

Lessons from Five Years of Making Michigan Micro Motes

Pat Pannuto[†], Yoonmyung Lee[‡], ZhiYoong Foo[†], Gyouho Kim[†], David Blaauw[†], Prabal Dutta[†]

Electrical Engineering and Computer Science Department

[†]University of Michigan, Ann Arbor, MI 48109

{ppannuto, zhiyoong, gyohokim, blaauw, prabal}@umich.edu

Department of Semiconductor Systems Engineering

[‡]SungKyunKwan University, Seoul, South Korea

yoonmyung@skku.edu

ABSTRACT

It has now been over fifteen years since Kris Pister's call for "smart dust". Today, we are capable of building general purpose computing systems, including computation, storage, sensing, and communication, that fit in a cubic millimeter. In this work, we discuss the lessons learned in the design, manufacture, debugging, and preliminary deployment of millimeter-scale systems.

1. INTRODUCTION

The Michigan Micro Mote (M3) was recently inducted into the Computer History Museum as the "World's Smallest Computer" [6]. This recent success is the culmination of five years of development on the M3 platform and nearly fifteen years of work in low power design. Figure 1 showcases the evolution from one of our first 3D-stacked micro-scale systems, an Intraocular Pressure monitor [1], to our most recent pressure sensing system.

After the first few designs, building modular and reusable components became a first-order design constraint. By baking in modularity and composability as a fundamental design consideration, we have been able to manufacture over a dozen unique systems on the M3 platform, pulling along improvements in each component. Supporting this modularity requires careful consideration of layer size and pad layout. The power states of communicating modules must be coordinated, and the demand on the PMU must be planned for and accommodated. Our recent work at ISCA'15 introduces MBus, a new system bus to help address composition and power management challenges for modular, millimeter-scale systems [5]. This work, however, takes a broader view, and discusses the challenges in designing M3 chips, physically manufacturing M3 systems, debugging and bootstrapping systems too small to connect wires to, and some preliminary results from deploying M3 systems in non-lab settings, with a focus on *in vivo* applications.

Permission to make digital or hard copies of part or all of this work is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

Copyright is held by the authors.

WARP '15, June 13-17, 2015, Portland, OR, USA.

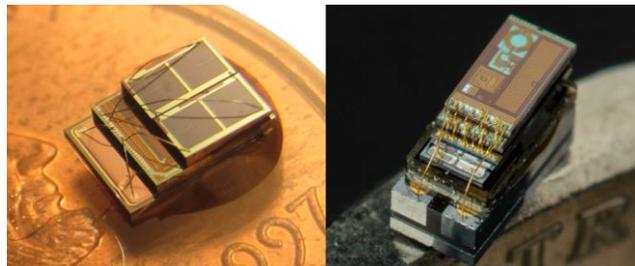


Figure 1. Stack Evolution. The left image is one of our first forays into 3D-stacked, micro-scale systems on a U.S. penny [1]. The right image is our most recent pressure sensing system, on the edge of a U.S. nickel.

2. DESIGN CHALLENGES AND METHOD

At millimeter-scale, energy is critical. To minimize static leakage, M3 chips often contain several power domains. Currently, ensuring proper isolation, level conversion, and isolation of level conversion between power domains is a very manual task as tool support has not yet caught up with our aggressive power-gating. As a consequence, we often do not simulate all power states prior to fabrication. Each power domain has its own clock network, to prevent the clock tree from unexpectedly crossing power domains. Isolation and level conversion present a particularly interesting challenge during cold-boot, when even the nominally always-on signals are rising, and requires a special power-on reset circuit for isolation networks.

New M3 components usually follow a "3 spin" model. The first design featuring the new circuit or module is a debug chip designed to validate the new component, with numerous test points and override signals. This first spin is large and often debug pad area dominates. The second design cuts the majority of debug signals, targeting the final form factor, and integrates the M3 frontend. Usually, the majority of the chip functions correctly and the third design is needed only to fix a few minor issues and scale production of the new layer.

While we had established mechanisms for per-chip lead times, we needed to develop new methods for integrated systems. There is a cost, time, and risk tradeoff between the number of layers in the stack that we replace at any time. The yield and operation of the stacks are a function of each of the individual layer yields and operating behaviors as well as cross-layer interactions. Tracking, isolating, and debugging system-level issues in a mix of new and previously reliable chips is a continuously evolving process.

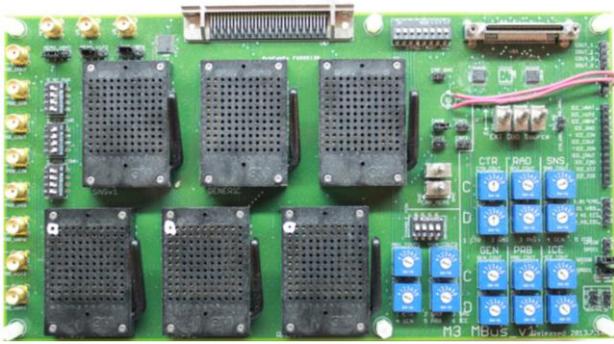


Figure 2. "Flat Stack" Debug Board. The switches at the bottom-right enable quick switching of stackup.

