AFRL-RY-WP-TR-2023-0141

# DISCRETE INTEGRATED CIRUCITS ELECTRONICS (DICE)

**Neil Gershenfeld**
**MIT Center for Bits and Atoms**

**OCTOBER 2023**
**Final Report**

**DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.**
***See additional restrictions described on inside pages***

**STINFO COPY**

**AIR FORCE RESEARCH LABORATORY**
**SENSORS DIRECTORATE**
**WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320**
**AIR FORCE MATERIEL COMMAND**
**UNITED STATES AIR FORCE**

# NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with The Under Secretary of Defense memorandum dated 24 May 2010 and AFRL/DSO policy clarification email dated 13 January 2020. This report is available to the general public, including foreign nationals.

Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RY-WP-TR-2023-0139 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//Signature//

CHRISTOPHER A. BOZADA
Program Manager
Aerospace Components and Subsystems Division

//Signature//

GENE M. WILKINS, Lt Col, USAF
Deputy Chief
Aerospace Components and Subsystems Division
Sensors Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

*Disseminated copies will show "//Signature//" stamped or typed above the signature blocks.

# REPORT DOCUMENTATION PAGE

**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.**

| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED | |
|---|---|---|---|
| | | START DATE | END DATE |
| October 2023 | Final | 23 May 2019 | 30 December 2022 |

**4. TITLE AND SUBTITLE**

DISCRETE INTEGRATED CIRCUITS ELECTRONICS (DICE)

| 5a. CONTRACT NUMBER | 5b. GRANT NUMBER | 5c. PROGRAM ELEMENT NUMBER |
|---|---|---|
| FA8650-19-2-7921 | N/A | 62716E |
| **5d. PROJECT NUMBER** | **5e. TASK NUMBER** | **5f. WORK UNIT NUMBER** |
| N/A | N/A | Y1ZW |


**6. AUTHOR(S)**

Neil Gershenfeld

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

MIT Center for Bits and Atoms
20 Ames St
Cambridge, MA 02139


**8. PERFORMING ORGANIZATION REPORT NUMBER**

| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | | 10. SPONSOR/MONITOR'S ACRONYM(S) | 11. SPONSOR/MONITOR'SREPORT NUMBER(S) |
|---|---|---|---|
| Air Force Research Laboratory<br>Sensors Directorate<br>Wright-Patterson Air Force Base, OH 45433-7320<br>Air Force Materiel Command<br>United States Air Forces | Defense Advanced Research Projects Agency (DARPA/MTO)<br>675 North Randolph Street<br>Arlington, VA 22203 | AFRL/RYD | AFRL-RY-WP-TR-2023-0141 |

**12. DISTRIBUTION/AVAILABILITY STATEMENT**

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**13. SUPPLEMENTARY NOTES**


This material is based on research sponsored by the Air Force Research Laboratory (AFRL) and the Defense Advanced Research Projects Agency (DARPA) under agreement number FA8650-19-2-7921. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory (AFRL), the Defense Advanced Research Projects Agency (DARPA), or the U.S. Government. Report contains color.


**14. ABSTRACT**


DARPA's Next Generation Microelectronics Manufacturing (NGMM)program is focused on three dimensional heterogeneously integrated technologies (3DHI). This refers to the stacking of separately manufactured components from different material systems within a single package. There is a gulf between the sophistication and investment in the chiplets that go into these packages, and obstacles in the systems around them, the Discrete Integrated Circuit Electronics (DICE) project aimed to improve their performance, reduce development times, shorten supply chains, and increase integration by developing a cross between a pick-and-place and 3D printer which can assemble VMD (volume) rather than SMD (surface) mounted devices. In this reporting period the project demonstrated DICE on CPU, MCU, CSP, and SMD scales, with examples of end-to-end workflows for their design, assembly and programming. Future work aims to focus on developing the most promising applications that have emerged.


**15. SUBJECT TERMS**

Three dimensional heterogeneously integrated technologies, 3DHI, surface mounted device, SDM, volume mounted device, VDM, pick-and-place, 3D printer, CPU, MCU, CSP

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES |
|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | C. THIS PAGE | | |
| Unclassified | Unclassified | Unclassified | SAR | 590 |

| 19a. NAME OF RESPONSIBLE PERSON | 19b. PHONE NUMBER *(Include area code)* |
|---|---|
| Christopher Bozada | N/A |

# Table of Contents

**Section** **Page**

List of Figures ..... ii
1 SUMMARY ..... 1
2 RECOMMENDATIONS ..... 13
3 APPENDICES ..... 14
LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS ..... 585


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# List of Figures

**Figure** **Page**

Figure 1: Part Scales and Their Applications Investigated in the DICE Project ..... 2
Figure 2: SEM Imaging (left), Custom Testing Instrumentation (center), and Cycling Measurements (right) of the Electrical Performance of Press-fit Joints ..... 3
Figure 3: Micro-EDM Commissioned for Interconnect below the COTS Limit in the DICE Project, and a Test Part ..... 3
Figure 4: Universal Set of Routing Tiles for the 4H DICE Geometry (right), and Sample Assembly and Size scaling (left) ..... 4
Figure 5: H-bridge (left) and Ring Oscillator (right) Designs in the DICE 4H Geometry ..... 4
Figure 6: Primary, Secondary, Tertiary, and Quaternary Assembly of DICE Chip-scale Packaging ..... 5
Figure 7: DICE CPU-scale Parts (left), Robotic Assembly (center), and Distributed Display (right) ..... 5
Figure 8: Custom DICE Assembler Developed for the 4H Geometry, a Cross between a Pick-and-Place and 3D Printer ..... 6
Figure 9: Custom Assembler Developed for the Tiny-DICE Geometry ..... 6
Figure 10: Robotic Assembly of CPU-scale DICE (left), and an Assembled Test Circuit (right) . 7
Figure 11: Design and Simulation tool Developed for 4H DICE ..... 7
Figure 12: Simulation of the Ring-up of a DICE Ring Oscillator ..... 8
Figure 13: Discrete Logic Simulation of a Parametric DICE Adder Circuit ..... 8
Figure 14: CAM Tool Developed for Designing the Construction of DICE Computation ..... 9
Figure 15: Abstraction of the Programming Model for DICE Development and Execution ..... 9
Figure 16: DICE Programming Development Tool Executing a Scatter-gather Calculation as a Space-filling Curve ..... 10
Figure 17: DICE Programming Development Tool Executing a Locally-connected Multiphysics Particle Method Simulation ..... 10
Figure 18: DICE Spinoff in Structural Electronics, used in a Robotic Assembler that is Assembled from the Parts that it's Assembling ..... 11
Figure 19: DICE Spinoff in the Development a Superconducting Asynchronous Adiabatic Quantum Flux Parametron ..... 11
Figure 20: DICE in Micromachining Direct-write MEMS Devices (top and left), and Single-crystal Diamond Micromachining for 10 Million RPM Rotors for NMR Magic Angle Spinning ..... 12


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# 1 SUMMARY

The DICE (Discrete Integrated Circuit Electronics) project developed and demonstrated the direct-write assembly of integrated three-dimensional electronics.

There is a hierarchy in electronics from devices, to chiplets, chips, packages, boards, blades, and systems. Their boundaries introduce inefficiencies in performance, dependency on global supply chains, and increasing development cost and time. The DICE project sought to span them with a consistent representation, by developing a cross between a pick-and-place assembler and 3D printer, using a generalization of SMD (Surface Mount Devices) to VMD (Volume Mount Devices).

Potential benefits include reducing power consumption, improving security, simplifying software by aligning it with hardware, enabling novel form factors, and field reconfigurability and reuse. These were demonstrated in an end-to-end workflow, with results that were shared through MIT theses (provided as appendices):

- *Assembling Integrated Electronics*, Zach Fredin (2021): https://cba.mit.edu/docs/theses/21.09.fredin.pdf.
- *Superconducting Asynchronous Logic for Ultra-low Power High Performance Computing*, Camron Blackburn (2021): https://cba.mit.edu/docs/theses/21.09.blackburn.pdf
- *Distributed Displays for Discrete Integrated Circuit Electronics*, Justin Christensen (2021): https://cba.mit.edu/docs/theses/21.09.christensen.pdf
- *Laser Direct-Write Fabrication of MEMS*, Prashant Patil (2019): https://cba.mit.edu/docs/theses/19.09.Patil.pdf
- *Discrete Robotic Construction*, Will Langford (2019): https://cba.mit.edu/docs/theses/19.09.Langford.pd

shared through publications:

- *Design and Simulation of Phase Synchronizer for Adiabatic Quantum Flux Parametron Circuits*, L. Camron Blackburn, Evan Golden, Alex Wynn, Andrew Wagner, and Neil Gershenfeld, IEEE Transactions on Applied Superconductivity, pp.1-5 (2023): https://cba.mit.edu/docs/papers/23.02.Blackburn.pdf
- *Self-Replicating Hierarchical Modular Robotic Swarms*, Amira Abdel-Rahman, Christopher Cameron, Benjamin Jenett, Miana Smith, and Neil Gershenfeld, Nature Communications Engineering, 1 1-10 (2022): https://cba.mit.edu/docs/papers/22.11.Nature.pdf
- *Discrete Integrated Circuit Electronics (DICE)*, Zach Fredin, Jiri Zemanek, Camron Blackburn, Erik Strand, Amira Abdel-Rahman, Premila Rowles, and Neil Gershenfeld,


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

IEEE High Performance Extreme Computing Conference (2020): https://cba.mit.edu/docs/papers/20.09.DICE.pdf

through collaborations with existing companies:

- *Molex*: https://www.molex.com/en-us/home
- *TI*: https://www.ti.com/
- *Toyota*: https://www.toyota.com/

and through the creation of new companies:

- *Atomic Machines*: https://www.atomicmachines.com/
- *Micromize*: https://micromize.com/
- *Volta Labs*: https://www.voltalabs.com/

## Results

DICE was investigated at scales for assembling materials, surface-mount devices (SMD), wafer-level chip-scale packages (CSP), CPUs, and GPUs:

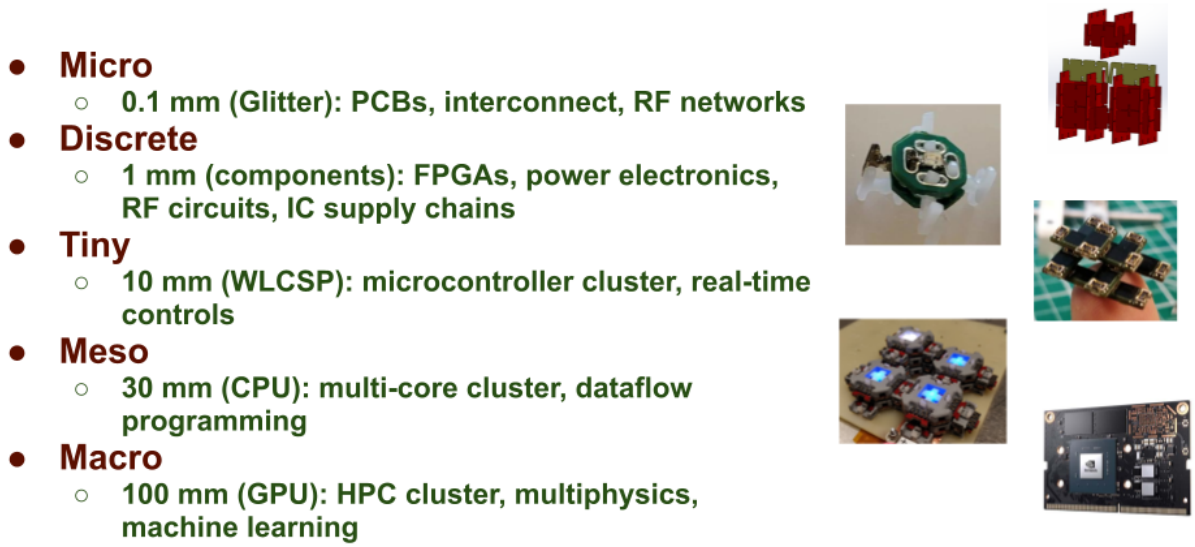



**Figure 1: Part Scales and Their Applications Investigated in the DICE Project**

The following sections describe work on parts, assemblers, design, programming, and applications, followed by recommendations for future work.

### *Parts*

The project began by studying the mechanical and electrical properties of press-fit joints, using custom instrumentation:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

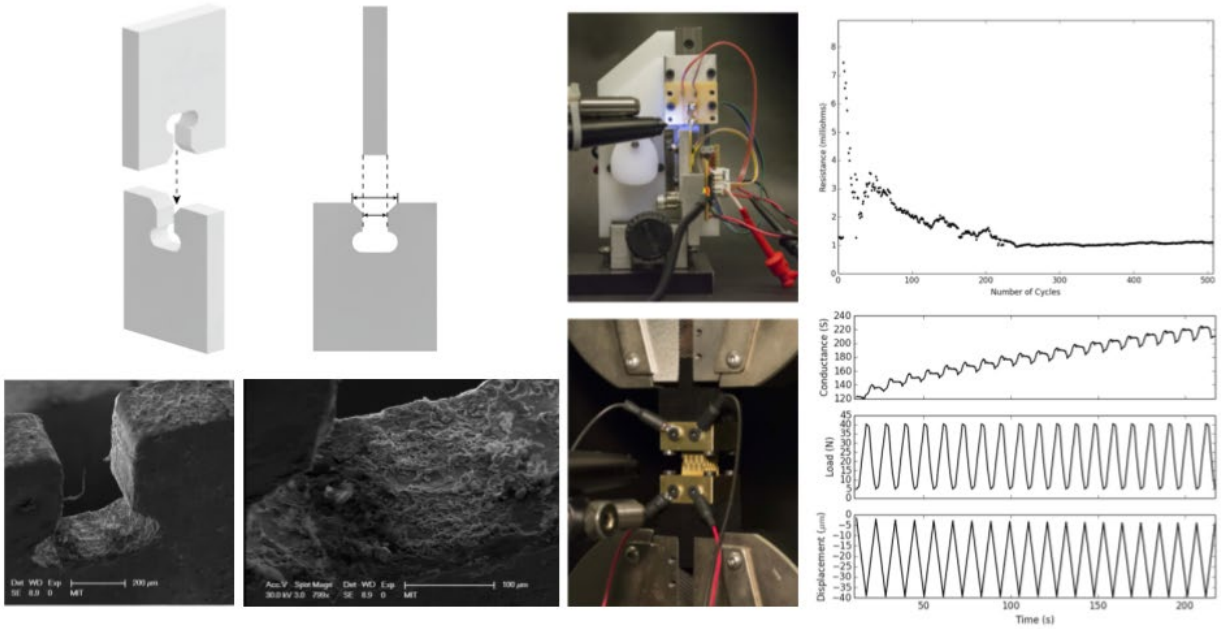



**Figure 2: SEM Imaging (left), Custom Testing Instrumentation (center), and Cycling Measurements (right) of the Electrical Performance of Press-fit Joints**

This highlighted the utility of pre-loading in optimizing joint performance, effectively eliminating their impact on target applications.

DICE development required work on interconnect below COTS limits; for that a custom micro-EDM was commissioned, capable of working down to 10$\mu$:

**Figure 3: Micro-EDM Commissioned for Interconnect below the COTS Limit in the DICE Project, and a Test Part**


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

State-of-the-art SMD packages are approaching the pitch of early generation ICs. For carrying these components a “4H” discrete-DICE geometry was developed and tested with COTS PCB fabrication:

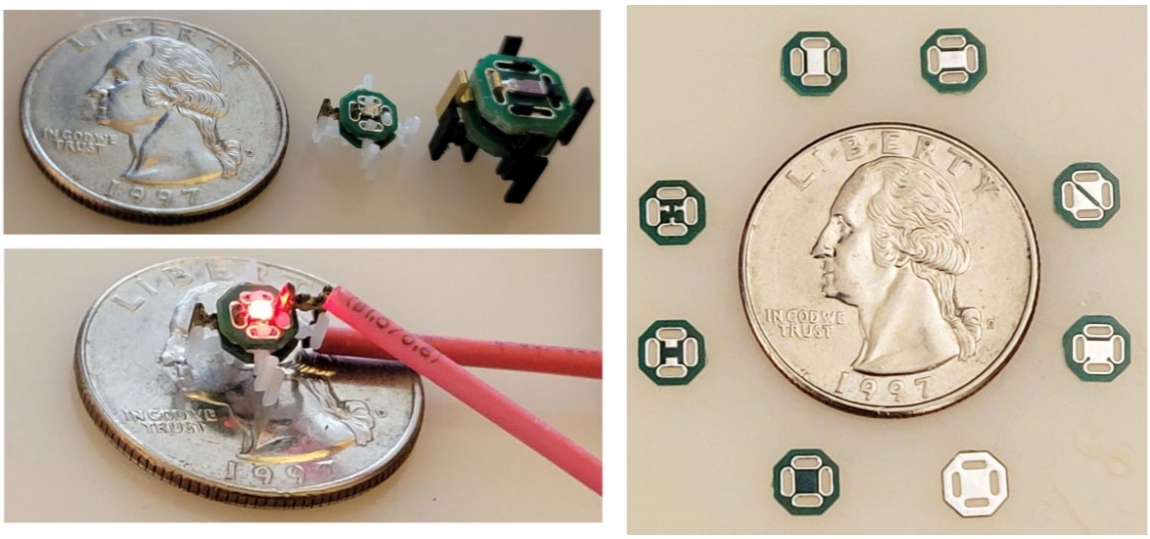


**Figure 4: Universal Set of Routing Tiles for the 4H DICE Geometry (right), and Sample Assembly and Size scaling (left)**

This set of carriers is universal for routing, illustrated here with an H-bridge (left) and ring oscillator (right):

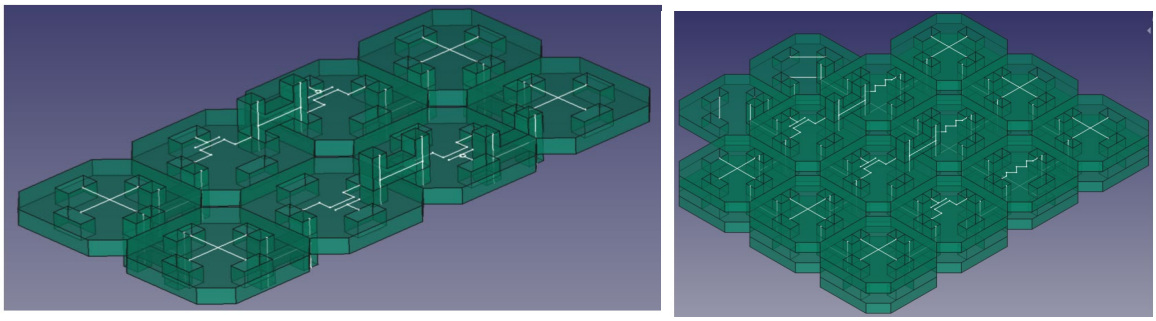


**Figure 5: H-bridge (left) and Ring Oscillator (right) Designs in the DICE 4H Geometry**

For carrying CSP components a tiny-DICE form factor was developed:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

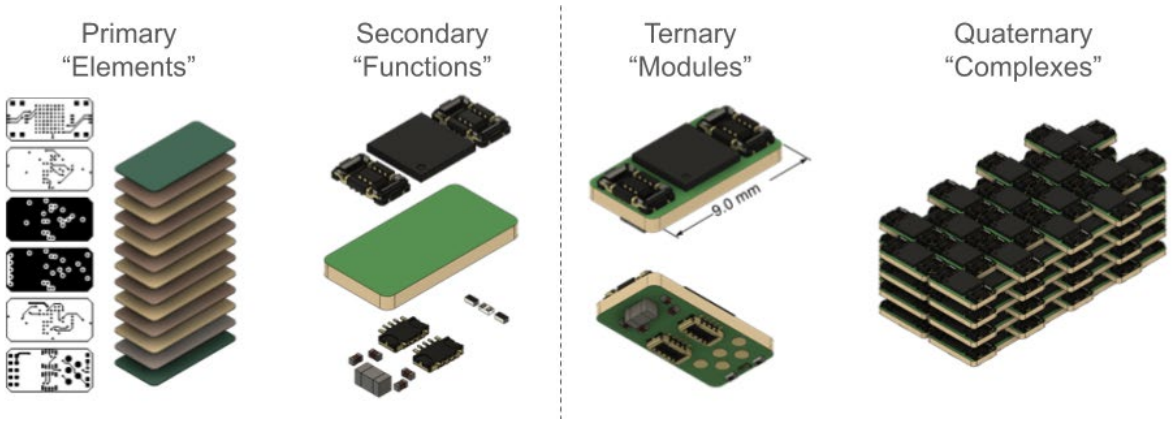



**Figure 6: Primary, Secondary, Tertiary, and Quaternary Assembly of DICE Chip-scale Packaging**

And for carrying CPU-scale components a meso-DICE form factor was developed:

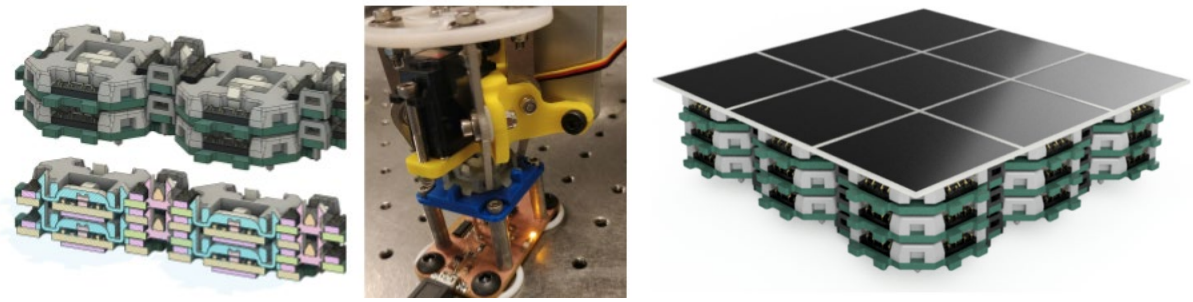


**Figure 7: DICE CPU-scale Parts (left), Robotic Assembly (center), and Distributed Display (right)**

***Assemblers***

Central to the DICE project are the corresponding assemblers at each of the scales. These were developed for discrete-DICE:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

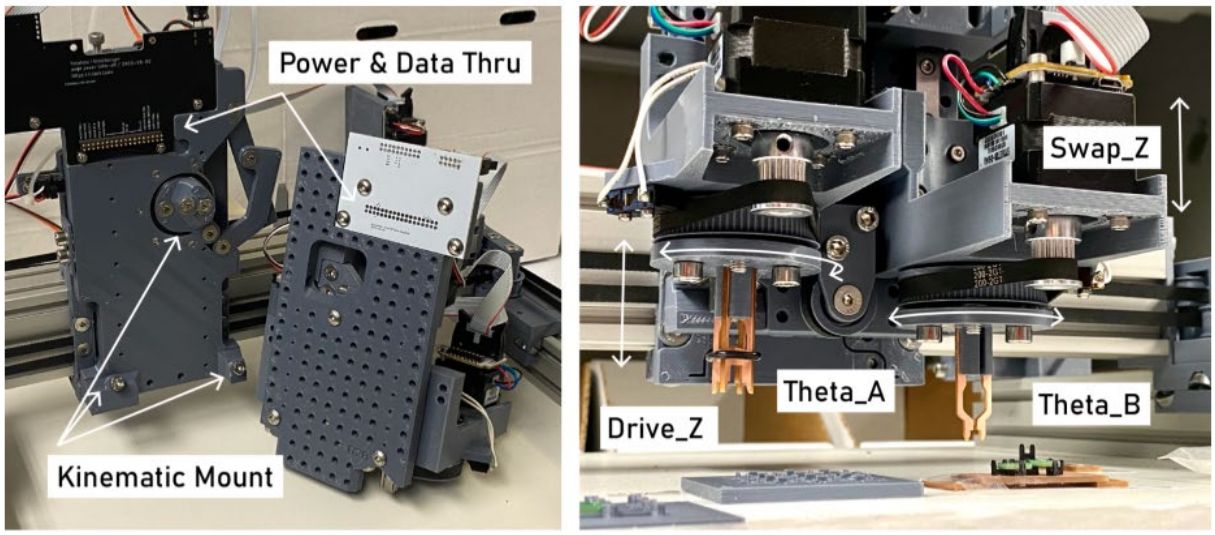



**Figure 8: Custom DICE Assembler Developed for the 4H Geometry, a Cross between a Pick-and-Place and 3D Printer**

for tiny-DICE:

**Figure 9: Custom Assembler Developed for the Tiny-DICE Geometry**

and meso-DICE:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

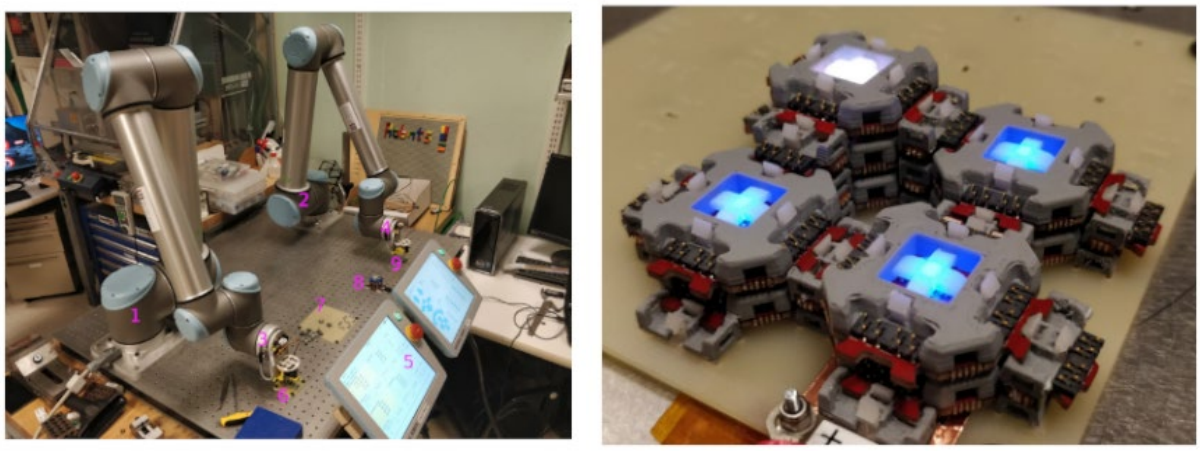


**Figure 10: Robotic Assembly of CPU-scale DICE (left), and an Assembled Test Circuit (right)**

***Design***

Because the DICE workflows differ significantly from existing electronics design automation (EDA), dedicated design tools were developed. These included hardware description languages and CAD user interfaces for three-dimensional electronics:

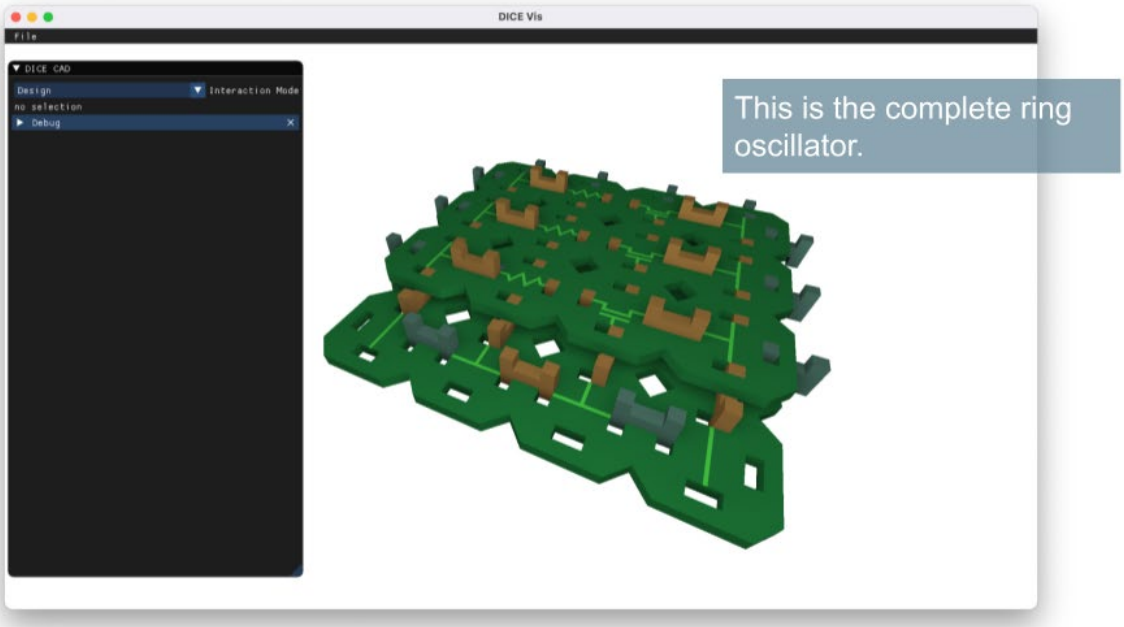



**Figure 11: Design and Simulation tool Developed for 4H DICE**

analog simulation:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

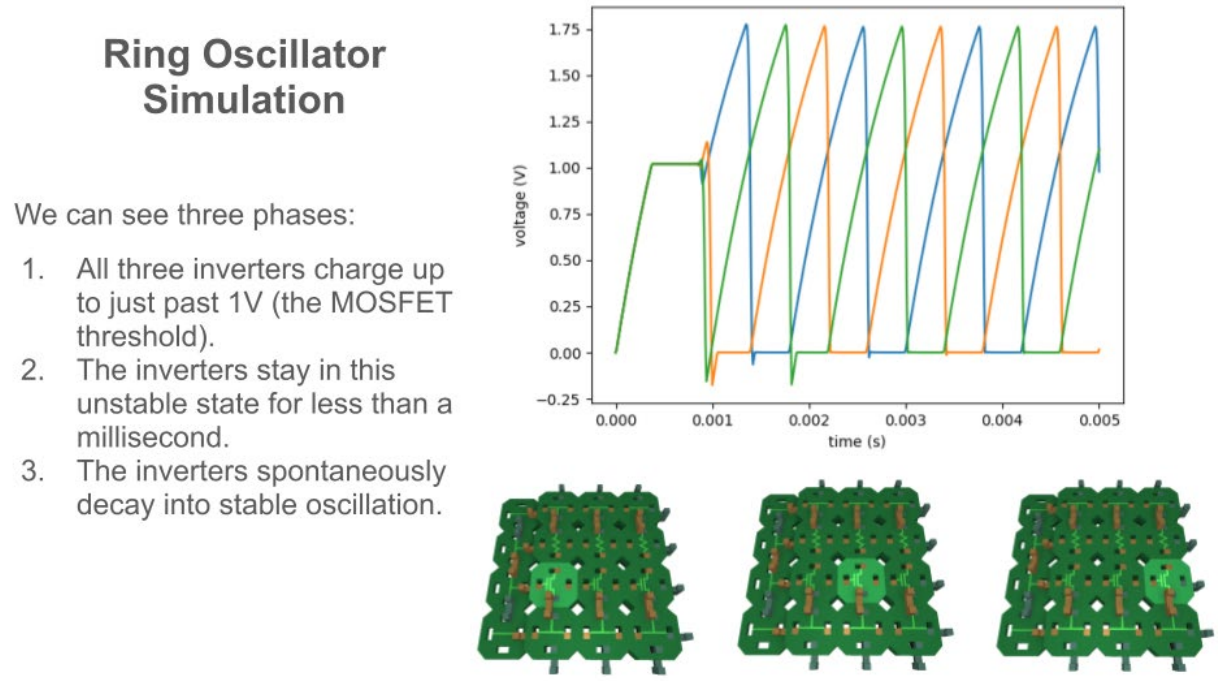



**Figure 12: Simulation of the Ring-up of a DICE Ring Oscillator**

digital simulation:

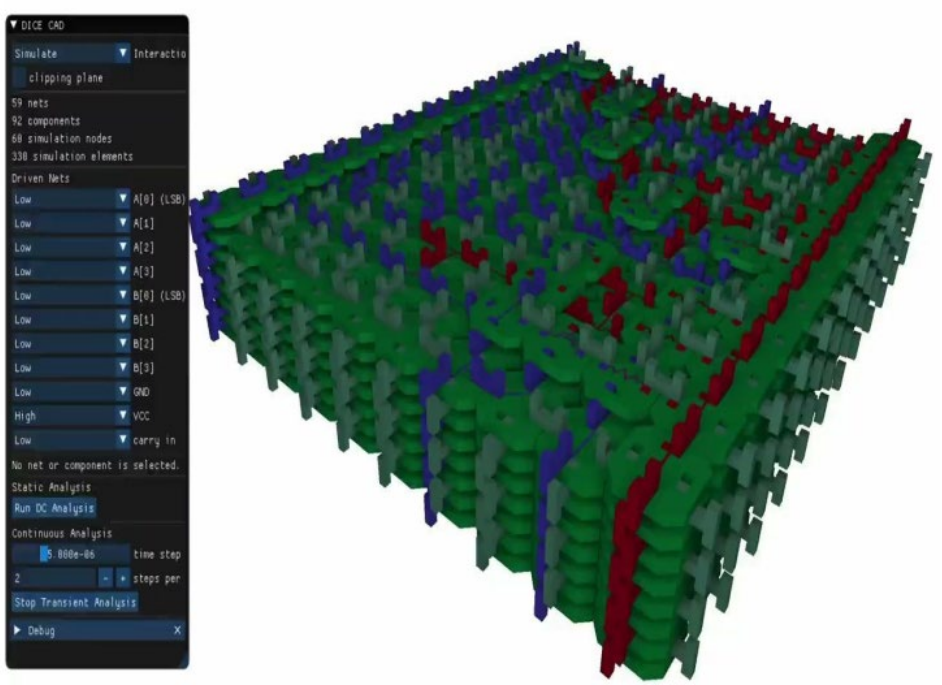



**Figure 13: Discrete Logic Simulation of a Parametric DICE Adder Circuit**

and CAM for assembly:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

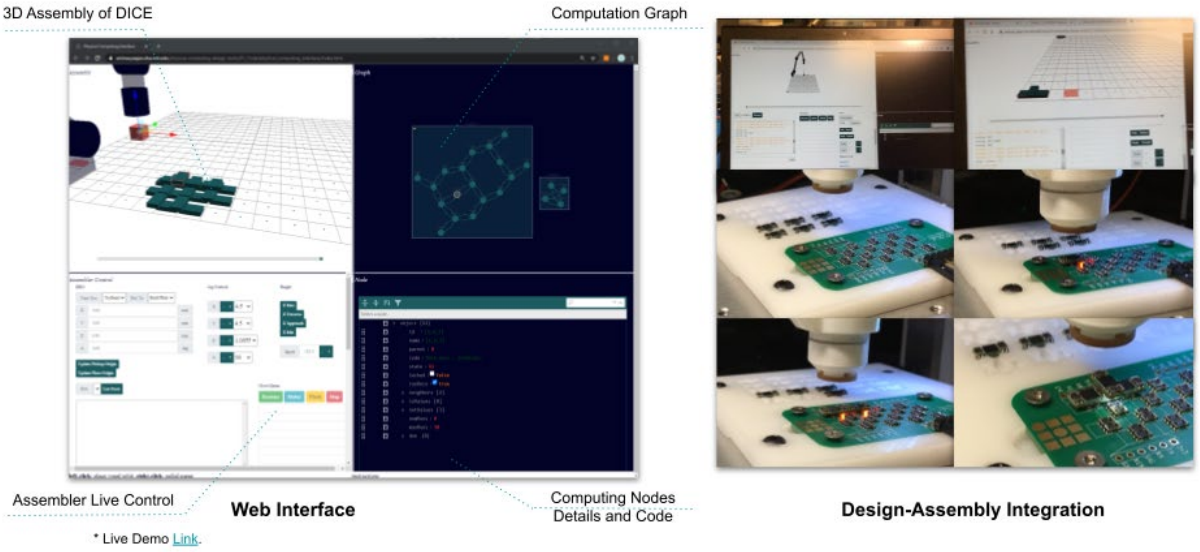



**Figure 14: CAM Tool Developed for Designing the Construction of DICE Computation**

***Programming***

DICE programming requires specifying the geometry of a computation. This was abstracted in a token-passing model that could be executed on CPUs and FPGAs as well as MCUs:

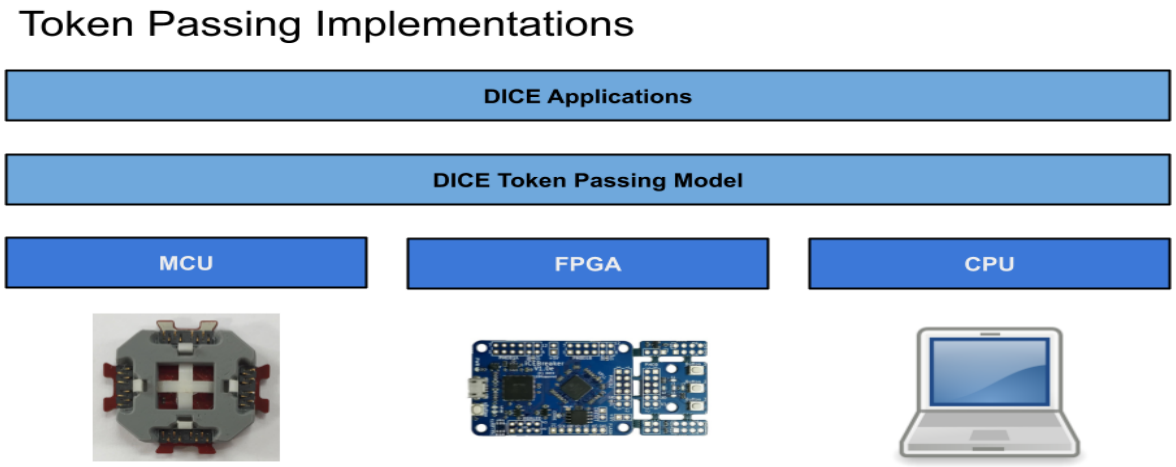



**Figure 15: Abstraction of the Programming Model for DICE Development and Execution**

This was tested on a DICE emulator for applications including a scatter-gather benchmark implemented as a space-filling curve:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

- A single SAMD51 achieves 16.8 MFlops
- The 16 SAMD51s on Static-DICE scale this performance linearly (communication overhead is negligible for this task)
- Calculation is distributed along a space filling curve for maximum scalability

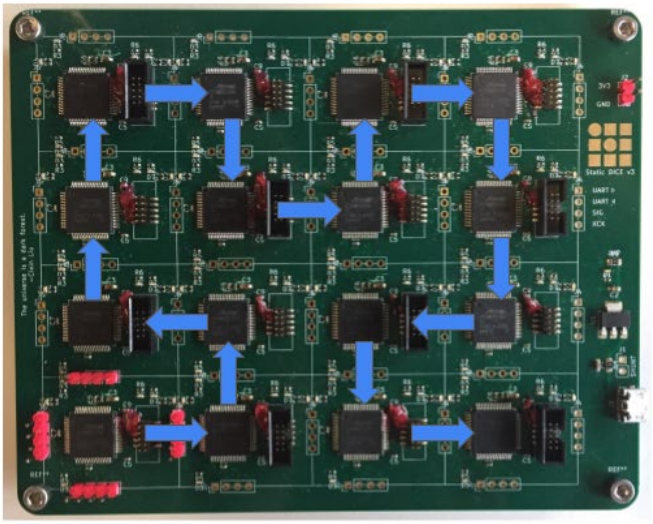


**Figure 16: DICE Programming Development Tool Executing a Scatter-gather Calculation as a Space-filling Curve**

and multi-physics modeling with particle systems, replacing global networks with local message-passing:

- DICE's local message passing is a perfect fit for 3D spatial simulation
- Simulations are spatially decomposed to run on arrays or volumes of DICE nodes

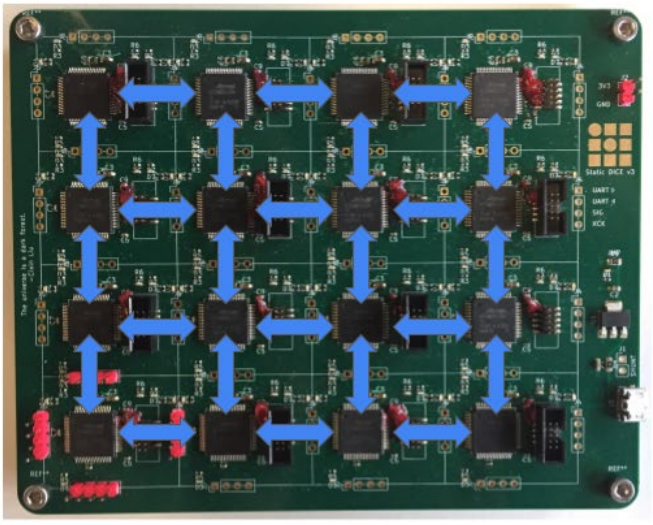


**Figure 17: DICE Programming Development Tool Executing a Locally-connected Multiphysics Particle Method Simulation**

### *Applications*

Work in the DICE project found applications in a number of related areas. One was the development of structural electronics, with modules that carry forces, power, and data, shown here in robots that can assemble themselves from the parts that they're assembling:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

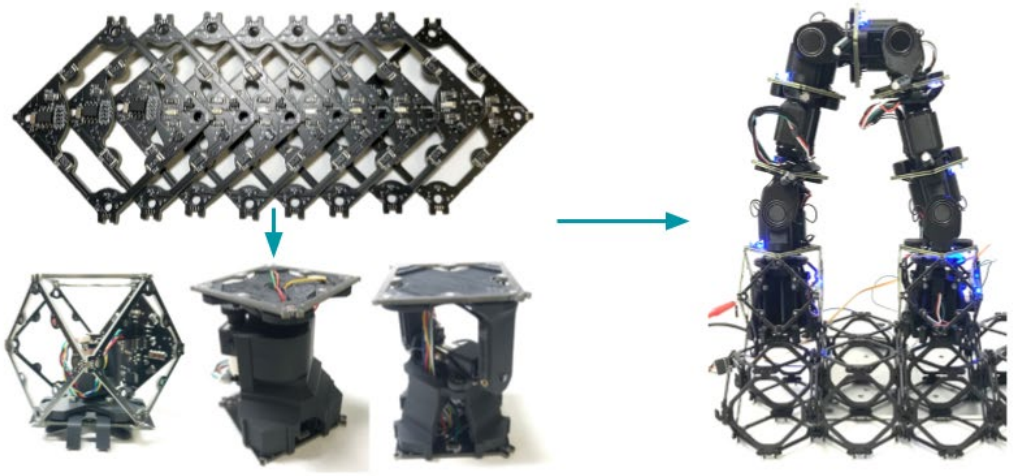


**Figure 18: DICE Spinoff in Structural Electronics, used in a Robotic Assembler that is Assembled from the Parts that it's Assembling**

A second was in superconducting electronics. Both yield and physics limits the chip scale; the Super-DICE project performed cryo-testing of DICE packages and used the Lincoln Labs fab to produce test chips, introducing asynchronous adiabatic quantum flux parametrons (AQFP). For scalability this eliminates the need for clocking, and operates on the scale of zepto-joules for potentially orders-of-magnitude reduction in computing power consumption (including the cryogenic overhead);

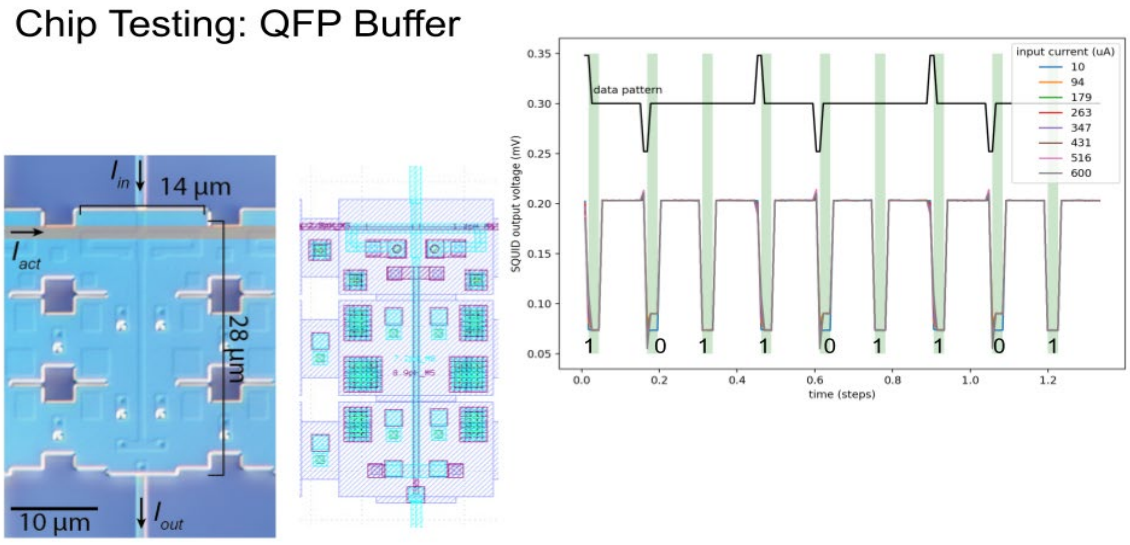



**Figure 19: DICE Spinoff in the Development a Superconducting Asynchronous Adiabatic Quantum Flux Parametron**


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

And the work on DICE prototyping led to the development of techniques for laser micro-machining that are being used for direct-write MEMS fabrication, miniaturized electronics, and single-crystal diamond machining to make 10 million RPM rotors for improving the resolution of solid-state NMR molecular spectroscopy:

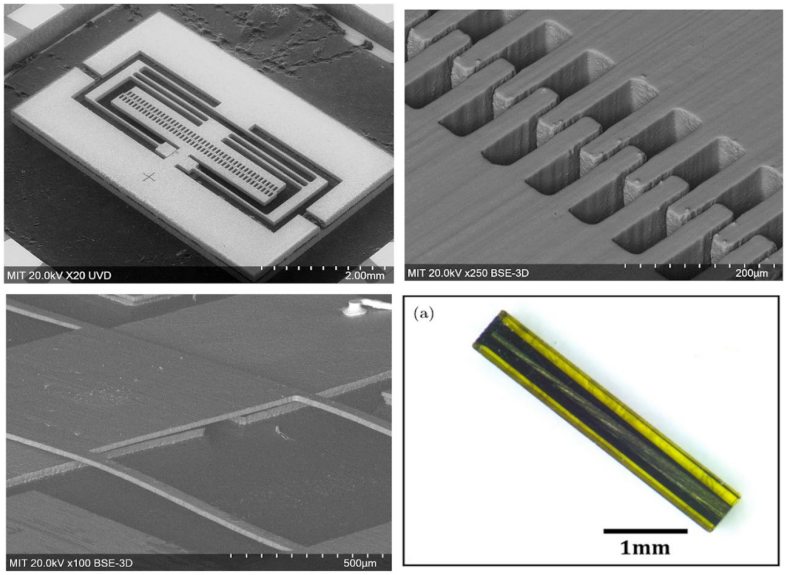



**Figure 20: DICE in Micromachining Direct-write MEMS Devices (top and left), and Single-crystal Diamond Micromachining for 10 Million RPM Rotors for NMR Magic Angle Spinning**


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# 2 RECOMMENDATIONS

The DICE project succeeded at demonstrating the feasibility of three-dimensional electronics assembly on multiple relevant scales, and produced prototypes of end-to-end workflows for their application. To develop those applications, promising areas for future work include:

- ***Packages***
  There are three levels of abstraction between a discrete device in a SMD package in a DICE VMD carrier; that overhead could be eliminated by developing VMD as a native form factor. An initial benefit would be eliminating the need for breadboards for prototyping circuits, and an eventual benefit could be eliminating the need for PCBs.
- ***Circuits***
  The DICE project demonstrated integration on the scale of tens of parts. Once that reaches thousands it will be possible to make in a table-top process the kinds of integrated circuits that shut down factories during the pandemic due to the failure of supply chains, such as microprocessors and motor controllers. Along with scaling all of the DICE results, this will require developing three-dimensional generalizations of two-dimensional place-and-route optimization.
- ***Chiplets***
  Compared to the enormous investment going into chiplet development and fabrication, very little is going into support for design and test around them; DICE carriers could be developed for chiplets that could bypass the hurdle of system-on-chip (SOC) integration for development.
- ***Chips***
  A number of important applications, such as electric vehicles, require voltages and currents beyond what is feasible to carry on-chip. That currently requires complex PCBs with hybrid on- and off-chip solutions; DICE offers an integrated alternative.

Each of these areas is potentially a multi-year program; the impact of each could be shown in a one-year demonstration.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# APPENDICES

# Discrete Integrated Circuit Electronics (DICE)

Zach Fredin,* Jiri Zemanek,*† Camron Blackburn,* Erik Strand,*
Amira Abdel-Rahman,* Premila Rowles,* Neil Gershenfeld*
*The Center for Bits and Atoms
Massachusetts Institute of Technology
Cambridge, MA, USA
first.last@cba.mit.edu
†Faculty of Electrical Engineering
Czech Technical University
Prague, CZ

***Abstract*—We introduce DICE (Discrete Integrated Circuit Electronics). Rather than separately develop chips, packages, boards, blades, and systems, DICE spans these scales in a direct-write process with the three-dimensional assembly of computational building blocks. We present DICE parts, discuss their assembly, programming, and design workflow, illustrate applications in machine learning and high performance computing, and project performance.**

***Index Terms*—Automated assembly, additive manufacturing, chiplets, system integration, machine learning, high-performance computing**

## I. Introduction

There is a familiar hierarchy from developing chips, to packages, to boards, to blades, to systems. Different skills are required for design at each layer, and different industries perform their production. We present an alternative to this prevailing paradigm, DICE (Discrete Integrated Circuit Electronics). DICE is based on the three-dimensional assembly of computational building blocks that can span these levels of description. It extends the role of modularity from design to construction, with a direct-write process using a cross between pick-and-place and 3D printing.

The immediate antecedant to DICE is the assembly of chips from chiplets [1]. These can offer very high interconnect speed and density, but are limited to two-dimensional chip-scale applications. DICE takes advantage of progress in packaging, fabrication, and automation to offer an architecture that spans from devices to systems, with acceptable performance overhead as described below. A closer antecedant is older, the post-war Project Tinkertoy that aimed to robotically assemble electronics from standard building blocks [2]. DICE revisits that goal with updated components.

Potential benefits of DICE include reducing the time and cost for system development, aligning applications with architectures, shortening and reducing vulnerabilities in supply chains, reducing waste with disassembly rather than disposal, and enabling greater system integration and novel form factors.

We gratefully acknowledge support from, and collaboration with, DARPA and the Air Force Research Laboratory under award number FA8650-19-2-7921, Lincoln Labs under ACC 767, the Fulbright Program, and the Center for Bits and Atoms consortium.

Realizing these benefits requires revisiting the historical boundary between hardware and software, in order to simultaneously specify the content and construction of a computation. In the following sections we present an end-to-end DICE workflow, including performance projections (for existing and prospective devices), modules (on multiple length scales), their programming (both embedded and application), assembly (using conventional and custom robots), design (both logical and physical placement), and applications (aligning geometry and algorithms for machine learning and high-performance computing).

## II. Projections

We start with projections of DICE performance before discussing their implementation. Table I compares the speed, power, and cost efficiency of existing and potential DICE devices with other computing platforms to illustrate their potential impact. Assuming a low communication overhead, supported by experimental benchmarking explained in sections III and VI, and adequate cooling and sparsity in the DICE lattice for power distribution, discussed further in section III, DICE nodes can be assembled to match complex computing structures. It would take roughly 3,000 SAMD51 microcontroller DICE nodes to match the compute performance of an Intel Core i7 and the DICE network would require about 30% more power. The same comparison using NXP RT1020 microcontrollers would need approximately 1,000 nodes and consume half of the power of the i7.

Looking beyond commercial-off-the-shelf (COTS) components, DICE has the potential to scale to state-of-the-art supercomputing performance with a dedicated node modeled after the Summit supercomputer's power and cost efficiency [3]. The Qualcomm Hexagon DSP processor has been shown to achieve a maximum power efficiency of 58mW/GHz on an architecture with 256KB of RAM, two 32 bit ALUs, two 64 bit vector execution units, and an asynchronous FIFO bus interface [5]. Taking this as a starting point, the logic units could be replaced with single precision FPUs shown to have 1.44mW power consumption without significant additional cost [6]. Combining these low-power techniques, a DICE node would be able to perform single-cycle floating-point operations


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

TABLE I
DICE PERFORMANCE PROJECTION AND COMPARISON

| | GFlops | GFlops/W | GFlops/$ | W | $ |
|---|---|---|---|---|---|
| Summit ORNL[(a)] | 148,600,000 | 14.7 | 0.45 | 10,096 kW | $325M |
| Nvidia V100[(b)] | 12,589 | 50.4 | 2.10 | 300 W | $5993 |
| Intel i7-8700T[(c)] | 46.9 | 0.34 | 0.155 | 140 W | $537 |
| SAMD51J20A[(c)] | 0.0168 | 0.233 | 0.01 | 72 mW | $2.91 |
| NXP i.MX RT1020[(c)] | 0.0417 | 0.593 | 0.02 | 70 mW | $1.87 |
| "Summit MCU"[(d)] | 1 | 14.7 | 0.45 | 68 mW | $2.22 |
| Super-DICE[(d)] | 5 | $10^6$ | – | 10 nW | – |

[(a)] November 2019 Green500 list [3]
[(b)] Manufacturer's specification.
[(c)] Flops and power measured with a pi calculation benchmark [4]
[(d)] Projected calculations for dedicated DICE hardware.

at 1 GHz while consuming 68mW with a cost of $2.22. Described here as the "Summit MCU" due to matching its 14.7 GFlops/W and 0.45 GFlops/$, just 50 nodes would outperform the Intel Core i7 while using 97% less power.

Development of rapid single flux quantum (RSFQ) superconducting electronics is limited by their cryogenic cooling overhead and a maximum of $O(10^4)$ Josephson junctions per computing unit due to bias current in the circuit [7]. DICE provides a promising solution for these limitations by building complex computing from simple low gate-count nodes and making better use of cryogenic refrigerator volumes with its 3D scaling. It has been shown that 8-bit superconducting ALUs can operate at 5 Ghz with 1.4 aJ power dissipation [8]. Extrapolating this to a custom Super-DICE node, a chip with $O(10^4)$ Josephson junctions operating at 5GHz would require roughly 10 nW per operation. Taking into account the $10^3$ W/W overhead for 4K cooling, superconducting DICE would perform at $10^6$ GFlops/W, a $10^5$ power efficiency improvement over state-of-the-art supercomputing.

## III. MODULES

### A. Design Considerations

Based on the preceding projections, DICE modules were developed with COTS components to demonstrate an end-to-end workflow rather than ultimate performance limits. A core element of DICE is the direct physical assembly of computational elements which can be configured to suit both planned and unexpected applications. Interconnection between elements supports the transfer of information, power, and mechanical loads, and must be durable enough to permit repeated re-assembly [9], whether by hand or by purpose-built machine.

Prior to developing dedicated DICE devices, a number of general-purpose microcontrollers were considered for prototyping the workflow. The search space was limited to devices with floating-point units (FPUs), since planned applications would require this capability. Microcontrollers were further evaluated based on estimated performance (net and per watt), along with minimum available package size, RAM, and built-in communication peripherals. FPGAs were also considered as they would allow considerable flexibility in terms of internode communication implementation and problem-specific gate configurations [10], but were avoided for these initial iterations due to space concerns and lack of direct portability with existing applications. Based on this evaluation, and the availability of a well-documented, open development toolchain [11], the Microchip SAMD51J20 ARM Cortex M4F microcontroller [12] was selected for initial prototyping of a general-purpose DICE computing module, with more specialized heterogeneous modules anticipated to follow.

### B. Physical Iterations

Initially, two hardware iterations were designed and tested in parallel. The first design, called Static-DICE, was not mechanically reconfigurable, but rather was intended to quickly provide a network of elements large enough to develop and test inter-node communication layers. Given von Neumann neighbors [13] (i.e Cartesian connections only, or four interconnections per node), a minimum of twelve nodes were needed if laid out in a two dimensional grid. Four additional nodes were added to the final design, forming a square 4x4 grid with sixteen elements.

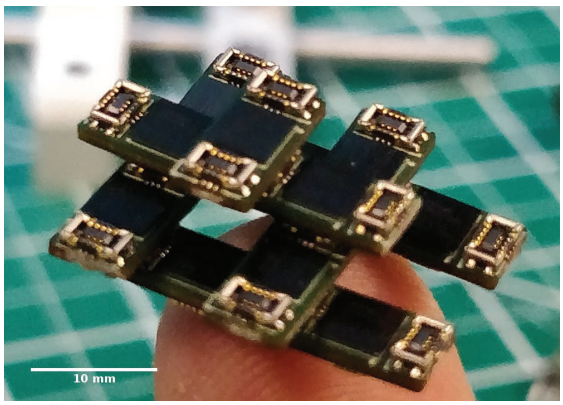



Fig. 1. Eleven Tiny-DICE nodes assembled.

The second design, called Tiny-DICE (see Fig. 1), consisted of mechanically reconfigurable modules capable of tiling with


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

four neighbors per node. Tiny-DICE was used to study the practical limits of commercially available hardware, both in terms of size (using the smallest parts possible) and suitability for mechanical assembly (as small-scale board-to-board connectors are typically not designed for parallel configurations or structural use). As with Static-DICE, Tiny-DICE used Microchip SAMD51J20 processors, this time in 3.5 x 3.6 mm wafer-level chip-scale-package (WLCSP) format. Power, information, and mechanical interconnection was provided by 6-pin Molex SlimStack [14] mezzanine-type connectors. The small ball pitch on the WLCSP microcontroller necessitated a commercially sourced high density interconnect PCB with six layers, 100 um traces, and blind vias. Final node dimensions of 4.5 x 9.0 mm were driven by the connector and processor size. Board outline tolerances were below manufacturer capabilities so PCB edge routing was performed in-house on a precision CNC mill prior to automated component placement and reflow soldering. After verifying solder joint integrity using X-ray inspection, a spring contact programmer was fabricated to flash Tiny-DICE nodes with test code.

Tiny-DICE devices were evaluated for power consumption, computation, and thermal performance as a fifteen-node cluster on a build plate as shown in Fig. 2. Each device was programmed with a simple pi calculation benchmark [4] set to run for one million iterations. Power consumption was measured using a current shunt and an oscilloscope, and the results compared to other hardware devices as shown in Table I. Convective cooling was adequate on this small scale; for system scaling, the sparsity of the DICE lattice offers a design degree of freedom that can vary the density of dissipation, and can interdigitate computation with a cooling medium.

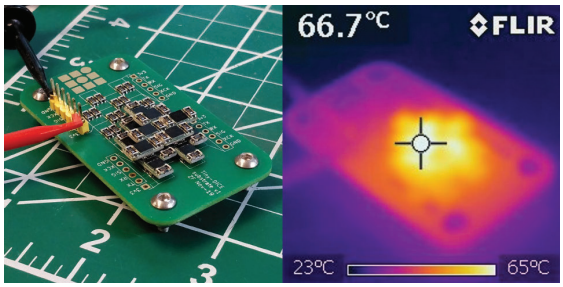



Fig. 2. Left, fifteen Tiny-DICE nodes connected in a cluster on a build plate with an external supply and instrumentation; right, the same cluster after reaching steady state running a pi calculation benchmark, viewed through a long-wave IR thermography camera.

Lessons from both Static-DICE and Tiny-DICE informed a third hardware iteration, called Meso-DICE, shown in Fig. 3. These devices are intended to quickly demonstrate the complete workflow, from design tools through automated assembly, application execution, and mechanical reconfiguration. To simplify construction and assembly mechanics, the hardware is scaled roughly to Static-DICE proportions, and includes fabricated mechanical features to self-align and latch nodes together. Meso-DICE elements tile in a simple cubic lattice with six neighbors; to ease assembly, passive strut elements link adjacent nodes using COTS spring terminals. As with Tiny-DICE, a spring contact programmer was fabricated to simplify firmware development.

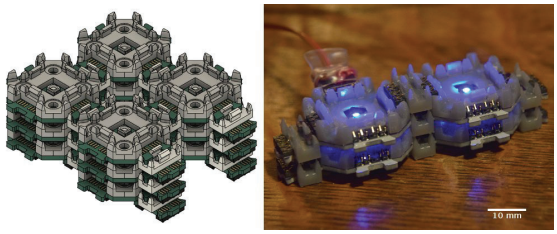



Fig. 3. Left, twelve rendered Meso-DICE nodes and struts are assembled to form a 2x2x3 lattice. Right, four functional test elements are assembled and powered on.

## IV. Assembly

### A. Tiny-DICE assembler

Two approaches were evaluated for automating DICE assembly. For Tiny-DICE parts, a dedicated machine was developed that is a cross between a pick-and-place and a 3D printer. The purpose of this automatic assembler is to build a physical structure according to the layout created in the DICE design tool. The machine has to pick Tiny-DICE parts from the stack and place them to the final positions. The build process starts with putting the parts on the substrate; the next layer is then placed on top of the modules on the substrate, and so on. Optionally, before the deposition, the assembler can put every part into a programming station to load firmware with a bootloader that allows later updates.

The Tiny-DICE assembler has a standard XYZ configuration with a rotational axis on the tool head. The assembler frame is machined from steel reinforced HDPE and aluminum. To limit friction, all axes use polymer linear guideways. XY axes are driven by standard stepper motors NEMA-17 with GT2 timing belt and the Z-axis is driven by a non-captive stepper motor with a leadscrew. A control board receives G-code commands from a computer and drives the motors with a resolution of 5 µm (possible repeatability is approximately 50 µm) and max speed 50 mm/s. The assembler also has an on-tool USB microscope that allows precise jogging and optical registration.

To reduce the size as much as possible, the Tiny-DICE parts have no features to facilitate assembly, and due to the dense packing of the final assembly, the space for grabbing of parts is limited. Therefore, the assembler utilizes a pair of mating SlimStack connectors to attach the part to the tool head. An extendable pin actuated by a linear servo (with max force of 35 N) is then used to release the part from the tool head and place it to the final destination.

We successfully tested building structures with several layers. Limiting factors are the self-aligning capabilities of the connectors and the absence of mechanical support under upward-facing connectors on Tiny-DICE parts. Therefore,


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

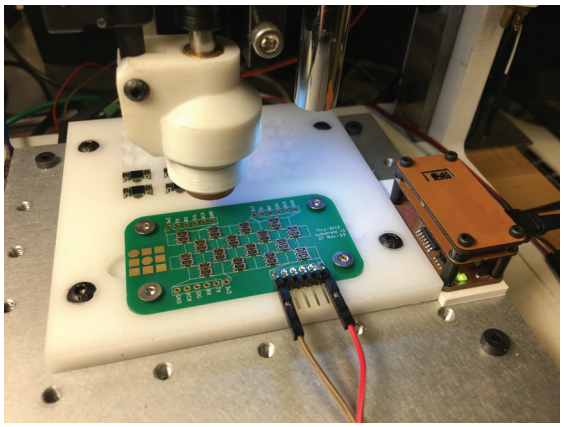


Fig. 4. Automatic Tiny-DICE assembler. The substrate is in the front, the stack of parts in the back, and the programming station on the right side.

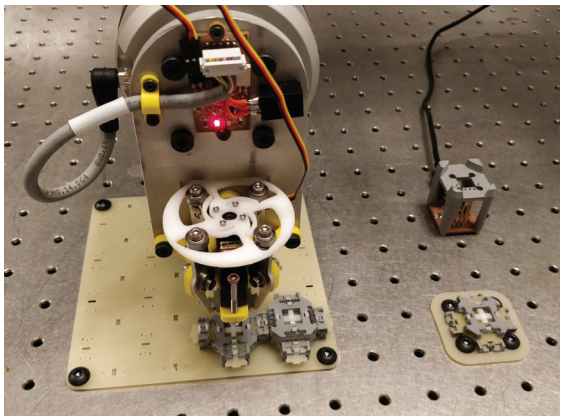


Fig. 5. Automatic Meso-DICE assembler. Left, the end effector and control circuit is mounted to the end of a UR-10 robotic arm, shown placing a Meso-DICE node on the build plate; bottom right, the Meso-DICE node pickup location; top-right, the integrated Meso-DICE programmer.

the successful placing of parts relies on the rigidity of the connection between already-placed parts.

### B. Meso-DICE assembler

For the larger Meso-DICE platform, a 6-axis Universal Robots UR10 arm was used to manipulate parts. An end effector was fabricated which uses a hobby servo motor to actuate a machined acetal cam, which then closes four aluminum jaws around a node part. The cam uses a flexural profile which gives each follower a small amount of compliance, allowing the end effector to adapt to imperfect pickup coordinates and slight misalignment in the built lattice. A small circuit mounted to the end effector translates native UR10 signals into servo commands while limiting stroke to avoid overloads, allowing the system to be controlled using the UR10's pendant, an open-source Python scripting library [15], or the DICE design tool.

In use, Meso-DICE parts are picked from a dedicated pickup location. As with the Tiny-DICE assembler, the Meso-DICE system has an integrated programmer for flashing nodes with new firmware during assembly. Computational structures are built up on a dedicated fiberglass build plate.

## V. Design

Traditional electronics design workflows follow a sequential linear process that starts with design, then analysis and simulation, and finally system fabrication and testing. Often these stages are executed independently from each other, using different tools, making it hard to translate the feedback from the simulation or testing stages into viable design amendments. This adds considerable inefficiency to an inherently iterative design workflow. As an alternative, we developed an integrated, closed loop DICE design tool where one can design, simulate, optimize and fabricate re-configurable computing systems (see Fig. 6).

The design workflow follows several steps. First, DICE nodes are placed to visualize the 3D structure of the computing system. The user then selects individual nodes to specify

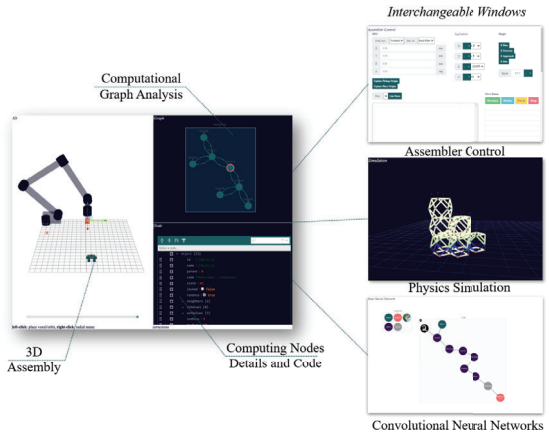



Fig. 6. Integrated DICE design tool with interchangeable windows to enable the simultaneous design of software and hardware, as well as assembly and testing in real time.

details, such as processor type, connected neighbors, program, number of operations, maximum token size, and so forth. During these steps, a connectivity graph is generated in real-time, showing the average computation and communication cost of each node based on these preprogrammed node details. In the future, these numbers could be automatically calculated by pre-compiling and profiling code for each node.

By treating the computation graph as a dataflow network, graphical algorithms are used to find and highlight computation and communication bottlenecks. Based on this analysis, design recommendations are generated and displayed, including suggestions for different processors, extra nodes, or alternative routing.

If the user wants to further optimize the reconfigurable computing system, the performance projection graph tool visualizes the effect of each hardware and software decision


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

on projected overall system performance. A generative probabilistic programming model is used to draw samples from these inputs and project the cost, energy use, and speed of the system. Using classic inference techniques one is able to infer a hardware configuration for a given software workflow. In the future this could be extended to additionally perform hardware architecture discovery.

After optimization, the simulation window allows the user to visualize the state and output of the DICE computation. Currently, this shows the output of a simulated computation provided on the host computer prior to physical assembly. In the future, this framework will be used to visualize real-time data from the DICE network, augmenting display nodes.

In parallel with the previous steps, the interactive assembly control window provides a simple mechanism for controlling the mechanical DICE assembler. Using this feature, systems can be fabricated, programmed, and tested to quickly pull in feedback and amend the system architecture accordingly.

This integrated design workflow paves the way for the design of both discretely integrated circuits and reconfigurable computing systems that can change and evolve during execution as needed.

## VI. Application

### A. Programming Model

The DICE programming model emphasizes the physical locality of both memory and compute resources. By reflecting the underlying compute hardware, this strategy facilitates reconfigurability, scalability, and energy efficiency. As a result, applications must be implemented with distributed algorithms that operate on distributed data. This requires a different set of tools than is conventionally used for writing parallel code.

Synchronization and communication between DICE nodes is achieved via the transfer of tokens. Tokens carry arbitrary data payloads and have application defined identifiers. DICE nodes can check for the presence or absence of a token with a certain identifier from a certain neighbor. Tokens are exchanged only between neighboring nodes, so if data is to be passed between physically separated nodes propagate along a chain of intermediate nodes to get there.

Conceptually, DICE nodes can be viewed as actors that process and exchange these tokens. Input and output data of computations are passed around as token payloads, and each DICE node greedily performs computations as inputs become available. Currently this model is implicit in the application codes discussed below. Future versions will employ a domain specific language that associates tokens with actions explicitly, ceding the main loop to the DICE runtime environment and allowing it to invoke actions as their tokens become available.

### B. Comparison with Existing Models

In the sense that DICE networks consist of regularly arranged data processing units, they are similar to systolic arrays [16]. However for DICE there need not be consistent directions of data transfer, and the complexity of the data and computations handled by each node is typically higher.

DICE networks are similar to Petri nets in that they produce and consume tokens along a graph [17]. However DICE networks impose a consistent lattice structure on the graph, and allow arbitrary data payloads to be associated to the tokens. It is also more common for connections between DICE nodes to be bidirectional than in Petri nets. In a similar manner, DICE networks resemble asynchronous logic automata (ALA) [18], [19], where ALA's single bit tokens are replaced with DICE's arbitrary payloads, and ALA's boolean algebra operations are replaced with DICE's arbitrary computations.

In Flynn's taxonomy [20], each DICE node is a SISD computer. A network of DICE nodes is a MIMD computer; future implementations could also include SIMD or MIMD nodes.

### C. Implementation

For the existing SAMD51-based prototype nodes, the DICE programming model and runtime environment is implemented as a C/C++ library organized into three layers. The first layer is a hardware abstraction layer that interfaces with the SAMD51's peripherals. In particular, a full implementation of the firmware stack requires use of the generic clock controller (GCLK), nested vector interrupt controller (NVIC), external interrupt controller (EIC), and the universal synchronous and asynchronous receiver and transmitter (SERCOM USART). The second layer is a data link layer that enables transfer of raw bytes between neighboring nodes. It serves an analogous purpose to the data link layer in the OSI model [21]. The third layer is the token layer, which implements the token passing model described above.

### D. Physical Simulation

DICE lends itself to problems which can be subdivided via local data parallelism. One such problem is physical simulation, which is also a leading use of national class supercomputers. Here, we describe a generalized discrete element method (DEM) simulation for DICE.

Each DICE node is responsible for simulating a certain region of space [22]. Adjacent regions within the simulation are computed by adjacent DICE nodes. As particles move between these regions, their data is transferred between the corresponding nodes. Computationally, each node computes local interaction forces between particles, and uses these forces to integrate the particles' velocities and positions forward in time. Particles near region boundaries may interact across them, requiring some data transfer. A critical factor for performance is the density of interacting particle pairs, since the calculation of interaction forces dominates compute time. In practice this density is limited by an interaction cutoff distance.

Within each DICE node, the memory required to store particle data and the computational effort required to compute interaction forces are both proportional to the volume of the simulated region. The data that must be transferred between nodes, however, is proportional to the surface area of the region. Thus the overhead of communication between nodes is reduced as the size of the regions simulated by each node


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

grows. This imposes a limit on the acceleration one can achieve for a single, fixed size simulation by distributing it across more nodes. But larger simulations (in physical extent and number of particles) can be run without any increase in run time if the number of nodes is increased proportionately.

Even for the existing generation of SAMD51J20 based nodes, the projected communication overhead is not prohibitive. Each SAMD51 has 256KB of RAM, can be clocked at 120MHz, and has six SERCOMs that can each transmit a maximum of three million bits per second. Given these constraints, and assuming an interaction cutoff distance of three times the characteristic particle spacing at maximum density, a single node can store approximately 1,700 particles. At maximum capacity, each node would compute ~135,000 interaction forces in ~1.3 seconds. Computing these forces would require transferring ~200KB of particle data between each node and its neighbors, which would take ~85ms if all six SERCOMs are used in parallel. Thus if data transmission and force computation happen serially, about 6% of each DICE node's time would be spent transferring data. In practice data transmission and computation can be overlapped using DMA.

The projected Summit MCU discussed earlier would enable much greater performance. In particular, with 1 GFLOPS and 256KB of RAM, each node could compute its particles' interaction forces in ~16ms. To keep the communication overhead manageable, the Summit MCU's SERCOMs should be proportionately faster than the SAMD51's. In particular, with six 128MBit/s SERCOMs, each node could transfer the necessary particle data in ~2ms. This would result in each node spending ~11% of its time transferring data, assuming no overlap with computation.

### *E. Machine Learning*

Another important application for DICE is to create reconfigurable computing systems for machine learning. Deep Neural Networks (DNNs) are currently used in countless applications, and through time, the models are becoming deeper and more sophisticated. In an attempt to benchmark and compare the training performance of variably sized and shaped DNNs on different hardware architectures (CPUs, GPUs or TPUs), its was concluded that there were no winners [23]. TPUs had the highest throughput and worked best with large batches, GPUs were more flexible for small problems and irregular computations, and CPUs were the most programmable and were the only ones to support larger models [23]. There is an increased need for accelerators and application-specific hardware in order to reduce data movement, one of the main bottlenecks of deep learning training, without compromising accuracy, throughput and cost [24].

Consequently, joint hardware/software design workflows are essential for developing a deep learning system, where spatial computing architectures are tailored to the depth and shape of the DNNs, as well as to the size of the training data. This will minimize the data movement and shared memory access, which dominates the energy consumption in traditional computing architectures.

As a first step to address this problem, a machine learning add-on was implemented as part of the integrated physical computing design tools (see Fig. 6). There, one is able to choose and add different kinds of DNN layers and specify their size, activation function, and parallelism strategy. The add-on also has a real-time graphical visualization of the training progress showing the updated accuracy of the model though time.

In order to benchmark and estimate the computing requirements for DICE to train DNNs, we chose AlexNet, a Convolutional Neural Network (CNN) that is widely used for benchmarking hardware as it was the first CNN to win the ImageNet challenge [25]. AlexNet consists of five convolutional (CONV) layers and three fully connected (FC) layers. For a 227x227 input image, it requires 61M weights and 724M multiply-and-accumulates (MACs). Similar to most DNN architectures, the FC layers have significantly more weights than CONV layers (58.6M vs 2.3M) and CONV layers are more computationally expensive than FC layers (666M MACs vs 58.6M MACs).

Assuming complete data parallelism, a number of challenges arise when naively trying to map AlexNet onto a DICE system that uses only one type of node (the SAMD51 processor prototype). Since each processor has only 256Kb of RAM, for FC layers, one might need up to 2300 nodes just to store the weights and perform the calculations, which will result in a communication overhead of more than 1600%. Therefore, specialized node types are required to efficiently map AlexNet, or any DNN, into a DICE architecture in an effort to minimize data movement, maximize number of parallel computations, and minimize the number of idle nodes. One is to design hierarchical memory allocation and access. Dedicated memory nodes could store data (filters, weights or input images) which is hierarchically broadcast based on the layer architecture. This enables temporal and spatial data reuse, where the data is read only once from the expensive memory and is sent to the small local cheap memory for reuse. Moreover, the number of weights stored and computation performed can be pruned by introducing specialized nodes that address the sparsity generated when using the ReLU as an activation function. For example, AlexNet's layers have around 19-63% sparsity. This has proven to reduce the energy cost by 96% using similar spatial architecture hardware [26]. If these changes were implemented, in addition to using the projected Summit MCU performance instead of the SAMD51, the computation speed will increase by 85x and the average communication overhead for FC layers will decrease to 33%.

Even though the performance of the first generation DICE demonstration for deep learning applications does not outperform current specialized HPC architectures, advances in Internet of Things and embodied computing require computation to be physically near sensors and data collection devices in order to minimize large data transfer. The modularity and ease of assembly of DICE will facilitates this system integration along with the addition of specialized robotic and electro-mechanical input and output modules for data collection, actuation and


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

output visualization.

Furthermore, one key aspect of DICE modules is their reconfigurability; this means that in addition to changing the hardware modules to fit different types of networks, one can reconfigure the hardware itself as a system learns. Recent research in deep learning is developing dynamic DNNs that not only optimize their weights but also their structure, which means that the hardware should be optimized to match at different stages of the training. This reconfigurability will also be essential for online learning and probabilistic inference tasks where the computing architecture grows as more data are presented or more analysis is needed.

## VII. Conclusion

The DICE framework provides a novel approach to system architecture for high performance computing. As power consumption and memory bandwidth continue to limit exascale system development [27], DICE blends a low-power node architecture, distributed memory hierarchy, 3D packaging, and intuitive design implementation to align hardware with software.

Beyond the potential raw compute power of DICE nodes detailed in section I, the design tools and automated assembly of the DICE framework introduces system scalability and structural flexibility that are not available with existing supercomputing racks and chassis. Building a custom supercomputer or expanding an existing cluster requires rework of middle layer code and complicated compiler optimizations [28]. This development bottleneck is bypassed in DICE due to its direct mapping between programming logic and spatial computing. Furthermore, the 3D packaging of DICE nodes and their unique interconnect design allows myriad physical configurations with potential for sparse lattice structures to optimize thermodynamics, for structural elements for robotic design, and for incorporating novel device physics.

DICE introduces a new paradigm for high performance parallel computing which realigns the relationship between design, assembly, and architecture to seamlessly transition from logic to construction, and lays the groundwork for significant HPC power gains with future dedicated device development.

## References


[1] S. K. Moore, "Chiplets are the future of processors: Three advances boost performance, cut costs, and save power," *IEEE Spectrum*, vol. 57, no. 5, pp. 11–12, 2020.

[2] R. Henry, "Project tinkertoy: a system of mechanized production of electronics based on modular design," *IRE Transactions on Production Techniques*, vol. 1, no. 1, pp. 11–11, 1956.

[3] [Online]. Available: https://www.top500.org/green500

[4] N. Gershenfeld, *The Nature of Mathematical Modeling*. Cambridge University Press; 1st edition, Nov. 1998.

[5] M. Saint-Laurent, P. Bassett, K. Lin, B. Mohammad, Y. Wang, X. Chen, M. Alradaideh, T. Wernimont, K. Ayyar, D. Bui, D. Galbi, A. Lester, M. Pedrali-Noy, and W. Anderson, "A 28 nm dsp powered by an on-chip ldo for high-performance and energy-efficient mobile applications," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 1, pp. 81–91, 2015.

[6] A. A. Del Barrio, N. Bagherzadeh, and R. Hermida, "Ultra-low-power adder stage design for exascale floating point units," *ACM Trans. Embed. Comput. Syst.*, vol. 13, no. 3s, Mar. 2014. [Online]. Available: https://doi.org/10.1145/2567932

[7] I. I. Soloviev, N. V. Klenov, S. V. Bakurskiy, M. Y. Kupriyanov, A. L. Gudkov, and A. S. Sidorenko, "Beyond moore's technologies: operation principles of a superconductor alternative," *Beilstein Journal of Nanotechnology*, vol. 8, p. 2689–2710, Dec 2017. [Online]. Available: http://dx.doi.org/10.3762/bjnano.8.269

[8] N. Takeuchi, T. Yamae, C. L. Ayala, H. Suzuki, and N. Yoshikawa, "An adiabatic superconductor 8-bit adder with 24kbt energy dissipation per junction," *Applied Physics Letters*, vol. 114, no. 4, p. 042602, 2019. [Online]. Available: https://doi.org/10.1063/1.5080753

[9] W. K. Langford, "Electronic digital materials," Master's thesis, Massachusetts Institute of Technology, 2014. [Online]. Available: http://cba.mit.edu/docs/theses/14.08.Langford.pdf

[10] R. Amerson, R. J. Carter, W. B. Culbertson, P. Kuekes, and G. Snider, "Teramac-configurable custom computing," in *Proceedings IEEE Symposium on FPGAs for Custom Computing Machines*, 1995, pp. 32–38.

[11] ARM, "Gnu toolchain," Accessed on: June 19, 2020. [Online]. Available: https://developer.arm.com/tools-and-software/open-source-software/developer-tools/gnu-toolchain/gnu-rm

[12] Microchip, "Atsamd51j20a - 32-bit sam microcontrollers," Accessed on: June 18, 2020. [Online]. Available: https://www.microchip.com/wwwproducts/en/ATSAMD51J20A

[13] A. W. Burks, "Von neumann's self-reproducing automata," *University of Michigan Technical Report*, 1969.

[14] Molex, "Slimstack board-to-board connectors," Accessed on: June 18, 2020. [Online]. Available: https://www.molex.com/molex/products/family/slimstack_fine_pitch_smt_board_to_board_connectors

[15] O. Roulet-Dubonnet, "python-urx," Accessed on: August 28, 2020. [Online]. Available: https://github.com/SintefManufacturing/python-urx

[16] H. Kung and C. E. Leiserson, "Systolic arrays (for vlsi)," in *Sparse Matrix Proceedings 1978*, vol. 1. Society for industrial and applied mathematics, 1979, pp. 256–282.

[17] J. L. Peterson, "Petri nets," *ACM Computing Surveys (CSUR)*, vol. 9, no. 3, pp. 223–252, 1977.

[18] D. Dalrymple, N. Gershenfeld, and K. Chen, "Asynchronous logic automata." in *Automata*, 2008, pp. 313–322.

[19] N. Gershenfeld, D. Dalrymple, K. Chen, A. Knaian, F. Green, E. D. Demaine, S. Greenwald, and P. Schmidt-Nielsen, "Reconfigurable asynchronous logic automata: (rala)," *SIGPLAN Not.*, vol. 45, no. 1, January 2010. [Online]. Available: https://doi.org/10.1145/1707801.1706301

[20] M. J. Flynn, "Some computer organizations and their effectiveness," *IEEE transactions on computers*, vol. 100, no. 9, pp. 948–960, 1972.

[21] J. D. Day and H. Zimmermann, "The osi reference model," *Proceedings of the IEEE*, vol. 71, no. 12, pp. 1334–1340, 1983.

[22] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," Sandia National Labs., Albuquerque, NM (United States), Tech. Rep., 1993.

[23] Y. E. Wang, G.-Y. Wei, and D. Brooks, "Benchmarking tpu, gpu, and cpu platforms for deep learning," *arXiv preprint arXiv:1907.10701*, 2019.

[24] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[26] Y.-H. Chen, T. Krishna, J. S. Emer, and V. Sze, "Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks," *IEEE journal of solid-state circuits*, vol. 52, no. 1, pp. 127–138, 2016.

[27] S. Gibson, "Revisiting the 2008 exascale computing study at sc18," HPC Wire, Nov. 29, 2018. Accessed on: June 15, 2020). [Online]. Available: https://www.hpcwire.com/2018/11/29/revisiting-the-2008-exascale-computing-study-at-sc18/

[28] E. M. Arvanitou, A. Ampatzoglou, N. Nikolaidis, A. Tsintzira, A. Ampatzoglou, and A. Chatzigeorgiou, "Investigating trade-offs between portability, performance and maintainability in exascale systems," in *46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA 20)*, 08 2020.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Assembling Integrated Electronics**

by

Zach Fredin

B.S., Case Western Reserve University (2007)
M.E.M, Case Western Reserve University (2008)

Submitted to the Program in Media Arts and Sciences
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

© Massachusetts Institute of Technology 2021. All rights reserved.

Author .....
Program in Media Arts and Sciences
August 20, 2021

Certified by .....
Prof. Neil Gershenfeld
Director, MIT Center for Bits and Atoms
Thesis Supervisor

Accepted by .....
Prof. Tod Machover
Academic Head, Program in Media Arts and Sciences



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Assembling Integrated Electronics**

by

Zach Fredin

Submitted to the Program in Media Arts and Sciences
on August 20, 2021, in partial fulfillment of the
requirements for the degree of
Master of Science

## Abstract

Modern high-performance computing (HPC) systems consist of static architectures built from monolithic components. Miniaturization driven by lithographic technology has pushed Moore's Law to its limit after more than half a century, to the point that new chips require multi-billion dollar investments and supercomputer systems are built on a decades-long planning horizon. At the same time, typical HPC workloads like physical simulation have inherent geometry which is not reflected in the compute architecture, leading to a broad range of issues from cache concurrency to programming difficulty. Beyond integrated circuits, adjacent problems exist in electronics generally; printed circuit board assemblies (PCBAs) are similarly static, and the production and recycling of these products is environmentally unsustainable and requires extensive infrastructure.

The solution is to modularize electronics and autonomously assemble 3-dimensional computing structures from asynchronous, reusable elements. Of course, this concept brings with it a host of new questions: how are the devices programmed, how is communication bandwidth conserved, how do the elements physically interact, and how are the structures fabricated and assembled?

This thesis provides insight on module design and assembly automation for 3-dimensional electronics through two distinct prototype iterations. Evaluation of these systems revealed the mechanical limitations of commercial connectors, so an alternative method called digital materials is described which merges electrical interconnect and physical substrate. This method discretizes substrates into the fundamental elements that make up interconnect systems: conductive and insulating parts which are properly arranged to route signals to asynchronous processing nodes. Along the way, a novel method for constraining motion in these discrete assembly systems using modular superelastic flexures is introduced, characterized, and used to rapidly fabricate several machines.

Thesis Supervisor: Prof. Neil Gershenfeld
Title: Director, MIT Center for Bits and Atoms



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## Acknowledgments

Thank you, CBA colleagues, past, present, and future. Alfonso, Camron, Chris, Jiri, Dave, Jake, Erik, Amira, Ben, Will, Prashant, Sam, Pranam, Sabrina, Jack, Patricia, Filippos, Justin, Alex, Miana, Eyal: You are the smartest, most capable people I have ever had the privilege of working with.

Thank you, Neil, for creating the CBA and letting us explore the limits of digital fabrication in a world-class shop.

Thank you, Joe and Santanu, for your support and feedback as I consolidated my work into this document.

Thank you, CBA staff: Joe, Kara, Tom, John, James, and Sherry. We could not do what we do without you.

Thank you, Magneteers: Salima, Dan, Natalie, Bob. Your work is so interesting, and your machines are so cool (insert liquid helium joke).

Thank you, MAS '21 cohort: we got through this together against sometimes overwhelming odds. If you did not get a hoodie, I have the extras from Nina under my desk.

Thank you, Haystackers: Lauren, James, Amanda, Jera, Paul, Tanya, Andrea, Matt, Adam, Ellen, and many more. You have created a wonderful place.

Thank you, CBA sponsors and granting agencies, for supporting our work and holding us to account.

Thank you, MAS staff, for putting up with late-night panicked emails and helping us navigate MIT's myriad systems and protocols.

Thank you, family: Becca, Mom, Bill, Daniel, and Peter. You mean the world to me, and I can't wait to spend more time together.

Thank you, Dad, for teaching me so much. I wish you could see what I have done.

