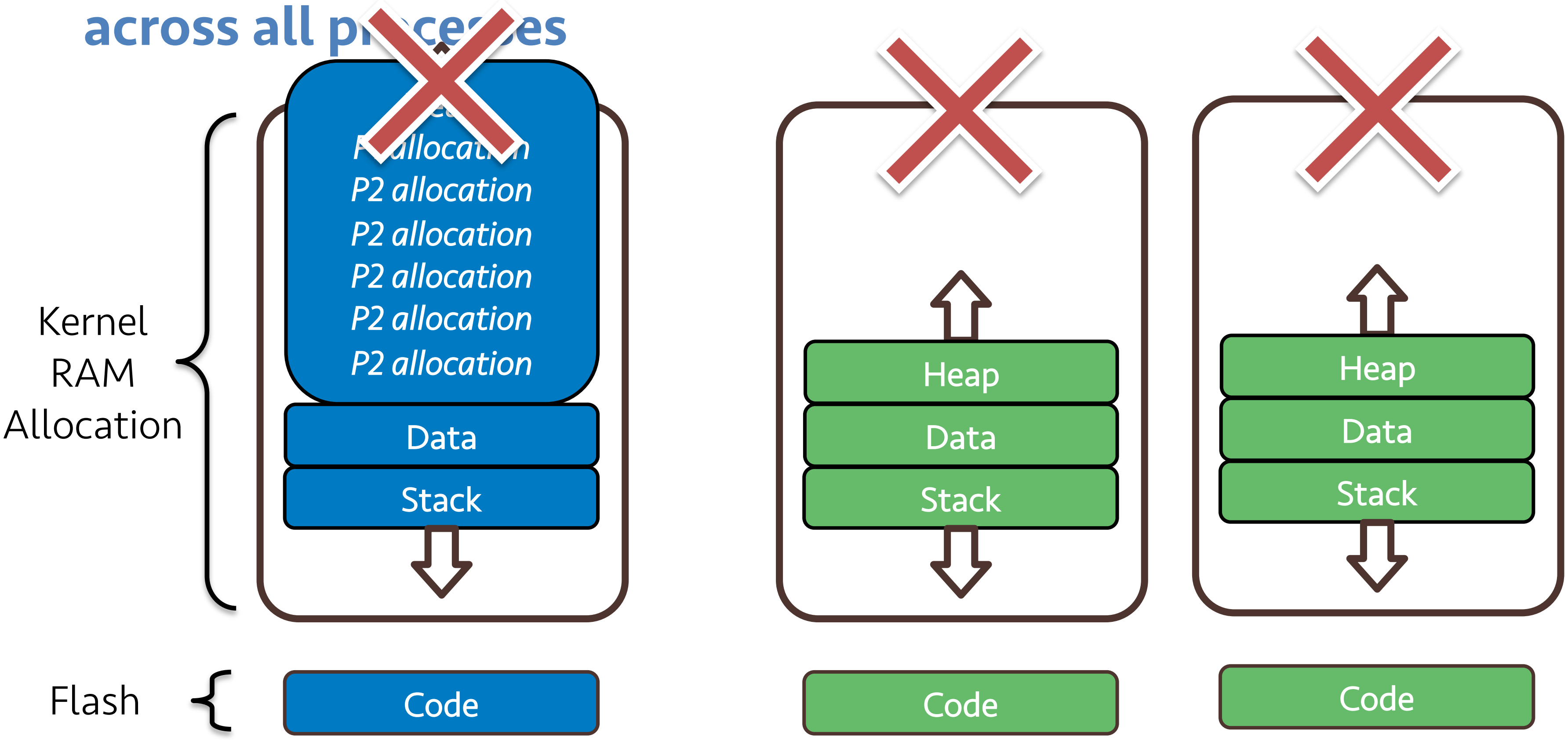
Traditional kernel heap allocation results in shared fate across all processes



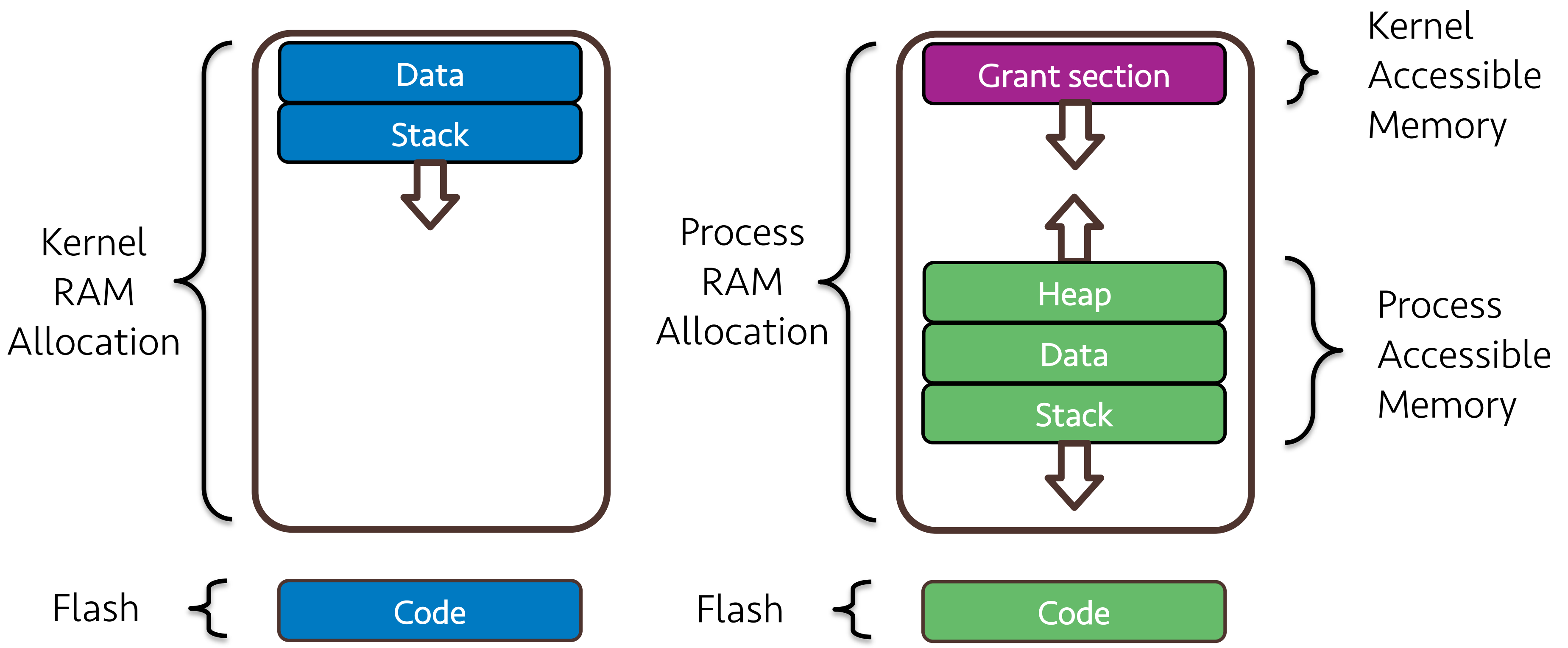
Traditional kernel heap allocation results in shared fate across all processes



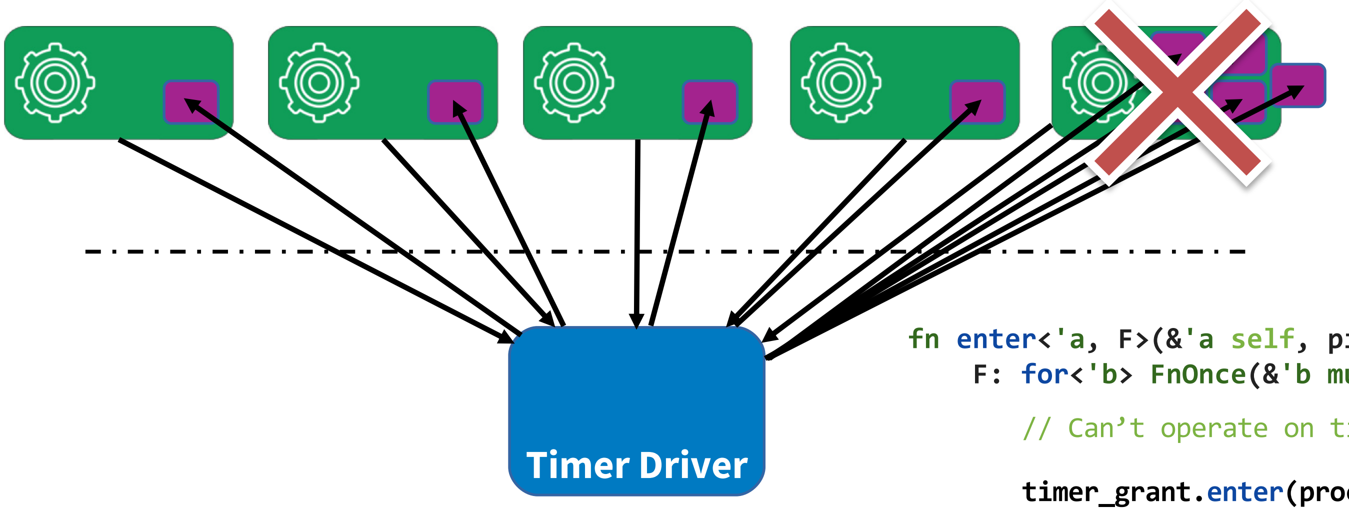
Traditional kernel heap allocation results in shared fate across all processes



Grants safely and efficiently fragment the kernel heap inside the process triggering the dynamic allocation



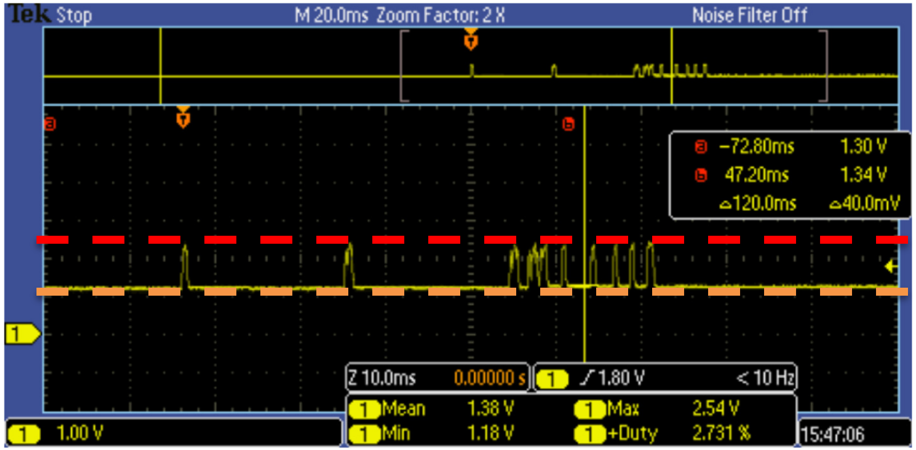
Grants: Kernel heap safely borrowed from processes

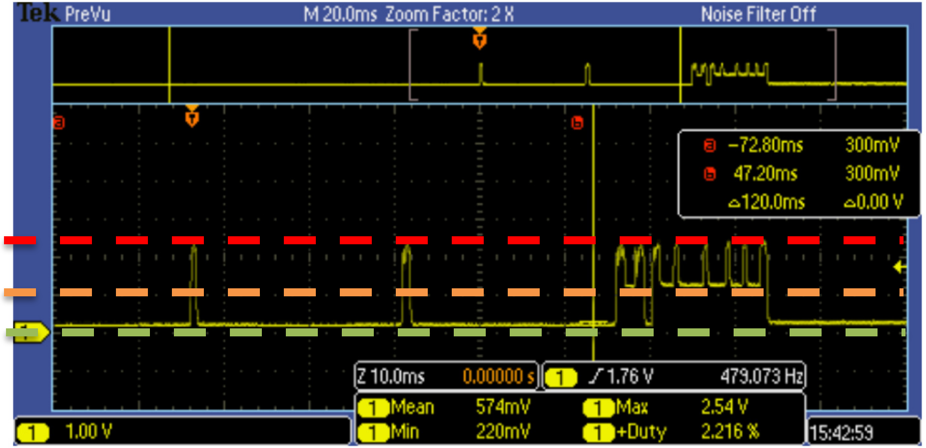



```
fn enter<'a, F>(&'a self, pid: ProcId, f: F) → where
    F: for<'b> FnOnce(&'b mut T)

      // Can’t operate on timer data here

      timer_grant.enter(process_id, |timer| {
          // Can operate on timer data here
          if timer.expiration > cur_time {
              timer.fired = true;
          }
      });

      // timer data can’t escape here
```

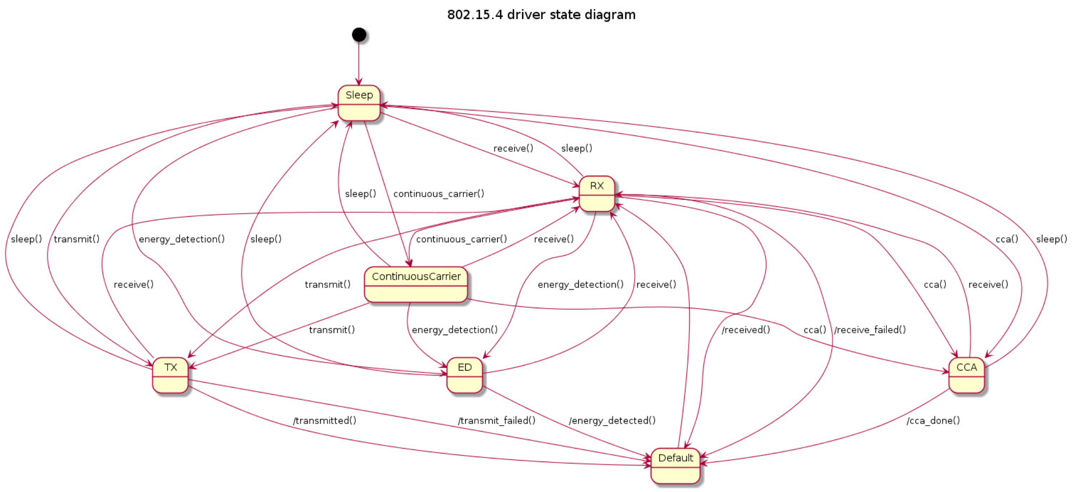
There are more resources than memory…
Correct-by-construction power management

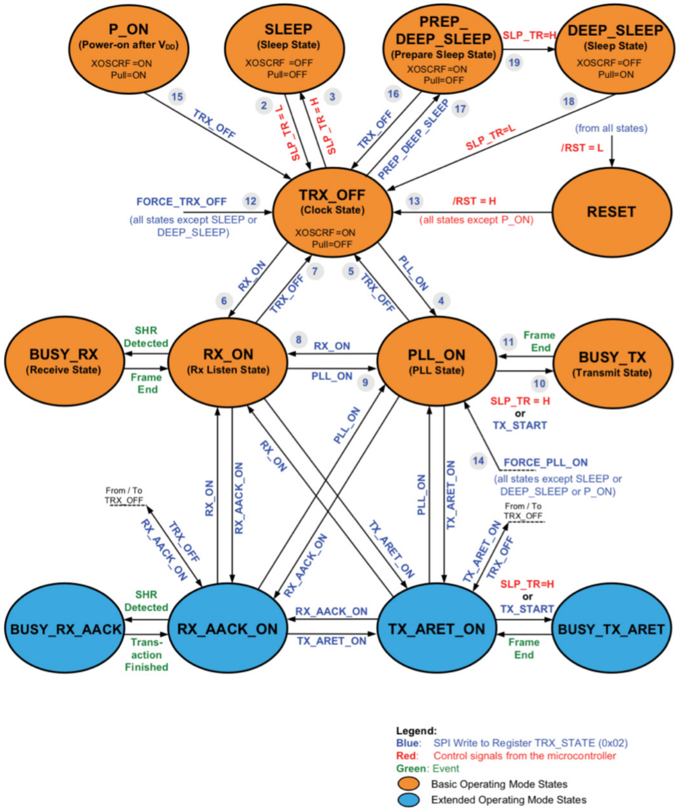
- Challenge: How to ensure peripherals are in the correct power states?
  - On all execution paths?





Modularity mismatch: Isolated peripherals have complex state machines, software mixes execution paths





Ongoing work: Leveraging rich type systems to capture peripheral state machines

Case Study: USART

- 11,000 SLOC
- `clock_enable()/clock_disable()` from **20** calls to just **1**
- Removed 35 `unsafe` blocks
- ~40 minutes of expert developer time
- **20x reduction in energy on test workload**

Deeply embedded platforms today can and should provide the types of safety and correctness guarantees, afforded by modular abstractions, expected from traditional compute platforms

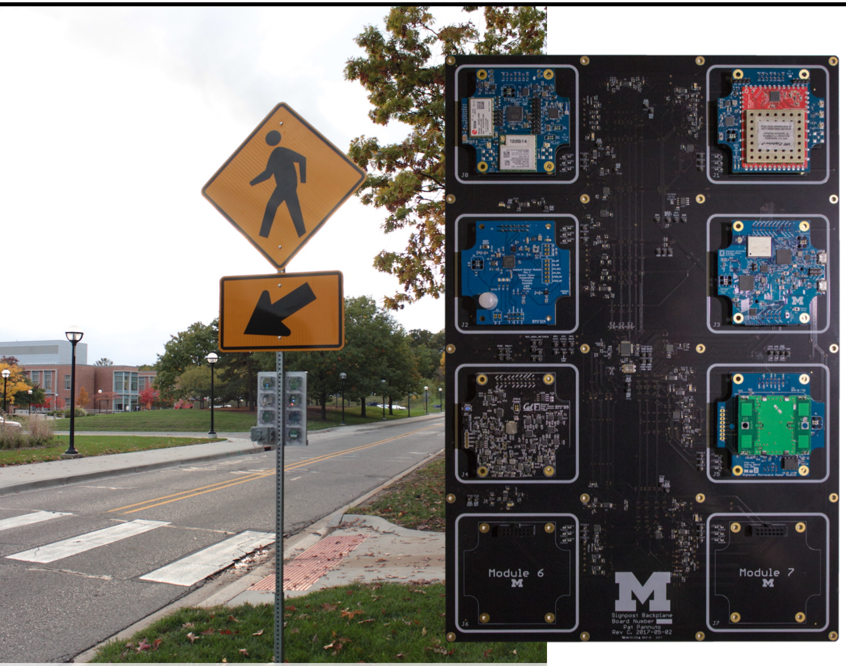
Foreshadowing future work

We've done some of this, but there is more to go

Tock today: In use by universities around the globe, major corporations, startups, and hobbyists. www.tockos.org

4,425 commits | 43 branches | 14 releases | 61 contributors

Unstar 1,155 | Fork 147



The Signpost Platform for City-Scale Sensing
Joshua Adkins, Branden Ghena, Neal Jackson, **Pat Pannuto**, Samuel Rohrer, Bradford Campbell, Prabal Dutta
Information Processing and Sensor Networks (IPSN '18)

Outline

- Introducing the nanopower computing class

- The development of a modular, nanopower computing platform

  Hardware — Composing micro-scale systems

  Software — Returning modularity to microcontroller software

  Services — Deployment and management of “dust”

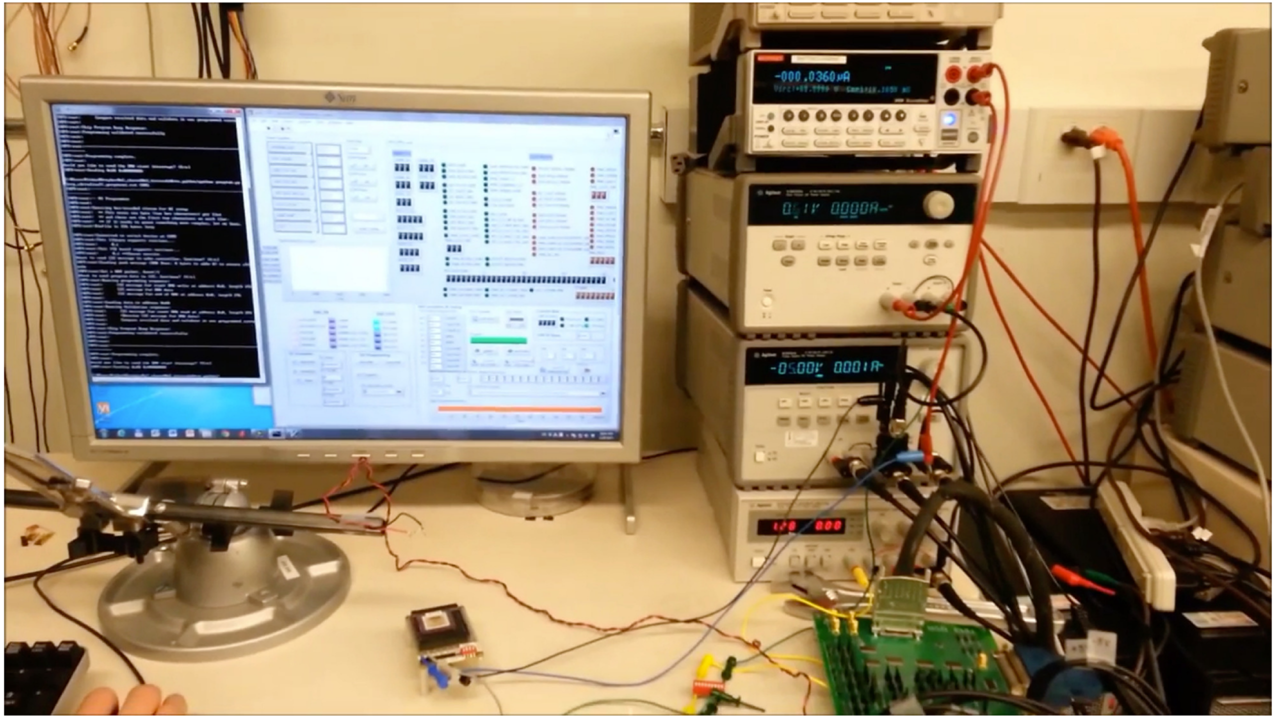
- The future of nanopower computing and its broader impact

Two management questions

1. How do you program something too small to attach wires to?

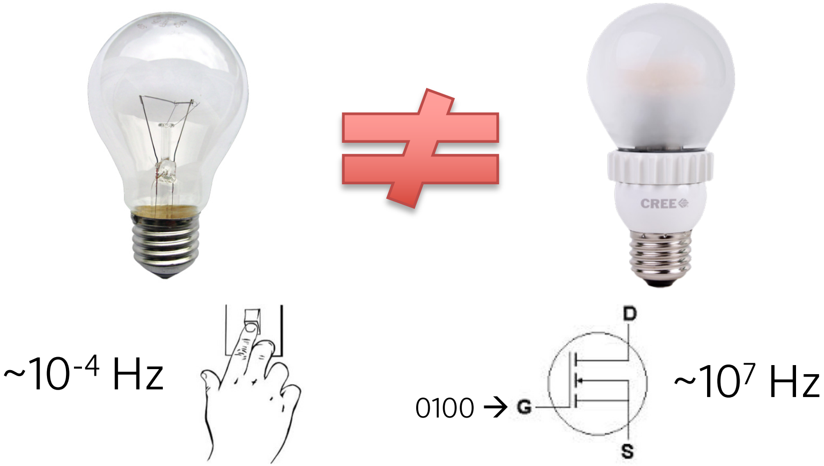
2. How do you keep track of 1,000s of millimeter-sized computers?