3. MANUFACTURING TECHNIQUE

M3 systems currently integrate chips from 65, 130, and 180 nm processes. To build a more compact stack, we first thin the wafers from 300 microns to 150 microns. We then dice on die attach film and stack the layers. As seen in Figure 1 (right), each layer is slightly smaller (or in some cases slightly offset) so that the pads form a series of steps. This stair-step design creates a loosely 45° edge and largely enables wirebonding between any layers.

While our fabrication, stacking, and bonding processes are fairly mature, packaging remains an active area of exploration. We have experimented with mounting in a glass package and more recently have transitioned to epoxy molds. Our 180 nm chips exhibit somewhat serious light sensitivity, which black epoxy protects from. Our systems also employ solar harvesting, however, requiring a "window" of clear epoxy on top of the encapsulation. Some systems include an imager and lens, while another includes a pressure sensor, neither of which can be wholly encased.

4. SYSTEM CONSIDERATIONS

4.1 Bootstrapping a millimeter-scale system

Freshly manufactured systems are unprogrammed and are too small, or too encapsulated, to physically attach wires to program. Mask ROM is unappealing as it is inflexible and area-expensive. Instead, we developed an ultra-low power optical receiver frontend, placing a photocell between the I/O pads on the processor layer [4]. In practice, the optical frontend is a write-only frontend to the system bus, enabling direct communication to any layer in the system and providing a robust mechanism to rescue corrupted systems.

4.2 Debugging millimeter-scale systems

Debugging remains an active challenge for M3. One approach is shown in Figure 2. This board allows us to work with partial stacks. Each chip is in its own package, allowing individual debugging or on-the-fly partial stack construction. The I/O drive strength of each M3 chip is relatively low, requiring very low impedance buffers (analog signal buffers in practice). Even then, we occasionally experience slew-related issues in the current debugging setup.

An open question is how to debug an assembled, or worse encapsulated, stack. While the optical frontend provides a nice mechanism to program and recover, the only output mechanism is the radio. Debugging stacks is further frustrated by the system energy budget. In normal operation, the system operates at very low duty cycles. The active power is 10's of μW while the harvester can only charge at 10's of nW. For debugging, the result is either frequent deep discharge of the battery, which significantly reduces its lifetime, or short, infrequent debugging sessions.

4.3 Deploying millimeter-scale systems

As manufacturing and yield mature, long-term (order months) longitudinal tests and out-of-lab deployments are beginning to reveal new issues. Our stacks use tiny, roughly 1 mm^2 and $0.5\text{--}5\ \mu\text{Ah}$, thin film batteries. At shallow discharges, 10% or less, the batteries last for 10,000+ cycles. At deeper discharges, 60% or more, however, battery capacity will fall off in as few as 10's of cycles.

Several of our initial deployment aims are in biomedical space, implanting M3 stacks. As visible light cannot penetrate the body, we recently developed a replacement harvester tuned to low-wavelength infrared [2]. Our earliest chips were only tested in a 25°C lab and implanting in a 40°C body revealed that designing for a wider temperature range, and the resulting variable power draw, is important to maximize the flexibility of all of our modules.

5. REFERENCES

- [1] G. K. Chen, "Power Management and SRAM for Energy-Autonomous and Low-Power Systems. Ph.D. Dissertation. University of Michigan, Ann Arbor, MI, USA. Advisor D. M. Sylvester.
- [2] W. Jung, S. Oh, S. Bang, Y. Lee, D. Sylvester, D. Blaauw, "23.3 A 3 nW fully integrated energy harvester based on self-oscillating switched-capacitor DC-DC converter," *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, Feb. 2014
- [3] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Next century challenges: Mobile networking for "smart dust"," in *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, (MobiCom '99)*. ACM, 1999.
- [4] G. Kim, Y. Lee, S. Bang, I. Lee, Y. Kim, D. Sylvester, D. Blaauw, "A 695 pW standby power optical wake-up receiver for wireless sensor nodes," *Custom Integrated Circuits Conference (CICC), 2012 IEEE*, Sept. 2012
- [5] P. Pannuto, Y. Lee, Y.-S. Kuo, Z. Foo, B. Kempke, G. Kim, R. Dreslinski Jr., D. Blaauw, and P. Dutta, "Mbus: An Ultra-Low Power Interconnect Bus for Next Generation Nanopower Systems," in *Proceedings of the 42nd International Symposium on Computer Architecture (ISCA '15)*. ACM, 2015.
- [6] D. Spicer. The World's Smallest Computer. 2015. *Online*. <http://www.computerhistory.org/atcm/the-worlds-smallest-computer/>