Thank you, Danica, for joining me on this wild journey and supporting me every step along the way. I love you.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Assembling Integrated Electronics**

by

Zach Fredin

The following people served as readers for this thesis:

Professor Neil Gershenfeld .....
Director, Center for Bits and Atoms
Professor, Media Arts and Sciences
Massachusetts Institute of Technology

Professor Joseph A. Paradiso .....
Director, Responsive Environments Group
Alexander W Dreyfoos (1954) Professor, Media Arts and Sciences
Massachusetts Institute of Technology

Professor Santanu Chaudhuri .....
Director, Manufacturing Science and Engineering, Argonne National Laboratory
Professor, Civil, Materials, and Environmental Engineering
University of Illinois Chicago



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Contents

**1 Introduction** **21**

1.1 Chip Fabrication . . . 23

1.1.1 The Yield Problem . . . 24

1.1.2 Chiplets . . . 25

1.2 Electronics Assembly . . . 26

1.2.1 PCBs . . . 27

1.2.2 Picking and Placing . . . 28

1.2.3 Breadboards . . . 28

1.2.4 Environmental Footprint . . . 30

1.3 High-Performance Computing . . . 31

1.3.1 Architecture . . . 32

1.3.2 The Programming Disconnect . . . 33

1.4 Contribution . . . 35

**2 Discrete Integrated Circuit Electronics (DICE)** **37**

2.1 Prior Work . . . 38

2.1.1 Project Tinkertoy . . . 38

2.1.2 Neuron Simulators . . . 38

2.1.3 Electronic Digital Materials . . . 39

2.2 Performance Projections . . . 39

2.2.1 Model . . . 40

2.2.2 Communication Peripherals . . . 41

2.2.3 Physical Modeling . . . 43

2.2.4 The DEM Node . . . 44



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**3 Modules** **47**

3.1 DICE Architecture . . . 47

3.1.1 Processor Selection . . . 48

3.1.2 Electrical Design . . . 49

3.1.3 Firmware and Programming . . . 50

3.2 Tiny-DICE . . . 51

3.2.1 Design . . . 51

3.2.2 Fabrication . . . 52

3.2.3 Testing . . . 54

3.3 Meso-DICE . . . 57

3.3.1 Design . . . 57

3.3.2 Fabrication . . . 60

3.3.3 Testing . . . 61

**4 Assembly Systems** **63**

4.1 Cartesian Assembly . . . 63

4.1.1 Design . . . 64

4.1.2 Evaluation . . . 65

4.2 6-DOF Assembly . . . 67

4.2.1 End Effector . . . 68

4.2.2 Infrastructure . . . 70

**5 Electronic Glitter Lattices** **73**

5.1 Electronic Digital Materials . . . 74

5.2 Glitter . . . 76

5.2.1 Fabrication . . . 76

5.2.2 Assembly . . . 78

5.3 Interconnect . . . 80

5.4 Scaling . . . 81

**6 Modular Superelastic Flexures** **83**

6.1 Flexures . . . 84

6.2 Superelastic Materials . . . 85



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

6.3 Modularity . . . 87
6.3.1 Orthogonal Taper Pin Joints . . . 88
6.4 Fabrication . . . 90
6.4.1 Flexural Elements . . . 90
6.4.2 Supporting Structures . . . 93
6.4.3 Assembly . . . 95
6.5 Fatigue . . . 96
6.5.1 Testing . . . 97
6.5.2 Experiments . . . 98
6.5.3 Next Steps . . . 99

**7 Compliant Machines** **101**
7.1 Single-Axis Flexure Test Machine . . . 101
7.1.1 Description . . . 101
7.1.2 Analysis . . . 103
7.1.3 Evaluation . . . 105
7.2 3-RRR CPM . . . 107
7.2.1 Description . . . 107
7.2.2 Inverse Kinematics . . . 108
7.2.3 Control . . . 111
7.2.4 Evaluation . . . 112
7.2.5 Computer Vision System . . . 115
7.2.6 Ruling Diffraction Gratings . . . 119
7.3 MicroPanto . . . 120
7.3.1 Description . . . 120
7.3.2 Micro-Engraving . . . 122
7.3.3 Implications . . . 123

**8 Future Work** **125**
8.1 Glitter Fabrication . . . 125
8.2 Micro-DICE . . . 126
8.3 Super-DICE . . . 128
8.4 Flexural Mechanisms . . . 128


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# List of Figures

1-1 Moore's Law, which suggests periodic doubling of IC transistor counts. . . . . 21
1-2 Semiconductor process nodes vs introduction year, graphed on a semi-log scale [14]. . . . . 22
1-3 A typical breadboard in use, demonstrating the difficulty inherent in reproducing or repairing work prototyped using this method [59]. . . . . 30
1-4 Diagram of one ORNL Summit AC922 HPC node [10]. . . . . 33

3-1 Two DICE hardware iterations assembled on their respective build plates. . . 47
3-2 Microchip ATSAMD51J20 microcontrollers. . . . . 49
3-3 A simplified one-dimensional DICE network with four modules. All the devices share a common 3.3 V DC power bus, but only communicate with adjacent neighbors. Practical implementations have 4 or 6 neighbors per module, and may include passive struts to maintain lattice geometry. . . . . 50
3-4 Nine Tiny-DICE modules assembled into a tetrahedral computational lattice and perched atop the author's pinkie finger. . . . . 51
3-5 Tiny-DICE module renders with labels, top and bottom. . . . . 52
3-6 Tiny-DICE module schematic, showing ATSAMD51J20 microcontroller, bypass and regulation capacitors, LED with current-limiting resistor, programming pads, and interconnect. Microcontroller symbol includes labels grouping SERCOM peripherals. . . . . 53
3-7 Tiny-DICE module 6-layer PCB layout, showing front traces and pads (red), rear traces and pads (green), internal traces (magenta and yellow), silkscreen marks (magenta and cyan), net names (white), PCB outline (blue), blind vias (gold crossed), and thru vias (gold and white). Power and ground pours indicated by hash marks around perimeter. . . . . 54



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

3-8 Tiny-DICE manufacturing steps. . . . . 54
3-9 Tiny-DICE post-assembly X-ray examination. . . . . 55
3-10 Tiny-DICE programmer with Free-DAP ARM-CMSIS-compatible firmware loaded on a built-in ATSAMD21 microcontroller. . . . . 55
3-11 Testing tiny-DICE modules using a 5-step pi series expansion, thermally imaged after reaching equilibrium. Optical image to right shows 11-module lattice. . . . . 57
3-12 Nine Meso-DICE nodes and seven struts assembled onto an early version of the build plate. . . . . 58
3-13 Meso-DICE node and strut, exploded to show 3D printed alignment part, milled Delrin latch, and assembled PCB. . . . . 59
3-14 Meso-DICE node and strut 4-layer PCB layout showing front traces and pads (red), rear traces and pads (green), net names (white), PCB outline (yellow), and vias (gold and white). Power and ground pours indicated by hash marks around perimeter. . . . . 60
3-15 Meso-DICE fabrication process. . . . . 61
3-16 Meso-DICE programmer. . . . . 62

4-1 Two commercially available Cartesian machines with principle linear axes labeled using green arrows. . . . . 64
4-2 An overview of the Tiny-DICE assembly machine, showing green PCB build plate, white 3D printed end effector, planar part storage plate, and integrated programming station to the right. Photo courtesy of Jiri Zemanek. . . . . 64
4-3 Tiny-DICE assembly machine end effector detail. Photos courtesy of Jiri Zemanek. . . . . 65
4-4 Close-up detail of Tiny-DICE modules after several assembly cycles showing broken mezzanine connectors caused by assembly over-constraint and misalignment. Photos courtesy of Jiri Zemanek. . . . . 66
4-5 The two Universal Robots UR10 6-DOF arms used for Meso-DICE assembly. The strut placement arm is on the left, while the node placement arm is on the right and includes green arrows indicating the six principle rotary axes. . 67


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

4-6 CAD renders of Meso-DICE node end effector, showing cam and arm action as gripper opens and closes around a module. Note that the compliant inner working surface of the Delrin cam does not distort in the closed render; on the fabricated end effector, these beams bend and bottom out, increasing the rigidity of the gripped part after pick-up. . . . . 69
4-7 Video stills of the Meso-DICE automated assembly process, showing strut and node pickup, programming simulation, and placement. . . . . 71

5-1 Tiny-DICE renders at various levels of deconstruction. . . . . 74
5-2 Two prior CBA projects to build electronic circuits from digital materials. . . 75
5-3 Will Langford's electronic digital materials "stapler" [49]. . . . . 75
5-4 Electronic glitter element design and lattice structure. . . . . 77
5-5 Conductive glitter fabrication process using micro-wire EDM. . . . . 78
5-6 Conductive glitter magazine with integrated stapler head. . . . . 79
5-7 Automated electronic glitter lattice assembly using stapler-type plate dispenser. 80
5-8 Flexural interconnect glitter part used to join adjacent DICE nodes. . . . . 81

6-1 Examples of monolithic flexure construction techniques. . . . . 85
6-2 2-axis flexural motion stage by Shorya Awtar [17]. Used with permission. . . . 85
6-3 Stress-strain curve of a typical metal, annotated to show recoverable and non-recoverable displacement after load is removed. From [62], with added annotations in red. . . . . 86
6-4 Stress-strain curve of a superelastic alloy, showing recoverable strain beyond the typical yield point of a conventional metal [69]. . . . . 87
6-5 Examples of wood joinery which makes use of simple wedges to secure pieces together. . . . . 89
6-6 Orthogonal taper pin joints at three stages of assembly. . . . . 90
6-7 Dimensioned drawing of a modular superelastic flexure intended for orthogonal taper-pin installation [31]. . . . . 91
6-8 Twenty-four modular superelastic flexures machined in two batches. Note blue paint marks, which indicate the larger side of the tapered hole [31]. . . . 92
6-9 Hand-reaming a waterjet-cut aluminum frame to produce a taper suitable for orthogonal pinning. Note custom wire-EDMed handle on taper ream [31]. . . 94


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

6-10 Two "pinsetter" tools, used to quickly place and remove taper pins for securing modular superelastic flexures [31]. . . . . 96
6-11 Automated fatigue tester used to evaluate modular superelastic flexures [31]. . 97
6-12 SEM micrograph of modular superelastic flexure as-machined surface, showing pitting and re-solidified debris from the wire-EDM process [31]. . . . . 98

7-1 Actuator test machine diagram showing flexures, motor, idlers, belt, and anchor. Motion is indicated with arrows. . . . . 102
7-2 Annotated image of actuator test machine [31]. . . . . 102
7-3 Several 250 ms steps of the linear actuator as measured with the laser displacement sensor, showing substantial ringing [31]. . . . . 105
7-4 A histogram of the actuator displacement step size across the laser displacement sensor's 10 mm range, showing a narrow normal distribution centered at 31.8 um [31]. . . . . 106
7-5 3RRR CPM diagram showing flexures, motors, idlers, belts, and anchors. Motion is indicated with arrows. . . . . 108
7-6 Image of installed 3RRR CPM, shown with control circuitry, build plate, and grating tool [31]. . . . . 109
7-7 3RRR CPM kinematic stage detail views. [31]. . . . . 109
7-8 3RRR CPM kinematic diagram from [90]. Used with permission. . . . . 110
7-9 3RRR CPM control system images [31]. . . . . 112
7-10 Testing the 3RRR CPM with a pen [31]. . . . . 113
7-11 Testing the 3RRR CPM with a sharpened bolt [31]. . . . . 113
7-12 3RRR CPM stiffness characterization setup [31]. . . . . 114
7-13 3RRR CPM planar stiffness test results in direction P1 [31]. . . . . 115
7-14 3RRR CPM planar stiffness test results in direction P2 [31]. . . . . 116
7-15 3RRR CPM vertical stiffness test results [31]. . . . . 117
7-16 ArUcO marker laser-engraved on copper, as imaged and identified by the computer vision system [31]. . . . . 118
7-17 ArUcO calibration setup with linear stage and laser displacement sensor [31]. 118
7-18 ArUcO calibration plot comparing laser displacement sensor values to computer vision measurements [31]. . . . . 119


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

7-19 Ruling a primitive diffraction grating using the 3RRR CPM and a diamond tool [31]. . . . . 120
7-20 Ruled diffraction grating images and micrographs. [31]. . . . . 121
7-21 MicroPanto overview, showing Handibot CNC router on the right, control laptop at center, and engraving mechanism at left. Blue and orange rods are pulltruded CFRP tubes covered in cable loom to avoid splinters. . . . . 122
7-22 MicroPanto construction details [32]. . . . . 123
7-23 MicroPanto example engraving, using a design by Lauren Fensterstock [5] [32]. 123
7-24 Three pieces of dry rice micro-engraved and inked to reveal detail. Two show Andrea Dezsö's [27] Forest Beings, while the third shows a hand-drawn "HAYSTACK" sign. . . . . 124

8-1 Render of a single Micro-DICE node, including a custom 10 GFlop DEM ASIC and electronic glitter lattice substrate/interconnect system. . . . . 126
8-2 Render of 768 Micro-DICE nodes in a corner-connected cubic lattice, roughly equivalent in overall computational power to a single V100. . . . . 127


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# List of Tables

3.1 Tiny-DICE manufacturing run yields. . . . . 56

6.1 Nitinol cost as compared to other metals and alloys. . . . . 88
6.2 Fatigue testing flexures with various characteristics. . . . . 99

7.1 Parameters used in the 3-RRR CPM inverse kinematics model [31]. . . . . 110
7.2 Results from 3RRR CPM stage calibration tests [31]. . . . . 120

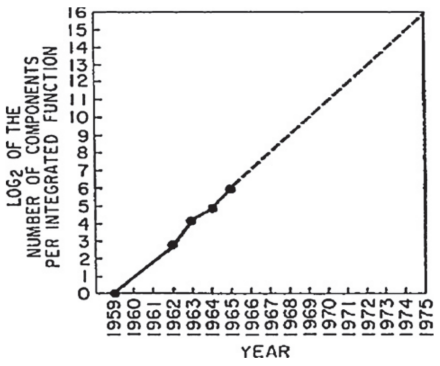


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

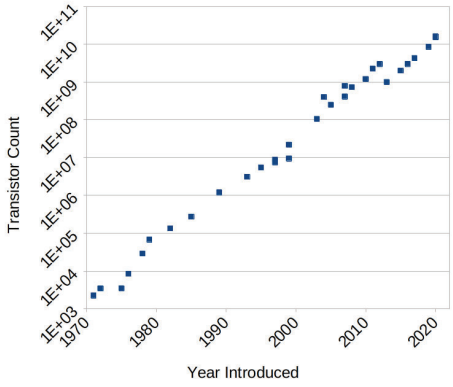
# Chapter 1

# Introduction

In 1965, only six years after the invention of the planar silicon transistor, Fairchild Semiconductor co-founder and future Intel CEO Gordon Moore plotted the number of circuit elements in four commercial integrated circuits (ICs) versus their release dates on a semi-log scale [58]. The resulting graph was remarkably linear, suggesting that the circuit elements in a given IC would double each year. Moore extrapolated his graph out another decade, which proved accurate if a bit modest; what later became Moore's Law held for another half-century with biennial doubling [54], as shown in Figure 1-1.




(a) Gordon Moore's original decade-long extrapolation of chip complexity [58].




(b) Moore's Law through 2021, adapted from [72].

Figure 1-1: Moore's Law, which suggests periodic doubling of IC transistor counts.

A good deal of ink has been spilled examining the push and pull of Moore's Law. Does the predictable trajectory of the last 60 years suggest that technological progress is inherently exponential? Have the applications for integrated circuits only required transistor



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

density to double on a biennial schedule? Is Moore's Law now the motivating force behind a trillion-dollar industry, as suggested in Intel's latest quarterly earnings release ("Inspired by Moore's Law..." [43])? One trend is clear: the capital costs for building new semiconductor fabrication facilities is growing at an unsustainable rate, with 3 nm fabs costing upwards of $20 billion USD in 2017 [28], as compared to a 90 nm fab costing $2.5 billion USD in 2007 [76]. This represents a compound annual growth rate (CAGR) of 23%.

In order to increase the number of transistors on a single chip from thousands in the 1970s to millions in the 1990s to billions in the 2010s, manufacturers could have taken one of three paths: increase the overall chip area; reduce the size of each transistor; or start building chips upwards into 3-dimensional space. While modern chips are larger than their early ancestors, and some designs use limited 3D features, Moore's Law is mostly the result of the ongoing shrinking of transistors, as seen in Figure 1-2.

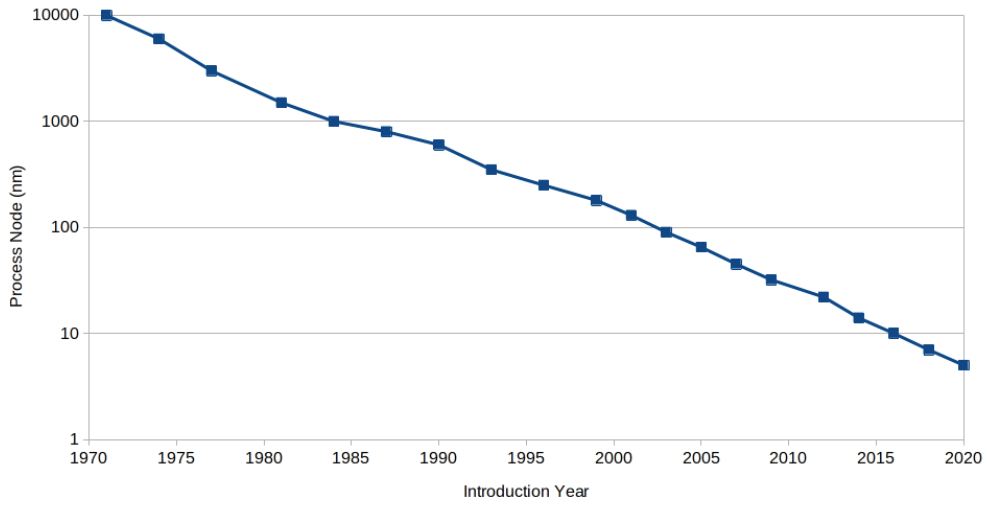



Figure 1-2: Semiconductor process nodes vs introduction year, graphed on a semi-log scale [14].

Transistor miniaturization is quickly reaching its practical limits. Smaller lithographic techniques require shorter wavelengths of light, which means even exotic extreme ultraviolet (EUV) sources [19] are nearing their resolution limit. Structures that may only be a few dozen atoms across are hard to make without defects at scale, one reason for the recent multi-year delays bringing new fabs online[43] [15]. In order to continue the half-century march towards ever-increasing computational density, a novel approach is needed. The work


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

presented here is an early step in this direction, and focuses on reversibly assembling true 3-dimensional computational structures, where the incremental cost to adding layers increases linearly as a simple function of material cost. These structures can then serve a wide variety of uses, from short-run desktop prototyping to large scale physics simulation.

The remainder of this introductory chapter briefly surveys the modern electronics and computing landscape with an eye towards scaling, cost-control, and environmental stewardship, and ends with a summary of this thesis' contribution. Chapter 2 introduces Discrete Integrated Circuit Electronics (DICE), an implementation of reconfigurable asynchronous 3D computing. Chapters 3, 4, and 5 focus on physical DICE module design and assembly, while Chapters 6 and 7 explore novel methods for constraining motion in DICE assembly systems. Chapter 8 concludes, and provides a roadmap for future research.

## 1.1 Chip Fabrication

Modern semiconductor fabrication is a complex, multi-step lithographic process that takes place in highly controlled clean rooms. The majority of devices are silicon-based, although other semiconductor materials such as gallium arsenide (GaAs) or gallium nitride (GaN) are used for specialized applications such as light-emitting diodes (LEDs). In order to function as a predictable substrate with stable electrical properties, the material must first be grown into a perfectly ordered single crystal called a boule, and then sawed into wafers which are polished flat. Typically, the boules are grown from a pure melt using the Czochralski method [86] or progressively solidified from a polycrystalline precursor in the Bridgman-Stockbarger method [61]. During the growth process, minute impurities such as boron are added to the melt to dope the silicon.

Once wafers are produced and polished, a variety of steps are performed to build up electronic structures. These steps can be additive, where new material is added to the wafer through deposition or oxidation; subtractive, where material is selectively removed using an etchant or mechanical method; electrical, where the substrate's electrical properties are changed using ion implantation to locally dope the material; or process-related, such as patterning a photoresist using UV light to mask a subsequent etching step. Between many of these steps the wafers must be carefully washed with ultra-pure water, requiring modern fabs to use 2-4 million gallons per day of the resource [24]. While the number of steps and


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

layers varies dramatically with chip complexity, a typical application processor may have 10 distinct layers and take 300 steps to fabricate over the course of several months.

After the final nano-scale features are etched, implanted, or oxidized into existence, the wafer consists of many identical rectangular devices on a single round substrate. Using a wafer probing apparatus, each device is electrically tested and the results recorded. Device yields vary dramatically depending on process node, device complexity, test stringency, and fab experience; one manufacturer reported 5 nm SRAM test chips being produced at 80% yield [93], but extrapolating these values suggest device yields of physically larger chips may be much lower [26]. In some cases, testing is used to characterize performance grades, allowing identically designed devices to be binned by quality and sold in tiers.

Finished and tested wafers are singulated into individual chips using a scribe, diamond saw, or laser. Individual dies are then packaged using a method dependent on the end user's requirements and technical capabilities. The largest integrated circuit packages were originally designed for through-hole electronics in which component leads are inserted through printed circuit boards (PCBs) and soldered. These ceramic or plastic packages, called DIPs (for Dual-Inline Pin), have leads at 2.54 mm pitch and are still used for prototyping with spring-terminal breadboards. Smaller surface-mount versions of DIPs, called SOICs (for Small Outline Integrated Circuit, 1.27 mm pitch) and SSOPs (for Shrink Small Outline Package, 0.65 mm pitch), are still seen on modern electronic devices. DFNs (Dual Flat No-Lead, 0.5 mm pitch) and QFNs (Quad Flat No-Lead, 0.5 mm pitch) reduce size by replacing the leads with flush tinned terminals, but still only electrically interface at the package perimeter. As processors have become more sophisticated, this dimensional interconnect limitation was solved by so-called grid arrays, where an entire face of the chip is covered in a regular pattern of through-hole pins (in a PGA, 2.54 mm pitch) or reflowable solder balls (in a BGA, 0.5 mm pitch). The Wafer-Level Chip-Scale Package (WLCSP, 0.4 mm pitch) extends this further by eliminating the ceramic or plastic package entirely, and applying the interconnect solder balls to the minimally-protected die directly.

### 1.1.1 The Yield Problem

Device yield is fundamental to fab economics. The Murphy model [60] begins by relating the overall yield of a monolithic device $Y_N$ to the individual yields of each component (i.e discrete transistors on a given chip) $Y_1$ and the number of components $N$:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

$$Y_N = Y_1^N$$

This equation alone quickly points engineers towards minimizing $N$ to maximize yield, so Murphy introduces several other factors that push optimal per-chip component counts higher. These factors include the difficulty in separating and handling extremely small devices made with only a handful of components, and the substantial expense of off-chip wiring that can be integrated with monolithic devices. In the 1960s, fabrication limitations meant the lowest per-component cost can be achieved with an $N$ value of 20 and a physical area of 30 mil square. Later yield models [22] combined both point defects and so-called parametric yield losses which affect entire wafers. And modern processors such as Apple's M1 have more than ten billion transistors [68], demonstrating half a century of steady fab yield improvements. Still, a clear negative correlation between device yield and overall device size exists.

One extreme exception is the fabrication of far larger chips that take up an entire wafer, as is the case with wafer-scale integration [44]) Cerebras' Wafer Scale Engine 1, with 1.2 trillion transistors and 400k cores, is 56 times larger than the largest GPU on the market. It is not known how Cerebras pushes the yield curve far enough to make such devices economical; one can speculate it is a combination of excellent process control, inherently fault tolerant design, and relaxed binning strategies. The Wafer Scale Engine benefits from fast on-chip interconnect and extremely local memory resources available to each core, particularly for some deep-learning applications. But even so, massive dies are delicate and awkward to package, and (like all chips) such devices are architecturally static after fabrication.

### 1.1.2 Chiplets

Chiplets represent another approach to increasing complexity by discretizing formerly monolithic devices into configurable 2-dimensional blocks [47]. These blocks are then permanently placed on a high-density interconnect called an interposer, which is another fabricated silicon chip that can be fabricated using lower-density design rules. Dividing large single devices into subsystems like this positively impacts yield, since for a given technology node the required defect-free area is smaller. Chiplets also allow physically closer integration of disparate devices, such as FPGAs and microprocessors, which is advantageous for high-bandwidth


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

communication [55].

Another important advantage of this strategy is that chip developers can focus on their core competencies, and use intellectual property (IP) licensing agreements to reuse chiplets from other vendors in their designs. Given standardized interfaces and specifications, one could imagine a marketplace of chiplets performing similar functions, where competitive pressures push costs down and encourage innovation. This type IP reuse and licensing is already prevalent in FPGA development, where vendors often provide developers with standard blocks to add common functionality such as USB or HDMI connectivity to a design [12]; chiplets merely extend the concept into the physical realm.

Chiplets, like System-on-Module (SoM) designs which perform similar functions but use conventional PCBs as interposers, are still architecturally static after assembly, meaning they cannot be reconfigured after packaging to fit novel problems. And chiplets are inherently 2-dimensional systems (technically up to 2.5-dimensional, when several chiplets are layered [87]), where power and data must flow from a single face, down to an interposer, laterally to another chiplet, and up to the device. Compared to a truly 3D system where all devices directly communicate with their nearest neighbors, this means chiplets must maintain communication bandwidth over a longer distance, while at the same time providing higher routing density since some elements may need to communicate with systems further away. These limitations strain interposer technology, requiring through-silicon vias and microwires that ultimately limit the complexity of systems [79].

## 1.2 Electronics Assembly

Once integrated circuits are designed, fabricated, and packaged, they must be further integrated into systems prior to use. Typically, ICs are soldered to PCBs along with myriad support components. For example, most chips require extremely stable power supplies, since even a brief sag in supply voltage can cause a fault or reset. Thus, most systems include dedicated power supplies that convert noisy external power (whether from a wall socket or a power bus) into one or more carefully regulated supply rails. At minimum, power supplies require a regulating device, such as a low-dropout regulator (LDO) or switching power controller, along with several passive components such as capacitors and inductors. Other ancillary equipment includes RF transceivers (Bluetooth, WiFi, cellular, etc) for commu-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

nicating with off-board networks; audio circuits for amplifying signals for headphones or speakers; external flash memory for storing configuration data or logs; user-operated controls such as switches and encoders; and electrical connectors for delivering power and data to the device.

### 1.2.1 PCBs

The vast majority of integrated electronic systems are built on PCB technology. PCBs serve as an intermediate interconnect that allows electrical designers to use monolithic integrated circuits in flexible applications without needing access to semiconductor fabrication facilities. Physically, a PCB is a stack of planar conductors and insulators, glued together into a flat rigid or flexible sheet. The top conductive layer is coated in gold or tin, allowing components to be easily soldered to the surface. Portions of the top layer along with additional conductive layers are selectively etched away during fabrication, leaving isolated conductive nets which allow the designer to electrically connect components at will. The nets are joined in the vertical direction using vias, which are physical holes through the insulating substrate that are plated during processing to allow current to travel between layers.

With these essential characteristics in mind, actual PCBs vary drastically in cost, complexity, and construction. At the lowest end, single-sided PCBs are made from 1.6 mm rigid phenolic FR1 sheet with a layer of 35 um copper foil glued to one side. Nets are milled into the copper sheet using 1/32" and 1/64" end mills on a desktop router. Rather than plating the copper, hand-soldering techniques are used along with flux to secure components to the PCB. This method, while primitive, is readily accessible with minimal equipment and is ideal for rapid circuit prototyping. Commercially, most circuit boards use fiberglass FR4 sheet as a substrate, and the patterns are etched rather than milled into the copper layer using photolithography techniques followed by acid etching. Such boards may have net traces as narrow as 75 um, and as many as 64 distinct insulating and conducting layers (although the vast majority are less than 10 layers).

The economics of modern PCB fabrication technology are interwoven with integrated circuit design. As discussed earlier, increased processor complexity has pushed chip designers to increase the density of off-board electrical interconnect, from DIPs (2.54 mm spacing, half perimeter) to QFNs (0.5 mm spacing, full perimeter) to WLCSPs (0.4 mm spacing, 2D array). This means there must be more signal wires reaching each chip across the 2D


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

surface of a given PCB layer. Increasing trace density by reducing trace width requires higher precision lithography machines and better etchant control. Similarly, smaller vias require tiny drill bits which operate slowly and are subject to breakage. Each additional layer adds an opportunity for failure caused by misalignment, contamination, or handling damage. All of these factors result in electronic assembly cost and complexity tracking increasing chip transistor count.

### 1.2.2 Picking and Placing

Once PCBs are prepared, components must be affixed to one or both sides. This is done by soldering, in which minute quantities of a conductive alloy with a low melting temperature permanently joints parts to pads. Historically, and in prototyping, this is done by hand with a soldering iron and flux-cored solder wire. In a production environment, solder paste and reflow techniques are used instead.

Solder paste is a spreadable mixture of minute solder balls mixed with gel flux. When the circuit board is fabricated, a matching stencil is also produced which includes apertures over the pads where components will be placed. The stencil is lined up on the PCB and solder paste is forced into the holes, leaving a pattern of flat paste-covered areas after the stencil is removed. Next, an automated machine called a pick-and-place pneumatically lifts components off of reels and precisely places them on the PCB, using an integrated vision system to correct alignment as needed. Once all of the parts are placed, the board is carefully transported to a reflow oven where the ambient temperature is slowly raised following a programmed profile. As the temperature increases, the flux boils off, cleaning any residual contamination off the pads and components. Next, the solder balls melt and coalesce, forming robust connections between the components and the board. The reflow oven holds the peak temperature for a time to ensure complete solder melting before gradually cooling to room temperature. At this point, the PCBA (for -Assembly) can be cleaned, tested, programmed, and further packaged into a complete system.

### 1.2.3 Breadboards

An interesting corollary to the ongoing miniaturization of IC packages is that fast reversible prototyping methods have not kept up. When DIP ICs and other through-hole components reigned, platforms called breadboards were commonly seen on workbenches for quickly test-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

ing circuit designs. Breadboards thick plastic plates with a grid of holes sized to accept DIP IC pins or passive component leads. The holes are spaced identically to ICs (2.54 mm) and have metal spring retainers at their base, such that items inserted into the breadboard are firmly retained but removable with additional force. Each spring retainer spans several holes, effectively creating a regular array of static nets which can be used to electrically connect components. Most breadboarding kits include a quantity of pre-stripped and bent wires of various lengths, so that users can quickly connect non-adjacent components.

Of course, anyone who has recently toured an electrical engineering department, a FabLab, or a widget company workspace has probably spotted plenty of breadboards hard at work testing circuit designs. While modern ICs are one-fifth the size of their DIP ancestors, breadboards have survived through the proliferation of breakout boards, adapter PCBs that convert tiny modern components into pseudo-DIP devices with breadboard-compatible 2.54 mm pin spacing. Some good has come from this ongoing trend; breadboards are cheap, easy to use and understand, and they enable backwards prototyping compatibility with half a century worth of electronic parts. Breakout boards are often open-source with robust community support, one key to the explosive growth of the electronics-focused parts of the maker movement.

The main reason breadboards are still in common use is that they are still the best, and only, method for reversibly prototyping electronic circuits. While some of this is due to the inertia of 2.54 mm pin spacing, the switch from through-hole to surface-mount components is mostly to blame. Surface mount devices have minimal (or no) leads, and as compared to standard DIP spacing the lead pitches vary dramatically between components. Even a given package configuration, such as QFP, is available in many pitches (0.8 mm, 0.5 mm, occasionally 0.4 mm) which would need to be accommodated with mechanical adapters. And surface mount devices are really designed for automated pick-and-place production, so they are often difficult to handle.

Unfortunately, breadboards have a few problems. First, as discussed above, they aren't actually sized for modern electronic components, limiting their ability to fully utilize modern processors. For example, the popular Teensy 4.0 prototyping breakout board uses the NXP MIMXRT1062DVL6A Arm Cortex M7 chip, which is packaged in a 10 x 10 mm BGA with 196 balls, 127 of which can be used as general-purpose input-output (GPIO) lines [77]. However, the breakout board only provides breadboard access to 24 GPIO lines,


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

despite taking up over five times the area of the core processor. Second, the long unshielded connectors which bus signals are unsuitable for high-speed electronics; while the exact limit depends on many factors such as wire length and adjacent signals, general design guidelines suggest breadboards shouldn't be used above 10 MHz. Such speeds are suitable for low-speed sensors and small displays, but fall short when routing closely coupled RAM chips or high-definition displays. Third, breadboards are rather delicate, particularly when connected to numerous off-board components; the plague of makers is the breadboarded circuit that functions properly on the workbench, but fails when transported to Maker Faire. And finally, breadboards hold a good deal of difficult-to-recover state, as seen in Figure 1-3. Even with a detailed photograph, it is hard to determine exactly how a breadboard is wired, so without careful documentation the results of a breadboarded test can be easily lost.

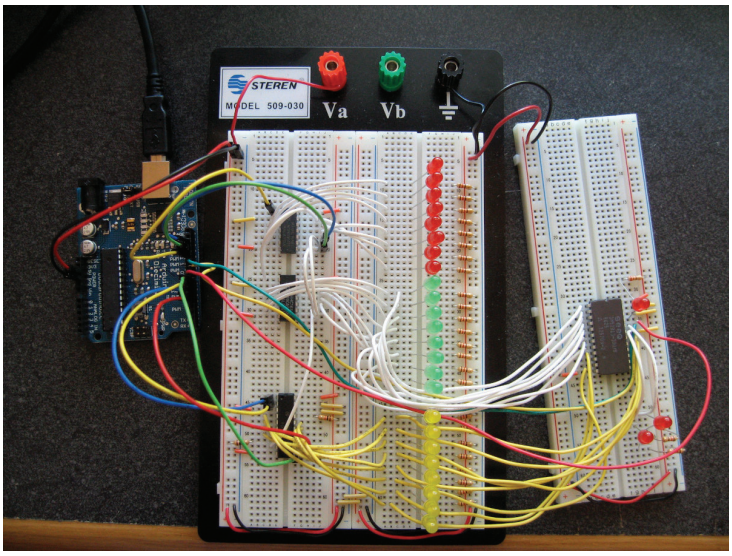



Figure 1-3: A typical breadboard in use, demonstrating the difficulty inherent in reproducing or repairing work prototyped using this method [59].

### 1.2.4 Environmental Footprint

A troubling side effect of PCB fabrication is water pollution. As discussed above, PCB production uses photolithography, which requires bare boards to be treated with various chemicals to mask and expose traces for subsequent etching. The etch process historically used ferric chloride; other compounds can also be used to selectively remove copper from the PCB substrate. After etching and washing, outer surfaces are plated, usually either with


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

tin or gold (which includes a nickel plating step to improve adhesion to the copper base). Boards are also coated in solder mask, which prevents solder from sticking to exposed traces.

Most of the aforementioned chemicals are aqueous, and in between many process steps (such as etching and plating) the PCBs must be thoroughly washed. Beyond PCB fabrication itself, the component placement process also uses some water to wash residual solder paste off assembled boards. While numbers from industry sources are difficult to find, various studies [39] have suggested that PCB fabrication uses on the order of 1000 L per square meter of board production. In other words, immersing a PCB in a column of water one meter tall shows roughly the amount of water polluted during the production cycle.

Industrial wastewater treatment is one clear answer to PCB production water use. Removing metals from wash water and neutralizing pH prior to discharge (or water reuse) is fairly trivial at industrial scale. However, since the infrastructure and raw materials needed for mitigation are far from free, an economic incentive exists for manufacturers to avoid water treatment. Such incentives can be counteracted by reasonable regulations, such as those outlined by the US EPA's Clean Water Act, but are dependent on local governmental oversight. Unfortunately, production globalization has far outpaced regulation, leading to the current reality where many PCBs are fabricated without proper consideration for environmental effects. A better solution is to find alternatives to aqueous processing entirely.

## 1.3 High-Performance Computing

Early computer systems used a mainframe and terminal model, where multiple users shared time on a centralized platform. Such architecture was borne of necessity; logic elements (whether they be relays, vacuum tubes, discrete transistors, or ICs) and memory (magnetic cores, punch cards, etc) were simply too large, inefficient, and expensive to provide dedicated systems for each user. As Moore's Law continued, minicomputers from DEC [21] and others enabled small workgroups to share systems, and personal computers (PCs) eventually gave us the 1:1 (or greater) computer:user ratio we enjoy today. A fundamental reason for this shift is that economical computational power now exceeds most user's needs, making it fiscally reasonable for a typical desktop workstation to rarely reach its performance limit (and, in most cases, sit idle when users are not physically present).

High-performance computing (HPC) refers to systems designed around a subset of prob-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

lems which continue to benefit from increased computational power. While problem types vary dramatically, many are based around modeling the real world. For example, in one study of the ORNL Kraken workload over one year [94], the vast majority of users came from academic disciplines focused on the physical sciences. More than 3/4 of the studied compute jobs were from atmospheric science, molecular bioscience, chemistry, materials research, or physics groups. While the specific problems likely varied dramatically in substance, it is reasonable to assume that many users were simulating real-world phenomena to predict its effects and compare them to observation.

Modern HPC systems include national-scale supercomputers, which are conceptually similar to last century's mainframes. Physically, a large HPC system may take up one or more floors of a building, weigh tens or hundreds of tons, cost many millions of dollars, and take a decade to develop and deploy [56]. Users share time on the system, which is often set up to accommodate several problems simultaneously; and unlike PCs, the system runs constantly, with controls and scheduling systems designed to maximize uptime.

### 1.3.1 Architecture

One of the key decisions during the HPC design process is how processors and memory are connected to one another. Oak Ridge National Labs' Summit [10] is a recent example of an HPC system that demonstrates the trade-offs system designers must make to accommodate memory and interconnect bandwidth and core processor speed.

Summit consists of roughly 4,600 nearly identical nodes housed in a room-filling grid of 19" server server cabinets. The nodes, officially IBM Power System AC922s [3], are each capable of roughly 40 TF (TeraFlops, or billion floating point operations per second). Nodes are physically distinct 30 kg devices, with redundant 2200-watt power supplies, water and air cooling systems, and physical enclosures allowing them to be removed and serviced or replaced. In addition to power, each AC922 has rear-mounted hose ports to connect to facility-provided cooling water, along with network connections that support 23 GB/s bandwidth via a protocol called EDR InfiniBand. Cabinets include InfiniBand switches that aggregate their installed nodes; multiple cabinets are then linked using higher level Director Switches to form the complete computing cluster.

A pair of IBM POWER9 processors control each Summit node. These 22-core devices are interlinked with a 64 GB/s connection, and they each connect to the InfiniBand controller


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

using a dedicated PCIe Gen4 16 GB/s link. The processors have their own dedicated 256 GB DDR4 memory modules connected at 170 GB/s. Each processor is further connected to three NVIDIA V100 GPU accelerators via 50 GB/s NVLINK; each accelerator then has its own dedicated 16 GB HBM2 memory, connected at 900 GB/s. A node diagram is shown in Figure 1-4.

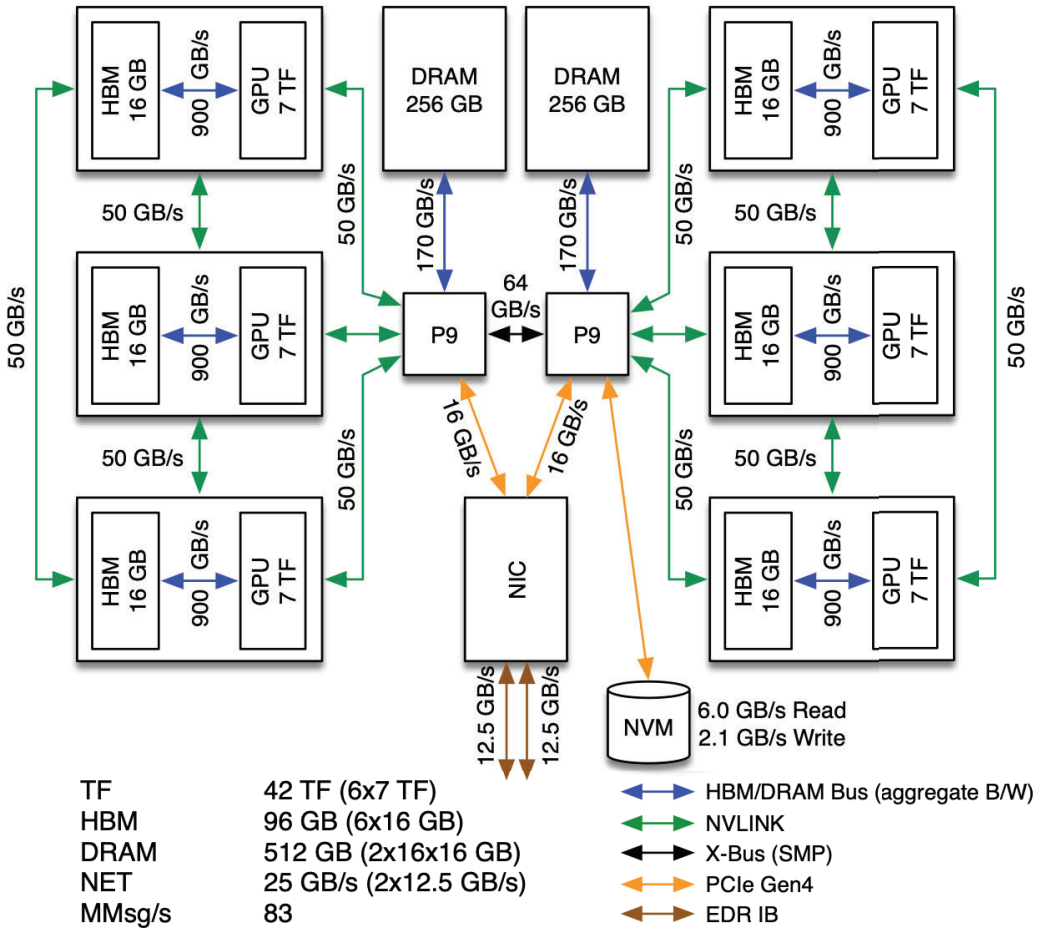



Figure 1-4: Diagram of one ORNL Summit AC922 HPC node [10].

### 1.3.2 The Programming Disconnect

HPC systems are architected hierarchical behemoths whose internal structure is inherently static. Importantly, while HPC users and system designers may attend the same conferences, they generally are not the same people; and given the decade-long planning horizon



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

for new HPC installations, it is unlikely that a given user's specific problem informed the overall system architecture. Of course, this disconnect is not unique to supercomputers; single-threaded desktop PCs are usually designed and programmed in different buildings (or companies) as well. And so programmers use compilers and interpreters so they can write abstracted code, like C or Python, and processors use common instruction sets so they can be readily programmed directly in assembly language without system design-specific knowledge.

For national-scale HPC systems like Summit with tens of thousands of cores, the analogs to abstract programming language are MPI and CUDA. MPI, for Message Passing Interface, is a communications protocol that defines how distributed nodes in a parallel computing system exchange information [6]. MPI is based around local memory that is not shared system-wide, and allows programmers to handle synchronization between compute threads in an abstract manner that is decoupled from physical processing node demarcation. CUDA, an NVIDIA application programming interface (API), extends this abstraction to GPUs, such as the six V100s in each Summit node [13]. Both MPI and CUDA are practically used in conjunction with familiar languages like C++. Unfortunately, there is not an "assembly language" for HPC systems. While users could theoretically write machine code for individual cores, the multiple layers of connectivity paired with a typical HPC system's workload of many simultaneous jobs from different users makes this impractical at any substantial scale.

Programming abstraction hides static architecture which inherently conflict with one of the most common HPC workloads, multiphysics simulation, because problem geometry is not reflected in compute geometry. Most physics is local; for example, the molecules that make up a fluid only interact with their neighbors. But the computational fluid dynamics (CFD) [89] model of a fluid system must keep track of all of the particles in a central synchronized database. Clearly, a more efficient model would allow nodes to only track particles to the range of their interaction distance. But this would require the computational geometry to align with the problem geometry, a virtual impossibility given the breadth of problem types, the disparate realms of HPC designers and users, and the fundamentally static nature of computational architecture.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 1.4 Contribution

This thesis presents a number of steps towards asynchronous reconfigurable 3D computing systems, outlined in the following chapter. The core work presented in Chapters 3-7 focuses on overcoming the physical challenges associated with reconfigurability, and lays the groundwork for future system iterations. Along the way, a number of novel concepts are explored, including an electronics prototyping method that goes beyond breadboards, and a modular system for building practical motion systems using flexural linkages rather than conventional sliding or rolling elements.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 2

# Discrete Integrated Circuit Electronics (DICE)

Monolithic processors present a fundamental limitation to computation due to the relationship between die size and fabrication yield. This observation is not novel; all high-performance computing systems spread tasks among a multitude of processing elements working simultaneously on different parts of a given problem. Building heterogeneous systems is physically advantageous, since packaging and interconnect can be designed around thermal management and fab capabilities. But such systems are inherently static, meaning their architecture does not reflect the geometry of a given problem set. Discrete Integrated Circuit Electronics [33] [50], or DICE, is a radical re-examination of how computation can reflect geometry by making reconfigurability a foundational part of system architecture.

A DICE system consists of many distinct but identical processing nodes arranged in a 3D lattice. The nodes provide structural support to one another, and share electrical connections that provide power and local communication channels. Like physical particles such as molecules interacting in the real world, DICE nodes are fully asynchronous, and interchange information through data tokens which are produced and consumed by adjacent devices. Crucially, DICE structures can be disassembled and reconfigured to suit new problem sets, and the physical assembly infrastructure is integral to the functionality of the overall system. External data can be sent serially to individual nodes to pass on, or in parallel to a lattice face using a suitably equipped build surface.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 2.1 Prior Work

While the combination of asynchronicity, reconfigurability, and three-dimensionality is unique to DICE, all of these concepts have been previously explored in the context of computing. A non-exhaustive summary of several projects that lead up to DICE is presented here.

### 2.1.1 Project Tinkertoy

Project Tinkertoy [45] was a project started in 1953 by the National Bureau of Standards to automate the production of electronic assemblies. The system consisted of many identical ceramic wafers that each carried a few passive components, such as resistors, capacitors, or inductors. The wafers had etched grooves that were filled with silver ink, which was then sintered into conductive nets analogous to modern PCB traces. After fabrication, wafers were assembled into stacks with vertical interconnecting wires, and often topped with a wafer equipped with a socket to receive a vacuum tube. While some final assembly steps required limited manual intervention, the Project Tinkertoy workflow was designed to be highly automated.

Remarkably, Project Tinkertoy came about before the widespread adoption of semiconductor logic devices such as transistors, and was not a modular computing system. The concept was instead motivated by the long development time for electronic assemblies; in particular, project supporters were concerned that it would take too long for the military to reactively develop new systems during wartime. Project Tinkertoy went through at least two hardware iterations during the late 1950s, including a follow-up effort with RCA in 1957 [46], but ultimately did not keep up with transistor- and IC-driven miniaturization trends.

### 2.1.2 Neuron Simulators

Neurons are biological cells that humans and animals use to process information. While neurons vary dramatically by function, they all have inputs, called dendrites, and an output at the end of a structure called the axon. Neurons maintain an electrical voltage potential across their membranes using ion channels and pumps. As a neuron's dendrites are excited by other neurons or sensory stimuli, the membrane potential increases according to the relative weighting of the signaling dendrite. If the membrane potential exceeds the neuron's action potential threshold, the neuron fires and sends a signal down its axon to other neurons


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

or output systems like muscles.

Neurons are naturally occurring asynchronous computing systems. Their independence was originally postulated by Santiago Ramón y Cajal near the end of the 19th century, who observed and drew neurons after staining them with Gogli's method [80]. Several decades later, Otto Schmitt built the first neuron simulator as part of his PhD work at Washington University [74]; as part of this work he also invented the now-ubiquitous Schmitt trigger, which unfortunately overshadowed his contribution to neuroscience. Much later, I started a company building electronic neuron simulators for educational purposes that included a number of sensory and motor devices so students could build complex asynchronous systems on their desks [65].

While neuron-like models called perceptrons [73] were used to further neuroscience research, their use as computational building blocks was never practical when compared to boolean transistor-based logic systems.

### 2.1.3 Electronic Digital Materials

More recently, Langford [49], Hiller [42], Ward [88], Popescu [67], MacCurdy [53], and others introduced the concept of electronic digital materials. These 3D lattices of conductive, insulating, and resistive elements are composed into functional electronic structures using a purpose-built assembly robot. Langford later proposed adding computation [50] nodes for robotic control, and prototyped early physical devices that later evolved into DICE. Notably, Langford's 2019 PhD thesis was the first document to use the term Discrete Integrated Circuit Electronics. As will become apparent in subsequent chapters, discretizing electronic components such as PCBs and connectors into insulating and conductive elements is beneficial beyond simply reducing part inventories to primitives based on physical properties.

## 2.2 Performance Projections

Asynchronous reconfigurable 3D computational structures like DICE have apparently stark advantages, such as the capacity to scale beyond physical clock phase limits, and clear issues, such as the added complexity and bandwidth loss of intra-node interconnect. These opposing characteristics can be quantified together to reveal optimal system configurations


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

for a given problem space. Specifically, where should the demarcation between inter- and intra-node exist? At one extreme, each node is a simple ALA [16] [36] device, performing single-bit operations on single-bit tokens. At the other extreme, the entire structure consists of a single massive internally synchronous node, analogous to a modern HPC system. It is informative to first consider a computing model in purely abstract terms, then bound it with a specific problem space, then add dimensionality to the configuration (i.e. 1D, 2D, and 3D lattices), and finally add real-world numbers based on currently available devices and near-term fab capabilities. From there, it is possible to project performance of a DICE structure on simple benchmark problems and compare it to current systems.

### 2.2.1 Model

Consider a computational structure build from asynchronous token-passing modules with a simple single-threaded internal processing architecture. On every clock cycle, each module can perform one of several tasks:

1. Perform a computation.
2. Store a value in local memory.
3. Retrieve a value from local memory.
4. Send a data token to a neighbor.
5. Receive a data token from a neighbor.
6. Do nothing.

To further simplify the model, one can assume memory retrieval is instantaneous; token-passing is bidirectional; and doing nothing is optimized out; thus:

1. Compute
2. Communicate

Given computation and communication cannot happen simultaneously, and only computation furthers completion of the problem (i.e communication is overhead), the computational efficiency $E$ of the system can be described as:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

$$E = \frac{T_{COMP}}{T_{COMP} + T_{COMM}}$$

where $T_{COMP}$ is time spent on computational tasks, and $T_{COMM}$ is time spent communicating with adjacent nodes.

It is well known that intra-chip electrical communication bandwidth exceeds inter-chip bandwidth by a substantial factor. This is due to several factors including physical proximity and lower transmission line capacitance. So it makes sense that internal computation can be clocked at a far higher speed than node-to-node communication. That means if the processor must send a bit with every computational cycle, communication time will dominate and the computational efficiency $E$ will plummet. The application being run on the processor is thus important to $E$; an example problem space, physical simulation, is analyzed later in detail.

### 2.2.2 Communication Peripherals

First, it is helpful to examine simpler ways to improve $E$ without changing communication speed or compute time, using a commercially available chip used for prototyping in later chapters as an example. The Microchip ATSAMD51J20A microcontroller has an ARM Cortex M4F core that runs at 120 MHz and can be reliably clocked upwards of 160 MHz, while its six built-in UART peripherals are only capable of 3 Mbps communication [1]. In other words, performing a single computational operation is always far faster than sending a single bit of data to an adjacent device.

In order to accommodate the difference in computation and communication speeds, the model can be iterated to include communication peripherals. The aforementioned ATSAMD51J20A microcontroller, along with most microcontrollers generally (and microprocessors, via off-chip devices), have a separate subsystem dedicated to managing intra-device communication. This peripheral mediates the speed differential between the faster core processor and the slower physical communication channel, allowing the processor to work on other tasks while the peripheral manages chatter with off-chip devices. In practice, this means the processor only needs to spend a single computational cycle to fill a local buffer with a data token. On the receiving end, the buffer independently receives data and flags the processor only when the buffer is full, at which point the processor spends another cycle to retrieve the data.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Again taking the ATSAMD51J20A as an example, assume a given problem takes 100 computational cycles to complete, and produces 10 bits of data that must be exported at the end. Without a communication peripheral, the computational portion would take:

$$T_{COMP} = \frac{1}{160MHz} \times 100cycles = 6.3 \times 10^{-7}s$$

while communication would take:

$$T_{COMM} = \frac{1}{3MHz} \times 10cycles = 3.3 \times 10^{-6}s$$

giving a computational efficiency of:

$$E = \frac{6.3 \times 10^{-7}s}{6.3 \times 10^{-7}s + 3.3 \times 10^{-6}s} = 16\%$$

Conversely, if the processor was able to offload the result to the communication peripheral at full speed (but still one bit at a time):

$$T_{COMM} = \frac{1}{160MHz} \times 10cycles = 6.3 \times 10^{-8}s$$

and thus:

$$E = \frac{6.3 \times 10^{-7}s}{6.3 \times 10^{-7}s + 6.3 \times 10^{-8}s} = 91\%$$

Clearly, this abstraction glosses over many details; the ATSAMD51J20A is a 32-bit microcontroller, for example, and there is some overhead involved in starting the communication peripheral and moving data to and from memory. It is also notable that if there were 100 bits of data to transmit and the calculation were performed repeatedly, the communication buffer would overflow since the peripheral wouldn't be able to keep up with the processor. This introduces an important inequality, which relates communication time $T_{COMM}$, processor speed $S_{COMP}$, communication peripheral speed $S_{COMM}$, and compute time $T_{COMP}$:

$$\frac{T_{COMM}}{T_{COMP}} <= \frac{S_{COMM}}{S_{COMP}}$$

As long as this inequality holds true, the processor will remain saturated and the com-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

munication peripheral will idle (or saturate, when the expressions are equal). Normalized to communication speed, devices should spend more time "doing" and less time "talking" to remain at peak efficiency. Put another way, as communication speed deviates from processor speed, each module should increase its computational workload with respect to its communication. Importantly, the physical electrical limitations of intra-module communication (as compared to inter-module) is analogous to the slower rate of a communication peripheral, even if a given DICE node is not so equipped. This fundamental notion can be used to scale DICE nodes to problem sets and vice versa, as will be shown later in this chapter.

### 2.2.3 Physical Modeling

Most classical physical modeling, be it finite element analysis applied to a structure or computational fluid dynamics applied to a turbine, is concerned only with local interactions. This is because the force laws that govern classical physics are also local; atoms push against one another in a crystal lattice, but they do not directly affect atoms beyond their immediate bonds. Fluids, likewise, consist of particles interacting locally and transmitting forces over distance indirectly. This observation, that physics is local and asynchronous, is central to DICE as applied in a simulation environment [78].

Consider a cube of computational nodes, each internally synchronous but externally asynchronous, used to simulate a 3-dimensional structure using the discrete element method (DEM). The nodes keep track of internal particles on a local clock, tracking velocities and interactions with each time step and subsequently adjusting velocity and position data as interactions occur. These interactions occur over some finite distance; when two particles are within such proximity of each other, the physics model must determine to what degree they affect each other. As particles travel to the edge of a given node, they eventually enter the interaction distance of the adjacent node. At this point, the two nodes must begin exchanging data to determine when the particle should eventually transition from one node to the next. Thus, the interaction distance forms a "shell" around each node; when particles occupy this space, they increase the communication time $T_{COMM}$ of the node.

Say each node keeps track of a space 10 units long on a side, and the interaction distance is 1 unit. The shell volume as a fraction of overall volume is thus $(10\times10\times2+10\times8\times2+8\times8\times2)/(10\times10\times10) = 48.8\%$ If the particles are randomly distributed, that means nearly half the particles contribute to $T_{COMM}$. As discussed above, depending on communication


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

peripheral speed this could be significant enough to bottleneck the processor and reduce its computational efficiency. But simply tripling the side length reduces this value by more than half, to $(30 \times 30 \times 2 + 30 \times 28 \times 2 + 28 \times 28 \times 2)/27000 = 18.7\%$. This classic surface-area-to-volume scaling continues to benefit larger nodes; a node with 100-unit length sides and 1-unit interaction distance, for example, has less than 6% of its as "shell".

For the ATSAMD51J20A used in early DICE prototypes, the communication peripheral runs at $3/160 = 1.9\%$ of the core processor speed. Based on a 1-unit interaction distance, each node would need to be upwards of 300 units per side to avoid bogging down communication. In tests, these devices had sufficient memory to keep track of roughly 1500 particles, or $5.5 \times 10^{-5}$ particles (on average) per volumetric unit. This suggests that such devices could be useful for 3D gas simulation. If one particle occupies each volumetric unit (if they were, say, simulating a solid cubic lattice of digital materials), each node would represent a volume with a side length of 11 (rounding down). In this case, the communication peripheral would bottleneck the processor unless it ran at 45% of the core processor speed, or 72 MHz. This is clearly out of the question with conventional UART communication systems, but in the realm of practicality for a device equipped with a low-voltage differential signaling (LVDS) peripheral.

### 2.2.4 The DEM Node

A better method for processor selection is to consider the ideal interaction distance for a particle system and then specify communication peripheral speed, core clock speed, and memory (which is needed to keep track of more particles) to size an ideal device. Since the work in subsequent chapters is focused primarily on physical DICE infrastructure and end-to-end workflow development, the ATSAMD51J20A is still ideal for prototyping as it is relatively low-cost, available in a compact package with six independent communication channels (for cubic tiling), and easy to program using ubiquitous open-source tools.

Looking forward, an application-specific integrated circuit (ASIC) for physical modeling using the DEM summarized above would look substantially different as compared to any commercially available chip. The UART peripherals from the ATSAMD51J20A would be changed for a system explicitly designed at the hardware level for asynchronous token-passing, so that these functions (producing and consuming tokens, checking for the presence of a token) would occur without processor intervention. UART is unclocked, instead de-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

pending on relatively matching clock domains across asynchronous components; this method relies on stable clock domains, so switching to a truly asynchronous single-wire method like Manchester encoding [48] would be beneficial.

Computationally, a dedicated DEM ASIC would balance memory resources with particle count, shell size, and the previously discussed relationship between compute and communication speed. For modeling solid materials, this likely means a dramatic increase in local memory as compared to the microcontroller-based nodes discussed above, since a close-packed lattice has a far greater particle density as compared to a gas. The DEM ASIC would also include a fast double-precision floating point unit (FPU) rather than the single-point FPU in the ATSAMD51J20A. Finally, the shell ratio (i.e, the ratio of the shell volume to the interior volume) and $T_{COMM}$ would be used to determine a $T_{COMP}$ sufficient to maximize $E$.

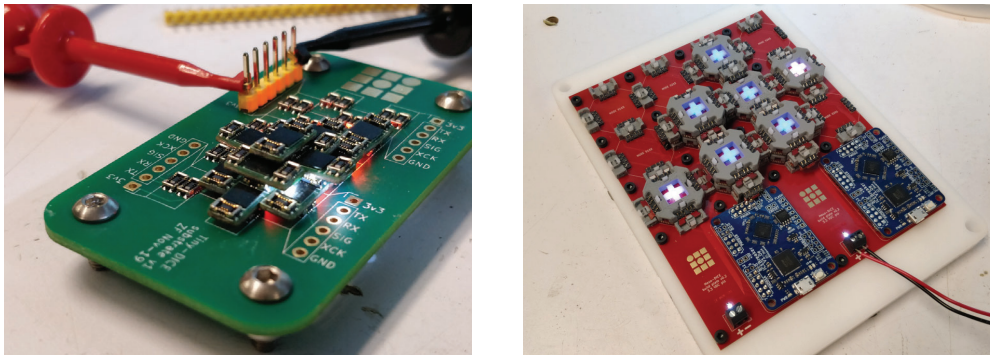
The physical packaging of a dedicated DEM ASIC would reflect the interconnect strategy discussed in Chapter 5. In particular, the bare chip would be designed for ready interfacing with glitter-scale electronic digital materials by increasing the relative size and thickness of the metallized interconnect pads. At the same time, the number of off-board electrical connections could be dramatically reduced, since the system would have no need for the broad applicability requirements that drive COTS microcontrollers to have high pin counts. The die would also include an on-board switching power controller designed for physical close-coupling to discrete inductors and capacitors, allowing the system to maintain full core voltage despite consecutive voltage drops across serially-connected discrete power rails.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 3

# Modules

In order to demonstrate the viability of reconfigurable computational systems, two distinct hardware platforms were designed and fabricated. The first generation, called Tiny-DICE, used the smallest available commercial-off-the-shelf (COTS) components and high-density interconnect (HDI) PCB fabrication technology to maximize computational density. The second generation, called Meso-DICE, was substantially larger and more mechanically reliable, and was intended to support the development of a full end-to-end DICE workflow, from design tools through applications.



(a) Tiny-DICE. (b) Meso-DICE.

Figure 3-1: Two DICE hardware iterations assembled on their respective build plates.

## 3.1 DICE Architecture

While the two physical DICE versions differ in some ways, such as mechanical configuration and neighbor count, they also share many characteristics which can be addressed together.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

This strategy was deliberate, since there was some overlap between development timelines of each design and sharing features (particularly the microcontroller) dramatically reduced firmware development overhead.

### 3.1.1 Processor Selection

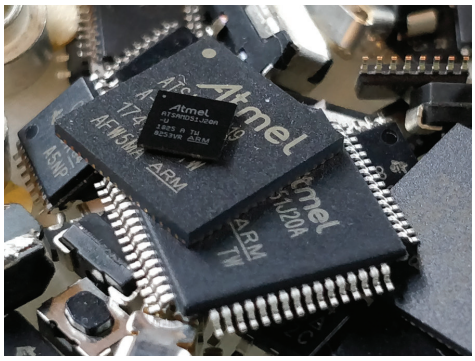
A multitude of axes exist to evaluate the suitability of a microcontroller for a project. Since the DICE platform is intended to be application-agnostic and performant, a fast 32-bit microcontroller with a large on-board memory bank and a hardware floating-point unit (FPU) is desirable. To minimize inter-node communication overhead, the device needs to have enough independent physical communication peripherals to manage data streams from all adjacent modules; in the case of a cubic lattice, this means six total channels (North, South, East, West, Up, and Down). Physically, the microcontroller needs to be available in a wafer-level chip-scale package (WLCSP), a method of packaging integrated circuits (ICs) that is minimally larger than the silicon chip itself. The device must be able to operate with minimal external components to reduce overall module size, including flash memory chips, power management ICs, and crystal oscillators. Finally, the part must be readily available from domestic commercial distributors in small quantities, and ideally should have well-documented support in the open-source hardware community to ease initial development.

A field-programmable gate array (FPGA) would provide the greatest flexibility since a custom-tailored soft processor core could be implemented based on application requirements. The requirement for dedicated communication peripherals could be satisfied in FPGA fabric; even better, custom protocols designed for token-passing could be used instead of standard serial standards like Universal Asynchronous Receive-Transmit (UART). However, FPGAs come with significant downsides, such as requirements for multiple staged power rails and external flash memory, both of which take up valuable PCB area. Furthermore, with some exceptions such as the Project IceStorm [91], most FPGA development environments are closed-source and difficult to extend to custom workflows.

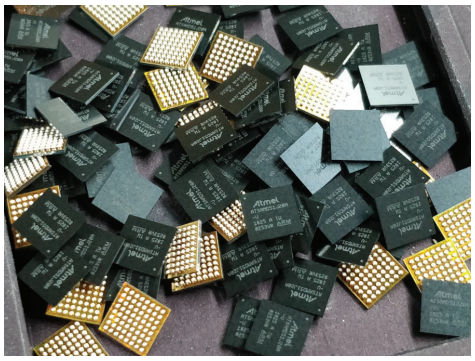
After considering several ARM Cortex M4F processors, such as the STM32F4 series and the Maxim MAX32660, the Microchip ATSAMD51J20 [1] was selected for DICE. This microcontroller has six independent serial communication channels, allowing for cubic lattice packing, while still being offered in a small 64-contact package in several form factors. The SAMD51 series forms the core of the Adafruit Feather M4 Express and Metro M4


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

products, open-source development platforms that are extensively documented and used by the electronics community, along with the networked dataflow machine controllers built by Jake Read [71], a CBA colleague.




(a) ATSAMD51J19s and -20s in WLCSP, QFN, and QFP packages from the pick-and-place "dropped part" bin.



(b) A pile of uninstalled 3x3 mm WLCSP ATSAMD51J20s randomly oriented to show 0.4 mm spaced solder ball contacts.

Figure 3-2: Microchip ATSAMD51J20 microcontrollers.

### 3.1.2 Electrical Design

DICE modules share common power rails, using parallel interconnect pathways and dedicated power and ground planes when possible to minimize node-to-node resistance (see Figure 3-3). This ultimately limits the distance between a given module and a power source; at some point, consecutive voltage drops will cause far-away modules to brown out. DICE build plates feed power into an entire face in parallel, so theoretically the length and width of a DICE structure could be infinite with a fixed maximum height. One method for increasing the maximum height of the computational structure would be to add local regulation to each module and supply a higher bus voltage; however, this adds significant complexity and cost to the design and reduces power efficiency, even if switch-mode buck regulators are used. Since computational efficiency (i.e W/Flops) is a key test for DICE, this strategy was not pursued.

In contrast to the common power rails, DICE modules only communicate with physically adjacent neighbors. To minimize overhead, the ATSAMD51J20's built-in SERCOM peripherals were used in UART mode for this task, so two communication wires for high-speed bulk data transfer connect the appropriate RX and TX port on each device. Two addi-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

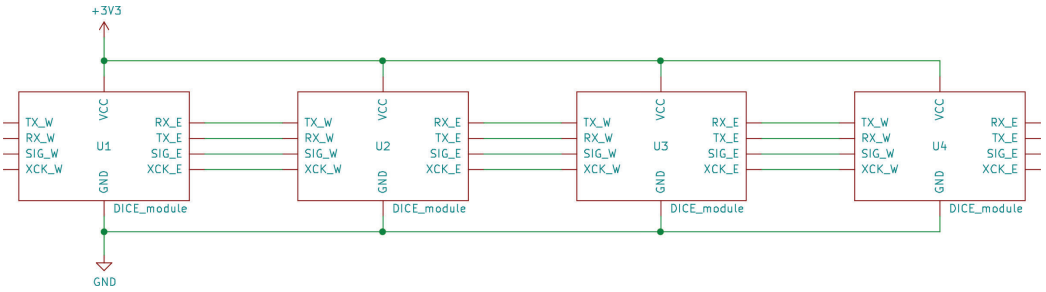



Figure 3-3: A simplified one-dimensional DICE network with four modules. All the devices share a common 3.3 V DC power bus, but only communicate with adjacent neighbors. Practical implementations have 4 or 6 neighbors per module, and may include passive struts to maintain lattice geometry.

tional general purpose input-output (GPIO) connections are included to provide flexibility for communication protocol development.

An important benefit of an intentionally asynchronous system is the lack of need for accurate local clocking. Because of this, the microcontroller support circuits lack the space-intensive crystal oscillator portion often seen on devices that need to manage timing-critical external communication protocols like USB. Thus, the only external components included beyond connectors are bypass and regulation capacitors and a status LED.

### 3.1.3 Firmware and Programming

DICE firmware is written in C and C++ and compiled using the factory-supported GNU Arm Embedded Toolchain. The build process uses a CMake framework developed by Erik Strand [78], a CBA colleague, which allows test and application code to be quickly ported between different DICE generations and other platforms.

The ATSAMD51J20 offers several options for loading, or flashing, firmware onto the device. The microcontroller includes a built-in USB peripheral, meaning the device can connect directly to the USB port on a computer and self-program if equipped with a suitable bootloader. However, this requires breaking out USB pins or adding a dedicated connector to each board in addition to the standard programming pins, which must still be accessible for initial bootloader installation. Instead, DICE uses the ARM SWD debugging protocol for firmware flashing and debugging, a clocked serial protocol that requires a total of five electrical connections to use (data, clock, reset, power, and ground).

Using SWD means inserting a dedicated programmer between the DICE module and the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

computer. Programmers are fairly cheap and ubiquitous, but they still represent another piece of hardware that can be misplaced or broken. An excellent alternative is to use the open-source Free-DAP project [82], a firmware image that turns a cheap Microchip ATSAMD21 microcontroller into a generic ARM-CMSIS programmer. For a few dollars and a simple milled PCB, one can then build the programmer into the physical programming rig, leaving the user to supply a common USB cable. In the same vein, one of several open-source command-line flashing utilities can be used to initiate the programming sequence, such as Open-OCD or edbg [81]. The latter, by the same author as Free-DAP, is used here.

## 3.2 Tiny-DICE

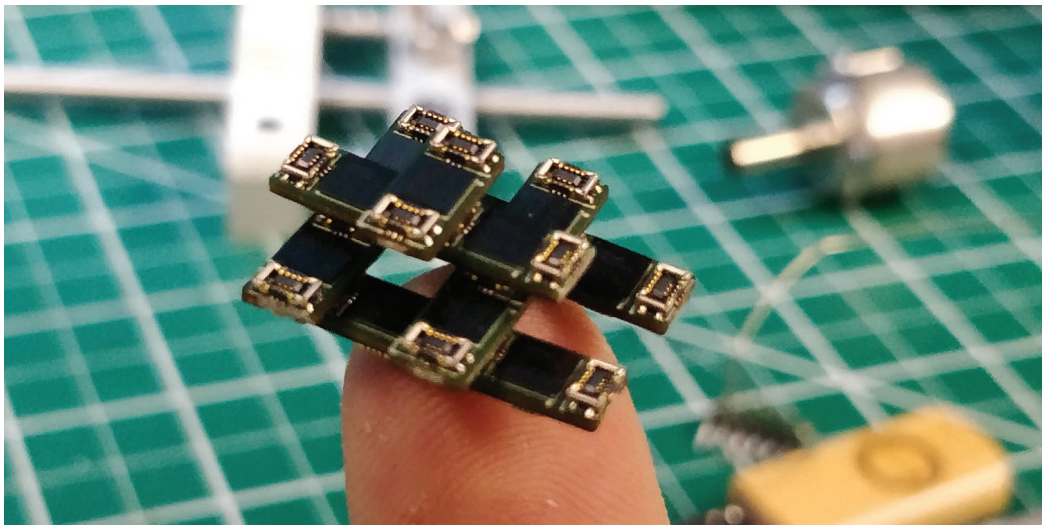


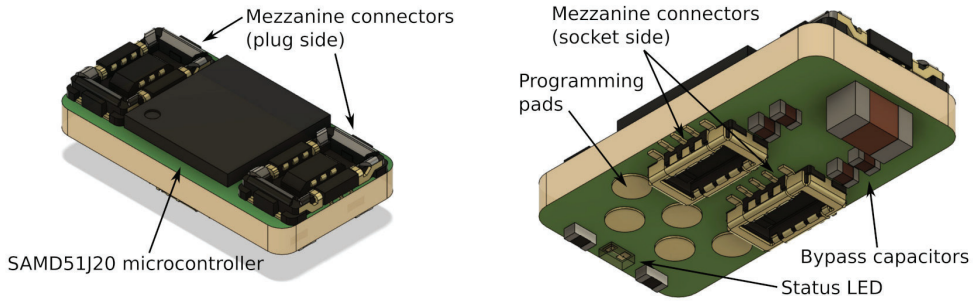
Figure 3-4: Nine Tiny-DICE modules assembled into a tetrahedral computational lattice and perched atop the author's pinkie finger.

### 3.2.1 Design

Tiny-DICE modules are designed around the smallest available COTS components. In addition to the aforementioned WLCSP ATSAMD51J20 microcontroller, each module includes four Molex SlimStack [7] mezzanine connectors which allow the devices to attach to a purpose-built base-plate and each other. The two bottom socket-style connectors are oriented side-by-side and centered, while the two top plug-style connectors are orthogonal to the bottom connectors and arranged at the edge of the PCB, shown in Figure 3-5. Overall,


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

each module measure 9 mm x 4.5 mm, the minimum dimensions required to accommodate the mezzanine connectors and microcontroller.




(a) Tiny-DICE module, top.

(b) Tiny-DICE module, bottom.

Figure 3-5: Tiny-DICE module renders with labels, top and bottom.

Beyond the microcontroller and connectors, the Tiny-DICE design includes numerous bypassing and power regulation capacitors as recommended by Microchip, along with a tiny 0201 status LED and current-limiting resistor (see schematic, Figure 3-6). The PCB was designed around 4 mil/4 mil space/trace rules for a 6-layer HDI PCB with 0.15 mm drills and 0.25 annular via rings (see layout, Figure 3-7). Due to the 0.4 mm pitch on the WLCSP microcontroller, numerous blind vias were required to fan out some inner balls which substantially increased the cost of the boards.

Each Tiny-DICE module directly connects to up to four neighbors in a flattened tetrahedral configuration. This configuration results in an overall lattice packing density of approximately 50%, and eliminated the need for a second part type to join adjacent modules. One shortcoming of this strategy is that devices at the edge of a given lattice are cantilevered well beyond the edge of the adjacent boards, resulting in a moment force during uniaxial assembly.

### 3.2.2 Fabrication

Bare Tiny-DICE PCBs were purchased from a commercial vendor; however, the modules were ordered as un-routed 2x8 panels as the vendor edge machining tolerance was greater than required for the project. Singulation was performed on a removable aluminum jig using a 1/16" carbide end mill on a Roland MDX-540 desktop CNC router. Since the same edge routing tolerance applied to the panel perimeter used to locate the boards on the jig's three


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

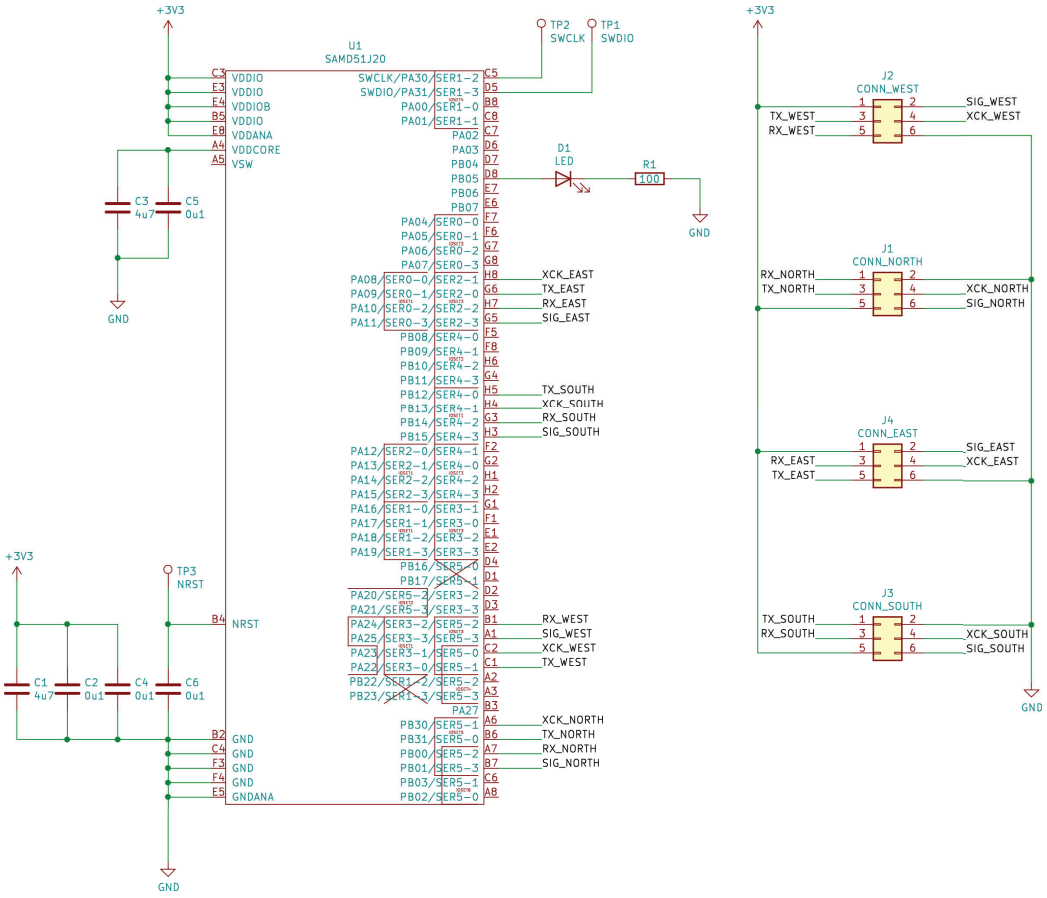



Figure 3-6: Tiny-DICE module schematic, showing ATSAMD51J20 microcontroller, bypass and regulation capacitors, LED with current-limiting resistor, programming pads, and interconnect. Microcontroller symbol includes labels grouping SERCOM peripherals.

ground locating pins, a panel was initially routed and examined under optical microscopy to adjust machining offsets for the singulation operation; it was then determined that these same offsets could apply to subsequent PCBs, so later milling operations yielded all 16 modules.

After edge routing, the panels were mounted on a milled phenolic fixture for solder paste application using a commercial press, and the back side (opposite the microcontroller) populated using a Mechatronika M10V pick-and-place machine. The panels were reflowed using a shop-modified reflow oven, cooled, and flipped so the front side could be similarly populated and soldered. A sample of these steps is shown in Figure 3-8.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

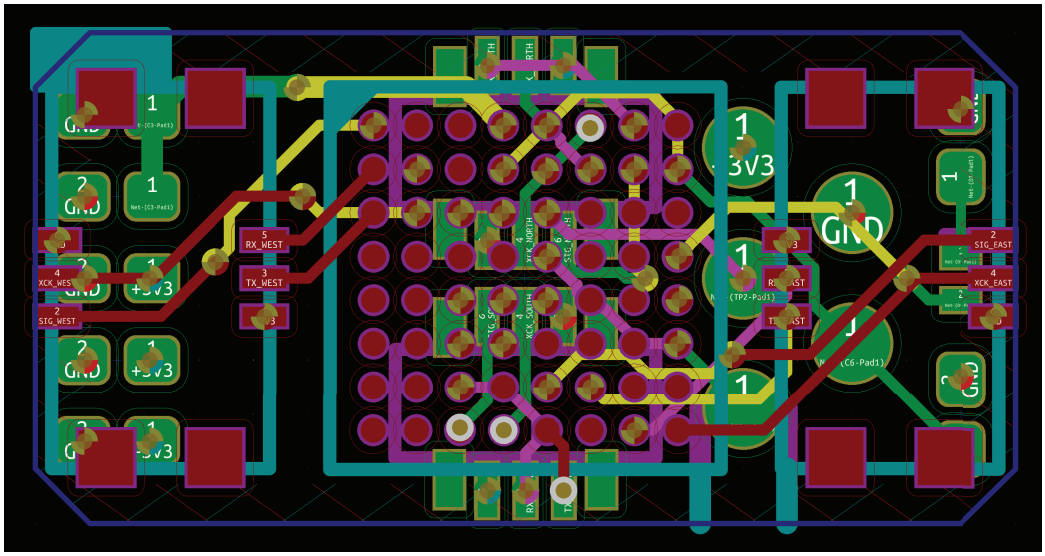
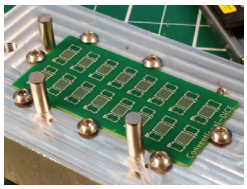
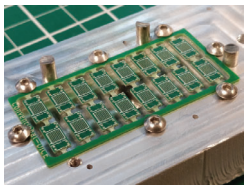


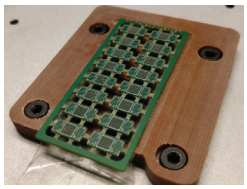

Figure 3-7: Tiny-DICE module 6-layer PCB layout, showing front traces and pads (red), rear traces and pads (green), internal traces (magenta and yellow), silkscreen marks (magenta and cyan), net names (white), PCB outline (blue), blind vias (gold crossed), and thru vias (gold and white). Power and ground pours indicated by hash marks around perimeter.



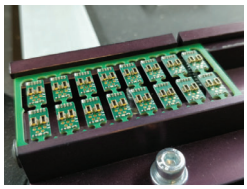
(a) Unrouted Tiny-DICE PCB panel, on singulation jig.



(b) PCB after singulation with Roland MDX-540.



(c) PCB after solder paste application, on paste jig.



(d) PCB after population of back side components.

Figure 3-8: Tiny-DICE manufacturing steps.

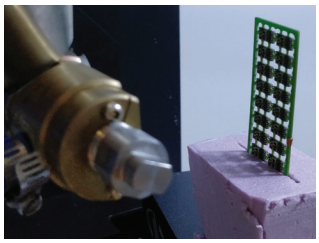
### 3.2.3 Testing

The majority of Tiny-DICE modules were assembled in one multi-day run, with each step being performed on multiple panels prior to moving on to the next. This saved a great deal of setup time but hid significant problems until after a large quantity of microcontrollers were used, an unfortunately wasteful practice that could have been avoided with a bit more patience.

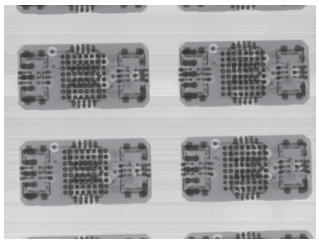
After the final reflow operation, one PCBA was visually examined on both sides using a Lynx Evo projection stereo microscope. Several modules appeared to have solder bridges between mezzanine connector pins; given the likelihood of similar but hidden defects under


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

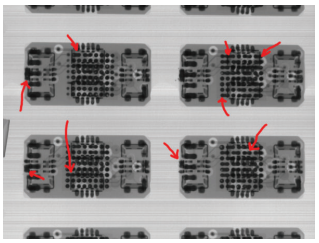
the microcontroller, all of the PCBAs were X-rayed using a Nikon Metrology XTH160 Micro CT scanner in single-image mode. Careful examination of the resulting micrographs, shown in Figure 3-9, showed frequent solder bridges, some of which were also in undetectable areas under the mezzanine connectors.



(a) X-ray setup, showing source in the foreground to the left. Note copper tape on PCB to identify board.



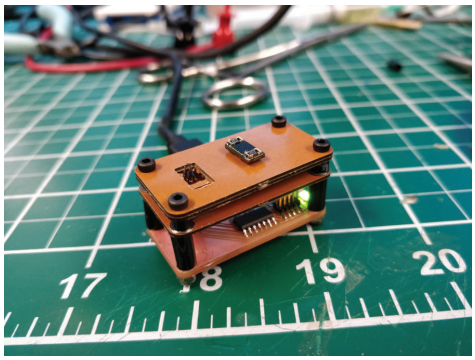
(b) X-ray micrograph of four good Tiny-DICE modules. Tabs are invisible due to lack of copper core.



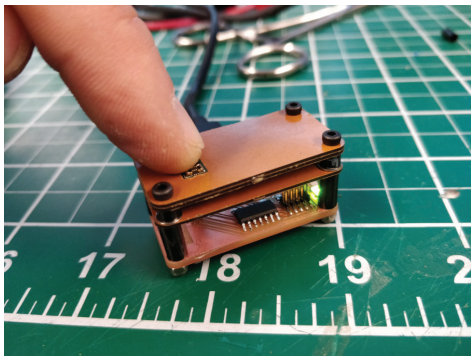
(c) X-ray micrograph of four faulty Tiny-DICE modules. Red arrows point to solder bridges.

Figure 3-9: Tiny-DICE post-assembly X-ray examination.

A custom programming jig, shown in Figure 3-10, was designed and fabricated using sheets of phenolic to constrain the pogo pins and provide a recess for a single module. The in-house-routed PCB which forms the programmer's base includes a ATSAMD21 loaded with Free-DAP firmware, a linear regulator, and a status LED.




(a) Tiny-DICE programmer with module.




(b) Programmer in use, testing a module.

Figure 3-10: Tiny-DICE programmer with Free-DAP ARM-CMSIS-compatible firmware loaded on a built-in ATSAMD21 microcontroller.

PCBAs that passed X-ray inspection were singulated by clipping the tabs holding the modules together and sanding the rough edges smooth with a small Proxxon disk sander. They were then flashed with a simple test program that blinked the on-board LED. Ulti-



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| Step | Input Qty | Output Qty | Yield | Comments |
|---|---|---|---|---|
| Singulate | 188 | 156 | 83.0% | fixture misalignment, test singulation cuts |
| Reflow | 156 | 58 | 37.2% | solder paste quality issues, stencil too thick |
| Rework | 58 | 15 | 25.9% | abandoned, remaining qty 43 pcs viable |
| Overall | 188 | 15 | 8.0% | yield up to 30.1% with finished rework |

Table 3.1: Tiny-DICE manufacturing run yields.

mately, of 188 modules attempted, only 15 functional modules were produced, for a final yield of 8.0%; an additional 43 modules required reasonably achievable but time-consuming rework that was not completed, so the best possible yield for the run would have been 30.1%. A breakdown of scrap sources can be seen in Table 3.1.

The vast majority of Tiny-DICE scrap was caused by unrecoverable solder bridges. Later tests showed that the Chip Quik SAC305 Thermally Stabilized paste, combined with a relatively thick stainless steel stencil, resulted in too much solder remaining on the PCB prior to component installation. Yields would likely improve dramatically with a finer sized, less viscous solder paste, such as Kester T5.

Tiny-DICE boards were loaded with a simple program [37] which uses a series expansion to calculate pi using a 5-step iterative loop:

$$\pi(N) = \sum_{i=1}^{N} \frac{0.5}{(i-0.75)(i-0.25)}$$

$N$, or the number of program iterations, was set to a large value such as 1,000,000, and the program was set to toggle a GPIO pin at the start and stop of the calculation. A digital storage oscilloscope was connected to the pin and set to trigger on a rising edge in single-shot mode, which allowed the total calculation time to be accurately calculated. Since each iteration involves exactly five floating point operations, calculating the processor speed in Flops was simply a matter of dividing 5,000,000 operations by the oscilloscope-recorded pin toggle delay. This yielded a value of 16.8 MFlops, roughly equivalent to an Intel Pentium Pro processor [35].

A number of Tiny-DICE nodes were programmed identically to iterate the pi calculation for a much longer period, so that they would run continuously as the lattice reached


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

thermal equilibrium. The onboard LED was also deactivated to reduce unnecessary current consumption. Eleven such nodes were assembled in a tetrahedral lattice and examined using a FLIR i5 infrared imaging system, as seen in Figure 3-11.

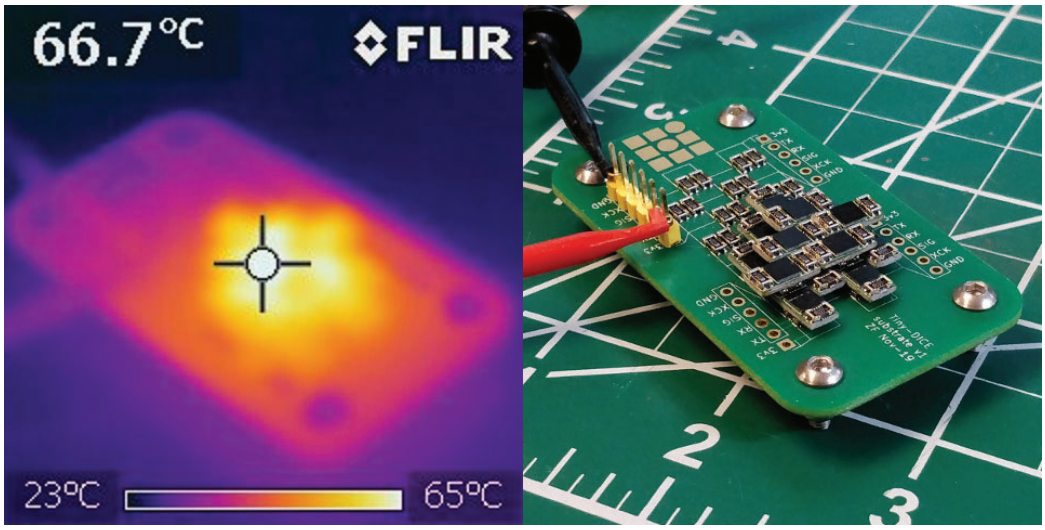



Figure 3-11: Testing tiny-DICE modules using a 5-step pi series expansion, thermally imaged after reaching equilibrium. Optical image to right shows 11-module lattice.

The thermal imaging revealed that the center of the lattice approached 70 C, near the ATSAMD51J20's 85 C operating limit. As such, larger lattice structures would likely benefit from increased sparsity or active cooling, either through forced air or fluid immersion. During the test, the lattice consumed 0.384 A at 3.28 V, both measured using an HP 34401A 5.5 digit bench multimeter. The structure thus required 1.26 W of power, which equates to 0.20 GFlops/W. This is on the order of an Intel i7-8700 at 0.34 GFlops/W [33], but well below GPUs like the NVIDIA V100 at 50.4 GFlops/W.

## 3.3 Meso-DICE

### 3.3.1 Design

In contrast to Tiny-DICE, the Meso-DICE iteration was less concerned with minimizing size as it was with maximizing utility, reliability, and manufacturing yield. As such, no physical parts were carried over between the two designs. Even though the processor remained the same, the packaging was changed to the Quad Flat No-Lead (QFN) version to lessen the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

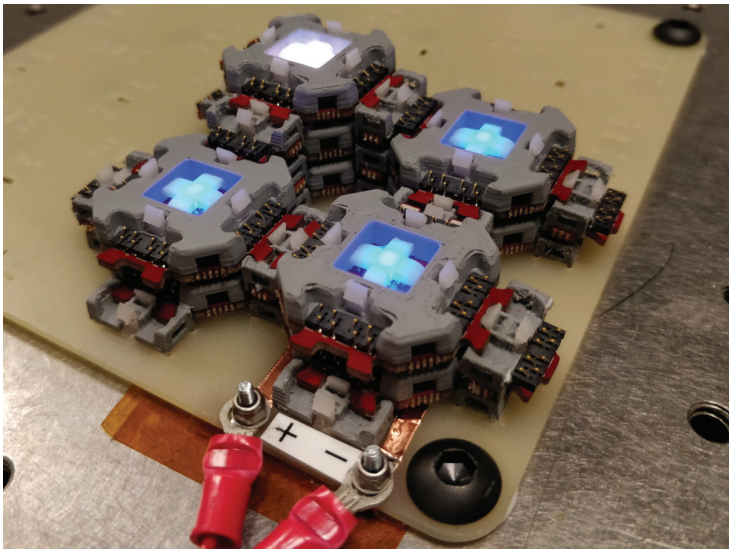



Figure 3-12: Nine Meso-DICE nodes and seven struts assembled onto an early version of the build plate.

chance for soldering defects and allow for more conservative PCB design rules.

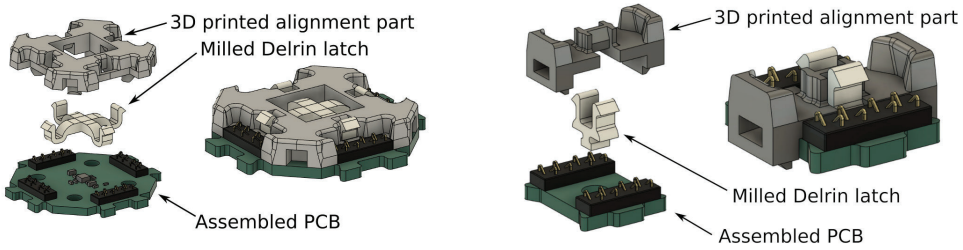
An important change for Meso-DICE is the addition of two neighbors to each node to form a cubic lattice. This configuration is advantageous for many reasons; in particular, a Cartesian layout is much simpler to map to a physical volume defined in X/Y/Z coordinates such as in a simulation environment. However, the flat planes that make up each layer of the lattice require modules to be joined edge-to-edge, which is difficult particularly if uniaxial assembly from the Z direction is used. One strategy is to use asymmetric modules that overlap edge-to-edge, but this requires electrical components at multiple heights which increases fabrication complexity and cost. Instead, the DICE modules were decomposed into active nodes and passive struts which join uniaxially and form a cubic lattice, as shown in Figure 3-12.

A number of integrated commercial connectors were considered for Meso-DICE; that is, connectors that both mechanically constrain and electrically interconnect two printed circuit boards. These types of connectors are generally designed without parallel insertion in mind, meaning they are not intended to be used in pairs or greater quantities on the same substrate. In some cases, part datasheets specifically call out this limitation, or specify that flexural elements should be added to resolve the mechanical overconstraint caused by the parallel configuration. While this limitation also exists for the mezzanine connectors used


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

in Tiny-DICE, the tetrahedral lattice configuration meant there were always at least six connection points between two connectors on the same board. Even when fully mated, these extra connectors provided sufficient flexibility to avoid putting too much stress on the solder joints or components. With Meso-DICE, vertically stacked nodes are separated by four connectors via struts; a brief test using a milled board showed that this configuration inadequately relieved the connector overconstraint and resulted in PCB and solder joint failure.

Instead of depending on integrated COTS connectors that manage mechanical alignment, latching, and electrical connectivity, these three requirements were separated into discrete parts, shown in Figure 3-13. The electrical interconnect uses surface mounted spring terminals that press against plated PCB pads. After PCBs are fabricated and populated, Meso-DICE struts and nodes are assembled with a 3D printed part which aligns and horizontally constrains mating parts, along with milled Delrin latches that vertically secure the parts together. The parts are secured using posts integrated into the 3D printed parts which are heat-staked to the PCBs, forming a tight and compact connection.




(a) A Meso-DICE node.

(b) A Meso-DICE strut.