We can realize an always-on optical wakeup frontend for just 100’s of picoWatts



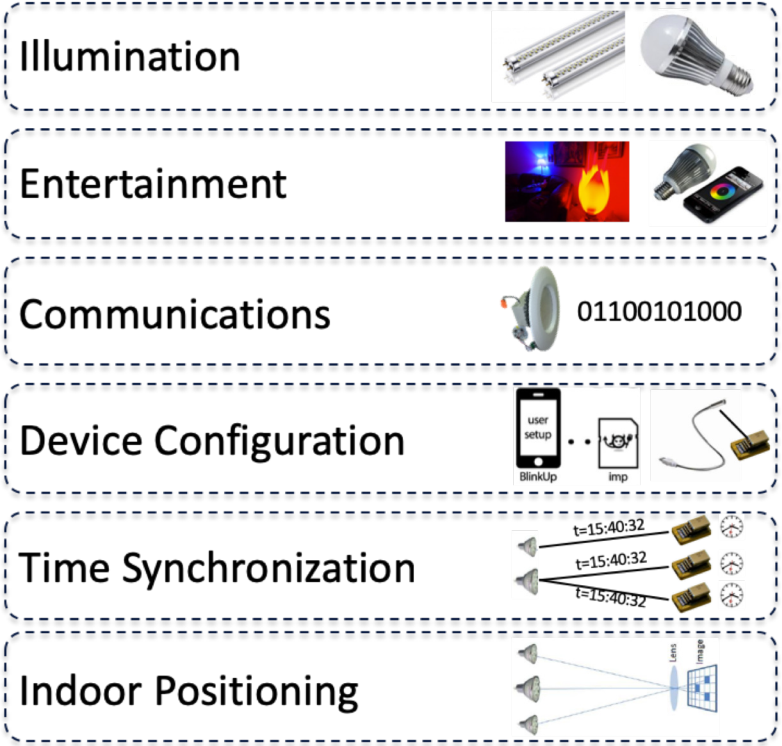
A Millimeter-Scale Wireless Imaging System with Continuous Motion Detection and Energy Harvesting
Gyouho Kim, ZhiYoong Foo, **Pat Pannuto**, Ye-Sheng Kuo, Benjamin Kempke, Mohammad Hassan Ghaed, Suyoung Bang, Inhee Lee, Yejoong Kim, Seokhyeon Jeong, Prabal Dutta, Dennis Sylvester, and David Blaauw
2014 Symposium on VLSI Circuits (VLSIC)

The solid state revolution: not just a substitute good



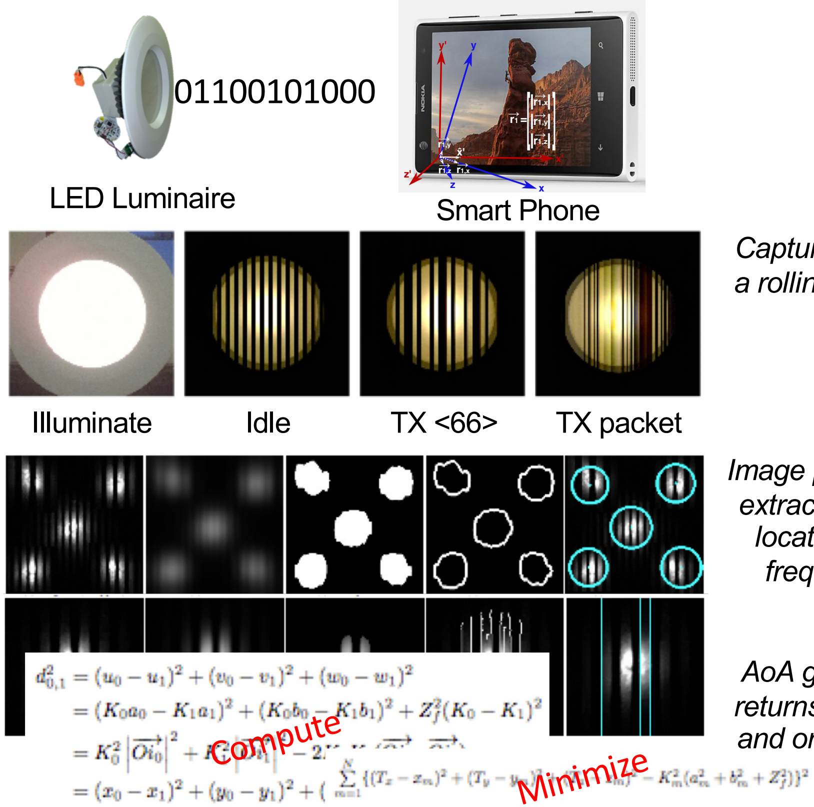
Global replacement of lighting infrastructure!

- European Commission. Commission adopts two regulations to progressively remove from the market non-efficient light bulbs.
- Australian Dept. of Industry. *Energy Labelling and MEPS Program Regulatory Ruling*.
- The 100th US Congress. *Energy Independence and Security Act of 2007. Public Law 110-140.*



System Architecture Directions for a Software-Defined Lighting Infrastructure
Ye-Sheng Kuo, **Pat Pannuto**, and Prabal Dutta
Workshop on Visible Light Communication Systems (VLCS '14)

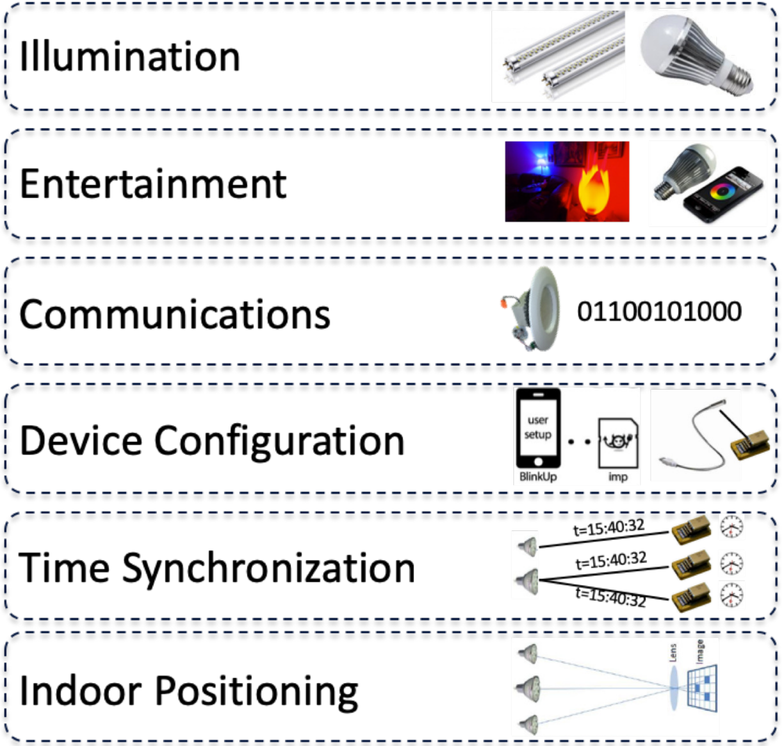
The solid state revolution: not just a substitute good



Captured using a rolling shutter

Image processing extracts beacon locations and frequencies

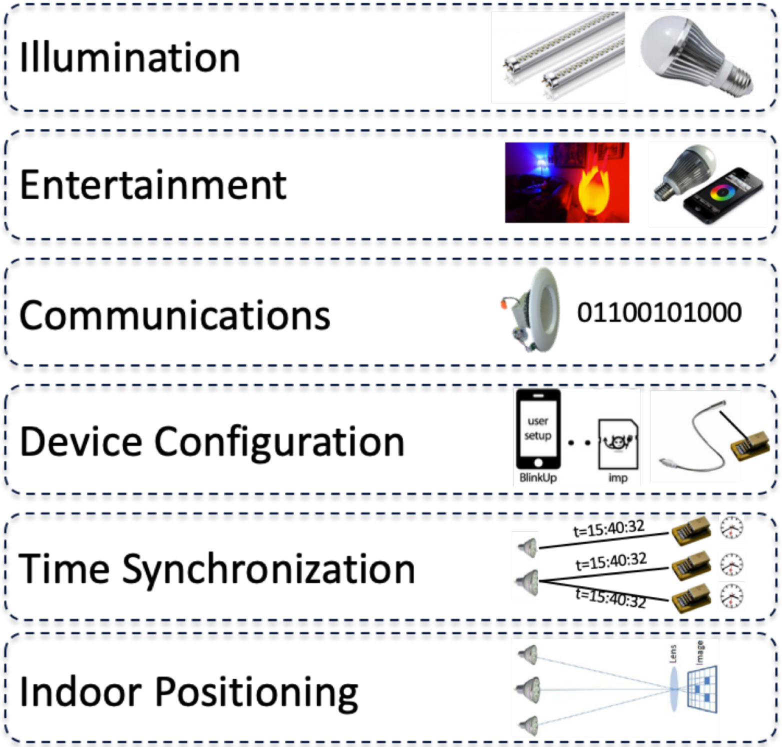
AoA geometry returns position and orientation



Luxapose: Indoor Positioning with Mobile Phones and Visible Light
Ye-Sheng Kuo, **Pat Pannuto**, Ko-Jen Hsiao, and Prabal Dutta
The 20th Annual International Conference on Mobile Computing and Networking (MobiCom '14)

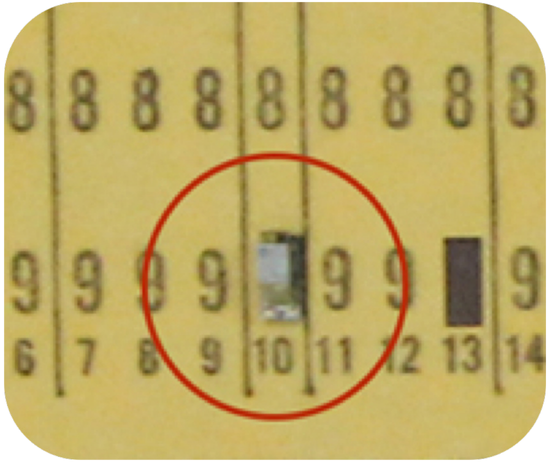
The solid state revolution: not just a substitute good

The multitenant lightbulb is an exciting platform for modular, composable services

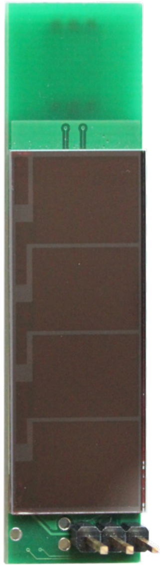


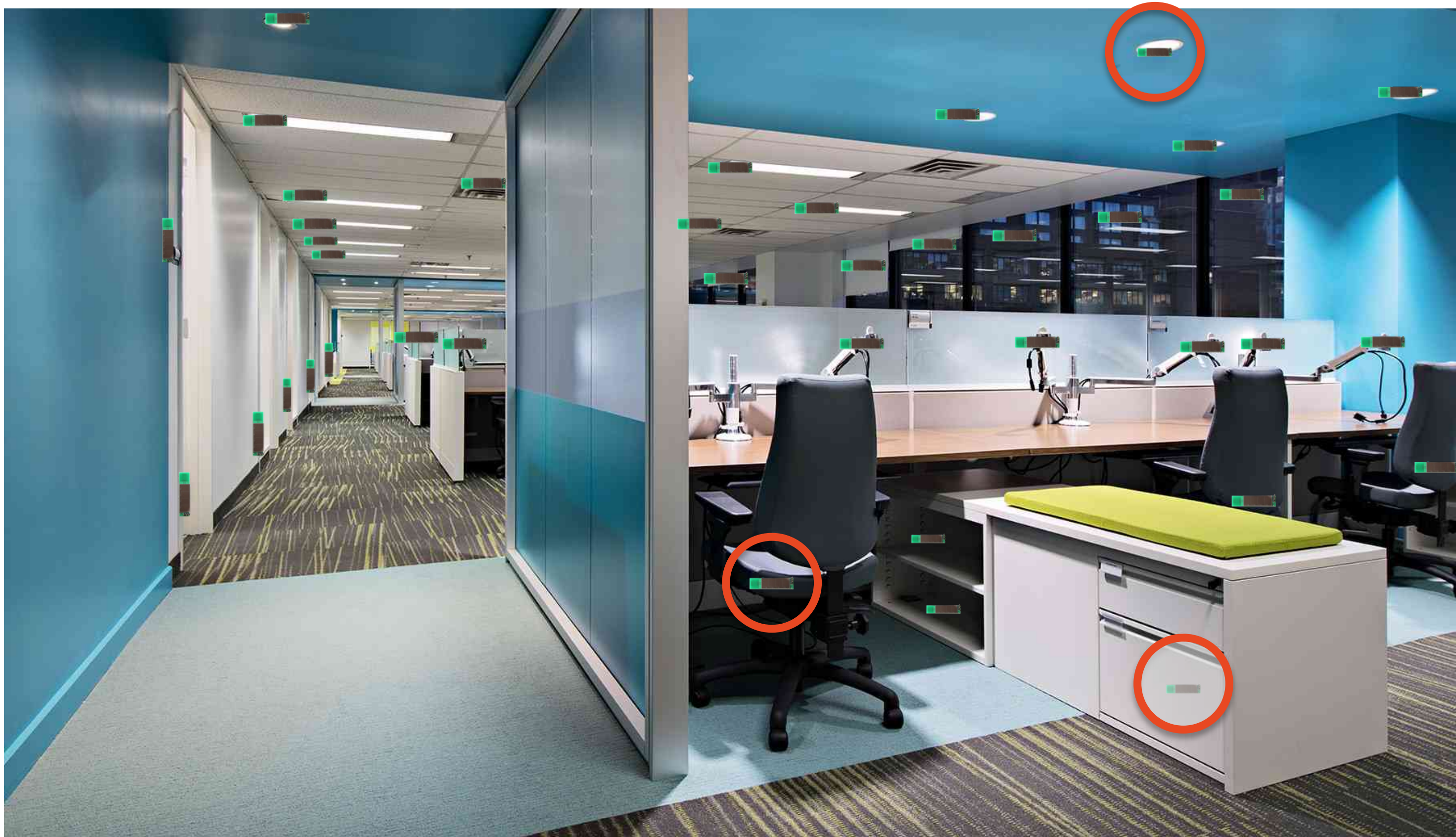
Two management questions

1. How do you program something too small to attach wires to?
   - A modular visible light platform supports programming, configuration, and synchronization of devices
2. How do you keep track of 1,000s of millimeter-sized computers?



Slocalization: Ultra wideband backscatter localization
(Not quite millimeter-sized yet)

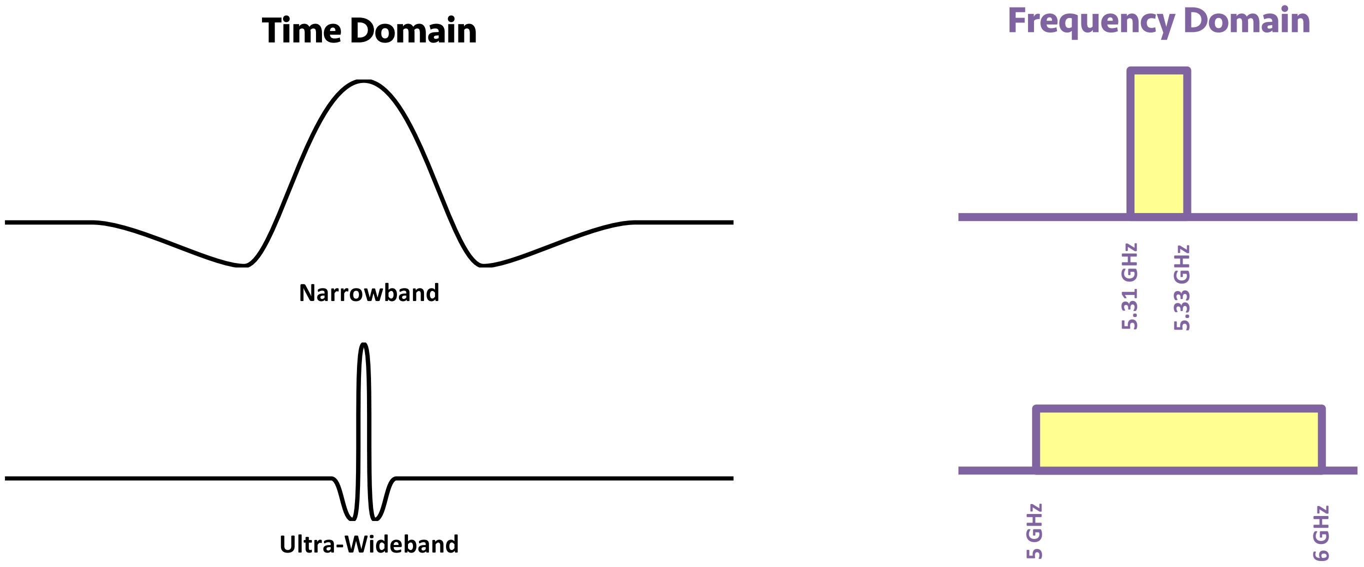




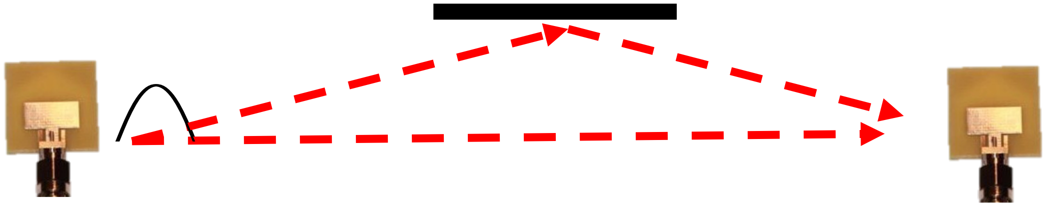
Why RF, why ultra wideband, why backscatter for ubiquitous localization?

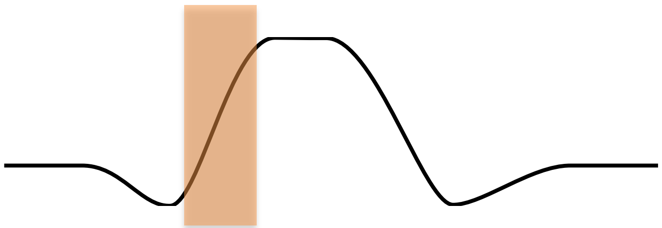
Mautz, Rainer. "Indoor positioning technologies." (2012).

Why RF, why ultra wideband, why backscatter for ubiquitous localization?

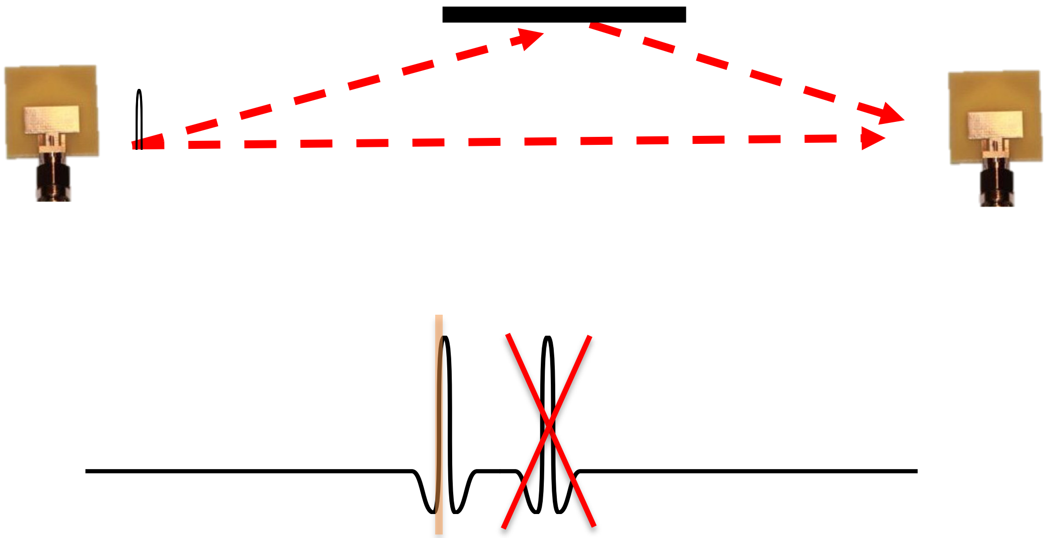


Reflections make time-of-flight estimation difficult and inaccurate

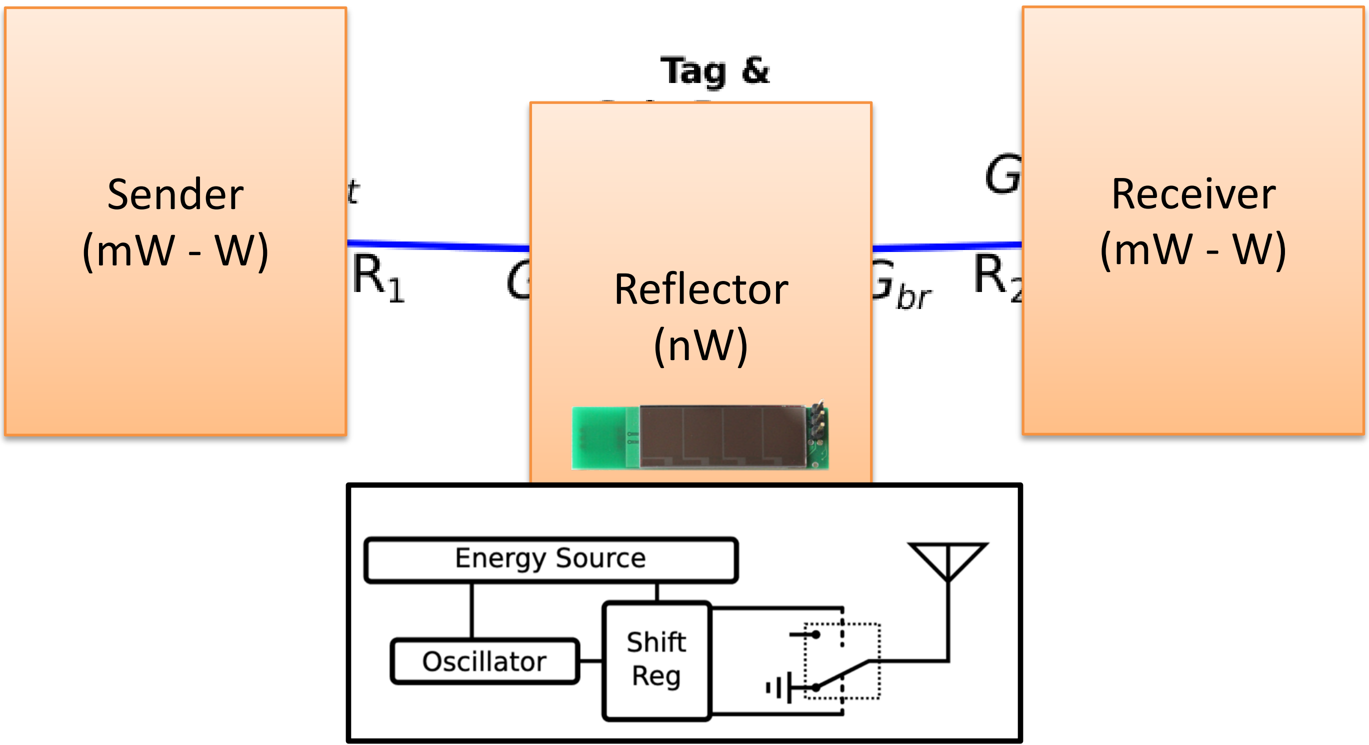




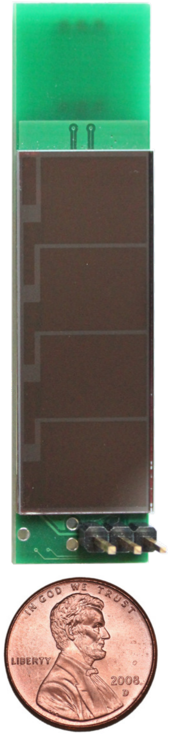
Ultra wideband can better disambiguate multipath and identify signal arrival time



Why RF, why ultra wideband, why backscatter for ubiquitous localization?



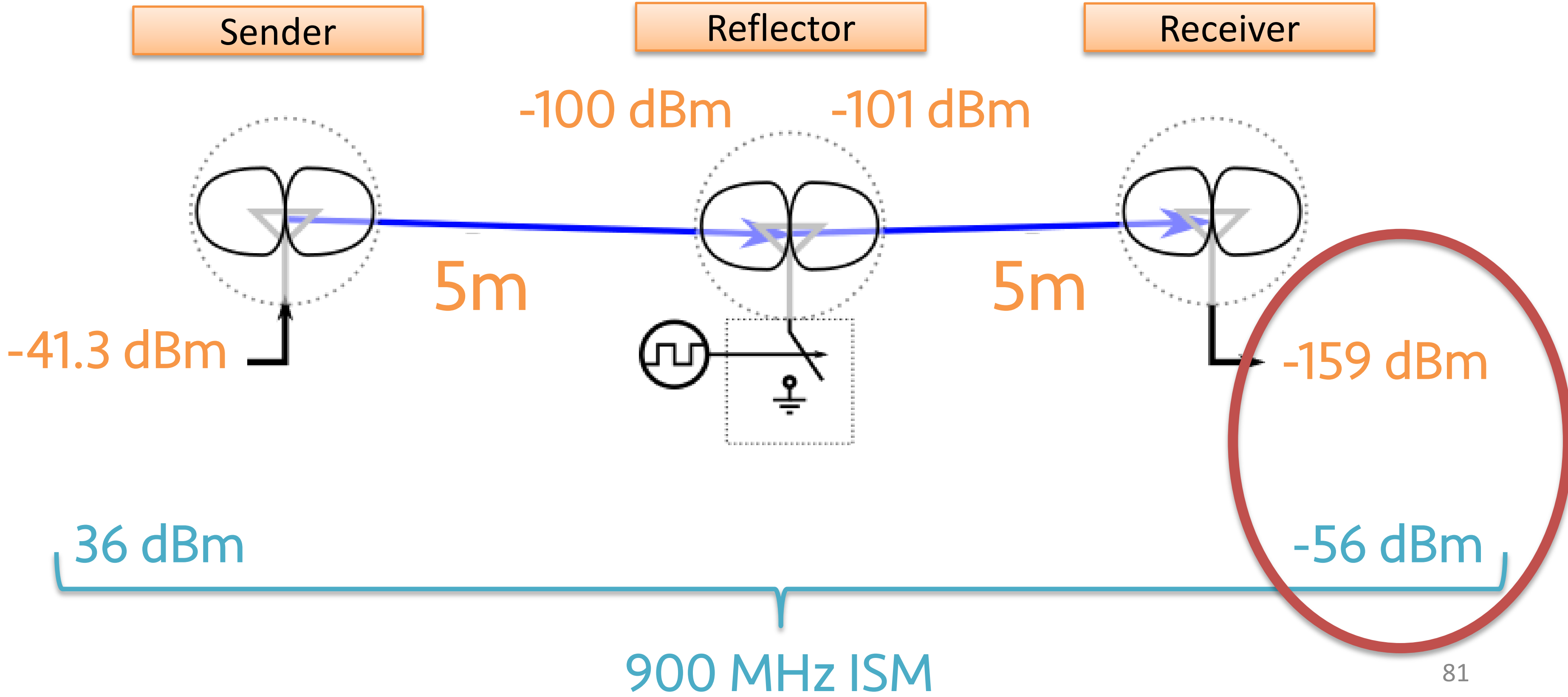
There is a new tradeoff to introduce to enable wide-area ultra-low power, high-quality localization



- Covers areas 30m+
  - "through walls"
- Decimeter accurate
- <1 μW tag
  - (COTS, can do order of magnitude or more better with VLSI)
- (Nearly) unlimited number of concurrent tags
- 1-15+ minutes per location fix
  - A latency/energy tradeoff for localization

Slocalization: Sub-μW Ultra Wideband Backscatter Localization
Pat Pannuto, Kempke, Benjamin, and Prabal Dutta
Proceedings of the 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN'18). **Best Paper Finalist.**

UWB Backscatter is passive reflection of a lot less energy than traditional communications



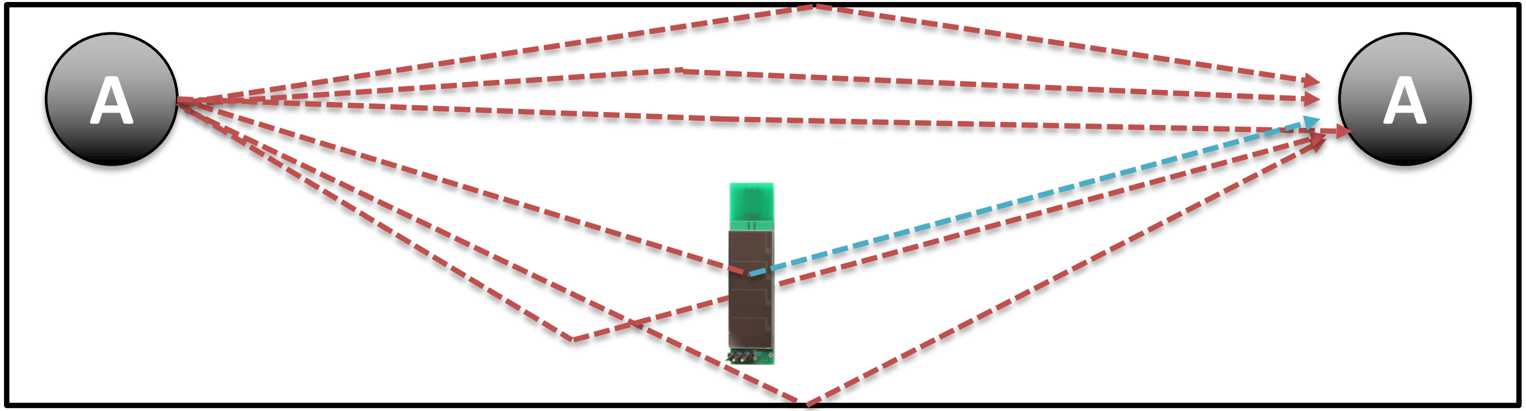
UWB Backscatter is passive reflection of a lot less energy

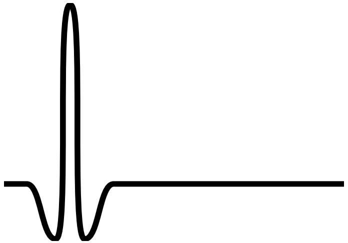
Packet Error Rate	Data Rate	Typical Receiver Sensitivity	Units
1%	110 kbps	-106	dBm/500 MHz
10%	110 kbps	-107	dBm/500 MHz
	[illegible] kbps	-102	[illegible]Bm/500 MHz
1%	850 kbps	-101	dBm/500 MHz
	6.8 Mbps	-93 (*-97)	dBm/500 MHz
10%	110 kbps	-106	dBm/500 MHz
	850 kbps	-102	dBm/500 MHz
	6.8 Mbps	-94 (*-98)	dBm/500 MHz

Typical receiver sensitivity ranges from -94 to -106

How do we recover a signal that is way below the noise floor?

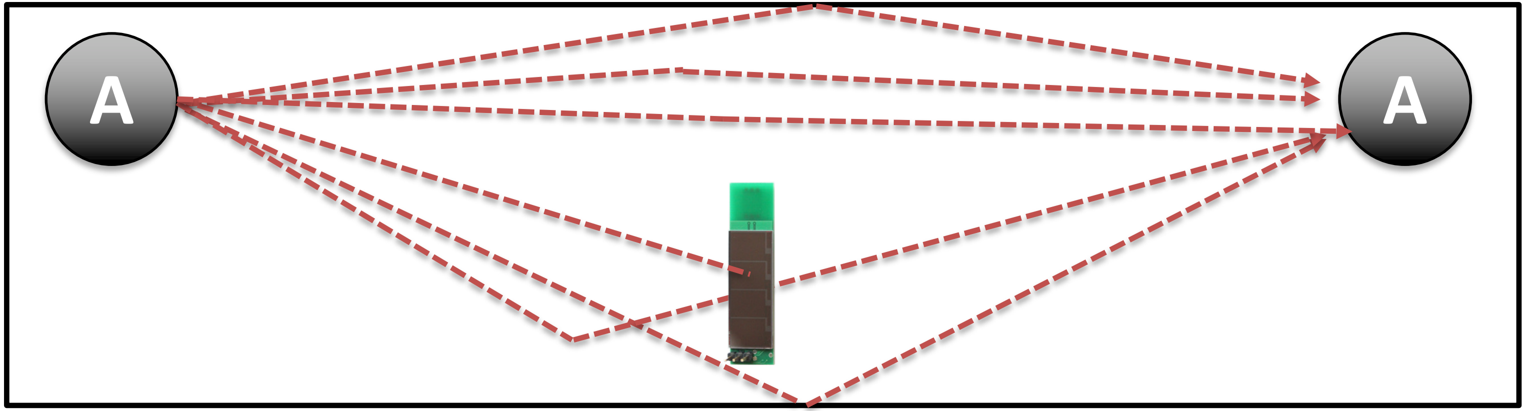
- Exploit tag stationarity and environmental stability

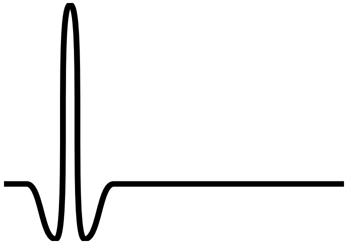


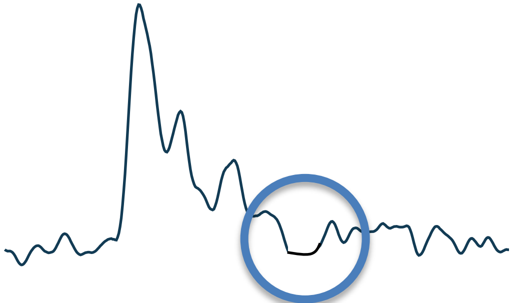


How do we recover a signal that is way below the noise floor?

- Exploit tag stationarity and environmental stability

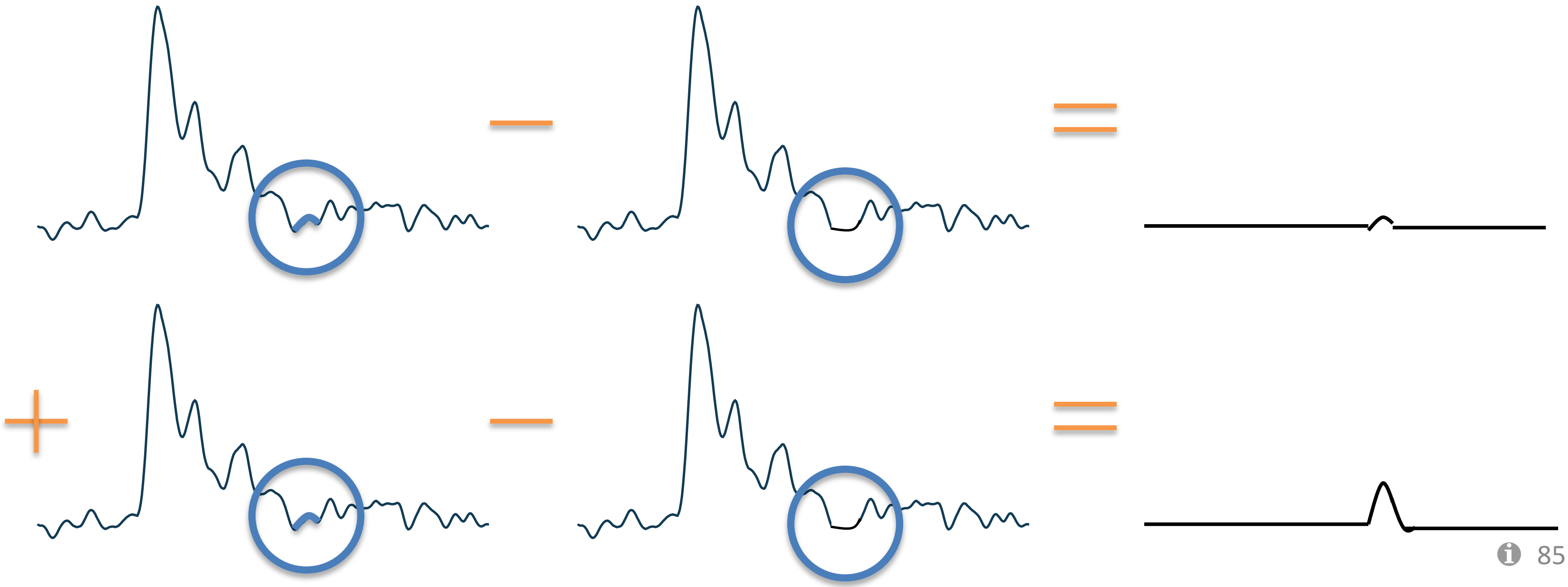




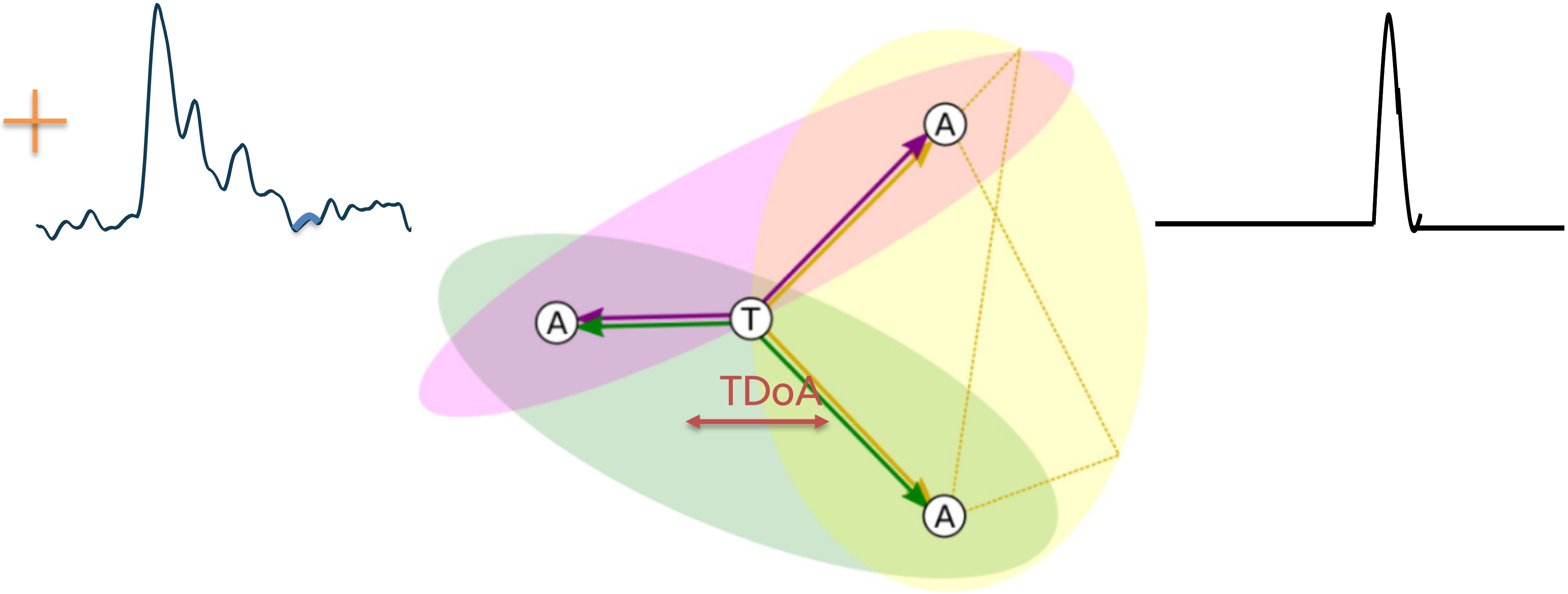


Ideally, the only change in the channel impulse response is the tag reflection

- Subtracting the environment finds the tag



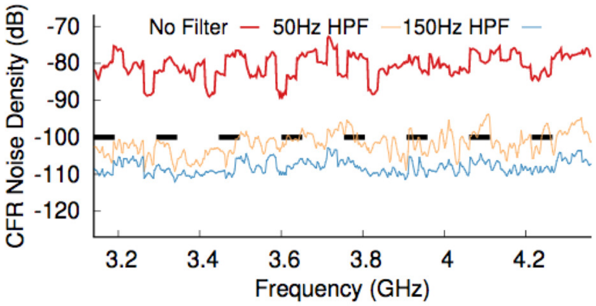
The goal is to estimate the time difference of arrival (TDoA) and laterate



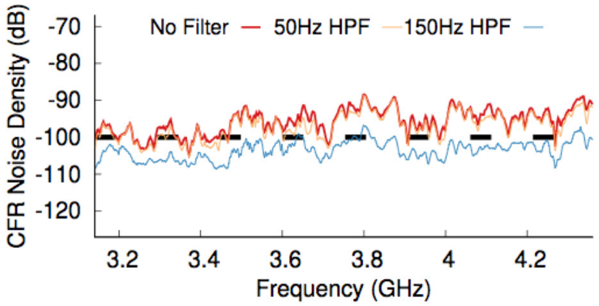
- First peak is anchor—anchor path, then anchor—tag—anchor

Extracting the tag signal in the real world has a few additional challenges

- The environment is not actually static
  - But noise is largely white & Gaussian
  - And we can filter out the rest (sets floor for tag frequency, active power)



(b) CFR Noise Walking



(c) CFR Noise Fluorescents

- Finding tag phase offset currently requires brute force search

Directly generating and recovering UWB signals is challenging (especially circa 2014-2018… changing fast!)