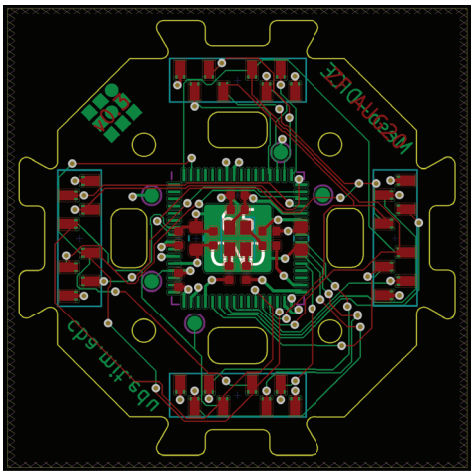
Figure 3-13: Meso-DICE node and strut, exploded to show 3D printed alignment part, milled Delrin latch, and assembled PCB.

Electrically, Meso-DICE nodes were nearly identical to Tiny-DICE but substituted an integrated red/green/blue (RGB) LED to increase local indication options. Each connector included eight electrical contacts: two for power and ground, four for horizontally adjacent nodes, and two for vertically adjacent nodes. In order to pass signals up or down, the struts simply pass these lines through; to form a complete connection, a pair of adjacent struts must be used to route all four data lines. The Meso-DICE PCB routing for the strut and node are shown in Figure 3-14. Both boards are 4-layer designs with internal power and ground pours; the struts required 4/4 design rules due to the extensive cutouts for Delrin

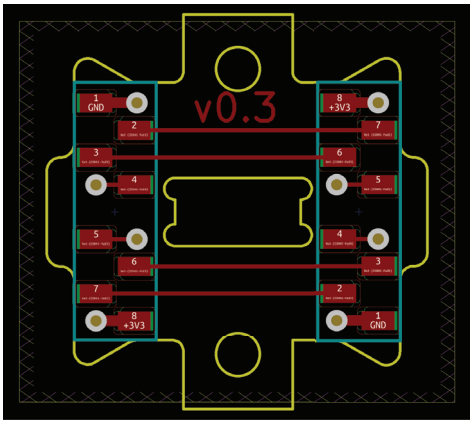


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

latch clearance.



(a) Meso-DICE node PCB layout.




(b) Meso-DICE strut PCB layout.

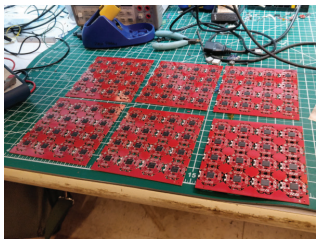
Figure 3-14: Meso-DICE node and strut 4-layer PCB layout showing front traces and pads (red), rear traces and pads (green), net names (white), PCB outline (yellow), and vias (gold and white). Power and ground pours indicated by hash marks around perimeter.

### 3.3.2 Fabrication

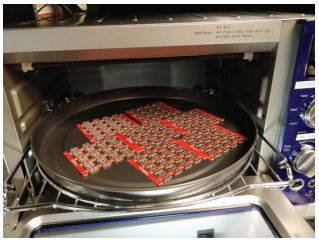
As with Tiny-DICE, bare Meso-DICE PCBs were purchased from a commercial vendor. The boards arrive tab-routed in 4x4 panels; since edge clearances were not critical, this work was left to the vendor. Due to the larger component position tolerances, a solder paste jig was not needed either, simplifying paste application setup. After paste application, the boards were populated using the Mechatronika M10V pick-and-place machine and reflowed as before. This process was repeated on the opposite side for nodes since they include parts on both sides. After cooling, the boards were visually examined for soldering defects and bridges were manually fixed.

Delrin latches were milled on a Roland MDX-540 desktop router from 2.5 mm sheet stock using a 1/32" carbide end mill. For nodes, the top and bottom pieces were snapped together prior to insertion into the PCB. The plastic alignment parts were 3D printed using Matter Hackers Tough PLA on a Prusa MK3S printer and manually cleaned after removal from the print bed. A spare extra-large chisel tip for a Weller WES51 soldering iron was milled into a concave hemisphere and used to heat stake the struts and nodes together at 260 C. This process worked sufficiently but required careful clamping to ensure no gaps formed between

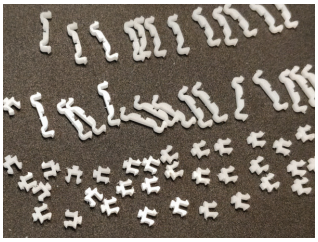

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.



(a) Examining 96 Meso-DICE node PCBs after soldering.

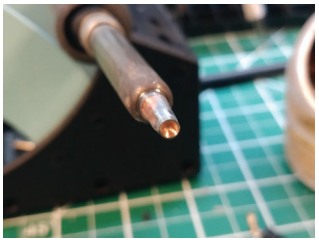


(b) Removing 112 Meso-DICE strut PCBs from the reflow oven.

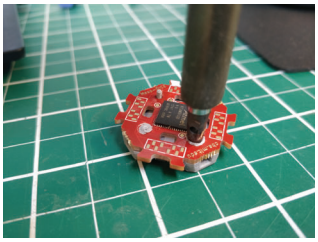


(c) A collection of machined Delrin clips for Meso-DICE struts and nodes.

(d) 3D printed alignment pieces before removing support material.



(e) Customized Weller WES51 iron tip for heat-staking 3D printed parts.



(f) Heat-staking a Meso-DICE 3D printed part onto a PCB.

Figure 3-15: Meso-DICE fabrication process.

the 3D printed parts and the PCBs.

### 3.3.3 Testing

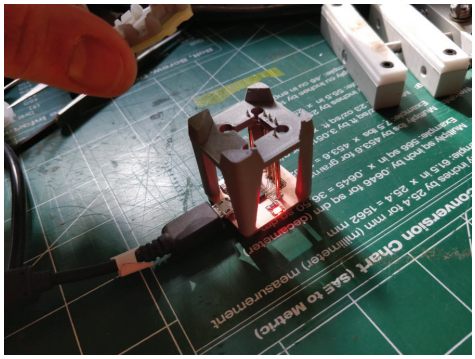
After reflow soldering, Meso-DICE PCBAs were visually inspected using a Lynx Evo stereo projection microscope. Solder bridges on the QFN microcontroller were somewhat common but quickly remedied using a bit of paste flux and a clean hot soldering iron tip. Yields were not tracked but were subjectively much higher than Tiny-DICE; occasionally a solder bridge was not recoverable, but nearly all boards with soldering defects were fixed. A notable exception occurred when the shop-modified reflow oven proved unable to adequately hold soak temperatures long enough to fully reflow the connector pads; after this, a commercial convection toaster oven was used manually to solder the remaining struts and nodes.

A programming jig, shown in Figure 3-16 was fabricated to quickly flash test firmware onto Meso-DICE nodes. The design is similar to the Tiny-DICE programmer in that it uses an integrated ATSAMD21 running Free-DAP. The top of the programmer has a 3D printed alignment jig that horizontally constrains the Meso-DICE nodes during flashing and keeps the pogo pins oriented correctly. Prior to final assembly with the Delrin latches and 3D

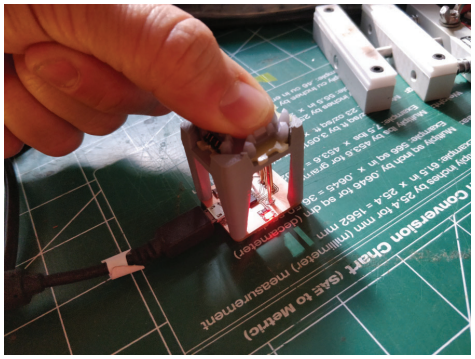


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

printed alignment jigs, all nodes were flashed with the a simple test program that flashed the onboard RGB LED white and blue.



(a) Programmer with 3D printed alignment jig, pogo pins, and Free-DAP PCB.



(b) Programmer in use with an early non-functional mock up node.

Figure 3-16: Meso-DICE programmer.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 4

# Assembly Systems

An important step in discrete modular computation is assembly. While hand-assembly is useful during testing, the goal of this project is to integrate automated assembly into the design workflow, just as compilation integrates into the traditional software development process. Building on the two module designs discussed previously (Tiny-DICE and Meso-DICE), this chapter presents two strategies for automated assembly of computational systems.

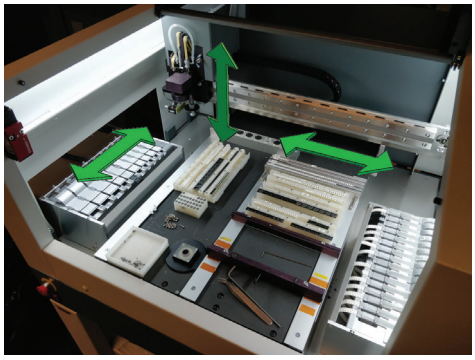
## 4.1 Cartesian Assembly

Most commercial computer numeric control (CNC) machines use some form of Cartesian position system. In other words, these systems have orthogonal X, Y, and Z axes which allow the user to address a rectangular work volume using simple coordinates. Two examples of commercially available Cartesian machines are shown in Figure 4-1.

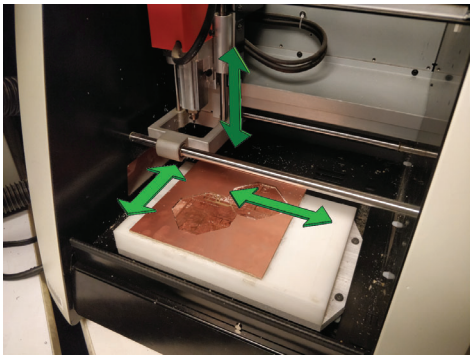
Cartesian motion systems benefit from fully decoupled axes, meaning each axis has its own actuation system which is not affected by the others. Moving in a straight line along one of these axes simply means rotating a single motor a set amount; coordinated movement requires path planning between several motors to account for acceleration. This stands in contrast to serial or parallel manipulators, where any straight line motion must be coordinated between multiple actuators and actuator rotation is related to position through joint kinematics.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.



(a) Mechatronika M20V pick-and-place machine, used for electronics assembly.



(b) Roland SRM-20 desktop mill, used for rapid PCB prototyping.

Figure 4-1: Two commercially available Cartesian machines with principle linear axes labeled using green arrows.

### 4.1.1 Design

An existing Cartesian machine originally fabricated by Will Langford was adapted for Tiny-DICE assembly. This work was carried out by Jiri Zemanek and is included here for the sake of completeness, and because some of the lessons learned during testing informed the Meso-DICE design.

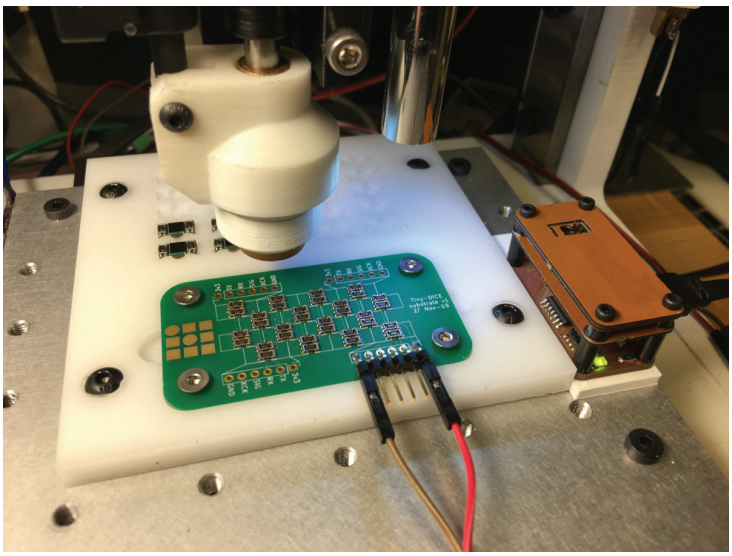
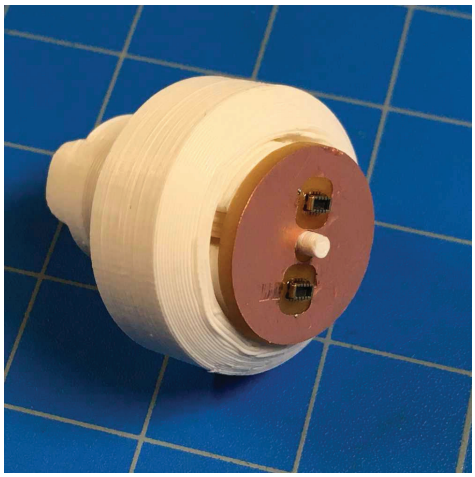


Figure 4-2: An overview of the Tiny-DICE assembly machine, showing green PCB build plate, white 3D printed end effector, planar part storage plate, and integrated programming station to the right. Photo courtesy of Jiri Zemanek.

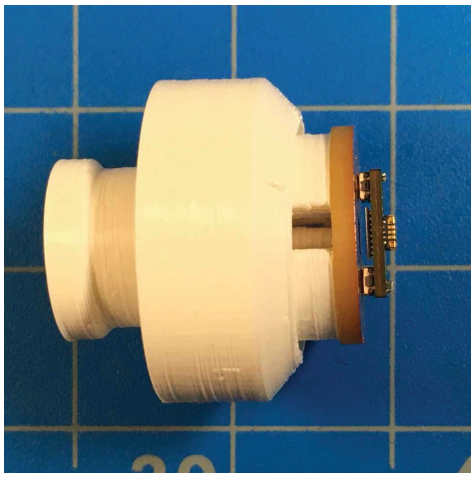


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

With Tiny-DICE, assembly automation was a secondary consideration that was not taken into account during the module design process detailed previously. As such, the modules lack features to assist with handling operations such as pick-up, alignment, or placement. Because of this, the 3D printed end effector simply uses a pair of Molex SlimStack connectors soldered to a milled FR1 PCB, as seen in Figure 4-3. A 3D printed ejector pin runs through a hole in the PCB and is actuated by the motion platform's existing linear servo.



(a) End effector isometric view showing mezzanine connectors and ejector pin.



(b) Side view with Tiny-DICE module installed.

Figure 4-3: Tiny-DICE assembly machine end effector detail. Photos courtesy of Jiri Zemanek.

The Tiny-DICE assembler is controlled by a web-based Javascript interface and a Tiny-G controller. More information about this implementation can be found in [50]. Minor modifications were made to the code to allow quick pick-up of nodes based on integer grid location rather than raw distance, along with rotation during placement since alternating layers are orthogonal.

### 4.1.2 Evaluation

The Cartesian Tiny-DICE assembler is able to build several layers of computational volume without direct user intervention. In order to quickly evaluate electrical connectivity, each node was programmed with a simple script to continuously blink the on-board LED as soon as it powered up. Each node took roughly 12 seconds to pick up, relocate, and assemble; this rate varied somewhat depending on the location of the node in the pickup grid and its

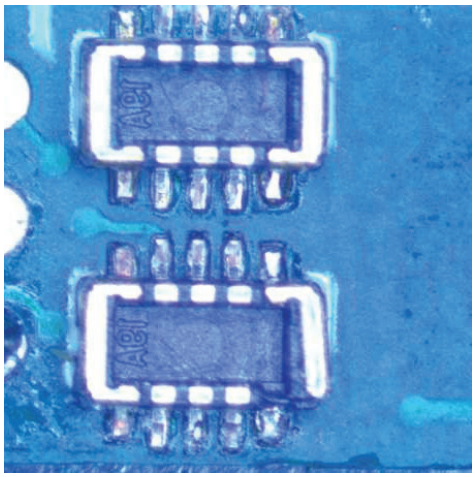


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

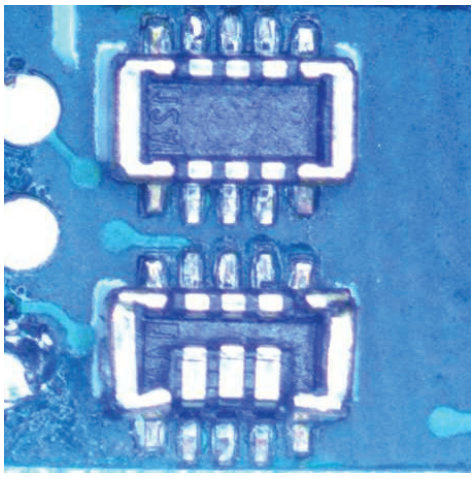
distance to the placement point. As discussed previously, the cantilevered end of each node meant that higher levels could not be assembled as reliably as the first layer, which is built upon a rigid PCB.

During automated assembly tests, inconsistent but alarmingly loud noises were observed during insertion of the nodes. Sometimes, the Tiny-DICE devices would seat quietly; however, in other cases the linear servo driving the insertion pin would audibly strain, and the entire end effector would visibly snap into position when the node seated. This usually occurred during node placement on the build plate, suggesting connector overconstraint due to the shorter kinematic chain connecting the two mezzanine connectors.

After a few automated assembly cycles, several of the nodes stopped working reliably. These failures were identified by observing the onboard LED; failed modules ceased blinking. In some cases, this was observed on the second assembly run; in others, a given node continued to work without any issues. The failed nodes were examined under a Lnyx Evo projection microscope and the connectors were imaged (see Figure 4-4).



(a) Tiny-DICE module with broken mezzanine connector.



(b) Tiny-DICE module with crushed mezzanine connector.

Figure 4-4: Close-up detail of Tiny-DICE modules after several assembly cycles showing broken mezzanine connectors caused by assembly over-constraint and misalignment. Photos courtesy of Jiri Zemanek.

The connector failures noted above are predicted in the SlimStack connector datasheet, which states:

> When mounting several board to board connectors on a same PWB [Printed



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Wire Board, synonymous with PCB in this context], ensure to mount the each mating connector on a separate PWB.

Prior to design of the Tiny-DICE modules, it was observed that such language exists in nearly all commercially available subcompact mezzanine connectors. The decision to move forward despite this concern reflected a disinterest in compromising miniaturization in favor of durability. However, the result directly informed the Meso-DICE design decision to separate connector alignment, latching, and connectivity into distinct sub-components.

## 4.2 6-DOF Assembly

An alternative to Cartesian locating mechanisms is a serial manipulator, where a string of rotary or linear axes are linked end-to-end. This configuration improves the flexibility of the motion system, since rotating joints can allow end effectors to reach under or around objects that would block an X/Y gantry. Because of this flexibility, serial manipulators in the form of 6-degree-of-freedom (6-DOF) robotic arms, such as those shown in Figure 4-5 are popular for factory automation applications where systems must be adapted for different jobs without structural changes.

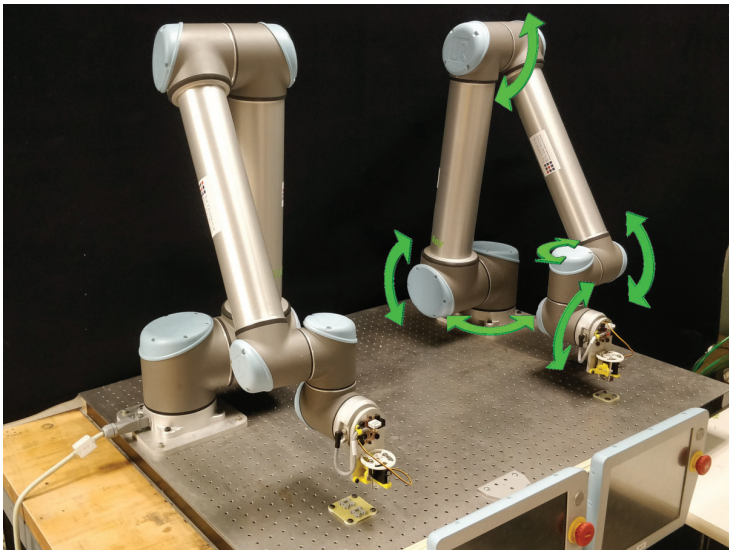


Figure 4-5: The two Universal Robots UR10 6-DOF arms used for Meso-DICE assembly. The strut placement arm is on the left, while the node placement arm is on the right and includes green arrows indicating the six principle rotary axes.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Serial manipulators are substantially more complicated than Cartesian machines to control, since the entire kinematic chain must be considered together to calculate the final location of the end effector. In particular, path planning is difficult since the degrees of freedom are inherently linked; it is relatively straightforward to get a tool from one point to another, but programming a specific path requires substantially more calculation. Fortunately, commercial robotic arms such as the Universal Robots UR10s used in this project include sophisticated inverse kinematic functions that allow users to simply input Cartesian or joint-space coordinates, and the control system translates between the two coordinate systems as needed.

### 4.2.1 End Effector

As seen in Figure 3-13, Meso-DICE struts and nodes include a 3D printed part that holds the Delrin flexural latches in place, aligns struts and nodes to each other, and serves as a gripping point for the assembly machine. The gripping point features consist of tapered rectangular blind cavities, with two such features on each strut and four on each node. The gripping points are located to avoid interfering with adjacent struts or nodes in a partially constructed computational lattice.

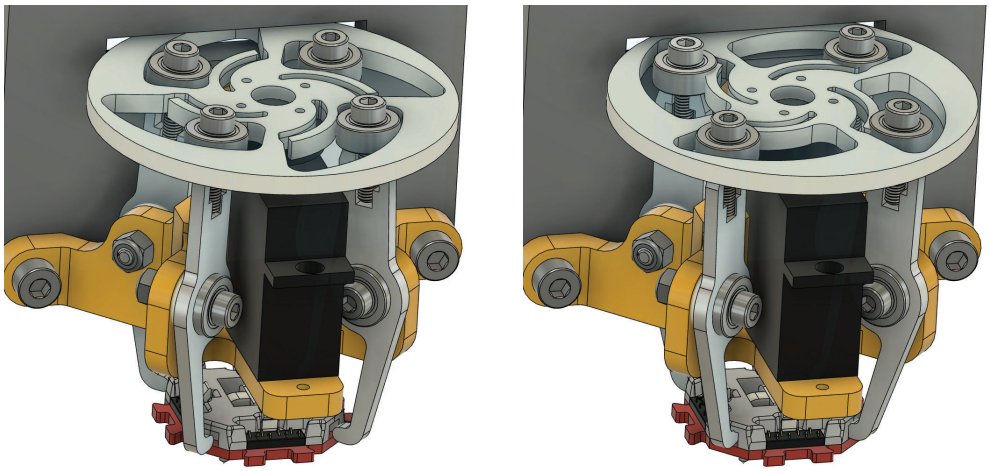
Since the 6-DOF assembly infrastructure uses a pair of UR10 arms and the Meso-DICE lattice consists of two unique parts, an arm and matching end effector was dedicated to placing each type of module. The end effectors use 2.5 mm aluminum arms with tapered rounded barbs to squeeze the gripping points on the modules. The tapered nature of these components accommodates a millimeter of misalignment during module pickup, dramatically simplifying the manual part feeder tray design. The aluminum arms were fabricated on the OMAX waterjet cutter, and press-fit onto ball bearings which are then connected to a 3D printed frame. In order to minimize slop in the arm pivots, stiff Teflon washers fit between the bearings and the frame, and the joint is preloaded and secured with a bolt and locknut. Once adjusted for tension, this arrangement proved to be simple, compact, reliable, and remarkably easy to actuate due to the Teflon's low friction against the bearings.

To minimize size and complexity, the two (struts) or four (nodes) aluminum jaws are actuated with a single Hitec micro hobby servo using a Delrin cam orthogonal to the arm pivot axes. Each arm has a second ball bearing mounted in line with the cam and opposite the barbs which ride along the cam grooves. Notably, the inner working surface of the cam


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

includes flexural beams which grant a small amount of compliance to the cam during pickup, which further helps the system tolerate minor misalignment. Once the servo fully actuates, these flexural elements bottom out and lock the part precisely in place for transport and placement in the computational lattice. This design strategy enabled rapid iteration of the pickup mechanism, as multiple cam profiles could be quickly fabricated and tested without redesigning and fabricating the other end effector components. Figure 4-6 shows a CAD render of the node end effector in both its open and closed state as it picks up a module.

During early testing, it was observed that the struts tended to rotate during placement when not supported on one side, such as at the edge of the lattice structure. Since the strut end effector only grips the modules at two points, this sometimes resulted in incomplete assembly of the parts which could then cause damage when subsequent nodes were added. To address this issue, a milled Delrin "foot" was bolted to the bottom of the strut end effector to provide uniform uniaxial assembly pressure across the entire module. This solved the problem and resulted in reliable assembly of struts at any point around the lattice.



(a) Open. (b) Closed.

Figure 4-6: CAD renders of Meso-DICE node end effector, showing cam and arm action as gripper opens and closes around a module. Note that the compliant inner working surface of the Delrin cam does not distort in the closed render; on the fabricated end effector, these beams bend and bottom out, increasing the rigidity of the gripped part after pick-up.

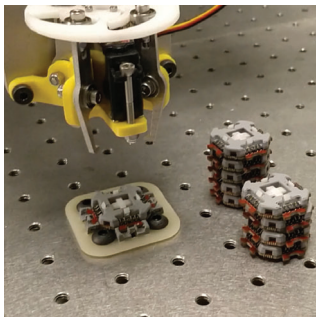

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### 4.2.2 Infrastructure

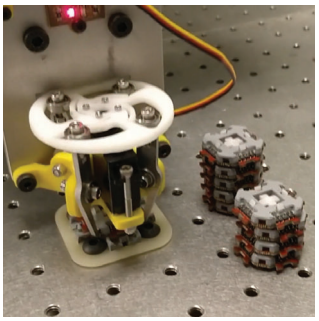
The two UR10s are rigidly mounted to a large stainless steel optical table with 1/4-20 tapped holes on 1" centers. This mounting scheme dramatically simplified system development, as it provided a known flat reference surface with convenient bolt holes for attaching various Meso-DICE accessories, including feeder trays, the node programmer (shown in 3-16), and the lattice build plate.

In a future DICE incarnation, a vibratory feeder or other sorting mechanism would collate modules and prepare them for pickup by the placement system, fully automating the mechanical assembly and leaving the user to simply pour a container of loose modules into a waiting hopper to "recharge" the system. For the Meso-DICE iteration, a simpler strategy was used in which the experimenter manually places a strut or node on a pickup plate while the UR10 is placing the previous module on the lattice. The pickup plate, shown in Figure 4-7, consists of several 3D printed alignment features from the opposite module type (struts or nodes) secured to a milled fiberglass plate without the associated Delrin latch components. In use, the experimenter simply must remain vigilant and replace parts as they are consumed, since the system has no feedback mechanism to determine whether a strut or node is actually picked up from the pickup point.

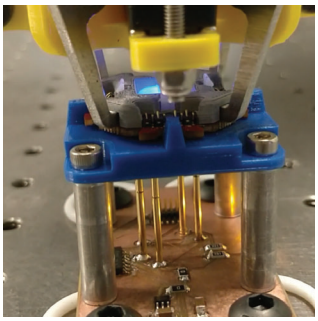

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.



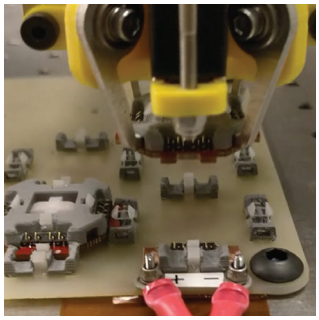
(a) End effector preparing to pick up a node from the pickup plate.



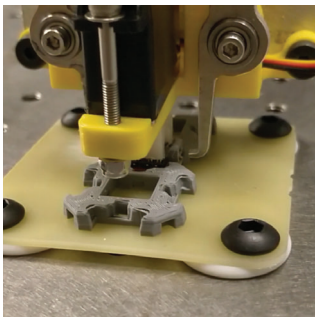
(b) Picking up a node; red LED above end effector indicates gripper actuation.



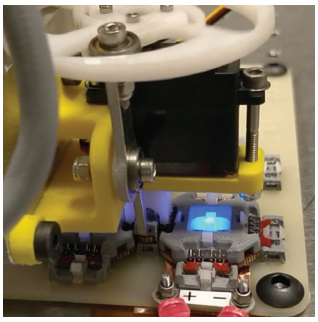
(c) Simulated node programming at powered programming station.



(d) Preparing to place a node on an early version of the build plate.



(e) Picking up a strut; notice milled Delrin placement foot.



(f) Placing a strut after rotating 90 degrees from pickup.

Figure 4-7: Video stills of the Meso-DICE automated assembly process, showing strut and node pickup, programming simulation, and placement.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 5

# Electronic Glitter Lattices

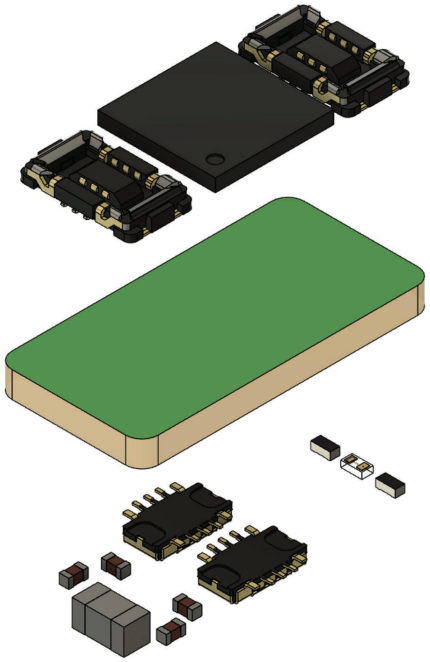
The preceding chapters showed a relatively conventional approach to module design and assembly for DICE, using COTS components and commercially fabricated PCBs. However, the shortcomings of this approach became clear, particularly when pushing the limits of miniaturization with Tiny-DICE. Commercial interconnect systems are application-specific; in the case of the mezzanine connectors used in Tiny-DICE, this application is single-use connection of mobile phone parts during factory assembly. As demonstrated by the frequent failures during automated Tiny-DICE assembly, such connectors do not self-align and cannot withstand the substantial forces caused by overconstraint from parallel insertion.

Clearly, one approach to solving this problem is to design a new connector to fit the novel DICE application. Such a connector would be designed around existing electronics assembly techniques, such as pick-and-place component placement and reflow soldering. The part would use stamped and plated metal features for electrical connectivity, and the structure would consist of overmolded mineral-filled liquid crystal polymer (LCP), a high-temperature engineered plastic commonly used for electronic parts. The LCP features could be designed with large approach chamfers to facilitate connector self-alignment, and the metal contacts could have sufficient compliance to relieve the overconstraint caused by parallel insertion. But a conventional approach to connector design requires a conventional approach to project management. Designing a novel miniature connector from scratch would require a multi-year effort and substantial investment in tooling and pilot-scale fabrication runs, an inflexible method that would risk decoupling the connector design from the unique requirements of the DICE project.

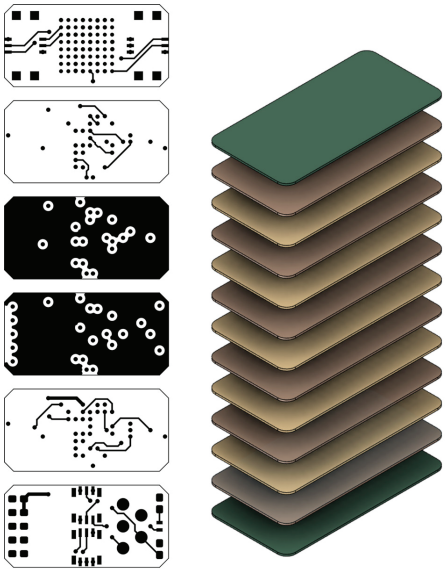

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 5.1 Electronic Digital Materials

Another approach starts by reducing connectors to their fundamental building blocks, which are simply conductive and insulating materials cleverly composed to pipe electrons in specific directions. Interestingly, this same definition could be applied to PCBs, where copper, sheet adhesive, and fiberglass-reinforced polymer (FRP) are marked, etched, drilled, and plated to isolate electrical nets from one another, as seen in Figure 5-1.



(a) Tiny-DICE PCB separated from its components.



(b) Tiny-DICE PCB blown up, showing six conductive copper layers interspersed with insulating fiberglass and solder mask.

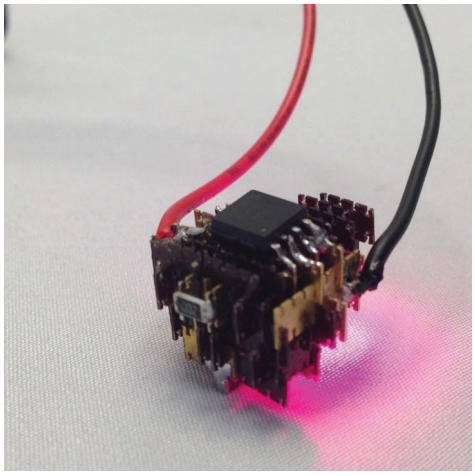
Figure 5-1: Tiny-DICE renders at various levels of deconstruction.

In a dramatic re-imagining of PCBs and connectors, conductive and insulating materials are discretized into two dimensional plates, which are then friction-fit together to build routing structures. This approach, called electronic digital materials, has been demonstrated several times at the millimeter to centimeter scale at CBA, as seen in Figure 5-2.

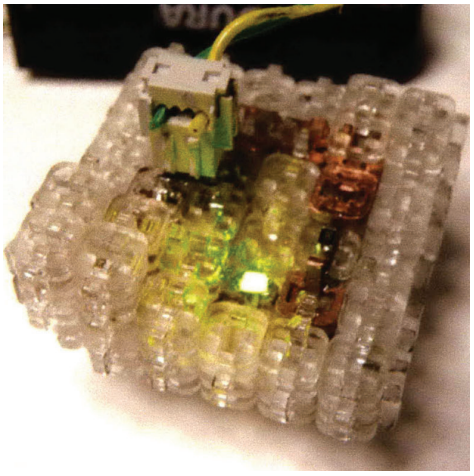
Using digital materials with friction fit joints means heterogeneous materials can be utilized, and structures can be deconstructed and reused after assembly. Two dimensional planar parts also benefit from simple uniaxial assembly, so automating the construction of



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.



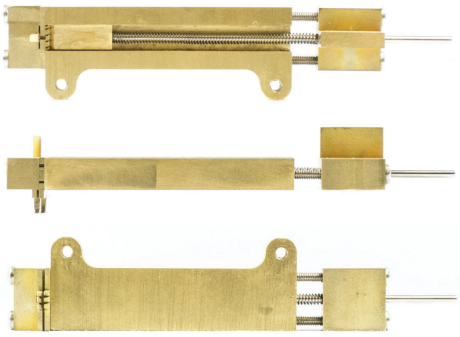
(a) Circuit by Will Langford, 2014 [49].



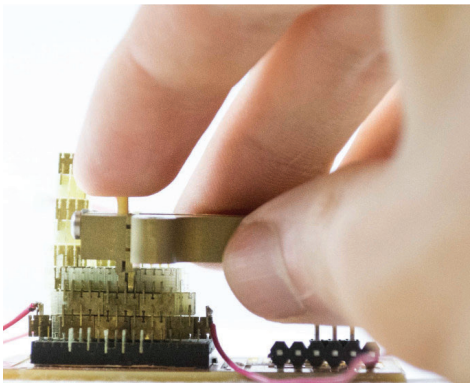
(b) Circuit by Jonathan Ward, 2010 [88].

Figure 5-2: Two prior CBA projects to build electronic circuits from digital materials.

the lattices can be achieved with a relatively straightforward three axis machine. Planar materials can be conveniently dispensed from a magazine, dramatically simplifying part handling. Langford used this strategy to design and fabricate a digital materials "stapler", shown in Figure 5-3.



(a) Manual stapler design.



(b) Stapler in use.

Figure 5-3: Will Langford's electronic digital materials "stapler" [49].

Langford and Ward's work on electronic digital materials was ultimately too large to effectively route complex electronic circuits. While both researchers were able to create simple demonstration circuits, the individual routing elements were too long to freely run circuit nets in any direction. This limitation stemmed from the lack of prototype-scale



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

fabrication methods; Langford suggested deep reactive ion etching (DRIE) as a potential method for scaling electronic digital materials down, but such a process requires lithographic techniques relegated to a clean room and results in unsorted parts which must be handled and prepared for assembly. He also showed that direct-write laser machining was simply too slow to practically produce sufficient parts for a demonstration circuit.

## 5.2 Glitter

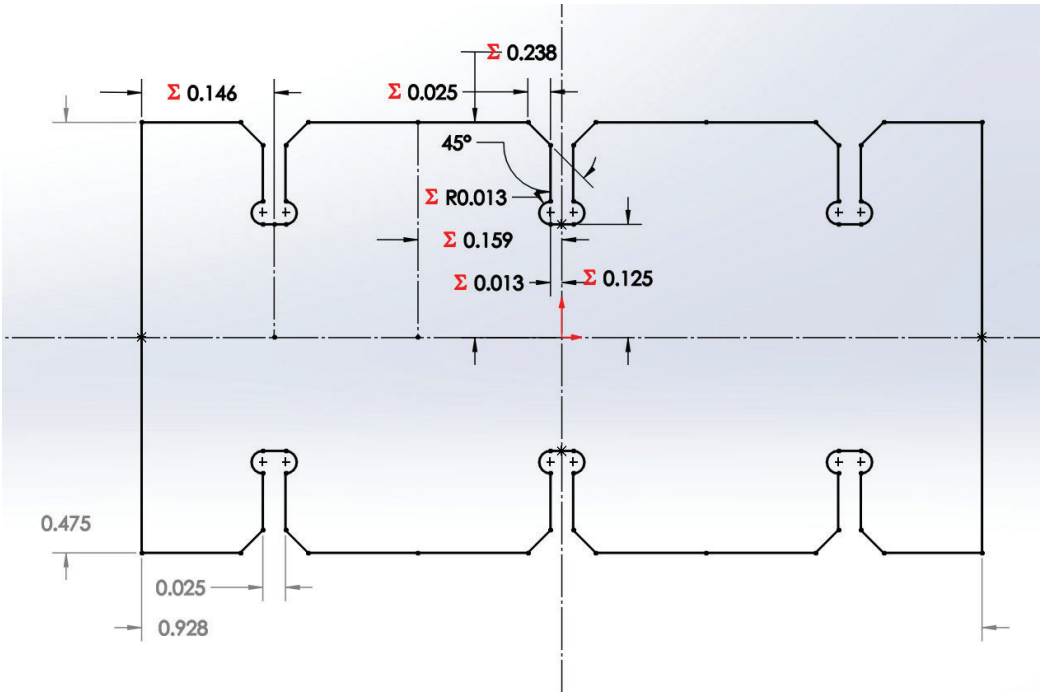
The largest commonly used surface-mount integrated circuits are provided in Small-Outline Integrated Circuit (SOIC) packages with a lead spacing of 1.27 mm. Denser components include Quad Flatpack (QFP) components with 0.8 mm or 0.5 mm lead spacing, along with Thin Shrink Small Outline Package (TSSOP) with 0.65 mm or 0.5 mm lead spacing. In order to freely access all of the pins on such a device, a square lattice with a pitch in the range of the lead spacing can be used, provided adjacent elements can be electrically isolated from one another. If four plates are joined near their corners to form such a routing element, each plate is around a millimeter along its longest dimension, roughly the size of fine glitter and a five-fold decrease in size compared to Langford's work. One design for such a material is shown in Figure 5-4.

Once isolated conductive nets can be routed at the same pitch as the smallest components, any arbitrary circuit can be constructed to connect components to one another. In conventional PCB fabrication, as complexity grows and more nets need to cross one another, additional layers must be added to the board during fabrication; adding such layers increases fabrication cost and turnaround time substantially, limiting real-world design complexity. The approach described here, in contrast, can simply grow vertically to accommodate additional routing with a corresponding linear increase in part count, cost, and fabrication time.

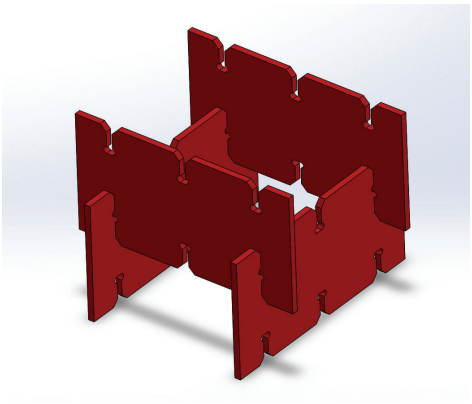
### 5.2.1 Fabrication

Conductive glitter is prepared by stacking multiple sheets of stock, such as 25 micron phosphor bronze foil, and sandwiching the stack between two thicker pieces of aluminum plate. The entire assembly is then cut using a micro wire-EDM [20] that uses 10 or 20 um tungsten wire, shown in Figure 5-5. Unfortunately, delays related to the Covid-19 pandemic pushed

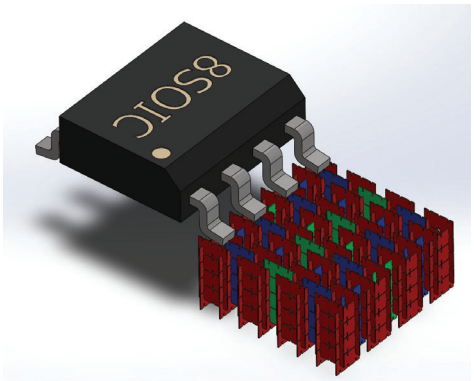

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.




(a) CAD sketch of electronic glitter with dimensions in millimeters.



(b) Four glitter plates fit together to form a routing node.




(c) Multiple nodes (red) joined with struts (blue, green) to route an SOIC-packaged IC.

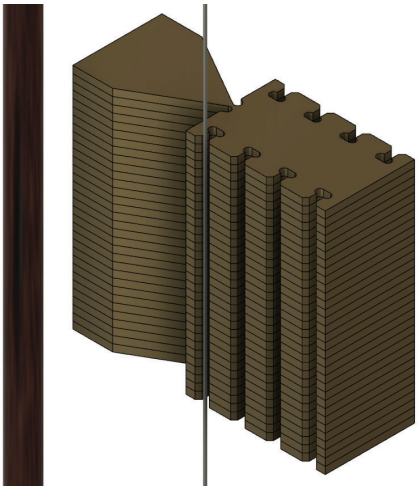
Figure 5-4: Electronic glitter element design and lattice structure.

back micro-wire EDM system fabrication and delivery, so reducing this concept to practice falls into the "future work" category.

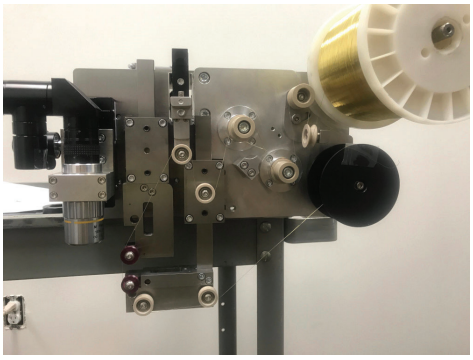
Insulating glitter cannot be cut using the same techniques as wire-EDM fundamentally requires a conductive substrate. While insulating elements could be prepared using laser micro-machining or another direct-write process, a better option is to use the micro-wire



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.



(a) Fabricating multiple conductive elements from 25 um phosphor bronze foil. Human hair shown for scale.



(b) Custom-built micro-wire EDM tool head as of August 2021. Photo courtesy of Viteris Technologies [20].

Figure 5-5: Conductive glitter fabrication process using micro-wire EDM.

EDM to machine precision stamping tooling out of hardened steel or tungsten carbide, and then rapidly stamp the material out of a strip of material. There are several polymeric materials that could be suitable and are available as thin 25 um films, but an interesting low-cost alternative is to use muscovite mica, a naturally occurring mineral with a long history of use in the electronics industry as a low-cost structural insulator. For example, most household pop-up toasters use sheets of mica as heating element supports [84], since the material is low-cost, heat resistant, electrically insulating, and reasonably strong. Mica has a distinct advantage in that its crystal structure is extremely anisotropic with a reliable cleavage plane, so the material can be readily split into precise thin sheets. In fact, one of the scientific uses for high-grade mica is as a sample surface for atomic force microscopy [29], since freshly split sheets are atomically flat. Once prepared, the material is then stamped using the aforementioned precision dies.

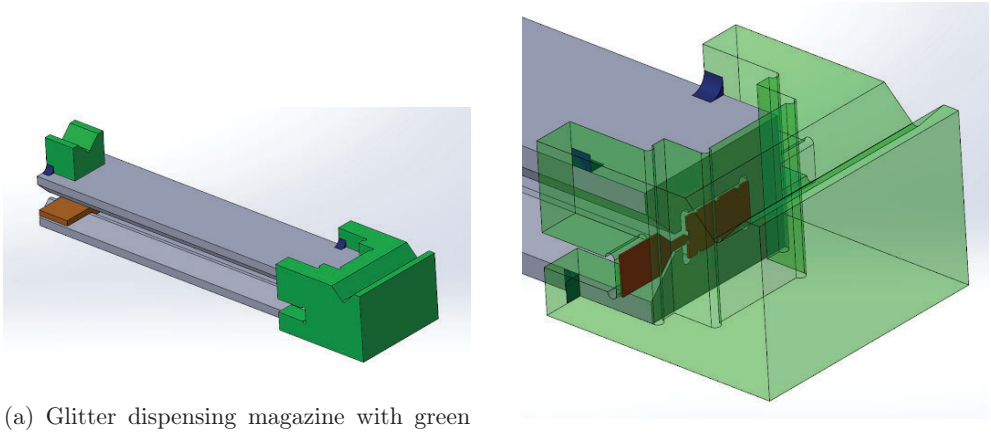
### 5.2.2 Assembly

As discussed previously, a compelling advantage to assembling 3D structures from press-fit 2D components is uniaxial assembly. This means parts can be dispensed from a magazine rather than picked up from a tray or tape as is done with electronics pick-and-place machines that are not of the high-speed "chip-shooter" variant. By including an additional break-



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

away tab on each conductive part, a stack of plates can be removed from the micro-wire EDM and inserted in bulk into a magazine; once secured with a follower, the tab can be snapped off and the parts are ready to dispense. This is shown in Figure 5-6.



(a) Glitter dispensing magazine with green kinematic locating features at top and orange magazine follower at left.

(b) Close-up view of stapler head, showing glitter with loading tab still attached in red.

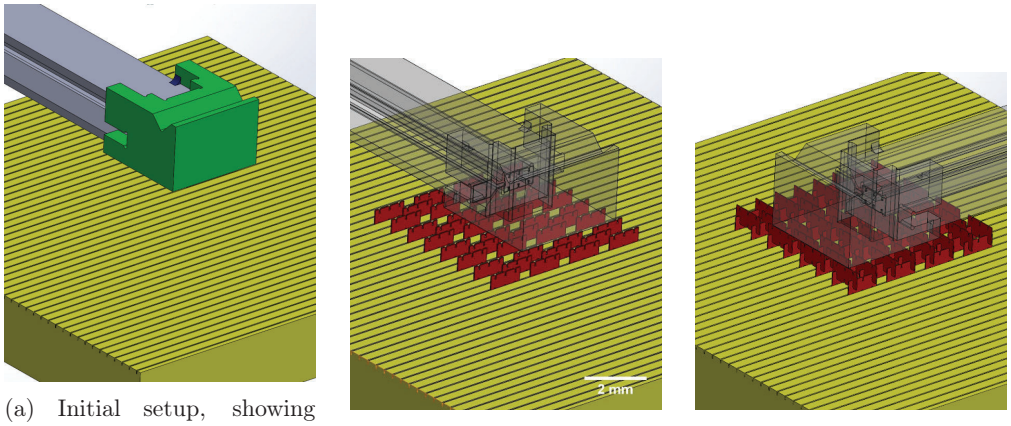
Figure 5-6: Conductive glitter magazine with integrated stapler head.

The magazine mounts kinematically using a Maxwell-type coupling, in which the three V-grooves on the magazine interface with three matching spheres on the assembly machine, constraining all six degrees of motion in a repeatable manner. The magazine is held in place with a powerful rare-earth magnet, allowing it to be quickly swapped to change between conductive and insulating parts. The assembly machine consists of the magazine and coupling, which translate vertically to add layers to the build lattice, and the build surface itself, which is mounted on an X/Y platform below the magazine. The build surface is a laminated assembly which has a series of grooves in one direction that match the lattice pitch and stock thickness, and are sized to provide a light friction fit for the assembled structure. During assembly, the stapler head moves horizontally relative to the build plate to assemble a complete layer, and then translates vertically to begin the next orthogonal layer. This process is depicted in Figure 5-7.

The maximum placement rate of the assembly machine will be limited by the stiffness of the motion control system and the speed of the actuators controlling the location of the placement head, along with the cycle time of the stapler mechanism itself.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.




(a) Initial setup, showing build plate in yellow and magazine in gray and green.

(b) Placing the first row of glitter (red).

(c) Placing the second row of glitter at 90 degrees (red).

Figure 5-7: Automated electronic glitter lattice assembly using stapler-type plate dispenser.

## 5.3 Interconnect

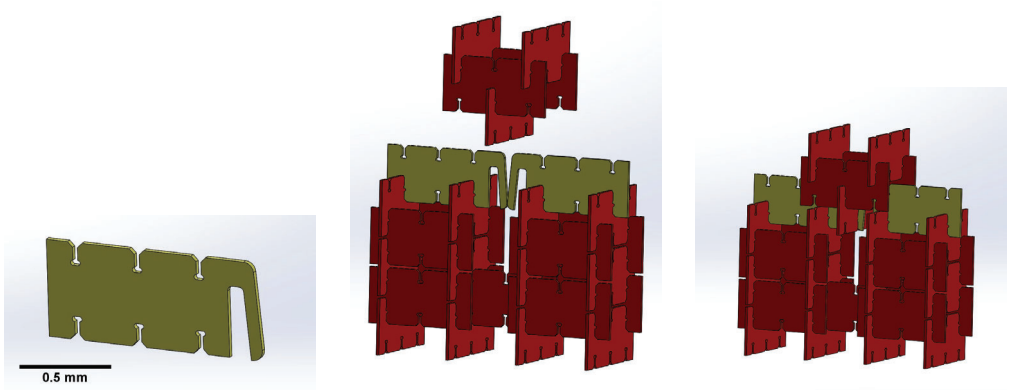
Adding new building blocks to the glitter assembly ecosystem must be approached cautiously. Each new part type requires a dedicated magazine for placement; even if these assemblies can be automatically swapped during the build process, they add significant complexity to the assembly machine. A reasonable approach is to only add parts that provide unique and interesting inherent qualities tied to their construction material; for example, the conductive and insulating parts described earlier could be supplemented by actuation and flexure to provide motion, as shown by Langford in 2019 [50], and by resistive and semiconducting parts to provide passive and active circuit elements, as shown by Langford in 2014 [49]. Other parts could include environmental sensing elements or output devices whose function depend on construction material. A comparison can be made here to biology, in which 20 amino acids form the basis for life and differ in fundamental physical property, such as affinity for water or pH [38].

A worthy exception to this guideline is a dedicated conductive part which enables reversible module-to-module interconnect using flexural extensions. On the surface, such a part is similar to the standard conductive part since it is manufactured similarly out of identical material. However, as shown in Figure 5-8, the interconnect part interfaces with the flat edge of normal conductive parts using a flexural sliding interface rather than an interlocking sliding joint, so the joint characteristics can be tuned separately from the inter-



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

locking parts. This means we can get around the LEGO problem: it is difficult to assemble a LEGO structure using tools made from LEGO blocks if the tools pick up blocks in the same manner they are held together. Mechanically distinguishing inter- and intra-DICE module assembly is fundamental to enabling computational reconfiguration without risking module deconstruction.




(a) Interconnect part detail, showing scale and integrated flexural beam.

(b) Pre-assembly view showing two interconnect parts installed facing each other.

(c) Post-assembly view showing reversibly assembled structure.

Figure 5-8: Flexural interconnect glitter part used to join adjacent DICE nodes.

## 5.4 Scaling

A typical circuit fabricated in a FabLab is 50 x 50 mm, and consists of only one routing layer. With square millimeter routing nodes and connecting struts, this means an equivalent substrate fabricated from electronic glitter would require $50 \times 50 \times 5 = 12,500$ discrete conductive and insulating elements. Assuming a 1 Hz assembly rate, such a circuit could be prototyped in a few hours. Clearly, this will require at least some consideration for magazine changes; with 25 um parts, a reasonably sized 25 mm magazine would only hold 1000 elements, so this process will also require automation. A more complex circuit such as Meso-DICE, which is a 4-layer design which is roughly 25 x 25 mm would take up to twice as long, since each layer would be separated by a full layer of insulating elements. Of course, this assumes the components are mounted conventionally on a planar top layer; in reality, the structure would only be as tall as it needs to be, so the actual quantity of parts might be half or fewer than that estimate.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Beyond prototyping, simply adding additional assemblers would only increase throughput linearly; even 100 machines would only assemble 400 circuits in a typical 8-hour shift. A far more compelling approach mirrors the biological ribosome by recursing the assembly machines themselves: in other words, assemblers building assemblers, which could then scale exponentially given materials, energy, and space. In this arrangement, assembly machines would be built out of the same conductive and insulating parts they use to build circuits, taking advantage of two additional fundamental elements: flexural, to constrain motion; and mechanically active, to create motion. Such a system would require attention to a great many details, including part handling (since magazines would be fabricated out of the parts they hold) and large-scale motion logistics (since the machines would need to leave the build platform eventually). But a clear next step is to build a glitter assembly system using flexural elements, which is the motivation for the following chapters.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 6

# Modular Superelastic Flexures

Most motion systems, such as the Cartesian and 6-DOF assembly systems presented previously, use rolling or sliding elements to constrain movement to desired rotary or linear axes. For macro- and meso-scale work such systems are ideal, since bearings and slides are ubiquitous, cheap, and sufficiently performant. In fact, the proliferation of low-cost 3D printers over the past ten years has meant a wide variety of linear rails, ball screws, and precision bearings are available in single-lot quantities through a multitude of international distributors. The DICE designs shown previously existed on the right scale for such machines; Meso-DICE in particular included self-alignment features large enough to tolerate millimeters of misalignment during assembly provided the tools included compliance. While that iteration used commercially available robotic arms for assembly, it would have been reasonable to quickly build a Cartesian assembler without too much effort.

Reducing resolvable size below around 10 microns, however, becomes difficult with low-cost commercial components. While rotary actuators can be geared down to increase resolution, this strategy starts to introduce backlash and complexity into the system. Bearing and lead screw quality starts to play a larger role, and simple roller-on-extrusion motion constraints reach their limits due to inconsistent surface finish. Of course, many of these issues can be ameliorated through grinding, lapping, and other precision fabrication techniques. However, these methods are highly operator dependent and can require expensive specialized machines to accomplish.

Notably, the electronic glitter lattices presented previously are near this complexity inflection point. The discussed part geometry uses 25 um feedstock with matching slots,


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

each of which has a 25 um chamfer to pull mating parts into alignment. Ideally, the assembly machine will have a positioning resolution several times smaller than that so fine adjustments can be made to minimize insertion force and error rate.

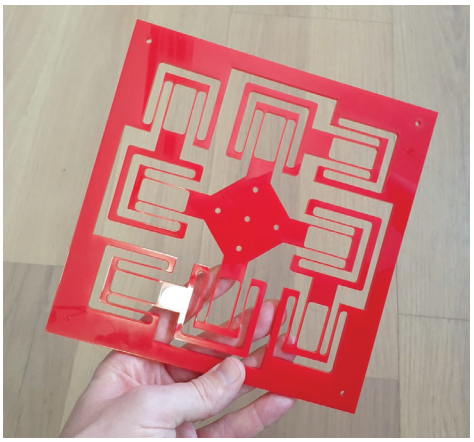
## 6.1 Flexures

Flexures are a method for constraining motion that provides an appealing alternative to conventional rotating or sliding joints, particularly at micron (or smaller) resolution [18]. These deliberately compliant features rely on reversible elastic deformation rather than sliding or rolling elements. The advantages of such a design are clear; flexures are highly repeatable, have practically zero backlash, require no lubrication, and can be modeled using simple beam equations. On the other hand, flexures cannot perfectly emulate rotary joint kinematics; even with exotic beam shapes, the virtual center of rotation of a flexure does not remain stationary through its whole range of motion [51]. While this motion is repeatable and predictable, the traced path of a flexural joint is not perfectly circular. Flexures are also non-trivial to fabricate; most are either bolted on pieces of flexible shim stock, or are integrated into complex monolithic assemblies fabricated using wire-EDM, milling, laser cutting, or injection molding, as seen in Figure 6-1.

A far more limiting aspect of flexures is mechanical range, particularly for elements fabricated from stiff metal. Typical stainless steel alloys yield at less than 1% strain; beyond this, unrecoverable plastic deformation takes place which adversely affects the repeatability of the system. Worse, fatigue concerns usually push designers to restrict mechanism displacement to one-third of the material's yield strain. Taken together, these limitations dramatically reduce the addressable workspace for a given mechanism. For example, Shorya Awtar's two-axis parallel XY stage [17], shown in Figure 6-2, measures 300 x 300 mm, but only has an addressable range of 5 x 5 mm. Notably, the aluminum beams that make up the flexural mechanisms have a length to width ratio of 76:1 (47.5 mm long and 0.625 mm wide). Decreasing this ratio would increase the stiffness of the machine and reduces overall size, but would also further restrict addressable range.

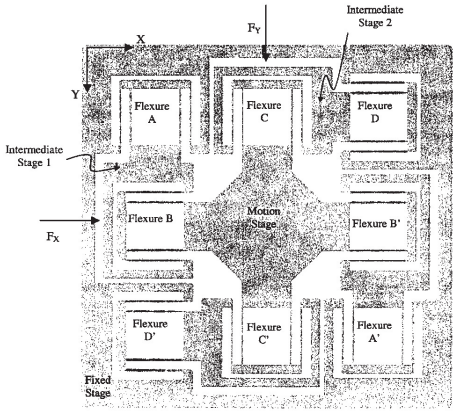

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

(a) The injection molded plastic lid from a Tic-Tac candy dispenser, showing an integrated flexural hinge [66].



(b) A 2-degree-of-freedom flexural stage from the Urumbu circuit mill project, laser-cut from acrylic [83].

Figure 6-1: Examples of monolithic flexure construction techniques.




(a) Stage diagram showing actuator locations (Fx and Fy), flexures, and intermediate stages.

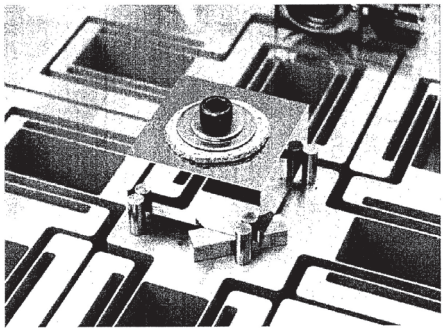


*Fig. 6.6 Metrology Target Block*

(b) Image of actual test flexure, fabricated using wire-EDM from a monolithic plate of 6061-T651 aluminum.

Figure 6-2: 2-axis flexural motion stage by Shorya Awtar [17]. Used with permission.

## 6.2 Superelastic Materials

The traditional flexural elements discussed above constrain motion by averaging many molecular deformations that occur throughout the material when loaded. As strain increases



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

beyond the yield point, irreversible plastic deformation starts to complement this elastic behavior; when the load is removed, the material only springs back partway. An annotated stress-strain curve for such a material is shown in Figure 6-3.

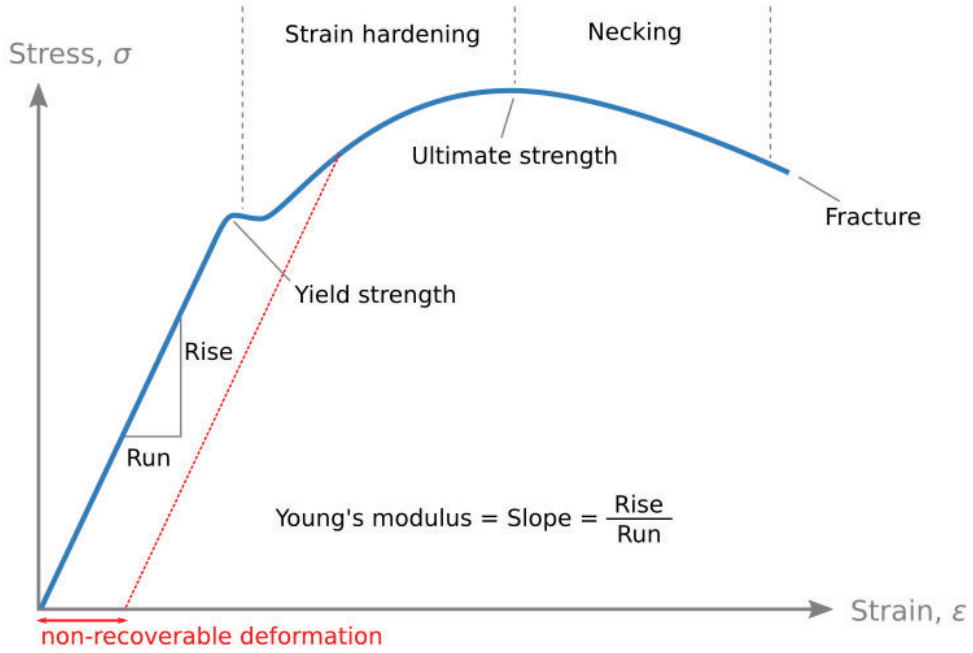



Figure 6-3: Stress-strain curve of a typical metal, annotated to show recoverable and non-recoverable displacement after load is removed. From [62], with added annotations in red.

A specialized type of materials called shape memory alloys (SMAs) exist which are capable of sustaining far greater reversible strain than conventional metals. This property, called superelasticity, is the result of a stress-induced crystallographic phase transformation. In the popular nickel-titanium SMA called nitinol (for Nickel Titanium Naval Ordinance Laboratory, where it was discovered), the stable austenitic phase phase reversibly changes to martensite, allowing for 6-8% strain recovery without permanent deformation. This behavior is illustrated in Figure 6-4.

Clearly, fabricating flexural elements out of superelastic materials would be beneficial, conceivably allowing far smaller flexural mechanisms to be designed. However, nitinol is difficult to manufacture and generally used in wire or sheet form; thicker stock is expensive and hard to find, with one vendor selling a 25 x 150 x 10 mm bar for $299. This is roughly 200 times more than extruded aluminum bar stock. In other words, Dr. Awtar's 2-axis stage


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

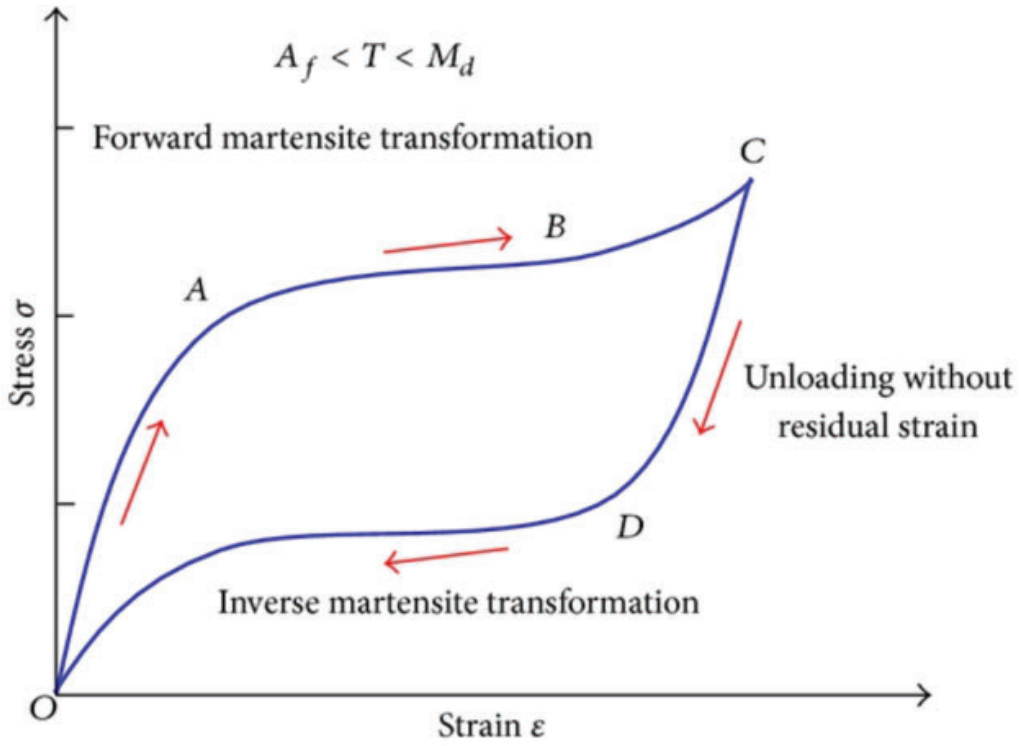



Figure 6-4: Stress-strain curve of a superelastic alloy, showing recoverable strain beyond the typical yield point of a conventional metal [69].

would increase in cost from $90 (if aluminum) to $18k (if nitinol), assuming 25 mm nitinol plate could be procured at the same volumetric cost. Calculations for nitinol, aluminum, and a few other metals are shown in Table 6.1.

## 6.3 Modularity

One approach to controlling high material costs is to introduce heterogeneity into the design. In other words, if the structural elements of a mechanism could be fabricated out of a cheaper material such as aluminum, the overall cost of the system could be dramatically reduced. However, this introduces the significant complication of joinery; the various materials must be connected in a sufficiently rigid manner so as to avoid introducing any unintentional flexibility into the system. Welding is easy to disqualify, due to the significant difficulty encountered in working with titanium alloys, along with the importance of precise heat treatment for superelastic materials (which a welding-induced heat-affected zone


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| Alloy | Cost per cubic centimeter | Cost of Dr. Awtar's stage |
|---|---|---|
| Al 6061-T6 | $0.040 | $90.60 |
| 304 Stainless | $0.084 | $190 |
| Invar 36 | $0.704 | $1590 |
| Nitinol | $7.97 | $17,900 |
| Gold | $1,130 | $2,530,000 |

Table 6.1: Nitinol cost as compared to other metals and alloys.

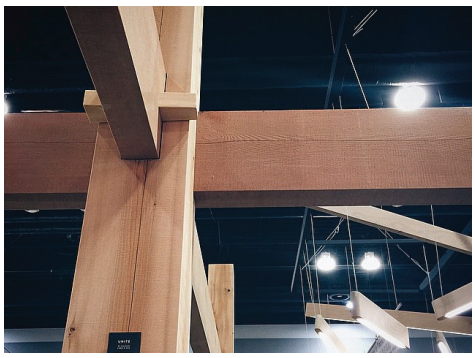
would damage). Bolts are frequently used to assemble blade-type flexures with spring steel, but require substantial design considerations and can suffer from micro-slip if not torqued properly. Adhesives can be highly performant, but require careful surface preparation and can exhibit viscoelasticity under sustained load.

### 6.3.1 Orthogonal Taper Pin Joints

An alternative to the conventional metal-to-metal connection techniques outlined above draws inspiration from two unrelated sources: the aerospace industry and traditional wood joinery. Since the dawn of the jet age, one of the most important joints on an airplane has been the interface between a turbine blade and a turbine disk. While fabrication methods vary, blades are generally investment cast from superalloys capable of withstanding high-temperature creep; disks, on the other hand, are forged and machined titanium. The blades are replaceable, but once installed must withstand massive centrifugal forces caused by the disk rotating about the turbine's axis. To minimize rotating mass, the joint must be as small as possible, but (particularly for compressor blades) must be able to accommodate the thermal expansion of the blade during service. To address all of these demands, blades are secured in machined slots around the disk perimeter, using a bulb, a dovetail, or most recently, a "fir-tree" arrangement.

Wood joinery also makes frequent use of dovetail joints, most clearly seen to attach drawer faces to sides. Angled dovetails are relatively easy to fabricate by hand (via chisel) or machine (via router), and result in far stronger joints than simple finger joints which do not mechanically lock together. In some cases, such as complex 3- or 4-way wood beam joints, internal dovetails are supplemented by tapered wedges which are tapped in place after assembly to pull the joint together. Two examples of wedge-secured wood joints are shown in Figure 6-5.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.



(a) Reproduced traditional Japanese joinery, showing wedge securing method [41].

(b) Tenon joints from a French granary (public domain image).

Figure 6-5: Examples of wood joinery which makes use of simple wedges to secure pieces together.

The wedge is an incredibly powerful simple machine. Discounting energy dissipation (via friction-induced heating) or storage (via elastic deformation of the wedge itself), a wedge is a linear force multiplier, where the mechanical advantage is simply the ratio of the width of the wedge versus its length:

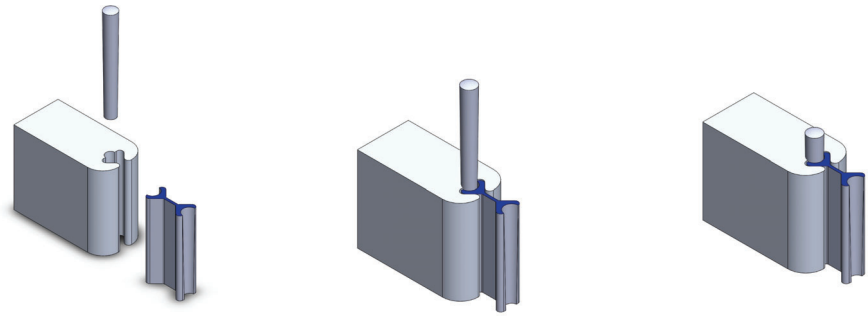
$$MA = \frac{length}{width}$$

This phenomenon is used extensively in the wood examples shown above, where a wedge is inserted orthogonal to the joint's principle axis and tapped home. This imparts a higher magnitude axial force along the joint, pulling the wood beams together. Friction between the wedge and the substrate secure the joint until the wedge is tapped out.

The metal joint design demonstrated here uses commercially sourced round taper pins, which are typically ground from steel and have a 48:1 taper, providing a far greater mechanical advantage as compared to wood wedges. Conventionally, such taper pins are used to precisely and repeatably align parallel sheets of material during mechanical assembly, or to secure rotating elements onto shafts. Here, taper pins are used like wood wedges, providing orthogonal force amplification to the dovetail and relying on friction to stay in place. A diagram of the orthogonal taper pin joint is shown in Figure 6-6, and a dimensioned drawing of the flexural element is shown in Figure 6-7.

Importantly, the taper pins themselves must be hard relative to the structural elements and flexures. Commercially, hardened stainless steel and carbon steel taper pins are readily


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.



(a) Taper pin, flexure, and structural piece awaiting assembly.

(b) Flexure inserted into dovetail and pin loose-fitted.

(c) Taper pin driven home, either via hammer taps or a tool.

Figure 6-6: Orthogonal taper pin joints at three stages of assembly.

available which fit this requirement. It is tempting to leverage the clock industry, which makes extensive use of soft brass taper pins for assembly tasks, as these pins are substantially cheaper than their ferrous counterparts. However, in testing the brass pins tended to plasticly deform and loosen over time, rather than elastically strain the dovetail elements and pin the assembly together with sustained force. Furthermore, brass pins are far easier to accidentally bend or break during assembly and disassembly, resulting in awkward and time-consuming pin extraction operations.

## 6.4 Fabrication

### 6.4.1 Flexural Elements

Flexural elements are prepared using a type of electrical discharge machining (EDM) that uses a hair-thin length of wire as the electrode. This method, called wire-EDM, can cut through virtually any conductive material with a high degree of precision. In use, the wire-EDM system maintains a voltage potential between the wire and the workpiece. As the wire is brought closer to the workpiece, the dielectric gap between the two conductors breaks down and allows a pulse of current to pass across the gap. This minute spark vaporizes a tiny section of the workpiece, which quickly condenses and solidifies in the dielectric fluid surrounding the cut. A high pressure jet of dielectric flushes the particle out of the gap to avoid shorting against the wire, and the process repeats thousands of times every second. Wire is continuously refreshed from a spool and collected in a waste bin for later recycling.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

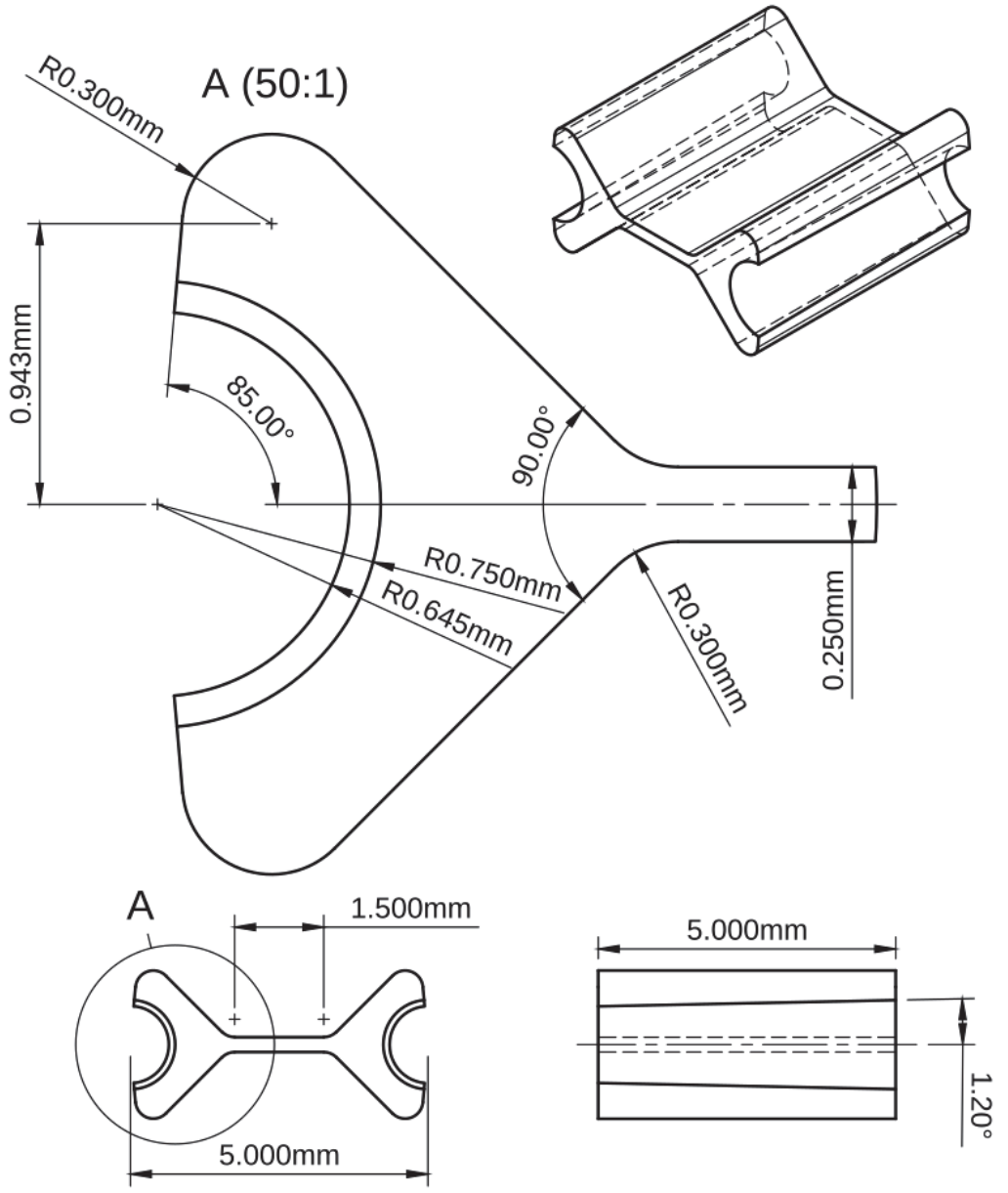



Figure 6-7: Dimensioned drawing of a modular superelastic flexure intended for orthogonal taper-pin installation [31].

As the gap between the wire and the workpiece grows, the voltage required to create the electrical discharge increases; a servoing circuit on the wire-EDM machine detects this change and advances the wire according to a predefined program.

Wire-EDM is good fit for prototyping nitinol flexures because it is a non-mechanical


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

method, so it avoids the normal pitfalls associated with milling titanium alloys: tool wear, breakage, and overheating. The wire feed system on most machines, including the Sodick SL400G at the CBA, can be selectively tilted to create tapers on parts; in this way, both the straight edges of the flexural area and the precise 48:1 taper of the joint can be fabricated in a single operation. Using 150 um brass wire and typical machining settings, it takes roughly 15 minutes to fabricate a single flexure. Fortunately, many flexures can be linked together with tabs and fabricated together, and the machine can be left to run overnight. An image of many nitinol flexures machined in this manner is shown in Figure 6-8.

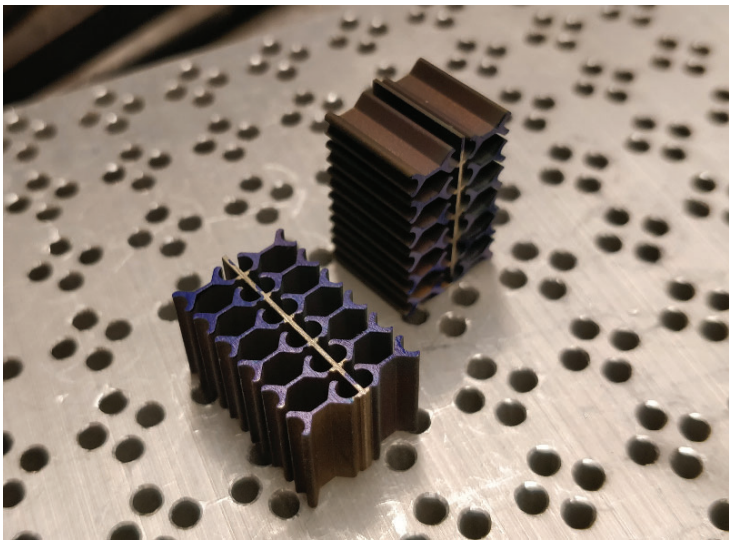


Figure 6-8: Twenty-four modular superelastic flexures machined in two batches. Note blue paint marks, which indicate the larger side of the tapered hole [31].

At scales beyond prototyping, other methods of fabrication would need to be explored. An obvious choice is investment casting, where a replica of the desired part is first fabricated from wax, which is then joined to a common "tree" with other identical parts. The assembly is then dipped in ceramic slurry and coated in sand; once this dries, the wax is burned out and the ceramic sintered in a furnace. Molten metal is then poured into the resulting cavity and allowed to cool, at which point the ceramic is broken away and the parts separated from the tree. One problem with this scheme is that molten titanium is highly reactive, so the melting and casting operation must take place in a vacuum furnace. While such facilities are commonly used to cast titanium elements for the aerospace industry, the capital costs of the equipment and the substantial pre-casting process makes investment cast parts extremely


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

expensive.

A far better fabrication solution is to use metal injection molding (MIM). In this process, a powdered precursor alloy is compounded with a polymeric binder and pelletized, and then fed through a conventional injection molding system to produce "green" parts, which are then washed in a recirculating solvent bath to remove most of the binder. These delicate and porous "brown" parts are then sintered in a vacuum furnace, which burns away the remaining binder and fuses the metal into a densified structure that usually requires minimal post-secondary operations to meet final dimensional specifications. This method has been used to fabricate nitinol parts [75], but mechanical tests showed lower-than-expected elongation at failure suggesting further process development would be needed. If these limitations could be overcome, the per-unit price for the flexures would likely plummet as compared to investment cast or wire-EDMed parts.

### 6.4.2 Supporting Structures

One advantage to modularizing flexural systems is that the rigid supporting sections can be made from a variety of convenient stiff materials. In this case, extruded aluminum plate is ideal, as it is relatively cheap and easy to machine. As with the flexures themselves, the structural parts of the flexure system can be wire-EDMed, which provides outstanding dimensional control and includes the 48:1 taper to match the assembly pins. However, the structural elements are generally far larger than the flexures themselves, so the machining time is unacceptably long.

Another option is to cut the structural parts using a precision waterjet cutter. This machine uses an ultra-high pressure jet of water, normally around 50,000 psig, to accelerate a stream of 40-grit sharp garnet grains to roughly Mach 2. The jet is controlled by a gantry, which directs it to follow a cutting path defined by a part file. As with other cutting methods using lasers or plasma jets, the kerf tends to expand slightly throughout the cut; to counter this, higher-end waterjet cutters (like the one used here, an OMAX 5555) can be equipped with a pair of tilting axes near the nozzle to counteract taper. This produces high quality parts out of thick stock; importantly, the cutting rate exceeds wire-EDM by two orders of magnitude or more, turning overnight jobs into coffee-break-scale tasks.

Waterjet cutting is fundamentally a 2D process; while kerf angle can be reduced to negligible values using the tilting nozzle scheme outlined above, the mechanism is unsuitable


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

for producing the precision 48:1 taper needed by the pins. Fortunately, the proliferation of taper pins for other purposes has resulted in the availability of matching 48:1 precision reams, designed to add a subtle angle to a drilled hole. Even better, spiral-pattern taper reams can be procured which are suitable for use in discontinuous holes that may cause an ordinary parallel ream to jam. In use, an extra superelastic flexure is inserted into a freshly water-jetted part; this element acts as a guide for the ream, which is then inserted and gently turned until the matte finish in the bore is completely shiny. This process is shown in Figure 6-9.

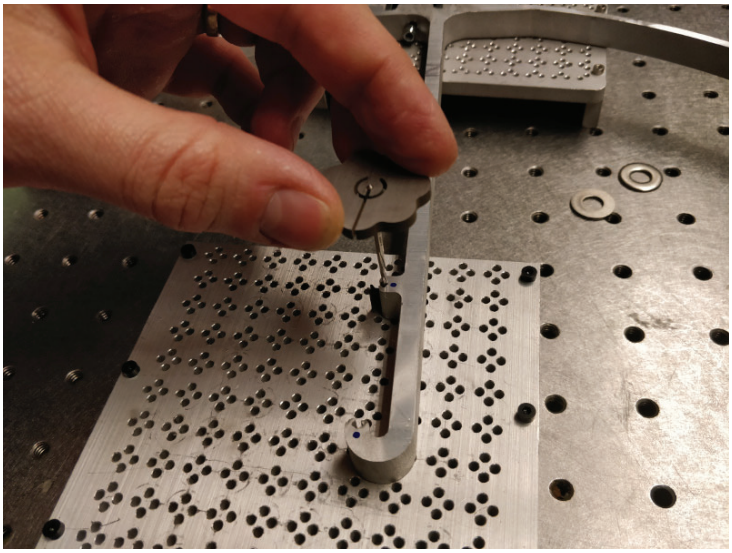


Figure 6-9: Hand-reaming a waterjet-cut aluminum frame to produce a taper suitable for orthogonal pinning. Note custom wire-EDMed handle on taper ream [31].

The fabrication methods described above are more suitable for mass production than wire-EDM, but water-jetting still suffers from two shortcomings: it requires a significant supply of garnet, which is relatively expensive, and it inevitably wastes raw material unless adjacent parts are perfectly tessellated on the stock sheet. To scale production economically, the clear choice here would be die casting, or for high performance applications, a more sophisticated but superficially similar process such as squeeze casting or thixoforming. Die casting and its relatives are ideal because dimensional control is more than sufficient for taper pin secured joinery, meaning parts would likely require little post-processing beyond aesthetic operations such as flash removal and light sanding. Another option would be to machine parts using conventional subtractive methods and then broach the dovetail features


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

with a custom tool; this could be feasible if ferrous alloys (which cannot be die-cast) are required. However, this would still require a post-fabrication taper reaming step which would add cost and complexity.

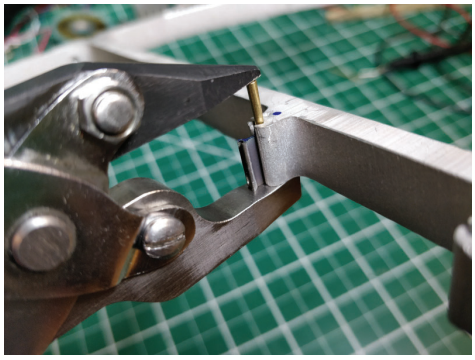
### 6.4.3 Assembly

Building assemblies from modular superelastic flexures is simply a matter of driving the taper pins far enough into the joints that they lock in place. The simplest method is to use a ball peen hammer and, optionally, a tapered punch. First, the structural part and the flexure are assembled on a flat surface that has holes which allow the taper pins to protrude from the bottom. Next, a taper pin is inserted and pushed into place to temporarily hold the structure together and verify final alignment. Finally, the top of the pin is gently tapped until the assembler deems the pin secure enough for their application. If necessary, the punch can be used to drive the pin sub-flush; however, this is only advisable if the structural parts are thicker than the flexures (for example, using 12.7 mm aluminum plate and 10 mm nitinol). Removing taper pins is simply a matter of flipping the assembly over and tapping the pin free, again using the punch if necessary to free the flexure. Depending on the length of taper pin used, the dis-assembler must exercise caution to avoid bending the pin and accidentally locking it in place.

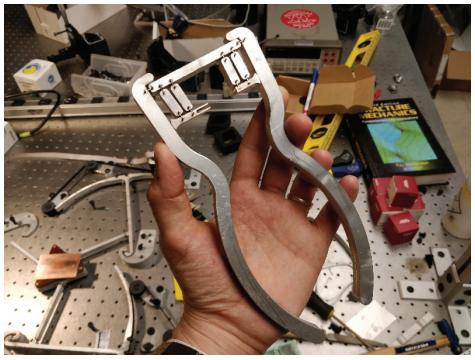
Assembling precision mechanisms with a hammer is not always favorable; in particular, the aforementioned method requires the use of a flat backup plate which may not be convenient. Another method is to use hand tools to apply precise force which gently pushes the pins home. Most "pinch-action" hand tools, such as channel-lock pliers, do not maintain parallel jaws throughout their range of motion, which quickly results in bent taper pins. However, the craft metalworking industry makes extensive use of parallel-jaw pliers that use sliding mechanisms to apply predictable parallel force across a 10 mm working distance. Using wire-EDM, a replacement jaw can be fabricated from 17-4 stainless steel which accommodates taper pins via a forked end, as seen in Figure 6-10. This arrangement can be conveniently used for both assembly and disassembly, provided the taper pins are long enough to remain proud of the structural elements.

An interesting extension of the custom jaw is a completely fabricated recursive assembly tool; in other words, a tool that can be used to assemble copies of itself. Such a design is also shown in Figure 6-10 and is made using the techniques outlined in this chapter:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.



(a) Pinsetter tool made by fabricating a custom lower jaw for a pair of common parallel-jaw pliers.



(b) Recursive pinsetter made with ten modular superelastic flexures and an aluminum frame.

Figure 6-10: Two "pinsetter" tools, used to quickly place and remove taper pins for securing modular superelastic flexures [31].

water-jetted aluminum plate for the handles, taper-reamed dovetail joints, wire-EDMed superelastic flexures, and taper pins (brass, in this case). The jaws themselves are also wire-EDMed out of 17-4 stainless steel, and take advantage of the material-agnostic nature of taper pin joints to join with the flexures in the same manner as the aluminum parts. This design is not particularly ergonomic (a bit of Plasti-Dip on the handles would go a long way), but it does work for assembly and disassembly of taper pin joints.

## 6.5 Fatigue

When designing around superelastic materials, the proverbial elephant in the room is fatigue life. By far the most common application for superelastic alloys is in the medical device industry, where they are used to fabricate highly collapsible stents that minimize incision size while maximizing arterial support [23]. In some cases, the effective fatigue cycle rate is the patient's heartbeat, since the stented artery expands with each blood pressure pulse and deforms the device. Such devices must reliably survive for the remaining life of the patient, which may end up being many millions of heartbeats. Unfortunately, the medical device industry is highly proprietary in nature and the methods used to guarantee high-cycle fatigue life in superelastic devices is a closely held trade secret. While academic literature points in the general direction of methods to increase fatigue performance, such as improving surface finish to reduce initial crack growth and carefully controlled heat treatment methods,


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

specific guidance is noticeably lacking [92].

During early testing it was noticed that modular superelastic flexures do occasionally fracture after repeated use. In some cases the reason is obvious; they are accidentally over-strained or twisted off-axis when a mechanism is relocated. However, in other cases the failure was clearly fatigue, so further study was required.

### 6.5.1 Testing

A simple automatic testing apparatus was fabricated to quantify flexure fatigue failures. The machine consists of a large hobby servo which rotates a lever arm, which then connects via bearings to a second and third arm, forming a 3-bar parallel linkage. The first and third bars are the same length, so the final linkage angle mirrors the servo rotation angle exactly. This final linkage is replaced with the superelastic flexure under test. To facilitate swapping in new flexures, the base and final linkage are quickly removable so the taper pins can be easily accessed as seen in Figure 6-11

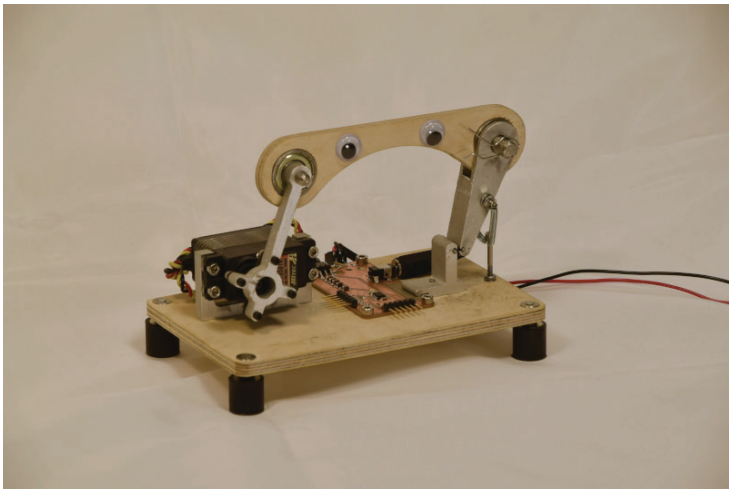


Figure 6-11: Automated fatigue tester used to evaluate modular superelastic flexures [31].

A Microchip 8-bit ATtiny412 microcontroller controls the servo motor, adjusting a pulse-width modulation (PWM) signal to rotate the actuator 20 degrees. Each time the servo cycles, the microcontroller sends the current count to a computer via a UART port and serial adapter. An automatic end-of-test function takes advantage of the conductivity of nitinol by monitoring electrical continuity across the flexure under test, and immediately stopping the servo when it detects an open circuit. Small tension springs attached to the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

third bar near the flexure ensure that it pulls away cleanly upon breakage.

Prior to building the automated testing apparatus, a conventional Instron 4411 material testing machine was used for the same measurements and produced limited qualitative results. However, the screw-driven nature of the Instron meant that the testing speed was quite slow; moving the test to the shop-built machine increased the testing speed twenty-fold to roughly 4 Hz, and added automatic end-of-test detection. Both of these changes dramatically improved experiment productivity and quality.

### 6.5.2 Experiments

As discussed above, the most commonly cited method for improving nitinol fatigue life is to improve surface finish. While raw wire-EDMed parts are dimensionally accurate to within a few microns and visually smooth, they are microscopically pitted and rough due to the periodic nature of the spark erosion process. A scanning electron microscope (SEM) micrograph of an as-machined flexure surface is shown in Figure 6-12.

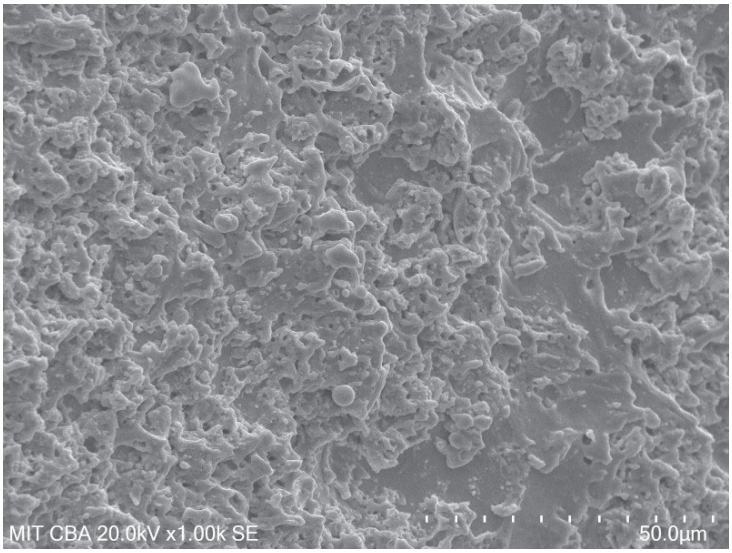



Figure 6-12: SEM micrograph of modular superelastic flexure as-machined surface, showing pitting and re-solidified debris from the wire-EDM process [31].

The first experiment using the Instron involved repeatedly pushing down on one side of a flexure with a polished plate, and using image analysis of a video recording to estimate the bend angle of the beam. During these tests, peak force was recorded and used to estimate the fracture point of the flexure, but determining the exact moment of failure was


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| Sample Type | Qty | Length | Cycles |
|---|---|---|---|
| Nitinol, 1 wire-EDM pass | 7 | 5 mm | 8223 ± 2510 |
| Nitinol, 1 wire-EDM pass, annealed | 2 | 5 mm | 9814 ± 280 |
| Nitinol, 2 wire-EDM passes | 5 | 5 mm | 9943 ± 3357 |
| Nitinol, 3 wire-EDM passes | 4 | 5 mm | 7881 ± 796 |
| Nitinol, 1 wire-EDM pass | 4 | 6.5 mm | 16,373 ± 4579 |
| Nitinol, 1 wire-EDM pass | 2 | 10 mm | 17,875 ± 2181 |
| 6061 Aluminum, 1 wire-EDM pass | 1 | 6.5 mm | 1504 |
| 17-4 Stainless, 1 wire-EDM pass | 1 | 6.5 mm | 1970 |

Table 6.2: Fatigue testing flexures with various characteristics.

highly subjective. Generally, the force curves decreased quickly by 10-15% and held steady for some time, and then gradually fell off to zero. The first test examined three samples deflected to various angles, while the second compared a raw flexure to one that has been hand-polished with a 400-grit stone. In both cases, drawing conclusive results is difficult; but it does appear that no order-of-magnitude changes in fatigue life occur, with all of the flexures failing between 1000 and 3000 cycles.

A second round of experiments were performed once the aforementioned testing apparatus was fabricated. These tests surveyed a variety of strategies for improving the fatigue life of the flexures, including lengthening the beams while not changing the bend angle; adding wire-EDM finishing passes, a common method for reducing surface roughness with the fabrication method; and annealing the machined flexures in an atmospheric tube furnace at 550 C for 30 minutes, which are typical values for such work [25]. For comparison's sake, flexures were also fabricated out of 6061 aluminum and 17-4 stainless steel. A complete table of results is shown in Table 6.2.

Clearly, additional wire-EDM passes and annealing had no statistically significant effect on fatigue life. Increasing the flexure length appears to have improved performance to a degree; however, given the jump from 5 to 6.5 mm, it is reasonable to hypothesize that this trend should continue to longer lengths which is not reflected in the data.

### 6.5.3 Next Steps

Fortunately, while $10^4$ cycles to failure is inadequate for a commercial product such as a medical stent, it is more than sufficient for building and testing flexural machines in a limited capacity, as seen in the following chapter. As such, continued work to improve the fatigue


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

life of modular superelastic flexures can occur in parallel with mechanism development and characterization.

A clear first step is to quantify the surface roughness of current wire-EDMed flexures using a micron-scale technique such as confocal microscopy. While the Sodick SL400G used here has a built-in tool to predict roughness for a given number of wire-EDM passes, the actual surface quality of multi-pass cuts should also be examined. If the value does not markedly improve (for reference, the machine estimates a 4-fold improvement in roughness average for three passes), machining parameters should be adjusted to improve that result prior to fatigue testing.

One technique for improving nitinol surface roughness and fatigue life cited in literature [57] is electropolishing. This method was previously used at the CBA by Prashant Patil [64] for deburring laser-machined MEMS structures, and a similar regime exists with wire-EDMed surfaces: micron-scale debris and sharp edges that must be smoothed down. This specific process is also called anodic leveling, and occurs because current densities are increased at sharp corners and peaks. Unfortunately, electropolishing is generally only useful for surfaces that are already quite smooth, so this technique will likely need to be combined with a mechanical polishing step if wire-EDM parameter adjustments are not able to sufficiently reduce surface roughness.

Due to limited equipment availability, annealing was only explored briefly and did not produce a statistically significant shift in fatigue life. But nitinol can be annealed across a broad range of temperatures [25], so a wider sweep of post-machining annealing conditions should be performed. Another note is that annealing can effectively reset lingering defects from the phase transformation that enables superelastic behavior. As such, a worthwhile experiment will be to cycle a flexure below its expected breaking point (i.e 2-5000 cycles), then perform an annealing cycle, and then cycle the flexure for the same number of times again. If this process can be repeated indefinitely, an interesting option could be to electrically anneal flexures in situ: since nitinol is electrically conductive, part of routine maintenance of a mechanism could be to simply run current through the joints to heat them up for a period of time to reset their fatigue life.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 7

# Compliant Machines

Compliant machines are assemblies of mechanisms built from flexural elements. This chapter presents three such machines of varying complexity which were used to characterize the performance of the previously described modular superelastic flexures, and to demonstrate their potential for use both in the electronics micro-assembly system proposed in Chapter 5 and in a community FabLab environment.

## 7.1 Single-Axis Flexure Test Machine

### 7.1.1 Description

The first machine is a four-bar flexural linkage driven by a stepper motor. This initial development iteration was used to validate the modular superelastic flexure concept generally, and to test for backlash (or lack thereof) and mechanical repeatability. Flexures are ideal for mechanically dividing using simple lever arms; in this case, the reduction ratio of the lever arm is 5:1, with the longer portion of the arm measuring 200 mm in length. A diagram the machine is shown in Figure 7-1, and an assembled image is shown in Figure 7-2.

A NEMA14 stepper motor drives a GT2 timing belt which provides the motive force to move the linkage. The far end of the long lever arm is an arc that sweeps about the flexural pivot through 90 degrees, onto which both ends of the timing belt are secured with an adjustable 3D-printed clip. Two idlers are arranged such that throughout a ±20 degree swing the timing belt stays tangent to the arc; in this way, the curvature functions as a wedge sliced from a circle, providing an additional mechanical reduction from the stepper's


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

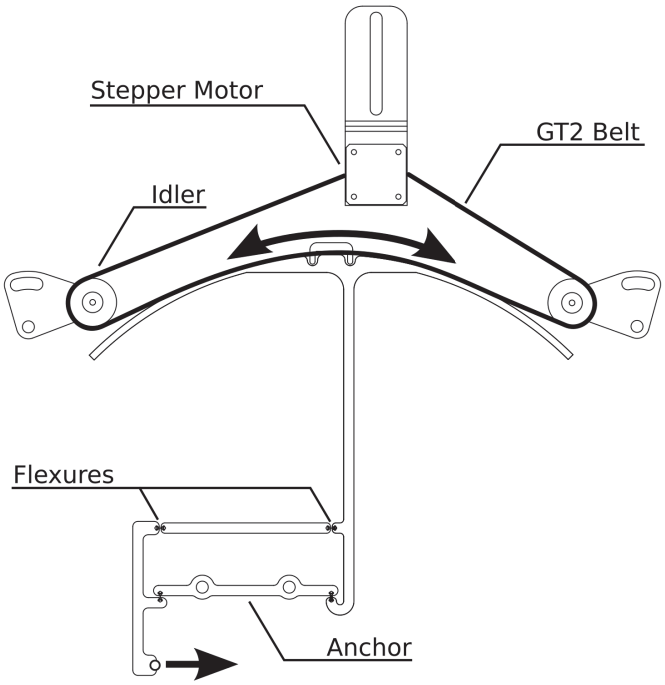



Figure 7-1: Actuator test machine diagram showing flexures, motor, idlers, belt, and anchor. Motion is indicated with arrows.

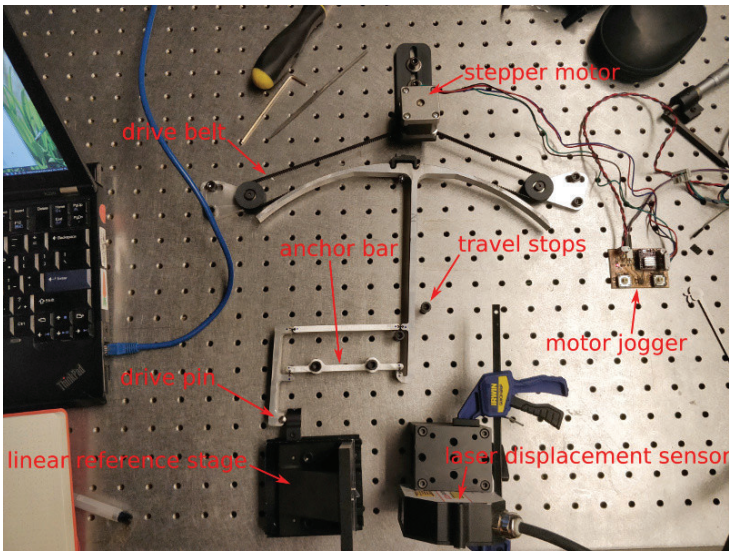



Figure 7-2: Annotated image of actuator test machine [31].

rotation. The constant-radius arrangement means the overall linkage reduction ratio does


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

not change through the mechanism's range. It also means the required belt length is fixed, so tensioning is simply a matter of rotating and securing the idlers.

A commercially sourced precision linear stage is positioned near the short lever arm with its motion axis parallel to the anchor linkage. The short lever arm applies force to the stage using a press-fit hardened dowel pin; this joint rotates freely to avoid over-constraining the flexural mechanism against the linear stage. The linear stage holds a steel target block with is used as a target for a Micro-Epsilon NCDT2300 laser displacement sensor, a non-contact measurement instrument with 150 nm resolution and 2 um linearity across its 10 mm working range. A laptop connects to the sensor via Ethernet to provide a real-time data readout and data logging for offline analysis.

Stepper motors are more complicated to drive than DC brushed motors, which simply require a constant voltage supply to spin. In this case, the stepper motor connects to a simple PCB controlled by an ATtiny412 microcontroller. Two switches mounted to the board allow the experimenter to jog the motor in either direction by a fixed number of steps defined in the microcontroller's code, or sweep continuously in single-step increments interspersed with quarter-second delays. The motor itself receives drive pulses from a dedicated stepper driver module made by Pololu which uses a Texas Instruments (TI) integrated circuit called the DRV8825.

The structural elements, such as linkages and anchors, were waterjet cut out of 12.7 mm aluminum plate, taper-reamed, and assembled with 10 mm wire-EDMed modular superelastic flexures as described in the previous chapter. The anchor linkage and linear stage were bolted to a flat optical table, and the laser displacement sensor secured using a magnetic mount.

### 7.1.2 Analysis

The stepper uses a 16-tooth pinion whose pitch diameter can be calculated as:

$$D_{pitch} = \frac{N_{teeth} \times P}{\pi}$$

where $N_{pitch}$ is the number of teeth on the sprocket and $P$ is the belt pitch. For a 2 mm pitch GT2 belt, the stepper's pinion thus has a pitch diameter of 10.19 mm. The resulting reduction ratio can be calculated as:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

$$X = \frac{R_{arm}}{R_{pinion}}$$

which solves to 39.25. Stepper motor rotation is broken into discrete steps based on the internal arrangement of permanent magnets and armature coils. The low-cost hobby motors used here have 200 steps per rotation, or:

$$\frac{360}{200} = 1.8 deg/step$$

Because of the mechanical reduction provided by the belt, the long lever arm thus rotates about its flexural pivot at:

$$A_{rotation} = \frac{1.8}{39.25} = 0.04586 deg/step$$

The 4-bar linkage uses equal-length opposing bars, so the mechanism is a simple parallelogram where diagonally opposed pairs of included angles are identical. Thus, the angle of the short lever arm which pushes on the linear stage is the same as the long lever arm's rotation. Since the end of the short lever arm rotates around a pivot, its path through space is not linear; as such, the stage does not move the same amount with each step.

When the flexures are straight and the long lever arm is exactly perpendicular to the linear stage, all four included angles of the linkage are 90 degrees. In this case, the short lever arm effectively only moves parallel to the linear stage. The predicted displacement of the stage can be calculated as:

$$Y = R_{arm} \times \sin A_{rotation}$$

In this case, $R_{arm} = 40mm$, so $Y = 32.02um$. The displacement of the stage per step will drop as the flexure angle becomes more extreme and cosine error is introduced. At 5 degrees, for example, the expected displacement per step is:

$$R_{arm} \times (\sin 5.04586 - \sin 5.00000) = 31.89um$$

so the error introduced by the rotating arm is 0.13 um. This error is on the order of the resolution of the laser displacement sensor.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### 7.1.3 Evaluation

First, the mechanism was evaluated for linearity by single-stepping the motor across the laser displacement sensor's 10 mm working range. A zoomed in detail of distance data from the sensor is shown in Figure 7-3. Notably, the mechanism shows significant "ringing" at the beginning of each step, later traced back to the spring loaded stage skipping off the dowel pin and slapping back down with each stop-start cycle. A simple Python script was used to convert this raw data into distinct step values. The script iterated through the data set and flagged any jump greater than 10 um in a single time step, and then threw out the following ten data points to eliminate ringing effects. The remaining values at each step were then averaged and the change with each step recorded. A histogram of the data is shown in Figure 7-4.

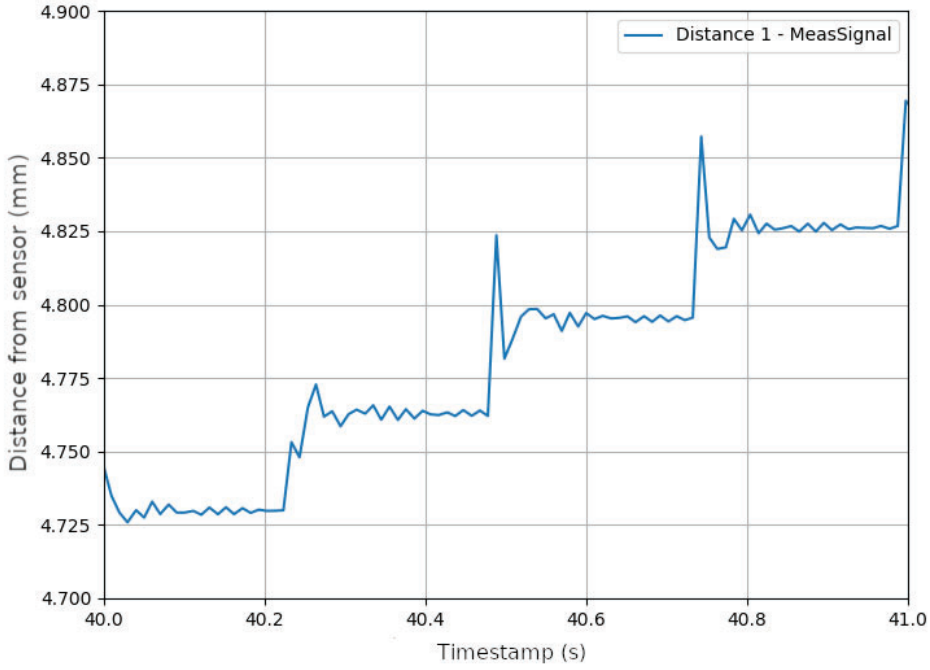



Figure 7-3: Several 250 ms steps of the linear actuator as measured with the laser displacement sensor, showing substantial ringing [31].

Clearly, measurement noise overwhelms the expected sub-micron error described above. Averaging the entire range of data results in a per-step displacement of $31.8 \pm 1.3um$ which


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

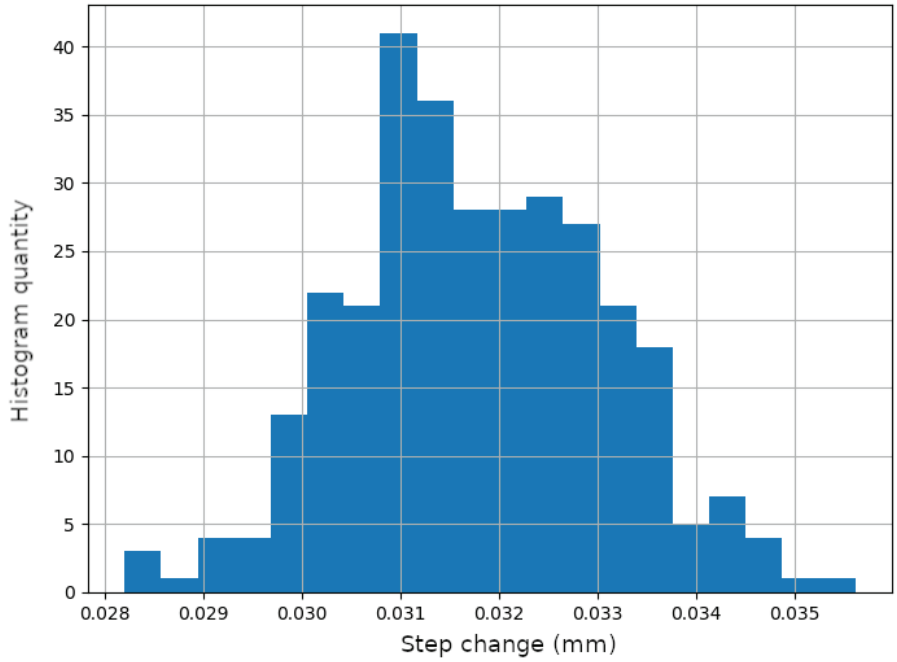



Figure 7-4: A histogram of the actuator displacement step size across the laser displacement sensor's 10 mm range, showing a narrow normal distribution centered at 31.8 um [31].

agrees closely with the predicted values. The relatively large standard deviation of the data could be the result of inadequate settling time for the ringing; alternatively, this could be a measurement-related error due to an imperfect target surface for the laser displacement sensor. Validation of the metrology strategy using a better measurement scheme such as a precision LVDT or an interferometer should be performed prior to assuming this result is due to unpredictable flexure behavior.

Subsequent tests were run in which the machine was cycled several steps from either direction towards a reference value to detect backlash. As a result of signal averaging, the variability of the measurement fell well short of the 150 nm device resolution again suggesting that backlash, if present, would require more precise tools to resolve.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 7.2 3-RRR CPM

The second machine is a prototype of the planar motion system that will be used to assemble the electronic glitter lattices discussed previously. The desired work area is in the range of a centimeter or two; this roughly lines up with the maximum displacement of the first machine, so several parts of its design was reused in triplicate to avoid repeated work. In addition to further demonstrating the feasibility of modular superelastic flexures as a machine-building component, this experiment aims to characterize the repeatable distortion caused by the non-ideal nature of flexural pivots; that is, the imperfect kinematics caused by non-static virtual rotational centers. The machine is also tested for stiffness in several directions, benchmark values that can be used to compare future iterations.

### 7.2.1 Description

This machine falls into a category of mechanisms called compliant parallel manipulators, or CPMs [40]. Its architecture draws on previous work [92], changing the modular flexure design and replacing the prismatic actuator with a rotary joint. The system is compliant, because it uses flexures as opposed to sliding or rolling contacts; and it is parallel, because the three actuators act simultaneously on the same output stage, versus robotic arms which are serially oriented strings of rotary joints. Each linkage uses three rotary linkages, including the drive linkage; and there are three independent linkages total, hence 3-RRR. The flexures constrain motion in the Z-axis or rotation around the X- or Y-axis, but the final stage is able to translate along the XY plane and rotate around the Z-axis. The three actuators fully constrain these three degrees of freedom. A diagram of the machine is shown in 7-5, and a photograph of the system assembled on a flat optical bench is shown in 7-6.

As with the first machine, structural components were waterjet cut from 12.7 and 6.3 mm aluminum plate, and the nine modular superelastic flexures were wire-EDMed out of 10 mm nitinol. The build plate itself was milled out of 7 mm phenolic, and mounts to the machine's stage using three 250 um/turn micro-adjustment screws from Kozak Micro [4]. The hardened ball tips of the three micro-adjusters interface with three D-shaped inserts press-fit into the aluminum stage. These inserts have 90 degree grooves cut in the top face which are oriented at 120 degree angles with respect to each other, forming a Kelvin-type kinematic coupling. The inserts were wire-EDMed in two operations out of pre-hardened


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

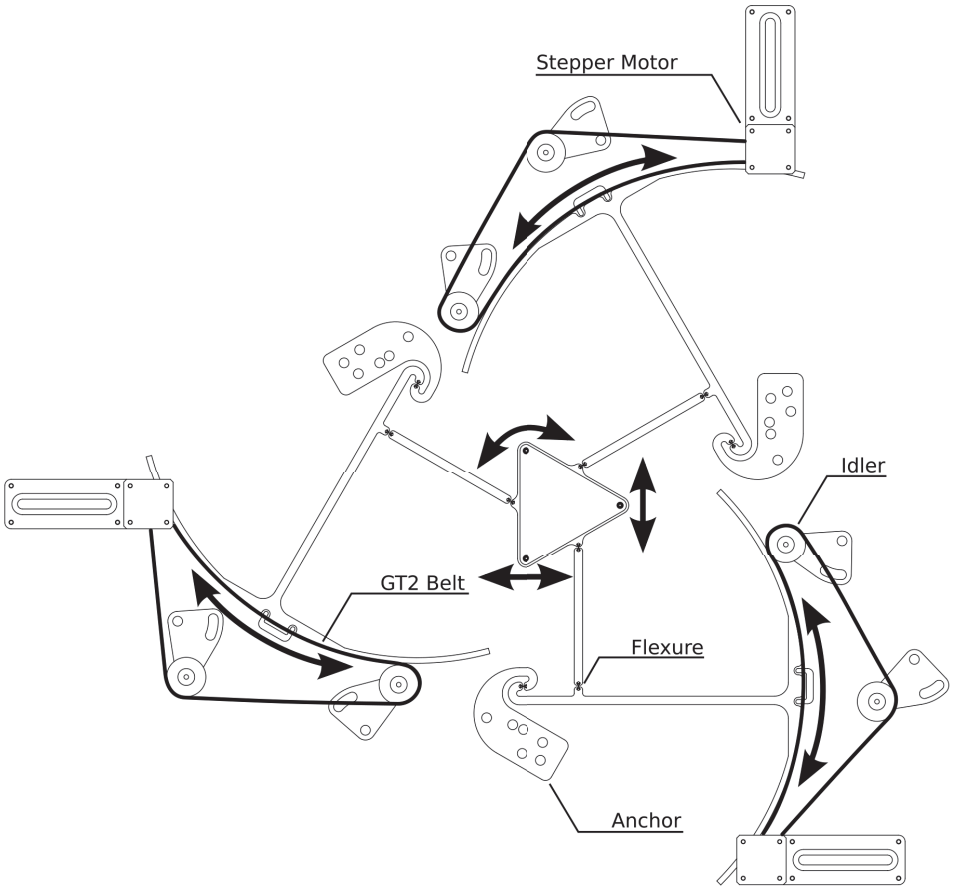



Figure 7-5: 3RRR CPM diagram showing flexures, motors, idlers, belts, and anchors. Motion is indicated with arrows.

4140 steel. The stage is kept in place by means of a pair of opposing 12.7 mm neodymium magnets, one press-fit into the stage and the other epoxied into a hole in the phenolic build plate. This arrangement allows the build plate to be quickly removed so samples can be mounted, and then replaced in a repeatable manner and leveled as needed using the micro-adjusters. A detailed view of the build plate and kinematic mount is shown in Figure 7-7.

### 7.2.2 Inverse Kinematics

Determining where the three actuators should go to produce a desired stage position (and angle) in Cartesian space means solving the inverse kinematic equations that describe the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

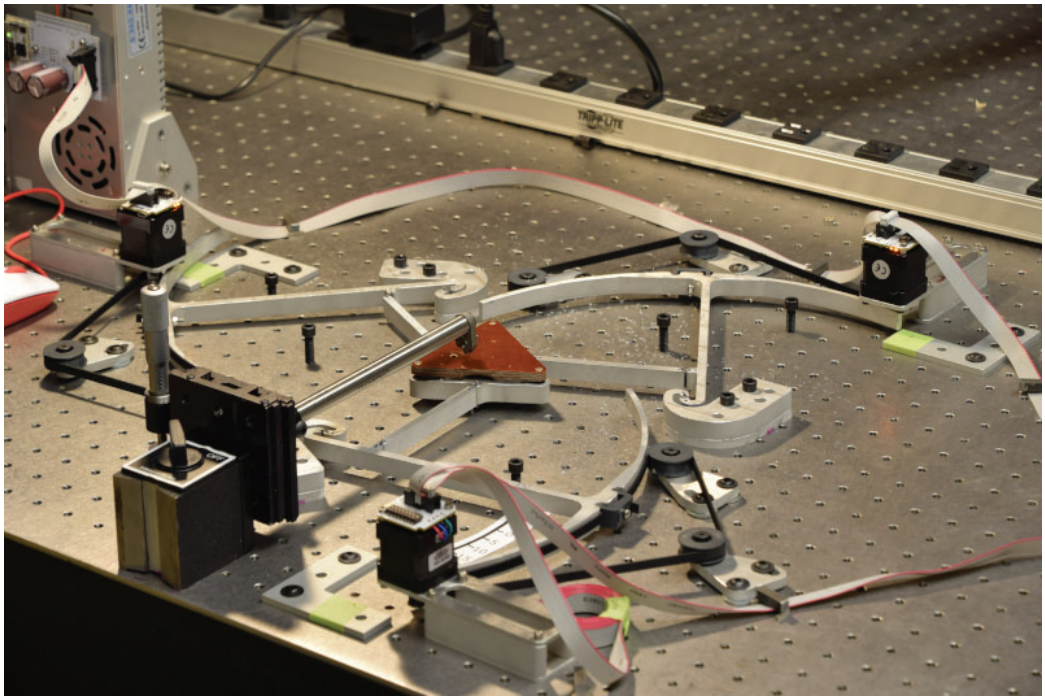
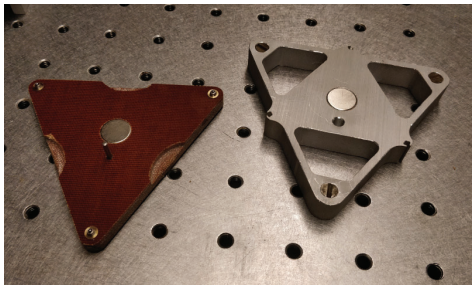
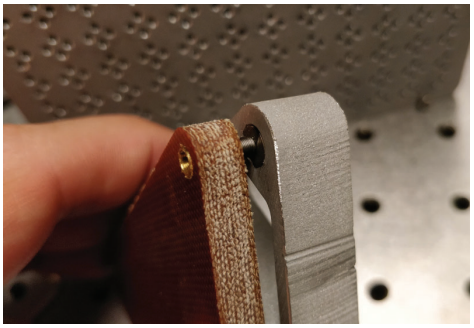


Figure 7-6: Image of installed 3RRR CPM, shown with control circuitry, build plate, and grating tool [31].



(a) Build plate (left) and stage (right) disassembled, showing kinematic inserts, magnet, and micro-adjusters.



(b) Close-up view of build plate mounted to stage, showing interface between micro-adjusters and kinematic inserts.

Figure 7-7: 3RRR CPM kinematic stage detail views. [31].

link system. The straightforward way to determine these values is via algebraic methods, discussed in [90] and used here.

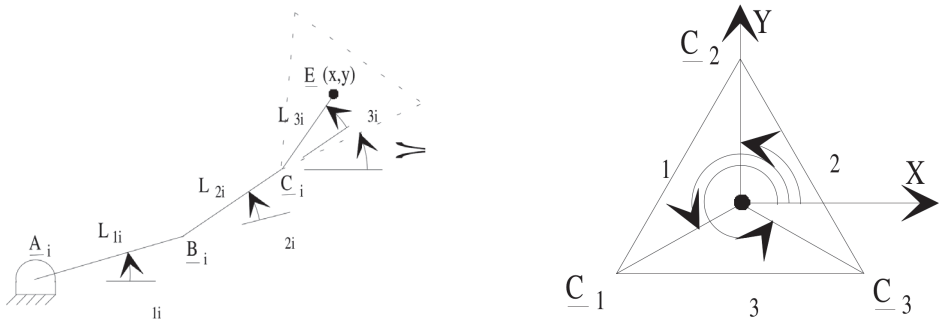
The 3-RRR CPM can be divided into three identical linkages, anchored on one side and connected on the other to the common stage. A diagram of each arm, along with parameter naming conventions and the location of the origin, is shown in Figure 7-8. The values for


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| Parameter | Value | Description |
|---|---|---|
| $\theta_{max}$ | 15 degrees | maximum allowable joint angle |
| $L_1$ | 40 mm | first link length |
| $L_2$ | 100 mm | second link length |
| $L_3$ | 33.6953 mm | distance from stage pivot to origin |
| $A_{1x}$ | 124.9124 mm | x position of first actuator pivot |
| $A_{1y}$ | 43.6453 mm | y position of first actuator pivot |
| $A_{2x}$ | -24.6582 mm | x position of second actuator pivot |
| $A_{2y}$ | 130.0003 mm | y position of second actuator pivot |
| $A_{3x}$ | -100.2545 mm | x position of third actuator pivot |
| $A_{3y}$ | 86.3546 mm | y position of third actuator pivot |
| $\psi_1$ | 57.0849 degrees | angle between x-axis and first link stage pivot |
| $\psi_2$ | 297.0849 degrees | angle between x-axis and second link stage pivot |
| $\psi_3$ | 177.0849 degrees | angle between x-axis and third link stage pivot |
| $\theta_{1_1}$ | 120 degrees | angle between x-axis and actuator 1 |
| $\theta_{1_2}$ | 0 degrees | angle between x-axis and actuator 2 |
| $\theta_{1_3}$ | 240 degrees | angle between x-axis and actuator 3 |

Table 7.1: Parameters used in the 3-RRR CPM inverse kinematics model [31].

each parameter were pulled from measurements in the CAD model, and are shown in Table 7.1. Note that the length values were all calculated to the nominal center of rotation for each flexure.




(a) Figure 3 from [90], showing the kinematic diagram for a single arm of the 3RRR CPM.

(b) Figure 4 from [90], showing the rotation values $\phi$ for the three arms.

Figure 7-8: 3RRR CPM kinematic diagram from [90]. Used with permission.

Starting from an initial desired location in Cartesian space $\{x, y, \theta\}$, the stage corner


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

locations $C_{ix}$ and $C_{iy}$ are calculated with respect to the origin for each link $i$:

$$C_{ix} = x + L_{3i}\cos(\phi + \psi_i)$$
$$C_{iy} = y + L_{3i}\sin(\phi + \psi_i)$$

Next, the two possible angle solutions ("elbow-up" and "elbow-down") are calculated using the half-angle substitution method [52]:

$$\theta_{1i_{1,2}} = 2\tan^{-1}\left(\frac{-F \pm \sqrt{E^2 + F^2 - G^2}}{G - E}\right)$$

where the intermediate values $E$, $F$, and $G$ are:

$$E = 2(C_{ix} - A_{ix})L_{1i}$$
$$F = 2(C_{iy} - A_{iy})L_{1i}$$
$$G = L_{2i}^2 - L_{1i}^2 - (C_{ix} - A_{ix})^2 - (C_{iy} - A_{iy})^2$$

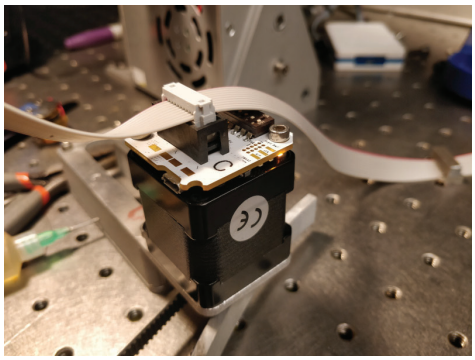
A Python script is used to iteratively call these functions on a given pandas dataframe of desired coordinates. Before storing the resulting angle values, the two possible solutions are compared to the maximum joint angle and the appropriate one is chosen. If the desired coordinates over-rotate a joint in both solutions, the script throws an error and stops.

### 7.2.3 Control

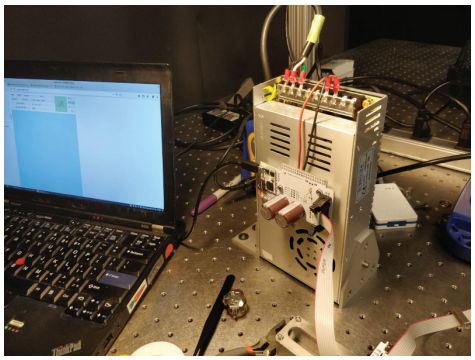
This machine makes extensive use of Jake Read's Clank [70] platform, which he developed for the remote pandemic version of the class How to Make (Almost) Anything. Clank is a desktop 3-axis Cartesian CNC router designed for fabricating small precise things like circuit boards and wax molds. The machine control architecture is unique, and builds on the work from [71]. The stepper motors are connected in a stateless dataflow network controlled by a web browser. Electrically, this takes the form of modules wired to each stepper that are then connected with a long piece of ribbon cable using insulation-displacement connectors (IDCs). This chain of control "nodes" is then connected to a "head" PCB, which connects via USB to the host computer (see Figure 7-9). A large 24 V DC power supply provides motor power to the nodes, while the USB connection provides logic power. Optionally, the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

motors can be equipped with shaft magnets that are picked up by a PCB-mounted high resolution magnetic encoder; with an added control loop, this allows them to function as closed-loop servo motors.



(a) A close-up of one NEMA14 stepper motor with control PCB and IDC cable connection.



(b) Head PCB mounted to power supply with control computer to the left.

Figure 7-9: 3RRR CPM control system images [31].

In this adaptation of Clank, the three motors corresponding to the three compliant linkages are connected and reprogrammed to identify as the X, Y, and Z axes. The Clank web user interface (UI) is used as-is; it includes buttons for quickly jogging the machine along an axis, and a window for loading Gcode files for fabrication.

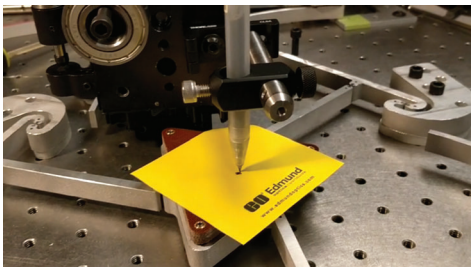
### 7.2.4 Evaluation

The machine was first tested with a pen writing on a sticky note adhered to the build plate. The pen was held vertically using a right-angle optical post adapter with a thumb screw, which was then screwed into a commercial linear stage. The stage was then held vertically with a magnetic clamp, such that the pen was directly over the sticky note. The stage includes a bias spring to hold it against a micro-adjuster, which was installed such that the spring tended to pull the pen off the paper. A series of heavy bearings were then hung on the stage, allowing for precise force control.

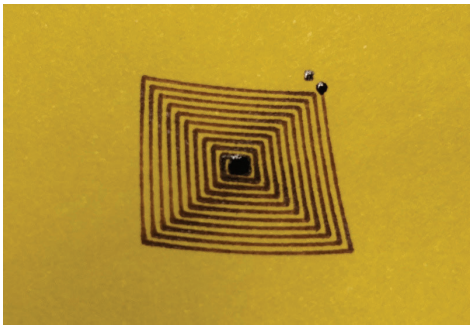
An additional function added to the Python script produced a simple square spiral pattern, which was then run through the inverse kinematics equations to produce stepper angles. These values were then written to a text file with the appropriate Gcode syntax characters and saved as a text file. A scaling value was also added to the inverse kinematics function so the same test could be executed at different scales. Figure 7-10 shows the setup


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

and resulting drawn pattern from the pen. As expected, the spiral is clearly visible but has some distortion caused by the non-ideal rotational characteristics of the flexure joints.



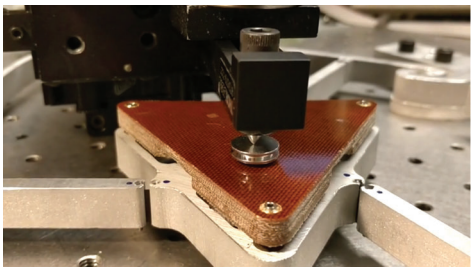

(a) Pen spiral test setup showing stage, pen holder, and bearings for weight.



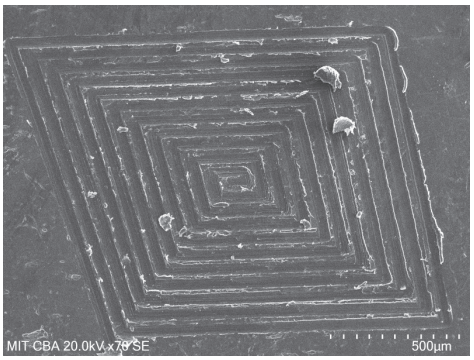
(b) Drawn square spiral showing some distortion. Spiral is roughly 15 mm across.

Figure 7-10: Testing the 3RRR CPM with a pen [31].

Next, the same spiral pattern was scaled down by a factor of ten, resulting in an overall width of roughly 1.5 mm. The pen was replaced with a sharpened bolt, and a small hole was drilled in the build plate to accommodate a press-fit aluminum SEM sample holder. This setup and resulting pattern micrograph are shown in Figure 7-11. Again, there is some expected distortion present from the non-ideal joints. The distortion does appear to line up with the pen drawing, suggesting the distortion may scale up and require correction only at the macro level.



(a) Bolt test setup showing stage and bolt holder.




(b) Scribed square spiral showing similar distortion to the drawn version.

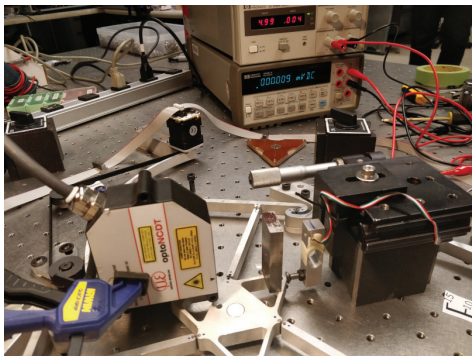
Figure 7-11: Testing the 3RRR CPM with a sharpened bolt [31].

The machine was also characterized for static stiffness as a benchmark for future iterations. To do this, the control system was first powered on to lock the motors. Next, a 5 kg

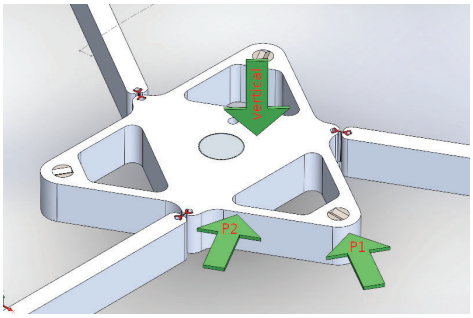

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

capacity load cell was attached to a linear stage with a micrometer adjustment mechanism, which was then bolted to a magnetic anchor. A bolt was attached to the other side of the load cell such that force could be applied to objects from various directions depending on the orientation of the magnetic base. Before use, the load cell was calibrated by applying a 5 V DC bias voltage to two terminals and measuring the output voltage with no load, and again with 500, 1000, and 1500 g calibration weights hanging from the module. Fitting a linear regression to the data returned a load cell sensitivity of 1.0081 V/g.

To measure the displacement of the stage under various loads, the laser displacement sensor was affixed to another magnetic base and aimed at a steel target temporarily secured to the magnet at the center of the stage. The load cell apparatus was then oriented to apply force to the stage in two directions, one parallel to a side of the stage and one normal to the face. Vertical stiffness was characterized by turning the laser displacement sensor on end and adding weights to the stage. An image of the setup and the test directions is shown in Figure 7-12.



(a) Stiffness characterization setup showing load cell, laser displacement sensor, bias voltage supply, and readout voltmeter.




(b) Diagram showing three stiffness test axes. Planar tests were performed with the load cell, while vertical tests simply used weights.

Figure 7-12: 3RRR CPM stiffness characterization setup [31].

The horizontal stiffness tests were recorded in roughly 100 mV increments, corresponding to approximately 1 N of force. The vertical tests used the same 500 and 1000 g calibration weights as were used to calibrate the load cell. Resulting plots with fitted linear regressions are shown in Figures 7-13, 7-14, and 7-15.

The regressions visually fit the data well. Taking their slopes, the stiffness in direction P1 is 143 N/mm; P2 is 237 N/mm; and vertical is 110 N/mm.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

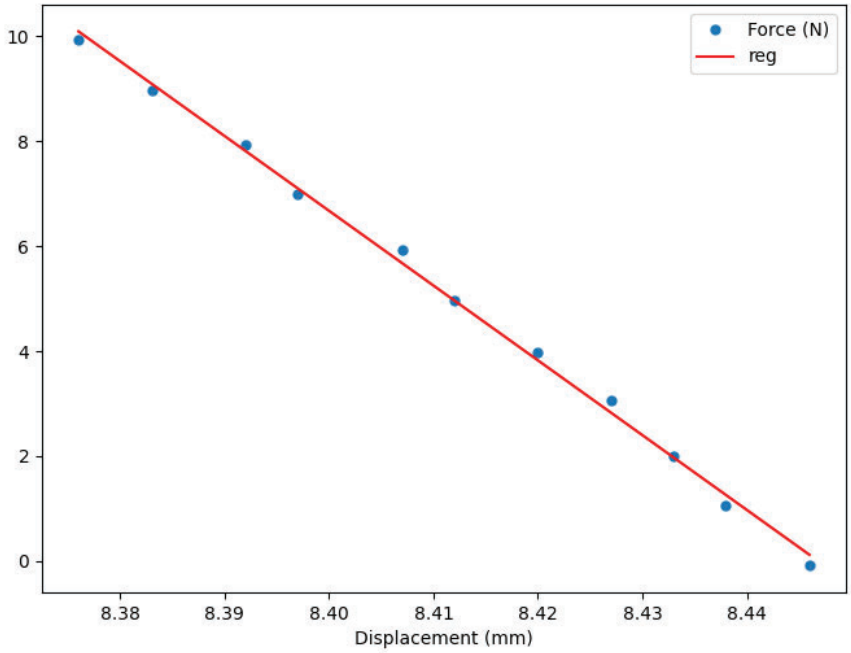



Figure 7-13: 3RRR CPM planar stiffness test results in direction P1 [31].

### 7.2.5 Computer Vision System

A computer vision system was installed in order to characterize and correct the distortion caused by non-ideal flexural pivots. The system consists of a Raspberry Pi 4 with 8 GB of RAM running Ubuntu 21.04; a Raspberry Pi High-Quality Camera with 12.3 megapixel resolution [30]; and a 50 mm CS-mount lens. The camera and computer were mounted to a rigid frame fabricated from modular extrusion.

The angular field of view $AFOV$ of a lens is:

$$AFOV = 2 \times \arctan(\frac{H}{2f})$$

where $H$ is the sensor size and $f$ is the focal length of the lens. The Sony IMX477 in the Raspberry Pi High Quality Camera measures 7.9 mm diagonally; thus, the lens and camera system has an angular field of view of 9.0 degrees. The field of view $FOV$ can then


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

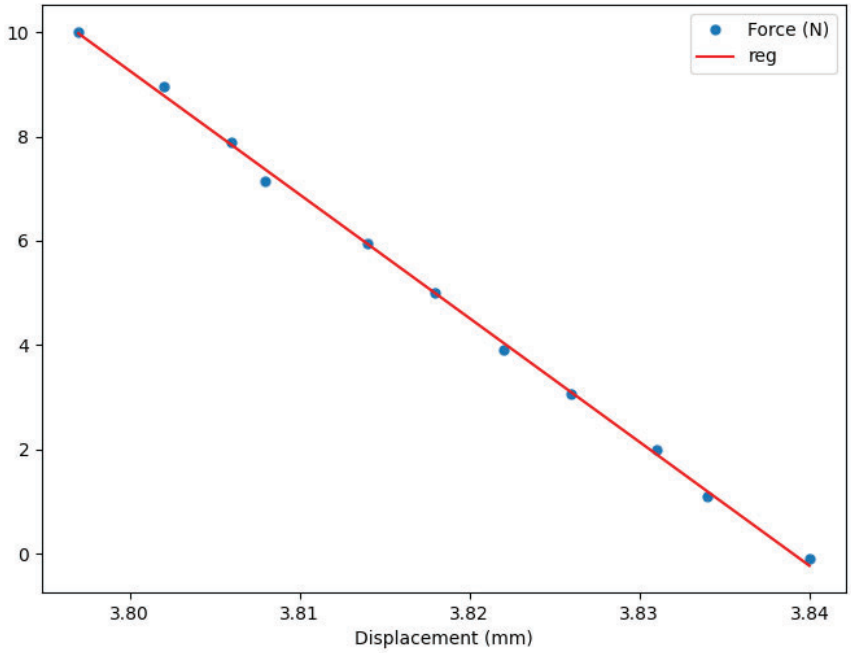



Figure 7-14: 3RRR CPM planar stiffness test results in direction P2 [31].

be calculated as:

$$FOV = 2 \times D_{working} \times \tan(\frac{AFOV}{2})$$

where $D_{working}$ is the working distance between the lens and the stage. For this setup, $D_{working}$ is approximately 400 mm, giving a diagonal field of view of 63.2 mm. Similarly, the effective resolution can be calculated from the IMX477's pixel size of 1.55 um square; in this case, the system should be able to resolve features down to 12.4 um.

In order to automatically detect the precise stage location and orientation, fiducials called ArUcO markers were used [34]. These markers are easy to scale and are often implemented for multi-object tracking; in this case, only a single marker was needed. A Python script was used to generate a valid marker pattern which was then raster-engraved on a polished piece of copper-clad FR1 circuit board material using a 532 nm nanosecond pulsed laser micro-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

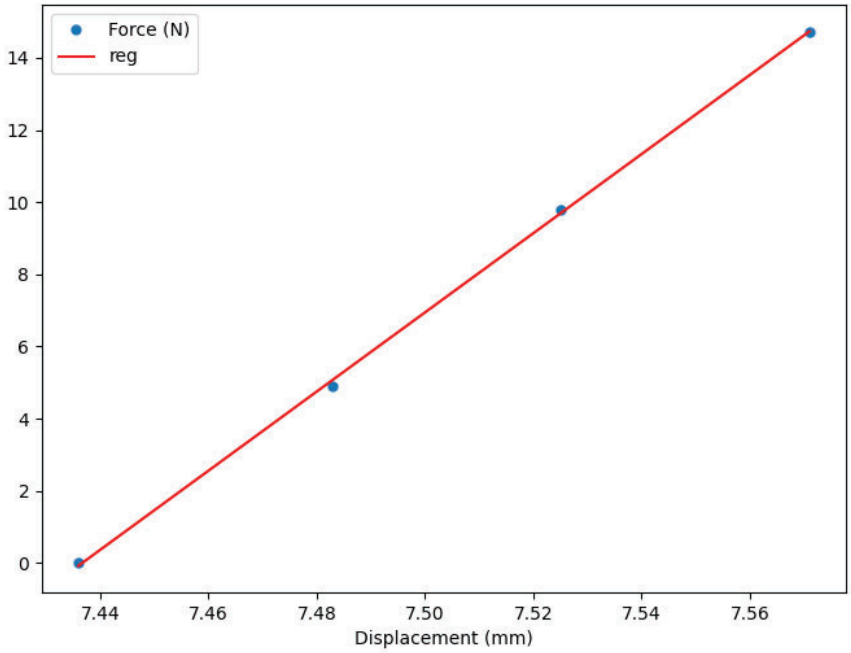



Figure 7-15: 3RRR CPM vertical stiffness test results [31].

machining system. OpenCV [8], an open-source computer vision application, was then used to identify the center location of the marker. An image from the computer vision system with an identification box around the ArUcO marker is shown in Figure 7-16.

To check the performance of the computer vision system, the 3RRR CPM was temporarily removed and replaced with a commercial linear stage, driven by a micrometer. The previously discussed laser displacement sensor was aimed at the stage to precisely measure its position. This setup is illustrated in Figure 7-17.

The linear stage was advanced in roughly 1 mm increments, and the laser displacement sensor data manually recorded and a corresponding image acquired. Plotting sensor data against ArUcO data yielded the expected straight line, as shown in Figure 7-18.

This process was repeated four additional times, covering most of the working area of the stage. The results are shown in Table 7.2, and suggest that the computer vision system is functioning at the level of performance predicted by the resolution calculations above.

Unfortunately, due to shortness of time the computer vision system was not utilized for


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

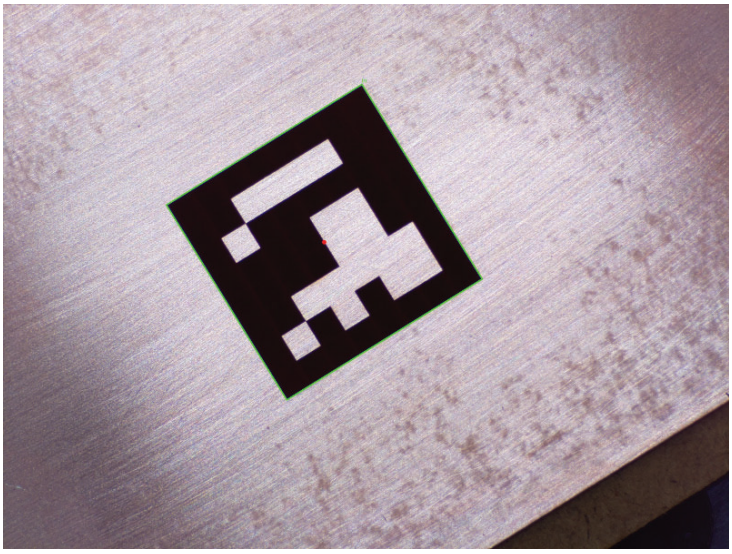


Figure 7-16: ArUcO marker laser-engraved on copper, as imaged and identified by the computer vision system [31].

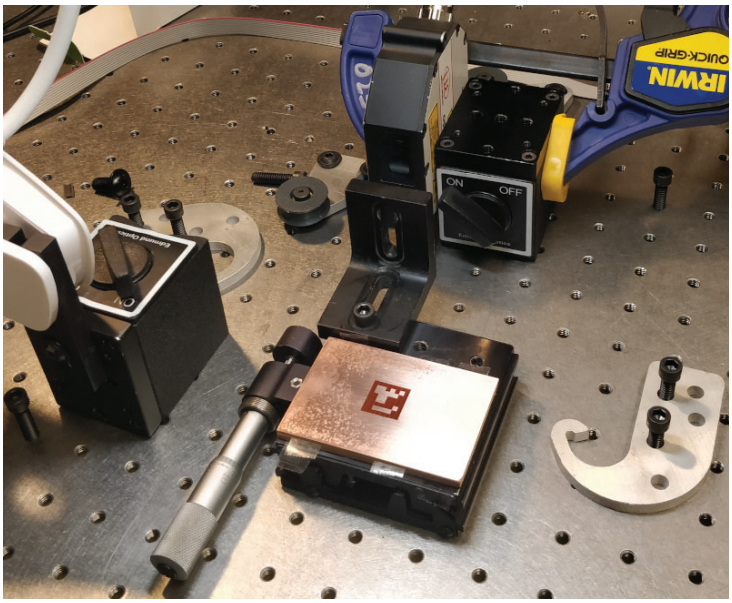



Figure 7-17: ArUcO calibration setup with linear stage and laser displacement sensor [31].

removing distortion from the motion system.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

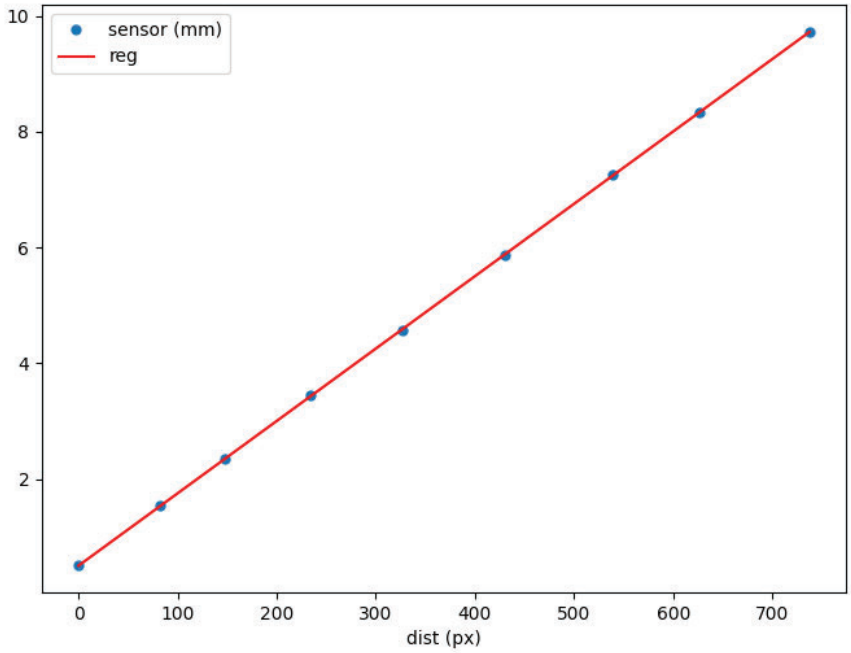



Figure 7-18: ArUcO calibration plot comparing laser displacement sensor values to computer vision measurements [31].

### 7.2.6 Ruling Diffraction Gratings

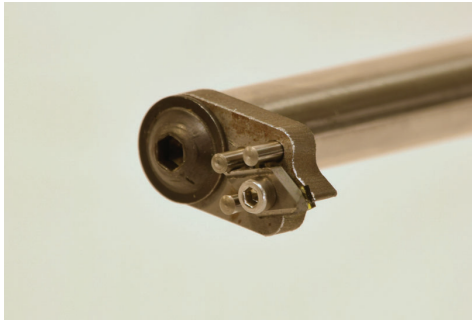
The 3RRR CPM was also evaluated qualitatively, to see if was sufficiently precise to mechanically rule a low-resolution diffraction grating capable of breaking white light into its constituent parts. In total, nine attempts were made with a variety of line spacing parameters and tool choices. Best results were obtained using a scrap of CVD diamond held in a purpose-built tool holder. This tool and a freeze-frame of the "ruling engine" in action is shown in Figure 7-19.

The fifth grating, with a groove spacing of roughly 40 um (or 25 lines/mm), proved functional if not particularly efficient or useful. Several images and SEM micrographs of the grating are shown in Figure 7-20. Notably, the head-on micrograph shows a pattern of micron-scale "waves", likely a result of individual motor steps at the resolution limit of the machine.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| Test | Location | Direction | Slope ($\mu$m/px) | Error (stdev, $\mu$m) |
|---|---|---|---|---|
| 1 | top left | horizontal | 12.50 | 8.34 |
| 2 | bottom right | horizontal | 12.50 | 6.15 |
| 3 | bottom left | vertical | 12.52 | 6.95 |
| 4 | top right | vertical | 12.57 | 10.53 |
| 5 | center | diagonal | 12.49 | 7.48 |

Table 7.2: Results from 3RRR CPM stage calibration tests [31].



(a) Grating ruling tool, initially a sharp carbide insert, later the piece of CVD diamond scrap shown here.

(b) Freeze-frame from grating ruling video showing tool scratching fine grooves into an aluminum SEM sample holder.

Figure 7-19: Ruling a primitive diffraction grating using the 3RRR CPM and a diamond tool [31].

## 7.3 MicroPanto

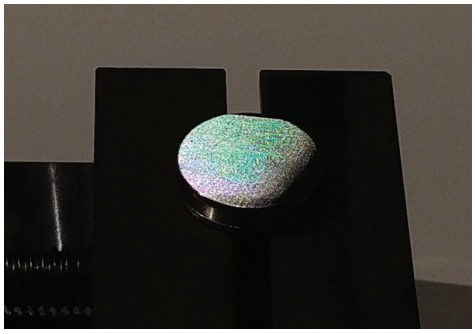
The third machine is a 50:1 pantograph reducer, designed shrink the working area of any desktop-scale CNC router with a corresponding decrease in resolution. MicroPanto [32] was designed and fabricated in roughly ten days, with half that time spent at the CBA preparing flexures and structural components and the other half spent at Haystack Mountain Craft School assembling and using the machine. This iteration was intended to examine the validity of the flexural pantograph concept as a micron-scale reduction system, and to further test modular superelastic flexures as a rapid machine prototyping platform.

### 7.3.1 Description

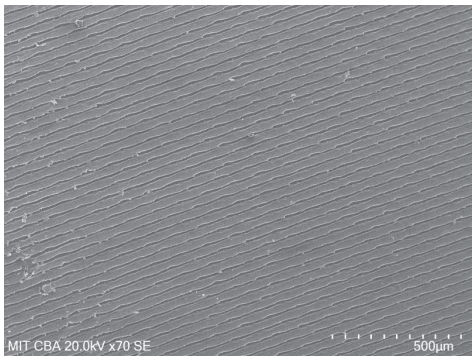
MicroPanto uses the same 12.7 mm aluminum structural stock as the previous two machines; however, to save weight, the pantograph arms are 8 mm pulltruded carbon fiber reinforced polymer (CFRP) tubes, which are epoxied into holes drilled in the aluminum parts. To


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

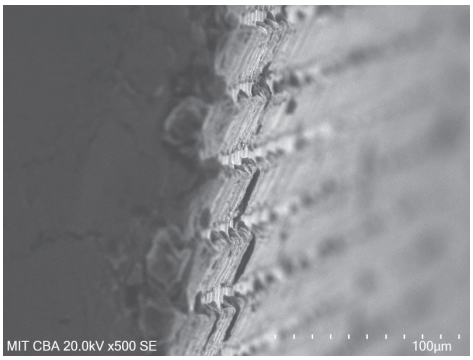
(a) Ruled grating splitting sunlight into a rainbow, which is projected onto a cup.



(b) Image of grating illuminated by a distant halogen point source.




(c) SEM micrograph of grating as viewed head-on, showing wavy distortion at the resolution limit of the machine.




(d) SEM micrograph of side of grating, showing consistent blaze angle and poor surface finish.

Figure 7-20: Ruled diffraction grating images and micrographs. [31].

achieve a 50:1 reduction ratio, the long arms of the pantograph measure 1000 mm from flexural pivot to flexural pivot, while the short arms measure 20 mm. On the drive ("large") side of the mechanism, another aluminum piece is fitted with a 6.3 mm ground shaft that can be mounted in a rotary tool equipped with a suitable collet. For the Haystack installation, the driving tool was a Handibot portable CNC router [9]. A picture of the overall setup is shown in Figure 7-21.

The "small" side of the mechanism is built on a 6.3 mm aluminum plate, which supports the pivot flexure through a stack of adjustable riser blocks. The plate also supports a miniature vise which is used to hold the workpiece. The working tool is a laser-turned CVD diamond mounted to a stainless steel flexure, which is actuated using a servo motor controlled by a simple ATtiny412-based circuit. The servo is activated by a microswitch temporarily mounted with magnets to the Handibot's Z-axis, allowing conventional tool-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

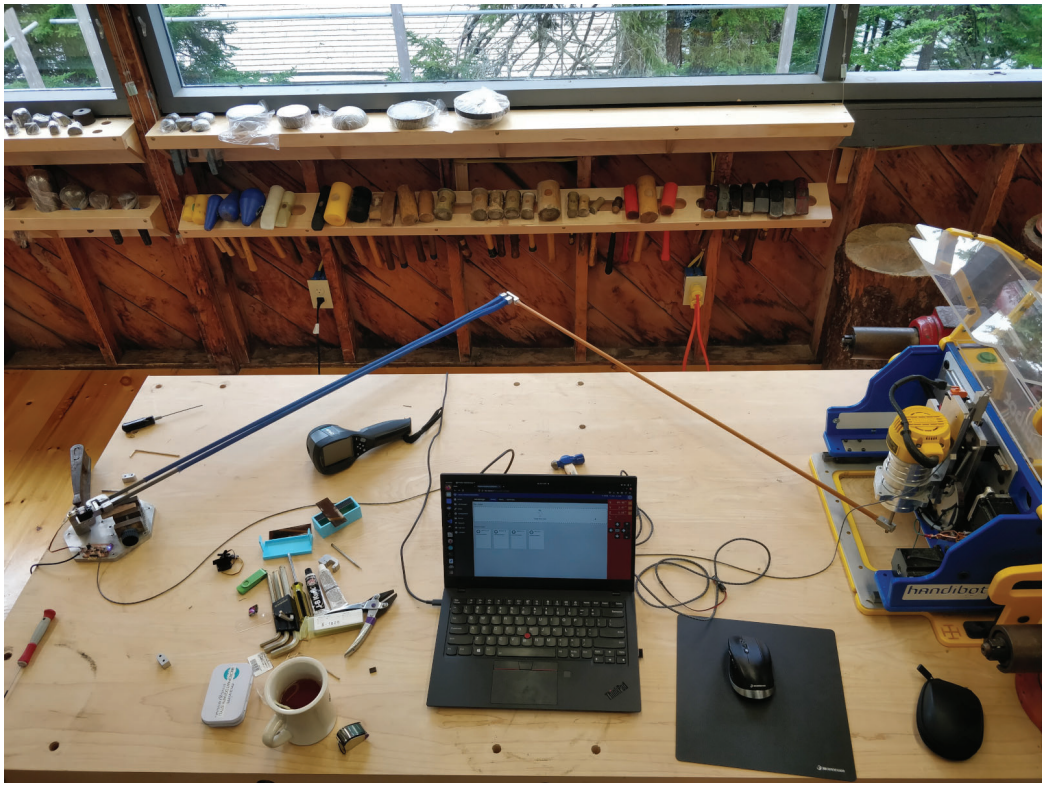


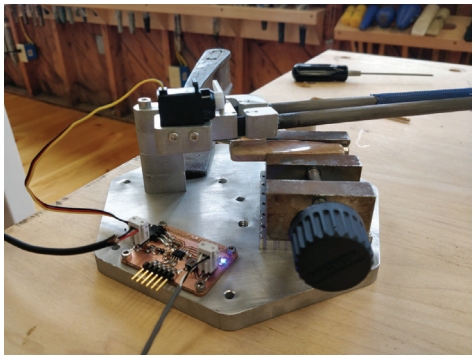

Figure 7-21: MicroPanto overview, showing Handibot CNC router on the right, control laptop at center, and engraving mechanism at left. Blue and orange rods are pulltruded CFRP tubes covered in cable loom to avoid splinters.

pathing techniques to be used without modification. A picture of the engraving head and an SEM micrograph of the diamond stylus are shown in Figure 7-22.

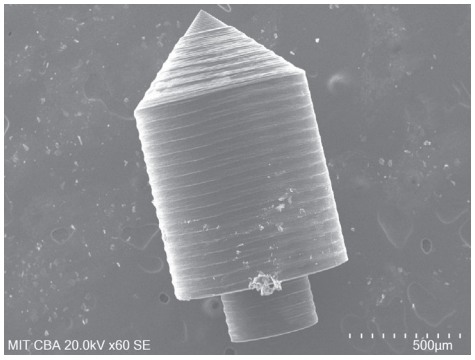
### 7.3.2 Micro-Engraving

MicroPanto was not rigorously characterized because that wasn't the point of the project. Instead, the machine was used to engrave various designs on a variety of brought and found substrates in collaboration with several of the artists that joined on the trip to Haystack. An example micro-engraving on black oxide coated steel is shown in Figure 7-23, and several engraved and inked pieces of dry rice are shown in Figure 7-24. Generally, the machine performed quite well when properly adjusted, but would benefit from a gentler stylus flexure and more precise height control.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

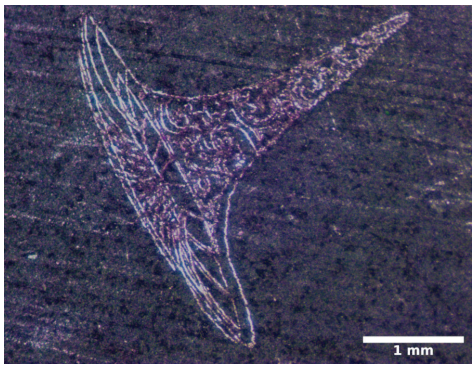


(a) Engraving head detail, showing base plate, control circuit, stylus servo, and workpiece.

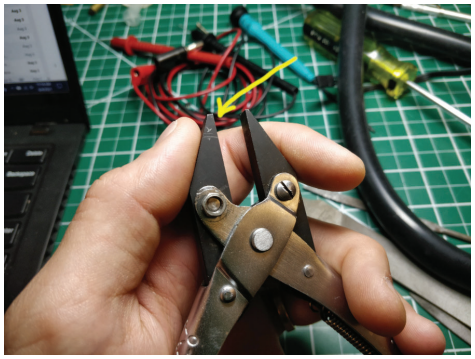



(b) SEM micrograph of laser-turned diamond stylus. Note finishing pass on tip, providing 2 um radius.

Figure 7-22: MicroPanto construction details [32].




(a) Close-up image of engraved black hole, showing good detail but some evidence of work shifting during the operation.



(b) Zoomed out view of the same engraving, demonstrating scale. Workpiece is a pair of ever-useful parallel jaw pliers.

Figure 7-23: MicroPanto example engraving, using a design by Lauren Fensterstock [5] [32].

### 7.3.3 Implications

While MicroPanto is a fairly limited tool (it has a binary Z-axis, for example), it does extend the mechanically addressable resolution available to FabLabs. Other than document scanners and paper printers, FabLab machines are generally limited to tens- or hundreds of microns in minimal feature size and addressable resolution. With more sophisticated tooling and perhaps a larger reduction ratio, the MicroPanto concept is a clear path toward bringing FabLab capabilities firmly into the single-micron realm.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Figure 7-24: Three pieces of dry rice micro-engraved and inked to reveal detail. Two show Andrea Dezsö's [27] Forest Beings, while the third shows a hand-drawn "HAYSTACK" sign.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 8

# Future Work

## 8.1 Glitter Fabrication

A clear next step is to commence fabrication and evaluation of the electronic glitter parts discussed in Chapter 5. As mentioned there, the covid-19 pandemic delayed delivery of a unique micro wire-EDM capable of cutting parts with 10 um tungsten wire. On paper, this plan seems sound; it's a straightforward scaling-down of previous electronic digital materials work [49]. On the other hand, scaling down is never that simple; the proposed feature sizes, for example, will be much closer to the grain size of the feedstock, which may lead to mechanical failures or unreliable results. Another concern is the current handling of the wire itself, which drops as the square of the diameter; this will reduce the cutting rate through the material and may require a tricky part transfer between the larger wire-EDM to minimize time spent cutting with tiny wire.

The proposed method for fabricating insulating parts is to micro-stamp them out of mica or a polymeric sheet stock. While early tests suggested that constraining tooling using flexural mechanisms may be feasible, the alignment requirements for parts at this scale are daunting and may require extensive mechanical development. Stamped parts also don't leave the machine nicely collated as with bulk wire-EDMed parts, so handling will need to be addressed. Again, the smaller scale of the parts means this will be complicated, since surface forces start to play a larger role in physical behavior.

Once a reasonable set of parts are produced, it will be exciting to explore automated assembly methods using the dispenser and assembly platform design outlined earlier. Here, perhaps the greater role of surface effects may be beneficial, if an interstitial material such


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

as solder can be used to thermally re-align the parts after assembly through capillary forces.

## 8.2 Micro-DICE

As a reliable workflow for mass-producing and assembling electronic glitter lattices comes online, it will make sense to look forward to the next DICE iteration. Given sufficient resources, taping out a DEM ASIC as discussed in Chapter 2 would allow DICE to move beyond a demonstration and workflow development platform and into the realm of real HPC work. For a first order approximation, one can scale the NVIDIA V100's computational performance (8 TFlops), power consumption (300 W), die size (800 $mm^2$), and transistor count (21 billion) to a more modest technology node, such as Intel's 22 nm node from 2011 [63]. Assuming a linear reduction and up-sizing from the V100's 12 nm FinFET node, a 10 GFlops DEM ASIC would include 30 million transistors, require a 2 $mm^2$ die, and consume 700 mW. The DICE carrier would use roughly $10^3$ conductive and insulating parts, or 250 $mm^2$ each of 25 $\mu$m-thick conductive and insulating material, worth $5 if purchased from research-grade sources [2] and taking 20 minutes to assemble at 1 Hz. A render of a proposed Micro-DICE node is shown in Figure 8-1.

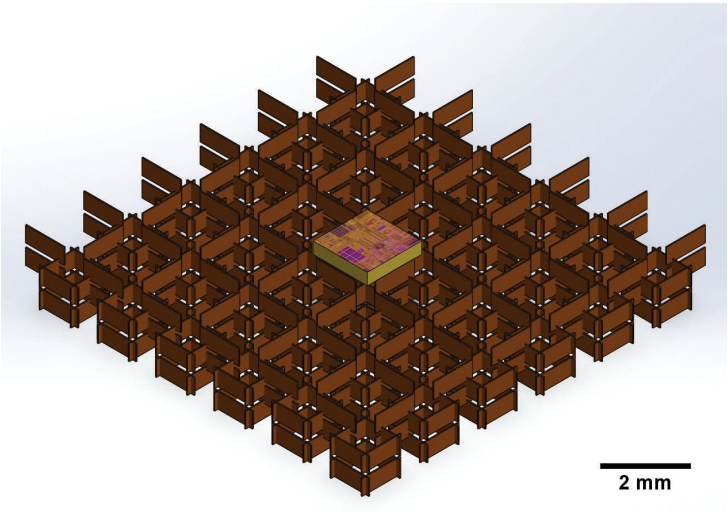



Figure 8-1: Render of a single Micro-DICE node, including a custom 10 GFlop DEM ASIC and electronic glitter lattice substrate/interconnect system.

It would take 800 Micro-DICE nodes to then equal the computational power of a single V100. Such a structure would measure 65 mm square and 36 mm tall, or roughly 13% of the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

volume of a dual-width PCIe V100 (which measures 40 mm x 111 mm x 267 mm [11]). While some additional space would be needed for a support plate and power input connection, it is notable that the V100's case does not include active cooling systems, so the comparison is surprisingly fair. Since the electronic glitter lattice has minimal cross sectional area in the vertical direction, forced liquid cooling through the lattice would be efficient and feasible. With the same economics as outlined above, the 800 Micro-DICE nodes would require 11 machine-days to assemble at a raw material cost of $4000. But notably, the scrap rate of the substrate material is many orders of magnitude cheaper than the research grade materials discussed here, and as discussed in Chapter 5, assembly machine recursion represents a clear potential path to an eventual reduction in module assembly time. A render of a V100-scale lattice of Micro-DICE nodes is shown in Figure 8-2.

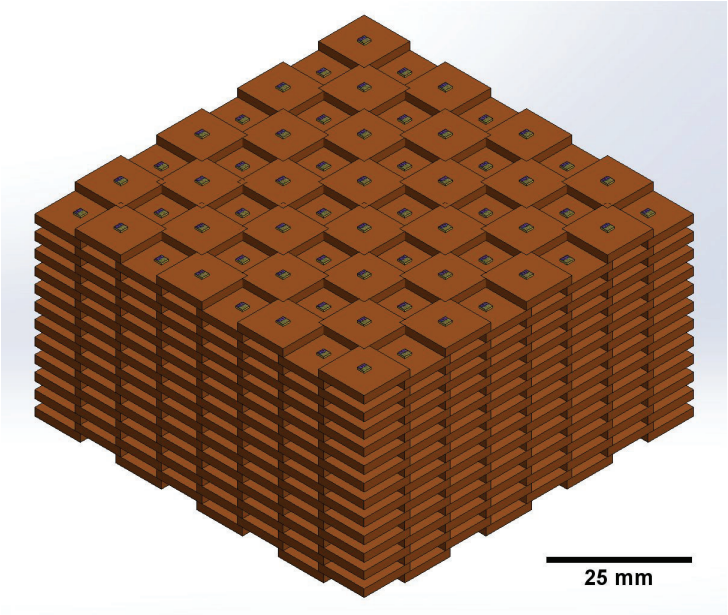



Figure 8-2: Render of 768 Micro-DICE nodes in a corner-connected cubic lattice, roughly equivalent in overall computational power to a single V100.

A rigorous examination of the economics of Micro-DICE, particularly as a V100 replacement, would consider many factors beyond the scope of this chapter. In particular, currently available process node capacity and design capabilities are clearly paramount; for example, given fab availability it may make sense to tape out a 45 nm chip. Furthermore, the chip designer's ability to integrate sufficient memory for DEM simulation may push the nodes to


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

be larger or smaller in terms of transistor count and resulting GFlops. Of course, increasing the process node size affects current consumption and efficiency. But notably, all of these trade-offs must be considered in light of Micro-DICE's stark advantages: the decreased development cost associated with a 30 million versus a 21 billion transistor ASIC, and the increased yield from a 2 $mm^2$ versus an 800 $mm^2$ die.

## 8.3 Super-DICE

Perhaps the most exciting result of the DICE project is the emergence of Super-DICE, which my colleague Camron Blackburn covers extensively in her concurrent Master's thesis. Briefly, superconducting electronics can be up to $10^5$ times more power efficient than CMOS devices when performing computational tasks, even including the roughly $10^3$ watt/watt cooling overhead needed to keep cryostats below 4 K. However, the latest superconducting fabrication technology has only recently resulted in putting a million junctions (analogous to CMOS transistors) on a single chip [85], meaning their fabrication technology is more than two decades behind the current state-of-the-art on a transistor-count basis [72].

Merging superconducting technology with DICE provides a clear path towards real relevance for superconducting electronics, since the concept would allow massive numbers of discrete asynchronous superconducting chips to work together on computational tasks. Even better, the inherently 3D nature of DICE is physically beneficial, since cryostats are three-dimensional spaces and heat transfer occurs as a function of surface area. We plan to continue our collaboration with MIT Lincoln Labs to further explore this concept.

## 8.4 Flexural Mechanisms

The frustrating fatigue problems with modular superelastic flexures will be solved, given enough time and (likely) electrochemistry. As this proceeds, a reasonable parallel effort will be to continue scaling down the working range of the machines, perhaps adding a stage to the MicroPanto to address nanometer-scale parts. Rather than attempting to fabricate devices at this scale, a good early step will be to build an atomic force microscope (AFM), likely using commercially available cantilevers and a fabricated feedback circuit.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Bibliography


[1] ATSAMD51J20A | Microchip Technology. https://www.microchip.com/en-us/product/ATSAMD51J20A.

[2] Goodfellow USA. http://www.goodfellowusa.com/. Accessed 8/18/2021.

[3] IBM Power System AC922: Technical Overview and Introduction. https://www.ibm.com/products/power-systems-ac922.

[4] Kozak Micro M3-0.25 Microadjuster Sets. https://kozakmicro.com/collections/m3-0-25-sets. Accessed 6/2/2021.

[5] LAUREN FENSTERSTOCK. https://laurenfensterstock.com/home.html. Accessed 8/10/2021.

[6] The Message Passing Interface (MPI) standard. https://www.mcs.anl.gov/research/projects/mpi/. Accessed 8/4/2021.

[7] Molex SlimStack Board-to-Board/Board-to-FPC Connectors. https://www.molex.com/molex/products/family/slimstack_fine_pitch_smt_board_to_board_connectors. Accessed 8/19/2021.

[8] OpenCV. https://opencv.org/. Accessed 8/10/2021.

[9] ShopBot HandiBot. https://www.shopbottools.com/products/handibot. Accessed 8/10/2021.

[10] Summit User Guide - OLCF User Documentation. https://docs.olcf.ornl.gov/systems/summit_user_guide.html. Accessed 8/4/2021.

[11] Tesla V100 PCIe Product Brief. https://images.nvidia.com/content/tesla/pdf/Tesla-V100-PCIe-Product-Brief.pdf. Accessed 8/18/2021.

[12] Xilinx Intellectual Property. https://www.xilinx.com/products/intellectual-property.html. Accessed 8/18/2021.

[13] CUDA Toolkit. https://developer.nvidia.com/cuda-toolkit, July 2013.

[14] List of semiconductor scale examples. https://en.wikipedia.org/w/index.php?title=List_of_semiconductor July 2021. Accessed 7/28/2021.

[15] btarunr . Samsung 3 nm GAAFET Node Delayed to 2024. https://www.techpowerup.com/283983/samsung-3-nm-gaafet-node-delayed-to-2024. Publication Title: TechPowerUp.





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[16] Andrew Adamatzky, R. Alonso-Sanz, and A. Lawniczak. *Automata-2008: Theory and Applications of Cellular Automata*. Luniver Press, 2008.

[17] Shorya Awtar. *Synthesis and Analysis of Parallel Kinematic XY Flexure Mechanisms*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2004.

[18] Shorya Awtar, Alexander H. Slocum, and Edip Sevincer. Characteristics of Beam-Based Flexure Modules. *Journal of Mechanical Design*, 129(6):625–639, May 2006.

[19] Vivek Bakshi. *EUV Sources for Lithography*. SPIE Press, 2006.

[20] Eberhard Bamberg. Viteris MW250 Micro-EDM. http://www.viteris.com/products/MW250.html. Accessed 6/2/2021.

[21] C. G. Bell, A. Kotok, T. N. Hastings, and R. Hill. The evolution of the DECsystem 10. *Communications of the ACM*, 21(1):44–63, January 1978.

[22] C.N. Berglund. A unified yield model incorporating both defect and parametric effects. *IEEE Transactions on Semiconductor Manufacturing*, 9(3):447–454, August 1996.

[23] Arani Bose, Marius Hartmann, Hans Henkes, Hon Man Liu, Michael M.H. Teng, Istvan Szikora, Ansgar Berlis, Jurgen Reul, Simon C.H. Yu, Michael Forsting, Matt Lui, Winston Lim, and Siu Po Sit. A Novel, Self-Expanding, Nitinol Stent in Medically Refractory Intracranial Atherosclerotic Stenoses. *Stroke*, 2007(38):1531–1537, 2007.

[24] Arthur Chuang. Water Management Innovation is Central to the Future of ICs, July 2021.

[25] William B Cross, Anthony H Kariotis, and Frederick J Stimler. *Nitinol characterization study*. NASA, Langley Research Center, 1969.

[26] Ian Cutress. Early TSMC 5nm Test Chip Yields 80%, HVM Coming in H1 2020. https://www.anandtech.com/show/15219/early-tsmc-5nm-test-chip-yields-80-hvm-coming-in-h1-2020.

[27] Andrea Dezsö. The home site of visual artist Andrea Dezsö. http://andreadezso.com. Accessed 8/11/2021.

[28] Samson Ellis, Yuan Gao, and Cindy Wang. TSMC Ready to Spend $20 Billion on its Most Advanced Chip Plant - Bloomberg. *Bloomberg*, October 2017.

[29] Sylvain Ferrero, Agnès Piednoir, and Claude R. Henry. Atomic Scale Imaging by UHV-AFM of Nanosized Gold Particles on Mica. *Nano Letters*, 1(5):227–230, May 2001. Publisher: American Chemical Society.

[30] The Raspberry Pi Foundation. Buy a Raspberry Pi High Quality Camera. https://www.raspberrypi.org/products/raspberry-pi-high-quality-camera/.

[31] Zach Fredin. Building machines with Zach. https://fab.cba.mit.edu/classes/865.21/people/zach/index.htm Accessed 6/10/2021.

[32] Zach Fredin. MicroPanto: a 50:1 flexural pantograph for micron-futzing with a ShopBot. https://gitlab.cba.mit.edu/zfredin/micropanto. Accessed 8/10/2021.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[33] Zach Fredin, Jiri Zemanek, Camron Blackburn, Erik Strand, Amira Abdel-Rahman, Premila Rowles, and Neil Gershenfeld. Discrete Integrated Circuit Electronics (DICE). In *2020 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–8, Waltham, MA, USA, September 2020. IEEE.

[34] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280–2292, June 2014.

[35] Neil Gershenfeld. CBA pi benchmark database. https://gitlab.cba.mit.edu/pub/pi. Accessed 8/20/2021.

[36] Neil Gershenfeld, David Dalrymple, Kailiang Chen, Ara Knaian, Forrest Green, Erik D. Demaine, Scott Greenwald, and Peter Schmidt-Nielsen. Reconfigurable asynchronous logic automata: (RALA). *ACM SIGPLAN Notices*, 45(1):1–6, January 2010.

[37] Neil A. Gershenfeld and Neil Gershenfeld. *The Nature of Mathematical Modeling*. Cambridge University Press, 1999. Google-Books-ID: zYAcGbp17nYC.

[38] Amanda Paige Ghassaei. Rapid Design and Simulation of Functional Digital Materials. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2016.

[39] Martin Goosey. Water use in the printed circuit board manufacturing process and approaches for reducing consumption. *Circuit World*, 31(2):22–25, January 2005. Publisher: Emerald Group Publishing Limited.

[40] C. Gosselin and J. Angeles. The Optimum Kinematic Design of a Planar Three-Degree-of-Freedom Parallel Manipulator. *Journal of Mechanisms, Transmissions, and Automation in Design*, 110(1):35–41, March 1988.

[41] Gaile Guevara. Traditional Japanese joinery techniques. https://www.flickr.com/photos/gaileguevara/17173017438. Accessed 8/1/2021.

[42] Jonathan D. Hiller, Joseph Miller, and Hod Lipson. Microbricks for Three-Dimensional Reconfigurable Modular Microsystems. *Journal of Microelectromechanical Systems*, 20(5):1089–1097, October 2011. Conference Name: Journal of Microelectromechanical Systems.

[43] Intel. Intel Reports Second-Quarter 2021 Financial Results. Technical report.

[44] Michael James, Marvin Tom, Patrick Groeneveld, and Vladimir Kibardin. ISPD 2020 Physical Mapping of Neural Networks on a Wafer-Scale Deep Learning Accelerator. In *Proceedings of the 2020 International Symposium on Physical Design*, ISPD '20, pages 145–149, New York, NY, USA, March 2020. Association for Computing Machinery.

[45] W. James. Characteristics of Modular Electronics Components. *IRE Transactions on Component Parts*, 3(2):69–72, September 1956. Conference Name: IRE Transactions on Component Parts.

[46] Ronald K. Jurgen. Whatever happened to project tinkertoy? *IEEE Spectrum*, 24(5):20–21, May 1987. Conference Name: IEEE Spectrum.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[47] Gordon Keeler. Common Heterogeneous Integration and IP Reuse Strategies. https://www.darpa.mil/program/common-heterogeneous-integration-and-ip-reuse-strategies.

[48] V Lalitha and S Kathiravan. A Review of Manchester, Miller, and FM0 Encoding Techniques. *The Smart Computing Review*, 4(6), December 2014.

[49] William Kai Langford. Electronic Digital Materials. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2014.

[50] William Kai Langford. *Discrete Robotic Construction.* PhD Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2019.

[51] Nicolae Lobontiu, Jeffrey S. N Paine, Edward O'Malley, and Marc Samuelson. Parabolic and hyperbolic flexure hinges: flexibility, motion precision and stress characterization based on compliance closed-form equations. *Precision Engineering*, 26(2):183–192, April 2002.

[52] Hamilton H. Mabie and Charles F. Reinholtz. *Mechanisms and Dynamics of Machinery.* John Wiley & Sons, January 1987.

[53] Robert MacCurdy, Anthony McNicoll, and Hod Lipson. Bitblox: Printable digital materials for electromechanical machines. *The International Journal of Robotics Research*, 33(10), 2014.

[54] Chris A. Mack. Fifty Years of Moore's Law. *IEEE Transactions on Semiconductor Manufacturing*, 24(2):202–207, May 2011.

[55] P. Y. Martinez, Y. Beilliard, M. Godard, D. Danovitch, D. Drouin, J. Charbonnier, P. Coudrain, A. Garnier, D. Lattard, P. Vivet, S. Cheramy, E. Guthmuller, C. Fuguet Tortolero, V. Mengue, J. Durupt, A. Philippe, and D. Dutoit. ExaNoDe: Combined Integration of Chiplets on Active Interposer with Bare Dice in a Multi-Chip-Module for Heterogeneous and Scalable High Performance Compute Nodes. In *2020 IEEE Symposium on VLSI Technology*, pages 1–2, Honolulu, HI, 2020. IEEE. ISSN: 2158-9682.

[56] Morgan McCorkle. ORNL Launches Summit Supercomputer | ORNL. https://www.ornl.gov/news/ornl-launches-summit-supercomputer. Accessed 8/18/2021.

[57] Wei-dong Miao, Xu-jun Mi, Xin-lu Wang, and Hua-chu Li. Electropolishing parameters of NiTi alloy. *Transactions of Nonferrous Metals Society of China*, 16:s130–s132, June 2006.

[58] Gordon E Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8), April 1965.

[59] Seth Morabito. Messy. https://www.flickr.com/photos/twylo/2556770440/, June 2008. Accessed 8/18/2021.

[60] B.T. Murphy. Cost-size optima of monolithic integrated circuits. *Proceedings of the IEEE*, 52(12):1537–1545, December 1964.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[61] D. Nicoară and I. Nicoară. An improved Bridgman-Stockbarger crystal-growth system. *Materials Science and Engineering: A*, 102(2):L1–L4, July 1988.

[62] Nicoguaro. Stress strain ductile curve. https://commons.wikimedia.org/w/index.php?curid=89891144, May 2020. Accessed 8/20/2021.

[63] Nilay Patel. Intel announces 22nm chips for 2011. https://www.engadget.com/2009-09-22-intel-announces-22nm-chips-for-2011.html, September 2009.

[64] Prashant Patil. *Laser Direct-Write Fabrication of MEMS*. PhD Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2019.

[65] Andrew Petto, Zachary Fredin, and Joseph Burdo. The Use of Modular, Electronic Neuron Simulators for Neural Circuit Construction Produces Learning Gains in an Undergraduate Anatomy and Physiology Course. *Journal of Undergraduate Neuroscience Education*, 15(2):A151–A156, June 2017.

[66] Polyparadig. Empty Tic Tacs box. https://commons.wikimedia.org/wiki/File:Mint_box_polypropylene_l Accessed 5/2/2021.

[67] George A Popescu. Digital Materials for Digital Fabrication. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2007.

[68] Jon Porter. Apple says new Arm-based M1 chip offers the 'longest battery life ever in a Mac'. https://www.theverge.com/2020/11/10/21558095/apple-silicon-m1-chip-arm-macs-soc-charge-power-efficiency-mobile-processor, November 2020.

[69] Hui Qian, Hongnan Li, Gangbing Song, and Wei Guo. Recentering Shape Memory Alloy Passive Damper for Structural Vibration Control. *Mathematical Problems in Engineering*, 2013:e963530, November 2013. Publisher: Hindawi.

[70] Jake Read. clank-lz. https://gitlab.cba.mit.edu/jakeread/clank-lz. Accessed 5/2/2021.

[71] Jake Robert Read. Distributed Dataflow Machine Controllers. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2020.

[72] Max Ritchie and Hannah Roser. A logarithmic graph showing the timeline of how transistor counts in mircochips are almost doubling every two years from 1970 to 2020; Moore's Law. https://commons.wikimedia.org/wiki/File:Moore%27s_Law_Transistor_Count_1970-2020.png, November 2020. Accessed 7/28/2021.

[73] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[74] Otto Schmitt. *An electrical theory of nerve impulse propagation*. PhD Thesis, Washington University, 1937.

[75] E. Schöller, L. Krone, M. Bram, Hans Buchkremer, and D. Ståaver. Metal injection molding of shape memory alloys using prealloyed NiTi powders. *Journal of Materials Science*, 40:4231–4238, August 2005.

[76] Anand Lal Shimpi. AMD's Fab 36 Grand Opening - 90nm and 300mm in Germany. https://www.anandtech.com/show/1821. Accessed 7/28/2021.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[77] Paul Stoffregen. Teensy® 4.0 Technical Information. https://www.pjrc.com/store/teensy40.html#tech. Accessed 7/31/2021.

[78] Erik Steven Strand. Inverse Methods for Design and Simulation with Particle Systems. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2020.

[79] Masahiro Sunohara, Takayuki Tokunaga, Takashi Kurihara, and Mitsutoshi Higashi. Silicon interposer with TSVs (Through Silicon Vias) and fine multilayer wiring. In *2008 58th Electronic Components and Technology Conference*, pages 847–852, May 2008. ISSN: 2377-5726.

[80] Larry W. Swanson, Eric Newman, Alfonso Araque, and Janet M. Dubinsky. *The Beautiful Brain: The Drawings of Santiago Ramón y Cajal*. Abrams, 2017.

[81] Alex Taradov. edbg: Simple utility for programming Atmel MCUs though CMSIS-DAP protocol. https://github.com/ataradov/edbg. Accessed 5/2/2021.

[82] Alex Taradov. Free-DAP. https://github.com/ataradov/free-dap, August 2021. Accessed 8/19/2021.

[83] Denis Terwagne. Urumbu flexible X/Y stage. https://gitlab.fabcloud.org/academany/fabacademy/2020/bootcamp/spicy/-/tree/master/Urumbu/flexible-XYstage. Accessed 7/1/2021.

[84] Thomas Thwaites. *The Toaster Project: Or a Heroic Attempt to Build a Simple Electric Appliance from Scratch*. Chronicle Books, September 2011.