RTLS Systems are black box

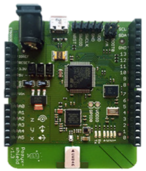
Time Domain P440
(now Humatics)

Ubisense Research Package
(now Investcorp Technology Partners)

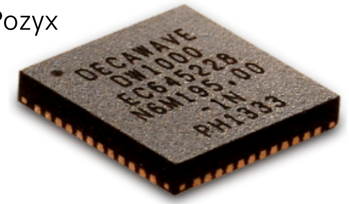
802.15.4a has protocol expectations & overhead

Pixie

Ciholas DWUSB

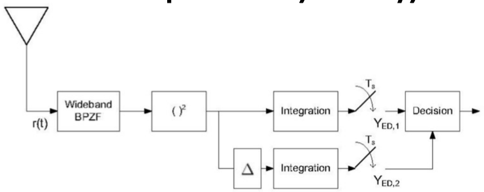


Pozyx

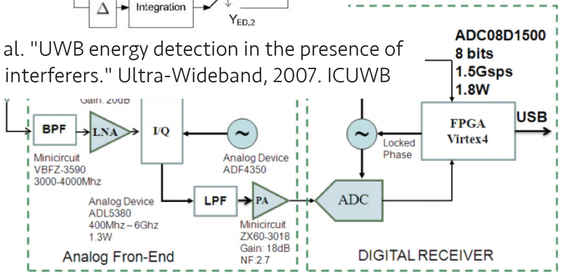


DecaWave DW1000

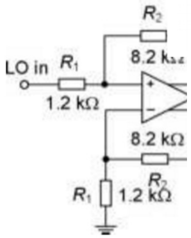
Research receivers: expensive, noisy, or niche



Rabbachin, Alberto, et al. "UWB energy detection in the presence of multiple narrowband interferers." Ultra-Wideband, 2007. ICUWB



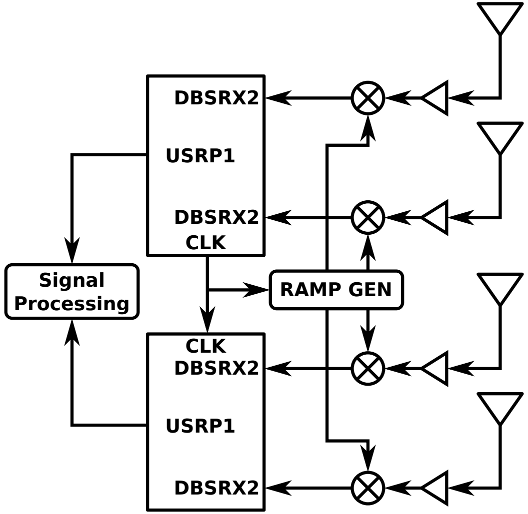
Segura, Marcelo, et al. "Experimental demonstration of self-localized ultra wideband indoor mobile robot navigation system." Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on. IEEE, 2010.

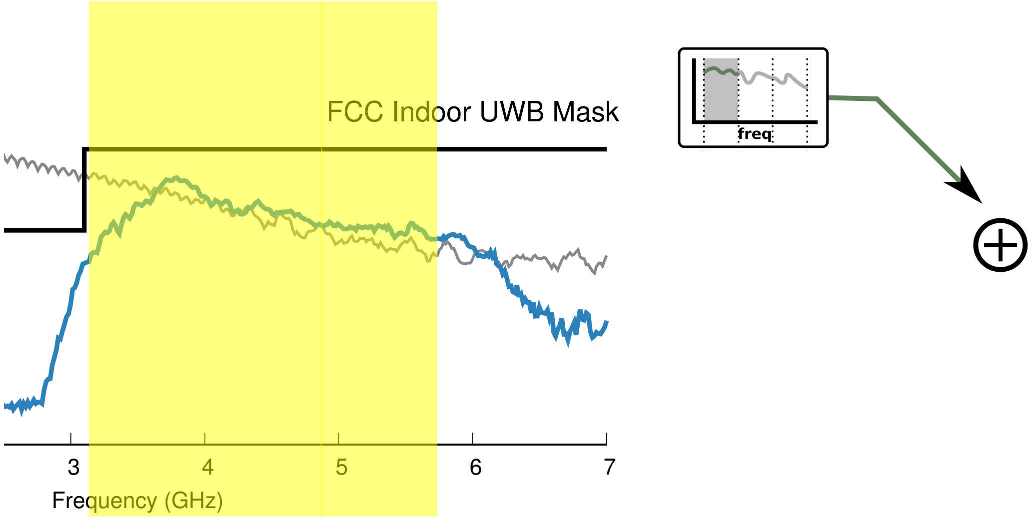


Hantscher, Sebastian, et al. "Hardware concepts for the sequential sampling of repetitive pulse radar echoes in cost-efficient ultra-wideband transceivers."Microwave and Optical Technology Letters 52.3 (2010): 585-591.

Directly generating and recovering UWB signals is challenging (especially circa 2014-2018… changing fast!)

- Developed bandstitched UWB transceiver architecture
  - Generic narrowband SDR (USRP)
  - Measure Channel Frequency Response in 20~25 MHz chunks





Directly generating and recovering UWB signals is challenging (especially circa 2014-2018… changing fast!)

- Developed bandstitched UWB transceiver architecture
  - Generic narrowband SDR (USRP)
  - Measure Channel <u>Frequency</u> Response in 20~25 MHz chunks

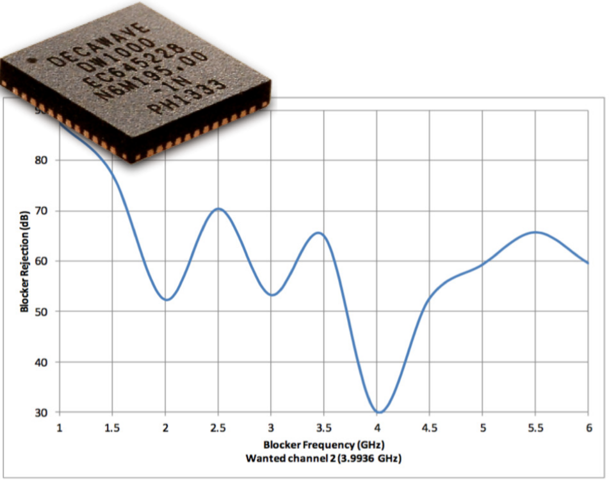
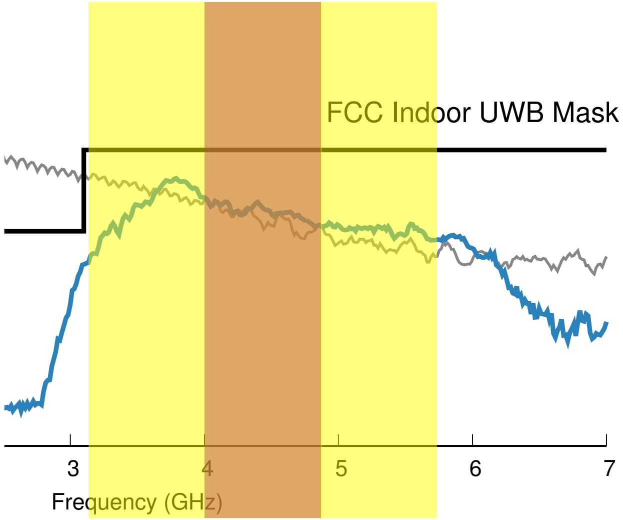
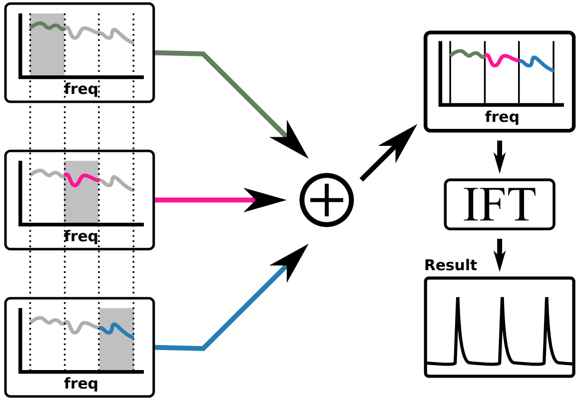


Figure 3 : RX Interferer Immunity on Channel 2

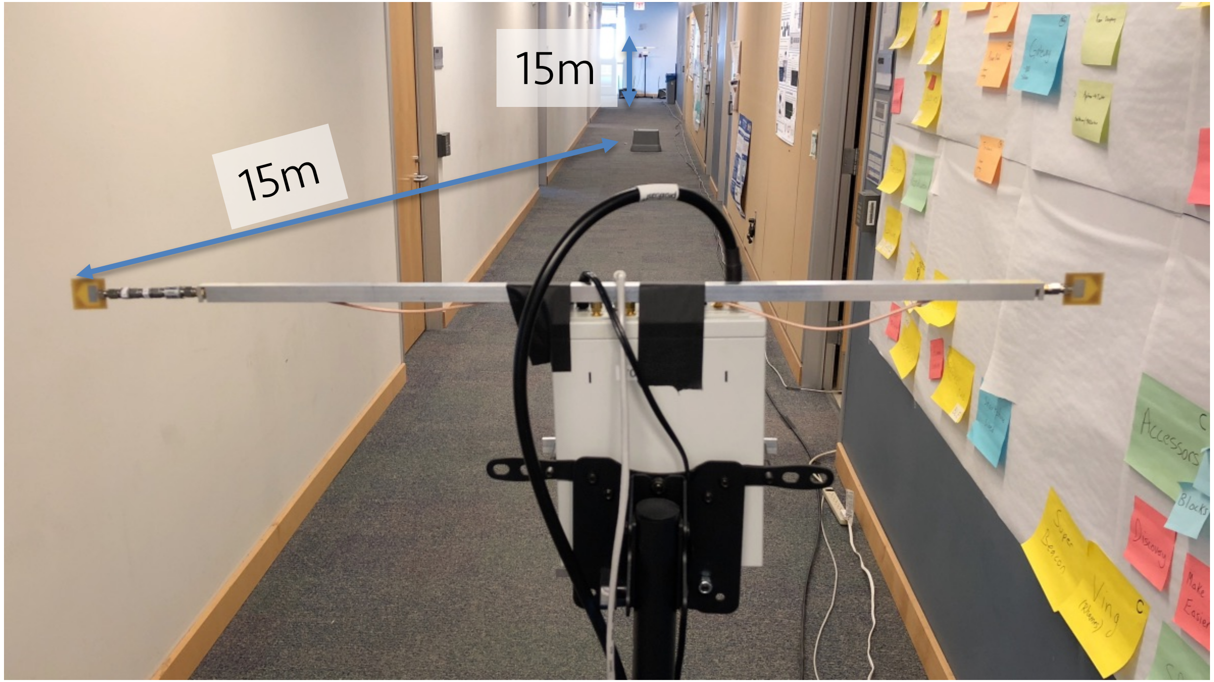
DW1000 Datasheet



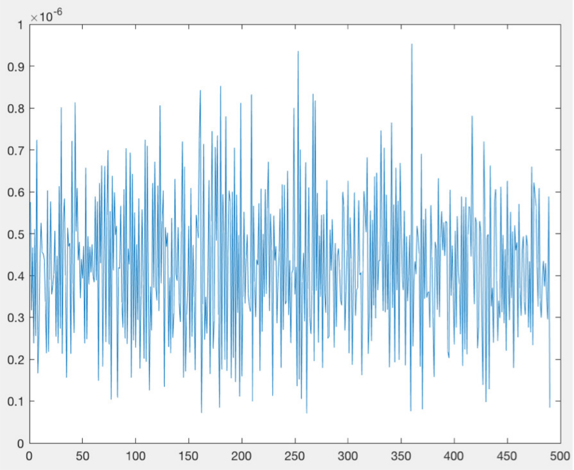


Does it really work?

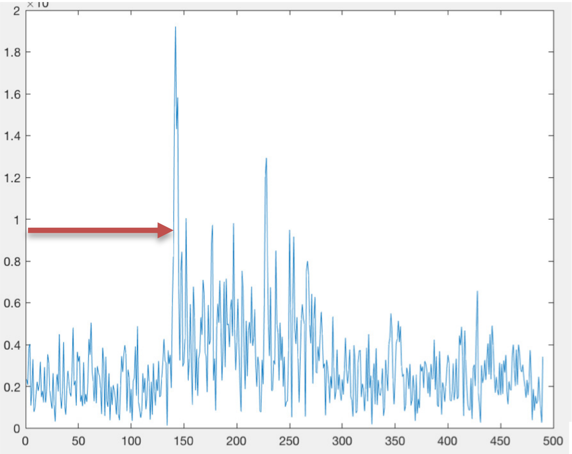
- 15 minutes can cover 30 meters
- 7 cm error (3D Euclidean distance)



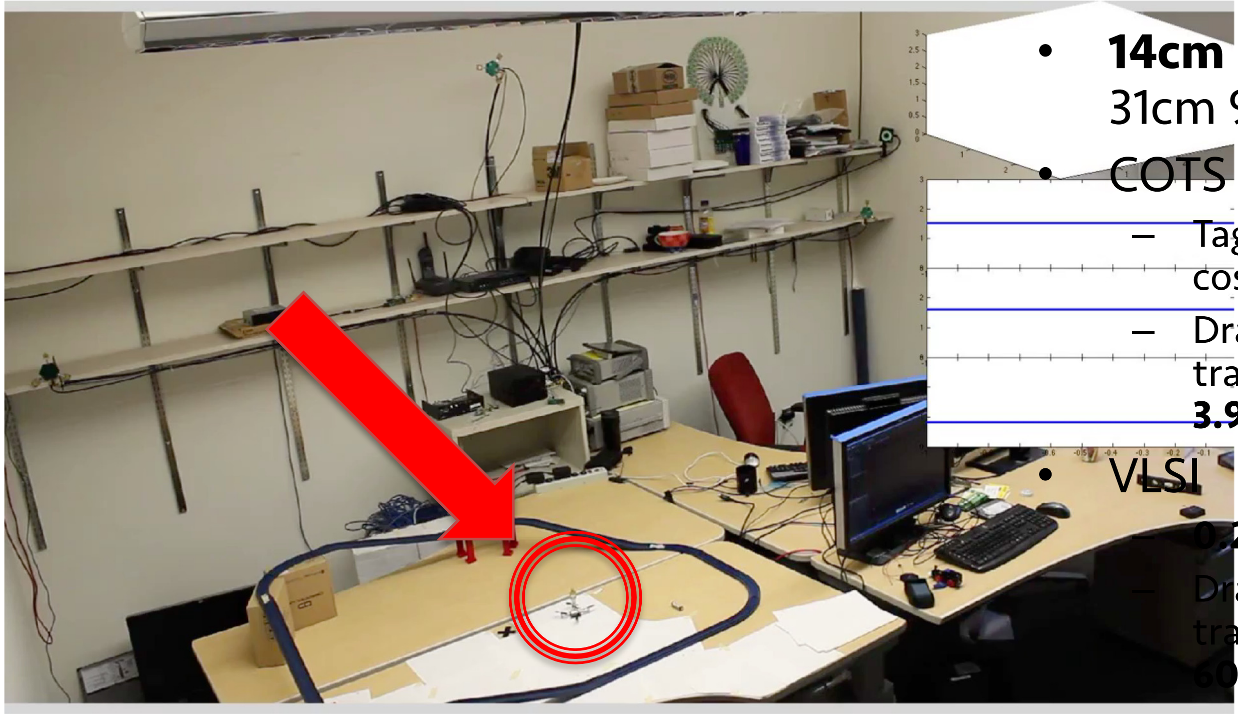
After a few seconds



After a few minutes



The same infrastructure can track moving devices on-demand, enabling adaptivity



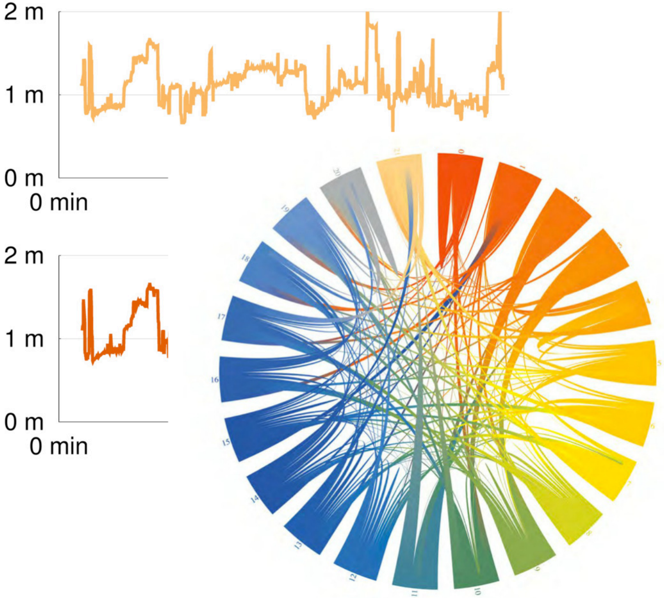
- **14cm median** 31cm 90%ile
- COTS
  - Tag weighs **3g** and costs **~$4.50**
  - Draws **75mW** transmitting, **3.9mJ / fix**
- VLSI
  - **0.2 mm^2** IC
  - Draws **1mW** transmitting, **60µJ / fix**

Harmonia: Wideband Spreading for Accurate Indoor RF Localization (HotWireless’14)
Harmonium: Asymmetric, Bandstitched UWB for Fast, Accurate, and Robust Indoor Localization (IPSN’16)
Harmonium: Ultra Wideband Pulse Generation with Bandstitched Recovery for Fast, Accurate, and Robust Indoor Localization (TOSN’18)

Part of a broader collection of localization technologies

Human interaction tracking

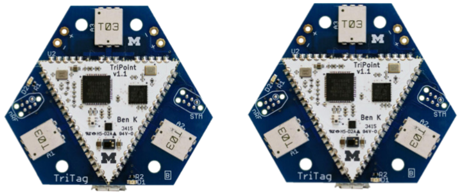
– Opo. 93h battery, 5cm, 0.5 Hz

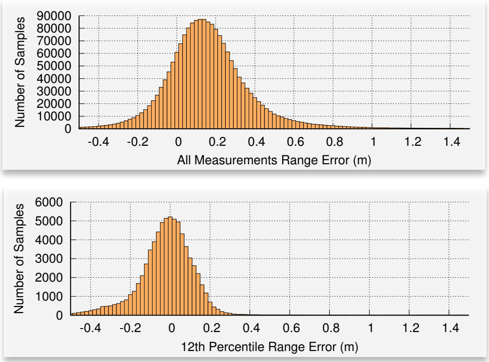


Opo: A Wearable Sensor for Capturing High-Fidelity Face-to-Face Interactions (SenSys'14)

Robust Ranging

– SurePoint. 53cm 99th %ile





SurePoint: Exploiting Ultra Wideband Flooding and Diversity to Provide Robust, Scalable, High-Fidelity Indoor Localization (SenSys'16)

Two management questions

1. How do you program something too small to attach wires to?
   – A modular visible light platform supports programming, configuration, and synchronization of devices

2. How do you keep track of 1,000s of millimeter-sized computers?
   – Perhaps with ultra wideband backscatter

Outline

- Introducing the nanopower computing class
- The development of a modular, nanopower computing platform

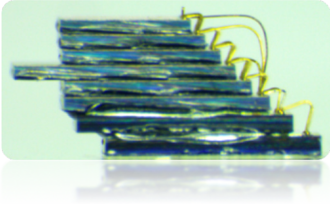
  Hardware — Composing micro-scale systems

  Software — Returning modularity to microcontroller software

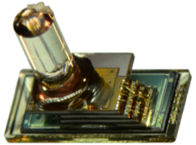
  Services — Deployment and management of "dust"
- The future of nanopower computing and its broader impact

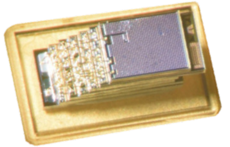
Smart Dust Delivered.
Delivering the millimeter-scale computing class

- Communication is always the hardest part
  - Backscatter is promising but carries infrastructure demands
  - First systems used VLC! Can we bootstrap communication with localization?

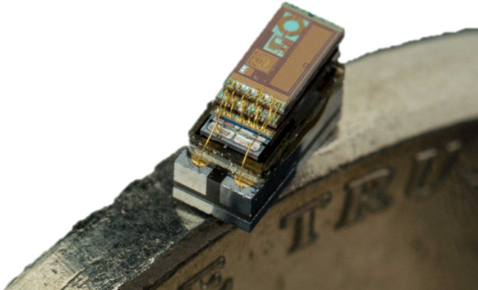


- Advanced computation at the edge
  - Executing ML in resource-constrained environments
  - Training ML from physically challenging deployments (federated learning?)





- Security, Privacy, Ownership, and Enforcement
  - Discovery and interaction paradigms for owners, visitors, or opportunists
  - Long-running tasks across an evolving physical compute fabric
  - Detecting illicit devices

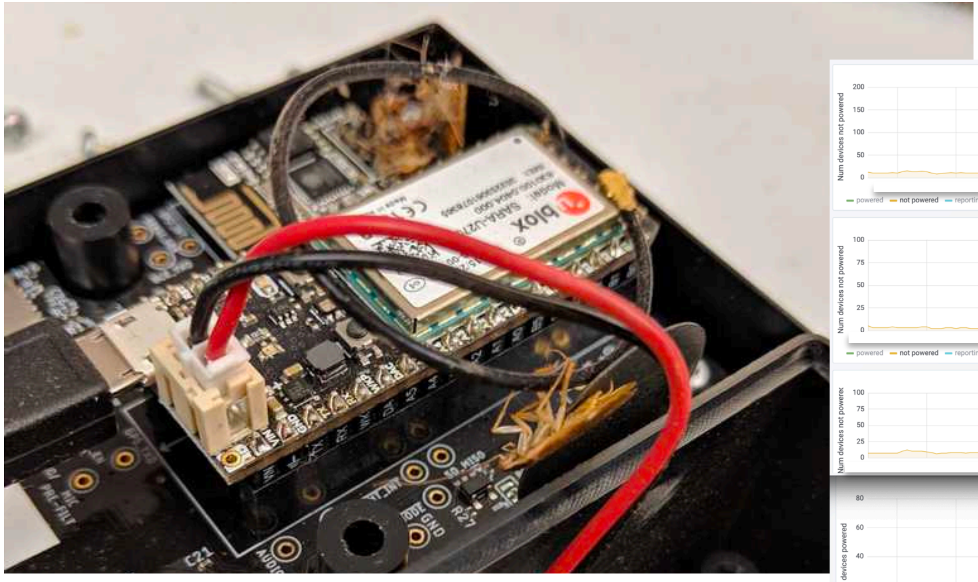


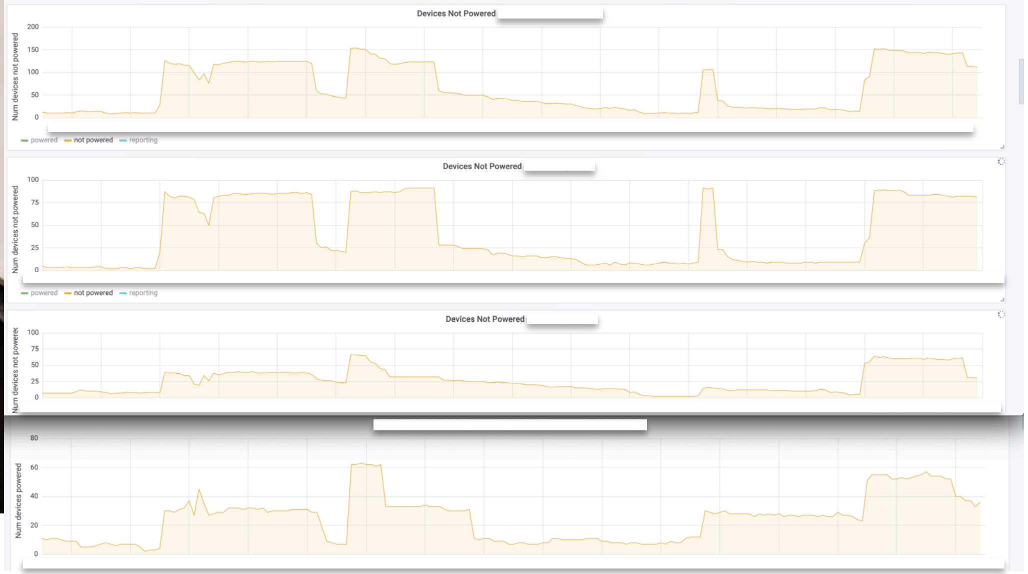
The first two theses?

- Revitalizing microarchitecture for microcontrollers
  - MMUs aren't actually expensive if you only have 64k of memory…
  - Where are the 64-bit MCUs? What about an IOMPU?
  - RISC-V is a clean slate opportunity!
- How to guarantee a battery-powered system will last for 10 years?
  - Correct-by-construction power management
  - Hardware and OS runtime monitoring
  - Energy-adaptive applications; programming environments?

“Digitizing the Physical World,” aka what can’t your computing access that you really want to?

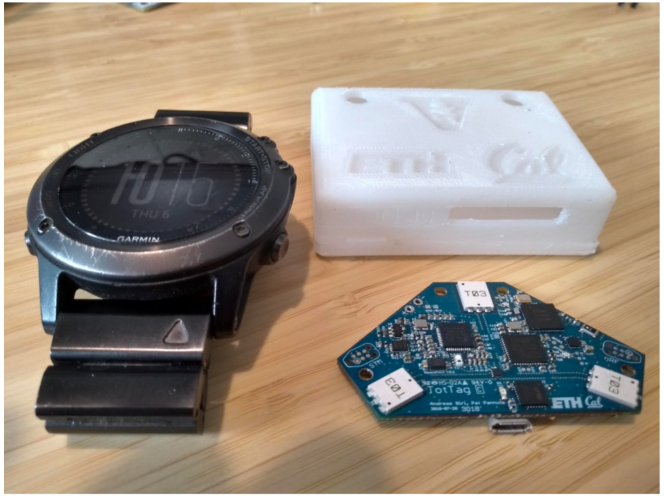
- Application verticals are a joy of embedded research, currently:
  - City/country-scale power grid health monitoring with Berkeley economists

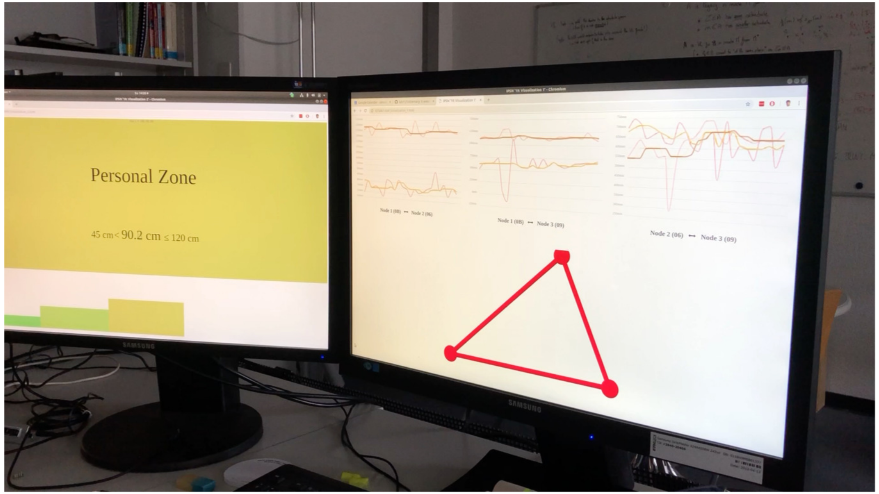




“Digitizing the Physical World,” aka what can’t your computing access that you really want to?

- Application verticals are a joy of embedded research, currently:
  - City/country-scale power grid health monitoring with Berkeley economists
  - Longitudinal parent/child interaction tracking with Vanderbilt psychologists





“Digitizing the Physical World,” aka what can’t your computing access that you really want to?

- Application verticals are a joy of embedded research, currently:
  - City/country-scale power grid health monitoring with Berkeley economists
  - Longitudinal parent/child interaction tracking with Vanderbilt psychologists
- Democratizing embedded systems & untethering the maker movement
  - Shrinking the chasm between Arduino tutorials and engineered product
  - Freshman engineers should be exposed to hardware!

Enhancing access to computer science more broadly

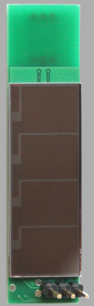
- Course for early-career EECS students (600+ students over 2 semesters)
  - This course is invaluable for students who want to be in EECS but don't have the background that independent programmers come in with. Classes like this are what enable students who feel behind to enter classes at the same level as others with more experience. I really appreciate everyone who set up this course – Student 2
  - Simply put, it's hard to be a newcomer to CS. I think the lecturers did a really good job of erasing some of the barriers of learning new skills – Student 5

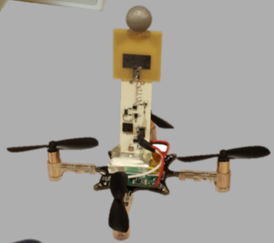
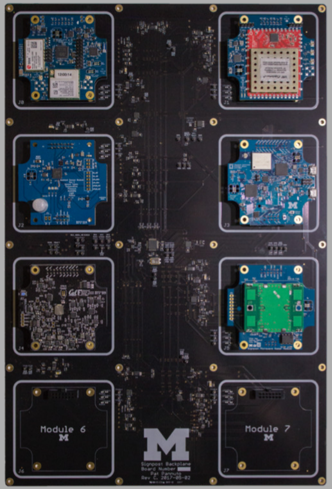
- What I am thinking about next
  – Empower people to interact with their physical world

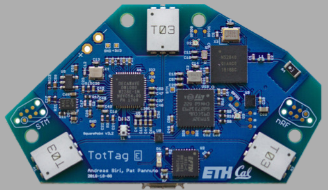
A Modular Platform for Nanopower Computing

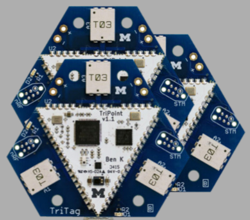
Pat Pannuto, UC Berkeley

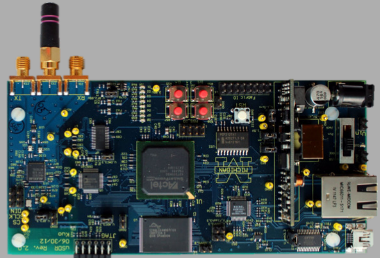
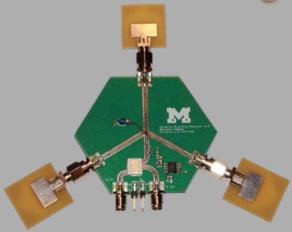
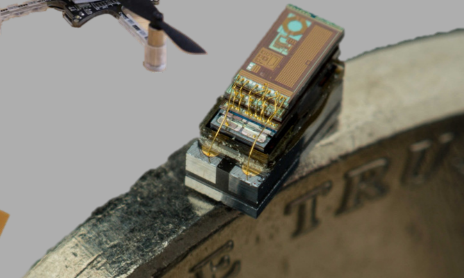
https://patpannuto.com
ppannuto@berkeley.edu

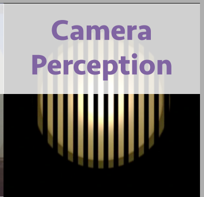






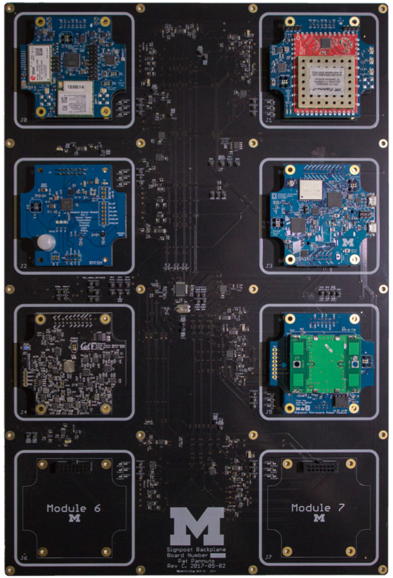



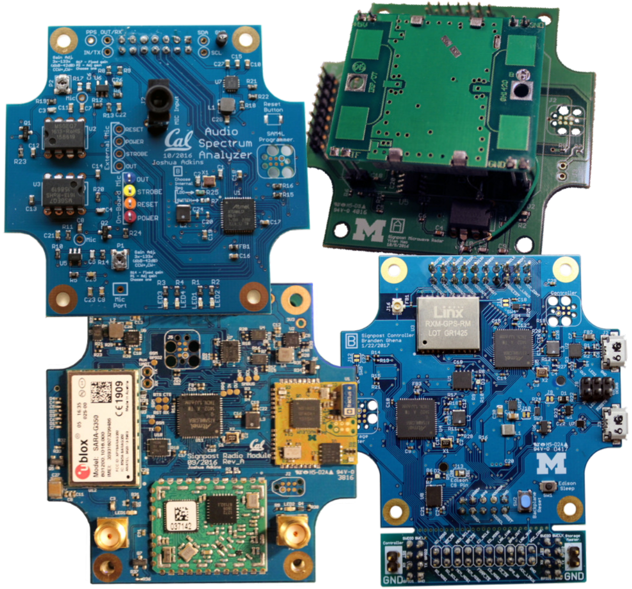
What I'd like to hear from you

- What do you (or your algorithms) want to know about the world?
  - What data do you want that you don't have?
  - And why can't you get it?

The Signpost platform: Infrastructure-free infrastructure for city-scale sensing applications

- This means multiple, independent, untrusted applications must share

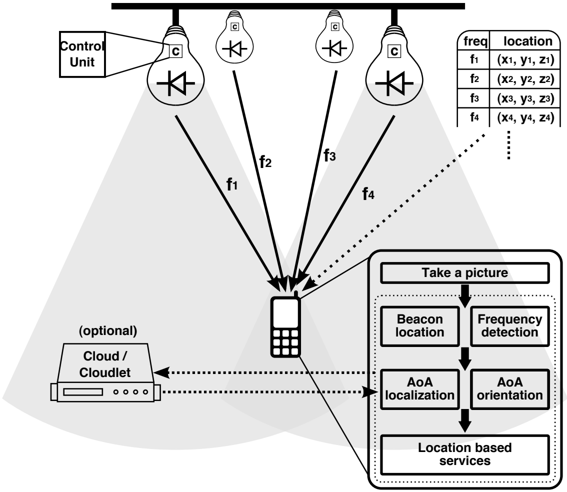


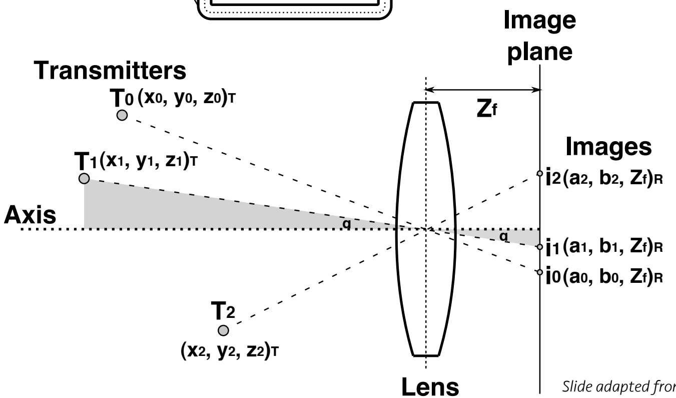


The Signpost Platform for City-Scale Sensing
Joshua Adkins, Branden Ghena, Neal Jackson, **Pat Pannuto**, Samuel Rohrer, Bradford Campbell, Prabal Dutta
Information Processing and Sensor Networks (IPSN '18)

VLCP: Visible light communications and positioning

- LED luminaires
  - Slightly-modified
  - Transmit beacons
  - Identities or coordinates
- Smart phones
  - Run background mobile app
  - Take images periodically
  - Perform local processing
  - Offload to cloud/cloudlet
- Cloud/cloudlet server
  - Do photogrammetry
  - Do AoA Localization
  - Estimate location
  - Estimate orientation
  - Provide location-based services



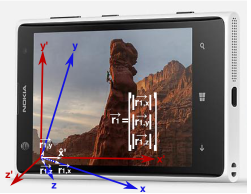


Luxapose: Indoor Positioning with Mobile Phones and Visible Light
Ye-Sheng Kuo, **Pat Pannuto**, Ko-Jen Hsiao, and Prabal Dutta
The 20th Annual International Conference on Mobile Computing and Networking (MobiCom '14)

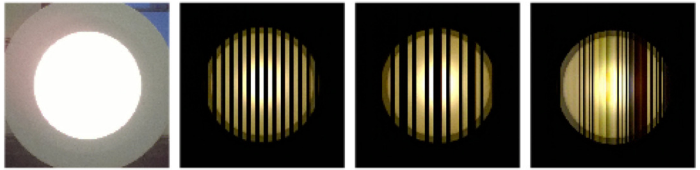
Slide adapted from Prabal Dutta, adapted from Ye-Sheng Kuo

Indoor localization with VLC and astral navigation

LED Luminaire

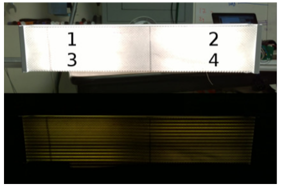


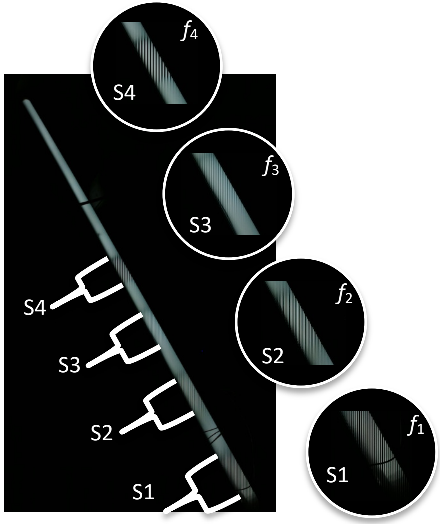
Smart Phone



Illuminate | Idle | TX <66> | TX packet

Captured using a rolling shutter





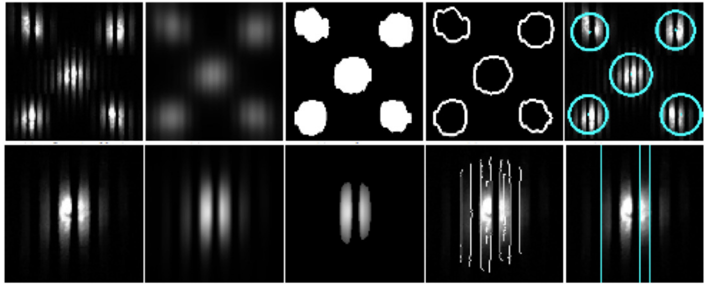


Image processing extracts beacon locations and frequencies

$$
\begin{aligned}
d_{0,1}^2 &= (u_0-u_1)^2+(v_0-v_1)^2+(w_0-w_1)^2\\
&= (K_0a_0-K_1a_1)^2+(K_0b_0-K_1b_1)^2+Z_f^2(K_0-K_1)^2\\
&= K_0^2\left|\overrightarrow{Oi_0}\right|^2+K_1^2\left|\overrightarrow{Oi_1}\right|^2-2K_0K_1(\overrightarrow{Oi_0}\cdot\overrightarrow{Oi_1})\\
&= (x_0-x_1)^2+(y_0-y_1)^2+(z_0-z_1)^2,
\end{aligned}
$$

Compute

$$\sum_{m=1}^{N}\{(T_x-x_m)^2+(T_y-y_m)^2+(T_z-z_m)^2-K_m^2(a_m^2+b_m^2+Z_f^2)\}^2$$

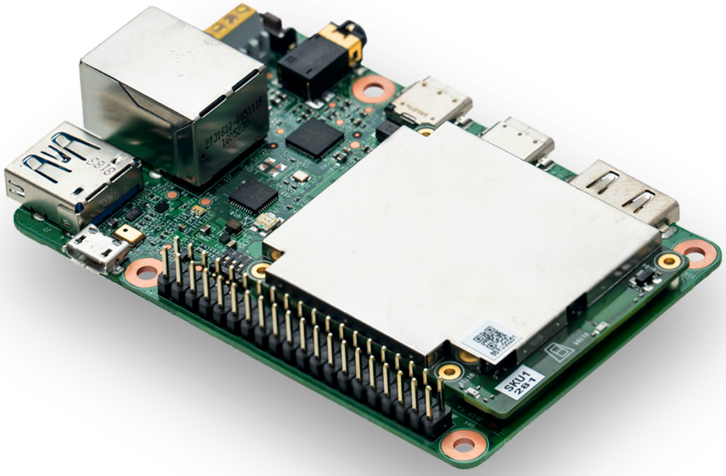
Minimize

Luxapose: Indoor Positioning with Mobile Phones and Visible Light
Ye-Sheng Kuo, **Pat Pannuto**, Ko-Jen Hsiao, and Prabal Dutta
The 20th Annual International Conference on Mobile Computing and Networking (MobiCom '14)

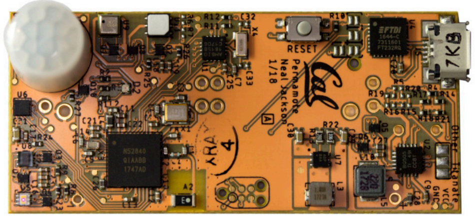
Slide adapted from Prabal Dutta, adapted from Ye-Sheng Kuo

Backup: How does machine learning fit into all this?