[85] Sergey K. Tolpygo, Vladimir Bolkhovsky, Daniel E. Oates, Ravi Rastogi, Scott Zarr, Alexandra L. Day, Tarence J. Weir, Alex Wynn, and Leonard M. Johnson. Superconductor Electronics Fabrication Process with MoNx Kinetic Inductors and Self-Shunted Josephson Junctions. *IEEE Transactions on Applied Superconductivity*, 28(4):1–12, June 2018. Conference Name: IEEE Transactions on Applied Superconductivity.

[86] Reinhard Uecker. The historical development of the Czochralski method | Elsevier Enhanced Reader. *Journal of Crystal Growth*, 401(1):7–24, September 2014.

[87] Pascal Vivet, Eric Guthmuller, Yvain Thonnart, Gael Pillonnet, Guillaume Moritz, Ivan Miro-Panadès, Cesar Fuguet, Jean Durupt, Christian Bernard, Didier Varreau, Julian Pontes, Sebastien Thuries, David Coriat, Michel Harrand, Denis Dutoit, Didier Lattard, Lucile Arnaud, Jean Charbonnier, Perceval Coudrain, Arnaud Garnier, Frederic Berger, Alain Gueugnot, Alain Greiner, Quentin Meunier, Alexis Farcy, Alexandre Arriordaz, Severine Cheramy, and Fabien Clermidy. A 220GOPS 96-Core Processor with 6 Chiplets 3D-Stacked on an Active Interposer Offering 0.6ns/mm Latency, 3Tb/s/mm2 Inter-Chiplet Interconnects and 156mW/mm2@ 82%-Peak-Efficiency DC-DC Converters. In *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, pages 46–48, San Francisco, CA, 2020. IEEE. ISSN: 2376-8606.

[88] Jonathan Ward. Additive Assembly of Digital Materials. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2010.

[89] John F. Wendt, John D. Anderson, and Von Karman Institute for Fluid Dynamics, editors. *Computational fluid dynamics: an introduction*. Springer, Berlin ; [London], 3rd ed edition, 2008. OCLC: ocn288984495.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[90] Robert L. Williams and Brett H. Shelley. Inverse Kinematics for Planar Parallel Manipulators. American Society of Mechanical Engineers Digital Collection, February 2021.

[91] Claire Wolf and Mathias Lasser. Project IceStorm. http://bygone.clairexen.net/icestorm/. Accessed 8/19/2021.

[92] Miao Yang, Chi Zhang, Guilin Yang, and Wei Dong. Optimal Design and Tracking Control of a Superelastic Flexure Hinge Based 3-PRR Compliant Parallel Manipulator. *IEEE Access*, 7:174236–174247, 2019. Conference Name: IEEE Access.

[93] Geoffrey Yeap, S. S. Lin, Y. M. Chen, H. L. Shang, P. W. Wang, H. C. Lin, Y. C. Peng, J. Y. Sheu, M. Wang, X. Chen, B. R. Yang, C. P. Lin, F. C. Yang, Y. K. Leung, D. W. Lin, C. P. Chen, K. F. Yu, D. H. Chen, C. Y. Chang, H. K. Chen, P. Hung, C. S. Hou, Y. K. Cheng, J. Chang, L. Yuan, C. K. Lin, C. C. Chen, Y. C. Yeo, M. H. Tsai, H. T. Lin, C. O. Chui, K. B. Huang, W. Chang, H. J. Lin, K. W. Chen, R. Chen, S. H. Sun, Q. Fu, H. T. Yang, H. T. Chiang, C. C. Yeh, T. L. Lee, C. H. Wang, S. L. Shue, C. W. Wu, R. Lu, W. R. Lin, J. Wu, F. Lai, Y. H. Wu, B. Z. Tien, Y. C. Huang, L. C. Lu, Jun He, Y. Ku, J. Lin, M. Cao, T. S. Chang, and S. M. Jang. 5nm CMOS Production Technology Platform featuring full-fledged EUV, and High Mobility Channel FinFETs with densest 0.021µm2 SRAM cells for Mobile SoC and High Performance Computing Applications. In *2019 IEEE International Electron Devices Meeting (IEDM)*, pages 36.7.1–36.7.4, December 2019.

[94] Hao Zhang and Haihang You. Comprehensive Workload Analysis and Modeling of a Petascale Supercomputer. In *Workshop on Job Scheduling Strategies for Parallel Processing*, pages 253–271, Berlin, May 2012. Springer.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Superconducting Asynchronous Logic for Ultra-low Power High Performance Computing

by

L. Camron Blackburn

B.A. Physics, New York University (2017)

Submitted to the Program in Media Arts and Sciences
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

© Massachusetts Institute of Technology 2021. All rights reserved.

Author .....
Program in Media Arts and Sciences
August 20th, 2021

Certified by .....
Professor Neil Gershenfeld
Director, MIT Center for Bits and Atoms
Thesis Supervisor

Accepted by .....
Professor Tod Machover
Academic Head, Program in Media Arts and Sciences



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Superconducting Asynchronous Logic for Ultra-low Power High Performance Computing

by

L. Camron Blackburn

Submitted to the Program in Media Arts and Sciences
on August 20th, 2021, in partial fulfillment of the
requirements for the degree of
Master of Science

## Abstract

High performance computing is bottlenecked by increasing power demands and memory bandwidth, while superconducting electronics are bounded in circuit complexity due to a limit on the number of switching devices on a single chip. This thesis proposes a modular, asynchronous superconducting computing framework which aims to solve both of these problems. A discrete set of logic gates are proposed and implemented using Adiabatic Quantum Flux Parametron (AQFP) logic. AQFP logic devices can achieve picosecond gate delays with zeptojoule ($10^{-21}$ J) switching energy, just bordering the theoretical Landauer limit for computing energy demands, by adiabatically switching the location of a single flux quanta in a double-well potential. The heart of the project lies in the modular architecture design that realigns hardware layout with software dataflow to allow for scalable, distributed computing systems from basic circuit building blocks. Projecting the simple circuit design performance to large-scale high performance computing systems, Super-DICE aims to achieve a $10^3$ order of magnitude improvement in power consumption, while still accounting for the cryogenic cooling overhead of the superconducting electronics. Beyond the dramatic power performance improvement with this logic technology and architecture, it also allows for designers to rapidly prototype hardware computing optimizations without needing to go through the expensive and time consuming process of fully custom ASIC design.

In this thesis, I review the device physics of the Quantum Flux Parametron and present a set of basic AQFP combinatorial logic gates. I then propose a circuit design for asynchronous token buffering between these modular gates and describe how they can be assembled as digital materials to create scalable, complex 3D computing structures. I simulate the proposed circuit designs in SPICE and project performance of a potential superconducting supercomputer using this framework. Motivated by the energy efficiency of superconducting electronics, the heart of this thesis radically proposes to redefine traditional processor architecture by discretizing large-scale system integration into a heterogeneous set of building blocks which blur the line between hardware and software with a reconfigurable, asynchronous spatial computing sys-



148
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

tem.

Thesis Supervisor: Professor Neil Gershenfeld
Title: Director, MIT Center for Bits and Atoms




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Superconducting Asynchronous Logic for Ultra-low Power High Performance Computing**

by

L. Camron Blackburn

This thesis has been reviewed and approved by the following committee members:

Neil Gershenfeld .....
Director, MIT Center for Bits and Atoms
Professor, Media Arts and Sciences
MIT

Alexander Wynn .....
Research Scientist
MIT Lincoln Laboratory

Deblina Sarkar .....
Assistant Professor, Media Arts and Science
MIT



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Acknowledgments

First off, I want to thank my advisor, Neil Gershenfeld, for providing the space, resources, and support to make this thesis possible. The Center for Bits and Atoms is a truly unique place, and the extent to which I am grateful for my time here continues to grow each day.

Thank you to Alex Wynn for getting me started with the basics of superconducting electronic design a little over a year ago and continuing to be a invaluable guide throughout this project.

To my thesis readers, thank you for taking the time to provide your input and guidance throughout the thesis procedures.

I gratefully acknowledge funding support from, and collaboration with, MIT Lincoln Laboratory under ACC 767, DARPA and the Air Force Research Laboratory under award number FA8650-19-2-7921, and all of the sponsors under the Center for Bits and Atoms consortium.

The last two years have been an unprecedented time, with Coivd-19 interrupting the 2021 Masters program a little less than half way through. Given this context, I'd like to make an extra large thank you to Neil, the CBA staff and admin, and the MAS department, for keeping the lab running smoothly during a tumultuous time and providing a sense of safety and security which I feel so lucky to have had.

To my fellow students and researchers at CBA, thank you for the inspiration, comedic relief, and advice. Special thanks to Zach and Alfonso for providing some sanity while we were the only three people in 023 for a few months doing covid research, and for starting our CBA journey together.

To all the inhabitants and friends of Das Haus, thanks for always being up for an adventure and for all the family dinners and shenanigans.

To Alfonso (again) thank you for being the best flatmate, labmate, and kitten co-parent anyone could ask for - I am beyond grateful to have your friendship, your company, and your cooking.

And a throwback to my undergrad physics crew - Nikitas, Chris, Nick, Iraj -





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

thanks for all the late nights we spent falling in (and occasionally out of) love with physics and for inflicting strong enough grad school FOMO to get me to where I am today.

Finally, a huge thank you to my family - Mom and Jeff, Evie, Danny, and Brian - for always cheering me on, making me laugh, and putting life in perspective. Thanks to my grandpa for all the love and support, and for not having held it over my head (too much) that I ended up at MIT instead of Stanford. Lastly, to my Dad thank you for instilling in me a curiosity and excitement for innovation from a young age - I wish more than anything he were alive to share even a fraction of the projects I've been able to learn about and work on during my time at CBA, because there's no doubt he would be in awe of the state of technology today.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Contents


**1 Introduction** **17**

1.1 Motivation . . . . . 17

1.2 Related Work . . . . . 19

1.2.1 Electronic Digital Materials . . . . . 19

1.2.2 DICE Framework . . . . . 20

1.2.3 Distributed & Asynchronous Circuit Design . . . . . 22

1.2.4 Superconducting electronics . . . . . 24

1.3 Contributions . . . . . 26

**2 Ultra-low Energy Superconducting Logic** **29**

2.1 Josephson Effect . . . . . 30

2.2 Superconducting Logic Families . . . . . 33

2.3 Junctions and Loops . . . . . 35

2.4 Quantum Flux Parametron . . . . . 39

2.5 QFP Logic Cells . . . . . 44

**3 Asynchronous Superconducting Circuits** **47**

3.1 Asynchronous Logic Automata . . . . . 47

3.1.1 ALA Cell Library . . . . . 49

3.2 Superconducting ALA . . . . . 52

3.2.1 AQFP Logic Gates . . . . . 53

3.2.2 Asynchronicity in AQFP . . . . . 54

3.3 Token Buffers . . . . . 55





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

3.3.1 C-Element Coincidence Buffer . . . . . 57
3.3.2 Precharge Full Buffer (PCFB) . . . . . 60
3.3.3 QFP Full Binary Buffer (QFBB) . . . . . 63
3.3.4 Variable Activation QFP (VAQ) . . . . . 65

**4 Logic Modules** **73**
4.1 Token Boundary . . . . . 73
4.2 Adder . . . . . 74
4.2.1 Vanilla ALA Adder . . . . . 74
4.2.2 Synchronous Adders . . . . . 76
4.3 Multiplier . . . . . 77

**5 Evaluation** **79**
5.1 SPICE Simulation . . . . . 79
5.2 Energy Dissipation . . . . . 82
5.3 Scalability . . . . . 83

**6 Conclusion** **87**
6.1 Future Plans . . . . . 87
6.2 Impact . . . . . 89


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# List of Figures


1-1 Power consumption comparison between state-of-art supercomputers and Super-DICE . . . . . 18
1-2 Functional digital material breakout for SOIC interconnect [1] . . . . 20
1-3 Langford's Electronic Digital Material parts [2] . . . . . 21
1-4 DICE framework at different node size scales and (c) shows end-to-end design tool . . . . . 23
1-5 Sycamore: 53-bit superconducting quantum processor [3] . . . . . 25

2-1 Josephson Junction diagram and schematic symbol . . . . . 31
2-2 RCSJ Equivalent Circuit Diagram . . . . . 32
2-3 Junction I-V SPICE Plot . . . . . 33
2-4 Superconducting Logic Family Tree . . . . . 34
2-5 Superconducting Loop with Josephson Junction . . . . . 37
2-6 Mechanical QFP Analogy, from [4] . . . . . 40
2-7 QFP Schematic . . . . . 41
2-8 QFP energy-phase plots . . . . . 42
2-9 $\phi$ output vs activation signal, $\alpha$, from [5] . . . . . 42
2-10 Three-phase activation signal propagating through QFP buffers . . . 43
2-11 AQFP Cell Library. Junction-level schematics and logic-level schematics on the left and right, respectively. . . . . 46

3-1 Architecture diagrams of traditional computing vs distributed computing (like ALA) . . . . . 48
3-2 ALA cell library . . . . . 50





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

3-3 LFSR ALA schematic and layout from [6] . . . . . 51
3-4 AQFP Logic Gates for ALA Library . . . . . 53
3-5 AQFP data input phase synchronizer . . . . . 55
3-6 Asynchronous Linear Pipeline . . . . . 56
3-7 C-element Schematics: NAND gate design, Majority gate design, and circuit symbol . . . . . 58
3-8 Asymmetric C-element schematics and circuit symbol . . . . . 59
3-9 QFP-level C-element schematic . . . . . 60
3-10 Precharge Full Buffer Schematic . . . . . 61
3-11 Precharge Full Buffer (PCFB) QFP-level implementation . . . . . 62
3-12 QFP Full Binary Buffer (QFBB) for single data input . . . . . 63
3-13 NOR Flip-Flop . . . . . 64
3-14 QFP Full Binary Buffer (QFBB) for two data input . . . . . 66
3-15 I/O Transformer QFP . . . . . 67
3-16 QFP with Puller for activation signal boost . . . . . 69
3-17 Junction-level schematic of Variable Activation QFP with Driving QFP 70
3-18 Dual-rail to binary token converter circuit . . . . . 71

4-1 ALA Serial Adder schematic given on the left and an example of it used in the Fibonacci sequence generator to the right. . . . . 75
4-2 ALA Carry Adder . . . . . 75
4-3 Synchronous Half Adder . . . . . 76
4-4 Synchronous Full Adder . . . . . 76
4-5 Synchronous Full Adder without XOR gates . . . . . 77
4-6 ALA Multiplier . . . . . 78

5-1 SPICE plots for AQFP logic gates . . . . . 80
5-2 Phase synchronizer. Input passed on activation phase 1, 2, and 3 and data out makes it to the buffer loop each time. . . . . 81
5-3 SPICE plots for synchronous adders . . . . . 81
5-4 QFP bit energies for 10 buffers in series [7] . . . . . 82


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

6-1 Pizza mask test chip currently in fabrication at MIT LL . . . . . 88




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# List of Tables

2.1 Bit-Energy Comparison of SCE Logic Families . . . . . 36

3.1 C-element Logic Table . . . . . 57

3.2 Asymmetric C-element Logic Table . . . . . 58

5.1 Circuit Performance Projections . . . . . 84



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 1

# Introduction

## 1.1 Motivation

High performance computing is bottle-necked by increasing power demands and memory bandwidth, while superconducting electronics are bounded in circuit complexity due to a limit on the number of switching devices on a single chip. The work in this thesis proposes a superconducting computing framework that aims to solve both of these problems. It is part of a larger project, which we call Super-DICE for superconducting Discrete Integrated Circuit Electronics, in collaboration with fellow colleagues at the Center for Bits and Atoms (CBA) and MIT Lincoln Laboratory (MIT LL).

High performance computing is crucial to a wide range of research fields, such as machine learning, multiphysics modeling, and climate science. While so many areas of expertise require more computational complexity, supercomputers are limited by their expensive power consumption. According to the Top500 benchmarking list, as of June 2021, Supercomputer Fugaku in Japan is the worlds fastest supercomputer and requires 29.9 MW of power with 15.42 GFlops / Watts [8]. The accompanying Green500 benchmarking list ranks the worlds most energy efficient supercomputers, and as of June 2021 the MN-3 supercomputer cluster in Japan is the worlds most energy efficient supercomputer with 29.7 GFlops/Watts [9]. This places state of the art supercomputers at an energy efficiency on the order of $10^{-10}$ W/Flops.



159
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

On the other hand, superconducting electronics provide a compelling ultra-low power alternative for classical computation. It has been shown that an 8-bit superconducting ALU with $10^3$ JJs operating at 5 GHz has an energy dissipation of 1.4 aJ per operation [10]. This implies 10 nW power dissipation ($1.5 \times 10^{-18}$J/op$\cdot 5 \times 10^{9}$s$^{-1}$). Assuming this power performance could be extrapolated to a chip with the maximum number of switching devices, $O(10^6)$, with $10^{10}$ op/s, then a superconducting supercomputer could have a power efficiency of $10^{-18}$ W/ops. Accounting for the $10^3$ W/W cryogenic cooling overhead cost, the net power efficiency would be on the order of $10^{-15}$ W/flops.

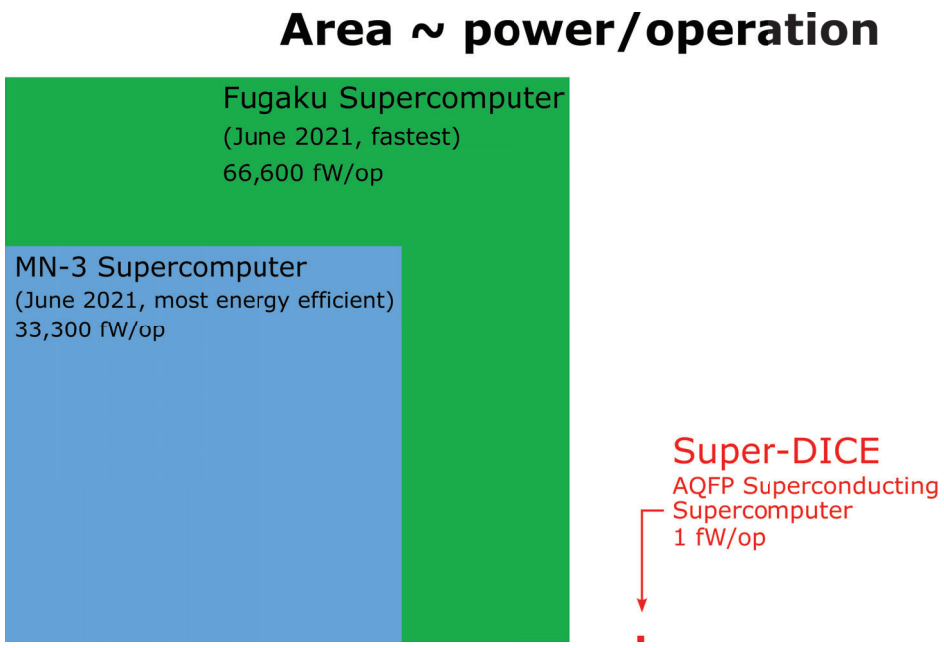



Figure 1-1: Power consumption comparison between state-of-art supercomputers and Super-DICE

This quick back of the envelope case study implies a $10^5$ order of magnitude improvement in power efficiency for superconducting supercomputing. Figure 1-1 visualizes the scale of this improvement. The difficulty in realizing this projection comes when scaling the superconducting circuit design - there are physical limitations to how small and dense the Josephson junctions (a crucial element to the superconducting switching device) can be on a single chip. Therefore, any reasonably complex


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

superconducting processor needs to extend to a multi-chip module and take on the added design complexity and communication overhead that entails. The solution to this problem lies in Super-DICE's modular, asynchronous computing architecture that realigns hardware layout with software dataflow to allow for scalable, distributed computing systems from basic circuit building blocks.

## 1.2 Related Work

Many of the ideas put forth in this thesis are an extension of previous and concurrent CBA projects, while also pulling from a large body of literature in asynchronous circuit design, spatial computing, and superconducting electronics. Some of these related projects and backgrounds are explored here.

### 1.2.1 Electronic Digital Materials

Digital materials, discussed deeply in Jonathan Ward's 2010 Master's thesis, are 3D structures made of a finite number of building blocks with discrete joints [1]. By their nature, discrete joints are self-aligning and error-correcting which allows for robust and scalable assembly. LEGO block construction can provide a descriptive example of digital materials: there is a predefined set of parts with plug and port discrete interconnects and a child with about 0.2mm resolution hand placement can assemble complex structures with about 5 $\mu$m precision [1]. Ward focused on digital materials for making reconfigurable 3D physical structures, and extended this idea to functional structures with the design of a 3D electrical interconnect for SOIC (small outline integration circuit) packages, displayed in Figure 1-2.

The work in Will Langford's masters thesis expanded this concept to the idea of Electronic Digital Materials which consist of a conductive, insulating, and resistive part-type to build passive electronic components and four semiconducting parts to build active electronic components [2]. Some of these discrete electronic components are shown in Figure 1-3.

The inspiration for Super-DICE comes two generations after electronic digital




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

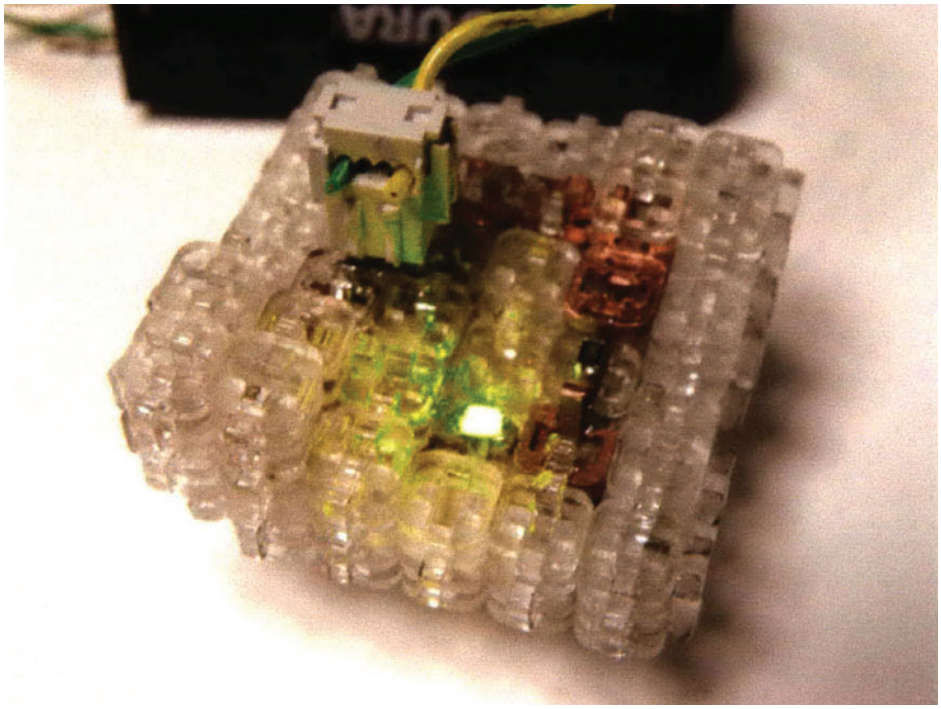


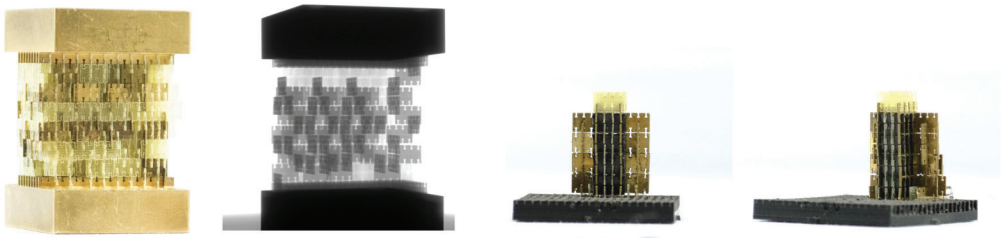
Figure 1-2: Functional digital material breakout for SOIC interconnect [1]

materials, with closer relation to the DICE project which will be described in the next section. However, the spirit of digital materials providing the building block for computational structure remains in the super-DICE project, in both the interconnect packaging and computer architecture. Although the work in this thesis focuses on intra-chip circuitry of super-DICE components, the project as a whole also focuses on interconnect and packaging design for 3D computing structures, and this work is currently being documented in Zach Fredin's Master's thesis.

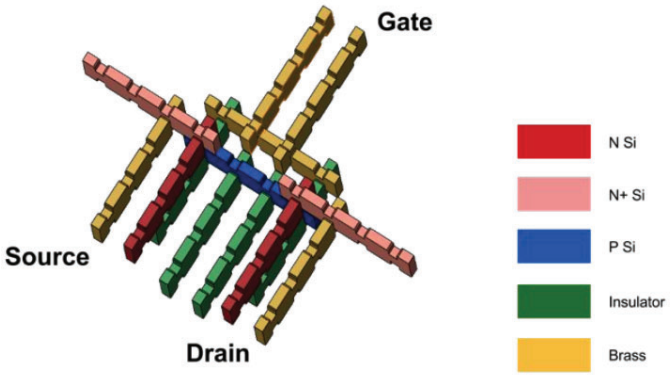
### 1.2.2 DICE Framework

Super-DICE came into being as an offshoot of the DICE project, which is simultaneously in development. Inspired by Langford's electronic digital materials, the DICE project takes a more complex starting node of commercial off-the-shelf microcontrollers to create three-dimensional LEGO-like building blocks for application specific hardware design [11]. DICE nodes scale in a reconfigurable direct-write au-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.



(a) Inductor (b) Resistor




(c) MOSFET design

Figure 1-3: Langford's Electronic Digital Material parts [2]


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

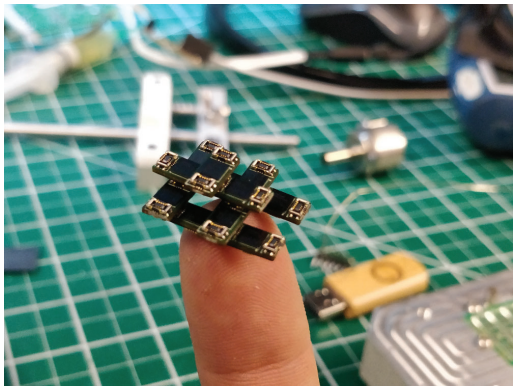
tomated assembly process and communicate through an asynchronous token passing protocol which allows hardware design to adapt and grow to optimize software tasks. DICE parts and the design workflow are shown in Figure 1-4. A driving motivation for DICE, along with Super-DICE, is the spatial alignment of software instructions with hardware layout. For example, if you're programming a deep neural network, you want the weights calculated from one layer to be stored physically close to where the next layer is computed, in order to save time and energy in retrieving values from memory. As will be explore more in Section 3.1, in traditional CPU architecture the programmer does not easily have control over where information is stored in the hardware; however, the DICE framework and design tool allows the programmer to align the physics of their software with the physics of their hardware, resulting in speed and power optimizations. Furthermore, providing the programmer with control over spatial location of memory storage also has important applications in secure and trusted systems, to allow better protection against buffer overflow vulnerabilities. The design tool also offers a creative sandbox for rapid prototyping of custom hardware design.

As will be shown throughout this thesis, Super-DICE maintains many of the benefits provided by the DICE project, with the added gain in power performance due to the superconducting technology. The DICE project is currently developing much of the end-to-end workflow which will be required for Super-DICE, such as packaging and interconnect design, custom automated assembly machines for the discrete parts, and user-friendly design tools for system development. This work is further documented in the Master's theses of Erik Strand [12], Zach Fredin, and Justin Christensen.

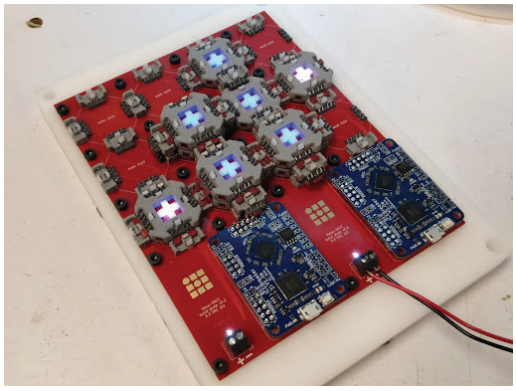
### 1.2.3 Distributed & Asynchronous Circuit Design

Both distributed and asynchronous circuit design are not new concepts. The intra-chip node design of Super-DICE parts follows from previous work on Asynchronous Logic Automata (ALA) [13, 6, 14]. ALA is a specialization of asynchronous cellular automata [15] and Petri nets [16] which consists of Boolean logic gates at each node to design a spatially coherent computing structure; this will be explored in much greater

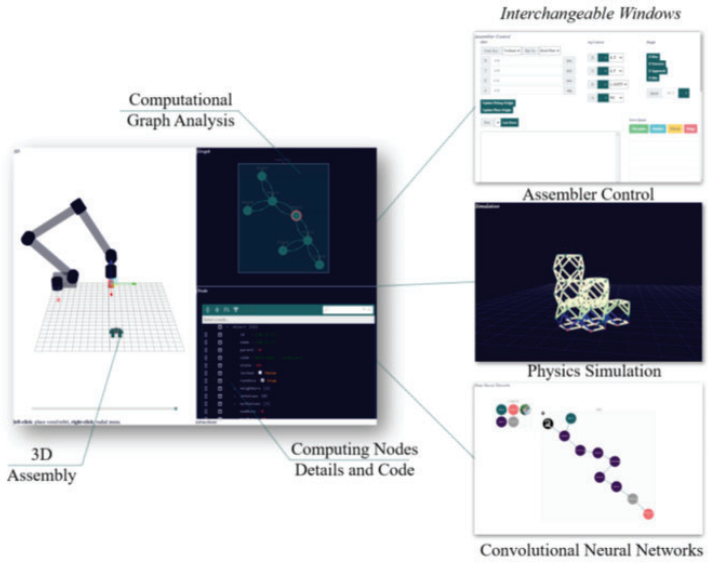

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.



(a) Tiny DICE



(b) Meso DICE




(c) DICE design tool

Figure 1-4: DICE framework at different node size scales and (c) shows end-to-end design tool


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

detail in Section 3.1.

Previous attempts at distributed, cellular automata computing systems for better spatial dataflow alignment were popularly explored in the 1980s and 90s, such as Thinking Machines' Connection Machine [17] and the CAM8 architecture [18] from MIT CSAIL. Similar to ALA, both of these architectures aimed to embed realistic spatial constraints into computer architecture; however, neither were asynchronous. Furthermore, their erasure in the space of parallel computing was arguably not due to technical feasibility, and instead poor business decisions [19].

Super-DICE is also similar to systolic arrays, which is a parallel computing architecture of homogeneous data processing units popular for matrix multiplication and LU-decomposition [20]. However, Super-DICE building blocks pull from a heterogeneous set of nodes and it's asynchronous nature allows information to flow through the system with more freedom than systolic array architectures.

### 1.2.4 Superconducting electronics

The idea of using superconducting materials to build computing devices has been around since the 1950s when Dudley Buck's cryotron device from MIT received a lot of attention and federal funding as the potential future of computing [21]. However, the cryotron ended up being more difficult to fabricate en masse than expected, while semiconductor transistors were just taking off, and the rest is history – the cryotron was forgotten [22]. Since then, there have been waves of excitement around superconducting devices, which will be explored deeper in Section 2.2; however, most of these projects were not developed beyond basic physical characterization and proof-of-concept circuit design, for reasons that will also be explored in greater depth in Chapter 2.

Today, most of the excitement around superconducting electronics is for application in quantum computing systems [23, 24]. Google's experimental realization of quantum supremacy, i.e. a quantum computer exceeding the capabilities of a classical supercomputer [25], was achieved with 53 programmable superconducting qubits. Google's Sycamore processor is shown in Figure 1-5 [3]. The active device proposed in


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

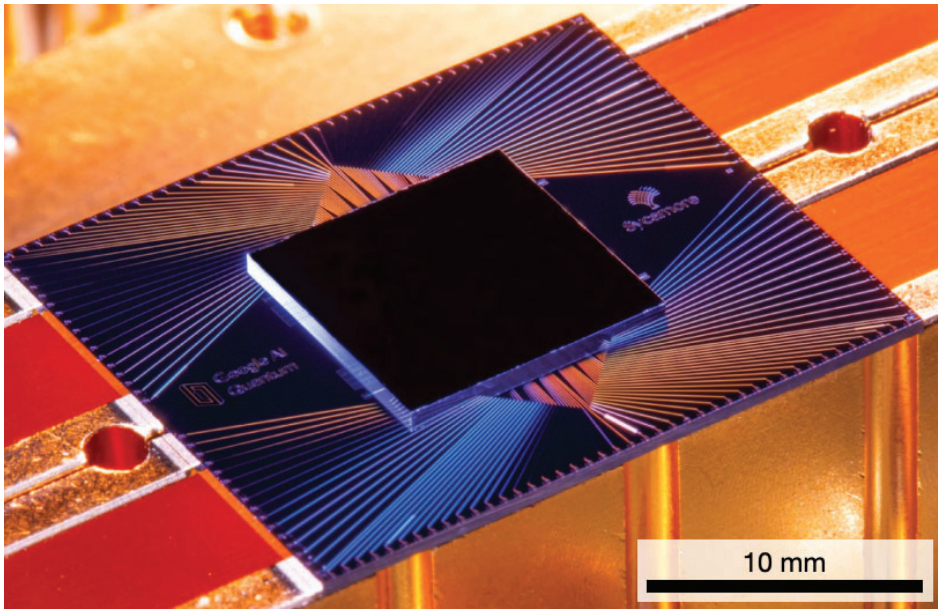



Figure 1-5: Sycamore: 53-bit superconducting quantum processor [3]

this thesis, the Quantum Flux Parametron [5], uses similar building blocks to popular qubit devices, particularly the flux qubit circuit; however, we exploit the quantum properties for their ultra-low energy operation in classical computation.

Similarly in classical superconducting electronics, Ayala et. al. published results on a microprocessor using AQFP logic called MANA (Monolithic Adiabatic Integration Architecture) [26]. This design modified a conventional RISC architecture for AQFP logic using custom EDA tools for automated synthesis and place and route [27, 28]. MANA provides a promising future for AQFP circuit design - it is the most advanced AQFP chip to be made so far and is about 80 times more energy efficient than 7-nm FinFET equivalent technology today, while accounting for cooling overhead [26]. The work in this thesis also proposes VLSI from AQFP logic, although the asynchronous and modular architecture of Super-DICE is very different from the RISC-like approach, and allows for a more scalable end-to-end workflow without complicated custom EDA tools.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 1.3 Contributions

This thesis contributes a set of superconducting logic building blocks along with an asynchronous token buffer mechanism for communicating between these logic blocks, which serves as a toolkit for building modular, asynchronous, and scalable ultra-low power superconducting computing systems.

In Chapter 2, I walk through a deeper background on superconducting electronics and the physical mechanisms of Josephson junctions and flux-transfer devices. I describe the current and energy characteristics of the quantum flux parametron, which is the core building block of the entire Super-DICE circuit library. And lastly, I describe the simple modifications to a QFP which can be made to turn the buffer into single-bit logic devices.

Chapter 3 provides more background on Asynchronous Logic Automata with schematics for logic gate building blocks. I also explore two different token buffer designs: the Precharge Full Buffer (PCFB) which directly implements a popular asynchronous buffer from CMOS design in AQFP logic, and the QFP Full Binary Buffer (QFBB) which I've designed for AQFP-optimized token passing.

Chapter 4 then explores more complex circuit building blocks, referred to as logic modules, and discusses where to place the token passing boundary. The token passing boundary draws the line between asynchronous and synchronous computing in the system and determining where to place it is a trade-off between modularity and communication overhead. Designs for full and half adders with varying granularity of token passing are also presented in Chapter 4.

Chapter 5 evaluates all of the circuit designs put forward in the previous chapters. Circuits are simulated in SPICE to verify logic operation and QFP energy performance is extrapolated for circuit level energy projections. It also discusses scalability of these designs to supercomputing size systems and considers cryogenic cooling overhead in detail.

Finally, Chapter 6 covers future plans, potential challenges, and prospective impact.





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

In sum, I hope that the work presented can provide a theoretical basis for scalable superconducting supercomputers that will continue to experimental development in future work.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 2

# Ultra-low Energy Superconducting Logic

This chapter will provide a brief background to the history of superconducting electronics, the basic physical principles behind superconductive devices and circuit logic families, and a deeper dive into the quantum flux parametron (QFP) logic device. The QFP is the building block for the computing architecture proposed in this thesis.

A material is classified as superconductive if a current can flow through it with no electrical resistance and it expels magnetic fields, i.e. it exhibits the Meissner effect. Superconductivity was initially discovered by Heike Kamerlingh Onnes in 1911, when he and his team reported no measurable resistance in mercury when it was cooled below 4.2 Kelvin [29]. For decades there was confusion as to what physical phenomena gives rise to superconductivity, but a theory was eventually explained and widely accepted by Bardeen, Cooper, and Schrieffer in 1957 [30]. At a low enough temperature, the critical temperature, $T_c$, it's more energy efficient for the free electrons in conductive metals to form quasiparticle pairs, commonly called cooper pairs. The cooper pairs are bosons (fermion + fermion = boson), which means that they no longer follow the Pauli exclusion principle, like normal electrons, and can all share the ground state energy. This means there is no scattering between cooper pairs and the crystal lattice phonons, and, therefore, no resistance. Once a current is started, it will continue forever. In depth analysis on the solid state nature of





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

superconducting materials is presented in [31].

Low temperature superconductors (LTS), also referred to as ordinary superconductors, are usually metals that have a transition temperature below 10 Kelvin. There is also a newer class of high temperature superconductors (HTS), mostly made from oxides or ceramic layered materials, which have critical temperatures above 77K. Even warmer, room temperature superconductivity was discovered in 2020 with a carbonaceous sulfur hydride that has a critical temperature of 288K when held at a high pressure of 267 GPa [32].

Although high temperature superconductors are promising for the future of superconductivity, they are not yet relevant for superconducting circuits because tunneling junctions exhibiting the Josephson effect are not mature enough to be stable switching devices in HTS. Therefore, the work in this thesis will focus on LTS, and all proposed chip design requires cryogenic cooling to operate properly.

## 2.1 Josephson Effect

Josephson junctions (JJs) make up the fundamental active device for most superconducting electronics (SCE). JJs consist of two superconducting layers separated by an insulating material which is thin enough for cooper pairs to easily tunnel through. The wave functions on each side of the junction are weakly coupled, resulting in the following fundamental properties for the current, I, and voltage, V, across the junction [31].

$$I = I_c \sin \phi \tag{2.1}$$

$$V = \frac{\Phi_0}{2\pi} \frac{d\phi}{dt} \tag{2.2}$$

Where $I_c$ is the critical current, i.e. the maximum supercurrent, determined by the fabrication process and geometry of the junction; $\phi$ is the phase difference of the electron wave functions across the junction; and $\Phi_0$ is the magnetic flux quantum




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

defined as

$$\Phi_0 = \frac{h}{2e} = 2.07 \times 10^{-15}\text{Wb} \tag{2.3}$$

From these relations, we see that when no voltage is applied, there is a nonzero DC current proportional to the phase difference between the superconductors - this is referred to as the *DC Josephson effect*. And when a constant voltage is applied across the junction, there is an oscillating current with frequency proportional to the applied voltage ($f_j = \frac{1}{\Phi_0}V$) - this is the *AC Josephson effect*. Josephson junctions can be designed with different types of geometry, but a diagram of a common design, along with the circuit schematic symbol, is shown in Figure 2-1.

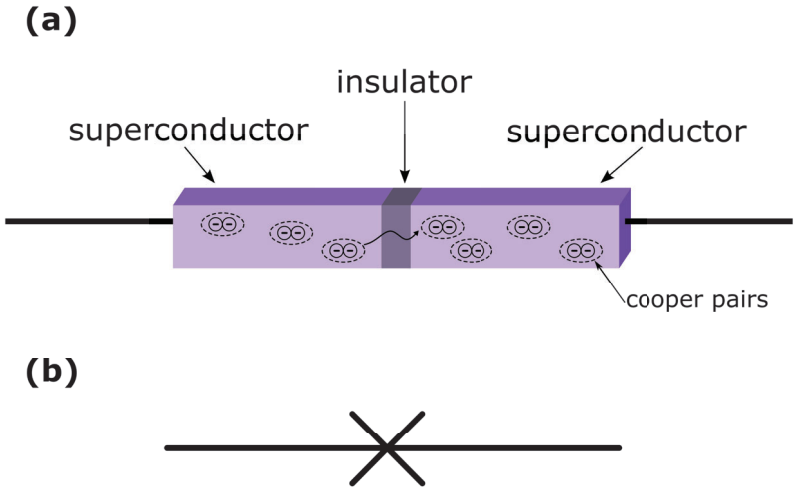



Figure 2-1: Josephson Junction diagram and schematic symbol

In reality, the quantum tunneling behavior through the junction is more complicated than Equation 2.1 describes. At finite temperatures, there is a bit of resistive current due to the insulator and Bogoliubov quasiparticle tunneling (Bogoliubov quasiparticles are excitations caused by single electrons paired with opposite energy holes, instead of being paired with another electron like in cooper pairs). There is also an added displacement current due to capacitive effects from the junction geometry, because its basically a parallel plate capacitor. We define an equivalent circuit


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

model, referred to as the Resistive and Capacitively Shunted Junction model (RCSJ), which is the simplest approximation for a realistic junction, composed of an ideal resistor, capacitor, and junction [31]. The RCJS model schematic is given in Figure 2-2. Therefore, a more accurate description of the junction current is

$$I = I_c \sin\phi + GV + C\frac{dV}{dt} \tag{2.4}$$

Which can be simplified to terms of $\phi$ using the second Josephson relation

$$I = I_c \sin\phi + G\frac{\Phi_0}{2\pi}\frac{d\phi}{dt} + C\frac{\Phi_0}{2\pi}\frac{d^2\phi}{dt^2} \tag{2.5}$$

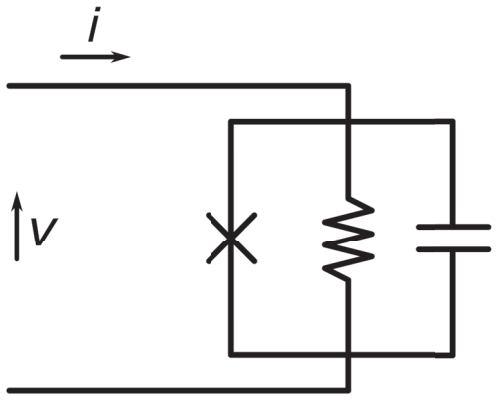



Figure 2-2: RCSJ Equivalent Circuit Diagram

When the current across the junction exceeds the critical current, then the junction is no longer superconducting and a nonzero voltage, described by the second term in Equation 2.4, appears. Figure 2-3 shows the voltage-current characteristics of a Josephson junction as simulated with the standard RCSJ model in WRSPICE, an open-source circuit simulation tool with support for superconducting electronics. The nonlinear, amplifying characteristics of the JJ I-V curve makes them a viable switching device for voltage-level encoded logic. However, as I will explore in the following sections, voltage levels are not the only way to flip energy states in a JJ.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

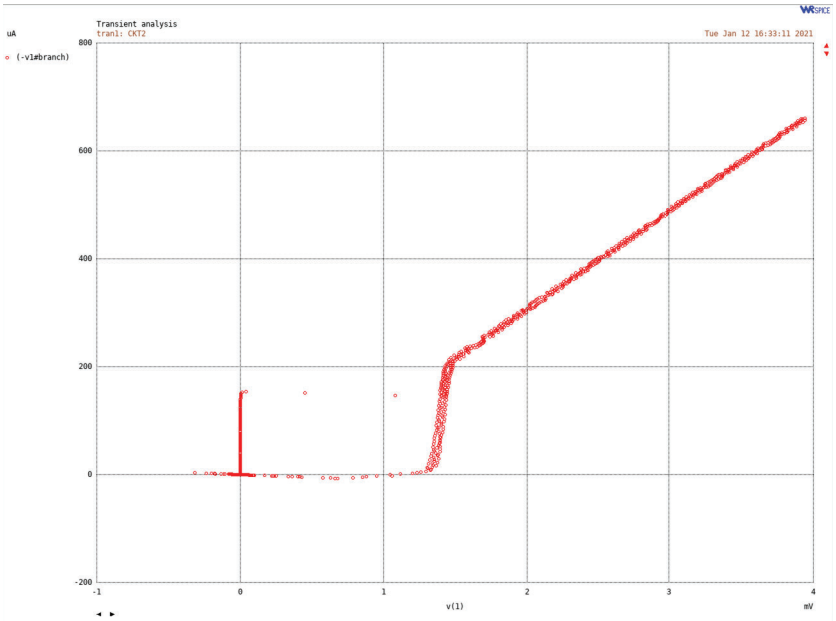



Figure 2-3: Junction I-V SPICE Plot

## 2.2 Superconducting Logic Families

Superconducting logic families can be grouped into two different classes based on how they store information: there are flux-transfer devices and voltage-transfer devices. Flux transfer devices encode information in units of magnetic flux stored in superconducting loops, while voltage transfer devices use the JJ I-V characteristics to switch the junction between it's superconducting and resistive state. A diagram of existing logic families and how they give way to one another is shown in Figure 2-4.

A driving motivation for most classical computers made with superconducting electronics is the ultra-low power dissipation and ultrafast switching speeds that superconducting conditions can maintain. Historically, a lot of work has gone into the promise of ultra-low power superconducting electronics, but there has been minimal reward. Most notably, was the IBM superconducting project which started in the early 1960s and ran until it was officially canceled as a billion-dollar "failure" in 1983 [33]. The program designed and fabricated the first Josephson junction switches, originally called tunneling cryotrons, which alternated between cooper pair tunnel-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

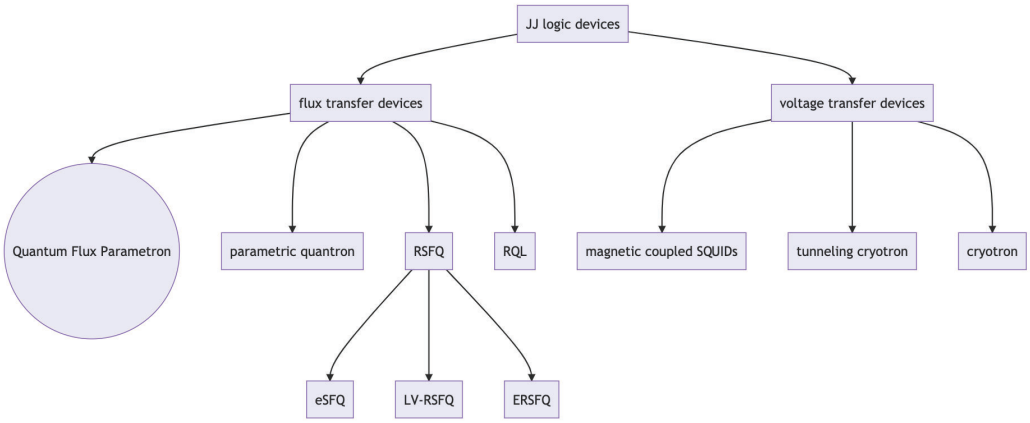



Figure 2-4: Superconducting Logic Family Tree

ing at V=0 and single-particle tunneling at a non-zero, but sub-gap, voltage [34]. Although it did drive a lot of monumental research discoveries for IBM, it was eventually abandoned because of size scaling limitations in the inductor-dependent device and "punch-through" errors (the junctions needed to be reset after each clock cycle, when the reset fails it's refered to as a punch-through error). All the while, semiconductor bipolar transistors and eventually MOSFET transistors were gaining in popularity with seemingly minimal foreseeable contingencies, so Josephson junctions lost the race [33].

However, superconducting fabrication research did not come to a full stop despite IBM's expensive and public failure, the Japanese Ministry of International Trade and Industry (MITI) funded a superconducting project from 1981 through 1989, which set the stage for much of the parametron-based superconducting logic used in the rest of this thesis [35].

For most of the 90s and early 2000s, the rapid single flux quantum (RSFQ) [36] was the most common superconducting electronic logic. The RSFQ had, and still has, a strong following because of it's robust stability, low energy flux-driven switching, and unique state storing abilities. Each RSFQ cell stores information in a superconducting loop with a Josephson junction, and the presence or absence of a single flux quanta (or fluxon) determines the state of the bit. Computation is done by transferring fluxons between RSFQ cells. RSFQ logic has also long promised energy efficiency superior to


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

CMOS technology at scale; however, after decades of research, performance returns on a RSFQ supercomputer is still met with skepticism. The skepticism originates from low fabrication margins, difficulty in scaling bias currents which are needed to reset the RSFQ cells, and power performance gains that are too small to justify the effort [37]. To improve some of the power performance drawbacks with RSFQ, new energy efficient spin-offs have been developed in more recent years, such as LR-RSFQ [38] and ERSFQ [39], and eSFQ [40]. However, all of these merely change the dc-biasing schema for RSFQ circuits with tricks to improve static energy dissipation and doesn't address the root of RSFQ issues or change the dynamic power dissipation.

Reciprocal Quantum Logic (RQL) [41] is another flux-transfer logic family. It is lower in power dissipation than any RSFQ variant because it removes static power dissipation by biasing junctions with an ac-current transformer instead of a resistor. However, it still performs computation by transferring fluxons between junction loops, so it cannot be as low energy as Adiabtic Quantum Flux Paraemtron (AQFP). AQFP logic devices can achieve picosecond gate delays with zeptojoule ($10^{-21J}$) switching energy by adiabatically switching the location of a single flux quanta in a double-potential well.

Given that the driving motivation for this work is ultra-low power dissipation distributed across multiple asynchronously communicating chips, the decision of superconducting logic technology was determined based on bit energy, which is compared in Table 2.1. AQFP is the clear leader in lowest power requirements, operating almost at the theoretical Landauer Limit [42], on the order of $kT$. Detail and operation of AQFP design will be explored through the rest of this thesis. Mukhanov et. al. provides a more comprehensive overview and comparison of SFQ logic families [43], although AQFP is left out of this analysis.

## 2.3 Junctions and Loops

When describing flux transfer devices, it's useful to define the generalized flux angle, $\Phi$, as the time integral of voltage, and analyze circuits based on their current-




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Table 2.1: Bit-Energy Comparison of SCE Logic Families

| Logic Family | Switching Energy (aJ) |
|---|---|
| 90 nm CMOS [44] | 2,620 |
| 7 nm CMOS [44] | 111 |
| RSFQ [43] | 0.15 |
| RQL [41] | 0.68 |
| AQFP [26] | 0.0014 ($23k_BT$) |

The values above are given for different circuit designs and taken from various energy review papers, so they do not provide a convenient apples-to-apples comparison; however, general performance is clear. Note that static power is also not account for in these values and it can account for significant overhead. This is why practical RQL circuits are much more energy efficient than RSFQ, even though this table shows RQL having a larger switching energy.

flux characteristics.

$$\Phi = \int v dt \tag{2.6}$$

Using the second Josephson relation (Equation 2.1) and the magnetic flux quantum, the generalized flux angle can also be described by [4]

$$\Phi = \frac{\Phi_0}{2\pi}\phi \tag{2.7}$$

To understand future analysis of QFPs, it's easiest to begin by describing the current-flux characteristics of a superconducting loop, large enough to hold a non-zero amount of flux quanta, with a single Josephson junction, as described by the schematic in Figure 2-5. Note that by convention, a positive phase, $\phi$, moves in the opposite direction of the current. Also, this figure includes an externally applied current, $i$, which, for now, will be considered to be zero.

Like all circuits, Kirchoff's law must hold such that $I_{junc} + I_{ind} = 0$. Assuming static conditions, the only flux in the circuit arises from the self inductance of the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

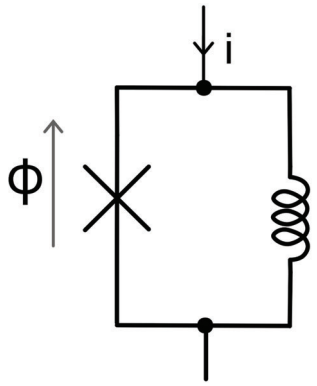



Figure 2-5: Superconducting Loop with Josephson Junction

loop

$$\Phi = LI_{ind} \tag{2.8}$$

And the junction current is given by the first Josephson relation (assume static conditions, which simplifies the RCSJ model). Therefore, the current equation for this circuit can be expressed as

$$I_j \sin\phi + \frac{\Phi_0}{2\pi}\frac{\phi}{L} = 0 \tag{2.9}$$

To determine the stable state of the circuit, we can analyze the energy, given that $U = U_{ind} + U_j$. The energy for an inductor is defined as

$$U_{ind} = \frac{1}{2}LI^2 = \frac{1}{2}\frac{\Phi^2}{L} \tag{2.10}$$

$$U_j = \int v \cdot i \, dt = -I_c \frac{\Phi_0}{2\pi}\cos\phi \tag{2.11}$$

Introducing an energy normalization factor, $E_j$, and inductance normalization factor, $L_j$, to simplify the equation, the Hamiltonian of a superconducting loop with a


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

tunneling junction is

$$u = \frac{U}{E_j} = \frac{1}{2}b\phi^2 - \cos\phi \tag{2.12}$$

where

$$E_J = I_c\frac{\Phi_0}{2\phi} \tag{2.13}$$
$$L_j = \frac{\Phi_0}{2\pi I_c} \rightarrow L = L_j/b$$

Local extrema can be found by taking the first derivative of the Hamiltonian with respect to phase, which also returns the current equation.

$$\frac{du}{d\phi} = 0 \rightarrow b\phi + \sin\phi = 0 \tag{2.14}$$

And stable points are determined by

$$\frac{d^2u}{d\phi^2} > 0 \rightarrow b + \cos\phi > 0 \tag{2.15}$$

When the loop is driven by some external current, the system is shifted from one stable state to another, such that

$$\sin\phi + b\phi = i \tag{2.16}$$

The applied current can be expressed in terms of a flux angle, $\beta$, given $i = b\beta$. Which corresponds to a shift of $b\phi$ by some amount $\beta$ on the load line.

$$u = \frac{1}{2}b(\phi - \beta)^2 - \cos\phi \tag{2.17}$$

The loop can also be driven by an external flux through a transformer, i.e. coupled inductor. When an external flux is added, the generalized flux angle for the circuit is no longer solely dependent on the loop's self inductance. The applied flux imposes a




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

bias to the phase difference across the junction.

$$\Phi + \Phi_{ext} = \frac{\Phi_0}{2\pi}\phi \longrightarrow \Phi = \frac{\Phi_0}{2\pi}(\phi - \alpha) \tag{2.18}$$

This bias behaves like a shift in the negative sine plot, and with a large enough phase can cause the current to jump states.

$$u = \frac{1}{2}b\phi^2 - \cos(\phi - \alpha) \tag{2.19}$$

## 2.4 Quantum Flux Parametron

The adiabtic quantum flux parametron (AQFP) is the building block of the superconducting architecture proposed in this thesis. Unlike RSFQ which passes fluxons from loop to loop, the QFP encodes information through the location of a single fluxon in a double-well potential. Information propagates when neighboring QFPs push the the other's fluxon into the appropriate well for computation. The QFP is derived from the parametron device, first proposed by Eiichi Goto as a digital logic component in 1954 [45].

It's most intuitive to describe the parametron through a mechanical analogy [4]. Imagine a marble inside a flexible bowl resting on a fulcrum. When no force is applied to the top of the bowl a marble sits at the single minimum above the fulcrum. Then, when a force is applied, the bowl folds around the fulcrum such that there are now two potential minimums for the marble to roll into, and the selection of state is based on some preset input direction of the marble. This analogy is animated in Figure 2-6. Applying the force to the top of the bowl sets the parametron into the active state. When activated, the input signal is amplified with theoretically infinite gain and is stored in one of the two potential minimums for as long as the parametron remains active. When the force is removed, the parametron is inactive and the state is erased, like the bowl returning to it's initial shape with the marble rolling into the single minimum.

QFPs operate similarly in that they are driven by ac activation signals which





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

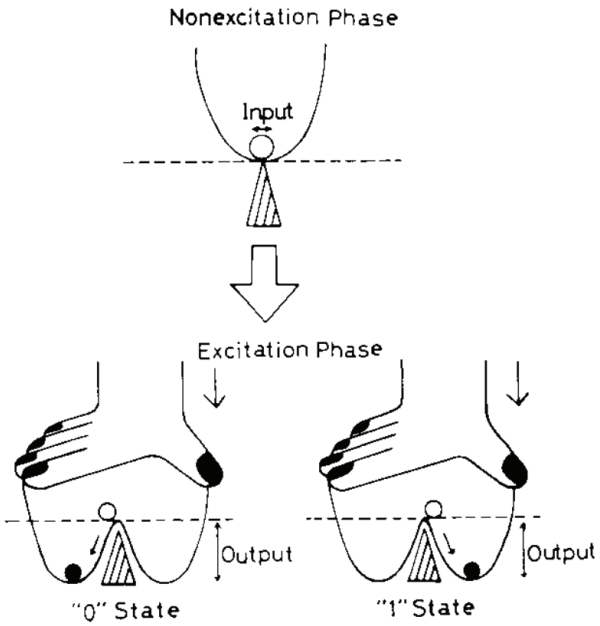



Figure 2-6: Mechanical QFP Analogy, from [4]

create an unstable equilibrium at a previously stable point, causing the system to "fall" into one state or another based on some input phase angle. The device is made of two superconducting loops, each with a Josephson junction, a load inductor, and an activation transformer with opposite parity biases. A schematic for the QFP is shown in Figure 2-7.

Following similar analysis from the previous section, the total energy for the QFP is

$$U = \frac{1}{2}b(\phi - \beta)^2 - \frac{\Phi_0}{2\pi}I_j\cos(\phi - \alpha) - \frac{\Phi_0}{2\pi}I_j\cos(\phi + \alpha) \tag{2.20}$$





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

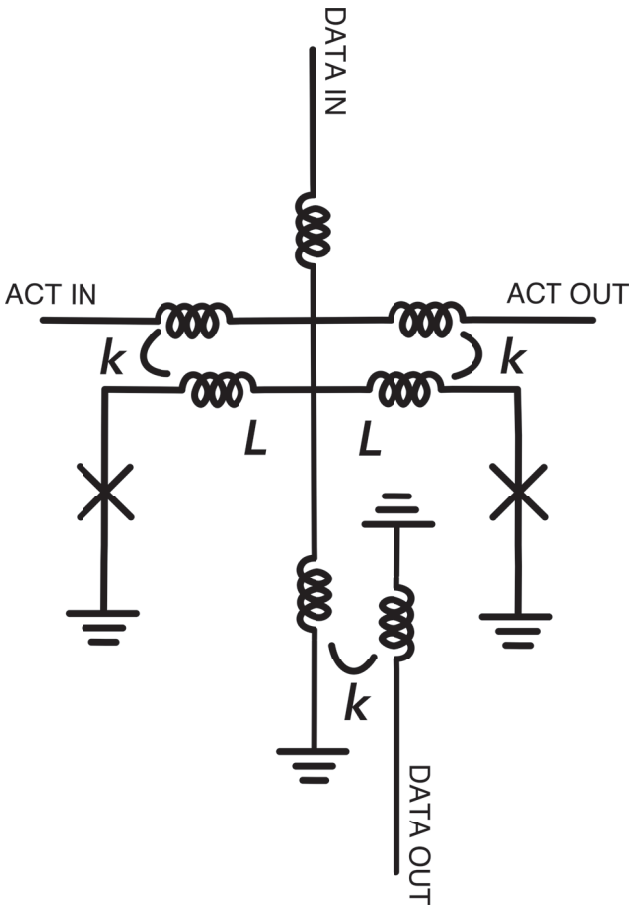



Figure 2-7: QFP Schematic

Introducing new normalization parameters to simplify the equation gives

$$u = \frac{1}{2}b(\phi - \beta)^2 - \cos\phi\cos\alpha \tag{2.21}$$

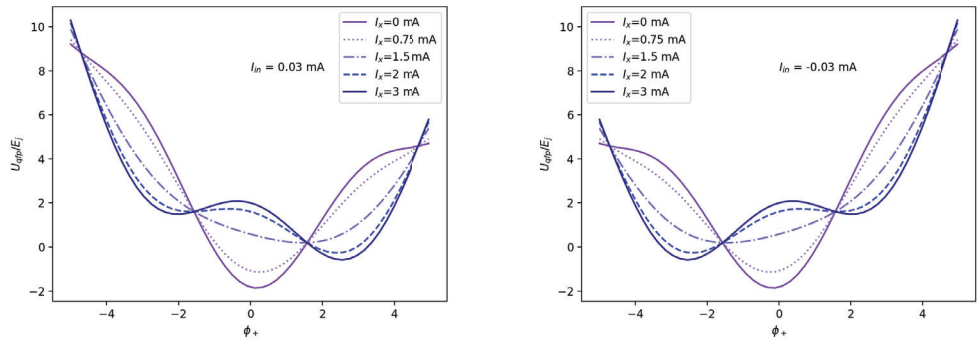
where

$$E_Q = 2I_c\frac{\phi_0}{2\pi} \tag{2.22}$$
$$B = \frac{L_Q}{L} \rightarrow L_Q = \frac{\Phi_0}{2\pi}\frac{1}{2I_c}$$

where $\beta$ is the input flux angle, $\alpha$ is the transformer activation flux angle, and $\phi$ is the output flux angle. A plot of the Hamiltonian is given for varying activation phase


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

angles in Figure 2-8. Figure 2-8a shows the QFP switching to the “1" state and 2-8b show the “0" state, based on the input current signals being high (positive) and low (negative), respectively.




(a) Value "1" (b) Value "0"

Figure 2-8: QFP energy-phase plots

The relationship between the output flux and the activation flux angle for various input values is shown in Figure 2-9. This plot demonstrates how the QFP is able to store memory when the activation angle remains high, since the value of the output flux remains high even when $\beta$ is removed.

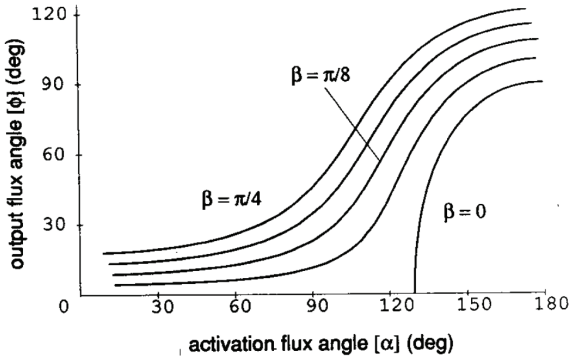



Figure 2-9: $\phi$ output vs activation signal, $\alpha$, from [5]

It’s important to note that QFPs are two-terminal devices, meaning that there’s no distinction between input and output. Because of this, they can be used for reversible computing, but it also complicates combinational circuits since back propagation of


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

outputs, aka relay noise, can interfere with the correct operation. Therefore, when propagating flux values through a chain of QFP devices, it's best to use at minimum a three-phase activation signal so that QFPs alternate between active, quenching, and blocking stages. Each buffer passes through cyclical stages: holding (holding data value), firing (receiving data signal), and blocking (inactive to stop back propagation of data). These stages and where they occur in the activation signal cycles are shown in Figure 2-10

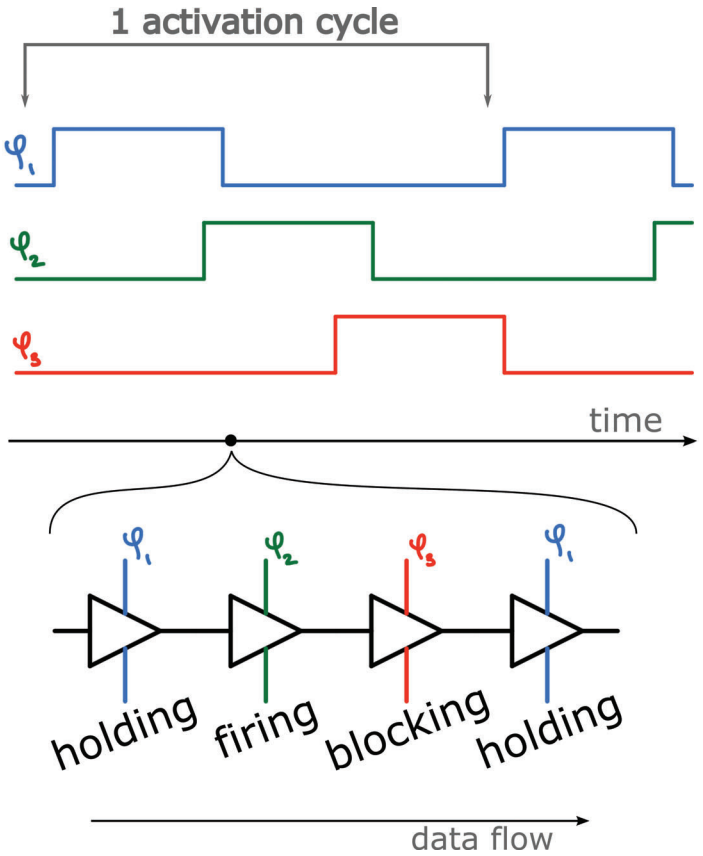



Figure 2-10: Three-phase activation signal propagating through QFP buffers

All previous analysis and explanation has assumed ideal conditions for QFP operation. However, in reality, there are various internal and external issues that can lead to errors and improper output. A common issue is imbalance between the junction critical currents, referred to as $\delta I_c$ noise, arising from various inconsistencies in the fabrication process. Large $\delta I_c$ can cause the stable state of the superimposed JJ loops


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

to be shifted from 0, meaning that the wrong output could be amplified if the input signal is not strong enough to overcome the imbalance bias. Other common internal issues arise from activation transformer imbalance, transformer self inductance, and antagonistic coupling between transformers. Additionally, external issues from neighboring QFPs or additional circuit elements such as the relay noise discussed earlier, homophase noise (i.e. clock skew between QFPs sharing activation phases), and input signal fluctuations can also cause errors. Most of these errors can be accounted for by strengthening input signals and activation transformers and improving fabrication engineering, although most corrections introduce important performance trade offs. More details on QFP optimizations through auxiliary circuits can be found in Hioe and Goto's QFP textbook [46].

## 2.5 QFP Logic Cells

Slight modifications can be made to the original QFP cell to change it's function from an amplifier to an inverter or constant output. Combining this with the ability for wired majority logic through branches of inductor lines, QFP circuits are capable of universal logic. The basic building blocks for a QFP cell library were originally optimized by Takeuchi et. al. in [47]. Junction-level schematics and abstracted symbols for all elements are shown in Figure 2-11.

The original QFP schematic serves as the buffer cell, amplifying the input signal with the same parity. The inverter cell is the same as the buffer, except that the coupling constant for the output signal is inverted so that the final output has the opposite parity of the initial input. And finally, the constant cells will return either a 0/low or 1/high output regardless of input signal due to an imbalance between the activation transformers.

Additionally, a 3-to-1 branch cell is used for a wired majority gate. Expressed in boolean logic, the output of the branch cell is determined by $x = MAJ(A, B, C) = AB + BC + CA$. With the majority gate AND, NAND, and OR gates can all be easily developed by passing $A$, $B$, and $C$ input values through different combinations


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

of buffers, inverters, and constants. These higher level logic gates are described in Section 3.2.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

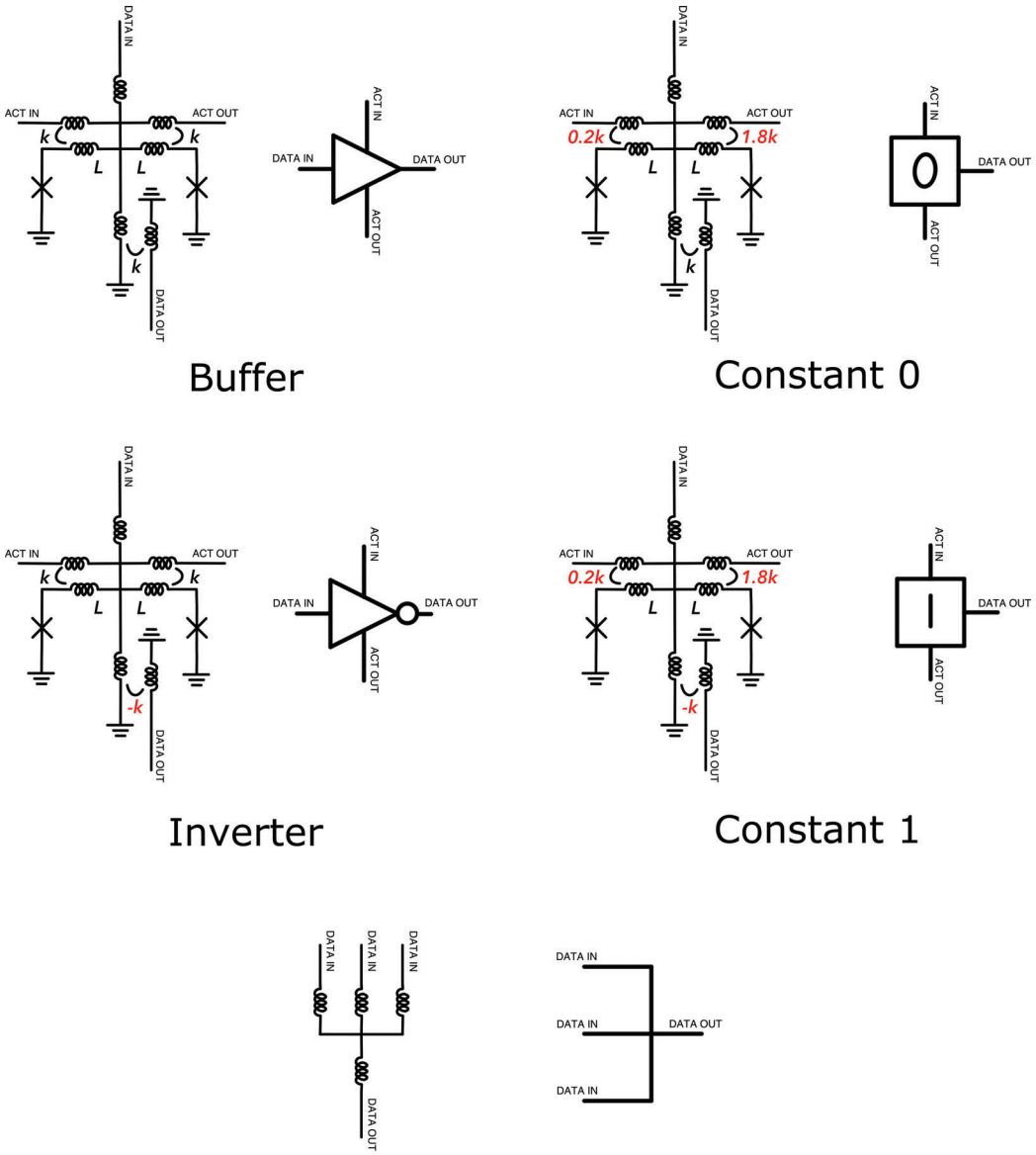



Figure 2-11: AQFP Cell Library. Junction-level schematics and logic-level schematics on the left and right, respectively.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 3

# Asynchronous Superconducting Circuits

This chapter begins by describing previous work done at CBA on asynchronous architectures; then extends the superconducting technology described in the previous chapter to fit this asynchronous framework; and finally, focuses on the token passing (aka handshaking) mechanism required for asynchronous communication.

## 3.1 Asynchronous Logic Automata

Asynchronous Logic Automata (ALA) is a spatial computing framework composed of finite volumes of information communicating through locally exchanged state tokens [13]. Inspiration for the design comes from Neil's infamous quote: "Computer science is one of the worst things to happen to computers or to science" because a fundamental error was made when computer science broke away from physics by directly implementing the Turing machine dataflow with the, now ubiquitous, von Neumann architecture [48]. The universal Turing machine is the ideal theoretical model of a computer. It consists of a finite state machine that performs operations based on input instructions read from an infinite memory tape. Most modern processors today implement the von Neumann architecture or Harvard architecture, the difference between the two being shared or separate data buses for reading/writing data from/to




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

memory. They both separate the memory unit from the control/processing unit, inspired by the Turing machine hierarchy.

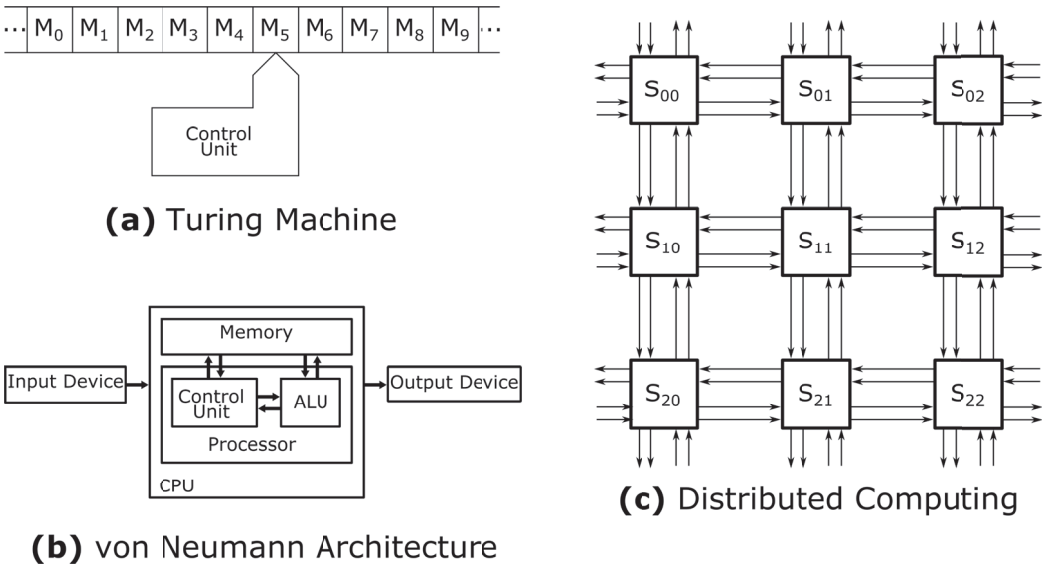



Figure 3-1: Architecture diagrams of traditional computing vs distributed computing (like ALA)

This separation of memory and processing is vastly different from how we understand physical information to propagate through distributed networks. In any biological computational system, state is not stored separately from interactions, but modern silicon computers do exactly that by separating processing from memory. ALA better mimics the physics of information processing by performing computations through a lattice (2D or 3D, depending on design) of heterogeneous automata all asynchronously exchanging state through tokens. As discussed in the introduction, cellular automata architectures have been explored and built before [17], however ALA is unique in it's ability to scale due to it's uniform representation across many layers of abstraction.

Figure 3-1 shows a block diagram of a distributed computing system, like ALA, compared to the von Neumann processor dataflow.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### 3.1.1 ALA Cell Library

The heterogenous automata that comprise ALA are simple logic gates and token manipulators shown in the cell library, Figure 3-2. The logic gates behave has simple combinatorial finite state machines. They receive one or two input data values, perform some computation, and pass along the output. Each cell completes on a single unit time step. Computation is only executed if the input/s is/are present and the output is empty, which is verified by an acknowledgment signal between the cells input and output. This ensures that proper sequential logic is maintained, with the added benefit that power is not required to do nothing (as is the case in most clocked architectures).

Communication lines between cells indicate present/absent and high/low data values. This can be done with dual-rail logic or binary logic, both requiring at least two wires for communication. Dual rail logic encoding has each wire representing a different data value, meaning that the data signal will always be high and the value depends on which line the signal comes from. In this protocol, it's not valid for both wires to be high at the same time, and if both are low then the value is null and the token is absent. On the other hand, binary logic encoding has one wire encoding data with true/1/high and false/0/low while the other indicates the present/high or null/low value. In this protocol, the data line is only checked when the present line is high. The method of encoding depends on the underlying technology implementation and does not change the functionality of the ALA cells, so more attention will be given to this in the following sections.

Finally, there are also token manipulating cells which perform actions like copying or deleting tokens. Both these cells have a control input and a data input - the control input determines whether or not the delete/copy action is performed while the data line holds the token value. If a token is being copied, then the token will be propagated without sending an acknowledgement signal at the input, causing the same token to be propagated again at the next time step. If a token is deleted, then the reverse happens; the acknowledgment signal will be sent to the input without propagating


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

the token to the output.

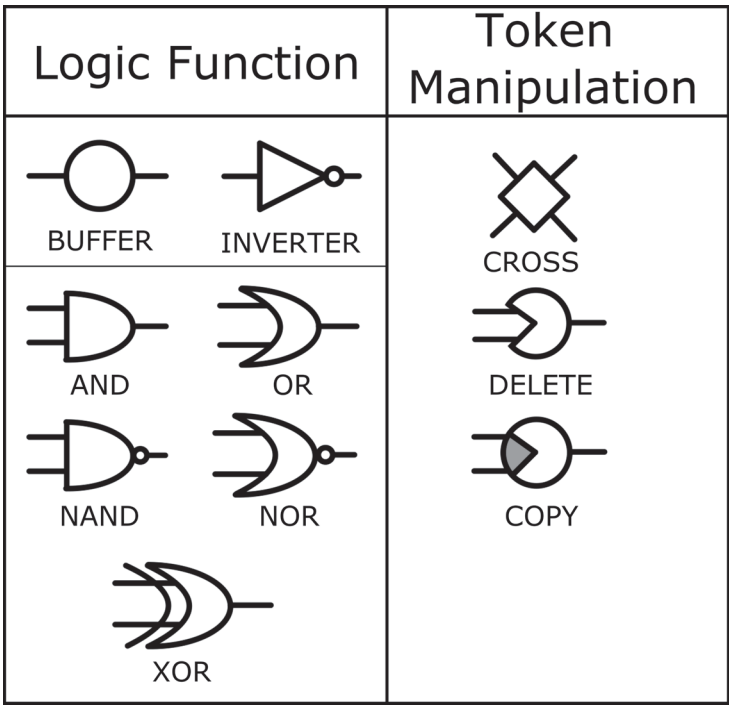



Figure 3-2: ALA cell library

Importantly, ALA is a representation of computation that can be implemented in any underlying fabrication technology. Forest Green's masters thesis work designed an ALA cell library in 90nm CMOS technology [6], which has served as a nice example for the work in this thesis with superconducting electronics. Green's thesis laid the groundwork for ALA implementation, while demonstrating the benefits of simple, scalable ASIC design. Figure 3-3 demonstrates the one-to-one mapping between ALA schematic design and layout with an LFSR (Linear-Feedback Shift Register) circuit. Unlike traditional circuit design, once each cell type is designed and optimized, nearly any circuit can be built by tiling the heterogeneous set of cells through straight forward component integration. This dramatically reduces the work required for designing ASICs (application-specific integrated circuits) by making layout and logic design one in the same.

A major drawback of CMOS ALA was the power and speed overhead that gate-




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

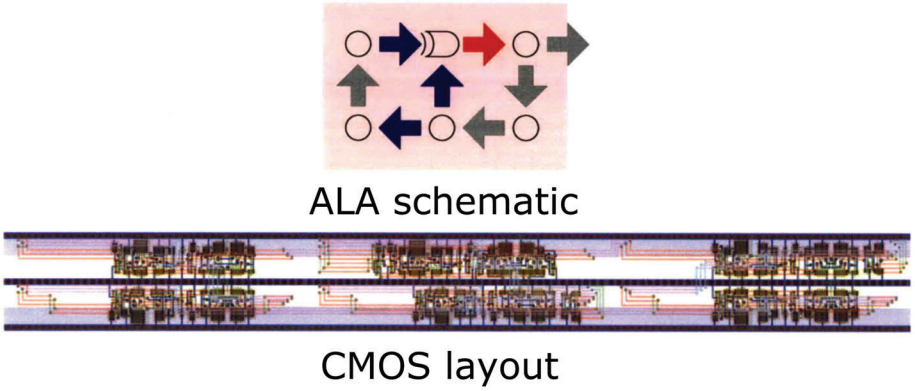



Figure 3-3: LFSR ALA schematic and layout from [6]

level token passing introduced. Projected power performance from circuit simulations for the ALA LFSR and a custom FPGA LFSR design resulted in the ALA circuit requiring 36.6 times more energy and 4.36 times more transistors than the FPGA circuit [6]. This is why much of the foreword looking work for the ALA project proposed better token passing mechanisms with SRAM cells for lower overhead.

The key benefits of CMOS ALA are clearly not in power, area, or speed performance of the finished circuit, but rather the design scalability, the ability to rapidly prototype custom chip designs, and the larger algorithmic power wins provided by the hardware and software alignment. Importantly, these are exactly the strengths needed for superconducting electronics. The QFP provides dramatic improvement in power and speed compared to traditional CMOS switches, so we don't mind an order of 10 times overhead introduced by tokens. Instead, we need the scalability for rapid designing and prototyping of AQFP circuits and we need the asynchronous modularity to remove the logical boundary between on-chip vs off-chip communication to bypass JJ count limitations on a single chip. These trade-offs will be explored deeper in the next section.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 3.2 Superconducting ALA

As explored in Section 2.2, there is a long history of interest in superconducting electronics with promising projections for ultra-low energy computers; however, making those promises into a mass-manufacturable reality has repeatedly failed. A majority of the failures have been due to design and fabrication engineering delays that couldn't compete with CMOS scaling wins, i.e. Moore's law. With the power and area scaling of CMOS transistors slowing down today, superconducting electronics reemerge as a promising alternative for ultra-low power circuits. However, billions, if not trillions, of dollars and decades of research and commercialization have built up the CMOS ASIC design pipeline used today. SCE does not have the same Electronic Design Automation (EDA) toolchain, history, or talent pool to compete with this right away. This is a significant bottleneck for superconducting circuit design, made clear by the SuperTools IARPA program launched in 2016 and still in operation today [49]. Furthermore, SCE designs are limited in the josephson junction density that can fit on a single chip, therefore large scale designs require complicated multi-chip modules.

ALA provides a solution to these bottlenecks by transforming the task of hardware design into a discrete, LEGO-like puzzle. The one-to-one mapping between schematic logic and hardware layout allows for simplification of automated place-and-route EDA tools, while improving design verification, tape-out costs and yield. Automated place and route tools in use currently accrue significant area overhead for dedicated routing tracks to make layout match schematic design; ALA significantly minimizes this overhead. Furthermore, the asynchronous, modular nature of ALA cells provides an easy ability to break design across multiple chips [6]. Finally, ALA abstraction allows complex logic architecture to happen in conjunction with device design; therefore, VLSI architecture design for superconducting circuits does not need to wait for fabrication issues to be solved.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### 3.2.1 AQFP Logic Gates

The first step to implementing ALA with AQFP technology, is to build the basic logic gates needed for the ALA library using the QFP logic cells from Section 2.5. QFP-level implementation schematics are given for each logic gate in Figure 3-4

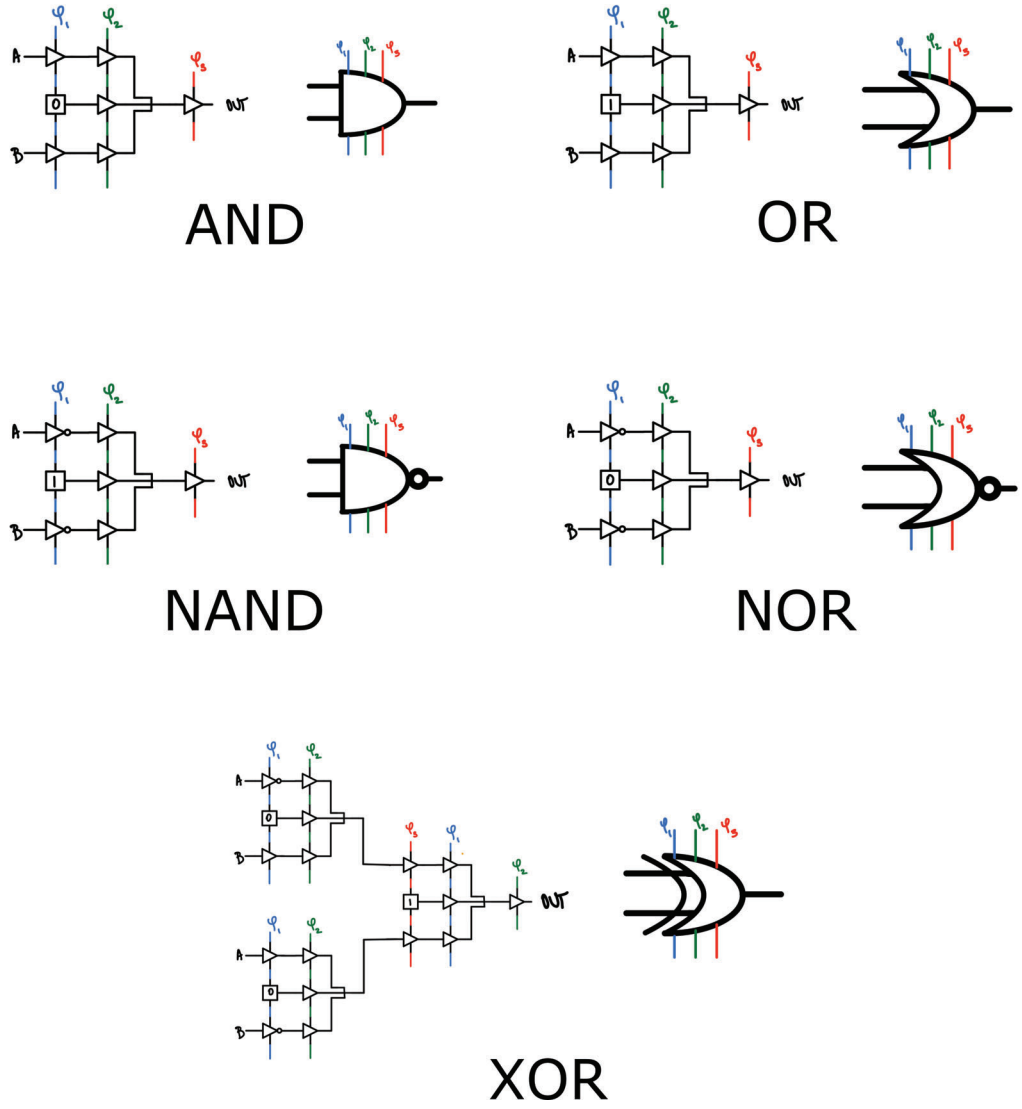



Figure 3-4: AQFP Logic Gates for ALA Library

Each of the designs, except the XOR which is slightly more complicated, consists of two inputs paired with a predefined constant and passed through a majority gate


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

to produce the proper logical output. Three-phase activation signals are designated by $\varphi_1$, $\varphi_2$, and $\varphi_3$ input and output lines. The constant cells have a slightly smaller output current, so an extra buffer firing on $\varphi_2$ is added to each of the majority gate input lines to makes sure that each value has an equal vote. Each AND, NAND, and OR gate takes 1 full activation cycle to run. The XOR gate is a bit more complicated because it's made from two AND gates and an OR gate: $XOR(A, B) = A*B+AB*$. The XOR gate is therefore more expensive in power and area, and requires 2 full activation cycles to complete.

Note that, all of these logic gates expect input data to arrive on the first activation phase, $\varphi_1$. The simple gates (NAND, AND, and OR) will send their output on the following $\varphi_1$, and the XOR gate will output data on the second $\varphi_1$. A "phase synchronizer", which would allow data values to enter the circuit design on any activation phase, is required to remove these design assumptions. This will be expanded on in the following section.

### 3.2.2 Asynchronicity in AQFP

Asynchronous design in AQFP logic is somewhat of a counter intuitive concept and, to the best of my knowledge, has not been done before. Asynchronous superconducting circuits are rather common in RSFQ logic (and its low-energy associated offshoots) due to the way that each RSFQ can save state until a reset is triggered in its superconducting loop. However, as noted in Section 2.2, RSFQ logic has much larger energy requirements than AQFP, and their role in the future of superconducting supercomputers has already been explored in detail with less promising results than AQFP [37].

AQFP technology is not a natural pair for asynchronous circuits because of the required activation signal. The activation signal is also commonly referred to as the clock signal. I have explicitly chosen to not call it a clock signal so that it does not get confused with the type of global clock signal required for synchronous computing logic. The three-phase activation signal needs to be locally ordered as $\varphi_1$, then $\varphi_2$, then $\varphi_3$, with sufficient overlap in order to properly pass bits; however, with the




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

addition of the above-mentioned phase synchronizer and proper data token buffers, there is no requirement that the local phase of a logic cells needs to be aligned with neighboring cells. Therefore, global clock synchronization is completely avoided, which is an important requirement to avoid limitations in system scaling beyond the wavelength of a global clock.

As mentioned with the logic gate, care must be given to ensure that data values are properly passed between neighboring QFPs and are not dropped between clock phases. To facilitate the activation phase agnostic data input, I designed a phase synchronizer, shown in Figure 3-5. It consists of an array of QFP buffers that propagates an input signal on any activation phase to the first phase output of the next activation cycle. SPICE simulation logic verification for the proposed design with be shown in Section 5.1.

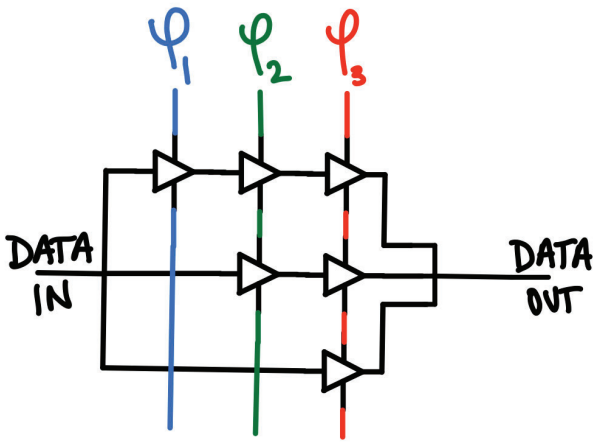



Figure 3-5: AQFP data input phase synchronizer

## 3.3 Token Buffers

In order for sequential logic to execute properly in asynchronous circuits, some type of token passing or handshaking mechanism must be implemented between logic blocks to control the flow of information. Most commonly, traditional asynchronous logic design can be thought of as pipelines of functional blocks separated by buffers


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

which control data flow. Figure 3-6 shows a block diagram of an asynchronous linear-pipeline structure from [50]. Nonlinear pipelines are similar in implementation, except they can have buffers with more than one input and/or output for forks or merges in the data path. The pipeline imposes flow control because at most one token can be stored along each channel and no tokens are lost, so data propagates along the structure relative to gate and wire delay [51].