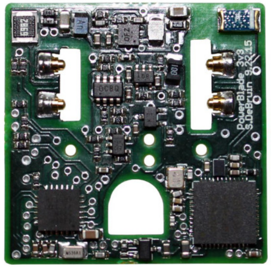
- Bringing EdgeML all the way to the edge

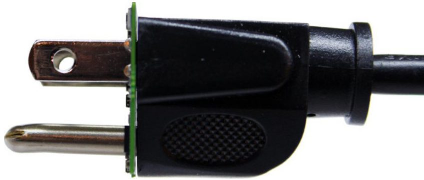


- What does battery-powered ML look like?

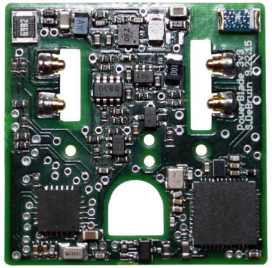


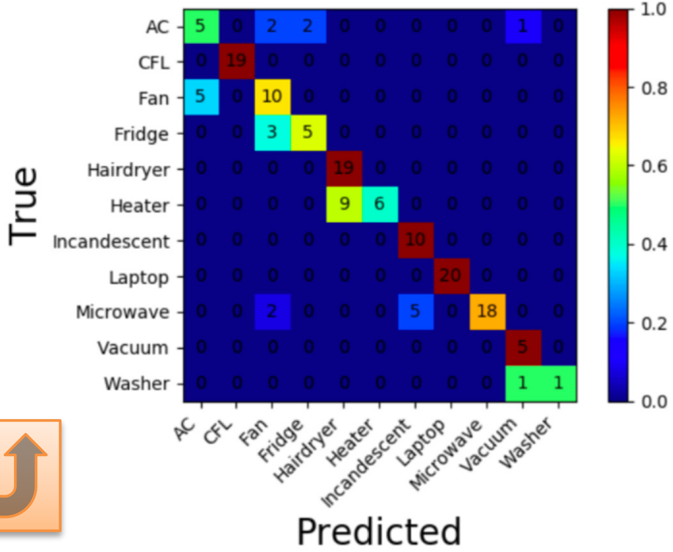
- What does battery-free ML look like?

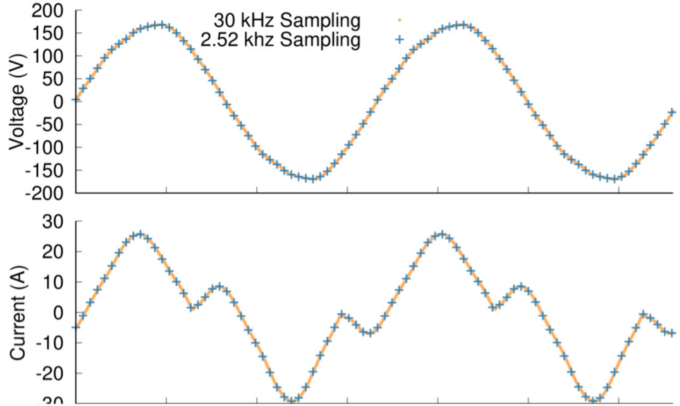


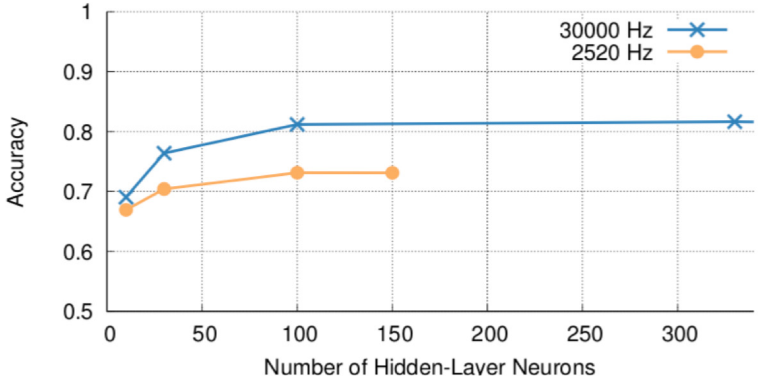


Backup: Pushing ML into sensors









Initial Size (Neurons / Bytes)	Initial Accuracy	Compressed Size (Neurons / Bytes / % Size)	Compressed Accuracy
10 / 2.15 kB	74.8%	8 / 1.73 kB / 80%	73.7%
30 / 6.37 kB	77.1%	16 / 3.42 kB / 54%	78.2%
100 / 21.1 kB	77.6%	35 / 7.43 kB / 35%	80.5%
150 / 31.7 kB	77.3%	59 / 12.5 kB / 39%	80.5%

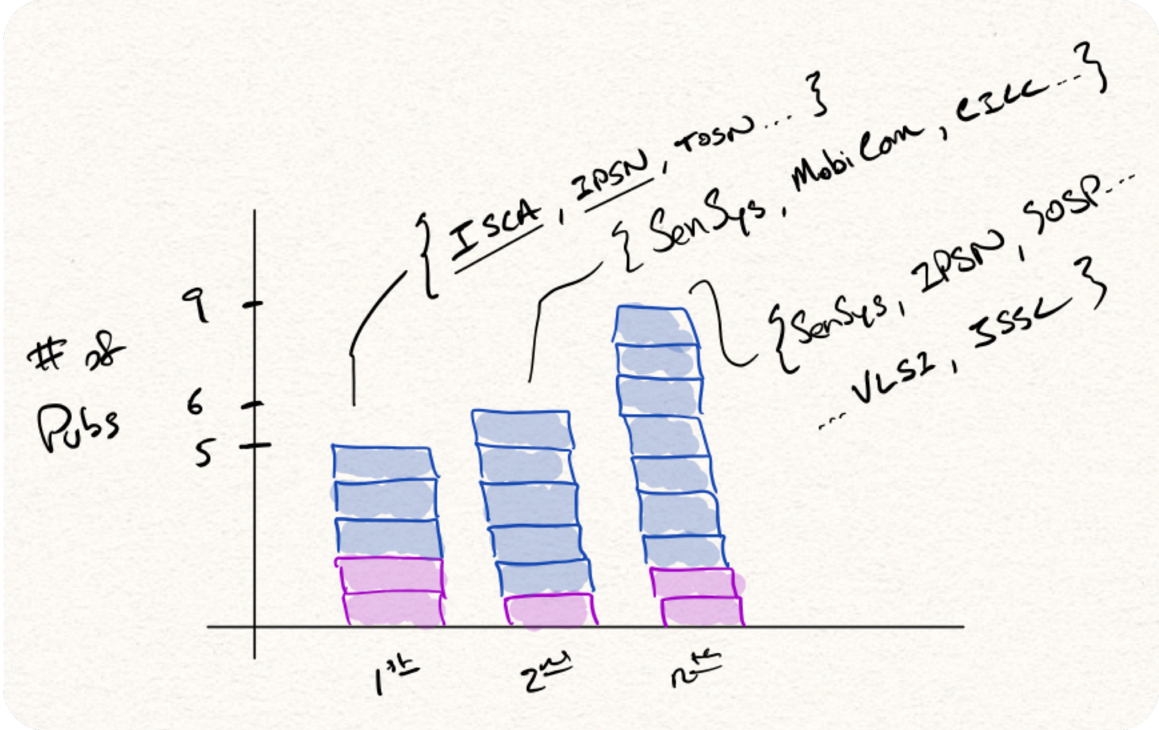
Backup: What about embedded OSes?

- Extant embedded OSes are really more like libraries

System	Concurrency	Memory Efficiency	Dependability	Fault Isolation	Loadable Applications
Arduino [6]		✓			
RIOT OS [5]		✓			
Contiki [14]	✓	✓			✓
FreeRTOS [8]	✓		✓		
TinyOS [33]	✓	✓	✓		
TOSThreads [28]	✓		✓		✓
SOS [23]	✓	✓			✓
Tock	✓	✓	✓	✓	✓

Backup: What's your home community?

- SenSys + IPSN (3+3)
  - MobiCom (2)
- Highly collaborative
  - 40 co-authors
  - 9 institutions
  - 4 continents



Backup: Computing for Computer Scientists

- Pedagogical goals for CS versus CS evaluation metrics
  - Computer science principles vs software engineering?
- High variance in student background
  - Achievement gap, belonging gap
- Created course for first-year EECS students
  - 1,500+ students and counting (I've taught around 600)
  - Now permanent course at Michigan (EECS 201)
    - This course is invaluable for students who want to be in EECS but don't have the background that independent programmers come in with. Classes like this are what enable students who feel behind to enter classes at the same level as others with more experience. I really appreciate everyone who set up this course – Student 2
    - Simply put, it's hard to be a newcomer to CS. I think the lecturers did a really good job of erasing some of the barriers of learning new skills – Student 5

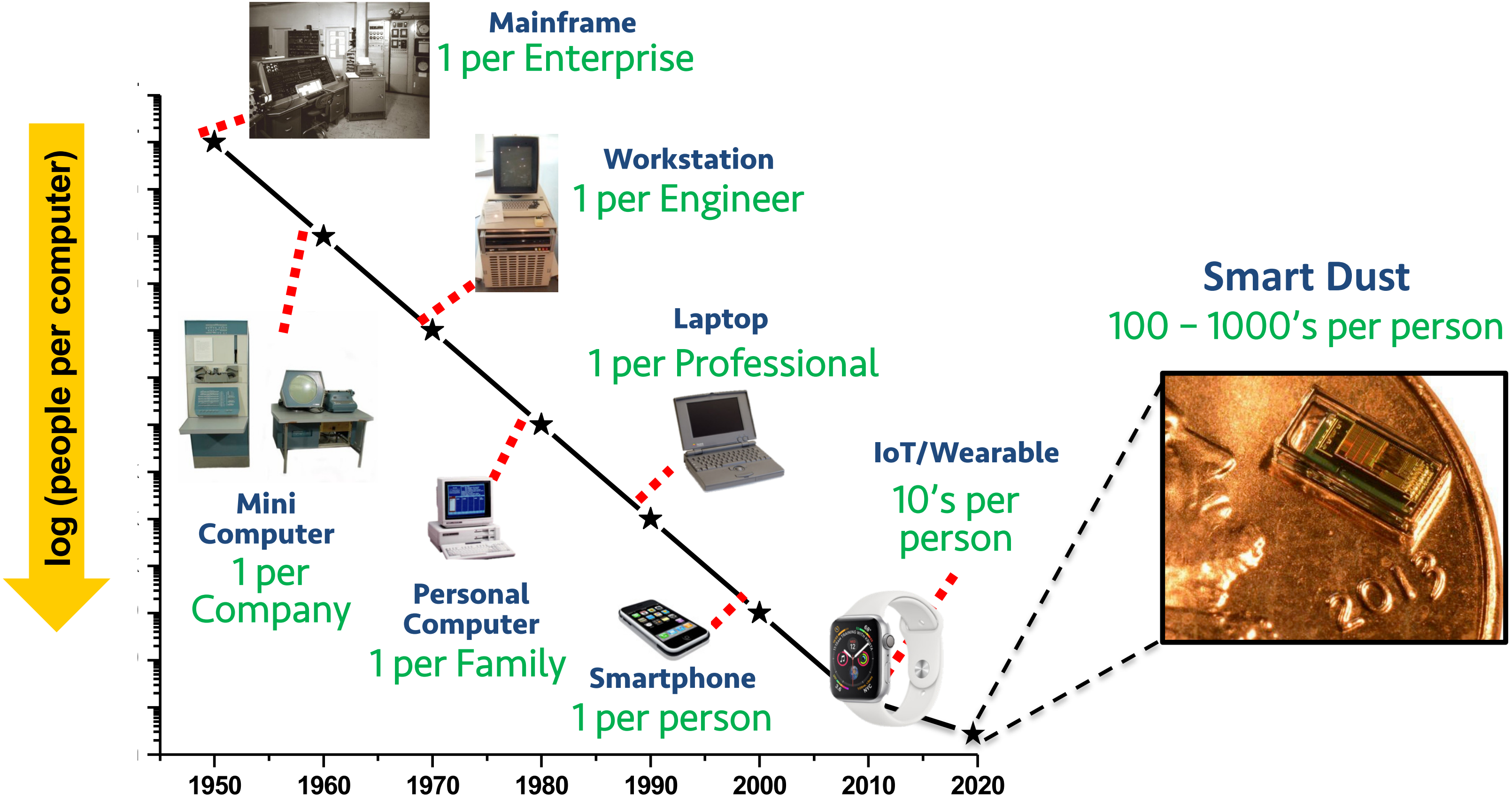
Backup: Teaching

- Undergrad
  - Core embedded [CSE145, 237A-D]
  - Broad background
    - Digital Logic [140/ECE25], Computer Architecture [30/141], Operating Systems [120/121], Networking [123/124], Uibcomp [118/218]
      - less so: Compilers [131], Synthesis/Optimization [143], Signals & Systems [ECE45]
  - Computing for Computer Scientists (C4CS)
- Graduate
  - "Resource Constrained Computing"
  - Systems (Embedded, Operating), Networking, Wireless

Backup:
Pedagogical goals of an undergraduate embedded course

- Understand the mechanics of how software interfaces with hardware
  - MMIO, peripheral buses, etc
- Understand how hardware interacts with the physical world
  - ADCs/DACs, quantization, uncertainty
- Understand how communication works
  - On-device (UART/I2C/SPI)
  - Off-device (Wireless – *maybe* some of the why/how of 15.4/LoRa w.r.t. energy)
- Understand how system design affects energy

Corollary to Bell's law: The number of computers per person is growing

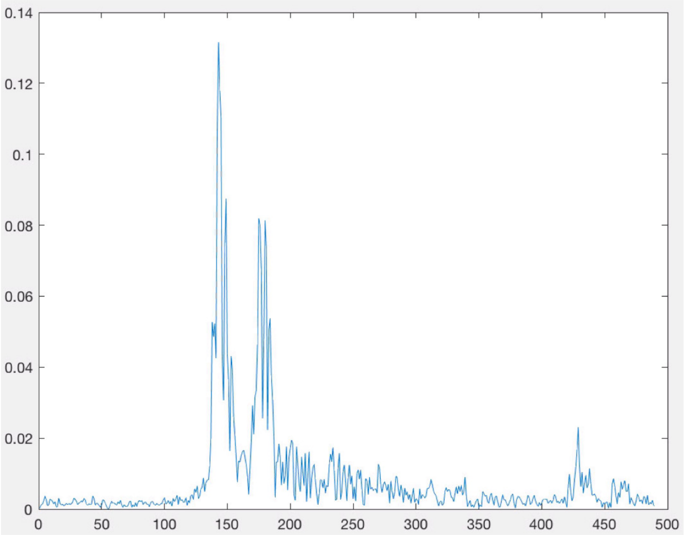


Backup: Extracting the tag signal in the real world has a few additional challenges

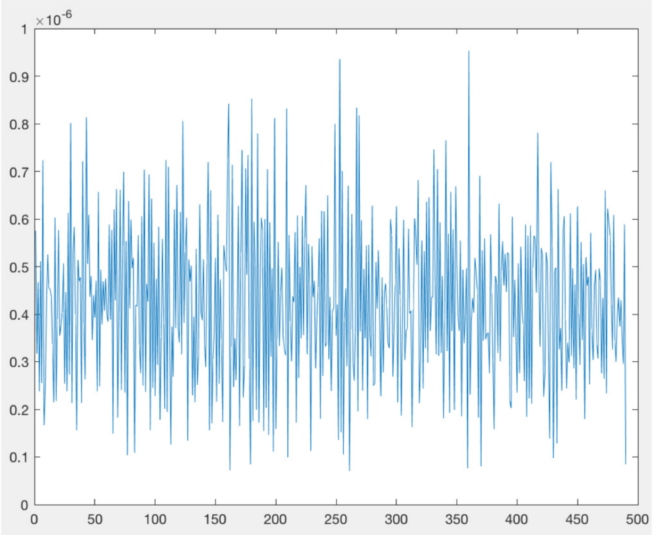
- The environment is not actually static
  - But noise is largely white & Gaussian
  - And we can filter out the rest
- When to add and when to subtract?
  - Problem: Need to know when the tag is reflecting or absorbing
  - Solution: Guess and brute force search
    - Tag stability limits Slocalization range

Backup: Recovery of modulated signal improves with greater integration time

- R_1 = 6 m, R_2 = 24 m



Raw Channel Impulse Response

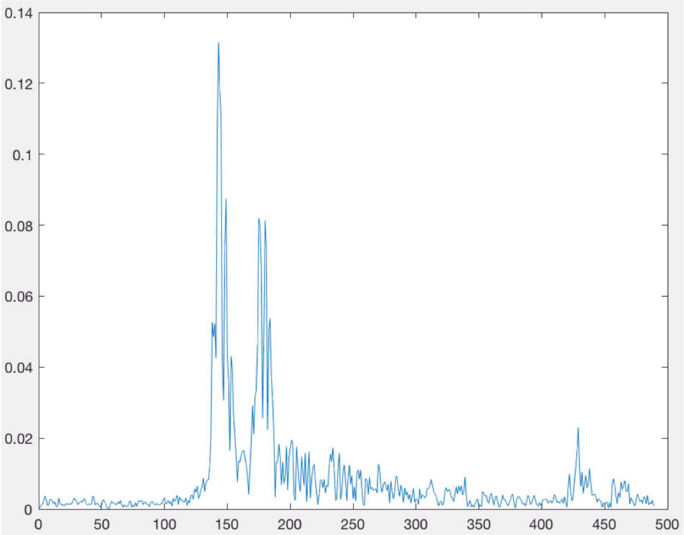


Correlated Channel Impulse Response

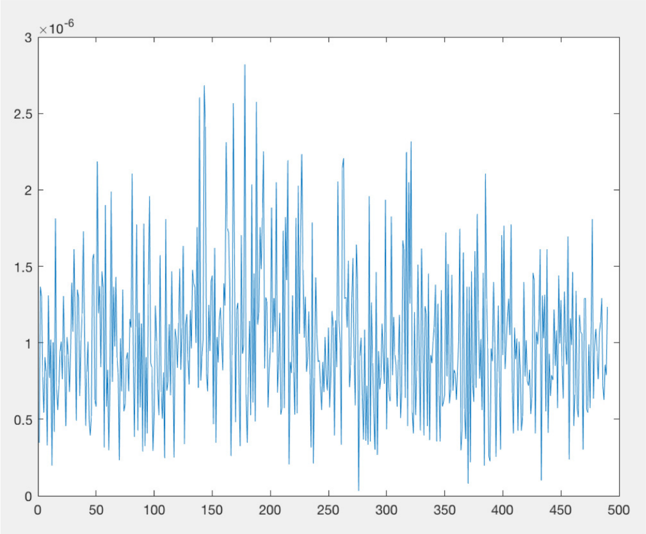
6 seconds

Backup: Recovery of modulated signal improves with greater integration time

- $R_1 = 6$ m, $R_2 = 24$ m



Raw Channel Impulse Response

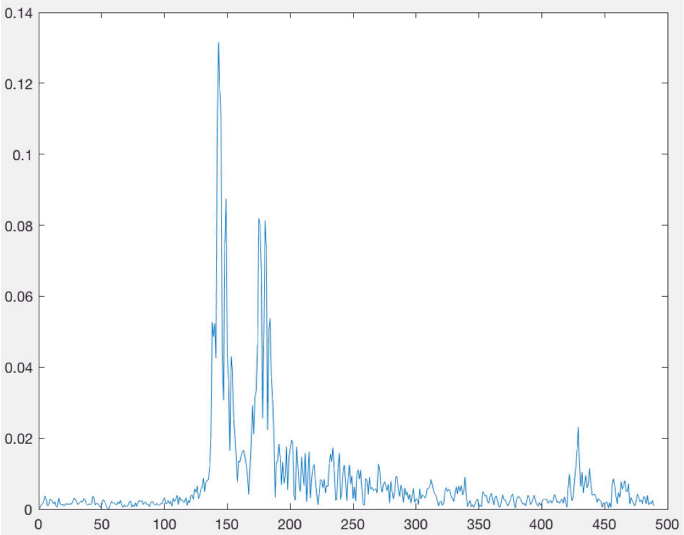


Correlated Channel Impulse Response

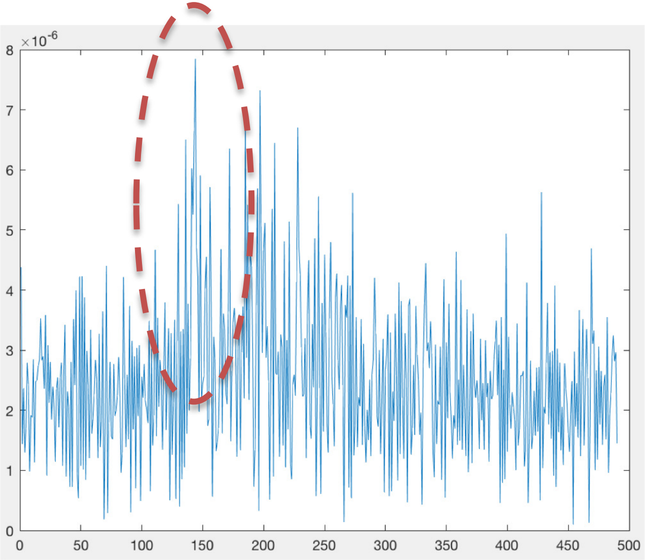
50 seconds

Backup: Recovery of modulated signal improves with greater integration time

- $R_1 = 6$ m, $R_2 = 24$ m



Raw Channel Impulse Response

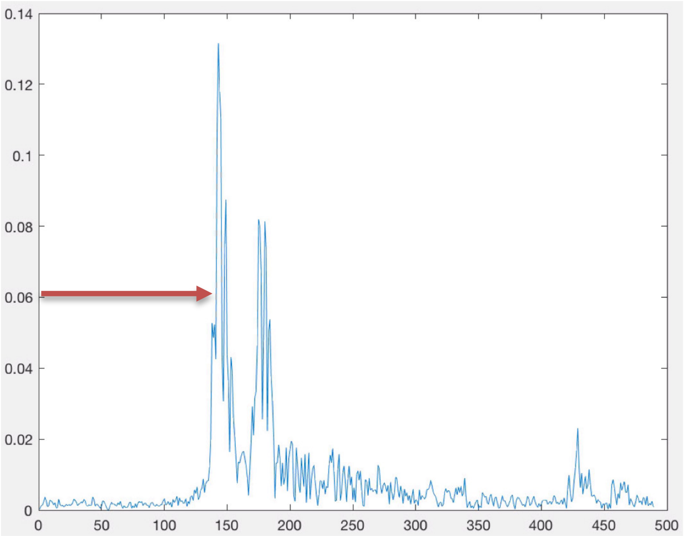


Correlated Channel Impulse Response

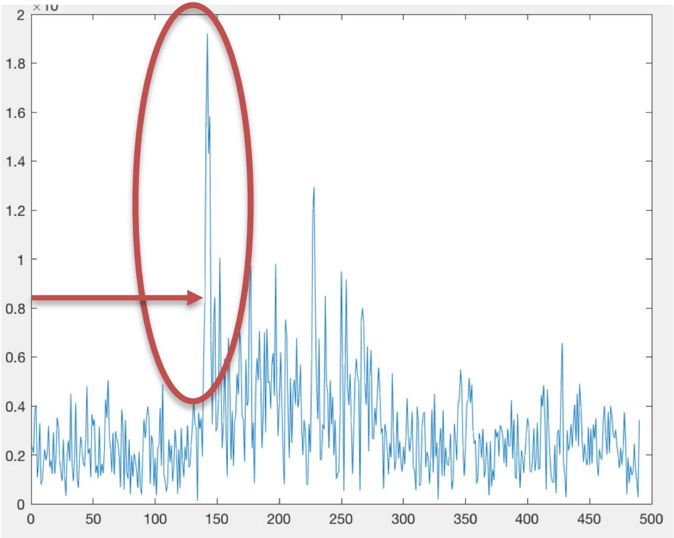
250 seconds

Backup: Recovery of modulated signal improves with greater integration time

- $R_1 = 6$ m, $R_2 = 24$ m



Raw Channel Impulse Response



Correlated Channel Impulse Response

500 seconds
(8.3 minutes)

Backup: Numerous diversity sources allows Slocalization to scale to very many tags

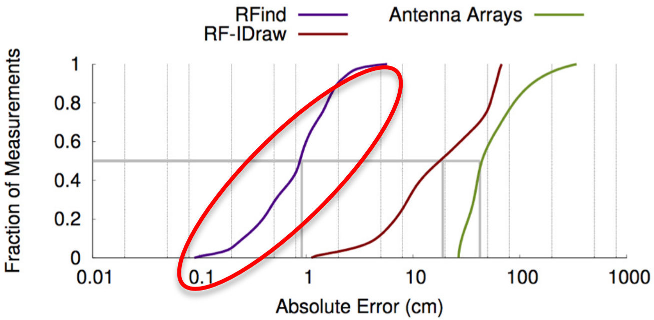
- Frequency division scales linearly in frequencies
  - Caveat: 256.00 Hz low-power RTCs exist, less so 256.20 Hz, etc
  - Caveat: Power draw scales linearly with switching frequency
- PN codes scale linearly in tag length
- Temporal code rotation scales *factorially*, but is very slow
  - Idea: Exploit tag stationarity further, rotating PN codes over time

$$f_{\text{STEPS}} * \text{PN}_{\text{bits}} * \text{Codes} = \underbrace{1{,}280}_{\text{Feasibility? Power?}} * \underbrace{63}_{\text{Time}} * \underbrace{4!}_{\textit{Time}} = 1{,}935{,}360 \text{ concurrent tags}$$

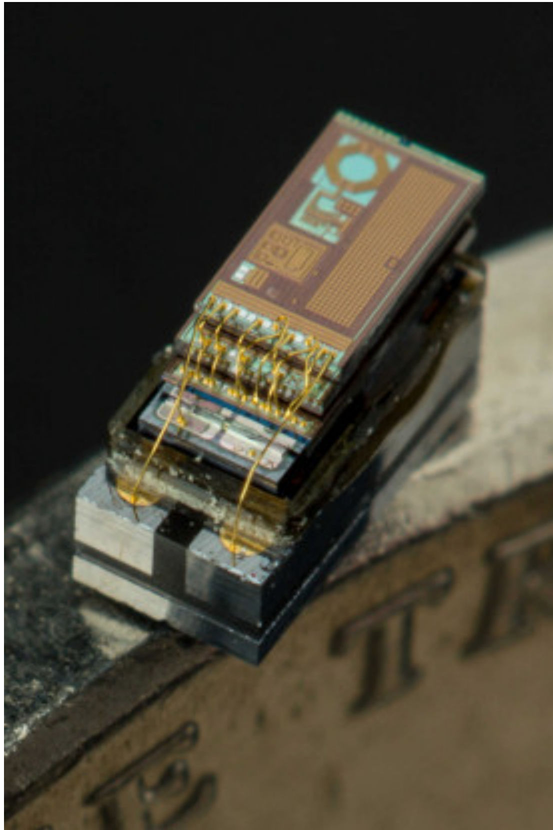
in roughly a few hours

Backup: Can we localize every physical thing? Even dust?

- Super-resolution technique from MobiCom'17
  - Use frequency information to refine localization



- Key idea:
  - If traditional localization can get close (7cm), refine estimate based on the estimate from each frequency
- What about mm-scale antennas?
  - Today, in frequencies of interest, fall from 0 to -15 dBi



Minding the Billions..., Ma, Selby, Adib. MobiCom'17

In the next few years...

Energy is a deployment concern and a first-order resource

Q: What should OS policy be for energy as a resource?

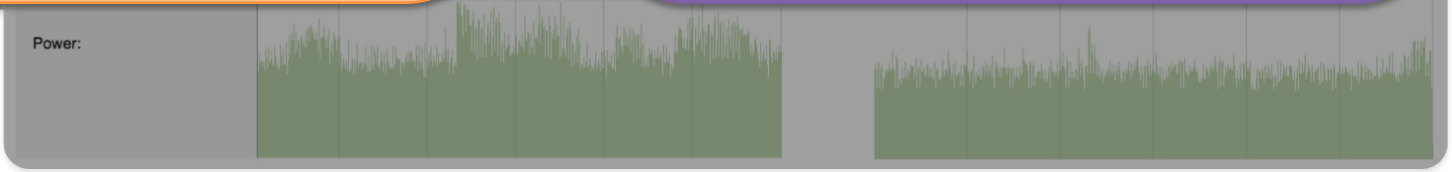
Q: What is the abstraction for cross-platform energy performance?

Q: What does graceful degradation look like, how do we support it?

Q: What are the other salient resources: time, bandwidth, ?

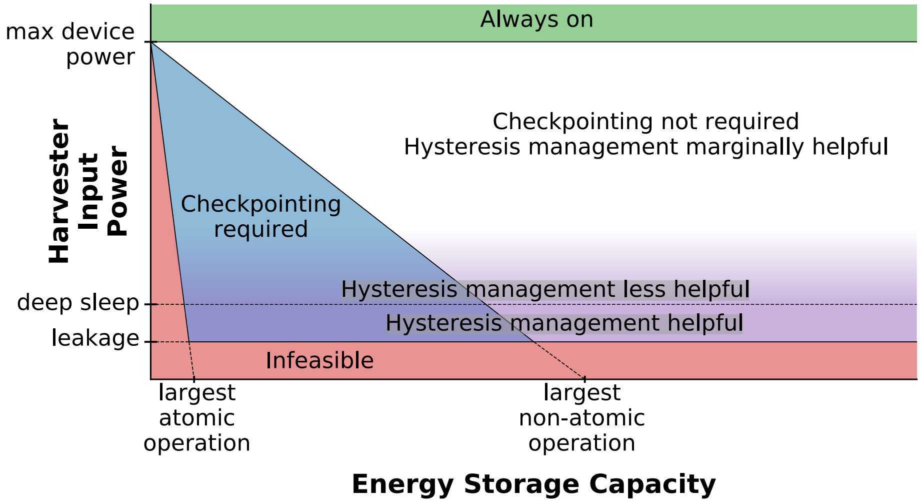
Q: What is the role of hardware support (e.g. PRET machines)?

Q: How do we capture all of these constraints for app developers?



In the next few years…
From EdgeML to Peripheral ML

- Embedded == weird computers
  - Megasample / second sensors
  - 10-100's MHz processors
  - 10-100's kB of RAM
  - 1-100's kbit/s communications
- With weird execution environments



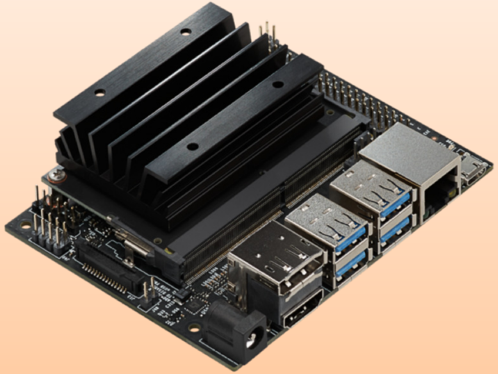
Neal Jackson, Joshua Adkins, Prabal Dutta
To Appear (IPSN'19)

In the next few years…

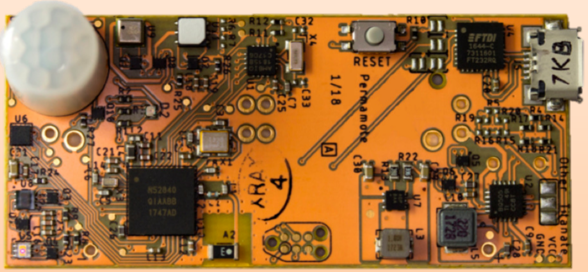
From EdgeML to Peripheral ML

- Embedded are weird computers
  - Megas
  - 10-10
  - 10-10
  - 1-100
- With weird execution environments

required
narginally helpful
elpful
lpful
Energy Storage Capacity



EdgeML is still plugged in

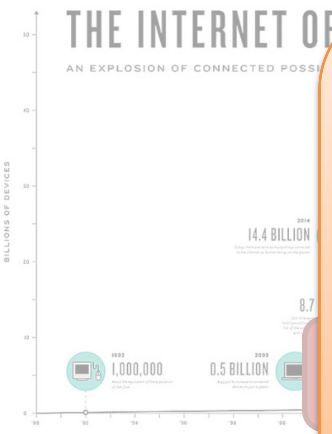


Q: What does ML look like on peripheral compute?

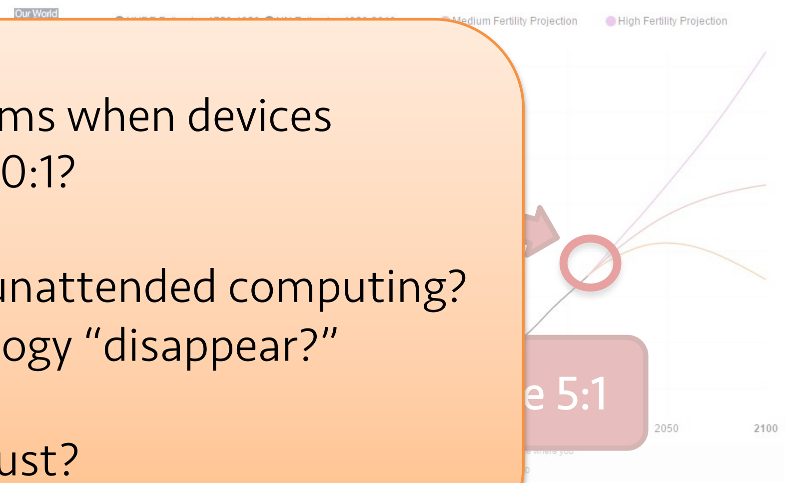
Neal Jackson, Joshua Adkins, Prabal Dutta
To Appear (IPSN'19)

In the longer term…

How does the world change with billions of devices?



Source: https://www.ncta.com/sites/default/files/plat



Q: How will we manage systems when devices outnumber people 100 or 1000:1?

Q: How do we push towards unattended computing? How do we make this technology “disappear?”

Q: How do we interact with dust?

Teaching at ETH

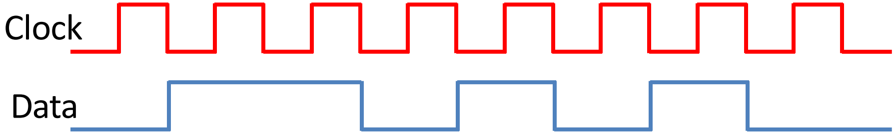
- Undergrad Qualifications
  - Core embedded
  - Broad background
    - Circuits, Signals & Systems, Digital Logic, Computer Architecture, Operating Systems, Networking; less so Compilers, Synthesis/Optimization
    - D-ITET Courses
      - Networks and Circuits, Digital Circuits, Signal and System Theory
      - Communication Systems, Communication Networks, Embedded Systems
- Graduate Seminars
  - Embedded, Systems, Networking, Wireless

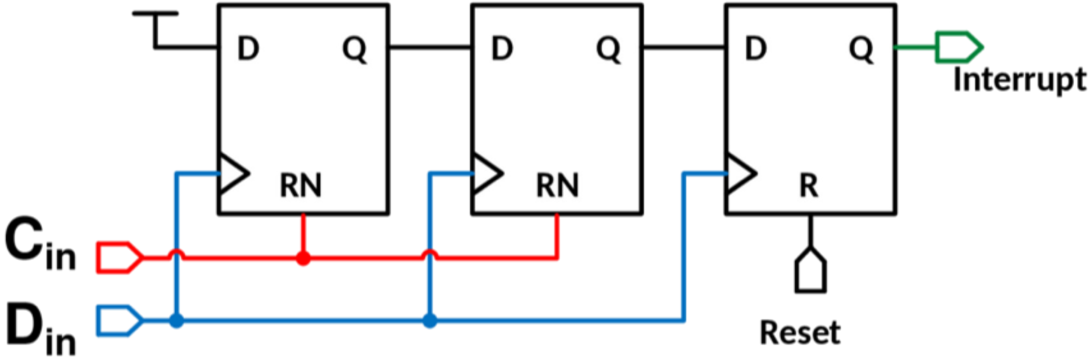
Hard Problem 1:
What's the best way to terminate messages?

- An MBus message is 0…N bytes of data
- ~~Embed length in message~~
  - Imposes large overhead for short messages
  - Forces fragmentation of long messages
- ~~"End-of-message" sentinel byte(s)~~
  - Imposes large overhead for short messages
  - Requires escaping if sentinel is in transmitted
  - Data-dependent behavior, hard to reason about
    - Worst case 2x overhead!