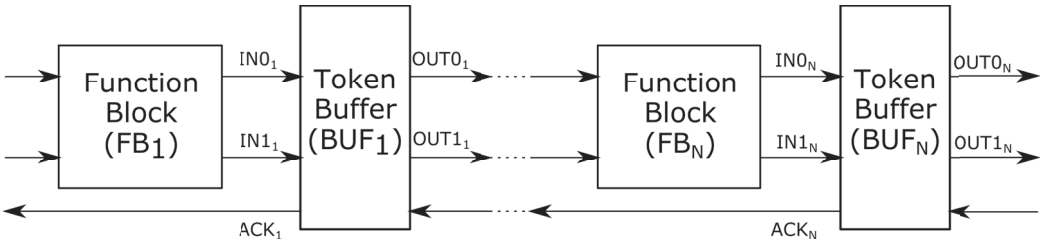



Figure 3-6: Asynchronous Linear Pipeline

Token buffers can be described as being full buffers or half buffers. A full buffer can support distinct tokens on their input and output, while half buffers cannot. For example, in an N-stage FIFO (first in first out) linear pipeline with half buffers, there can be a maximum of N/2 tokens in the system. On the other hand, the same system made of full buffers would be able to hold N tokens.

In its purest form, ALA is a nonlinear asynchronous pipeline with full buffers embedded in each logic gate so that tokens are manipulated and propagated at every step. As I will elaborate on in the token boundary section (Section 4.1), better performance can be achieved when the synchronous functional blocks are expanded and the buffers are not as dense because this decreases the token passing overhead. Regardless of how frequently the buffers are placed, they need to be capable of receiving a data input, passing that input to the logic block, and sending an acknowledgement signal when the block is ready for a new token. To not confuse these buffers with the QFP buffer or the ALA buffer gate, I will refer to these as token buffers.

The decision and design of the token buffer drives circuit power and area performance because it determines the communication overhead required for tokens. Various token buffer implementations have different payoffs. Generally, the most ro-




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Table 3.1: C-element Logic Table

| A | B | C |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | $C_{t-1}$ |
| 1 | 0 | $C_{t-1}$ |
| 1 | 1 | 1 |

bust token buffers make no assumptions about gate or wire delay, which makes them more resilient to noise and deviations in fabrication parameters. However, this robust design comes at the cost of power and area overhead. Token buffers that make assumptions about minimum and maximum gate delays are simpler and less expensive, but require stricter fabrication margins and design assumptions.

### 3.3.1 C-Element Coincidence Buffer

A common component to token buffers is the Muller C-element [52]. The C-element is a coincidence buffer, meaning that it keeps track of whether or not all of its inputs have arrived. It does this by updating it's value only once A and B are equal, otherwise remaining in the previous state value. The truth table is given in Table 3.1 and Boolean logic shown below.

$$C = A.B + A.C_{t-1} + B.C_{t-1} \tag{3.1}$$

The C-element can be naively implemented with NAND logic gates and more optimally implemented with a single majority gate with a feedback loop, shown in Figure 3-7. However, this majority gate implementation is difficult to time and properly verify in simulation design.

There is also an asymmetric c-element, which has extra inputs that trigger a switch in the c-element depending on the direction of data, i.e. low to high vs high to low. The input with a plus (minus) symbol needs to be high (low) for the c-element to


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

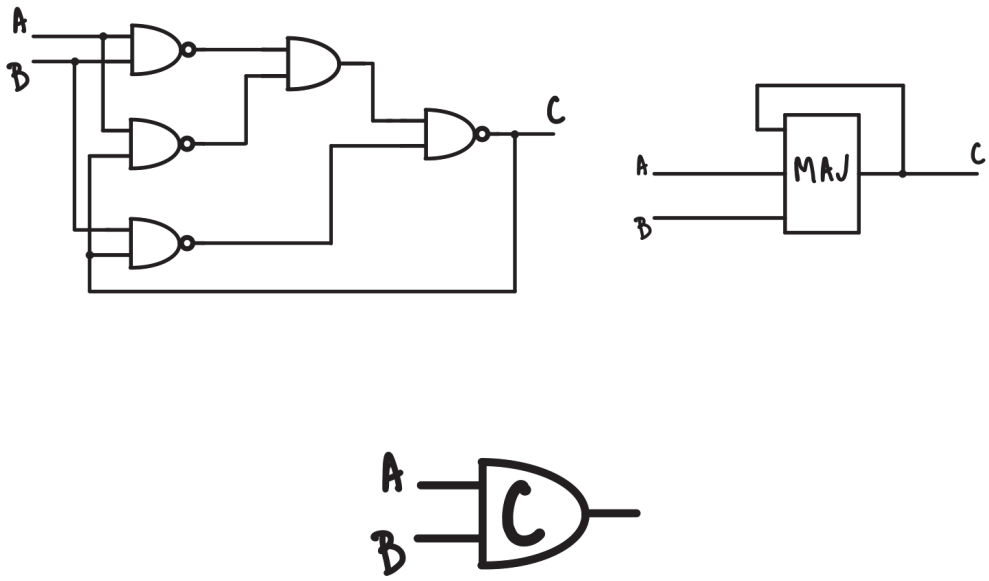



Figure 3-7: C-element Schematics: NAND gate design, Majority gate design, and circuit symbol

change from 0 to 1 (1 to 0). The asymmetric c-element is shown in Figure 3-8, and the modified truth table in Table 3.2.

Table 3.2: Asymmetric C-element Logic Table

| M | P | A | C |
|---|---|---|---|
| 0 | 0 | 1 | $C_{t-1}$ |
| 1 | 0 | 1 | $C_{t-1}$ |
| 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | $C_{t-1}$ |
| 1 | 0 | 0 | $C_{t-1}$ |
| 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 |


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

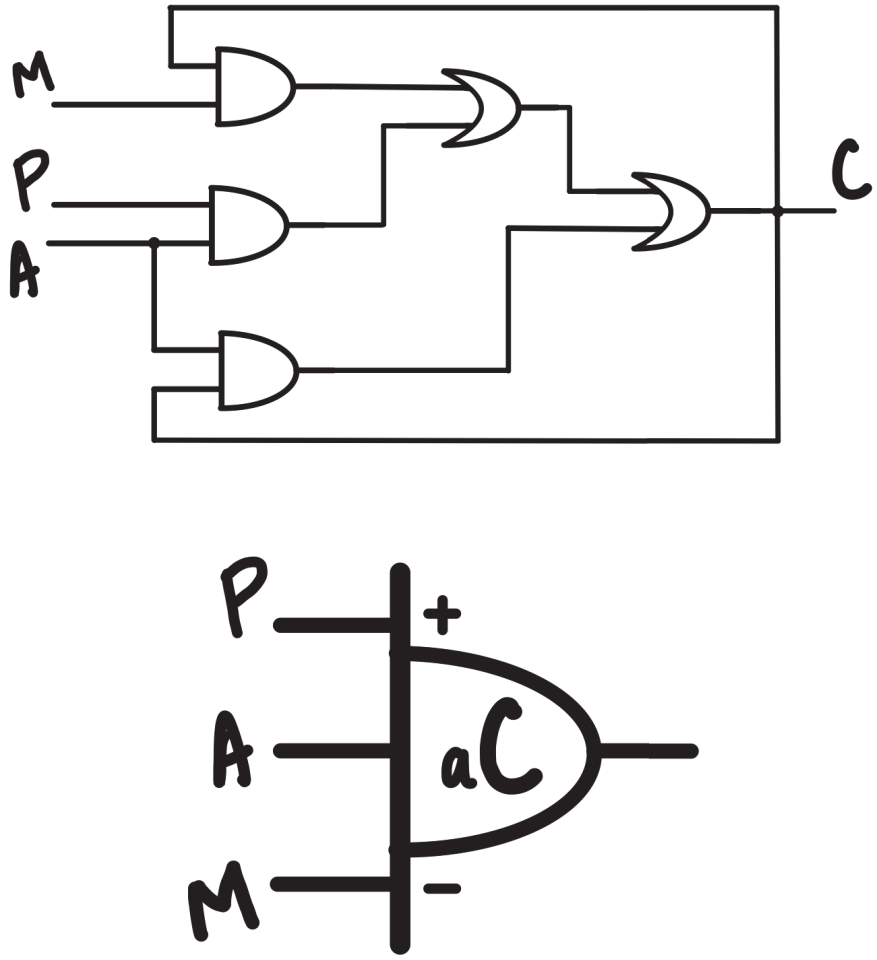



Figure 3-8: Asymmetric C-element schematics and circuit symbol

The C-element gate level schematics shown above are not designed for AQFP logic gets because they do not account for the activation phase timing corrections. Figure 3-9 shows QFP-level schematics for the C-element, where QFP buffers are added throughout the circuit to make sure that data are not dropped. The phase synchronizer array could also be added to each of the inputs if needed; however, this was not done because the C-element alone is not a token buffer because it does not incorporate feedback from neighboring gates or have an acknowledgement signal. The C-element is merely an important building block of the Precharge Full Buffer which


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

will be described next.

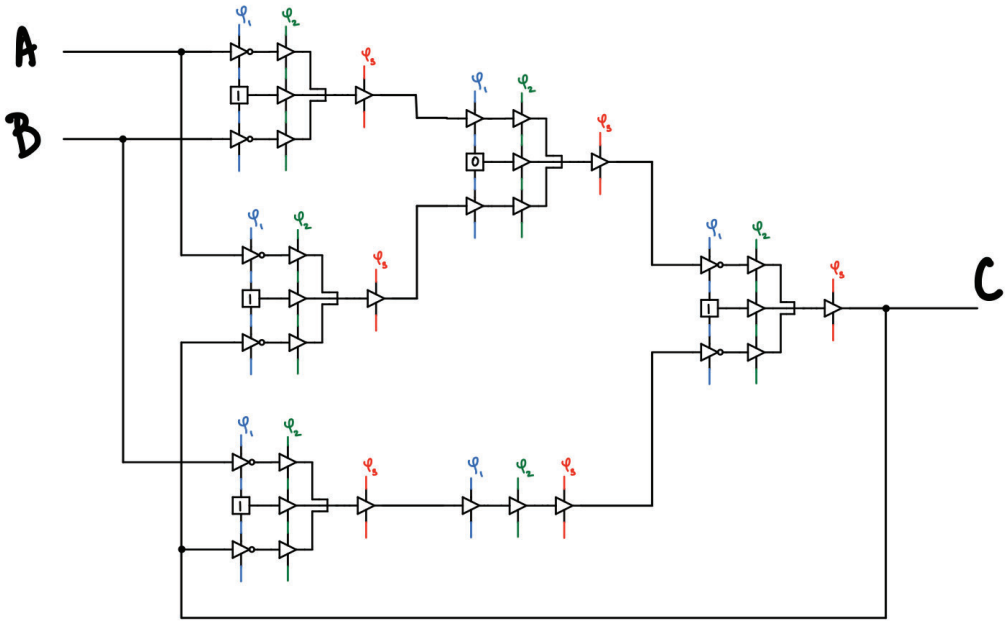



Figure 3-9: QFP-level C-element schematic

### 3.3.2 Precharge Full Buffer (PCFB)

Contrary to the name, the PCFB is not always a full buffer, by the two-token capacity definition [50]; however, as Green showed, it functions as a full buffer for ALA CMOS cells [6]. This token buffer is very much optimized for CMOS transistor design (indicative by "precharge" in the name, implying domino logic), but the basic Boolean logic functionality can be implemented with QFP logic cells. Figure 3-10 gives a logic-level schematic of the token passing mechanism.

The token passing mechanism starts with all data lines low and ack lines high. When a signal comes in on the In0/In1 wires, it triggers C1/C2 to switch from 0 to 1. This brings C3, and therefore InAck, low via the NOR gate on the outputs. Pulling InAck low resets the data input values because it is linked to the OutAck signal in the left-hand environment. Once the data is reset, InAck pulls back high via the NOR output on the input signals. Similarly, the output signals are reset by OutAck once




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

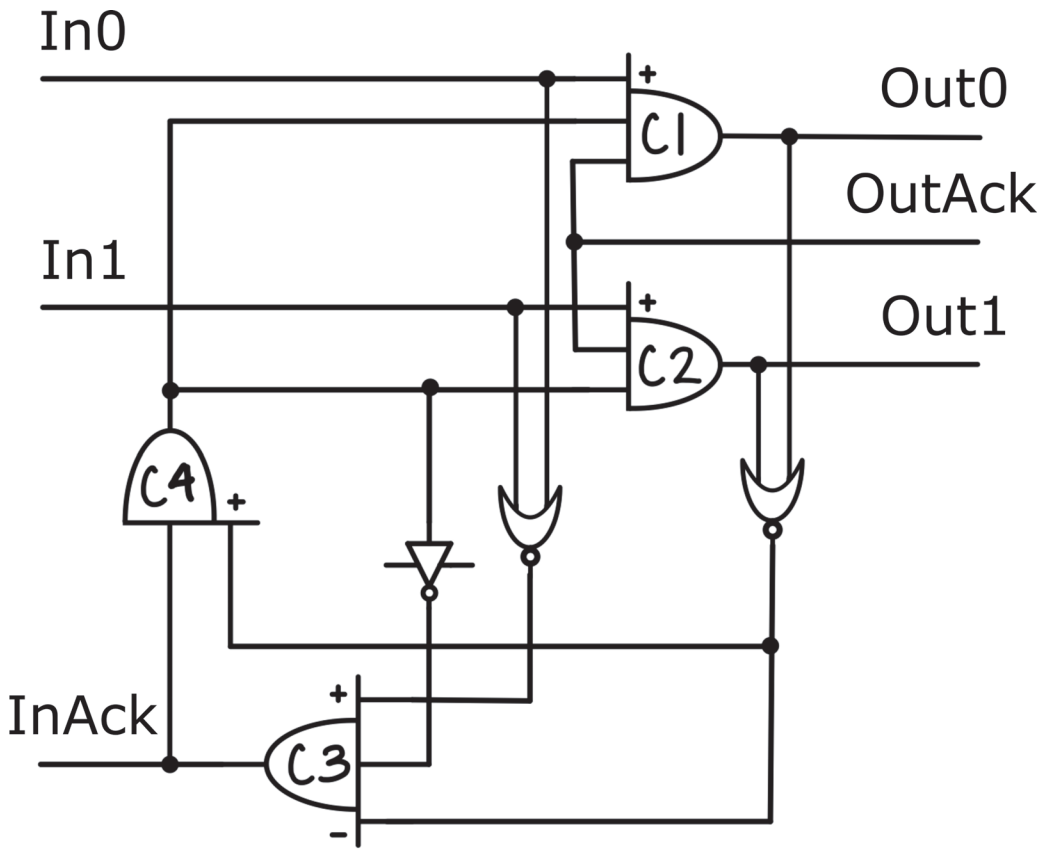



Figure 3-10: Precharge Full Buffer Schematic

the data pulls the InAck signal low in the right-hand environment.

One buffer is needed for each single-bit data input to each functional block. Modifications could be made to the design to allow for two inputs, such as what is needed for the AND, NAND, OR, etc. gates, but a better token buffer was explored before continuing with this design.

A QFP-level implementation for the PCFB is shown in Figure 3-11. The design is clearly very area and power expensive. Furthermore, all of the AQFP logic gates expect binary encoded data, meaning that a single data line passes high and low values, so a conversion element needs to be added between the dual rail encoded tokens from the PCFB and the AQFP logic gates. This element will be discussed in the final section. It's also important to note that the PCFB expects input on


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

activation phase 1, so each functional logic gate cell must output values on phase 3. This can be a simple thing to keep track of for small functional blocks, but for more complicated blocks, the buffer array synchronizer could be added before each input so that values on each activation cycle can arrive regardless of phase.

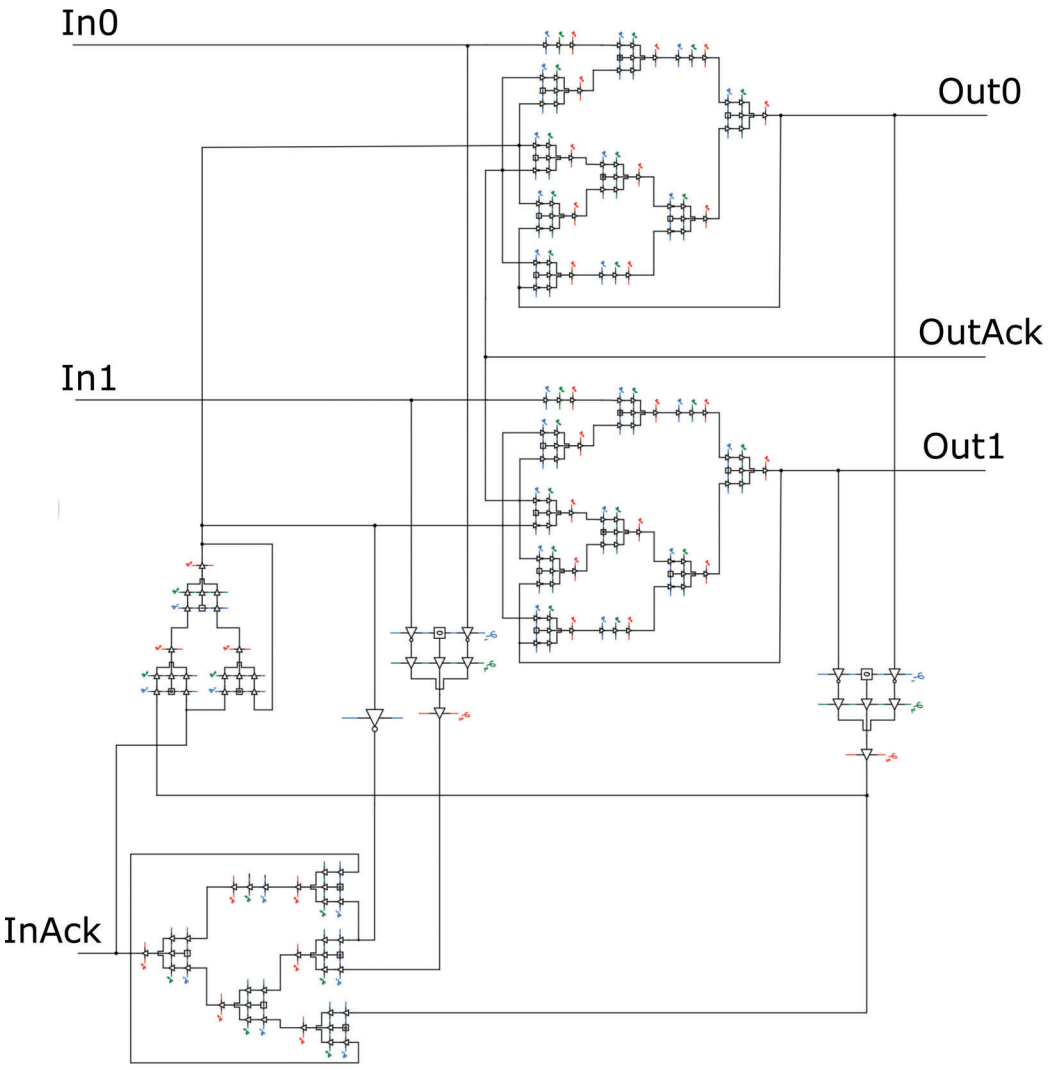



Figure 3-11: Precharge Full Buffer (PCFB) QFP-level implementation

The PCFB was explored first, because it is a direct transfer from Green's previous CMOS work to low-energy SCE implementation; however, it's clearly not the best design for the specialized superconducting technology. A better optimized token


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

buffer is explored next.

### 3.3.3 QFP Full Binary Buffer (QFBB)

As discussed, the PCFB is designed for CMOS technology and although the basic logic can be recreated in AQFP technology, it's not a very efficient way of passing tokens. Furthermore, the dual rail encoded token passing that is most commonly used in asynchronous pipelines, are not a good match for the majority-gate logic cells designed in AQFP. Therefore, I propose a more resource-efficient design which I call the QFP Full Binary Buffer (QFBB), Figure 3-12.

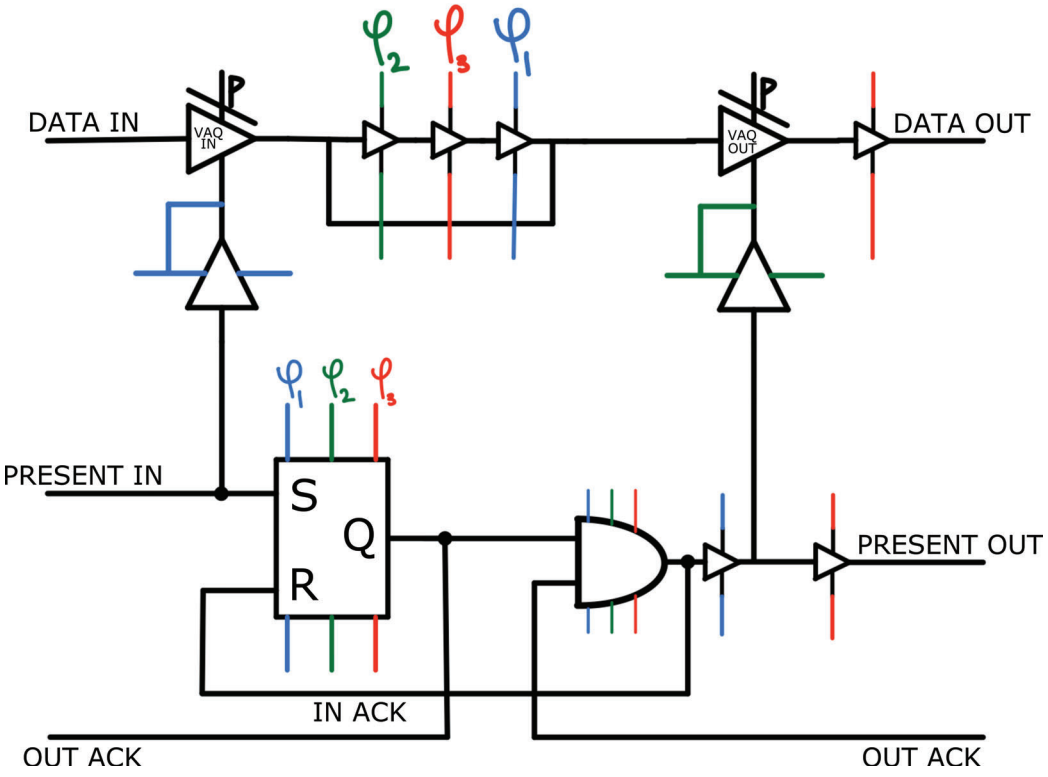



Figure 3-12: QFP Full Binary Buffer (QFBB) for single data input

Binary encoded asynchronous data means that one wire has the data value (0/low and 1/high) while the other wire indicates present/high or null/low, and the value line is only valid when the present line is high. Regardless of one vs two inputs, the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

data value arrives at the QFBB with a present signal and is stored in a buffer loop while the present signal is stored in a NOR gate flip-flop until it is either propagated by an output acknowledgment signal or reset by an input acknowledgment signal. The NOR flip-flop is described in Figure 3-13. The communication and coordination between the data value and present signal is where the QFP-specific trick comes into play. As discussed in Section 2.3, QFPs are two terminal devices with two symmetric input signals. I've described QFPs so far as only being activated by an external ac-signal, however the QFP could also be a logic device on it's own if we allow the output from one QFP to provide the activation signal to another. This is the heart of the QFBB: the present signal triggers a *driving QFP* which activates a *variable activation QFP* so that the data value can be stored into the buffer loop or propagated to the function block input. The inner workings of the variable activation QFP (VAQ) will be explored in the next section. As in the PCFB, it is assumed that the data input always arrives on phase 1, and that the functional element also expects input on phase 1; and again, this can be generalized by adding the phase synchronizer.

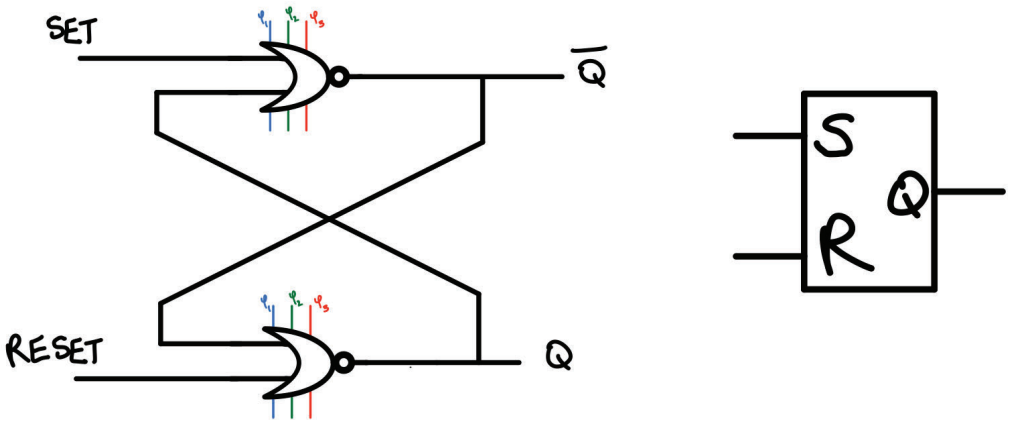



Figure 3-13: NOR Flip-Flop

The single input QFBB, Figure 3-12, design is useful for a strict implementation of ALA (meaning that bit-level token passing around gate-level cells are used) because we need flow control around the single input buffer and inverter cells. It operates as follows:


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

When no tokens are present, both acknowledgment signals are low. When data arrives with the present signal high, VAQ IN is triggered so that the data value is propagated to the ring oscillator, where it is stored across activation cycles. The high present signal is then also stored across activation cycles in the flip-flop. If OUT ACK is low, indicating that there is no token blocking the next cell block, then the present signal passes the AND gate. This triggers VAQ OUT so that the value in the ring oscillator can propagate to the function block. The present signal is also passed to the function block and it must be propagated with every activation phase that occurs inside the function in order to stay properly paired with the output bit to trigger the next token presence. A high output from the AND gate also resets the IN ACK signal with the flip-flop, so that an incoming token from the left-hand environment can arrive and the signal flows.

The two input design is very similar; although, it blends a coincidence buffer with the token flow so that functional blocks are only initiated when both inputs are present. This is done by adding an AND gate between the two present signals which must be high before triggering VAQ OUT. The two input design, shown in Figure 3-14, is more useful for wrapping logic gates or larger math modules since single-bit level token flow control may be overkill, as I will explore in the next chapter.

### 3.3.4 Variable Activation QFP (VAQ)

Although my idea for a QFP activated by another QFP came independently, the concept was originally proposed as a variable activation QFP (VAQ) by Hioe, Hosoya, and Goto in 1991 [53]. It is the building block for their D-gate design, which generates combinatorial logic from networks of QFPs without the majority-gate wire logic, and will be explored in future work (Section 6.1). The output of one QFP does not have a strong enough output current on it's own to drive another similar QFP. This can be solved by adding a "puller" to the VAQ QFP [46].




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

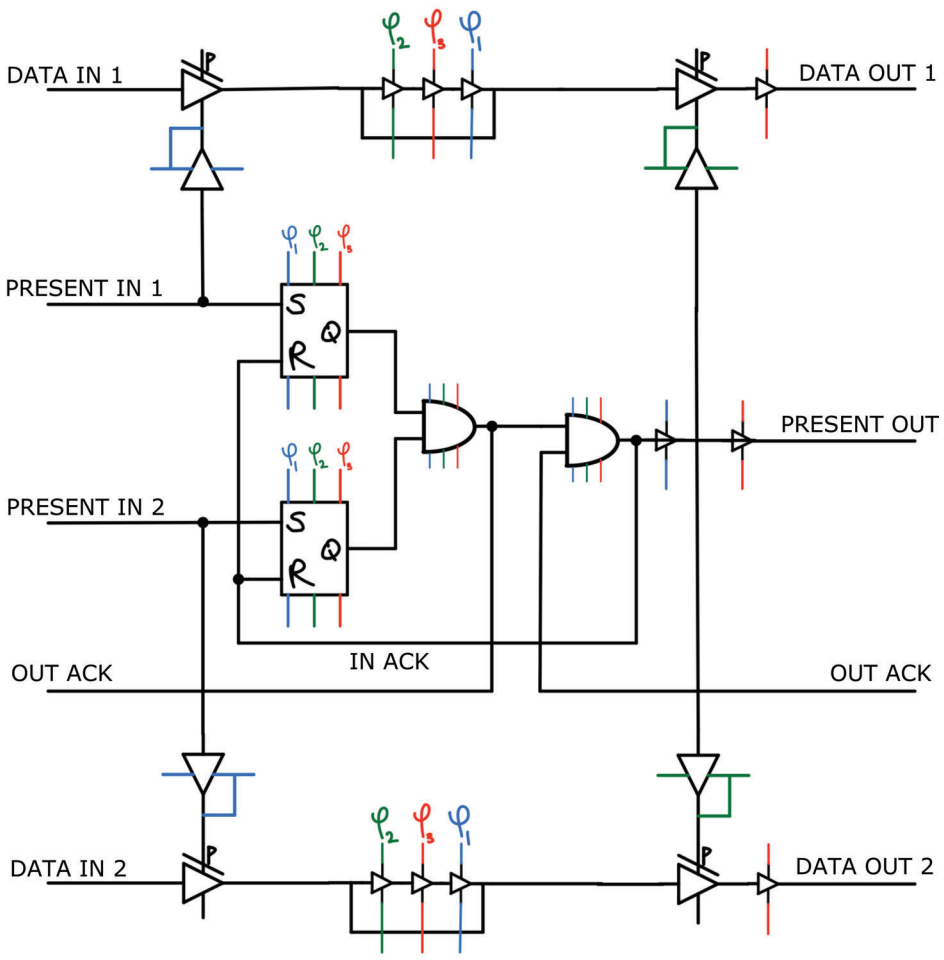



Figure 3-14: QFP Full Binary Buffer (QFBB) for two data input

A puller is a QFP attached in series to another QFP's (in this case, the variable activation QFP) activation line. A puller needs to be added to the I/O transformer QFP, shown in Figure 3-15, which is a different schematic/layout, but same theoretical operation, as the QFP buffer from Figure 2-7. The original QFP can be specified as an "activation transformer" QFP because the activation signal is applied through coupled inductors and the I/O signal is a current wire input. The "I/O transformer" is reversed, so that the I/O signal is applied to both junctions through the transformer and the activation signal is on a line; therefore, the puller can be added in series. The puller is biased by $\pi$ with its own input transformer. When it's placed between the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

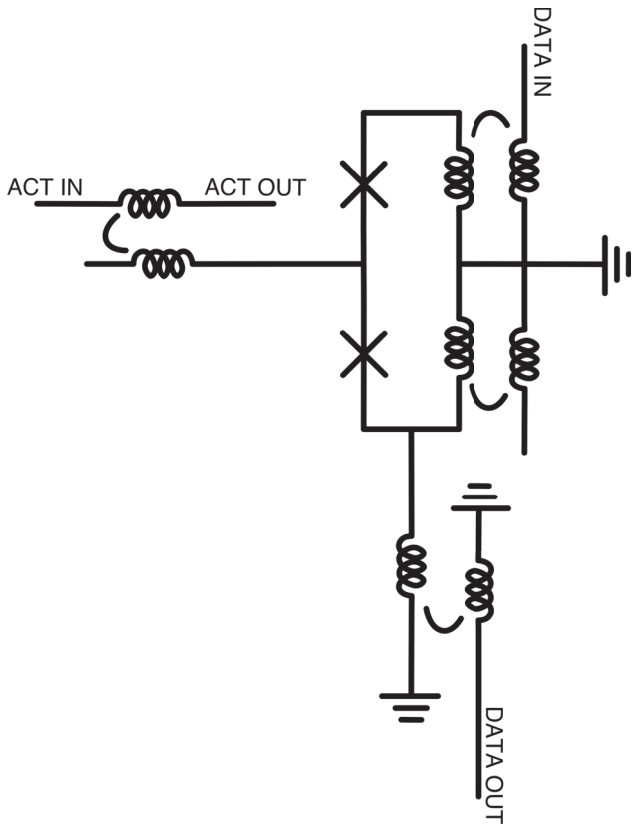



Figure 3-15: I/O Transformer QFP

QFP and an activation transformer, as shown in Figure 3-16, the puller will provide a current boost to the activation signal, $\alpha$, so that the QFP can still fire even though it receives a smaller activation. More details on the I/O transformer QFP and puller operation can be found in [46].

Turning our attention to the driving QFP, the typical output flux angle of a QFP is around 120 degrees (it depends on $\beta$), and the VAQ needs an activation signal of at least $\pi$ to fire [46]. Therefore, we can pair the driving QFP output with the usual ac activation signal so that the VAQ is fired whenever the driving QFP output is the same polarity as its activation signal. Luckily, in the design I've made, we only need a VAQ triggered by a high value, which matches the polarity of the ac activation signal already in use, but in theory you could also have a low triggered VAQ. A junction-level schematic diagram of the VAQ and the driving QFP along is shown in Figure


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

3-17.

As was hinted at in the PCFB section, the VAQ is also the element needed to convert between dual rail encoding and binary encoding of asynchronous data lines. This is because we need some way to have if/then logic between the present and value line. An example of a dual rail to binary encoding converter is given in Figure 3-18. Again, data value is stored in a buffer ring oscillator, and only propagated forward with a paired present signal when the VAQ fires.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

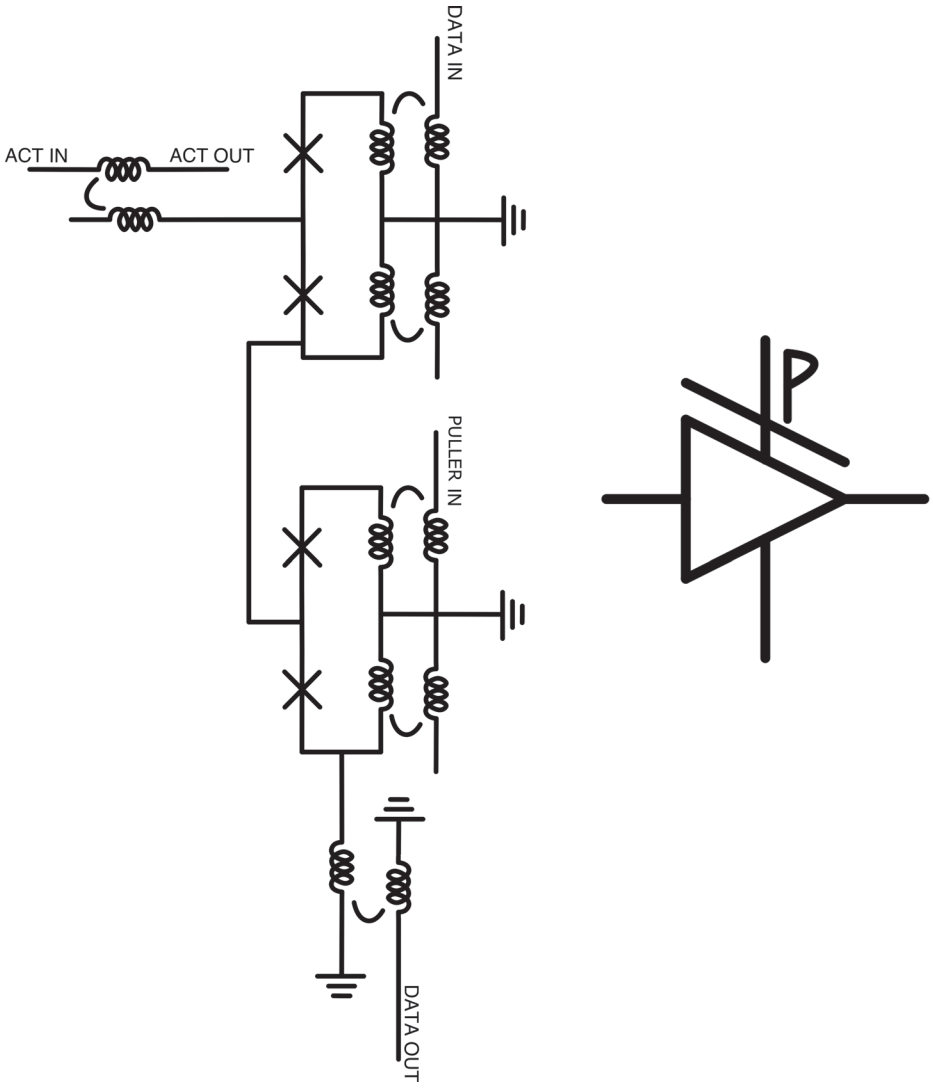



Figure 3-16: QFP with Puller for activation signal boost


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

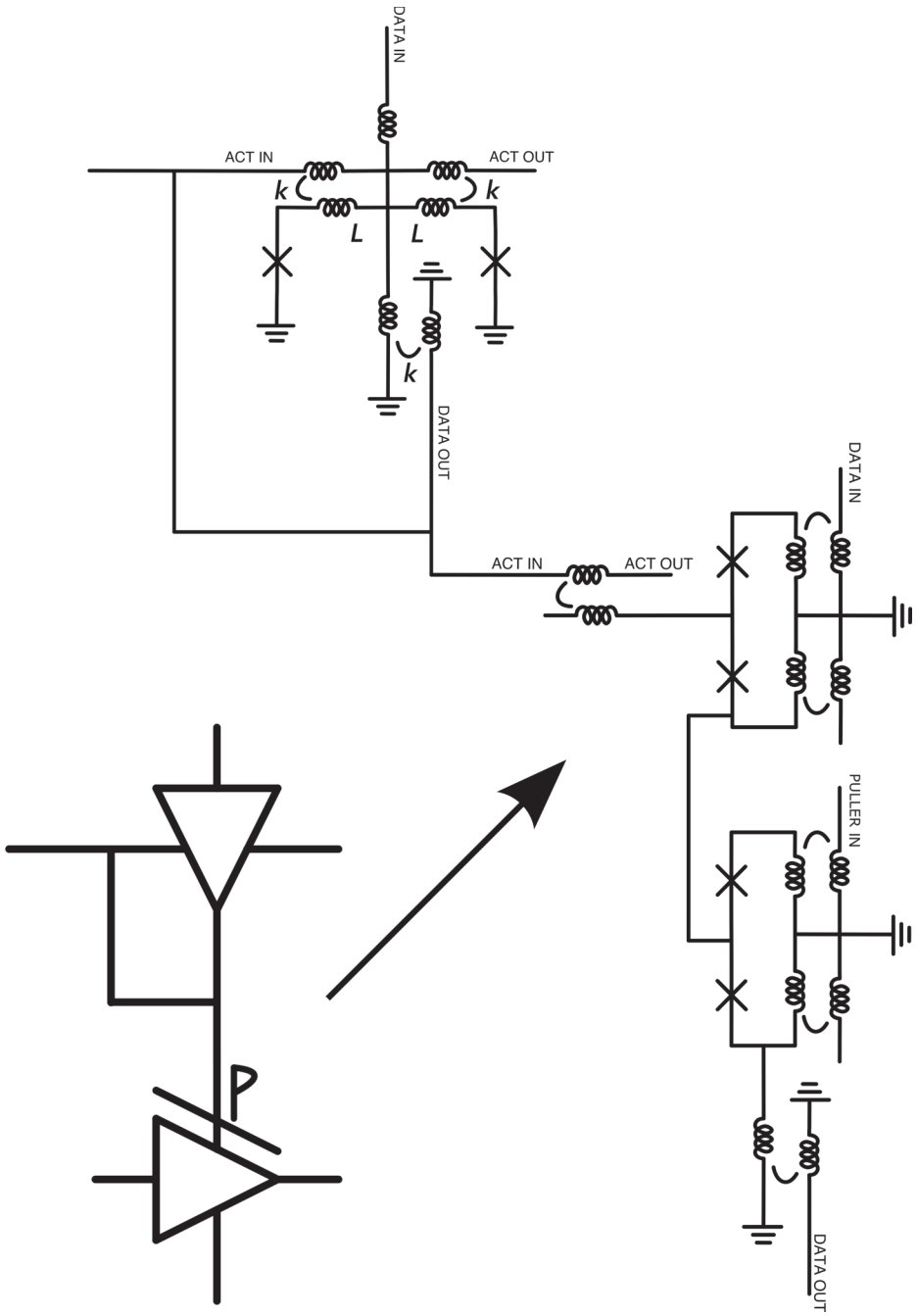



Figure 3-17: Junction-level schematic of Variable Activation QFP with Driving QFP





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

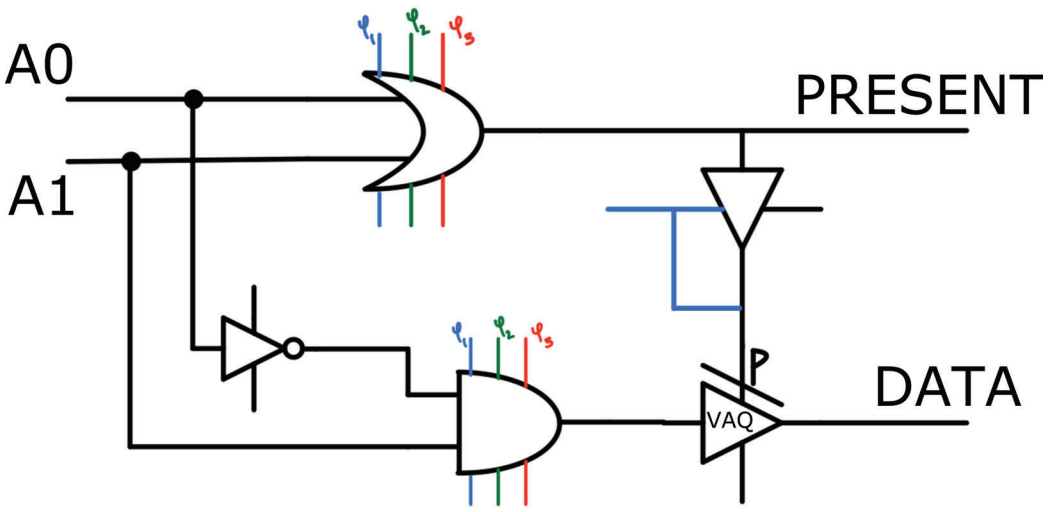



Figure 3-18: Dual-rail to binary token converter circuit


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 4

# Logic Modules

This chapter explores how to design more complex computing functionality using the building blocks introduced so far. These larger logic modules compose the next layer of modular design abstraction, since they could be taken as more complex starting nodes for asynchronous assembly.

## 4.1 Token Boundary

The token boundary is the distinguishing line between synchronous versus asynchronous regions of the architecture. Determining where to impose the token boundary is a crucial design study being done with Super-DICE. At one extreme is ALA, where the tokens pass bit level information at each logic gate. This has the most modular design, and therefore the most flexible and scalable architecture requiring as little technology-specific layout optimizations as possible. Although, it also has the highest token passing energy and area overhead. Especially for the AQFP gates, wrapping every ALA QFP buffer with a full token buffer is grossly inefficient - it adds, at minimum, an extra 66 Josephson junctions (the amount required for the single input QFBB design) for every QFP buffer cell, i.e. a 33 to 1 junction overhead.

At the other extreme is DICE, where data tokens are passed at chip boundaries because each chip is composed of multi-purpose synchronous microcontrollers. The DICE framework has a smaller power budget dedicated to asynchronous communi-




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

cation than the bit-level token passing, however the architecture is not as flexible or efficient for modular customization. Especially for the superconducting circuits, the large synchronous microcontroller node design requires intense technology-specific optimizations and custom EDA tools, as demonstrated with the MANA AQFP microcontroller [26], which still cannot compete with CMOS performance complexity.

Therefore, the Super-DICE token passing boundary lives in between these two extremes. The exact location depends on the chosen token buffer and requires more design work and results from fabricated chips. The token buffer choice plays an integral role in determining the token boundary. Larger buffers may be more robust and make less restrictive assumptions about timing requirements, but then they should be accompanied by larger synchronous nodes to decrease the density of expensive buffers; whereas, smaller and more efficient token buffers can be placed more often. Therefore, the improvements that the QFBB offers over the PCFB is significant. This trade-off will be explored quantitatively in the next chapter.

At this stage of the project, it makes sense to build super-DICE circuits with the token boundary at the logic module. These designs are explored next and compared to vanilla ALA implementations.

## 4.2 Adder

### 4.2.1 Vanilla ALA Adder

To explore an extreme example in detail, a straight-forward ALA serial adder is shown in Figure 4-1. I'll refer to this design as the ALA Serial Adder. It is a full adder with two inputs and a carry, slightly modified from Green's thesis work [6]. If the carry out is linked directly to the carry-in, then this serial bit adder can scale to any bit precision. Furthermore, if the output is connected to one input with a short ALA buffer delay and the other input with a longer ALA buffer delay, then the simple circuit becomes a fibonacci adder, as in Figure 4-1.

There is also a more efficient ALA adder design explreod be Greenwald [14], which


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

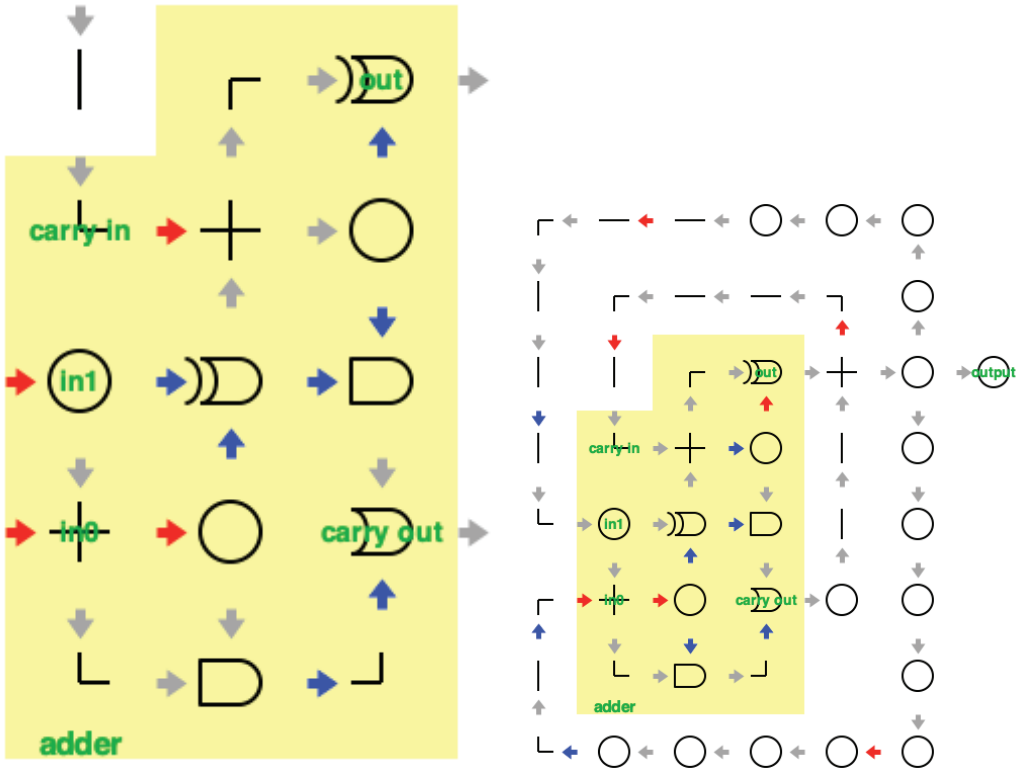



Figure 4-1: ALA Serial Adder schematic given on the left and an example of it used in the Fibonacci sequence generator to the right.

I'll refer to as the ALA Carry Adder, shown in Figure 4-2. More efficient adder designs may be possible in ALA, however it's clear that this granular of token passing is not needed, so attention was given to other QFP designs.

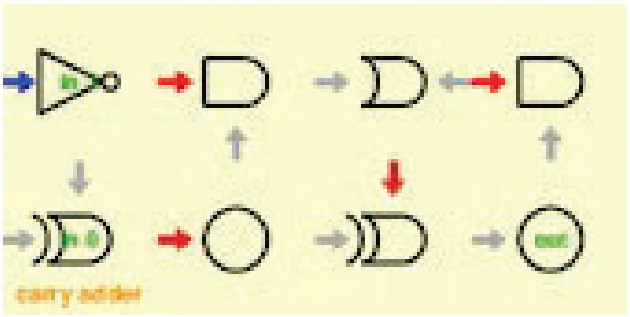


Figure 4-2: ALA Carry Adder


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### 4.2.2 Synchronous Adders

It's much more energy and area efficient to make synchronous AQFP adders with token buffers added to the input. However, this design removes the ability for a single schematic to be an adder of arbitrary position, as in the case of the ALA adders.

An example of a synchronous half adder is shown in Figure 4-3. The figure does not show the entire "module" which would also include the phase synchronizer and token buffer at the input, so it can be placed anywhere and expect any arbitrary asynchronous input. Similarly, a full adder is shown in Figure 4-4.

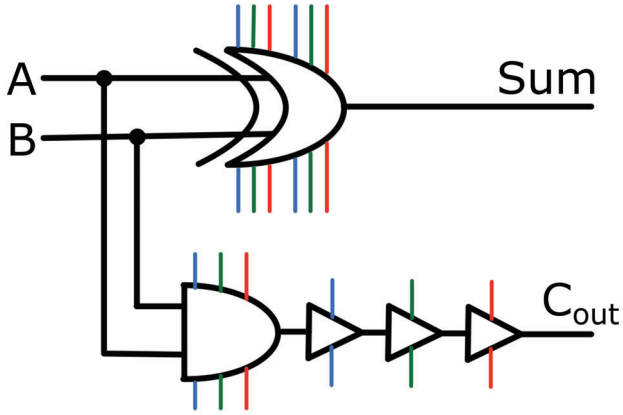



Figure 4-3: Synchronous Half Adder

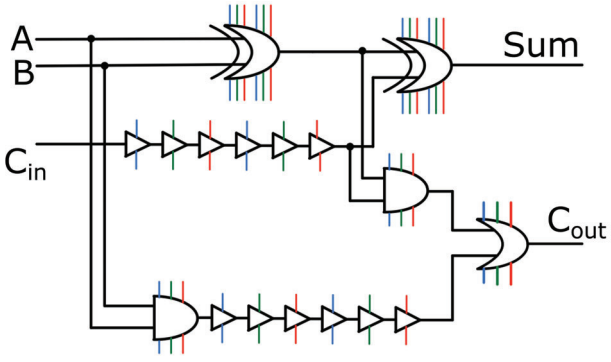



Figure 4-4: Synchronous Full Adder

Both of these logic modules will be compared quantitatively to the ALA serial adder in the next chapter. However, it's clear just from schematic designs, that the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

synchronous adder with logic module token passing is superior in power performance without losing too much granularity in architecture control. Therefore, super-DICE will build on a modified ALA framework.

However, trading off bit-level token passing makes the design more difficult to scale because the design is no longer parametric. For example, the ALA serial adder can be arbitrary bit length if the output is hooked up to the input properly and parallelism is not lost because ALA scales in both time and space. Meanwhile, the synchronous adder needs to take up more space and be re-designed for each word-length change. Therefore, when dealing with higher precision numbers, it may be more useful to use the ALA design, or a tighter blend of the ALA serial adder and the traditional ripple carry adder. Answers to these type of larger scale design trade-offs are currently being worked on.

Finally, XOR gates are expensive in QFP design, so it may be more energy and/or area efficient to implement the full adder in NAND logic. An example of this equivalent circuit design is give in Figure 4-5.

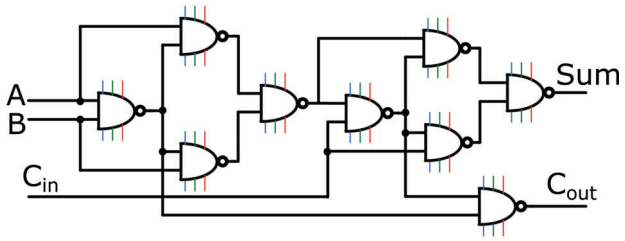



Figure 4-5: Synchronous Full Adder without XOR gates

## 4.3 Multiplier

Designs for a multiplier are currently being developed. Figure 4-6 shows an ALA multiplier design from Greenwald's previous thesis work building and benchmarking an ALA matrix multiplier [14]. However, in Super-DICE the multiplier will be optimized for AQFP performance so the design can be more efficient.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

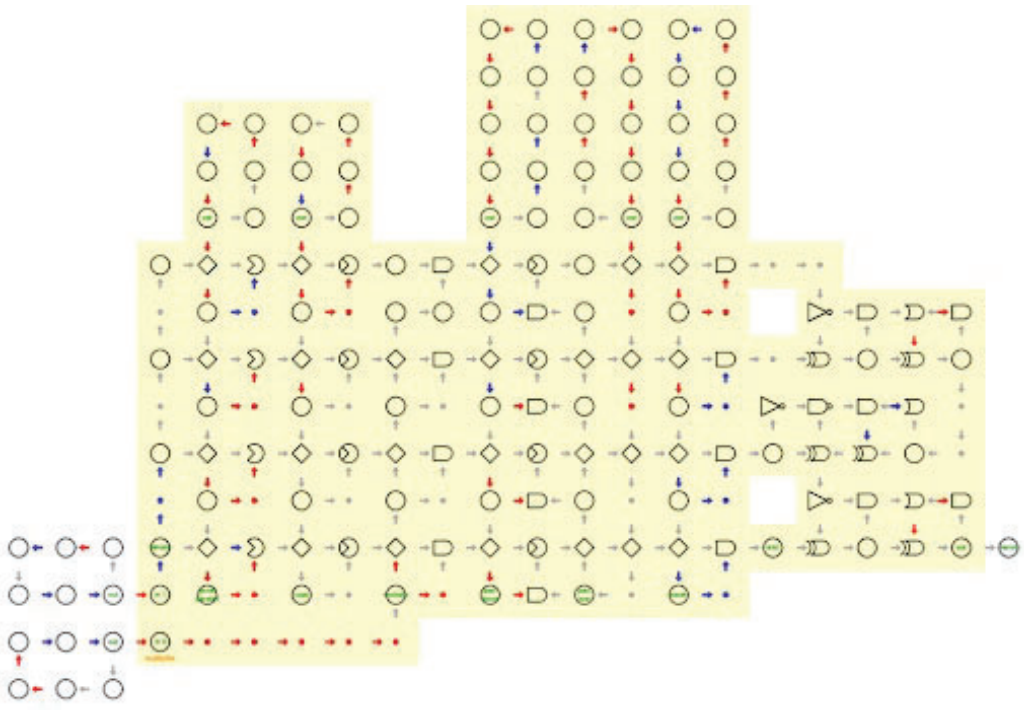


Figure 4-6: ALA Multiplier


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 5

# Evaluation

## 5.1 SPICE Simulation

Superconducting circuit schematics were simulated using WRSPICE [54], an open source integrated circuit simulation tool which includes Josephson junction models. The default WRSPICE Josephson junction model is based on the RSCJ equivalent circuit described in Section 2.1, however other models can also be loaded and customized which may be useful for future custom QFP development [54].

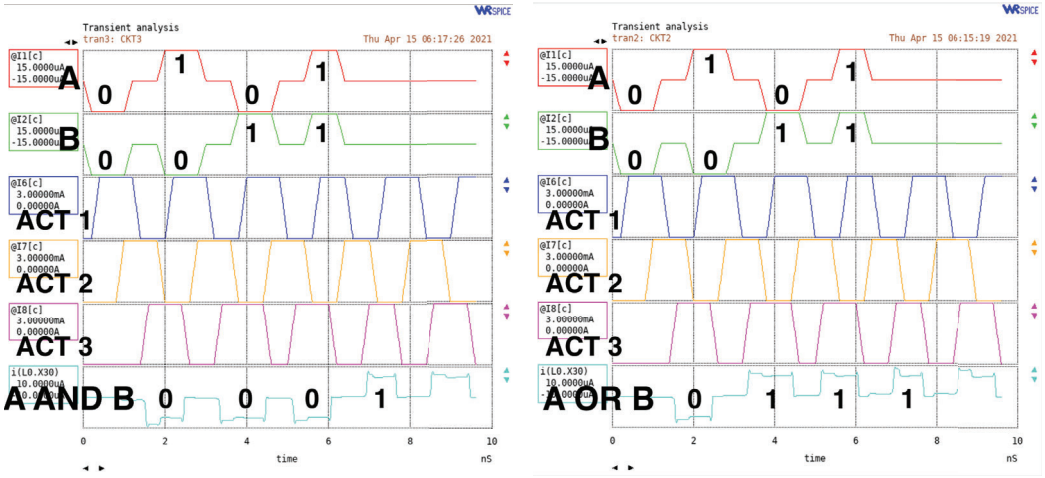
SPICE simulations were used for logic level verification of Super-DICE circuit designs. Looking back to the AQFP logic gates, Figure 5-1 shows the SPICE output for the AND, OR, and XOR gates. The top two plots show the input current values, A and B. The next three plots show each of the activation current phases (ACT1 for $\varphi_1$, ACT2 for $\varphi_2$, and ACT3 for $\varphi_3$). Then the final plot shows the output current from the logic gate.

Figure 5-2 shows the results of the phase synchronizer with a buffer ring oscillator attached to the output. The schematic is shown on the top and SPICE simulations for an input on each of the activation phases are shown below. A 0 signal is passed on the data line at different times, and regardless of activation phase the 0 is repeated on the data output line after the first activation cycle. It's clear that no data are dropped, regardless of input phase.

Finally, the synchronous adder designs provide the expected output and are shown

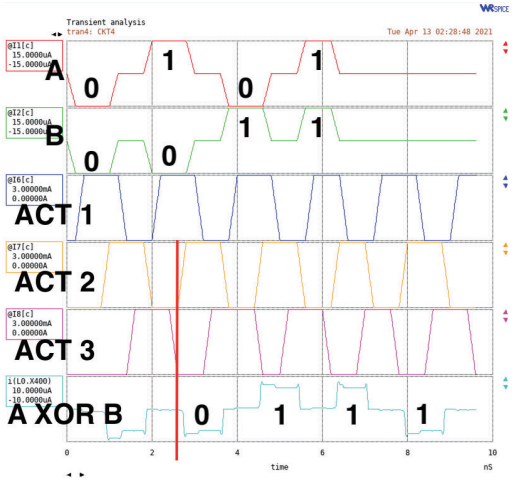

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

in Figure 5-3.




(a) AND SPICE

(b) OR SPICE




(c) XOR SPICE

Figure 5-1: SPICE plots for AQFP logic gates


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

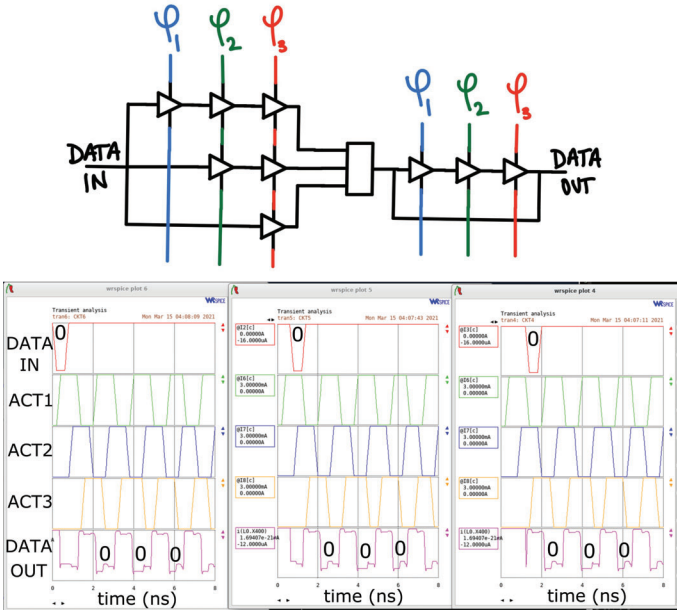
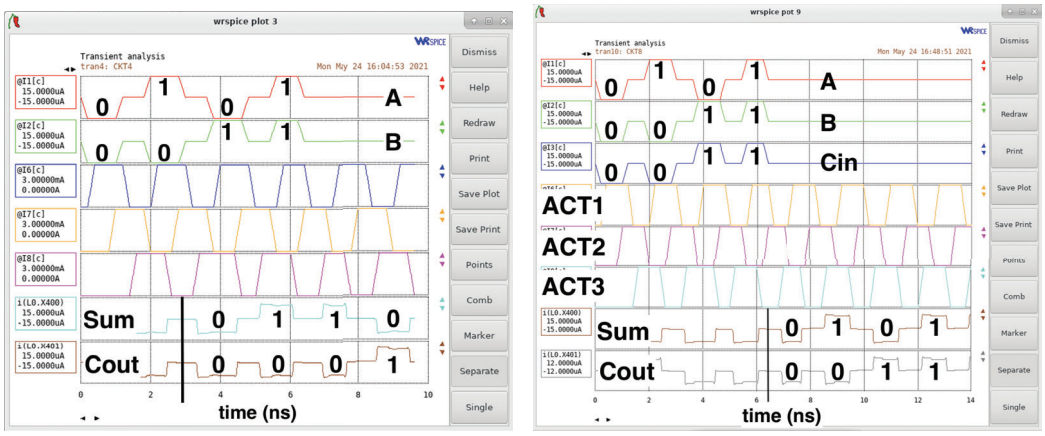



Figure 5-2: Phase synchronizer. Input passed on activation phase 1, 2, and 3 and data out makes it to the buffer loop each time.




(a) Half Adder SPICE

(b) Full Adder SPICE

Figure 5-3: SPICE plots for synchronous adders



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

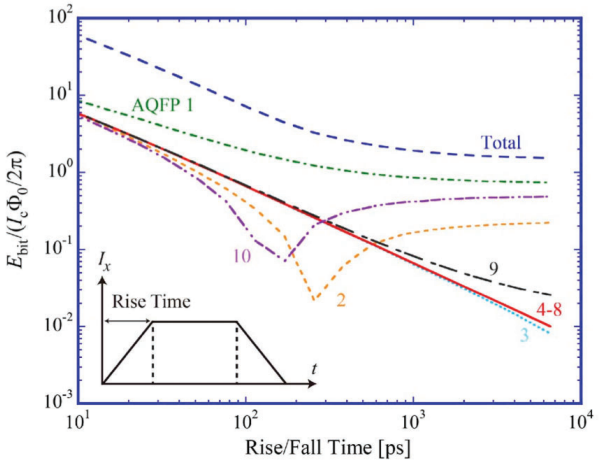



Figure 5-4: QFP bit energies for 10 buffers in series [7]

## 5.2 Energy Dissipation

With the logic operation verified for QFP circuit design, we now revisit energy performance to determine the feasibility of the initial $10^5$ order of magnitude power improvement claim. The dynamic switching energy of a single QFP is on the order of $100k_bT$. Figure 5-4 shows simulation data from Takeuchi et al. displaying the bit energy dissipation across a single buffer QFP when 10 QFPs are fired in series for a range of rise/fall activation times [7]. The results are used to optimize QFP design parameters for the lowest bit energy, while maintaining wide enough operation margins for robust functionality at a finite temperature. The resulting optimized bit energy is $6.40 \times 10^{-21}$ J [7]. The junction design parameters found in this paper are similar to what can be implemented at MIT LL with the SFQ5ee fabrication technology. Therefore, it's reasonable to assume that QFPs we design and fabricate can achieve similar energy dissipation values. Furthermore, given the adiabatic nature of AQFP, the energy dissipation depends on the switching speed, so a faster rising/falling time will result in higher energy dissipation while a slower rising/falling time will be lower energy. This can be thought of mechanically like the ball in the double well potential, the faster the energy landscape changes the more kinetic energy, and therefore friction energy dissipation, the ball will have.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Table 5.1 gives a comparison of area (through JJ count), timing, and power performance of each circuit design explored earlier. AQFP gates have energy interactions, so there will be slight differences in QFP switching energies depending on the circuit schematic, however these differences are ignored for now. The energy projections are done assuming that each QFP has a dynamic switching energy of $100k_bT$, implying that activation cycles are run at about 5 GHz. Note that for all of these values, I've assumed that all inductors cooled at 4K are lossless and values ignore dielectric energy loss in the activation ac-biasing lines. It's important to emphasize that the dielectric losses are a nontrivial factor to ignore because it's very likely that they dominate power dissipation for the chip, but this is being explored and optimized currently so will be accounted for in future work. Also note that all of these values are the energy dissipation at 4K, cryogenic cooling power overhead is not taken into account, but multiplying each value by $10^3$ would estimate room temperature energy dissipation.

The energy dissipation value projects the total energy dissipated for one complete operation of the circuit. For example the XOR gate consists of 20 QFPs which need to each fire twice for a full XOR computation (2 activation cycles), therefore an XOR operation costs 4000 $k_bT$.

It's clear from Table 5.1, that the QFBB is preferred to the PCFB, because of its 84% shrink in number of Josephson junctions and 98% decrease in energy dissipation. Furthermore, it's clear that a vanilla ALA token boundary for super-DICE is not preferable to the logic module level boundary, as discussed in Chapter 4.

Future work will continue to for accurately estimate these energy projections through simulation and experimental testing with fabricated chips.

## 5.3 Scalability

The Super-DICE circuits presented so far are a long way away from VLSI for superconducting supercomputers; however, a more detailed projection for large system Super-DICE performance can now be explored. The analysis for system power budget follows a similar narrative to feasibility projections in [55], but making ad-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Table 5.1: Circuit Performance Projections

| Circuit | JJ count | Activation cycles | Energy dissipation(aJ) |
|---|---|---|---|
| BUF/INV/CONST | 2 | 1/3 | 0.00580 |
| AND/NAND/OR/NOR | 14 | 1 | 0.0406 |
| XOR | 20 | 2 | 0.232 |
| Phase Synchronizer | 12 | 1 | 0.0348 |
| C-element | 76 | 3 | 0.661 |
| Asymmetric(+) C-element | 130 | 4 | 1.51 |
| PCFB | 406 | 14 | 16.5 |
| QFBB 1 input | 66 | 2 | 0.383 |
| QFBB 2 input | 128 | 4 | 1.48 |
| Dual to Binary | 36 | 2 | 0.209 |
| Sync Half Adder | 60 | 2 | 0.348 |
| Sync Full Adder (XOR) | 146 | 4 | 0.696 |
| 1-bit Half Adder Module | 200 | 7 | 4.06 |
| 1-bit Full Adder Module | 286 | 9 | 7.46 |

justments for AQFP technology and ALA architecture. Power dissipation in high performance computers today is roughly split between logic, memory, and interconnect energy loss. A superconducting supercomputer needs to also be cryogenically cooled, so also needs to budget for power dissipation through heat leakage (e.g. conduction leaks through structural linkages, convection through gases and liquids in the cooling system, and thermal radiation). The ALA architecture also does not impose a distinction between logic and memory power dissipation, and therefore also decreases the interconnect power dissipation because data is not constantly trafficked between memory and logic. Although, there will be more chip-to-chip interconnects for the Super-DICE 3D structure. Therefore, let's assume that our Super-DICE system power budget needs 10% for heat leaks, 70% for logic/memory, and 20% for interconnect. The Linde LR280 helium reliquefier refrigeration system can cool 1020 W at 4.4K with 395 W/W efficiency and requires 2 MW total power for operation [55].

Assuming we can fabrication chips with 2 million Josephson junctions, which is just at the limit of MIT LL fabrication abilities, then each chip could have 1


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

million QFPs. 1 million QFPs would have a switching energy of about 6 fJ ($10^6$ QFPs $\cdot 5.8 \times 10^{-21}$ J/QFP) operating at 5 GHz, so each chip would have 0.03 mW power dissipation. Allowing 70% of the 1020 W refrigeration capacity to go towards logic/memory chip energy dissipation, then 23.8 million chips could be cooled by this system. That means that 23.8 million Super-DICE nodes could fit in the refrigerator, only accounting for thermal budgeting and ignoring chip area and refrigerator volume. With 1 million QFPs per Super-DICE node, this system would have 23.8 trillion QFPs. Comparatively, the Summit supercomputer, the world's fastest supercomputer before this most recent Top500 benchmark in June of 2021 [8], has 27,648 NVidia V100s. Each Nvidia V100 has 21.1 billion transistors [56], so Summit has about 583 trillion transistors for quantitative GPU computation.

An energy efficient CMOS fused multiply-adder (FMA) requires about 30,000 gates [57], and assuming one gate takes about 7 QFPs, an FMA would need 21,000 QFPs. For a rough performance estimate, let's assume that all 23.8 trillion QFPs would go towards parallel FMAs, then the Super-DICE supercomputer would have $1.7 \times 10^{18}$ op/s (equal to $(23.8 \times 10^{12}/21000)\text{op} * 5 \times 10^9\text{Hz} * 3$ for 5 GHz activation cycling with 3 phases) for about 1 kW power dissipation, if the refrigeration capacity is maximally used. Therefore, this Super-DICE supercomputer would operate at $5.88 \times 10^{-16}$ W/ops at cryogenic temperatures, and $5.88 \times 10^{-13}$ W/ops accounting for the cooling overhead. Although this is lower than the initial $10^5$ magnitude improvement from the motivating case study (Section 1.1), the $10^3$ order of magnitude improvement remains significant over the state of the art CMOS energy efficiency at $10^{-10}$ W/ops. Therefore, future design work is warranted, especially with the goal of designing and optimizing Super-DICE FMAs.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 6

# Conclusion

## 6.1 Future Plans

A major next step in proving out the Super-DICE concept and performance is fabrication of the proposed circuit designs. At the moment, a test chip is in fabrication at the MIT LL superconducting foundry. It's layout shown in Figure 6-1. The chip is a pizza mask consisting of a few different data buffering methods - there are QFP cells passing data on and off chip, SFQ cells passing data on and off chip, and a SFQ-DC converter for reading data off the chip and into room temperature electronics. From this chip, we can experimentally verify power performance for QFP buffering and compare it to SFQ logic. The distinction between on and off chip data buffering will also provide useful information for interconnect overhead from the multichip super-DICE structure. Future test chips will include the token buffering mechanisms, as described in Section 3.3, and logic circuit designs for experimental verification of the power results projected in Table 5.1.

There are also many more design decisions to be explored for the Super-DICE architecture. For example, the logic gates could be redesigned using QFP D-gates [53], instead of the majority-gate logic proposed in Section 3.2. Majority gate logic is a linear input logic which makes it very susceptible to unwanted interactions between QFPs and line inductances, therefore decreasing its input margin. QFP D-gates are made up of two parts: the activation group and the logic group. The logic group


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

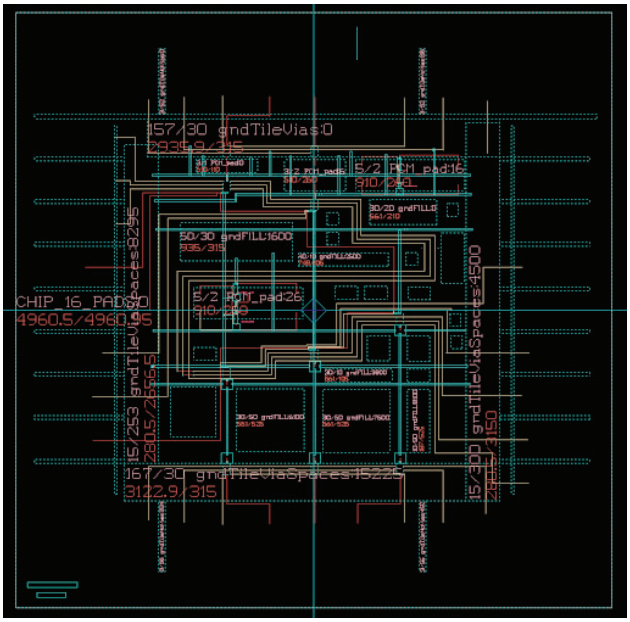



Figure 6-1: Pizza mask test chip currently in fabrication at MIT LL

consists of two VAQs which generate the output of the D-gate. The logic group is then driven by two QFPs which make up the activation group and behave as controls for the logic output. D-gates can realize any combinational logic function and have a much higher input margin than the gates from Section 3.2 [53]. Although they take up more area and require more Josephson junctions, the D-gate may be preferred as chip designs get more complicated and inductance crosstalk becomes more of an issue; therefore, they will be kept in mind for future development.

As mentioned in Chapter 4, the adder and multiplier design could be more optimized for power and area performance, along with parameterized scalability. The obvious next steps in development is to design a robust and efficient FMA (fused multiply-add), which can then be compared to state-of-the-art CMOS processors for finer-grained performance projections and benchmarking.

Additionally, in ALA designs, data tokens are destroyed and generated at each token buffer, this is an obvious waste of energy if tokens could be propagated instead. Token propagation would be similar to data propagation in QFP buffers, i.e. fluxons are not created or destroyed, but rather nudged into new locations by input currents to


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

encode information. Propagating tokens would behave like a billiard-ball reversible computer. Asynchronous Ballistic Reversible Fluxon Logic (ABRC) [58, 59] and Josephson transmission lines [60] are being explored for this task.

The fact that when nothing is being computed no work is being done and no power is consumed, is one of the major benefits of ALA. However, this is not the case with Super-DICE ALA because in the current design, all the QFPs are fired on their respective activation phase, regardless of whether or not there is data present at their input. This is why there is no activity factor added to chip performance projections, like in typical CMOS dynamic power projections, because each QFP is active every cycle. This could be mitigated by introducing some type of activation signal gating mechanism to the token buffer, similar to clock gating methods in CMOS VLSI design.

The Super-DICE project will also continue to develop an end-to-end workflow for system design, development, assembly, and application. Chip package and interconnect design using digital materials for 3D, reconfigurable construction is currently being engineered. Importantly, design tools for DICE will be extrapolated to ALA instrachip design as well, introducing an all-in-one programming and layout design tool, similar to that shown in Figure 1-4c.

## 6.2 Impact

This thesis provides a first step to developing a scalable end-to-end workflow for ultra-low power superconducting computing systems. I presented a set of computing building blocks implemented with Adiabatic Quantum Flux Parametron logic and proposed a token buffering mechanism for asynchronous communication between these modular parts. I discussed how more complex circuits can be built and projected energy performance for large scale systems. In sum, I hope this work can provide a toolkit for future superconducting device design.

With the decline of Moore's law and increasing computing demands, cutting edge software development is increasingly bleeding into custom hardware design. We see this as companies which traditionally operated in software applications have increased


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

investment in custom chip design, such as Google's TPU, Tesla's self-driving car chip, or Apple's custom M1 processor. The Super-DICE framework serves as a unique tool for this rising application not only due to it's improved power performance, but also because of it's reconfigurable and modular design methods. The Super-DICE toolkit is ideal for rapid prototyping of ASIC development due to the simple and scalable relationship between schematic and layout design. This greatly decreases time and development costs, and therefore decreases the barrier to entry to hardware design. Additionally, the spatial computing system blurs the line between hardware and software, while bypassing inefficient instruction paths imposed by traditional von Veumann processor architecture.

In conclusion, much more work needs to be done to experimentally verify the power performance projections of a Super-DICE supercomputer; however, the work shared so far indicates promising enough returns to continue with more involved design studies and provides a compelling framework to growing applications beyond just power performance.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Bibliography

[1] J. Ward, "Additive assembly of digital materials," Master's thesis, Massachusetts Institute of Technology, 2010.

[2] W. K. Langford, "Electronic digital materials," Master's thesis, Massachusetts Institute of Technology, 2014.

[3] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Quantum supremacy using a programmable superconducting processor," vol. 574, no. 7779, pp. 505–510.

[4] Y. Harada, W. Hioe, and E. Goto, "Flux transfer devices," vol. 77, no. 8, pp. 1280–1286. Conference Name: Proceedings of the IEEE.

[5] M. Hosoya, W. Hioe, J. Casas, R. Kamikawai, Y. Harada, Y. Wada, H. Nakane, R. Suda, and E. Goto, "Quantum flux parametron: a single quantum flux device for josephson supercomputer," vol. 1, no. 2, pp. 77–89. Conference Name: IEEE Transactions on Applied Superconductivity.

[6] F. Green, "Ala asic: A standard cell library for asynchronous logic automata," Master's thesis, Massachusetts Institute of Technology, 2010.

[7] N. Takeuchi, K. Ehara, K. Inoue, Y. Yamanashi, and N. Yoshikawa, "Margin and energy dissipation of adiabatic quantum-flux-parametron logic at finite temperature," vol. 23, no. 3, pp. 1700304–1700304. Conference Name: IEEE Transactions on Applied Superconductivity.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[8] E. Strohmaier, J. Dongarra, H. Simon, M. Meuer, and H. Meuer, "Top 500 the list." Top 500, Jun. 2021. Accessed on: Aug 24, 2021).

[9] E. Strohmaier, J. Dongarra, H. Simon, M. Meuer, and H. Meuer, "Green 500 the list." Green 500, Jun. 2021. Accessed on: Aug 24, 2021).

[10] N. Takeuchi, T. Yamae, C. L. Ayala, H. Suzuki, and N. Yoshikawa, "An adiabatic superconductor 8-bit adder with 24kbt energy dissipation per junction," *Applied Physics Letters*, vol. 114, no. 4, p. 042602, 2019.

[11] Z. Fredin, J. Zemanek, C. Blackburn, E. Strand, A. Abdel-Rahman, and P. Rowles, "Discrete integrated circuit electronics (dice)," in *2020 IEEE High Performance Extreme Computing Conference (HPEC)*, pp. 1–8, 2020.

[12] E. Strand, "Inverse methods for design and simulation with particle systems," Master's thesis, Massachusetts Institute of Technology, 2020.

[13] N. Gershenfeld, "Aligning the representation and reality of computation with asynchronous logic automata," *Computing*, vol. 93, pp. 91–102, 2011.

[14] S. Greenwald, "Matrix multiplication with asynchronous logic automata," Master's thesis, Massachusetts Institute of Technology, 2010.

[15] N. Fatès, "A guided tour of asynchronous cellular automata," in *Cellular Automata and Discrete Complex Systems* (J. Kari, M. Kutrib, and A. Malcher, eds.), (Berlin, Heidelberg), pp. 15–30, Springer Berlin Heidelberg, 2013.

[16] C. A. Petri and W. Reisig, "Petri net," *Scholarpedia*, vol. 3, no. 4, p. 6477, 2008. revision #91647.

[17] W. D. Hillis, "The connection machine." Accepted: 2005-08-08T21:52:32Z.

[18] N. Margolus, "CAM-8: A computer architecture based on cellular automata,"

[19] G. A. Taubes, "The rise and fall of thinking machines | inc.com." Inc. Magazine, September 1995.

[20] C. L. H. T. Kung, "Systolic arrays for (vlsi)," Tech. Rep. ADA066060, Carnegie-Mellon University, December 1978.

[21] D. A. Buck, "The cryotron-a superconductive computer component," *Proceedings of the IRE*, vol. 44, no. 4, pp. 482–493, 1956.

[22] D. C. Brock, "Dudley buck's forgotten cryotron computer." Section: History of Technology.

[23] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, "Superconducting qubits: Current state of play," *Annual Review of Condensed Matter Physics*, vol. 11, p. 369–395, Mar 2020.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[24] H.-L. Huang, D. Wu, D. Fan, and X. Zhu, "Superconducting quantum computing: A review," 2020.

[25] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, "Characterizing quantum supremacy in near-term devices," *Nature Physics*, vol. 14, p. 595–600, Apr 2018.

[26] C. L. Ayala, T. Tanaka, R. Saito, M. Nozoe, N. Takeuchi, and N. Yoshikawa, "MANA: A monolithic adiabatic iNtegration architecture microprocessor using 1.4-zJ/op unshunted superconductor josephson junction devices," vol. 56, no. 4, pp. 1152–1165.

[27] Q. Xu, C. L. Ayala, N. Takeuchi, Y. Murai, Y. Yamanashi, and N. Yoshikawa, "Synthesis flow for cell-based adiabatic quantum-flux-parametron structural circuit generation with HDL back-end verification," vol. 27, no. 4, pp. 1–5. Conference Name: IEEE Transactions on Applied Superconductivity.

[28] T. Tanaka, C. L. Ayala, Q. Xu, R. Saito, and N. Yoshikawa, "Fabrication of adiabatic quantum-flux-parametron integrated circuits using an automatic placement tool based on genetic algorithms," vol. 29, no. 5, pp. 1–6. Conference Name: IEEE Transactions on Applied Superconductivity.

[29] D. van Delft and P. Kes, "The discovery of superconductivity," vol. 63, no. 9, pp. 38–43. Publisher: American Institute of Physics.

[30] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, "Theory of superconductivity," vol. 108, no. 5, pp. 1175–1204. Publisher: American Physical Society.

[31] T. V. Duzer and C. W. Turner, *Principles of Superconductive Devices and Circuits, (Second Ed.)*. USA: Prentice Hall PTR, 1998.

[32] E. Snider, N. Dasenbrock-Gammon, R. McBride, M. Debessai, H. Vindana, K. Vencatasamy, K. V. Lawler, A. Salamat, and R. P. Dias, "Room-temperature superconductivity in a carbonaceous sulfur hydride," vol. 586, no. 7829, pp. 373–377.

[33] C. C. M. Mody, "Between research and development: IBM and josephson computing," vol. 69, no. 10, pp. 32–38.

[34] J. Matisoo, "The tunneling cryotron—a superconductive logic element based on electron tunneling," vol. 55, no. 2, pp. 172–180. Conference Name: Proceedings of the IEEE.

[35] K. Loe and E. Goto, "Analysis of flux input and output josephson pair device," vol. 21, no. 2, pp. 884–887. Conference Name: IEEE Transactions on Magnetics.

[36] K. Likharev and V. Semenov, "RSFQ logic/memory family: a new josephson-junction technology for sub-terahertz-clock-frequency digital systems," vol. 1, no. 1, pp. 3–28. Conference Name: IEEE Transactions on Applied Superconductivity.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[37] S. K. Tolpygo, "Superconductor digital electronics: Scalability and energy efficiency issues (review article)," *Low Temperature Physics*, vol. 42, no. 5, pp. 361–379, 2016.

[38] Y. Yamanashi, T. Nishigai, and N. Yoshikawa, "Study of LR-loading technique for low-power single flux quantum circuits," vol. 17, no. 2, pp. 150–153. Conference Name: IEEE Transactions on Applied Superconductivity.

[39] D. E. Kirichenko, S. Sarwana, and A. F. Kirichenko, "Zero static power dissipation biasing of RSFQ circuits," vol. 21, no. 3, pp. 776–779. Conference Name: IEEE Transactions on Applied Superconductivity.

[40] M. H. Volkmann, A. Sahu, C. J. Fourie, and O. A. Mukhanov, "Implementation of energy efficient single flux quantum digital circuits with sub-aJ/bit operation," vol. 26, no. 1, p. 015002.

[41] Q. P. Herr, A. Y. Herr, O. T. Oberg, and A. G. Ioannidis, "Ultra-low-power superconductor logic," vol. 109, no. 10, p. 103903.

[42] R. Landauer, "Irreversibility and heat generation in the computing process," vol. 5, no. 3, pp. 183–191. Conference Name: IBM Journal of Research and Development.

[43] O. A. Mukhanov, "Energy-efficient single flux quantum technology," vol. 21, no. 3, pp. 760–769. Conference Name: IEEE Transactions on Applied Superconductivity.

[44] A. Stillmaker and B. Baas, "Scaling equations for the accurate prediction of cmos device performance from 180nm to 7nm," *Integration*, vol. 58, pp. 74–81, 2017.

[45] E. Goto, "The parametron, a digital computing element which utilizes parametric oscillation," vol. 47, no. 8, pp. 1304–1316. Conference Name: Proceedings of the IRE.

[46] E. G. Willy Hioe, *Quantum Flux Parametron: A Single Quantum Flux Superconducting Logic Device*. P.O. Box 128, Farrer Road, Singapore 9128: World Scientific Publishing Co. Pte. Ltd., 1991.

[47] N. Takeuchi, Y. Yamanashi, and N. Yoshikawa, "Adiabatic quantum-flux-parametron cell library adopting minimalist design," *Journal of Applied Physics*, vol. 117, no. 17, p. 173912, 2015.

[48] N. Gershenfeld, "Programming a new reality [video file]," 2006.

[49] M. Heiligman, "Supertools," June 2016.

[50] E. Yahya Tawfik and M. Renaudin, *QDI Latches Characteristics and Asynchronous Linear-Pipeline Performance Analysis*. Pages: 592.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[51] P. A. Beerel, R. O. Ozdag, and M. Ferretti, *A designer's guide to asynchronous VLSI*. Cambridge University Press. OCLC: ocn459209613.

[52] D. E. Muller, University of Illinois at Urbana-Champaign. Graduate College. Digital Computer Laboratory, and University of Illinois at Urbana-Champaign. Department of Computer Science, *Theory of asynchronous circuits*. Urbana, Illinois : University of Illinois, Graduate College, Digital Computer Laboratory.

[53] W. Hioe, M. Hosoya, and E. Goto, "A new quantum flux parametron logic gate with large input margin," vol. 27, no. 2, pp. 2765–2768. Conference Name: IEEE Transactions on Magnetics.

[54] S. R. Whiteley, *WRspice Reference Manual*. Whiteley Research Inc., Sunnyvale, CA 94086, 4.3.13 ed., January 2021.

[55] D. S. Holmes, A. L. Ripple, and M. A. Manheimer, "Energy-efficient superconducting computing—power budgets and requirements," *IEEE Transactions on Applied Superconductivity*, vol. 23, no. 3, pp. 1701610–1701610, 2013.

[56] "Nvidia tesla v100 gpu architecture," tech. rep., NVIDIA, August 2017.

[57] E. C. Quinnell, "Floating-point fused multiply-add architectures," Master's thesis, University of Texas at Austin, 2007.

[58] M. P. Frank, "Asynchronous ballistic reversible computing," in *2017 IEEE International Conference on Rebooting Computing (ICRC)*, pp. 1–8, 2017.

[59] M. P. Frank, R. M. Lewis, N. A. Missert, M. A. Wolak, and M. D. Henry, "Asynchronous ballistic reversible fluxon logic," *IEEE Transactions on Applied Superconductivity*, vol. 29, no. 5, pp. 1–7, 2019.

[60] D. V. Averin, K. Rabenstein, and V. K. Semenov, "Rapid ballistic readout for flux qubits," *Phys. Rev. B*, vol. 73, p. 094504, Mar 2006.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Distributed Displays for Discrete Integrated Circuit Electronics

by

Justin Browning Christensen

B.S., Brigham Young University (2019)

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, in partial fulfillment of the requirements for the degree of

Master of Science in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2021

© Massachusetts Institute of Technology, 2021. All rights reserved.

Author .....
Program in Media Arts and Sciences
August 20, 2021

Certified by .....
Neil Gershenfeld
Director, The Center for Bits and Atoms
Thesis Supervisor

Accepted by .....
Tod Machover
Academic Head, Program in Media Arts and Sciences



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Distributed Displays for Discrete Integrated Circuit Electronics

by

Justin Browning Christensen

Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, on August 20, 2021 in partial fulfillment of the requirements for the degree of

Master of Science in Media Arts and Sciences

## Abstract

I present a distributed display architecture that integrates with a set of asynchronous mechanically and electrically re-configurable computing nodes, otherwise known as Discrete Integrated Circuit Electronics (DICE). Each singular display is physically and electrically connected to a DICE node which transmits useful data to be displayed. By integrating these displays with the DICE nodes, a multitude of applications are enabled, starting with real-time data visualization and debugging, and going on up to more complex applications, such as locally computed ray tracing and graphics rendering, as well as structural and volumetric displays. The advantages and implementation of the displays into the DICE architecture, as well as various examples of their applications are demonstrated and discussed. While the DICE nodes themselves address issues with locality in computing, these integrated distributed displays will help them overcome some of their limitations and enhance their capabilities. Together, these integrated devices and their scalability can lead to iterative improvements to graphical processing, form into spatial 2D grid (structural) and 3D mesh (volumetric) displays, and overall reduce the cost and complexity of distributed display and computing systems.

Thesis Supervisor: Neil Gershenfeld
Title: Director, The Center for Bits and Atoms



236
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Distributed Displays for Discrete Integrated Circuit Electronics**

by

Justin Browning Christensen

This thesis has been reviewed and approved by the following committee members:

Neil Gershenfeld .....
Director, The Center for Bits and Atoms
Professor, Media Arts and Sciences
Massachusetts Institute of Technology

Joseph Paradiso .....
Professor, Media Arts and Sciences
Massachusetts Institute of Technology

Cardinal Warde .....
Professor, Electrical Engineering
Massachusetts Institute of Technology



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Acknowledgments

I would not be where I am today without the tremendous love and support provided to me by so many people. I hope to express even a fraction of the gratitude I feel to those that made all of this possible.

First off, thank you to my advisor, Neil, for all that you've done for me. From opening up your group to me when I needed to find a new home in the lab, to all of the advice and counsel you've given to me. Beyond the technical portions that you helped me with, the habits and life skills you taught me are sure to be of great value in my life. Thank you for providing such a wonderful place to learn and grow in the Center for Bits and Atoms. I've had access to tools and skills I never would have before, and my mind has been expanded greatly because of it.

Thank you to my readers, Joe Paradiso and Cardinal Warde. Your suggestions and guidance over the past months have been invaluable in shaping this work and keeping me on track. Joe, thank you for being such an inspiring figure within the Media Lab with all the work you do. What stands out to me the most, however, is your willingness to meet with any student, no matter how busy your schedule is. Cardinal, thank you for all of the chats we've had about holography and display technology, and for being such a great instructor. I'm very glad I took your class when I first arrived at MIT.

Thank you to the rest of CBA for inspiring me to be better. Kara, Sherry, Joe, Candace, and James, you do a tremendous job keeping the lab and its subsequent functions running. Tom and John, thank you so much for being incredible teachers and friends in the shop. I loved spending my time in there with you learning to make new things and fix the machines that make them. Pranam, Ben, Jake, Amira, Filippos, Patricia, Jiri, Chris, Alfonso, Camron, and David, thank you for being so helpful and welcoming me into the group. I've really enjoyed seeing the work you do and being a part of it.

Thank you in particular to Zach and Erik. You both have been absolute life-savers to me. Many sleepless nights potentially banging my head against the wall (perhaps both figuratively and literally) were avoided due to your generosity and helpfulness with this project. Zach, you taught me so many tricks and tips to hone in my hardware development skills. Erik, you helped me understand the code and programming as a whole so much better. You two deserve a lot of credit for empowering me to clear the obstacles I was faced with. You were so responsive and helpful at any time of day, and even while on vacation. For that, I express my utmost appreciation. Thank you for being my friends and being so kind and generous to me with your time.