MBus "interjections" provide an in-band end-of-message with minimal overhead

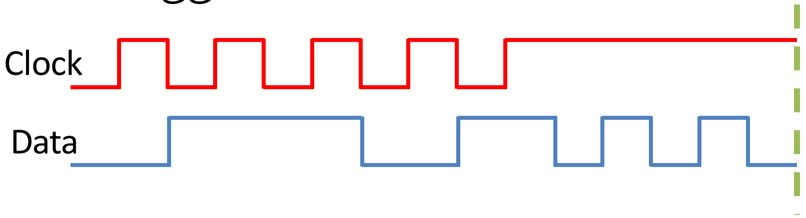
- During normal operation, Data toggles slower than Clock

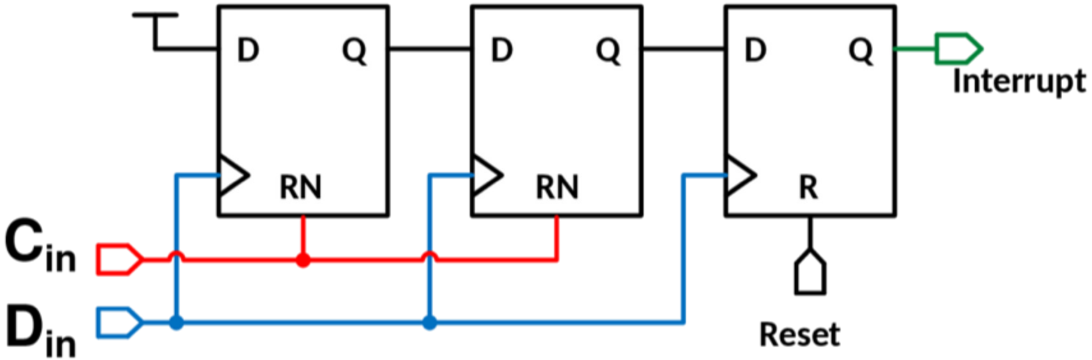




MBus “interjections” provide an in-band end-of-message with minimal overhead

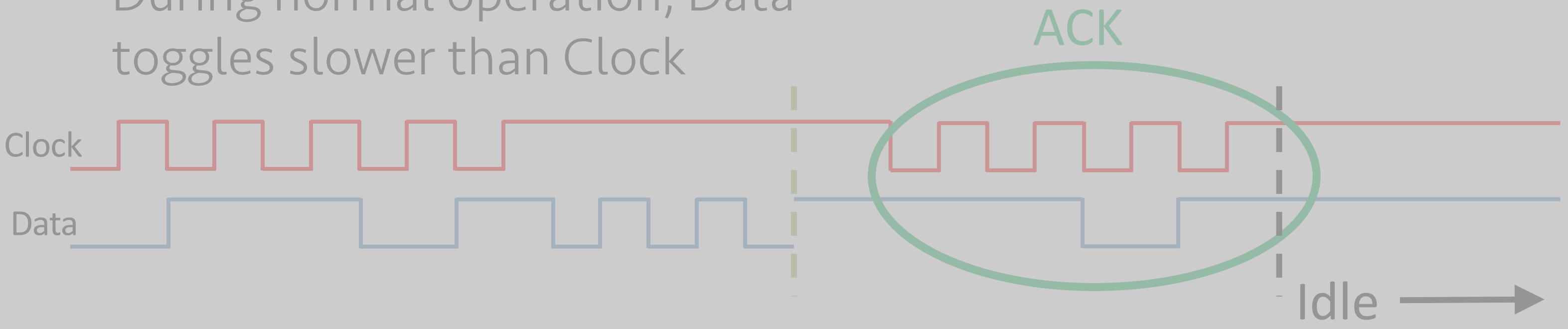
- During normal operation, Data toggles slower than Clock

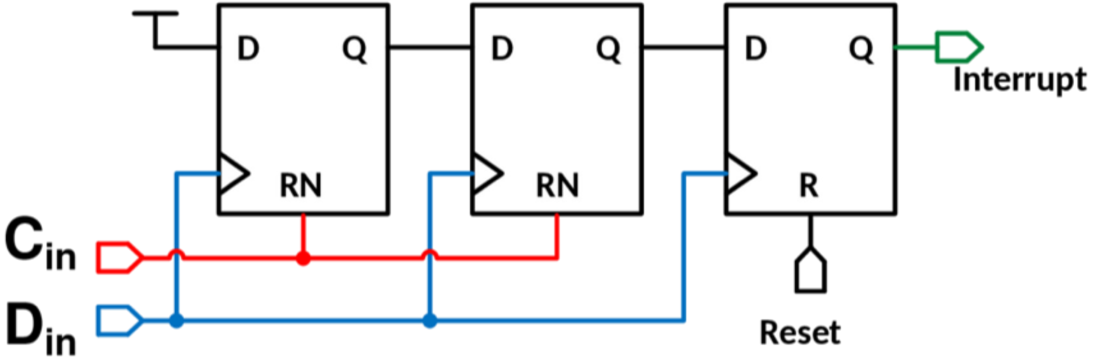




MBus “interjections” provide an in-band end-of-message with minimal overhead

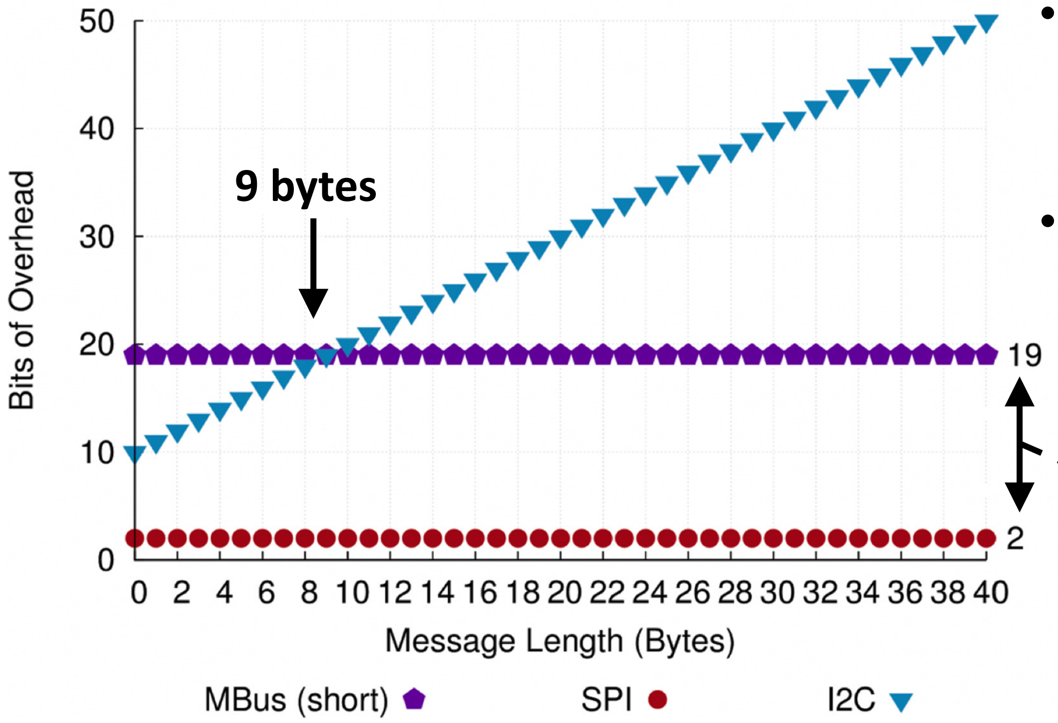
- During normal operation, Data toggles slower than Clock





Independent circuit provides reliability

Transaction-level ACKs minimize common-case overhead while interjections preserve flow control



- I²C acknowledges every byte
  - How often do NAKs happen?
    - To a random byte?
  - 12.5% overhead
- MBus ACKs transactions
  - Receiver can interject too

Arbitration
+
Address
+
Interjection

Location context is fundamental to a bevy of ubiquitous computing applications

– *Little is more basic to human perception than physical juxtaposition, and so ubiquitous computers must know where they are*
  - *Mark Weiser, The Computer for the 21st Century*

- But we do not have uniform means of expressing location
- Nor do most computational elements posses it
- Suggestion: This has inhibited effective creation of computational systems for "smart spaces"
  – How to make a space smart with no sense of space?

Ultra Wideband affords extremely high-fidelity localization

Ultra Wideband Radios (or impulse generators) are energy-hungry

System	Technology	Precision	Accuracy	Update Rate	Multiple Tags?	Top Tag Speed	Tag Power	Tag Volume	Max Tag/Anchor Dist
WASP [35]	NB (5.8 GHz) ToA	16.3 cm	50 cm (82%ile)	10 Hz	Yes	Several m/s	2-2.5 W	*Not Published*	*Not Published*
UbiSense [38]	UWB TDoA+AoA	99% w/in 30 cm	15 cm	33.75 Hz	Yes	*Not Published*	*Not Published*	24.5 cm^3	160 m
TimeDomain [2]	UWB TW-ToF	2.3 cm	2.1 cm	150 Hz	Yes	*Not Published*	4.2 W	97 cm^3	"hundreds of m"
Lazik et. al [27]	Ultrasonic TDoA	*Not Published*	3 cm (med) 12 cm (90%)	0.9 Hz	Yes	*Not Published*	1.1 W¶¶	88 cm^3	100 m
Harmonia [19]	UWB TDoA	*Not Published*	39 cm (med) 82 cm (90%)	56 Hz	No	*Not Published*	120 mW**	*Not Published*	*Not Published*
Tagoram [42]	NB (UHF) SAR	*Not Published*	12.3 cm (med)	At most 30 Hz	No	0.5 m/s	N/A	8 cm^3	10 m
WiTrack [3]	UWB ToF	*Not Published*	12 cm (med) 31 cm (90%)	At most 400 Hz	No	*Not Published*	N/A	32,700 cm^3 (avg torso [6])	*(Not Published)* > 11 m
RF-IDraw [40]	NB (UHF) Interferometry	3.6 cm (med) 3.7 cm (90%)	19 cm (med) 38 cm (90%)	At most 53 Hz	No	0.5 m/s*	N/A	8 cm^3	9 m
PolyPoint [20]	UWB ToF	31 cm	39 cm (med) 140 cm (90%)	16 Hz	No	*Not Published*	150 mW	9 cm^3	50 m
Harmonium [21]	UWB TDoA	9 cm (med) 16 cm (90%)	14 cm (med) 31 cm (90%)	19 Hz	No	2.4 m/s††	75 mW	1.5 cm^3	78 m
Chronos [39]	Bandstiched UWB ToF	*Not Published*	65 cm (med) 170 cm (90%)	12 Hz	No	*Not Published*	1.6 W§	2.7 cm³¶	*Not Published*
SurePoint	**UWB ToF**	**12 cm (med) 28 cm (90%)**	**29 cm (med) 50 cm (90%)**	**1-12 Hz**	**Yes**	**at least 2.4 m/s**	**280 mW**	**3 cm^3**	**50 m**

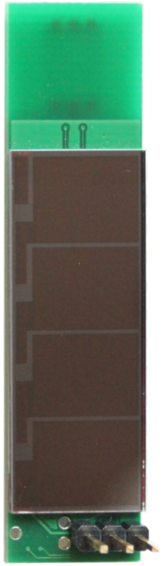
Lowest power UWB is 75 mW

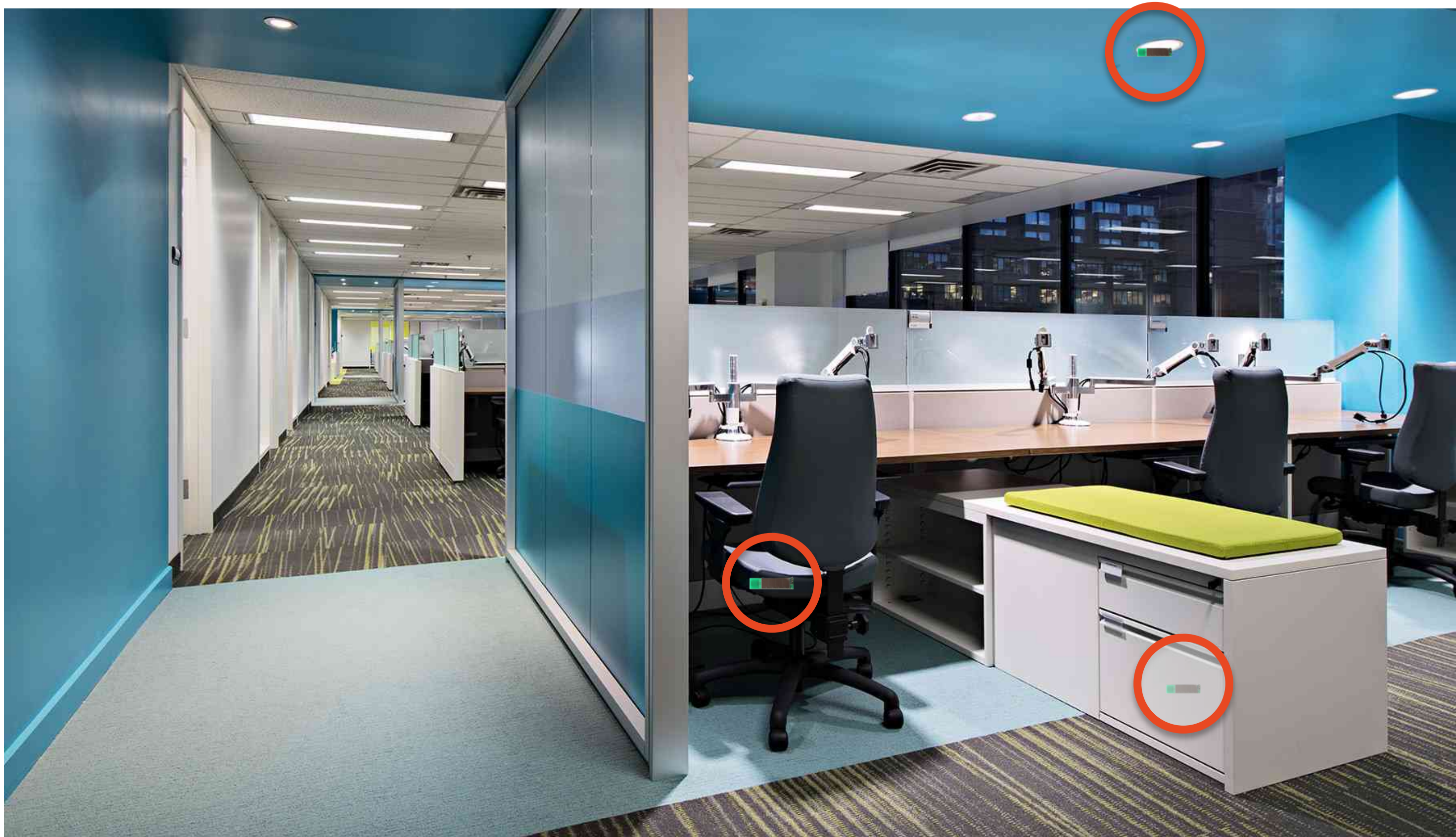
Table from: SurePoint: Exploiting Ultra Wideband Flooding and Diversity to Provide Robust, Scalable, High-Fidelity Indoor Localization. Benjamin Kempke, Pat Pannuto, Bradford Campbell, and Prabal Dutta. SenSys'16.

Backscatter renaissance is redefining low-power for wireless

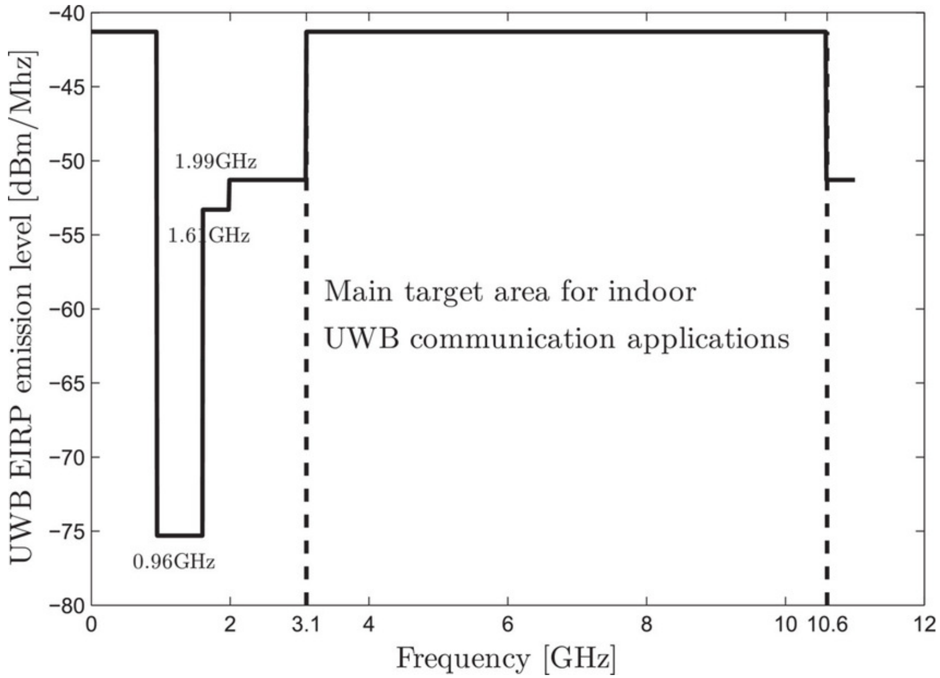
- Wireless communication from W—mW to µW—nW
  - Zhang, Pengyu, Jeremy Gummeson, and Deepak Ganesan. "Blink: A high throughput link layer for backscatter communication." MobiSys'12
  - Kellogg, Bryce, Aaron Parks, Shyamnath Gollakota, Joshua R. Smith, and David Wetherall. "Wi-Fi backscatter: Internet connectivity for RF-powered devices." SIGCOMM'14
  - Zhang, Pengyu, Pan Hu, Vijay Pasikanti, and Deepak Ganesan. "Ekhonet: High speed ultra low-power backscatter for next generation sensors." MobiCom'14
  - Zhang, Pengyu, Dinesh Bharadia, Kiran Joshi, and Sachin Katti. "Hitchhike: Practical backscatter using commodity wifi." SenSys'16
  - Ma, Yunfei, Nicholas Selby, and Fadel Adib. "Minding the billions: Ultra-wideband localization for deployed RFID tags." MobiCom'17
  - Varshney, Ambuj, Oliver Harms, Carlos-Perez Penichet, Christian Rohner, Frederik Hermans, and Thiemo Voigt. "LoRea: A Backscatter architecture that achieves a long communication range." SenSys'17
  - Carlos Pérez Penichet, Claro Noda, Ambuj Varshney, Thiemo Voigt. "Battery-Free 802.15.4 Receiver" IPSN'18

Slocalization: Ultra wideband backscatter whole-room localization for <1 µW





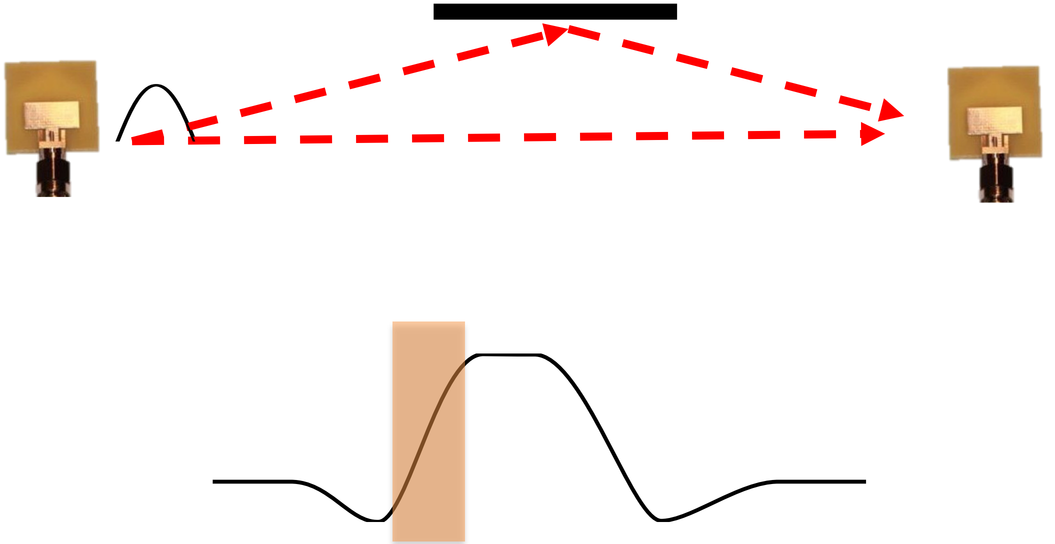
UWB can transmit 54 million times less power than traditional narrowband devices



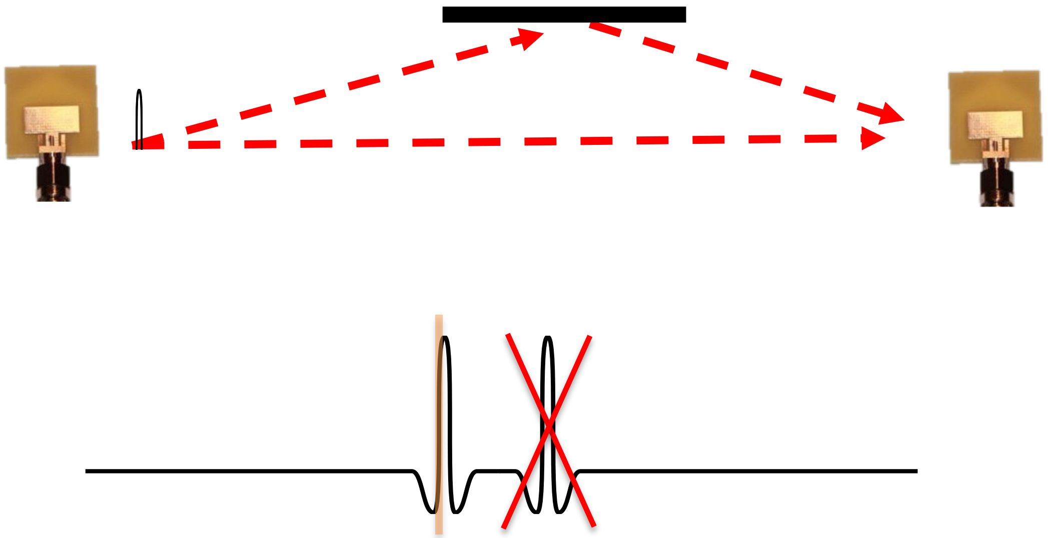
- 3-10 GHz UWB → -41.3 dBm
- 900 MHz ISM → 36 dBm
- 900 MHz unlicensed
  - Control → -13.3 dBm
  - Periodic → -21.2 dBm

Vijaya Yajnanarayana. "Spectral efficient IR-UWB communication design for low complexity transceivers" (2014).

Indoors, reflections make time-of-flight estimation difficult and inaccurate



UWB can better disambiguate multipath and identify signal arrival time



Let's think bigger about localization: Can we locate every physical thing?

Can we make location a piece of first-class context, available to every device?