Thank you to my Object-Based Media family. As my original home in the lab, you helped me to reach for the stars and develop that sense of wonder that is so prevalent at the Media Lab. Dan, Kristin, and Catherine, thank you for running the ship and making sure all of us were taken care of. Thank you for all our chats that helped expand my mind and fall in love with the magic of the lab. Pip, Everett, Laura, Bianca, Pedro, and Ali, thank you for being my big brothers and sisters at the lab, helping me get settled in, and being a great place of support for me as I learned



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

the ropes. Nina, Tyler, and Aubrey, we went through a lot together, and I couldn't ask for a better Masters Squad to be a part of. Thank you so much for everything, you will forever be my dear friends.

Thank you to all the other friends I made at MIT and to the Cambridge 1st Ward. Jonny and Annie, it was so fun for our kids to play together, and I am so grateful for the deep friendships that were formed between our families. I hope it will continue for a long time. Also, thanks for letting us borrow your car so often... Zach and Hailey, thank you for the rides to church, the fun times hanging out, and for always being willing to lend a helping hand. Ryan and Sarah, thanks for being such great neighbors and helping us get settled into Westgate, we're so glad to have met and gotten to know you.

Thank you, Dr. Daniel Smalley, for everything you've done for me. You pioneered the path for this kid from Utah to make it all the way to the MIT Media Lab. From when I first started working with your research group, I had countless memories that I'll cherish forever. I loved making my first hologram, my first holo-chip, being part of publishing a paper for the first time, and so much more. Most of all, though, I'll always remember that day in your office when we first met, which was the moment that I knew that I wanted to become an electrical engineer. You've changed my life for the absolute best, and I'm eternally grateful for it.

Thank you to the absolute best friends in the entire world. Cameron, Amy, Colby, Elias, and Inaya, you have been the greatest support group, cheerleading team, and shoulders to stand on. I still marvel to this day how lucky I am to have you in my life. Whether back in Utah or right there with me in Massachusetts, you all helped motivate me to be my best, and I wouldn't be anywhere close to the person I am today without you. I'm looking forward the rest of our lives and the friendship that will surely make it all so great.

Finally, thank you so much to my parents, grandparents, siblings, in-laws, and the rest of my family. Mom and Dad, you showed me how to work hard and do the best job possible from the very beginning. I strive each day to be the best person I can be and to make you proud, and I hope I've done so. Eric and Terri, thank you for being so supportive, even as I took your daughter and moved her all the way across the country with me. You've been such a blessing in my life, and I can't thank you enough for all you've done. Lia and Millie, Daddy loves you so much, and I'm so grateful for all the laughs, smiles, and play time that we've shared together. You brighten my day always and give me so many reasons to keep pushing and working. Kayla, you are the love of my life and I'd be completely lost without you. We've navigated so many decisions and trials in life together already, and there's no one else I'd rather experience it all with. Thank you for loving me and inspiring me to be better, thank you for putting up with and helping me overcome my faults, and thank you for helping me achieve this great accomplishment. I don't know where the rest of life will take us, but it doesn't matter so much as long as I get to hold your hand to walk beside you down that path as we let God prevail in our lives.

Thank you, to each and every one of you, and to any others that have been a part of my journey. I wouldn't be here without you, and I hope to continue on to make you all proud and to make this world a better place.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Contents

**1 Introduction** **11**

1.1 Background . . . . . 11

1.2 Locality . . . . . 12

1.3 Need for Visualization . . . . . 15

1.4 Distributed Displays . . . . . 16

1.5 Contributions . . . . . 17

**2 Hardware Integration** **19**

2.1 DICE Design . . . . . 19

2.2 Display Integration . . . . . 24

2.3 Sending Data . . . . . 25

2.4 Packaging . . . . . 28

**3 DICE Debugging and Data Visualization** **29**

3.1 Debugging . . . . . 29

3.2 Data Visualization . . . . . 30

3.3 Code Example . . . . . 34

**4 Application: Particle Simulation** **37**

4.1 Motivation . . . . . 37

4.2 Implementation . . . . . 38

4.3 Results . . . . . 39

4.4 Improvements and Variations . . . . . 43



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**5 Application: Ray Tracing** **47**

5.1 Motivation . . . . . 47

5.2 Proposed Implementation . . . . . 49

**6 Evaluation** **53**

6.1 Debugging Enhancement . . . . . 53

6.2 Data Visualization . . . . . 54

6.3 New Applications . . . . . 55

6.4 Performance Projections . . . . . 56

**7 Conclusion and Future Work** **59**



241
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 1

# Introduction

The Discrete Integrated Circuit Electronics (DICE) project is an effort by the MIT Center for Bits and Atoms to create an alternative computing architecture that is comprised of a series of interconnected "nodes" that are discretely assembled to perform a wide variety of parallel computational tasks [12]. Rather than the traditional definition of the point where two wires connect together, the word "node" here means a singular, self-contained device capable of performing computations, but with the capability of communicating its results and other data to other nodes. The work of this thesis is to take this new architecture and enhance it by integrating a set of distributed displays, enabling real-time debugging and data visualization, along with the plethora of applications opened up by these additions. The merits of these integrated displays and the issues that they overcome will be discussed, but first, the motivation and history behind the DICE project will be explained.

## 1.1 Background

Modern computational architectures have various limitations, such as minimum feature size, tunneling effects, and the ongoing pursuit of mastery over quantum effects. Advances in the industry have been able to avoid some of these limitations by focusing on refining other aspects and features. For example, with the focus on shrinking integrated circuit components according to Moore's Law, the complexity of comput-





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

ing devices has increased while the size or cost have either decreased or stayed the same. However, as technology approaches the limits of what is currently possible, it's not exactly clear which direction new advances can even take.

One of these limitations, which has been largely avoided by focusing on shrinking component size, comes in the form of how fast data can travel. Light can travel approximately one foot in a single nanosecond. Considering that up to gigahertz clock rates are often used in modern processors, that doesn't give a signal much time to access the appropriate memory cache and complete its assigned tasks. As the size of a circuit increases, the likelihood of being able to access a certain cell of memory anywhere on the device within a given clock cycle deteriorates. This can especially be seen in large super-computing centers. Because of the physical distance between different servers, quite complex routing and networking is required to ensure that each unit has access to the appropriate memory lines.

The above example, along with the rest of current common architectures, highlights a flaw that has largely gone under the radar: dependence on being able to access the same global memory from anywhere within a system. Because of developments according to Moore's Law, the distance between components within a circuit have decreased fairly proportionally to the increase in the complexity of the systems. Therefore, that physical limit of two components being too far apart to communicate within the clock cycle has been skirted. Thinking more in depth about this, though, causes one to question the necessity of this dependence on a global memory, which is present even among most modern architectures. While we haven't had to worry about running into this limitation, achieving even smaller feature sizes is becoming increasingly difficult and perhaps there isn't much more room to grow (or rather, shrink) down this current path. [25]

## 1.2 Locality

Over the years, various scholars have called into question this dependence on a global memory line and have sought to come up with architectures that bypass the need for



243
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

such universal access. Imagine if instead of one large system with different specialized areas that all needed to be in communication with one another, there were multiple instances of the same computing structure that were self-contained but could communicate with its closest neighbor. This method brings up the idea of "locality" in computing. If a single unit is able to handle computations at a local level, then share its results only with units that need them, not only does it remove the need for a global memory shared between every single unit, but it also means that the system can be scaled to meet the needs of the application.

One very notable example of this is the CAM-8 computer architecture [20] introduced in the early 1990s. This system's architecture is based on cellular automata, which essentially means that the system is made up of multiple smaller structures that tessellate together both physically and computationally to form one larger, more complex, system. Such a model was particularly useful for a handful of specific applications, most notably of gas particle simulations. Each physical "cell" is responsible for computations within the same area that it represents spatially. Just as gas particles interact with each other locally, each cell simulates the interactions locally, then shares information only with neighboring cells to which the effects of the particles "cross over" to effect them. As each cell runs its local computations, they all come together to form one larger environment that is capable of many feats that aren't as possible or are less efficient when using traditional computing architectures.

An iteration that evolved from this concept was that of a "paintable" computer, [6] which is essentially the idea of filling a sort of paint-like substance with hundreds of tiny and independent microprocessors. These "computing particles" are randomly distributed throughout the paint and are capable of asynchronous computations and local communication. This mixture can then be "painted" onto a substrate, forming the computer's structure according to the topology of the surface. The computing particles can distribute computational tasks to process in parallel, then report their results to their nearest neighbors, summing into a finalized calculation or completed overall computational task.

In a similar sense, there have been a variety of implementations of systems where


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

multiple computational or sensing devices are connected together to make computational substrates. One example of this is the Chainmail project, where a series of circuit boards with embedded processors and arrays of sensors are all connected together using conductive, but flexible materials to create a flexible skin. [21] A more recent example of this idea that evolved into full on textiles can be found in the Electronic Textile *Gaia* project, which draws upon the rich history of computational substrates to create a system that implements smart textiles at large breadth of scales. [26] Both of these examples implement structures that work independently in parallel, then share the data throughout for computational or sensing tasks.

While these are all very specific implementations of this concept, there is a large history of other work in this field, sometimes referred to as processor-in-memory. These systems work to maximize the efficiency of memory bandwidth by combining the processor and memory instead of keeping them separate like in traditional computational architectures. There have been various successful implementations of these systems, such as Computational RAM, which competes in cost and performance with traditional DRAM, [9] as well as the Terasys prototype, which was capable of supercomputer performance at a fraction of the cost. [13] One more successful implementation is Gilgamesh, which uses a "massively parallel architecture" to achieve up to petaflops of computing performance. [23]

Finally, like all these examples, the Discrete Integrated Circuit Electronics (DICE) project [12], which this work is based upon, completely challenges the idea of having one global clock cycle orchestrating all of the computational tasks. This comes in the form of a set of asynchronous, mechanically and electrically re-configurable computing nodes that are only capable of sending data to and receiving data from their immediate neighbors. By embracing locality in the computations at its foundation, this architecture handles each task by breaking it up into smaller and localized chunks, then bringing it all together for a final result. This system can easily be scaled simply by adding more nodes. In their current form, however, their are some limitations to their effectiveness in developing intriguing applications that need to be addressed. This work focuses on overcoming one of those major obstacles: the need


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

for visualization within the applications.

## 1.3 Need for Visualization

While much work has been put into validating and optimizing these nodes and how they communicate, there is still a big drawback within the current implementation of DICE: the lack of visualization of the performance/computations of each node. It is possible to view the data post-computation by porting it out through added signal lines in the system, but it isn't possible to see it in real time. This makes troubleshooting the system difficult, as well as makes it near impossible to visualize any meaningful data until after the simulation is complete, which misses out on perhaps the most interesting and useful portions of certain applications.

The work presented in this thesis is integrating into the system a series of distributed displays at an individual node level, enabling the real time visualization of node data. Using commercial off the shelf displays and connecting them to the already established system, the data is sent to and shown on the displays, enabling real time visualization and debugging of the nodes themselves. On top of these simple functions, these distributed displays provide even more potential applications, a couple of which are explored in depth in this thesis, such as particle simulations and locally computed ray tracing in graphics rendering. The performance of this system versus other commercially available graphics-focused processors is then discussed, showing both the benefits and drawbacks of this system.

The DICE architecture already helps in solving the issue presented by modern computing architectures in being overdependent on having access to a global memory, but by adding these displays to the system, a wide variety of applications are enabled to be explored in the future. With all of the many local computations and visualizations happening in parallel, the nodes can effectively share the load and fundamentally change the way we approach computing. This thesis shows the first steps in that direction, providing concrete examples, as well as theoretical applications to be implemented, explored, and evaluated in the future.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 1.4 Distributed Displays

As the main tangible addition to the project that this work brings about, it is necessary to mention current advances in distributed display technology and applications. Distributed displays, in their most basic state, are "computer systems that present output to more than one physical display." [17] Most often, they are seen as large, tiled displays in public spaces or even in TV studios, usually breaking down one source into a separate stream for each display, which combine together to make one large displayed version of the original source. This is typically done by using some sort of software to take the signal, cut it up visually and map it out to the various displays, then send all that data to some hardware that then multiplexes the signals out to each of the appropriate displays. Some use cases that have been further explored is using such displays to be used in offices and conference rooms or similar types of environments. [16] These use cases, as well as others, have typically been approached by researchers implementing their own distributed graphics systems using various computers and other devices networked together to render their displays, even making their process and algorithms scalable for larger displays. [16, 15]

Many of these systems still run into issues, though, especially when dealing with extremely large data sets to process through to render the displays. Often times, they must utilize high-performance computing (HPC) centers to handle the complex data, which takes time and doesn't allow for easily implemented real-time systems. Some have used real-time graphics accelerators, such as those found in supercomputers and gaming consoles, to try to overcome this problem. [10, 19] While a viable solution, the dependence on various external hardware can get messy, and interfacing with it all can be an even bigger problem.

While many modern implementations of distributed displays go big, the implementation with the DICE system will be at a much smaller form factor to match the size of the DICE nodes they will be integrating with. The reason for this is so that the software applications implemented within the hardware can align, especially in applications where the physical node represents the spatial computational volume it





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

is representing. Many of the same traits found between this system and others are still present, however, such as ability to be tiled and scaled, and general approach to distributing the computation and rendering among the different displays. By utilizing the DICE architecture for computation, the efficiency of the parallel processing should simplify and help overcome the limitations faced by typical distributed display systems.

## 1.5 Contributions

The contributions that this work has added are as follows: 1) a simple hardware/firmware integration process between the DICE nodes and the displays, 2) a basic framework for visualizing data and debugging information from the nodes to the displays, 3) demonstrable examples of various visualized applications, and 4) a proposed method of computing ray tracing algorithms locally within each node and displaying the subsequent results, as well as performance projections for how this method compares to commercially available components at its current state and in future scaled up versions of this project. While there is still much to be explored on the topic, hopefully these contributions will help continue down the path to discovering more efficient display and computational systems.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 2

# Hardware Integration

## 2.1 DICE Design

Before diving into how the displays are integrated into the system, a description of the existing DICE hardware is necessary. DICE has evolved as a project, having various iterations depending on the specific application/size constraints. Static-DICE is a set of nodes that are connected in four directions, all on a single PCB. Tiny-DICE is a set of modular nodes made using the smallest commercially available connectors that tessellate in a tetrahedral configuration. Finally, Meso-DICE, which this work builds on, is also a set of modular nodes and strut pieces to connect them, but made at a larger scale, designed to be easily programmed and assembled. The three iterations of the DICE project can be seen in Figure 2-1.

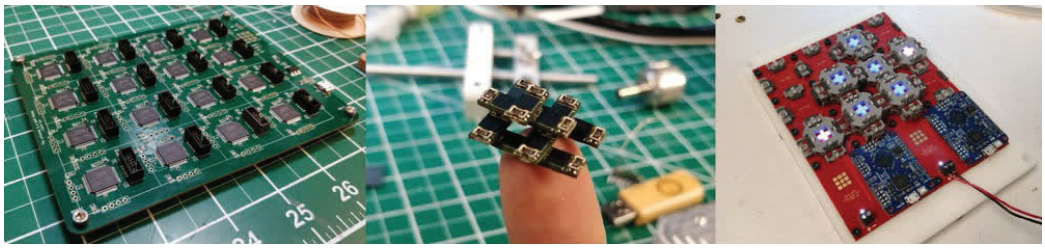


Figure 2-1: Current iterations of the DICE project. Left: Static-DICE; Middle: Tiny-DICE; Right: Meso-DICE. Images courtesy of Zach Fredin.

Each of these iterations of the DICE project are created with the purpose of im-



249
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

plementing a distributed computing architecture, with the function of breaking down any computing task into a parallelized one on the discrete hardware. They excel in certain applications that benefit from parallel computations, many of which are discussed in this work, but are meant as an all-purpose prototype to eventually become an alternative for modern computing structures. One huge benefit of this discretized architecture is the ability to align the software and hardware together. As the computations are parallelized, the nodes can better spatially represent what is being conducted within the software, which is especially true in specific applications. The processors embedded within each node depend on the scale and desired performance of each of the iterations, as do the interconnect design. Because this work focuses on the Meso-DICE iteration, a deeper explanation of Static-DICE and Tiny-DICE are not included here.

Meso-DICE consists of three main components: node, strut, and build plate (see Figure 2-2). The heart of each node is a SAMD51 microcontroller from Microchip, which is wired to a custom-made PCB that routes power, ground, and the various pins required for communication with the nearest neighboring node in each direction: north, west, south, east, top, and bottom. The bottom of the nodes have contact pads on all four sides and the top of the nodes have spring-pin connectors to press up on the aforementioned pads. These PCBs then have milled delrin and 3D printed PLA parts slotted into them, with the PLA pieces heat staked to secure everything in place. These pieces allow proper spacing and latching capabilities in order to connect with other components. The struts also have custom-made PCBs that simply have spring-pin connectors on top, contact pads on bottom, and several wires to route signals between the two sides of the strut to connect two individual nodes together. More delrin and PLA parts are assembled and heat-staked to the PCB to complete the strut. Finally, the build plate is required to have a solid base for the nodes and struts to stack on, and also delivers power to the nodes. Connectors utilizing strut components are fastened to the build plate PCB through heat staking in between the empty spaces left for nodes, and the whole structure is bolted to a larger block of milled plastic for stability. Power is routed to the nodes through a simple screw-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

terminal power block, which is easily connected to a power supply.

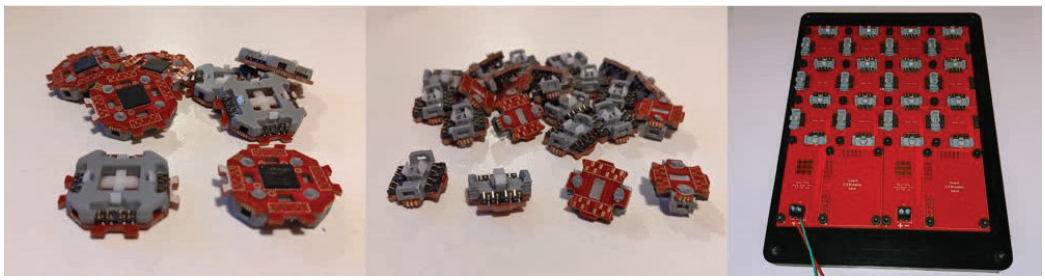


Figure 2-2: Meso-DICE components. Left: nodes; Middle: struts; Right: build plate

The specific build plate used for testing this project had some additional capabilities built into it, chief among them being an added section for an FPGA to be soldered to, which would then be routed to each of the nodes so that it can read the data that is being passed between nodes. This functionality has proven beneficial in other aspects of the project overall, but is not put in use for the work at hand, and will therefore not be utilized.

To program the nodes, a specialized programming stand was created by adding spring pins, spacing rods, and a 3D printed platform to a custom-made microcontroller PCB. Each node has a specific pattern of contact pads on the bottom that only line up in a specific orientation (the "top" of a node is the side with the white strip, and the "top" of the programmer is the side with the notch in it). By pressing a node onto the stand in the correct orientation, it can then be programmed. This specialized stand was specifically created with robotic assembly of nodes and struts in mind. End effectors on the robot arm can grasp a node, move it to the programming stand, press down and program, then move it to the build plate and press it in place. The programming stand can be seen in Figure 2-3.

As mentioned before, each Meso-DICE node has connection pads that carry various signals throughout each node and in between nodes. A simplified representation of these sets of pads is included in Figure 2-4 and Tables 2.1 and 2.2. When any given set of pads on the top side a node is oriented upwards, you can see that the footprint is the same no matter the direction. On the bottom side of the nodes, this footprint


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

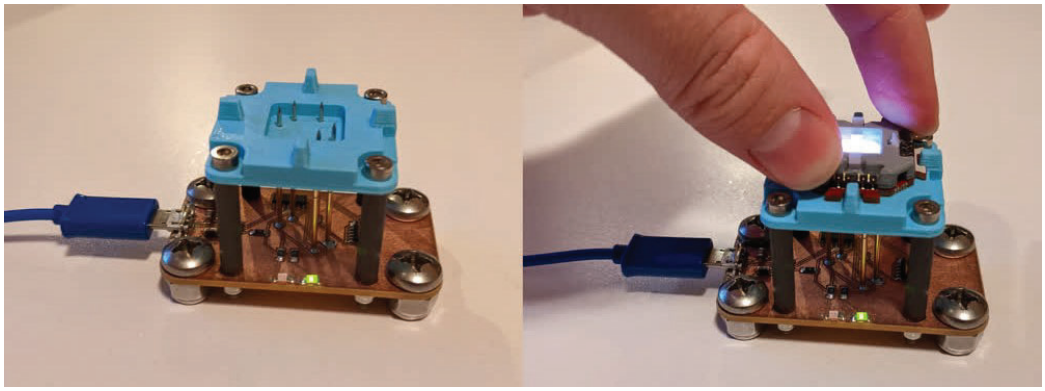


Figure 2-3: Meso-DICE node programming stand. Left: inactive; Right: programming a node

is mirrored so that signals can be ported easily from the top side to the bottom side through vias in the PCB. So, while the symmetry makes signal routing and tiling nodes together much easier, it can still be confusing to figure out exactly where each signal line can be accessed.

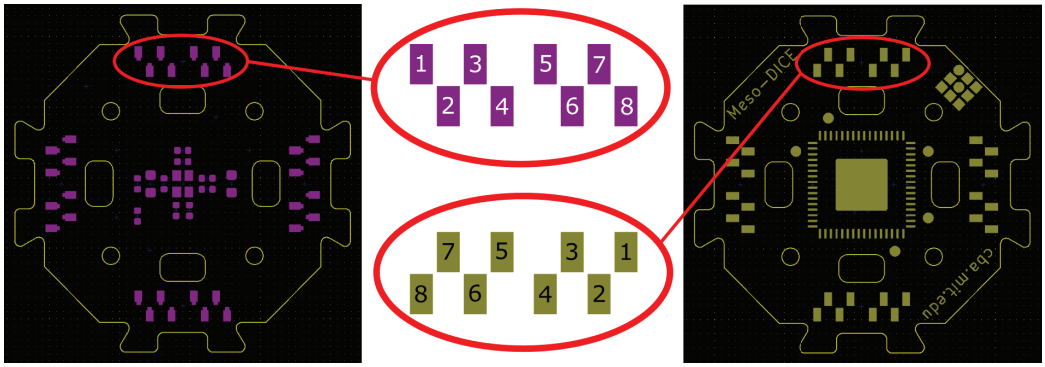



Figure 2-4: Meso-DICE node pinout diagram. Left: top side of node; Middle: zoomed in view of sets of both top and bottom node pads; Right: bottom side of node. Note that the pads and corresponding pin numbers appear mirrored to help demonstrate the connection between top and bottom pads. Meso-DICE node schematic designs courtesy of Zach Fredin.

As was mentioned previously, Meso-DICE was envisioned as an iteration of the DICE project that could be designed for end-to-end demonstrations of robotic assembly and programming [12]. Issues with the other iterations of the project included the lack of reconfigurability with Static-DICE and consistent failures of connector pieces


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| Pin | East | North | West | South |
|---|---|---|---|---|
| 1 | GND | GND | GND | GND |
| 2 | SIG_EAST | TX_NORTH | TX_WEST | SIG_SOUTH |
| 3 | XCK_EAST | RX_NORTH | RX_WEST | XCK_SOUTH |
| 4 | TX_TOP | TX_TOP | XCK_TOP | XCK_TOP |
| 5 | RX_TOP | RX_TOP | SIG_TOP | SIG_TOP |
| 6 | TX_EAST | XCK_NORTH | XCK_WEST | TX_SOUTH |
| 7 | RX_EAST | SIG_NORTH | SIG_WEST | RX_SOUTH |
| 8 | +3V3 | +3V3 | +3V3 | +3V3 |

Table 2.1: Meso-DICE node pinout (top of node).

| Pin | East | North | West | South |
|---|---|---|---|---|
| 1 | GND | GND | GND | GND |
| 2 | SIG_EAST | TX_NORTH | TX_WEST | SIG_SOUTH |
| 3 | XCK_EAST | RX_NORTH | RX_WEST | XCK_SOUTH |
| 4 | RX_BOTTOM | RX_BOTTOM | XCK_BOTTOM | XCK_BOTTOM |
| 5 | TX_BOTTOM | TX_BOTTOM | SIG_BOTTOM | SIG_BOTTOM |
| 6 | TX_EAST | XCK_NORTH | XCK_WEST | TX_SOUTH |
| 7 | RX_EAST | SIG_NORTH | SIG_WEST | RX_SOUTH |
| 8 | +3V3 | +3V3 | +3V3 | +3V3 |

Table 2.2: Meso-DICE node pinout (bottom of node).

with Tiny-DICE. Meso-DICE appears to get the best of both of those iterations, while minimizing the flaws of each, to make it a viable path to continue on. For the idea of integrating displays in particular, however, the nodes are in the correct size domain to easily attach a single display to each node (or at least each node on the top layer).

At this time, it is important to recognize those responsible for the DICE project's current implementation and design. Zach Fredin, a current Master's student and researcher at the Center for Bits and Atoms, is primarily responsible for the design and fabrication of all Meso-DICE hardware, including programming and testing equipment. For more information on the development of the different DICE iterations, it will be published in his thesis, being written alongside this one. [11] Erik Strand, a current PhD student and researcher at the Center for Bits and Atoms, is primarily responsible for the code and development platform used to run and program the DICE nodes, including many of the applications discussed in this work. Examples of the first applications developed for and run on DICE hardware can be found in his



253
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Master's thesis. [24] Now, with the design of Meso-DICE sufficiently explained, it is time to go over the displays and how they're integrated.

## 2.2 Display Integration

The display used for this project is a commercially available TFT LCD display, specifically the Adafruit 1.8" Color TFT LCD display with the ST7735R TFT driver. [1] The microcontroller used to control the display and interface with the DICE nodes is the Adafruit Feather M4 Express. [2] This combination was selected for a couple of reasons, including ease of programming, relative size, availability, and the fact that the Feather M4 uses the same ATSAMD51 chip that is contained in each of the DICE nodes. In all honesty, however, the exact display and microcontroller are fairly trivial components of the project as a whole. The goal with connecting the display to the Meso-DICE nodes was straightforward, so there was no need to reinvent the wheel there. To be thorough, however, the connections between the TFT display and the Feather M4 are listed in Table 2.3. Also, the connected devices are shown in Figure 2-5.

| **TFT Display** | **Feather M4** |
|---|---|
| LITE | 3V |
| MISO | MI |
| SCK | SCK |
| MOSI | MO |
| TFT_CS | 5 |
| CARD_CS | N/A |
| D/C | 6 |
| RESET | 9 |
| VCC | 3V |
| GND | GND |

Table 2.3: TFT display to Feather M4 connection guide.

The display itself is programmed through the Arduino IDE. After installing the correct libraries to handle both the Feather M4 and the display itself, it is quite easy to control. Utilizing the Adafruit GFX graphics library, as well as the Adafruit ST7735 display driver library, manipulating pixels and drawing shapes in various colors is


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

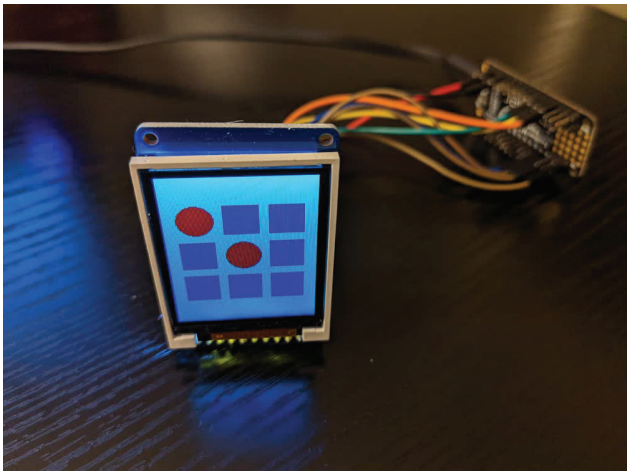


Figure 2-5: TFT display connected to the Feather M4 running some test display code.

very straightforward. Again, the wheel isn't being reinvented, so further explanation here is not necessary.

Connecting the display to the Meso-DICE nodes, on the other hand, requires some specialized hardware due to the self-contained nature of the DICE nodes and struts. A special header pin strut was fabricated specifically for this purpose. The typical PCB board used for all other struts was taken and epoxied to an electrical protoboard with wires connecting the PCB pad connections to the header pins soldered to the protoboard. This header pin strut can then be placed in the location of any other top layer strut in the structure. The contact pads on the bottom allow data to still be sent from one node to another, while also routing those signals up to the header pins, which allows that data to now be accessed by a simple jumper cable. This header pin strut is then connected to the RX pin of the Feather M4 to begin transmitting data that will be interpreted and translated to pixels on the display. This pin header strut and its subsequent pinout diagram can be found in Figure 2-6 and Table 2.4.

## 2.3 Sending Data

The way DICE nodes transmit and receive data between themselves is by utilizing a token passing scheme specifically created for this architecture. There are various


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

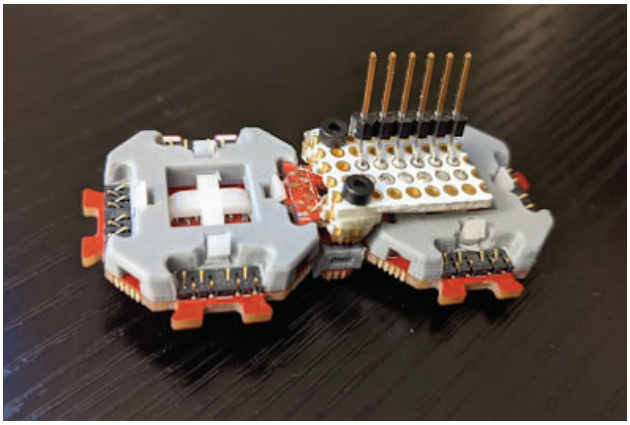


Figure 2-6: Pin header strut mounted between two Meso-DICE nodes used for signal interrogation by oscilloscope and to connect to the display.

| **Pin (Strut)** | **Pin (DICE)** | **Signal (DICE)** |
|---|---|---|
| 21 | 1 | GND |
| 20 | 2 | SIG or TX |
| 19 | 3 | XCK or RX |
| 18 | 8 | +3V3 |
| 17 | 7 | RX or SIG |
| 16 | 6 | TX or XCK |

Table 2.4: Pin header strut pinout with corresponding Meso-DICE node signals.

handshakes that occur, as well as the attaching and detaching of buffers to properly manage the data passing back and forth. At first, the idea was to implement this same token passing scheme between the display's microcontroller and the DICE nodes. However, a simpler solution was to just use the Feather M4's serial line to receive specific data directly from the nodes.

By matching the baud rates between the nodes and display, the display's microcontroller can properly read data that is received from the nodes, then according to the application, determine how to translate that into what to show on the display. Testing was conducted to ensure that the appropriate data was being sent out by the nodes and properly read into the displays. For example, Figure 2-7 shows a test bit pattern being sent out from a DICE node, which was then matched with the same bit pattern being sent out from the Feather.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

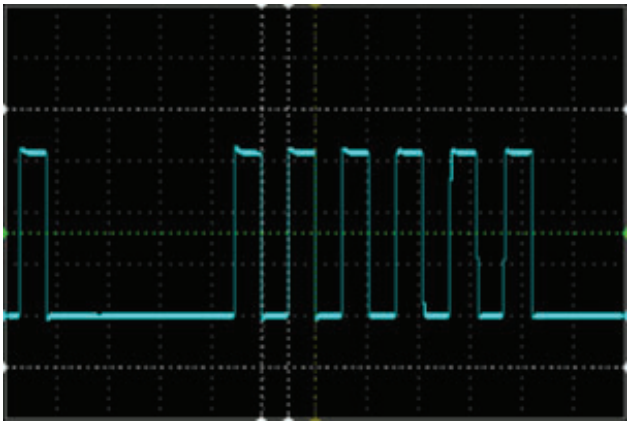


Figure 2-7: Zoomed in oscilloscope reading showing proper bit pattern being sent from node at proper frequency.

Because the data transmissions aren't happening using the same token passing scheme that the nodes typically use, we do have to use specialized code for the top layer nodes that are connected to the displays. However, depending on the application, the nodes don't transmit all of the information that would be required by the displays amongst themselves anyway, so specific top layer node code would still need to be written. Therefore, bypassing the token passing scheme for this simpler serial communication doesn't throw things off too much.

A secondary strut to intercept data from the nodes to be sent to the display was also created. By simply using some very fine 34 AWG magnet wire soldered directly to one end of a jumper cable and the pad of a strut with one set of Molex connectors removed, a solid connection was established. This secondary testing strut, as pictured in Figure 2-8 was used only for tests which required more than one display to be connected at a time. While it is definitely not as robust as the original pin header strut, it had the benefit of being connected to the TX_TOP data line, which allows the node that data is being pulled from to still have access to its surrounding neighbors. Node data needs to be properly formatted and directly sent to the pin connected to the display. If typical token passing data is sent over that line to another node, the Feather will attempt to read that data and display something according to those values. Because data between nodes doesn't necessarily contain all the pertinent


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

information being stored in a particular node, the Feather must opt to be sent data packaged specifically for it to interpret, rather than the smaller amounts of data being transferred.

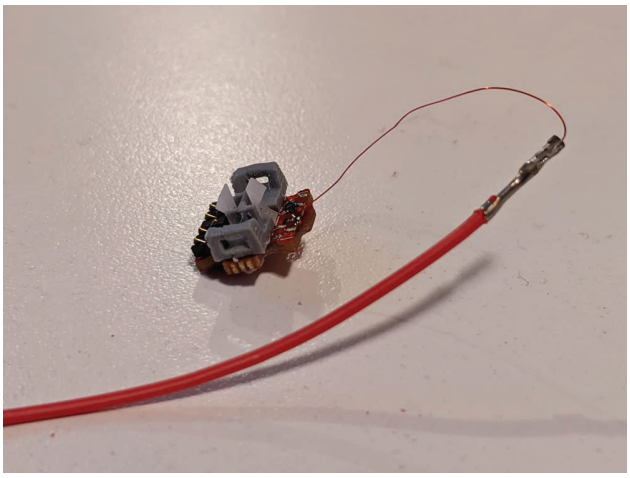


Figure 2-8: Secondary testing strut with access to TX_TOP signal line.

## 2.4 Packaging

Ideally, everything needed for the display could nicely fit into the same approximate volume of a Meso-DICE node and be easily connected as a top layer of the structure itself. At its current state, however, attention and efforts have been focused on determining the capabilities and unique applications that are enabled by these displays within the DICE architecture, so this ideal packaging has yet to be achieved. More work needs to be conducted to determine the best displays and microcontrollers to interface them with, fabricate Meso-DICE sized and shaped "display node" connectors with the appropriate data lines routed, and design the proper physical components needed to hold it all in place and look pleasing. This is definitely an area that needs to be explored further in the future.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 3

# DICE Debugging and Data Visualization

## 3.1 Debugging

With the displays now connected to the DICE nodes, they can now be utilized to debug the computations being performed by the nodes. Much like print statements are used to debug code that is run on a computer with an attached monitor, the displays can be connected to any given node and have the data or any other useful information printed on the display. The nodes can either be programmed to send certain parts of the computation to the display to verify that it's working at that point in the code, or the display can intercept the data as it's being passed between one node to the other to get a sense of the interaction between the nodes (this sounds contradictory to what was stated in the last chapter, but for debugging, the specific data being passed might be of more interest than anything else).

Up until this point, there were only two ways to attempt to troubleshoot or debug code run on the DICE nodes. One way is to run the code on a simulated version of the DICE nodes. By running this code on a laptop or desktop computer, the simulation could then print out or display any useful debugging information required on the screen. This method is extremely useful in that it doesn't require any physical hardware to develop and test the code. However, there are some obvious limitations.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

If the simulation isn't a perfect replication of the hardware, there's no guarantee that the code will work just as well on the hardware as it does in simulation. Sometimes there is a fault in the actual hardware that doesn't correspond to what the simulation does, so it is impossible to diagnose that problem when you don't know it even exists. By using the displays to debug the code, they can be used as a diagnostic tool to determine if there is a problem with the physical hardware or pinpoint the exact location in the code that is causing issues when run on the hardware.

The other method that has been used in debugging is logging all of the data exchanges that take place by using a connected FPGA to pull the data out of or being exchanged between each node. That data log can then be parsed for any errors and when they were encountered. The big issue with this method is that it isn't done in real-time. The data has to be parsed after the computations are completed. Errors can still be located, but not without parsing through all of the data in the whole system. Using the displays can help to see the error when it occurs and subsequently iron out that bug. Seeing it in real-time helps you know which part of the code to look at and having the display physically connected helps you even more to know which node is causing the issue.

While these other methods do have their merits and are being improved upon, the displays provide an excellent additional tool to debug any sort of code being developed for the DICE platform. They are easy to use and can show a wide variety of useful information customized to the specific needs of the application. The displays are sure to increase efficiency and ease of future development.

## 3.2 Data Visualization

A primary motivation of the DICE project was the various applications that it can adapt to. Some of these applications include particle simulations, machine learning, and graphics rendering. [12] Some of these applications will be explored in greater detail later on, but to illustrate the variety of use cases for these nodes and their displays, a few smaller examples will be noted here.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

One concept for machine learning applications within the DICE framework is that each node represents the different input, output, and hidden layers and neurons of the system. As the nodes make their respective computations and exchange their results between each other, there is a chance that the specific structure of the neural network isn't as effective as it could be, and could require a rearrangement. Because of the modularity and robotic assembly capabilities of the structure, the nodes could be rearranged automatically and continue in the machine learning process. An idea for visualization can be easily understood by imagining that the system is learning to identify different handwritten letters or numbers. The input, in this case a specific handwritten letter or number, could be shown, as well as the paths of the interconnections between the different neurons and layers, and finally, the letter or number that the system has identified at the end. Being able to physically see the entire network working, almost appearing as to be living and breathing, is sure to add better insights and understanding to how the network is performing, as well as how it can be improved.

Another simpler application is that of parallel computing. In one instance of this, the DICE nodes work together to perform a simple calculation of pi. Each node is given a specific section of the calculation to perform and upon completion of its portion, it sends its results along the chain until finally terminating at the final node that sums each portion up to determine the final result. A display can be attached to the final node in order to see the final result, or it could be connected anywhere along the chain in order to see the information being cascaded down the line. The DICE implementation of this calculation was developed by Erik Strand.

The above example was actually put into practice as a demonstration of the visualization and debugging capabilities of the DICE display hardware. In this test, multiple chains of nodes, starting with 2 and going up until 12, were assembled and connected. As the calculation moved up and down the line of nodes, the final result was sent to the display. This demonstration and its results are shown in Figures 3-1 and 3-2 and Table 3.1. The differences in the received values for each test show that with the increasing number of nodes, the precision of the answer increases, as well.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

While what is displayed isn't necessarily the most exciting thing to look at, it does show how simple it is to connect the display and read out data being processed within the nodes. The time it took for each calculation to complete was also measured. As can be seen, the addition of each node increases the amount of time it takes to complete the computation fairly linearly. Every single node is still doing the same amount of operations as the last, so the total number of operations scales proportionally with the number of nodes.

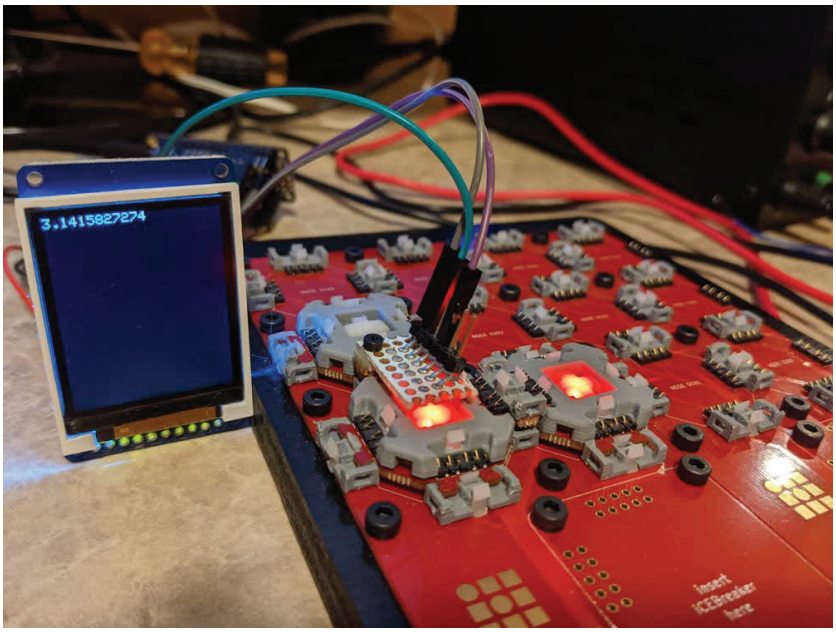



Figure 3-1: Pi is calculated across two nodes with the final result sent to and shown on the display.

In terms of graphics rendering, there are many different ways to visualize the relevant information. One idea is to have the displays render cross-sectional views of an object according to the different layers of physical nodes that represent the different spatial layers of the simulated object. Another more abstract idea is to have multiple surface layers of displays assembled together to act as faces of a cube so that each face of displays would be able to show a rendered object from its respective direction. A deeper look into a specific graphics rendering technique will be explored in the next chapter. Also, while more advanced displays would be required to give a true


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

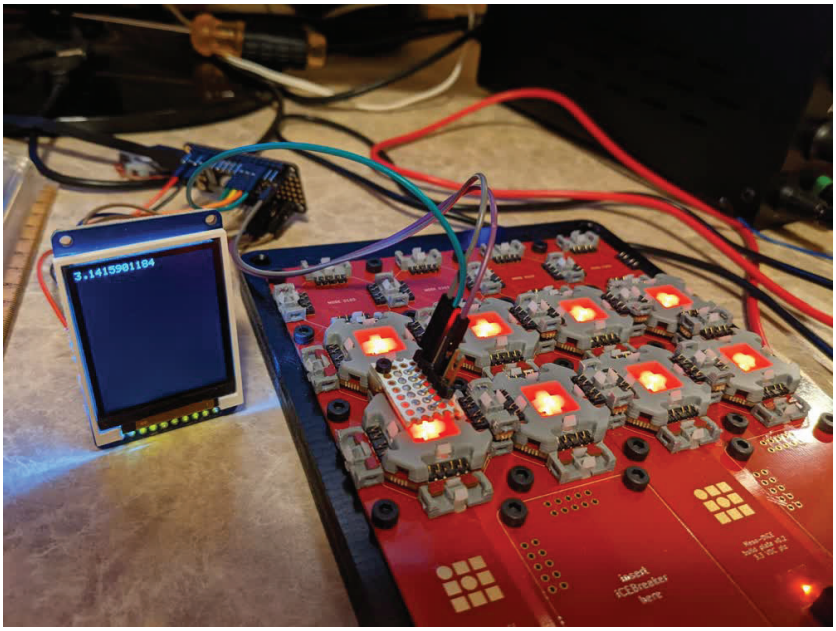



Figure 3-2: Pi is calculated across eight nodes with the final result sent to and shown on the display.

| # of Nodes | Pi Calculation Result | Time Elapsed (s) |
|---|---|---|
| 2 | 3.1415827274 | 1.1 |
| 3 | 3.1415860653 | 1.3 |
| 4 | 3.1415877342 | 1.5 |
| 5 | 3.1415886879 | 1.7 |
| 6 | 3.1415894032 | 1.9 |
| 7 | 3.1415898800 | 2.2 |
| 8 | 3.1415901184 | 2.4 |
| 9 | 3.1415903568 | 2.6 |
| 10 | 3.1415905952 | 2.8 |
| 11 | 3.1415908337 | 3.1 |
| 12 | 3.1415910721 | 3.4 |

Table 3.1: Pi calculation test results with varying number of nodes.

three-dimensional effect, the parallel computing power of the nodes could fairly easily compute different volumetric and holographic data that could then be showcased on a given face or multiple faces of a display structure.

These, along with all of the other countless applications that have yet to be explored, demonstrate the wide variety of new capabilities that not only the DICE


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

nodes themselves enable, but also the integrated displays. As this project is still so recent, more development of DICE applications will be needed in order to continue maximizing these capabilities. It is hoped that others will see the usefulness of this technology and desire to develop and test their own applications and subsequent visualizations using the platform. While the exact effectiveness and value that these displays add will be quantified and discussed later, it is apparent that this could be the foundation for a whole new way to approach computing and visualization.

## 3.3 Code Example

Revisiting the pi calculation example, a specific adaptation that is needed within the code should be discussed. While the specific method used may seem simple, it proved to be quite the non-trivial development in getting both hardware devices (node and display) communicating properly with each other. Within the DICE code, there is an object called "token_payload" that stores that individual node's calculated pi value, summed with what was shared with it from the next node in the chain. This value's type is a float, but the DICE hardware was developed to transmit uint32_t type data. A simple data conversion here isn't possible due to rounding errors causing a complete loss of precision. Therefore, the "reinterpret_cast" function is utilized with a combination of pointers and pointer addresses to preserve the data for the transmission. The data is then sent over one of the node's TX lines to the Feather for processing. This whole process can be seen in the following code snippet:

```
// Prepare the data that will be sent
float pi_actual = 3.1415926536;
float pi_result = token_payload->pi[index];


// Convert it to a uint32_t to send
uint32_t displayData[2];
displayData[0] = *reinterpret_cast<uint32_t*>(&pi_actual);
displayData[1] = *reinterpret_cast<uint32_t*>(&pi_result);
```


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

```
// Send both values over the TX_NORTH line
tx_north(displayData, 2);
```

On the receiving end, there is a bit of work to receive the entirety of the uint32_t data, in the form of reading in each byte individually and then concatenating it all together. However, once it is received and processed correctly, it can then be converted back into a float using the same "reinterpret_cast" function, but replacing the variable type specified. Some tricks are then implemented to erase the previous value from the display before finally printing the received value to the display.

```
// Make sure the whole data packet is available
if (Serial1.available() >= 4) {

    // Read in the data one byte at a time
    byte byte0 = Serial1.read();
    byte byte1 = Serial1.read();
    byte byte2 = Serial1.read();
    byte byte3 = Serial1.read();

    // Combine the bytes into the whole uint32_t data
    uint32_t incomingData = (byte3 << 24) | (byte2 << 16) |
        (byte1 << 8) | (byte0);

    // Convert the uint32_t data into a float
    float piData = *reinterpret_cast<float*>(&incomingData);

    // Erase the previous result from the display
    tft.setCursor(0, 0);
    tft.setTextColor(ST77XX_BLACK);
    tft.println(prevData, DEC);
```


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

```
        // Print the new result to the display
        tft.setCursor(0, 0);
        tft.setTextColor(ST77XX_WHITE);
        tft.println(piData, DEC);

        // Update the prevData value (same type as piData)
        // Note: prevData is initialized outside of this function
        prevData = piData;
    }
```

Note that in this example, there would be two values being sent over and received, the actual value of pi and the calculated result. However, only the calculated result would remain long enough on the display to actually be able to see it. It's trivial to implement more code to keep both values visible, but this code snippet is meant as a simple example of how data can actually be transmitted from node to display. With that simple demonstration out of the way, we now move on to a more complex example.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 4

# Application: Particle Simulation

## 4.1 Motivation

As the DICE project was inspired by the CAM-8 project, it only seems right to tackle one of the main issues that system was built for: particle simulations [20]. The way one particle interacts with another can be pretty straightforward to simulate, but once you add in hundreds, even thousands of particles into one system, the simulation can get bogged down pretty easily due to the strain on the computing resources. However, when applied within the CAM-8 or DICE architectures, the parallel and locally driven computing processes make such a task much more feasible. From simply determining how a set of gas particles will move about in a given space, to analyzing the strain experienced by particles in a complex structure under duress, physical particle simulations can provide a lot of beneficial information before a physical experiment is conducted. Being able to run such a simulation without needing the pure processing power of something like a high performance computing center is quite the alluring prospect.

Rather than just running these simulations and checking the results after the fact, however, it is much more informative to see the full simulation running in real time. Without that capability, there is a possibility of missing out on vital information, such as the exact interactions that happened before a strain-induced fracture occurred, or to be able to notice an overall trend or pattern that would have been much harder to





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

decipher with a more general look at numbers on a page. For all of these reasons, it is desirable to implement a structure for performing particle simulations within the DICE display framework.

## 4.2 Implementation

A simple particle simulation can be construed where each particle has simple interaction forces with each other particle, and the system on a whole can be observed in its movements. As mentioned before, each DICE node is given a specific region of space which it is responsible for. Any one particular node will compute the interaction forces and trajectory for each given particle within its spatial boundaries. When a particle's trajectory is headed for a boundary between nodes, the data for that particle is then sent to that neighboring node, which then assumes responsibility for that particle as it simply crosses over between nodes and their respective spatial boundaries. If a particle approaches a boundary without an adjacent node, that boundary is treated as a wall which the particle will just bounce off of and continue within that region.

A few different versions of this sort of particle simulation have been developed for use with the DICE architecture, all by Erik Strand. The most robust example of these is that of a simulated DICE node environment that utilized a grid of virtual nodes, each running the same general simulation code, then displayed the results of the simulation on the screen. [24] This simulation successfully proved that the ideas behind DICE node interaction worked and that the data could be successfully computed at the individual node level, then passed on to its appropriate neighbors to correctly model particle behavior. While this particle simulation performs exceptionally well, it is only virtual, and has not been directly ported to code that will properly run on the physical DICE nodes. Much of it has been adapted, however, and recent results show the simulation running properly on a single Meso-DICE node.

It is that current iteration of the developed code that the particle simulation with display functionality is tested. Using that version of the code to guide the implementation, a demonstration was created that would run the single-node particle


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

simulation and after each step, send the location of each particle to the display, which would then update continuously to show the movement of the particles. Similar code adaptations to those presented in the previous chapter in the pi calculation example are used to pass along the precise location of each particle, as well as to erase and redraw the particles according to their new positions.

The display is first initialized by drawing a simple bounding box square to represent the boundaries of the simulation. This can be adapted as needed or desired, but a simple box of size 100 pixels by 100 pixels seemed appropriate. Then, like the pi calculation example, the display starts listening for the particle information to be sent over. Every time that a new set of data is sent over, it reads and interprets the data, then draws a small point at the particle's location, mapped to the display size. This data that is read into the display comes in sets of x- and y-coordinates for each of the particles and is therefore quite simple to store and parse. The simulation code is then set on a loop so that the DICE node continuously updates the positions of the particles according to the velocities of the particles, then sends that positional data to the display so the simulation visualization can be continuously updated.

While simple, this setup can be used with even more complicated particle simulations, including when node to node communication is implemented fully on the hardware. The display code just expects its connected node to update it with particle positions, and will update the display accordingly. So long as the nodes send that data over, it doesn't matter what the interaction forces with other particles, boundaries, or whatever, the simulation can still be successfully displayed.

## 4.3 Results

After working through some bugs in the initialization of the display, the first particle was able to be seen on the display. It moved about fairly slowly (that particular test particle has a relatively low velocity), but there did not appear to be any glitching or sudden movements, and the display updated quickly enough for a smooth picture. As more particles were added to the simulation, the display was able to show each of





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

them moving around smoothly, as well. Faster particles appeared fast, slower particles appeared slow.

In terms of the DICE code, each of the particles appeared to obey the laws of the interaction forces given to them and to appropriately bounce off of any node boundaries present. As it was a very basic simulation, no unexpected or erratic behavior occurred. The code and algorithms run on the DICE node has proven to be robust, but that is of less interest to this test than the performance of the display.

Rather than seeing strings of numbers continuously updating over the serial monitor (which happened often in the initial development of the node to display code), it was much easier to determine that the simulation was working in its entirety when it was fully visualized. Each particle can be easily isolated and tracked. Interactions between particles and the boundaries could be clearly seen. Figures 4-1, 4-2, and 4-3 show the simulation and subsequent visualization running at various particle counts. It all was a great showcase of the merits that the display bring to these simulations overall.

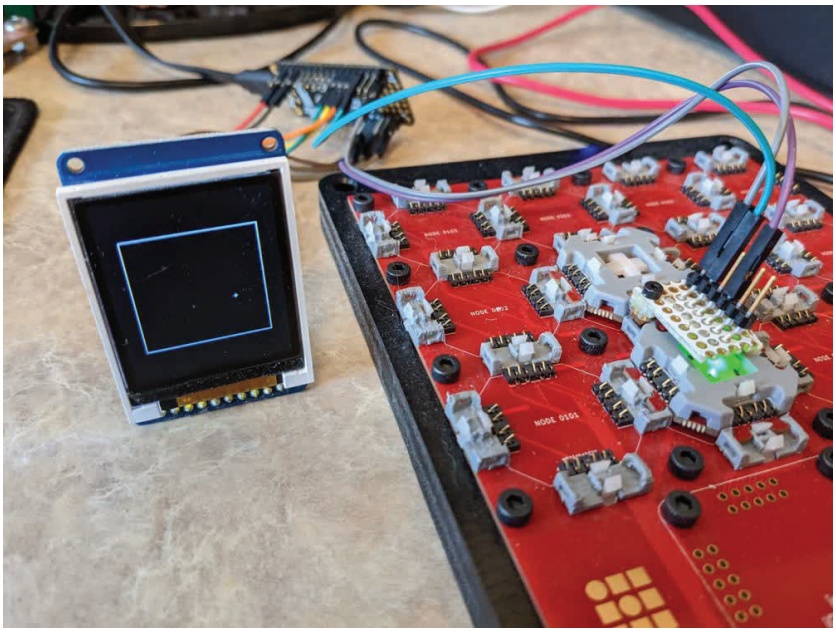


Figure 4-1: Particle simulation being run and visualized on the display with one particle.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

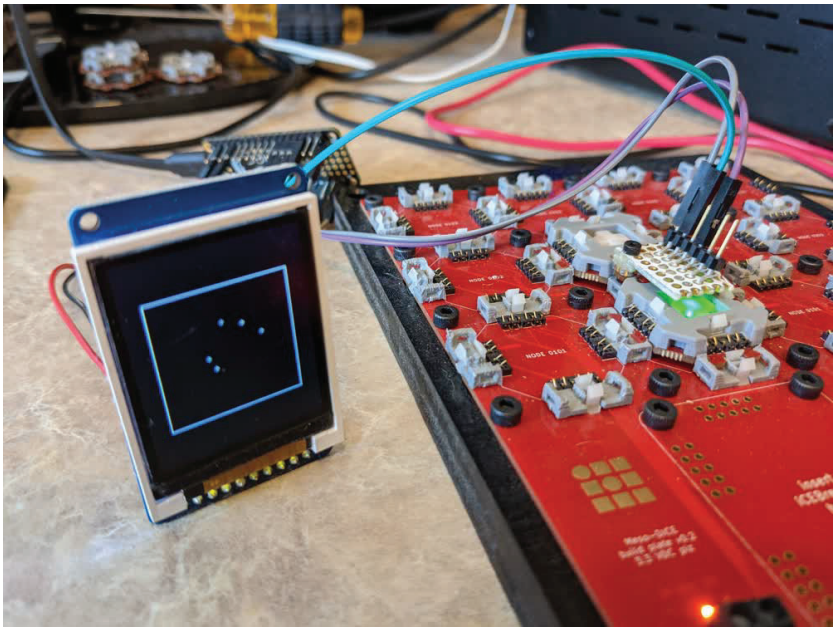


Figure 4-2: Particle simulation being run and visualized on the display with five particles.

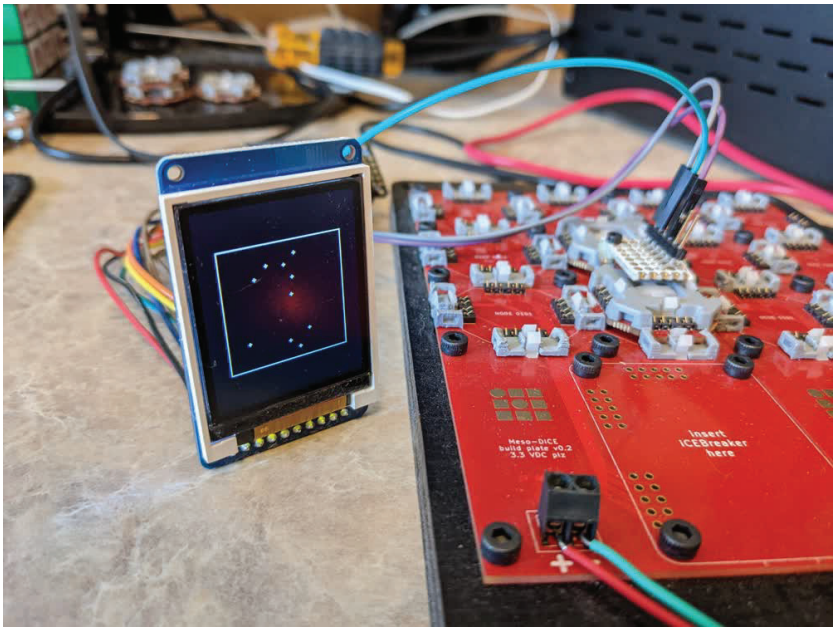



Figure 4-3: Particle simulation being run and visualized on the display with ten particles.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

However, that's not to say that the visualization was absolutely perfect. The main area of concern with the performance of the display is the speed at which it's able to update, and if that's fast enough to be a satisfactory visualization. If the node is calculating faster than the display can receive, interpret, and display data, then the data won't be organized properly and is therefore useless for the visualization. In most applications, there will be many more computational steps for the nodes to compute than the displays, so sufficient time should be allowed between display updates.

Unfortunately, this was the case in many of the tests, and the node's simulation code needed to be slowed down in order for proper communication and organization to be maintained. The simulation code already had a function it would call to wait so many clock cycles before running the loop of updating positions and velocities again, but that amount of time wasn't enough depending on how many particles were present in the simulation.

In fact, to determine the performance of the display in the simulation, a simple test was conducted to see how big of a delay (how many clock cycles to wait) were needed in order for the display to reliably keep up with the simulation, all while varying the number of particles present in the simulation. The results of this test can be found in Table 4.1.

| **# of Particles** | **# of Clock Cycles** | **Delay Time (ms)** |
|---|---|---|
| 1 | 100000 | 2.083 |
| 2 | 100000 | 2.083 |
| 3 | 150000 | 3.125 |
| 4 | 230000 | 4.792 |
| 5 | 250000 | 5.208 |
| 6 | 600000 | 12.500 |
| 7 | 350000 | 7.292 |
| 8 | 800000 | 16.667 |
| 9 | 520000 | 10.833 |
| 10 | 570000 | 11.875 |

Table 4.1: Delay time required per particle for display to reliably update all positions without the data getting unorganized. Delay time based off of DICE node's 48 MHz clock cycle.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

As can be seen, as more particles are introduced to the simulation, the more time the display needs to process all of the positional data and update the display. At lower particle counts, the delay isn't very substantial, but that is not the case as more particles are introduced. While the inherent slow feeling could be attributed to some of the test particles having relatively low velocities, the delay does become quite noticeable, regardless of the speed at which the particles are moving. What is really interesting is the two outliers in the measurements. For every particle count, the delay time increases slightly (or stays the same in the case of 1 and 2 particles). However, for a particle count of 6 or 8, the delay time increases substantially. This trend was validated on more than one node, and is quite puzzling. There doesn't seem to be any major differences in the code for those particle particles in terms of their initial positions and velocities, yet the delay required for them to run properly is quite erratic. More testing will need to be conducted to figure out what is going on in those instances.

## 4.4 Improvements and Variations

The first obvious improvement that needs to occur is increasing the speed at which the display can operate. This must be done by optimizing the code and improving any glaring time-consuming functions that cause everything to take so long. More of the processing of the data could be offloaded to the the nodes themselves, but it still needs to be converted, sent, converted back, then displayed. More tests would be required to determine if the display's microcontroller or the DICE node is faster at certain tasks in order to determine exactly what needs to be done where. The whole point was to access data already being stored within the DICE node and make a compelling visualization out of it. One final more drastic solution would be to completely revamp how data is shared between node and display. For example, other communication protocols, such as I2C or SPI could be used, which can be faster than the UART protocol implemented throughout this work.

For the single-node simulation, all particles are located in the same node's spatial




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

region, and the node has wall boundaries set on all sides. In theory, the multiple-node simulation could go in one of two directions. Each node can either be responsible for a specific set of particles to keep track of over the whole simulation, or they can represent the specific spatial region and keep track of any particles that enter that region. The latter method makes it easier to scale the simulation, however, so that is the route that will be taken for the application.

Therefore, next obvious thing to do is to continue developing the code so that it is capable of fully simulating particles crossing boundaries between nodes using the necessary node to node communication there. It has been shown to work on virtual instances, but not yet on the hardware. The display code should still work by displaying whatever particle position data it is fed, so development would be more on the DICE side of things. As a quick concept of what such a test could look like, Figure 4-4 shows two nodes running the simulation next to each other (but not technically passing data between each other), with two separate displays visualizing the particles from their respective node. The simulations were started at different times to show some variety in particle positioning. Also, the green lines of the boundary boxes are representative of the shared boundary between the nodes. Again, this is just a concept of what it could look like, especially in terms of scaling up the simulation, but more work needs to be done.

There are a variety of other more complicated visualizations that can occur, all possible because of the simplicity of the parallel calculations and data transmissions between nodes and to the displays. For example, colors can be applied to the various particles to represent relative velocity to one another. More complex interaction forces can be introduced with different signifiers of different physical phenomena being coded into the visualization based on variations in size, color, speed, etc.

Another feature than can be implemented is by utilizing the third dimension that the DICE nodes are able to stack in and transmit and receive data. DICE nodes can be stacked on top of one another, allowing for a 3D particle simulation. One way this could work is to have all nodes within a stack of nodes send the coordinates for each particle in their region up to the above node. Once all coordinates for each particle


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

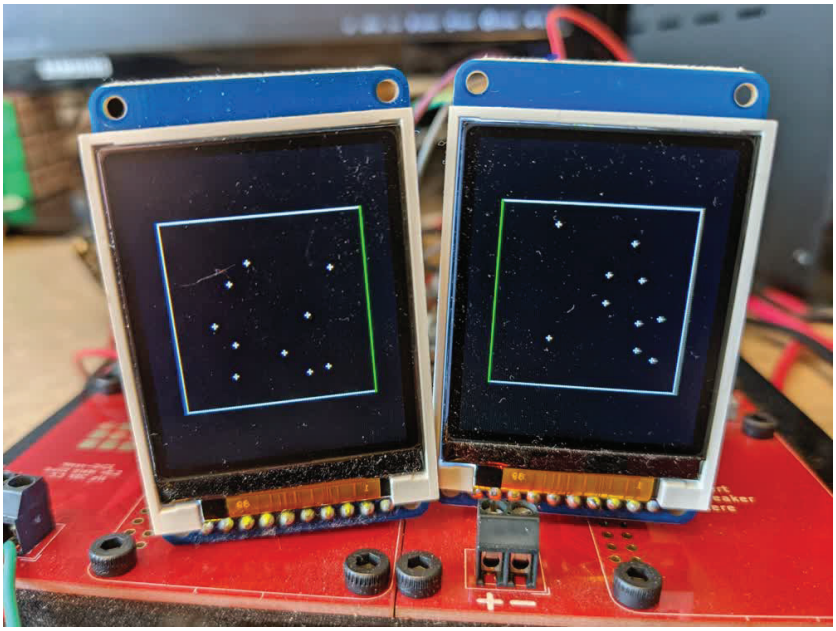


Figure 4-4: Concept visualization of what a particle simulation with two nodes communicating could look like. The green lines are representative of the shared border between the nodes.

in that stack of nodes is collected by the top layer node and sent to the display, the varying depths of the particles could be easily visualized using a simple shader. For example, the closer to the top, the more opaque the particle, and the closer to the bottom of the stack, the more transparent. As particles move higher and lower within the stack's spatial region, the corresponding point on the display can get more or less visible.

The two simulations run for this demonstration are in both two and three dimensions. The 2D simulation is a 4 x 3 node area, and the 3D simulation is a 3 x 2 x 2 node volume. Both simulations contain the same number of particles, with the height, width, and depth units being equal. The particles have just the basic interaction forces present, allowing them to bounce off of one another upon collision. Each particle is given a random initial velocity to incite movement and interaction.

One final thing worth mentioning is that of the characteristic of the displays that they currently work according to how well the data they receive matches up with the data that they're expecting. This can absolutely be both a bug and a feature,


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

depending on how you look at it. If mismatched data shows up, it breaks everything and ruins the visualization. On the other hand, though, it allows the visualization to be customized specifically for the application that it is working with, and can also help debug the problem of figuring out why the mismatched data occurred. It's a two-edged sword there, and there's sure to be more developments with it as this platform is used more and for more applications.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 5

# Application: Ray Tracing

## 5.1 Motivation

In the world of graphics rendering, the next iterations of the highest end commercially available products are locked behind years-long multi-billion dollar development cycles. These Graphical Processing Units (GPUs) specialize in delivering the best graphical quality for whatever sort of application they're applied to, being able to handle large amounts of data that make up the polygons, meshes, textures, colors, and everything else that is involved in a given scene or simulation environment. Like other high-end modern day computing products, they rely on an architecture that doesn't build in locality from the foundation, but still do extremely well with the extremely specialized and optimized architectures that they do implement.

Trying to build a GPU using DICE nodes and displays is an idea based on the desire to overcome this time consuming and expensive iterative process. Taking advantage of the scalability of the DICE architecture, a more complex or higher performing GPU could be achieved simply by scaling up and adding more nodes. For an application that requires just a little more or less rendering power, this flexibility in scaling could introduce a solution to providing the most efficient needs-based GPU for each application. It wouldn't be necessary to wait for the next, more powerful iteration of a device or get more power than is truly needed. It is this idea that makes DICE an attractive solution for graphics rendering applications.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

In particular, however, there is one graphical application that stands out from the rest: ray tracing. Ray tracing is the idea of computing the optical paths that each ray of light in a scene will travel in order to determine the exact lighting, color, and reflections that would naturally occur from that object. The general idea was first introduced by Albrecht Dürer in 1525, [14] but has been worked on extensively over the past few decades as the field of computer graphics has grown substantially. Many companies in the industry have developed all sorts of sophisticated software and programs to simulate and calculate all sorts of computer generated scenes, particularly those involved in making animated movies and effects. [7]

Because ray tracing algorithms are so computationally expensive, use cases for them have typically been limited to those of static images or pre-rendered scenes. However, advances in the hardware and in the algorithms themselves have led to enabling of real-time ray tracing capabilities. While it would seem that devices capable of real-time ray tracing would be limited to the most cost-prohibitive and specialized equipment, the technology can be found in the most recent generation of home video gaming consoles. However, even in those instances, extremely specialized hardware is used and took years of research to attain.

Therefore, it would be greatly advantageous to implement such a complex application on the simplified DICE hardware and use its parallel processing power to handle the heavy computations required. Utilizing the DICE architecture, the complex ray paths that help realistically render a scene optically could be divided up and computed locally to share the load and not be limited by overall memory speed and size. There are some limitations, such as resolution, that can arise from not using the most high-end chip-sets for the local computations, but the size and complexity of the scene shouldn't be an issue as the work can continue to be divvied up and executed in parallel by the entirety of the system.

The greater implication of executing such a system is, as was mentioned, the scalability of the system. By adding more nodes and displays, the resolution of the rendered image, the complexity of the scene, and the ability to perform it all in real-time is enhanced. The fundamentals that will work on the most basic of demon-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

strations can feasibly be made to work on a much larger system. The scaling iterative systems that could be created could combat the aforementioned years-long development cycle, all without having to build everything from the ground up. For these reasons, this application has been conceptualized and will hopefully help overcome these challenges.

## 5.2 Proposed Implementation

Before digging in on the exact implementation of this application, it is necessary to note that at this point the threshold of what has been achieved on the DICE hardware has been crossed. No one has yet attempted to create a ray tracing simulation on the system yet, either virtually or physically. Due to time constraints and the need for further development of the DICE node and display code, it is necessary to approach this application from a theoretical point of view. As more progress is made on the development front, the transition from theory to actual experiment should be quickly realized.

The implementation of this attempt at locally computed ray tracing takes advantage of the duality of light in that it acts like both a wave and a particle. Essentially, this application would utilize the same framework established by the particle simulation methods found in the previous chapter, but the particles are instead replaced with photons and the interaction forces between the particles are replaced with equations that dictate optical forces and phenomena. Rather than a particle simulation, this would transform into a sort of "photon propagation" simulation.

The DICE nodes would represent a volume which can be filled with various objects to be depicted by the displays. As the photons propagate through the scene, the rays will reflect off the object and those interactions can be calculated and recorded. The top layer of the simulation volume will act as a viewing window that captures the depiction of the scene based off of the interaction between the photons and the objects.

It's easy to think of ray tracing as "tracing" each ray of light as it is emitted from a light source, reflects off of various object surfaces, then exits the viewing window


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

where the viewer's vantage point is located. However, in practice, this method is not very feasible because it requires keeping track of every single ray of light emitted from the light source. So, in most algorithm implementations, the calculation starts at the vantage point of the viewer, goes through the viewing window, then goes on to interact with the various surfaces. Any rays that do not interact with an object or light source are represented by darkness (or the chosen background color). The rays that do interact with the objects or light sources then have the color calculated based on the object and light interactions, and that color is then displayed on the portion of the viewing window that the ray "originated" from. This typical method of ray tracing is called "forward" ray tracing, while following the actual path that the light would travel, from light source and eventually to vantage point, is called "backward" ray tracing. While it was mentioned that it's not very feasible to use backward ray tracing, it is necessary to simulate the effects of diffuse reflection of indirect light. [4] However, this is beyond the scope of this implementation, for we are just aiming to get basic ray tracing algorithms working within the DICE architecture.

Concept art of what this all might look like within the DICE perspective is included in Figure 5-1. The entire surface layer of displays (or just one display) will act as the viewing window through which the rays from the vantage point are sent. Each node will represent the volume of the scene, including representations of objects and light sources. Depending on the resolution used for the calculation, the corresponding number of photons will begin to propagate, starting at their own local node, then moving throughout the volume, being transferred from one node to another, just like the particles do in the particle simulation. When a given photon comes across an object, the proper reflection or transmission is calculated based on the characteristics of the object's surface at that point, then it continues on. Based on the paths and interactions, the appropriate color is reported back to the portion of the display(s) that the photon started from, which all combine into one all-encompassing image.

Note that this concept is applicable to all forms of ray tracing, not just the potential DICE implementation. The concept art is therefore very similar to the typical diagrams found in any medium with an explanation of how ray tracing works. What


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

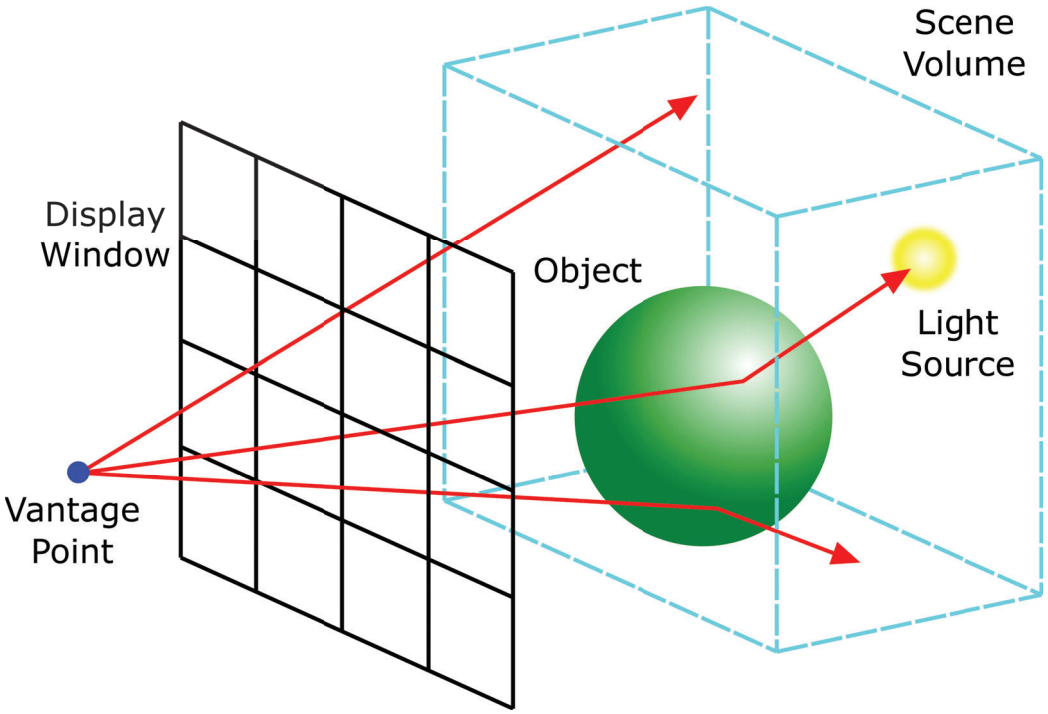



Figure 5-1: Concept art of the DICE ray tracing implementation. The display window in black represents pixels of a given display or a wall of full displays. The dashed cyan box represents the scene volume that will be broken up among various nodes to compute the paths of the rays. The red arrows show the paths of those photons as they propagate through the system, originating computationally from the blue vantage point. Objects and light sources (green and yellow spheres, respectively) will be represented in the scene volume.

sets this concept apart from all those other implementations, however, is the discrete methods that are employed within the DICE architecture. The volume containing the scene is broken into discrete spatial chunks that keep track of the photons propagating throughout the scene. The resulting rays that are traced and their determined colors are then represented on the displays at the point where the rays would have passed from the vantage point through the viewing window. This sort of "discrete ray tracing" has yet to be implemented, and the DICE architecture seems to be more than capable of achieving it.

Because the initial particle simulation worked on the virtual instance of the DICE hardware, [24] it suggests that by building upon that basic framework, this proposed


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

implementation should work, as well. Assuming success, this implementation would then have the capability to be scaled to create even larger displays or to increase the resolution or processing power of the computation. Each node has a certain computational capacity and by dividing up the work and with enough nodes, any heavy computation can theoretically be handled. There is obviously much work to be done to actually implement this within DICE. First off, finishing the development of the particle simulation to work on the physical hardware will help significantly as it builds the photon propagation framework that is required for this application. Next, adapting the optical equations into the code would allow for realistic path tracing to occur. Creating the objects and their surface characteristics is also a big part so that there is actually something to be reflected off of and rendered on the display. Simple tests with very simple objects will be the way to go once these different aspects are handled properly. From there, the viability of the DICE implementation of this application can be measured.

Beyond that, there are a variety of other ray tracing techniques that could be implemented to handle more complex scenes and ideas, such as distributed ray tracing to handle problems such as motion blur and fuzzy reflections. [8] Another technique that would be worth looking more into is that of "parallel rendering," which is essentially what we're already doing by breaking up the rendering tasks for each node, but more intricate and detailed methods exist. [22] Regardless of the complexity, for any further implementation, it would be wise to consult the plethora of information and guidance provided by the experts in the field. [18, 3]


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 6

# Evaluation

## 6.1 Debugging Enhancement

As was demonstrated in an earlier chapter, the ability to connect the display to any of the top layer nodes proves to be extremely beneficial when it comes to debugging the code running on the DICE nodes. Not only does it allow for detecting the problem in real time and determine which step of the code the issue occurs in, but it also pinpoints the exact node where the problem lies. While it's not as streamlined as using the virtual DICE nodes to emulate the code, when used in tandem with that method, code can be quickly developed and verified on the physical system that it will be running on when all is said and done.

Quantifying such an enhancement is a bit of a tricky task, but it can still be discussed. Much like having certain points in the code where a program can pause, verify the correct data, then continue on, the displays can act as such a troubleshooting diagnostic tool. Without actually having DICE code developers use the displays to test out its debugging enhancements, it is hard to determine exactly how much this tool will be utilized, but for the sake of this work, it has been of great use.

At its current state, debugging can be a little frustrating due to the fact that the node needing to be reprogrammed must be removed along with the pin header strut and other struts connected to it. Upon reprogramming, the structure must then be reassembled before more troubleshooting can occur. While this isn't necessary an




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

added issue because of the displays, it still can be time consuming. One thing that can overcome this is the end-to-end programming and assembly process that is being developed for the system using robotic assembly, however that may not be practical on an individual development basis.

Finally, as improvements in the packaging and optimization of the data transmission processes occur, the ease of use of the displays as a diagnostic tool will only increase. At that point, hopefully many more applications will have been explored and tested, giving many more opportunities for the refinement of the system and its effectiveness. At its current state, however, it does seem to be a worthwhile addition and contribution to the project overall.

## 6.2 Data Visualization

The DICE nodes on their own are capable of so many interesting applications and computations, and the addition of the displays into the system only amplifies the usefulness and intrigue of them. Being able to observe the workings of a system proves to be invaluable in gaining a better understanding of how each process unfolds. The amount of unique ways to showcase the data is seemingly limitless and can therefore elevate even the most mundane of measurements into more interesting observations that are more likely to unveil useful patterns hidden within the data.

There are some things to be aware of, however, when adding all of these visualizations to a given application. By adding more things for the system to do, it does tack on an additional time cost. While the system is still able to work asynchronously, the displays must be able to keep up with the data that the nodes are sending. If they don't, the data won't match up or the whole program must slow down to allow for the displays to catch up, making some applications less useful. This may not be an issue for all applications, especially for those with complex computational steps, but it can't be ignored outright.

It's also worth noting that some applications benefit more than others from having the displays in the first place, or some need only utilize one or two displays for the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

beginning and end of a chain of data, rather than the entire area of display space. The value that the visualizations truly add is quite subjective, and can really only be determined according to the developer and observer of whatever applications are run. All in all, there are some drawbacks to be aware of when using the displays and they must be considered in comparison to the benefits provided by them.

While the data that is shown on the displays is reliably correct, as in the displays only show exactly what is being read in and interpreted (which is why the unique "reinterpret_cast" function discussed in Chapter 3 is needed), the connections between the physical display and node, as well as the connections between the nodes, can be suspect at times. This causes incorrect behavior from the visualization, nullifying the usefulness of the visualized data. These unstable connections appear to be caused by iterative wear and tear from repetitive attachment and detachment from the build plate and other components. A sort of pre-loading structure to press down on each of the nodes and struts has been discussed, but is just one potential solution. This only seems to be an issue when more nodes are added together.

The added usefulness of the data that is visualized makes the integration of the displays a necessity for most applications. The cost to add the displays can be reduced by finding a display that's better suited for the DICE nodes in terms of size and design integration. More work will go into optimizing the data transferring to improve the speed of the visualizations, as well as improving the packaging so that the displays can each be seamlessly tessellated over the whole surface to form one larger display area.

## 6.3 New Applications

Aside from the applications that the DICE architecture is already capable of putting into practice, the addition of the displays enable other unique applications. Different forms of graphics rendering, model viewing, volumetric displays, and so forth were discussed, as well as the specific application of locally computed ray tracing. With such a recent project, the full array of potential applications has yet to be realized,




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

but there are surely more use cases that can be discovered. As these applications aren't possible without the DICE display framework, they serve as demonstrations of the merits of the system.

The ray tracing application in particular is one to get excited about because of the potential doors it opens in the world of graphics processing development. By scaling up the DICE-based GPU structure by adding more nodes and displays, intermediate steps between top of the line commercial product iterations could be achieved, bringing costs down and efficiency up. The whole idea of building the system to have just as much computational power as is needed, no more or no less, is now possible using this architecture.

## 6.4 Performance Projections

The basic DICE unit has been characterized and compared against various other computing chips and systems. According to performance projections, it would take roughly 3,000 DICE nodes to equal the performance of the Intel Core i7, while using 30% more power. [12] Based on these projections, it is worth exploring what additional computational and monetary costs are added on with the integration of the displays. Additionally, the basic performance of one singular DICE display node can be compared against top of the line GPU devices out on the market today, such as the NVIDIA V100.

First off, a discussion of monetary cost is required. Each DICE node costs roughly $3, and both the displays and the microcontrollers used for these tests cost approximately $20 each. Therefore, converting a top-layer node into a display node increases the cost more than tenfold. However, these specific devices were chosen for prototyping purposes and less expensive options could definitely pursued that could potentially fit the needs of the system even better.

In terms of power, a simple test was conducted where one node were placed on a build plate, which ran the particle simulation application drawing 60 mW of power. When connecting the display, the power consumption of the system increased by 130


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

mW up to 190 mW. Therefore, the power consumption of the system running a typical application can more than triple by adding in the display and its microcontroller. Again, by replacing these prototyping devices with something specifically designed for this system, this increase can be mitigated.

Coming back to the NVIDIA V100 comparison, however, shows how much work can continue to be done. Using the same math that compared DICE nodes to the Intel Core i7, we can determine that approximately 750,000 nodes would be required to match the performance of the NVIDIA V100, while using far more power. If 3,000 nodes didn't sound like it was pushing any limits, then 750,000 nodes is sure to sound absurd. Tack on the additional power consumption of the display layer and a potential decrease in practical computer power due to the need to slow the computations down so the displays can keep up, and it makes this system seem so inferior that its existence is called into question.

The thing to remember about all this, however, is that this system wasn't built to directly compete with these already existing products, but to provide an alternative architecture to potentially guide future developments in computing as a whole. These high-end computing devices are based on traditional non-locality based architectures. DICE is presented as an alternative that could potentially shape the next big wave of computing research.

It is worth noting, however, that there are other branches of this project being explored, one of them being the implementation of the DICE architecture using superconducting electronics. [5] According to current tests and projections, a custom-made "Super-DICE" node would utilize these superconducting advances to achieve computational and power consumption levels that are unheard of. Just 10 Super-DICE nodes could match the performance of the Intel Core i7, while using a fraction of a percent of the amount of power. Scaling up to the NVIDIA V100, it would take around 2,500 Super-DICE nodes to equal the performance, but again, only using a near-insignificant amount of power. These projections even show that if successfully implemented, it could lead to an improvement five orders of magnitude better than state-of-the-art super-computing systems. [12]


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

So again, while there is a lot of room for optimization, this system still gets the job done. Even though at its current state, these nodes would need to be scaled almost infinitely to match the levels of performance of modern computing structures, the fact that they work as well as they do is an improvement upon what has come before. There are various routes that the project can take in the future, but many, if not all, have the potential to achieve some very promising results and make progress in the fields of computing and graphics.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 7

# Conclusion and Future Work

The work done in this thesis is hopefully another step forward not only for the DICE project as a whole, but for the whole field of computing and displays. The displays integrated into this system not only enhance the capabilities of the DICE nodes themselves, but enable other useful applications that wouldn't be possible otherwise. One of these applications is that of locally computed ray tracing, which could shape up to be a great advance in the GPU industry. There are some obvious limitations from the system as a whole, but there is quite a bit of work that can be done to overcome or minimize those limitations, making this system even more efficient and beneficial for those who put it into practice.

First, an obvious area for improvement that could go a long way in making the system easier to use is packaging. While this may seem superficial, it would certainly improve the user experience with programming the devices. The specialized header pin strut did its job for the testing conducted, but it would be even better to have everything designed so that the displays could mount and latch on top of each node just like the other nodes do to stack on top of each other. In order to do this, however, more testing needs to be conducted on displays to determine which ones would work best for the system. As different applications are developed, it will be more clear what the needs are in a display, especially in terms of resolution and color capabilities, which would further guide the decision for which display is best for the system. It was mentioned that the Feather M4 microcontroller was chosen to drive





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

the display because of its compatibility with the display, but also because it uses the same ATSAMD51 chip as the DICE nodes. While it might not be possible to drive the display using only a DICE node, a special display node could be designed to include the extra components necessary beyond the Meso-DICE components, and then could mount on top of a typical DICE node to fit in with the rest of the assembly process. It would not only be aesthetically pleasing, but also very practical, to find a design that helped create a tight tessellation between each display to create a seamless layer of tiled displays. All these improvements would be sure to not only make things look nicer superficially, but also make them work together better and make the development and testing processes easier. Concept art of how this idealized packaging could look is included in Figure 7-1.

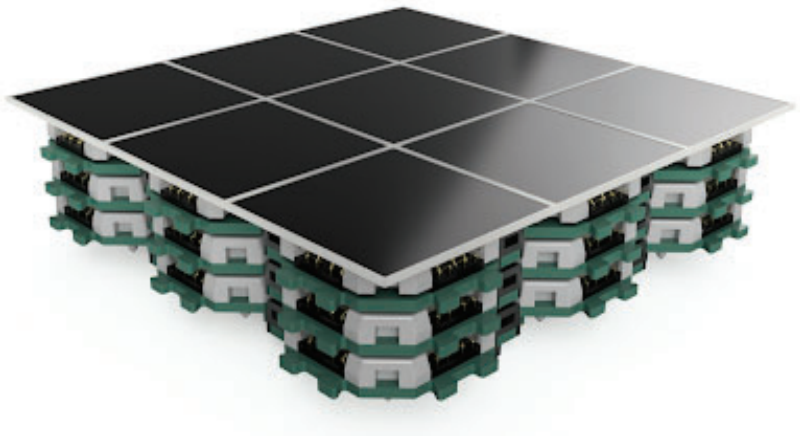


Figure 7-1: Concept rendering of an ideal Meso-DICE and integrated display packaging.

Next, work needs to be done on the transmission of data to the displays themselves. While the simple serial communication from the DICE nodes to the display's microcontroller seemed to do just fine for the testing conducted in this work, it is worth exploring how difficult yet beneficial it would be to have the same token passing scheme used between nodes put in place between the nodes and the displays. The handshaking and waiting functions would certainly help curb misread or unorganized data. The most important things that need to be maintained and improved, however,


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

is ease of use and speed. Whichever method can maximize both of those areas is the surefire winner to be adopted and maintained.

With those two areas improved, the system as a whole, especially in debugging and visualizations, will be lifted up. The visualizations will look much more appealing with displays seamlessly tiled and correctly placed above their respective node or node stack. By being able to just press in and power the displays in order to connect them and have them start displaying, the debugging process can be sped up, eliminating the current need for ensuring the header pin strut is in the correct location along with the jumper cables going to the correct pins. Soldered and secured connections are far more stable than a mess of jumper cables going between devices anyway, so all of this would sure to combine to lead to much more reliable performance.

The stability of the connections of the struts and nodes is something more to be explored in the overall DICE project. The connections held up extremely well for a given amount of time, but eventually started loosening up upon repetitive latching and unlatching. Many nodes lost one or more of their four heat-staked PLA posts holding the node together. Some struts integrated into the build plate also lost one side of the milled delrin latching piece. While not completely detrimental to operation, these breaks surely didn't make things more stable and secure.

One way that these structural issues could be mitigated is by having reprogramming capabilities implemented from node to node, and not just at the programming stand. By having a common backbone throughout the system, which partially exists in the build plate, the entire structure could be reprogrammed by sending the information to each node stack and having the nodes reprogram each other using the serial lines that connect them. This would make it so removing and replacing each component wouldn't be necessary, increasing the longevity of each of the components, not to mention the speed boost it would grant to development time.

As more time is dedicated to working through existing applications and improving them, the strengths of the system will become more apparent. The debugging capabilities should help any future developers work on different code and simulations within the system, as well as create appealing visualizations to make sense of the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

applications they're working on. While current applications may be limited, each of them can be adapted to implement the displays to improve upon them in some way.

The particle simulation in particular has a lot to offer for this system, and vice versa. Once code is developed to get the physical DICE nodes interacting with one another and sharing particle information, the entire visualization could be quickly realized. From there, 3D implementations, as well as more complex particle interactions simulating stress fractures and other physical phenomena can be easily introduced.

For the ray tracing application, it is quite the bummer that it didn't make it past the conceptual stages, so the obvious next step is to bring it from the theoretical to the experimental realm. That will take quite a bit of work, but as it builds upon the foundation of the particle simulation, propagating photons should be a compelling application of the capabilities of the system. From there, more complex scenes and objects can be successfully rendered.

Another application mentioned in earlier chapters included that of cross-sectional view rendering. For this application, the DICE nodes could represent a volume with an object inside. Depending on the layer of nodes selected, a cross-section view could be rendered of the object at that layer and displayed on the top surface displays. Different layers would represent different depths in the display volume. A concept of how this could look is included in Figure 7-2.

Some other incredibly intriguing ideas are volumetric and holographic display applications. Specialized display hardware would most certainly be required to make the most of these applications. However, the parallel processing provided by the DICE nodes is perfect for the complex and robust computations that must be performed for things like computational holography. Similar to the ray tracing application, it requires computing the paths of the object beam and reference beam as it propagates through an environment, then to determine the diffraction pattern caused by the interference of the beams. It would be quite something to see a hologram computed and displayed on the same system.

While performance projections show that the project has a long way to go to be directly compared to modern computers, the specific direction that it takes could


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

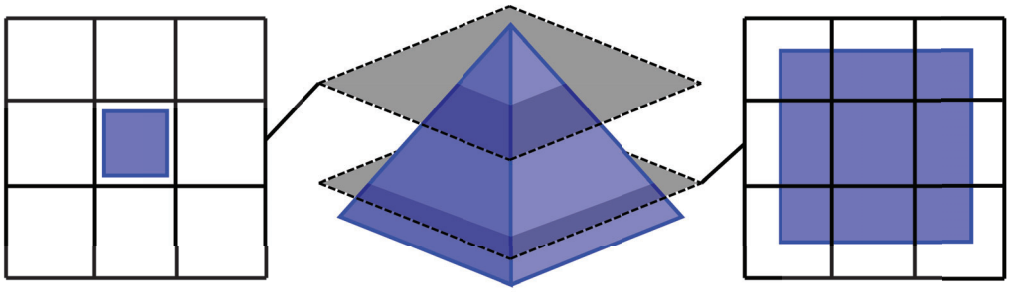


Figure 7-2: Concept of a cross-section view rendering application. The object in the middle would be included in the scene volume calculated by the display nodes. Depending on the layer selected, that layer of nodes would send data about the object from their layer to the displays, which would then render a cross-sectional view of the object. Different cross-sectional views can be seen represented in the display windows on the left and right.

potentially be among the first steps of a shift in computing architectures as a whole. Especially with the projections of the superconducting powered DICE nodes, the project could even be revolutionary. With the addition of the displays, it opens the doors to less conventional applications and implementations, but those should evolve with the project as a whole moving forward.

Again, it is anticipated that the contributions provided by this work are a great next step forward for all of the concepts explored. Much more research can and should be conducted to truly determine how impactful this system can be. As the system gets in the hands of more researchers and developers, hopefully its full potential can be realized. At the very least, this system is an interesting exploration of a unique computing and display architecture. However, if performance projections and other indicators are true, this could be another step towards a fundamental change in how the world looks at computing.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Bibliography


[1] Adafruit. "1.8" Color TFT LCD display with MicroSD Card Breakout - ST7735R." Accessed on: August 13, 2021. [Online]. Available: https://www.adafruit.com/product/358.

[2] Adafruit. "Adafruit Feather M4 Express - Featuring ATSAMD51 - ATSAMD51 Cortex M4." Accessed on: August 13, 2021. [Online]. Available: https://www.adafruit.com/product/3857.

[3] T. Akenine-Möller, E. Haines, and N. Hoffman. *Real-Time Rendering.* AK Peters/crc Press, 2019.

[4] J. Arvo. "Backward Ray Tracing." In *Developments in Ray Tracing, Computer Graphics, Proc. of ACM SIGGRAPH 86 Course Notes*, pp. 259-263, 1986.

[5] C. Blackburn. Concurrent master's thesis. Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences, 2021.

[6] W. J. Butera. "Programming a Paintable Computer." PhD dissertation, Massachusetts Institute of Technology, 2002. [Online]. Available: http://cba.mit.edu/docs/theses/02.02.butera.pdf.

[7] P. Christensen et al. "RenderMan: An Advanced Path Tracing Architecture for Movie Rendering." *ACM Transactions on Graphics (TOG) 37*, no. 3, pp. 1-21, 2018.

[8] R. L. Cook, T. Porter, and L. Carpenter. "Distributed ray tracing." In *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, pp. 137-145, 1984.

[9] D. G. Elliott, M. Stumm, W. M. Snelgrove, C. Cojocaru, and R. McKenzie. "Computational RAM: Implementing processors in memory." *IEEE Design & Test of Computers 16*, no. 1, pp. 32-41, 1999.

[10] A. J. Fewings and N. W. John. "Distributed Graphics Pipelines on the Grid." In *IEEE Distributed Systems Online*, vol. 8, no. 1, p. 1, Jan. 2007.

[11] Z. Fredin. Concurrent master's thesis. Massachusetts Institute of Technology, School of Architecture and Planning, Program in Media Arts and Sciences, 2021.





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[12] Z. Fredin, J. Zemanek, C. Blackburn, E. Strand, A. Abdel-Rahman, P. Rowles, and N. Gershenfeld. "Discrete Integrated Circuit Electronics (DICE)." In *2020 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, 2020.

[13] M. Gokhale, B. Holmes, and K. Iobst. "Processing in memory: The Terasys massively parallel PIM array." *Computer 28*, no. 4, pp. 23-31, 1995.

[14] G.R. Hoffman. "Who invented ray tracing?." *The Visual Computer, 6*, no. 3, pp. 120-124, 1990.

[15] G. Humphreys, I. Buck, M. Eldridge, and P. Hanrahan. "Distributed Rendering for Scalable Displays." In *SC '00: Proceedings of the 2000 ACM/IEEE Conference on Supercomputing*, p. 30. IEEE, 2000.

[16] G. Humphreys and P. Hanrahan. "A distributed graphics system for large tiled displays." In *Proceedings Visualization '99 (Cat. No.99CB37067)*, pp. 215-527. IEEE, 1999.

[17] D. R. Hutchings, J. Stasko, and M. Czerwinski. "Distributed Display Environments." In *CHI'05 Extended Abstracts on Human Factors in Computing Systems*, pp. 2117-2118, 2005.

[18] H.W. Jensen and P. Christensen. "High Quality Rendering using Ray Tracing and Photon Mapping." In *ACM SIGGRAPH 2007 courses*, pp. 1-es., 2007.

[19] K. Li et al. "Building and using a scalable display wall system." In *IEEE Computer Graphics Applications*, vol. 20, no. 4, pp. 29-37, July-Aug. 2000.

[20] N. Margolus. "CAM-8: a computer architecture based on cellular automata." 1993. [Online]. Available: https://people.csail.mit.edu/nhm/cam8.pdf.

[21] B. F. Mistree and J. A. Paradiso. "ChainMail: A Configurable Multimodal Lining to Enable Sensate Surfaces and Interactive Objects." In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, pp. 65-72, 2010. [Online]. Available: https://resenv.media.mit.edu/pubs/papers/2010-01-fp245-mistree.pdf.

[22] S. Molnar, M. Cox, D. Ellsworth, and H. Fuchs. "A Sorting Classification of Parallel Rendering." *IEEE computer graphics and applications 14*, no. 4, pp. 23-32, 1994.

[23] T. L. Sterling and H. P. Zima. "Gilgamesh: A Multithreaded Processor-In-Memory Architecture for Petaflops Computing." In *SC'02: Proceedings of the 2002 ACM/IEEE Conference on Supercomputing*, pp. 48. IEEE, 2002.

[24] E. Strand. "Inverse Methods for Design and Simulation with Particle Systems." Master's thesis, Massachusetts Institute of Technology, 2020. [Online]. Available: http://cba.mit.edu/docs/theses/20.09.strand.pdf.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[25] T. N. Theis and H.-S. P. Wong. "The End of Moore's Law: A New Beginning for Information Technology." In *Computing in Science & Engineering*, vol. 19, no. 2, pp. 41-50, Mar.-Apr. 2017.

[26] I. Wicaksono, J. Cherston, and J. A. Paradiso. "Electronic Textile Gaia: Ubiquitous Computational Substrates Across Geometric Scales." *IEEE Pervasive Computing*, 2021. [Online]. Available: https://resenv.media.mit.edu/pubs/papers/2021-Wicaksono-Cherston-EtextileGaia.pdf.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Discrete Robotic Construction

by

William Kai Langford

B.S., Tufts University (2012)
M.S., Massachusetts Institute of Technology (2014)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2019

© Massachusetts Institute of Technology 2019. All rights reserved.

Author .....
Program in Media Arts and Sciences,
School of Architecture and Planning
August 9, 2019

Certified by .....
Neil Gershenfeld
Director, MIT Center for Bits and Atoms
Thesis Supervisor

Accepted by .....
Tod Machover
Academic Head, Program in Media Arts and Sciences



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Discrete Robotic Construction**

by

William Kai Langford

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
on August 9, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

## Abstract

Robots, which require the integration of a wide variety of mechanical and electrical functionality, are seldom built in a single process, but are instead assembled from parts created using a variety of different processes. While fabrication has advanced significantly to enable the routine fabrication of complex and precise objects from computer designs, the assembly processes used to integrate these parts are still largely manual and are notoriously difficult to automate. Recent research in digital fabrication has looked for ways to avoid assembly altogether by manufacturing integrated devices in a single process but has often struggled to integrate more than a few materials or functionalities. Instead of avoiding assembly, this work embraces it. Inspired by the universality of amino acids that are the basis of molecular biology, I demonstrate an interchangeable set of building blocks that enable the construction of a wide variety of robotic capabilities, including machines that can assemble themselves. In this thesis I introduce a discrete approach to robotic construction that enables the fabrication of structure, mechanism, actuation, circuitry, and computation in a single process through the assembly of a small set of building blocks. This work is based on discretely assembled "digital" materials, in which parts are reversibly joined with a discrete set of relative positions and orientations, allowing for global geometries to be determined from local constraints, assembly errors to be detected and corrected, heterogeneous materials to be joined, and disassembly and reuse rather than disposal. This approach simplifies the fabrication of integrated electromechanical machines and points to the possibility of building technology that is able to grow (exponential self-assembly) and self-repair. Furthermore, this approach discretizes robotic systems at a finer granularity than prior work in modular robotics, offering benefits including the flexibility to integrate heterogeneous functions, agility to rapidly construct and modify designs, and incremental extensibility in both system size and performance. These benefits help lower barriers in the rapid prototyping of electromechanical machines, make designs more reusable by providing a physical representation that facilitates design automation and abstraction, and enable machines that are more integrated than would be practical with alternative methods.

Thesis Supervisor: Neil Gershenfeld
Title: Director, MIT Center for Bits and Atoms



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Discrete Robotic Construction**

by

William Kai Langford

This dissertation/thesis has been reviewed and approved by the following committee members

Sangbae Kim .....

Associate Professor of Mechanical Engineering

Massachusetts Institute of Technology



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Discrete Robotic Construction**

by

William Kai Langford

This dissertation/thesis has been reviewed and approved by the following committee members

Jennifer Lewis . . . . . .

Hansjorg Wyss Professor of Biologically Inspired Engineering

Harvard University



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## Acknowledgements

I'd like to acknowledge of the people who helped and supported me in this work. I wouldn't have been able to do it without you all.

Thank you to my committee members – Sangbae Kim and Jennifer Lewis – for taking time out of your very busy schedules to meet with me and to help me think through outcomes of this work beyond CBA's research roadmap.

I'd like to thank my advisor, Neil Gershenfeld. I certainly could not have done this work anywhere else. You're unwavering vision for the work we do is inspiring. Thank you for showing me a new way to look at the world. I'm certain that I will never again have access to quite the same level of resources again. Being able to design something in the morning, go to the shop in the afternoon, run jobs on the wire-EDM, precision CNC mill, and multiple 3D printers (all simultaneously), and then have a first prototype by dinner is a very special kind of power that I know I'll miss.

However, the tools are not nearly CBA's greatest feature. It is it's people. In the seven years that I've been here I've made great friends and had the opportunity to work with amazing collaborators.

I first want to thank the CBA staff – Joe, James, Kara, Jamie, John, Tom, and Ryan – not just for their support in managing projects, putting in orders, and doing all the behind-the-scenes work that makes our research work possible, but also for getting lunch, hanging out, and creating a sense of community.

I'd like to thank all of the CBA students – Sam, Prashant, Ben, Amanda, Grace, Jake, Sara, Erik, Amira, Filippos, Patricia, Noah, James, Nadya, Eric, Matt C., Matt K., and Kenny. I've learned so much just from being around you all for these years. Sam, in particular, thanks for our coffee-talks and whiteboard sessions.

I'd like to thank my friends and collaborators outside of CBA – Carlos, Jifei, Lining, Guillermo, and Sophia. In particular, Carlos, thank you for always making sure my priorities are straight with our weekly call (even when I'm "too stressed").

Thank you Sherry and Neil for taking me on Fablab trips all over the world. It's an amazing privilege to be able to travel to some of the coolest places in the world to help teach people about digital fabrication.

Thank you to the CBA UROPs – Uyen, Soma, Jason, Steven – for helping me experiment with ideas and being guinea pigs for my developing leadership and mentoring skills.

Thank you to my family – Mom, Dad, Ben, Grandma, and Oscar – for making me who I am today.

Thank you Iris for always believing in me and encouraging me to be my best self.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Contents


**1 Introduction** **15**

1.1 Motivation . . . . . 15

1.2 Background . . . . . 18

1.2.1 On-Demand Fabrication of Robotic Systems . . . . . 18

1.2.2 Self Replicating Machines . . . . . 20

1.2.3 Biology . . . . . 21

1.3 Approach . . . . . 24

1.3.1 Implications . . . . . 24

1.3.2 Goals and Metrics . . . . . 25

**2 Joints** **27**

2.1 Design Guidelines . . . . . 27

2.2 Interlocking Press-Fit Slots . . . . . 28

2.2.1 Electrical Characterization . . . . . 29

2.3 Alternative Joint Systems . . . . . 31

**3 Structure** **33**

3.1 Lattice Structures . . . . . 33

3.2 Digital Materials . . . . . 34

3.3 A New Geometry . . . . . 35

3.4 Sizing . . . . . 36

3.5 Production . . . . . 37

**4 Mechanism** **39**

4.1 Compliant Mechanisms . . . . . 39

4.2 Part Types . . . . . 40

4.3 Modeling . . . . . 40

4.3.1 Kinematics . . . . . 40

4.3.2 Quantization of a continuous range of transmission ratios . . . . . 42

4.3.3 Dynamics . . . . . 44

4.4 Production Methods . . . . . 45

4.5 Discussion . . . . . 47






DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**5 Actuation** **51**
5.1 Actuation at the millimeter scale . . . 51
5.2 Electromagnetic Lorentz Force Actuators . . . 53
5.2.1 Modeling . . . 54
5.2.2 Experimental Characterization . . . 57
5.2.3 Production . . . 60
5.2.4 Discussion . . . 61

**6 Walking Motor** **63**
6.1 Design . . . 63
6.2 Characterization . . . 63
6.2.1 Speed and Repeatability . . . 64
6.2.2 Force Additivity . . . 66
6.3 Discussion . . . 67

**7 Walkers** **69**
7.1 Electropermanent Feet . . . 69
7.2 MOTILE – Modular Tiny Locomotion Element . . . 71
7.3 Discussion . . . 73

**8 Positioning and Manipulation** **75**
8.1 Positioning . . . 75
8.2 Manipulation . . . 75
8.3 Discussion . . . 76

**9 Circuitry** **79**
9.1 Background . . . 79
9.2 Routing . . . 80
9.3 Logic . . . 82
9.4 Control . . . 84
9.4.1 Control in struts . . . 84
9.4.2 Control between struts . . . 84
9.4.3 Redefining struts for greater routing density . . . 86
9.4.4 Tiles – Decoupling mechanical and electrical routing . . . 87
9.5 Production Methods . . . 88
9.6 Discussion . . . 91

**10 Automation** **93**
10.1 Electronic Digital Material Stapler Assembler . . . 93
10.2 Design Overview . . . 93
10.2.1 Detailed Design . . . 94
10.2.2 Workflow and Characterization . . . 94
10.3 FABBR – Functional Assembler of Building Block Robots . . . 96
10.3.1 Design Overview . . . 96
10.3.2 Detailed Design . . . 97


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

10.3.3 Characterization . . . . . 99
10.3.4 Extensions . . . . . 101
10.4 Assembling Assemblers . . . . . 102
10.4.1 Current Limitations . . . . . 105

**11 Scaling** **107**
11.1 Part Production . . . . . 107
11.2 Actuation . . . . . 108
11.3 Assembly . . . . . 110

**12 Modeling** **113**
12.1 Modeling Approaches . . . . . 113
12.2 Optimization Approaches . . . . . 114
12.3 Compliant Mechanism Simulation and Optimization . . . . . 114

**13 Applications and Extensions** **117**
13.1 Discrete Integrated Circuit Electronics (DICE) . . . . . 117
13.2 Relaxing Constraints . . . . . 119

**14 Conclusion** **123**

**A Lessons Learned** **125**



304
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 1

# Introduction

## 1.1 Motivation

The design and fabrication of robots today can be expensive, time consuming, and inflexible. It often involves the integration of a wide variety of unique parts, each of which may be made using separate and unrelated processes. This assembly and integration step often represents the bottleneck in the time and flexibility in the development of a new robotic system [1] [2]. Achieving high performance often means designing from the ground up and fabricating a bespoke, application specific robot.

Researchers have looked at modularity as a way to increase the flexibility in the configuration of new robotic systems. By abstracting robotic capabilities to a small number of unique modules, a new robotic system can be configured and reconfigured to suit a particular task through the (relatively) simple arrangement of the modular parts. These modular parts, however, often end up needing to integrate actuation, sensing, communication, computation, and structure, and do so in a physically small volume. This results in modules that are expensive and difficult to manufacture, and ultimately limits the performance of the resulting robots made this way [3]. In doing so, the difficulty to develop a new robotic system has been reduced but at the expense of a simultaneous reduction in performance (Figure 1-1).

As a result, recent research in digital fabrication has looked to avoid the assembly and integration step altogether by fabricating devices in a single monolithic process. This includes processes that assemble intricate mechanisms through the lamination of various pre-machined layers [8], the integration of hydraulic channels in additively manufactured robot limbs [9], and even the direct printing of entirely soft robots [10]. However, each of these processes still has inherent limitations including the kinds of materials it can work with, the dimensionality of the end-product (2D vs 2.5D vs 3D), and the dynamic range it can support (maximum size per minimum feature size). As a result, I argue that no one process will be able to span the full range of desired robotic capabilities.

Biology offers an alternative inspiration. Rather than trying to integrate all of the desired functionality in a common process, it builds from a standardized set of building blocks that share a common interface. All of life is assembled from 20 standardized building blocks, amino acids (Figure 1-2). These building blocks are assembled one part at a time by a machine, the ribosome, which itself is made from the same building blocks. These biological assemblies are truly complex and integrated. They sense, actuate, move, flex, and act as structure and do so at a resolution





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

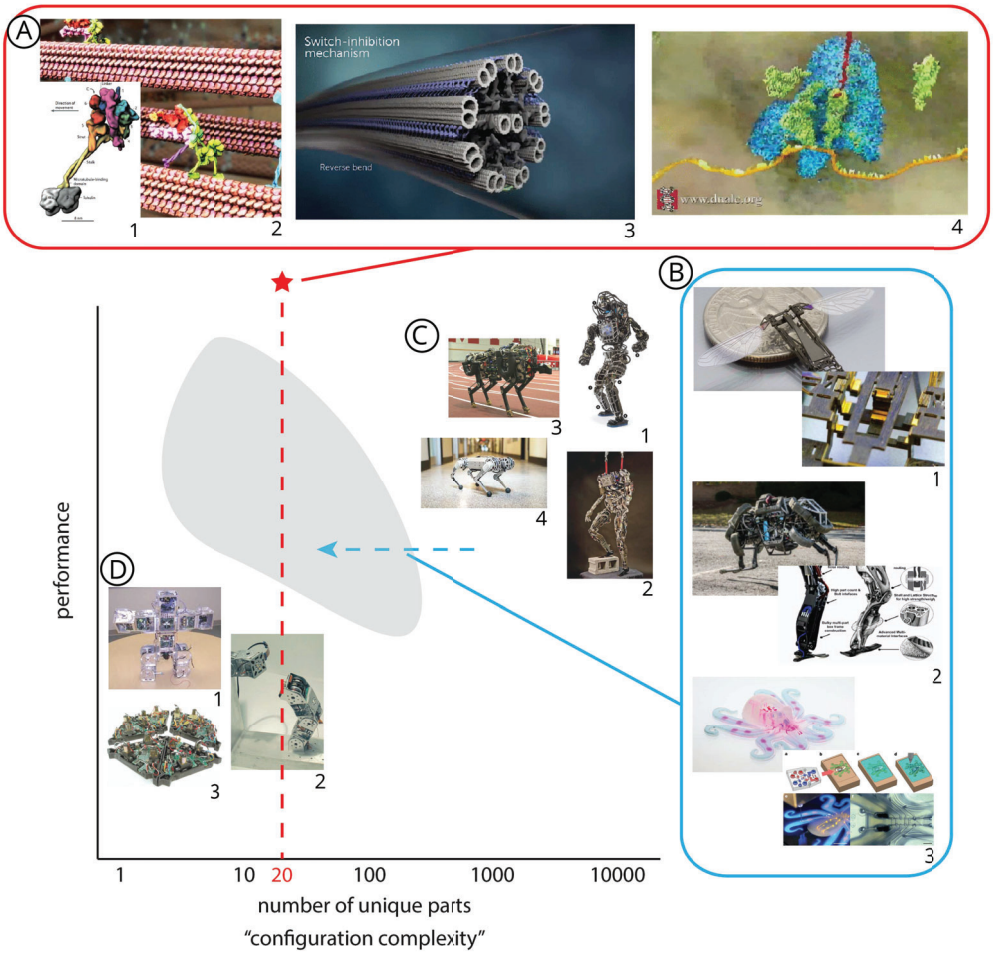



Figure 1-1: The robot design and fabrication space in terms of "configuration complexity" (number of unique parts) vs performance. Highly performant robots (C) are often bespoke and require ground-up design and development (1-4) [4] [5] [6] [7]. Modular robots (D) tend to tradeoff performance for versatility (1-3) [3]. Digital fabrication research (B) has looked for ways to avoid assembly by integrating functionality in monolithic processes (1-3) [8] [9] [10]. Biology (A) offers an inspiring alternative as it assembles all of life from (essentially) 20 building block parts (amino acids) (1-4) [11] [12] [13] [14].


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

and performance that far exceeds what we are currently capable of producing with human-made technology [15] [16].

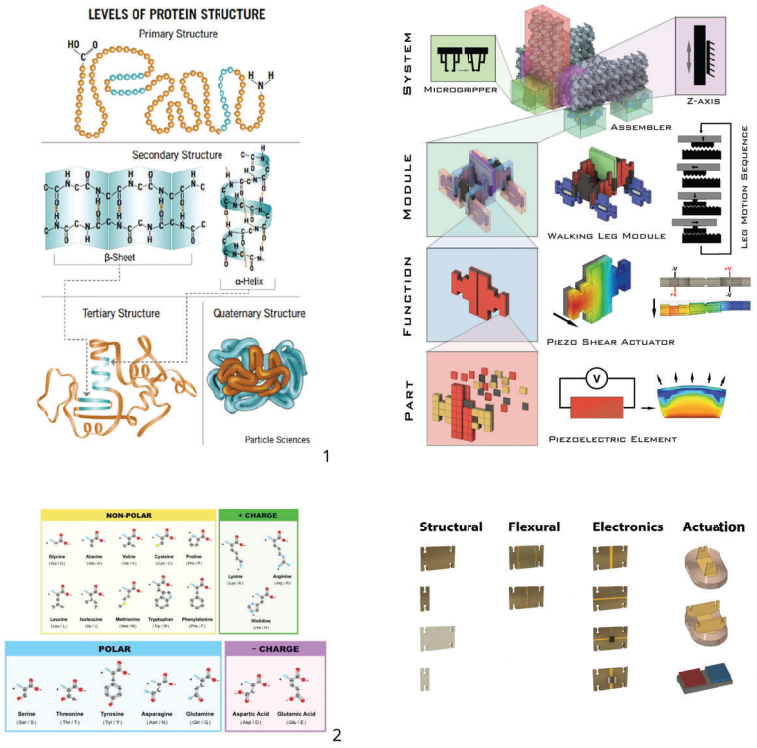



Figure 1-2: The twenty amino acids that enable the assembly of biological systems do so through levels of hierarchy: primary, secondary, tertiary, and quaternary structure. The "engineering amino acids" should likely exhibit a similar structural hierarchy. Individual bulk-material building blocks assemble functional parts, functional parts assemble into modules, and modules then assemble into systems (1-2) [16] [17].

This leads to the question that prompted the work developed in this thesis: **Can we develop a set of simple, functional parts that enable the assembly of a wide range of integrated machines?** In this thesis I introduce a discrete approach to robotic construction that enables the fabrication of structure, mechanism, actuation, circuitry, and computation in a single process through the assembly of a small set of building blocks. This approach takes inspiration from the amino acid building blocks that are the basis of molecular biology and offers benefits in the integration of heterogeneous functionality, agility in creating and modifying designs, and points towards the capability for technology to grow (self-assemble) and self-repair. Rather than avoiding assembly in pursuit of integrated manufacturing, this approach embraces it, and demonstrates that the interchangeability of building blocks enables the assembly of a wide variety of capabilities from a small set of part types.

The approach I introduce is based on prior work in discretely assembled "digital" material systems. Digital materials are based on a discrete set of parts, which are reversibly joined with a discrete set


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

of relative positions and orientations [18]. These properties allow global geometries to be determined from local constraints, assembly errors to be detected and corrected, heterogeneous materials to be joined, and disassembly and reuse rather than disposal [19]. Digital materials have been used to produce the highest reported modulus ultralight materials [20], shape morphing structures with the use of rigid and flexural parts [21], as well as electronic structures with the addition of conductive and insulating parts [22] [23].

This approach builds systems with integrated structure, mechanisms, and actuation in a way that can be incrementally extended and modified. Standardizing the assembly interfaces between parts and simplifying the assembly process, to require just a single vertical motion, means the assembly process can more closely resemble a digital fabrication workflow and represents a significant step towards enabling on-demand fabrication of a wide-range of robotic systems.

## 1.2 Background

The approach introduced in this thesis works towards two active goals in the areas of digital fabrication and modular robotics: (1) on-demand fabrication of robotic systems and (2) self-replicating machines. Additionally, this approach takes inspiration from the way that biological system manufacture across scales (grow).

### 1.2.1 On-Demand Fabrication of Robotic Systems

The goal of simplifying the manufacturing of highly integrated machines and robots to enable on-demand fabrication of robotic systems has been explored from a number of different perspectives. These include folded and origami-based manufacturing methods, additive manufacturing and soft-robotics, as well as digital material assembly.

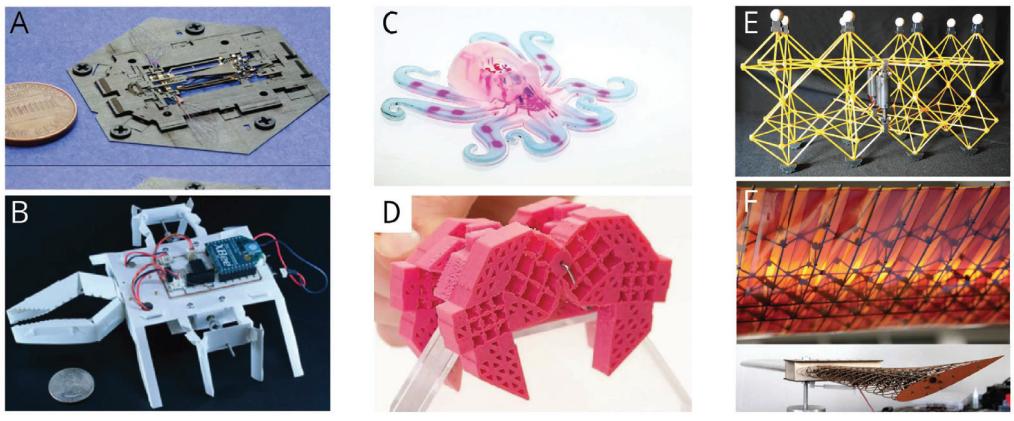



Figure 1-3: The goal of enabling on-demand robot fabrication has been approached by ways including: folding-based methods (A,B) [8] [24], soft robotics (C,D) [10] [25], and digital material (E,F) [26] [21].



308
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### Folding-Based Processes

Folding based processes look for ways to pattern functionality in a two-dimensional sheet and then fold to create a three-dimensional structure (Figure 1-3 A,B). These kinds of processes are well suited for on-demand robot fabrication because patterning and structuring two-dimensional sheets can be done quickly, precisely, and at a high resolution. These processes are often able to incorporate multiple materials to achieve various kinds of functionality within a single device. Whitney et al. demonstrated the assembly of complex mechanical linkages through the lamination of rigid and flexible layers [8]. Furthermore, an assembly scaffold, patterned along with the robot, was then used to coordinate the folding of the two-dimensionally patterned sheets into a three-dimensional robot, the Robobee. This approach has enabled the batch production of devices with a broad material set and a design flexibility that was previously inaccessible in millimeter-scale robotics. While these robots are still typically designed by hand, others have looked at the possibility of automating the design of folded robots [27]. The work of Onal et al. demonstrates how a library of design templates with parametric dimensions may be used to lower the barriers in the design and development of these robots [24]. Furthermore, these machines show the extent to which a single laser-cut shape can define complex three-dimensional body shapes and mechanisms.

While these folding processes are often able to incorporate both mechanical linkages and circuitry in a single process, actuation is typically integrated in a separate step. Onal et al., for example, attach off the shelf DC motors to the robot after it has been folded. In millimeter-scale laminate devices actuation often takes the form of a piezoelectric actuator that is made using a separate laminate process and assembled manually [28]. As these devices become more complex and integrate more functionality, the number of layers and process steps increases. Given the monolithic nature of these devices, an error in any single layer can ruin a whole device. As such, it is common for the designer to invest significant time upfront in the design of the device to ensure there are no errors, limiting the resulting turn-time and agility of the fabrication approach [27]. As a result, while monolithic designs could ultimately enable on-demand robot fabrication, practically, it often makes more sense to split a complex design into subsystems which are laminated and then integrated separately [28].

### Additive Manufacturing and Soft Robotics

Recent efforts to integrate actuation into 3D printing workflows have seen promise in the direct printing of soft robots and actuators (Figure 1-3 C,D). By structuring geometry, hydraulic or pneumatic forces can be converted into motion [29] [30]. These processes are able to produce devices with useful forces and motions in a nearly monolithic process. While these robots have been traditionally made with molding processes, more recent research has demonstrated the possibility of directly printing these kinds of soft robots and machines [10] [25]. These soft machines have a range of useful applications and are often applicable in scenarios that are not well served by traditional "hard" robots.

While these processes greatly extend the capabilities of robotic fabrication, they are still limited in the kinds of materials and functionalities that they can integrate. The material set is often limited by the requirements of the deposition processes, which constrains the range of functionality and, as a result, the potential applications. Furthermore, the scarcity and properties of compatible soft-materials limits the available functionality below the millimeter scale [31] [32].


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Other research has looked at ways to directly print hydraulic actuators for use in high-performance robots such as Boston Dynamic's Atlas robot [9]. These actuators are able to unify what used to be several separate components. But beyond hydraulic actuators, this process doesn't likely scale to include other kinds of functionality like electronics or large displacement compliant joints.

#### Digital Materials

Discretely assembled "digital" material systems are another avenue that has been explored towards the goal of enabling on-demand robot fabrication. Digital materials are a discrete set of parts, which are reversibly joined with a discrete set of relative positions and orientations [18]. These properties allow global geometries to be determined from local constraints, assembly errors to be detected and corrected, heterogeneous materials to be joined, and disassembly and reuse rather than disposal. The building block parts that compose a digital material can be produced using a range of different processes and in a range of different materials. This enables the integration of diverse functionalities including structure [20], mechanisms [21], and electronics [22] [23] within a single assembly framework.

The prospect of using digital materials to enable on-demand robot fabrication has been explored in theory [33] [34] but physical implementations have largely relied on either slow global thermal actuation [33] or on off-the-shelf actuation components [23]. In a more recent study, Cramer et al. demonstrated very simple locomoting robots made through the actuation of deformation modes of a discretely assembled lattice using commercially available linear servo motors [26] (Figure 1-3 E).

### 1.2.2 Self Replicating Machines

The idea that modular building blocks can be used to assemble self-replicating systems is not a new one and has been explored computationally since at least the 1940's. At that time Von Neumann developed the concept of a universal constructor, built around a set of cellular automata rules, that can construct itself and pass on a blueprint for further self-replication [35]. For a number of reasons, Von Neumann's concept is not particularly physically realistic. For example, blocks are not translated but rather destroyed at one location and created at another. However, others have built on Von Neumann's original idea and developed versions of the universal constructor that are more physically-based [36]. These ideas are still very far from a physical embodiment but demonstrate that a self-replicating machine is kinematically feasible. Researchers as recently as 2016 have written papers entitled, "Are Self-Replicating Machines Feasible?" [37], clearly showing the disparity between theory and practice. This more recent research still concludes that self-replicating machines are feasible, but notes that there are a number of practical barriers that must be overcome first.

#### Modular Robotics

Work in the field of modular and reconfigurable robotics has looked to address the challenges of self-replication by simplifying the feedstock to a set a of modular building blocks [42] [43] [40]. These systems typically integrate actuation, controls, communications, and structural connections within modular units, which are then configured (and reconfigured) to suit a particular task. While these systems illustrate the universality of a small number of modular building blocks, the resulting modules are relatively complex, involve dense integration of the various functions, and tend to be


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

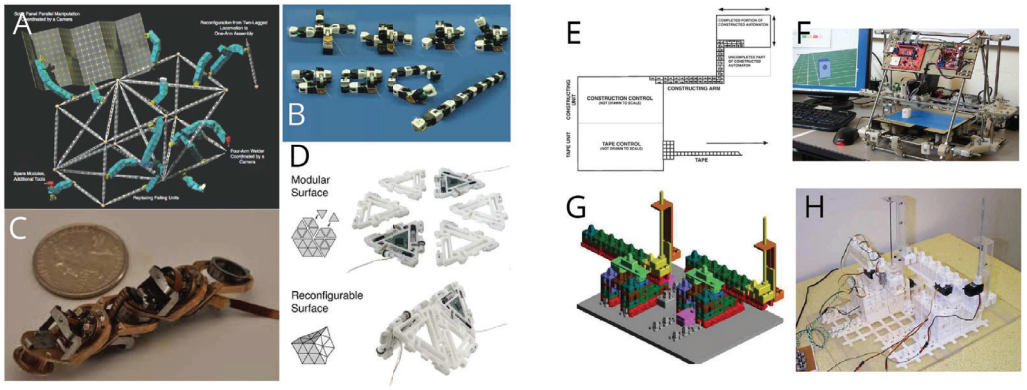



Figure 1-4: Modular robotic systems are configurable and capable of adapting to a variety of different applications. They may be used as manipulators to assemble truss structures [38]. Modular chain robots [39] [40] and surfaces [41] can be contorted and twisted to form a variety of shapes and structures.

expensive to fabricate in volume. As a result, they have typically found limited utility outside of the research lab.

#### In-Situ Resource Utilization (ISRU) Approaches

These self replicating machines are being pursued for a number of different applications, one of which is space exploration and development. The development of in-space manufacturing capabilities and modular robotics architectures are important steps towards self-replicating systems. Research in self-replicating hardware is in early stages and most prior work addresses high-level system design challenges including hierarchy and scale [44] as well as required ISRU capabilities and launch requirements for self-sustaining factories [45]. Other prior work has looked at lower-level implementation details and established methods of creating self-replicating systems with macroscale building blocks. Moses et al. built a universal constructor system, for example, using 18 part-types that is capable of assembling some of its own subsystems [46]. However, the parts have a high-degree of embedded complexity, including conventional actuators and processors, which limits the potential scalability of such a system.

### 1.2.3 Biology

#### Ribosomal Assembly (Growth)

Ribosomal assembly, or growth, is able to achieve a wide dynamic range between the size of the parts and the size of the assemblies because of error correcting processes during part assembly and because of its self-replicating machinery that enables an exponential ring-up in manufacturing capacity (ribosomes assemble ribosomes).

Amino acids are often referred to as the building blocks of life [15] and make up, when assembled as proteins, approximately 60% of the dry mass of mammalian cells. They are made up of a common backbone, which allow them to bond to form a string, and a side chain, which confers simple chemical, structural, or functional properties to the part such as hydrophobic/hydrophilic or basic/acidic.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

These twenty standard L-$\alpha$ amino acids, each with unremarkable properties on their own, are the building blocks that enable the assembly of all biological machinery, including a machine that can make itself, the ribosome. Ribosomes assemble amino acids at relatively unremarkable rates, ranging from 0.6 Hz to 20 Hz [47], but each cell may contain upwards of 72,000 ribosomes (in fast growing E. Coli), representing one third of the cell's dry mass and a volumetric density of approximately 50,000 per cubic micrometer [15]. The tens of trillions ($10^{13}$) of cells in the human body, each containing, on average, a few billion proteins ($10^9$), which are themselves made up of a couple hundred amino acids ($10^2$), means we are made up of approximately $10^{26}$ parts [15]. In comparison, the largest dynamic range (minimum feature size to build area) of our most advanced manufacturing techniques is only about $10^{13}$ (Nanoscribe [48]). This kind of complexity is made possible in reasonable time-scales not because ribosomes are fast (they assemble at approximately 1 Hz), but because ribosomes can assemble ribosomes, enabling massively parallel assembly.

Beyond the ability of ribosomes to assemble ribosomes, another critical aspect in being able to span large dynamic ranges in a single process is error correction. Before the mRNA instruction set reaches the ribosome, it has already gone through multiple stages of error correction and proofreading through the transcription process. During and after translation of the mRNA into the peptide chain, another set of error correcting steps is taken to discard, degrade, and recycle incorrectly translated proteins. Transcription has been reported to have error rates on the order of $10^{-5}$ while translation error rates are an order of magnitude more, or $10^{-4}$ [49] [15] [50].

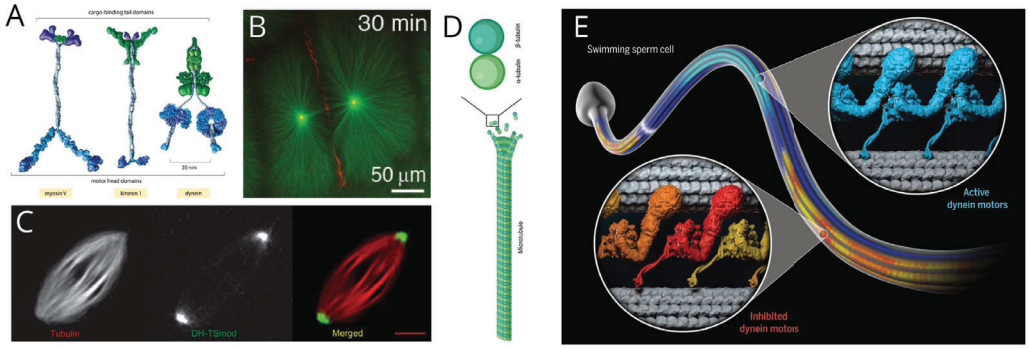



Figure 1-5: Biological machines exhibit modularity and reusability. The same motor proteins (A) [15] and microtubules (D) [51] that help organize the cell (B,C) [52] [53] are also used to actuate flagella mechanisms of swimming sperm (E) [13]

.

Amino acids enable the assembly of highly capable machines and behaviors. The proteins assembled from amino acids are present in a wide range of sizes but have an average length of approximately 400 amino acids in eukaryotic cells [15]. As the protein's primary structure is being formed through translation of mRNA by ribosomes, it folds into a three-dimensional structure based on hydrophobic/hydrophilic interactions, molecular chaperones, and environmental conditions such as pH. The structure of the protein exhibits hierarchy and its functionality and geometric motifs can be grouped into primary, secondary, tertiary, and quaternary levels [16]. These proteins are then used as building blocks for a wide variety of cellular mechanisms and behaviors including structure, elasticity, transport, and signaling [54]. Furthermore, as a result of assembling all of life from the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

same twenty building blocks, biological assemblies end up having a great deal of modularity and reusability across a number of different applications. The same kinds of microtubules and motor proteins that organize cells and give them their structure are also used to actuate flagella to propel swimming organisms like sperm [13] (Figure 1-5).

### Manufacturing Across Scales

Treating ribosomal assembly like a manufacturing process, it is possible to compare our existing manufacturing methods and see that all fall short of the precision, resolution, and complexity of biological assembly. In Figure 1-6 I present estimates of the resolution and throughput of ribosomal assembly in biological systems. These estimates should be viewed as establishing orders of magnitude rather than precise measured quantities. For each biological system I estimate the number of amino acids that compose all of the proteins that make up that specimen. For the three mammals presented (Mouse, Human, Blue Whale), I estimate amino acid content based on a protein fraction of their dry mass (60%) and assume a dry mass fraction (30%) of their total mass. I do the same for C. Elegans noting that the protein fraction is closer to 55% of their dry mass. HeLa cell, E. Coli, and Eukaryotic protein estimates are based on protein counts reported in literature combined with average protein length in Eukaryotes (400aa) [15]. In all of the above I assume that growth rate is linear with time, which will both overestimate and underestimate growth at certain development stages but therefore represents a reasonable average.

I do the same for manufacturing techniques including discrete assembly, additive manufacturing, and state of the art conventional assembly, estimating the effective number of resolvable elements and the throughput at which those elements can be assembled.

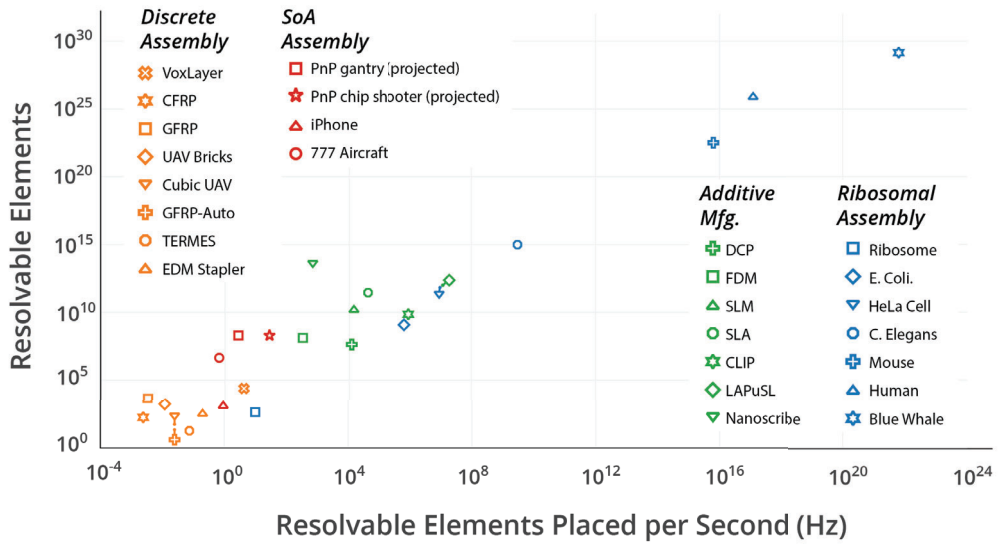



Figure 1-6: Manufacturing throughput vs dynamic range for across manufacturing processes. Biology's throughput scales with the size of the manufacturing problem and enables the fabrication of organisms with far more fabrication complexity than any state-of-the-art technological process.

From the plot, it's clear that the state of the art in digital fabrication roughly matches the





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

throughput and complexity of a human cell. Beyond that, there is no technological fabrication process that can extend into the space that biology occupies. We (humans) are made from on the order of $10^{26}$ amino acid parts. While our fabrication capabilities let us span from tens of nanometer silicon chips to hundreds of meter long aircraft, we do so in discontinuous processes with no shared representation. And it is the discontinuities in this approach that I argue largely limit the flexibility of manufacturing.

## 1.3 Approach

This thesis attempts to address the question: **can we develop a set of simple functional parts that enable the assembly of a wide variety of integrated machines?**

The hypothesis is that by standardizing the interfaces between parts, the assembly process can be treated like a digital fabrication process. Furthermore, by discretizing functionality at such a fine level, we can enable much tighter integration and coupling between functionality that is typically separate in conventionally made robotics. This merging of functionality represents a unified approach to robotic construction that could also help lower the barrier to things that are not yet possible, like machine self-replication.

In this thesis I focus on the development of capabilities necessary to build assemblers that can assemble assemblers. This application serves both as a case study for integrating various robotic functionalities within a machine, and represents the direction toward massively-parallel high-throughput assembly. In order to build assemblers that can assemble assemblers there are a number of capabilities that are needed and steps to achieve. The assembler will need to integrate structure, mechanism, actuation, circuitry, and computation. The following chapters will describe the design of the building block parts and how their arrangement enables the assembly of these various capabilities.

### 1.3.1 Implications

This work impacts the way in which robotic systems are designed and fabricated today and helps pave the way towards future goals such as high-resolution on-demand fabrication of robotic systems and self-replicating machinery. Specific implications of this work include:

- **Treating assembly like a digital fabrication process.** This work shows that by standardizing assembly interfaces, the assembly and configuration of functional systems can be treated like a digital fabrication process, enabling on-demand fabrication of robotic systems.

- **A unified approach to robotic construction enables a tighter coupling between structure, mechanism, actuation, circuitry, and computation.** By discretizing at a fine-grained functional level, capabilities that are typically distinct can be merged much more densely in this approach. Circuits can be routed through structural volumes, actuated degrees of freedom can be distributed rather than centralized, and computation and control elements can be spatially distributed closer to the components they drive.

- **A new way of digitally fabricating integrated machines and robots with a unmatched range of functionality.** Rather than trying to integrate multiple functionalities in


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

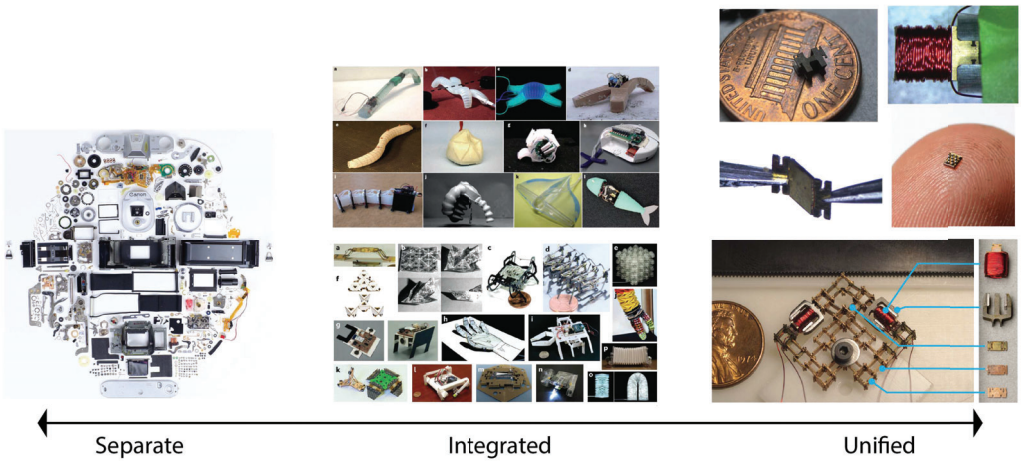



Figure 1-7: The approach explored in this work represents a unified approach to robotic construction. Conventional machines integrate diverse functionalities without any standardization and are often bottlenecked by boundaries between functionality. Rather than integrating the fabrication of functionality within a single process [55] [56], this approach enables the assembly and integration of wide variety of parts, each of which can be made using separate and dedicated processes. "Separate" image from [57].

a single process, this approach is able to achieve an unmatched range of integrated functionality through the integration of parts made in a wide range of different processes.

- **A viable way to close the fabrication complexity gap.** This approach helps lower barriers to assembling assemblers and provides a viable way to increase the resolution and complexity of technological systems through recursive assembly.

### 1.3.2 Goals and Metrics

This work is guided by the following metrics:

**Complexity** The aim of this work is to enable the assembly of relatively complex devices from relatively simple parts. For this assembly approach to be interesting, the dynamic range between the complexity of the parts and the whole should be larger than existing work in modular robotics and much closer to the dynamic range (minimum feature size to work envelope) of an additive manufacturing process.

**Universality** An additional aim of this work is to show how a small number of building blocks can be combined to form a large variety of assemblies. This metric discourages special-purpose parts and emphasizes finding part types that play a number of useful roles in a system.

**Performance** Finally, the assemblies fabricated using this process should be useful and competitive with existing techniques for on-demand robot fabrication. Power density, specific stiffness, and fabrication time are criteria that are particularly relevant to this work.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

In the following chapters I will discuss the components of this assembly framework: starting with and individual structural joint, and demonstrating the assembly of structure, mechanism, and actuation. Following that, I will demonstrate the assembly of machines that integrate these capabilities. From there, I will discuss the how circuitry and computation can be added to these assemblies, look at the automation and scaling of this work, and finally discuss future applications and promising results that are being explored.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 2

# Joints

## 2.1 Design Guidelines

One of the most fundamental aspects of this work is the identification of interlocking joint systems that can be used to reliably and robustly join parts in this assembly framework. This is also one of the most challenging aspects of this work: finding functionally compelling joints that are simple and small enough to "disappear" in the structure. The assembly framework presented in this work does not presuppose any particular joint type but does seek a number of desirable traits:

- **The parts should interlock mechanically.** One of the fundamental takeaways in [58] was the observation that "To build strong structure from bricks, they should be interlocked. This gives the structure the strength of covalent bonding in at least some directions."

- **The joints should function both as a mechanical connection and electrical connection.** This work hinges on the idea that parts can serve a dual purpose, providing structure as well as electronic functionality simultaneously.

- **The joints should be reversible.** They should be capable of being removed and rejoined a number of times before failing. Generally electronic connectors are rated to maintain a certain connection resistance over a certain number of cycles [59]. The challenge is in designing the connector to not wear significantly with use but still achieve a reliable contact that is able to circumvent insulating surface oxides [60].

- **The joints should be geometrically small and simple.** Because each part will feature at least two (and likely more) joints, the volumetric overhead of the joints adds up very quickly.

- **The joint should have some degree of error correction or self-aligning features.** Features that help guide the assembly reduce the precision required by the assembly machine (or person).

- **The joint should be easily "automatable."** For example, joining with only a single degree-of-freedom motion and not requiring the coordination of multiple end-effectors.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 2.2 Interlocking Press-Fit Slots

Based on these desirable traits, I selected a simple interlocking press-fit slot as the best suited joint system (Figure 2-1). The connection between parts is made with two orthogonal interlocking slots that are toleranced to be a tight fit with one another. This joint works well for this application first and foremost because it is geometrically small and simple. The geometry takes up a volume of $3t \times h/2.1 \times t$ given part thickness, $t$, and part height, $h$.

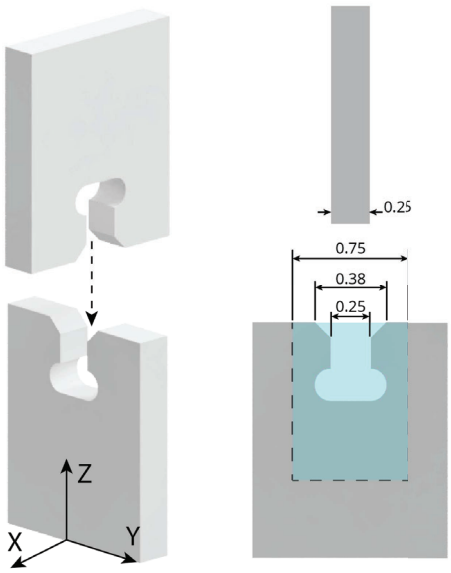



Figure 2-1: Interlocking Press-Fit Slot Design. The blue region indicates the area taken up by the single joint.

Cutouts at the bottom of the slots (often referred to as "dog-bone cutouts") are required to ensure that the parts can fully seat. Without the cutouts the radiused cutting tool (in this case the 0.23 mm kerf width of the Wire-EDM) will leave a radius in the corner of the slot and not allow the full insertion of the parts.

The length of the slot is chosen to be as long as possible with respect to the part geometry. In this case that length is slightly less than one quarter of the part-height, $h$. The length of the slot provides rigidity against torques about the X and Y axes. Rigidity against torques in the Z-axis is improved with an increase in the thickness of the parts.

The joints feature chamfers on the edge of the part geometry to aid in the relative alignment between mating parts. The chamfers are dimensioned to accommodate translational errors in X or Y assembly axes of up to $\pm 190$ $\mu$m.

The force required to insert a part scales with the number of simultaneous interlocking interfaces. For the parts used in this work, which are 250 $\mu$m thick and made of phosphor bronze, approximately 15 N of force is required per joint interface to insert (Figure 2-2). Because of the symmetry in the interlocking slots, the joint requires approximately the same amount of force to remove as it does to attach. This means these joints can resist tensile forces up to this insertion force.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

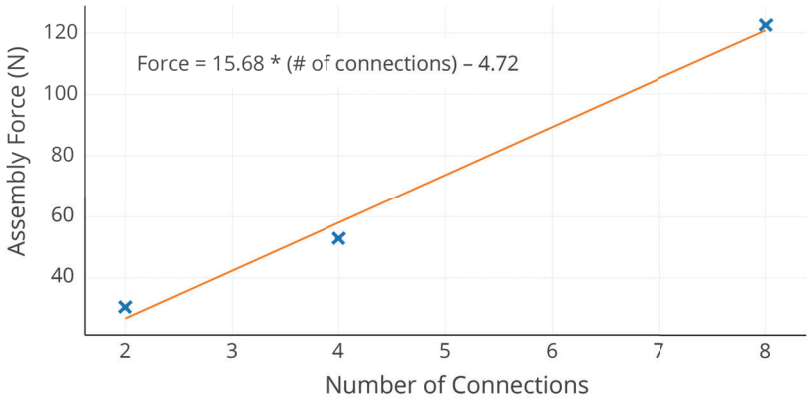



Figure 2-2: Assembly force is proportional to the number of simultaneous joints being made (15.7 N per joint).

### 2.2.1 Electrical Characterization

In my Master's thesis I characterized the electrical conductivity of these kinds of press-fit joints [61] (Figure 2-3). I'm including a summary here for completeness and convenience.

The joints are reversible meaning that the parts can be joined and unjoined a number of times while still maintaining a low contact resistance. To measure this, I constructed a machine that can join and un-join the parts rapidly and measure the resistance across the joint using a sensitive four-wire measurement. In their joined state, the parts are preloaded with a constant force spring. What I found is that the contact resistance is initially somewhat variable and then settles after a few cycles to a low ($\approx 1$ m$\Omega$) contact resistance.

Load cycling of a sample containing many joints led to an overall increase in conductivity. An electron micrograph of the surface revealed the likely cause of this conductivity increase was the flattening of surface asperities, effectively increasing the contact area.

#### Surface Preparation

Surface quality and preparation plays an important role in the development of good electrical contacts. It's common practice in the design and manufacturing of electrical contacts to plate the contacts with metals that form good electrical contacts. These metals are often soft (have low hardness), have high conductivity, and do not form hard oxide layers.

I tested an assembled sample before and after copper electroplating to determine its impact on the electrical conductivity of the sample. The assembled sample is composed of two nodes and three struts. The sample measures 20 m$\Omega$ using a 4-wire resistance measurement (measuring from outer strut to outer strut). This resistance is composed of two components: bulk resistivity of the bronze parts and the contact resistances of the joints. I can very roughly approximate the resistance due to the bulk resistivity of the parts by integrating the resistivity over the volume of the parts with the equation: $R = \rho L/A$. Based on this, I should expect the contribution of the bulk resistivity to be approximately 1 m$\Omega$, much less than the 20 m$\Omega$ that I measured. This indicates that, unsurprisingly,


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

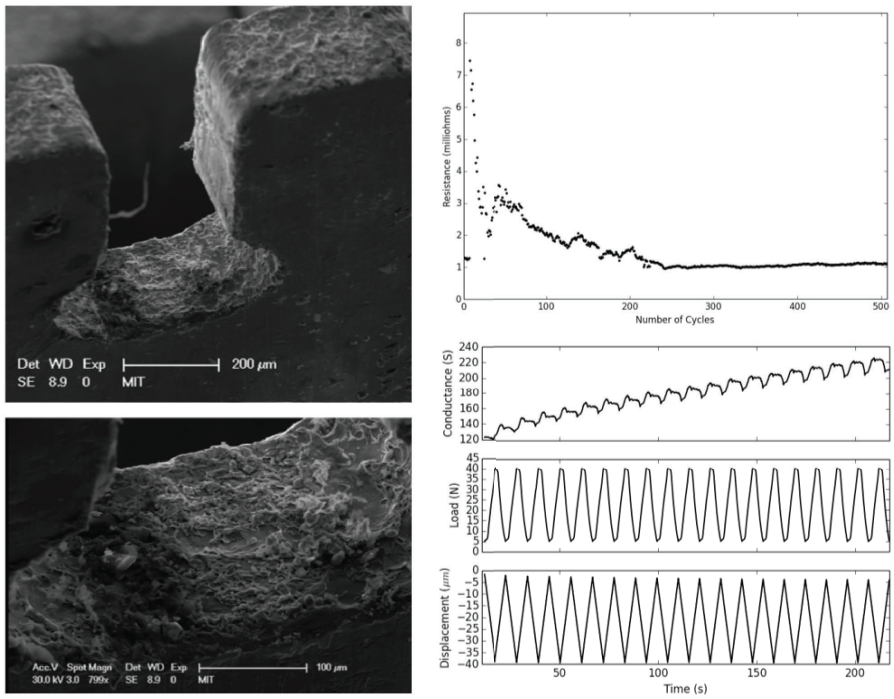



Figure 2-3: (Left) A scanning electron micrograph of an interlocking slot after having been used. Flattening of surface asperities is visible on the side of the slot walls. (Right-Top) The joint is reversible and has a 1 mΩ contact resistance even after 500 cycles. (Right-Bottom) Load cycling a structure composed of joints works to increase the conductivity of the bulk specimen.

the contact resistances contributes significantly to the resistance of the sample.

To account for the individual contact resistances within the sample it's important to have an idea of which resistances are in parallel and which are in series. Imagining the path of an electron through the the sample, I constructed an equivalent circuit to determine how the overall resistance is summed from the individual contact types (strut-node and node-node).

To determine the contribution of contact resistance on the overall resistance, I electroplated the specimen with copper (10 minutes at 200 mA). The idea being that the thin layer of extra copper has a negligible effect on the bulk resistivity but bridges the small gaps at the joints between parts, drastically reducing contact resistance. In fact, after electroplating, the resistance across the structure measures 1 mΩ (±1 mΩ).

Using electroplating to post-process assemblies is potentially possible but not ideal since it does not align well with the idea that discrete assembly can be a digital fabrication workflow. Instead, it is possible to just electroplate node parts prior to assembly and achieve many of the same benefits. Electroplating just the nodes (for 12 minutes at 50 mA) reduces the resistance of the assembled test structure to 1.5 mΩ (±1mΩ), roughly same as the fully electroplated sample.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 2.3 Alternative Joint Systems

While I selected the interlocking slot joint system as the most suitable to develop the functionality that I sought in this assembly framework, other joint types could certainly be worth exploring in more depth. Here, I'll give a cursory overview of some of the other joint types I've considered and their relative merits.

**"Snap-lock"** This joint is very similar to the press-fit slot but includes cutouts on either side of the slot to allow for the creation of flexural "snap-lock" features. This joint system allows for the possibility of tuning the insertion and removal force separately, potentially enabling a very low force insertion but high force removal. This joint suffers two primary downsides. For one, the cutouts on either side of the slot increase the size of the joint, certainly increasing the width of the joint by more than four times the minimum feature size of the manufacturing process (since a slot needs to be cut on either side of the flexure). For parts already produced at the minimum feature size of a process (where $\delta_{kerf} \approx t$), this increases the width of the joint by more than 2 times ($3t + 4\delta_{kerf} \approx 7t, 7/3 = 2.33$). The second major downside is that these flexural cutouts greatly reduce the torsional stiffness of the joint. While it may be possible to include geometric features that help address this concern, they will certainly increase the size of the joint greatly or else negate the potentially low insertion force.

**Directional Locking Features** Joints don't necessarily have to be made in a single motion, but can leverage a sequence of motions to insert and lock a part in place. This could be done, for example, with parts that slide in place (in the plane of the build front) and then are locked in place by a keystone element on the subsequent layer [62]. This kind of joint is attractive because it decouples assembly force from the ultimate strength of the joint. It is also a reversible joint; the assembly just needs to be disassembled in the right order to extract one part at a time. Another variant of this joint type uses micro-structured features (e.g. velcro) in order to attach parts in a controllable way [63].

**Clip** This kind of joint involves features for alignment and relative positioning (like loose interlocking slots) and an orthogonal clip that binds the two parts together. These have been shown to be capable of achieving high performance material properties at relatively small scales ($\approx$ 5 mm) [64]. A main advantage of this method is the decoupling of the force required for assembly from the ultimate joint strength because the clip can be designed to be inserted with very little force in the orthogonal direction but exhibit high strength otherwise. The challenge with this joining method is two fold: the size of the joint area can very easily increase dramatically with the introduction of this orthogonal clip and the clip itself represents an extra (often very small) part that needs to be picked up and manipulated, often greatly complicating assembly.

**Nut and Bolt** Nuts and bolts represent another way to join parts. The main advantage of this approach is that it is capable of making very strong joints [65]. The screw mechanism represents a mechanical advantage that transforms the relatively light installation torque into a large compressive force between the parts. While it has been shown that automated assembly of nut and bolt joints is possible ([66]), it is greatly complicated by the need to manipulate multiple small parts and often



321
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

necessitates complex toolheads or multiple end-effectors. This joint method is limited in its ability to scale down in size as the smallest commercially available nuts and bolts tend to be no smaller than 0.86mm (#000-120).

**Adhesive** Adhesive joints usually take the form of a liquid glue or tape. Liquid glues have the advantage of not requiring any assembly force and are able to join a wide range of different materials including plastics and metals. But they often need time to cure. Dispensing and applying liquid adhesives can be a challenge to do precisely at small small scales but has been shown to be possible, at least with some part geometries (e.g. [67]). For this reason, adhesive tapes may be a more attractive solution. Pressure sensitive adhesives require large assembly forces and may not provide the desired joint strength. B-staged thermoset adhesive tapes can provide a much larger joint strength but require a high temperature ( 200 °C) and long duration (≈2hrs) post-processing step [68]. Adhesives are less well suited for making conductive joints. While conductive epoxies, which include conductive fillers, can be used, they greatly limit the available options. Conductive Z-axis tape, which is only conductive through its thickness, may be used for the assembly of higher density electronic interconnect but suffers from the same problems as pressure sensitive adhesive, namely large assembly forces and low resulting strength [69].

**Solder** Solder is a joint method that's potentially capable of achieving high strength and high conductivity while still enabling disassembly. It has been used in modular robotics and controlled with embedded resistive heaters to activate and deactivate the joint [70]. It is also capable of scaling to relatively small sizes as it has been used with parts as small as 3 mm on a side [71]. However, solder requires the presence of flux to properly wet contact points and because this flux is often corrosive and in liquid form, it would likely need to be applied very soon before part placement and attachment, and may need a post-processing step to remove any excess flux. To activate the joint requires the application of heat. While low temperature solders exist with processing temperatures as low as 138 °C, most require upwards of 200 °C, which limits the materials that can be used in the assembly or means that the heat needs to be very locally delivered (e.g. with a hot-tip, laser, or focused hot-air).

**Welded** There are a number of welding methods that may be used to join parts but I will focus on two: resistance welding and ultrasonic welding. **Resistance welding** is potentially attractive for its simplicity. The part to be attached to the assembly is positioned in place, vertical pressure is applied to it, and then a current is run through the connector interface, locally melting the material and irreversibly joining the part to the assembly. Rather than interlocking slots, the parts may be designed with geometries that create current restrictions (like raised projections) to control where the weld energy is channeled [72]. The downside of this method is that it requires the parts to have conductive interfaces and possibly even requires the parts to be fully metallic depending on the locality of the melt zone and the degree of control in administering the current. **Ultrasonic welding** functions in a similar way but joins parts through melting generated by friction caused by ultrasonic vibrations applied to the interface between parts. While this technique is capable of working with both plastics and metals, it is not capable of working with both simultaneously and so restricts the joints to either plastic-plastic or metal-metal.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 3

# Structure

## 3.1 Lattice Structures

Recent advances in digital fabrication such as additive manufacturing have spurred interested in the study of architected and micro-structured materials, especially sparse lightweight lattice structures. While trusses and space-frames are often used in large scale structures such as bridges and buildings, their incorporation into smaller-scale structures and assemblies has been enabled by the ability to more precisely and economically pattern material in arbitrary ways. In particular, recent work has shown the possibility of printing lattices to build ultralight and stiff structures at micro-scales [73] and a number of commercial offerings have recently appeared to enable the automatic generation of complex lattice structures within user-defined bounding volumes [74] [75] (Figure 3-1 A).

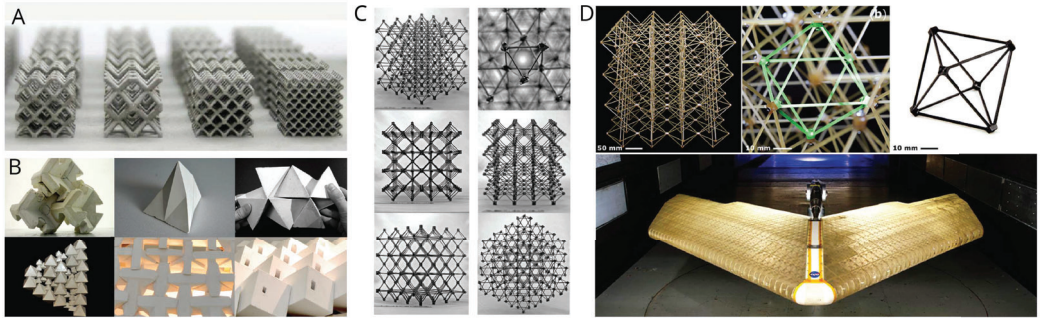



Figure 3-1: Recent research on lattice structures. Additive manufacturing has improved to enable the direct printing of complex lattices [76] (A). Topological interlocking materials use shape and frictional contacts to build tough materials from discrete parts [62] (B). Digital materials have set performance records for ultralight-but-stiff materials [20] (C) and have been used to assemble large scale structures like wings [77] (D).

While additive manufacturing focuses on building these structures monolithically and in a single process, it is possible to assemble architected materials from individual building blocks. Topologically interlocked materials (TIMs) is a field of study that looks at how structures can be assembled from shaped building blocks that interact through contact and friction interfaces. TIM structures often feature a boundary constraint and then generate all other constraints through the interlocking



323
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

shapes of the blocks. These structures have been shown to be highly damage tolerant and are able to absorb much more energy during impact than a monolithic structure of the same shape [78] [79] [62]. The design of these materials is often inspired by natural composite materials like nacre that exhibit incredible toughness, even with brittle constituent materials [80]. TIMs can sometimes be challenging to assemble, however, either requiring multiple degree-of-freedom motions or part stabilization until the global constraint is enforced.

The field of digital material systems represents another way to assemble highly performant structures that incorporate multiple materials or functionalities from individual building blocks. Digital material assemblies can be made using a wide variety of different part and lattice geometries. Digital materials designed for high-performance structural applications often take the form of sparse beam lattices such as cube-octahedra [20] [77] [65], Octet [20], or Kelvin [81] [82]. For functional electronic applications, the parts often take the form of tiles [23], dense interlocking assemblies [83] [84] [85], or sparser "Lego"-like assemblies [61].

## 3.2 Digital Materials

Digital materials have a number of desirable characteristics that make them well suited for the kind of high-density heterogeneous integration sought in this work.

- Digital material assemblies are capable of achieving high performance mechanical properties [20].
- They can incorporate heterogeneous materials or part types. For example, they can be used to assemble shape-morphing structures with the use of the rigid and flexural parts [21].
- Because the parts interlock with one another, there is a degree of error correction that happens during assembly. This means that the assemblies can be made more precise than the machine that assembles them [86].
- Furthermore, because digital material assemblies are often decomposable into simple two-dimensional parts, they can often be fabricated relatively easily at a range of different length scales, using a wide variety of processes, and made from a range of different materials.

Prior work has shown the ability to assemble electronic components from conductive, insulating, and resistive part types [22]. The parts used in these assemblies are referred to as "Lego-GIK" because they assemble in the same top-down fashion as Lego building blocks (Figure 3-2). The GIK acronym here refers to an earlier instantiation of these kinds of press-fit parts that was referred to as the "Great Invention Kit" [18]. These Lego-GIK parts feature four slots on the top of bottom of the part. The number of slots is the minimum number of slots to enable tiling in the X and Y axes.

In expanding upon these capabilities, an updated lattice geometry is beneficial for a few reasons. For one, while the multiple slots of Lego-GIK geometry serve the purpose of enabling a simple part to tile, they also make integrating functionality within the parts a challenge since there is very little extra area that can be used between each slot. Additionally, the redundant middle connections complicate the incorporation of mechanical functionality as they have no clear "input" and "output."

Based on this, the middle slots are removed such that the building block only contains two slots on the top and bottom. This opens up the central portion of each building block for the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

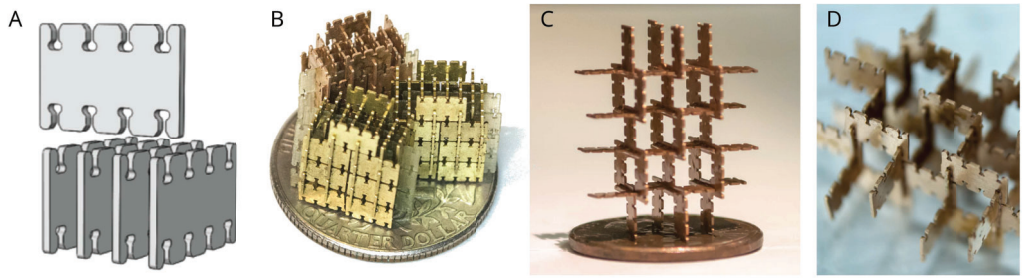



Figure 3-2: A series of structures built from "Lego-GIK" parts, which were used to explore the assembly of electronic digital materials.

development and incorporation of higher-level functions such as flexural hinges and actuation. This modified geometry has clear input, functional, and output regions within each building block.

This geometry still features a number of undesirable traits. For example, there is no way to tile these parts within a single plane. They rely on connections to parts above and below them and so require a minimum of four parts in the vertical direction in order to tile the plane. Furthermore, a number of the desired functions rely on the press-fit slots to impose an angular moment constraint. With only a single press-fit slot, repeated torques could cause unwanted wear and eventually result in degraded contact.

A further limitation can be observed by noting the minimum structural extension to the structure. That is to say, the minimum structural addition to the structure. In this case, that minimum structural extension is under-constrained, attaching to the structure through only a single press-fit connection. This under-constraint represents a challenge for automated assembly of the structure because it means assembly forces need to be controlled and directed in specific regions so as to not torque the part as it is being inserted.

## 3.3 A New Geometry

Based on these limitations, a further geometry was developed in order to better incorporate the mechanical and electrical functionality within the lattice. In my exploration of the design space of lattice-based discretely assembled structures, I restricted my search to regular rectangular lattices for their compatibility with press-fit construction methods and for their ease of laying out functionality in the cartesian axes (Figure 3-3).

Ultimately the new geometry relaxed the constraint of using a single part-type and introduced a secondary part-type, decomposing the lattice into edges and vertices, hereafter referred to as struts and nodes. In this decomposition, nodes are first assembled from four identical two-dimensional parts, connected through interlocking slots. Nodes are then connected to neighboring nodes through a pair of parallel struts (Figure 3-4).

This explicit decomposition into struts and nodes has a number of advantages. For one, it makes the parts more volumetric. Rather than two-dimensional panels, the parts are more cube-like, making them more amenable to automated assembly methods and allows them to be more easily manipulated. This volumetric nature also means that each slot is put under less torsional loading.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

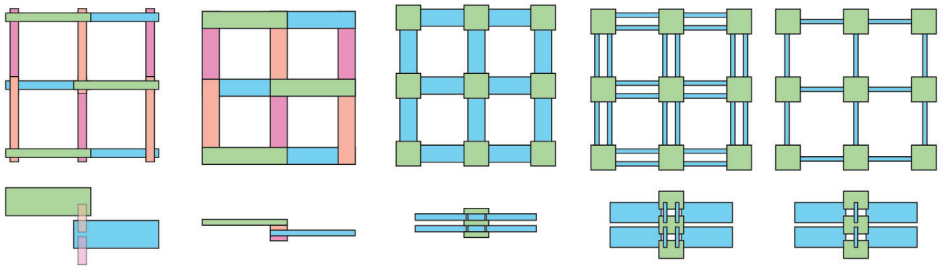


Figure 3-3: An exploration of the lattice design space. The leftmost design represents "Lego-GIK" construction while the second from the right represents the new geometry.

The rectangular nature of this lattice geometry allows these structures to tile the plane with a single layer and additional out-of-plane parts are not needed to give the layer its structure. The strut and node decomposition also provides a natural way to embed functionality in the lattice. As will be seen in the chapters to come, the struts often conveniently house functionality while the nodes provide connectivity.

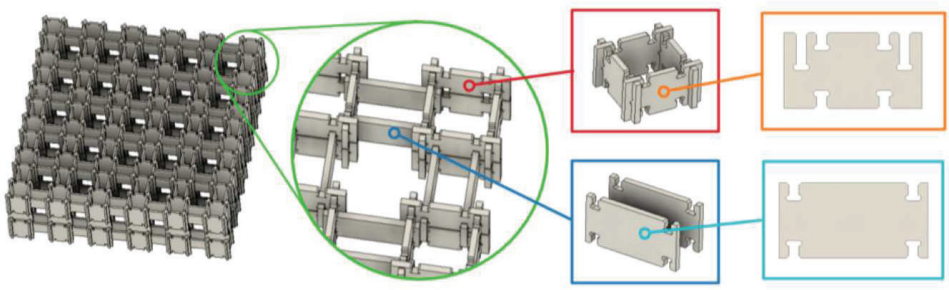


Figure 3-4: The rectangular lattice is composed to node and strut part types. Four node sub-parts are assembled to form a single node. Nodes are then connected to neighboring nodes through parallel struts.

With just these two part-types (node and strut), this assembly framework is 2.5D. This means that three-dimensional structures can be made through the stacking of individual planar layers. While this enables the assembly of a wide range of capabilities, it only enables functionality to be arranged within the horizontal plane. With two additional part types, these structural planes can be joined orthogonally. This capability enables the construction of orthotropic planes that are then assembled into three dimensional objects with functionality in any desired orthogonal plane. An example of this is discussed in Chapter 8.2.

## 3.4 Sizing

It is possible to make these parts and structures at a wide range of different scales. I've chosen to focus on the development of millimeter-scale parts for a few reasons.

- **Conventional millimeter-scale assembly is hard** [87]. The millimeter-scale is too large


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

to use photolithographic processes to produce integrated machines as is done in MEMS but too small to use any off-the-shelf components. Currently most machines made at this scale are produced in monolithic processes to avoid the assembly step. Automating assembly at this scale could help open up the fabrication of integrated a machines with a more diverse range of functionality.

- **It is still amenable to rapid prototyping.** Laser micromachining, wire-EDM, and some machining processes are able to produce the required feature sizes at this scale to make interlocking parts in a direct-write way.

Given this, I sized parts based on the minimum feature sizes of the fabrication tools available to me. In this case, that meant standardizing around 254 $\mu$m thick material and means the corresponding smallest feature size is also 254 $\mu$m. This enables relatively rapid production using the 150 $\mu$m wire on the wire-EDM.

The parts are sized to balance two objectives: (1) minimizing the pitch-spacing of the lattice (to maximize resolution) and (2) maximizing the functional area available to each part. The design space can be pictured by looking at two extremes. In prioritizing objective (1) one may seek to shrink the parts entirely until they are the minimum possible size to allow for the interlocking joints to function. On the other hand, prioritizing objective (2) would mean stretching the parts until the joint-interfaces become infinitesimally small compared to the body of the part.

Similarly, there's a need to balance the size of the nodes and struts. Shrinking the nodes allows for the lengthening of struts (and therefore functional area) without affecting the pitch spacing of the lattice. However, doing this also necessarily shrinks the spacing between the pair of struts, reducing the overall strut volume. Ultimately the geometry that is used in this work strikes a balance between these but ensures that the struts are longer than they are wide, which is beneficial for a number of the functions that are developed and described in coming chapters.

## 3.5 Production

The parts that make this structure are two dimensional with relatively simple geometric features, allowing them to be easily fabricated using a range of different processes.

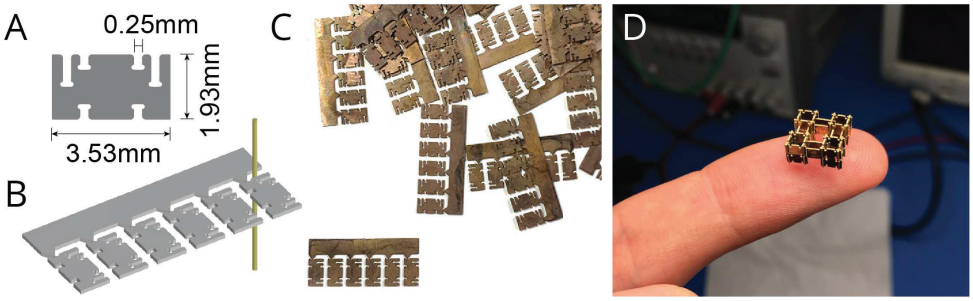



Figure 3-5: Node and strut parts are fabricated using wire-EDM, enabling the production of 108 parts in a single cut.

For prototyping, the basic structural (node and strut) parts are cut by wire-EDM (electrical




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

discharge machining) from 254 $\mu$m thick phosphor bronze. Arranging parts as a group (of six) and cutting a stack of (18) bronze sheets at the same time, enables the production of 108 parts with a single cut and takes just under one hour (Figure 3-5). These 108 parts are enough parts to assemble 27 nodes or 54 struts.

Nodes are assembled from four node parts manually using pliers. Automation of this assembly step is certainly possible and should be considered for scaling up the production rate of these parts.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 4

# Mechanism

## 4.1 Compliant Mechanisms

In order to create machines, we need ways to constrain motion. Conventional machines at the macroscale often using bearings to constrain motion. Assembling bearings from individual discrete parts is hard to imagine unless the building block parts are significantly smaller than the bearing contact surfaces. In our case with millimeter-scale parts, this would entail building bearing surfaces from thousands or millions of parts.

Biology, however, demonstrates the potential for creating mechanisms and motion constraint without the use of bearings. While some biological bearings do exist [88], much of the movement that biology accomplishes is through the arrangement of rigid and compliant structures.

Compliant mechanisms is an emerging field of research that looks at recreating a wide variety of mechanisms and linkages using embedded flexural degrees of freedom rather than pins and bearings. These mechanisms offer advantages over conventional bearing-based degrees of freedom such as increased performance (high precision, low friction, low mass, reduced wear), lower cost (simplified fabrication, low part count), and ability to miniaturize [89].

Compliant mechanisms made using subtractive processes are normally limited to a single material and so their mechanism (force/displacement properties) is purely a function of geometry. Without the ability to spatially vary material properties such as stiffness and strength, these mechanisms often end up working only for a limited range of applications [89].

New methods in additive manufacturing have recently enabled the fabrication of compliant mechanisms directly from digital designs with spatially varying material properties. These methods are still limited, however, in the range of material properties that are available. Additionally, the resolution of this technique limits the utility of mechanisms made this way at millimeter scales.

Instead, rather than printing or shaping material, the assembly approach I use enables the integration of mechanism part types that exhibit a high degree of compliance about a certain axis while remaining stiff in other axes. These flexural parts can be produced using laminate based processes that enable the heterogeneous integration of materials giving the designer much more control over the resulting properties of the mechanism like stiffness and range of motion. Additionally, compared to multi-material additive technologies, the assembled mechanisms are able to take advantage of better material properties and much finer resolutions (e.g. thinner flexible regions) than are typical


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

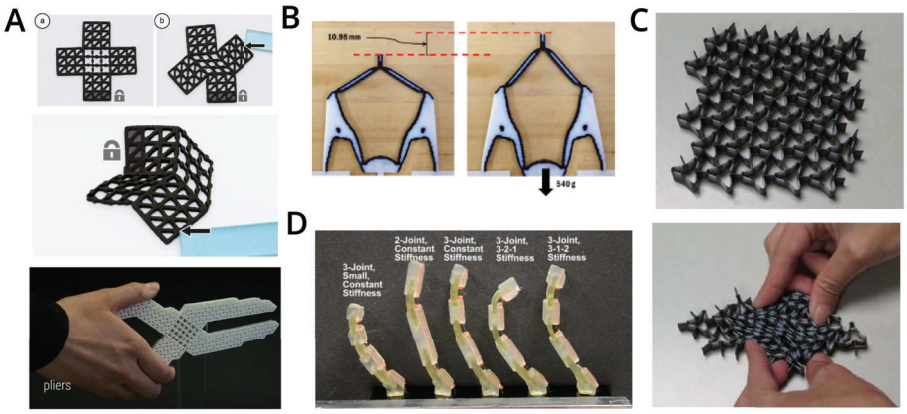



Figure 4-1: Recent research on compliant mechanisms. Compliant mechanisms can be designed as lattice metamaterials [25] and have been shown to be capable of being used as an amplifying and inverting mechanism [90], auxetic materials [91], and robot appendages [92].

for additive manufacturing approaches.

This laminate based approach to manufacturing compliant mechanisms has been explored increasingly in recent research [93][94][27]. The technique is often referred to as PC-MEMS (Printed Circuit Micro-Electromechanical Systems) [8]. Mechanisms made using this approach can be constructed to have a very high degree of compliance on-axis while remaining stiff off-axis [95].

## 4.2 Part Types

Two part types enable the assembly of a wide range of planar mechanisms (Figure 4-2): (1), a part with a single flexible degree of freedom and (2), a part with two flexible degrees of freedom. While it may not be possible to discretely assemble continuous rotary or linear constraints using discrete parts (at least at this scale and resolution), it is possible to assemble flexural mechanisms that approximate these motions, which exhibit a high degree of compliance within a certain axis or plane and high stiffness off-axis. Figure 4-2 shows common mechanism motifs that can be assembled from these two part types. In particular, two highlighted motifs are the simple rotary hinge and four bar linkage, which can be used to approximate linear motion.

## 4.3 Modeling

### 4.3.1 Kinematics

The discrete parts of this assembly framework provide a natural way of decomposing a complex linkage to evaluate its mobility. Parts assemble into a rectangular lattice framework in which pairs of nodes are connected by struts. This strut is what houses the mechanical degrees of freedom of the flexure hinges. Gruebler's equation can be used to characterize the mobility of some common


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

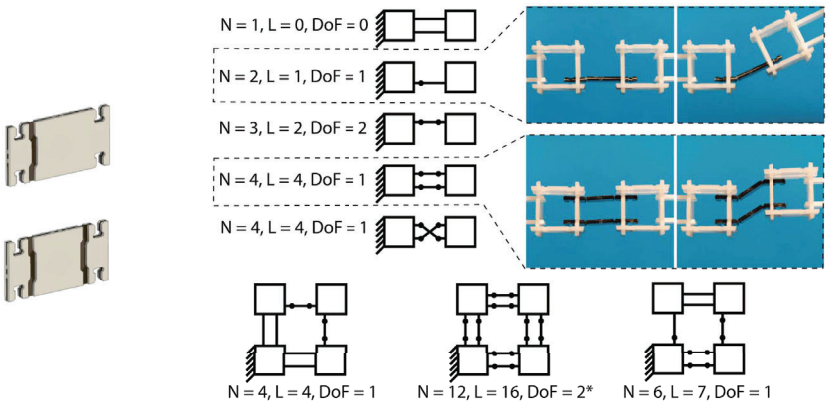



Figure 4-2: Two part types (left) are necessary to enable a wide variety of planar mechanisms. A series of strut motifs are characterized based on the Gruebler mobility criterion. Two motifs that are commonly used are highlighted: a revolute joint (top) and a parallelogram four-bar linkage mechanism that can be used for linear motion constraint (bottom).

strut arrangements (Figure 4-2). Gruebler's equation is defined as:

$$\text{DoF} = 3(N-1) - 2L - H$$

where $N$ is the number of bodies or links (including the ground link), $L$ is the number of lower order pairs (arresting two degrees of freedom), and $H$ is the number of higher order pairs (arresting a single degree of freedom).

The constraint of linear motion rests on the ability of the parallelogram flexure to approximate linear motion over small displacements. The parallelogram mechanism is straightforward to model and simulate using simple rigid-body models. In this case, I find that the parasitic motion is related to two parameters: $L$ (the distance between the flexure hinges) and $d$ (the magnitude of linear displacement). The error motion is given simply by:

$$\text{Error} = e = L - \sqrt{L^2 - D^2}$$

True straight-line linear motion constraints can be assembled with a few more parts. For example, by stacking two of the simple parallelogram linkages on top of one another, perfect straight-line motion constraint can be achieved. The caveat with this design, though, is that it is only well constrained for small displacements. With large displacements ($>$ 15 deg), it reverts from a single degree-of-freedom mechanism to a two-degree of freedom mechanism. Other researches have explored modifications and additions to these mechanisms that help constrain them over larger displacements [96] [97].

These two part types are able to assemble any planar mechanism (given unbounded size). Even within confined machines, they span a wide range of desirable mechanisms and transmissions. For example, Figure 4-3 demonstrates the assembly of a transmission that converts linear motion into the rotational motion of a gripper. It also shows the assembly of a transmission that converts rotary motion into straight line motion for use as a jumping leg. Finally it demonstrates the assembly of a




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

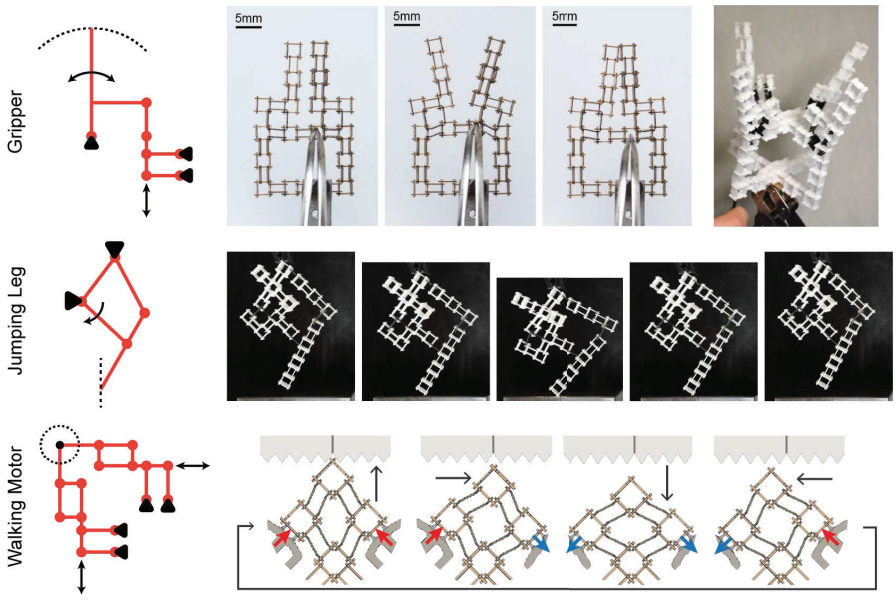



Figure 4-3: A variety of transmissions assembled from the two flexural part types. A gripper must convert linear motion into rotary motion. A jumping leg converts rotary motion into straight-line motion. And a walking motor decouples and distributes two independent degrees of freedom to a motor tip.

transmission that decouples and distributes two independent actuated degrees of freedom to a single drive tip in the case of a walking motor.

### 4.3.2 Quantization of a continuous range of transmission ratios

Given that these two part types enable the assembly of a wide range of mechanism, the question naturally arises: Can these two part types assembly any planar (pin-jointed) mechanism? To which, the answer is yes, but only if given an unbounded envelope. However, with volumetric constraints the answer is harder to define and a better question becomes: How closely can this discrete set of parts approximate a continuous range of transmission ratios?

A simple lever-based linear to linear transmission serves as a good representative test case for this analysis. With unlimited space constraints, pin-joints placed on a lattice can approximate any transmission ratio with arbitrary precision. To achieve ratios very close to certain desired ones, however, will require a very large number of parts. To achieve the exact ratio of 2.45, for example, would require $49 + 20 = 69$ parts (since the ratio is 49/20).

This is improved by introducing a sidedness to each hinge such that it is located on a lattice but is either shifted to the left or right by an amount, $D$. In these assemblies, in which a pair of strut-parts makes up a single strut, this is equivalent to picking which of the strut-parts has the hinge. This simple additional parameter drastically reduces the number of parts required to achieve a certain level of precision.

To demonstrate this, I run a series of optimizations that adjust parameters to achieve a desired


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

transmission ratio. The objective function is:

$$\min((desired - discrete)^2 + \alpha * (b + a))$$

where $\alpha$ is a weighting factor ($\alpha = 0.001$), $b$ and $a$ are the unit lengths of the lever arms, and $discrete = b/a$. The optimization is subject to constraints such that $a$ and $b$ are integers and are perturbed by $\pm D$ based on the binary values $p1$, $p2$, $p3$, representing the sidedness of each of the three flexures.

The result of a sweep between ratios of 1 to 6 with steps of 0.0625 is pictured in Figure 4-4. The first plot shows the desired ratio against the actual ratio and the second plot shows the absolute percent error between desired and actual. The third plot shows the overall size ($a$+$b$) of the required linkage.

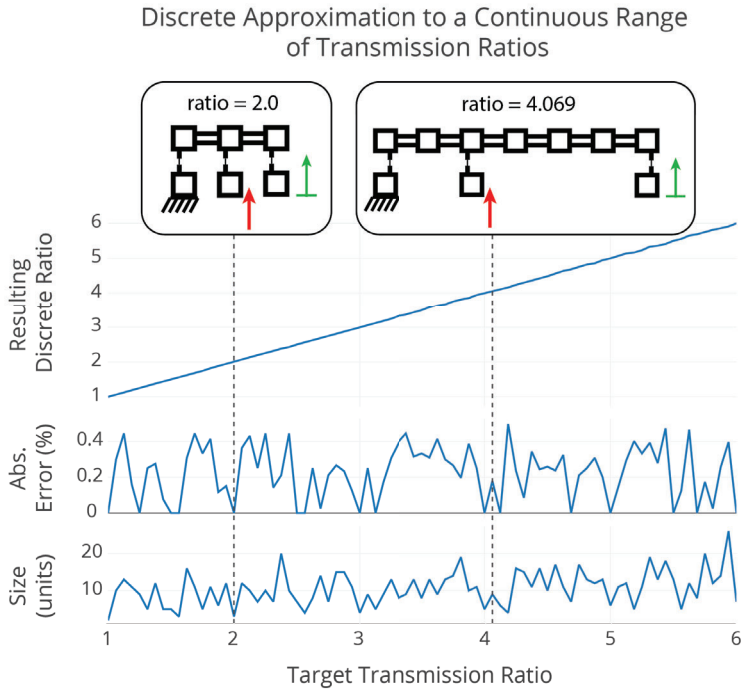



Figure 4-4: A plot showing that discrete parts on a lattice can representing a continuous range of transmission ratios. Ratios between 1:1 and 1:6 are sought at increments of 0.0625. By toggling which strut is made flexible, the continuous range can be represented in a finite area with less than 0.5% error.

What this shows is that these discrete mechanisms are able to represent a continuous range of transmission ratios with a small degree of error and in a finite length. Looking at a particular example is helpful to understand this. Take the desired transmission ratio to be 4.0625. With only discrete steps on the lattice in this straight lever-arm example, the smallest linkage that gets within 0.5% of the target ratio is 66 units long with a ratio of 4.077 ($b = 53$, $a = 13$). With a 5 mm lattice pitch, this entails a 0.33 m long lever arm, which is clearly not ideal. With the introduction




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

of the hinge-sidedness, the required size to achieve an error less than 0.5% is reduced by an order of magnitude to 7 units long with a ratio of 4.069 and entails only a 35 mm long lever arm.

Clearly this case of a straight lever-arm is constructed to illustrate this principle and real mechanisms can often take advantage of more than one dimension. For example, another way of achieving that 4.0625 transmission ratio is to construct a planar lever-arm whose hypotenuse approaches the target ratio. For example, without factoring in sidedness, a lever arm with $a = 2$ and $b = \sqrt{2^2 + 8^2}$ achieves a ratio of 4.062 within a 4 x 8 unit envelope.

### 4.3.3 Dynamics

Because these mechanisms must be capable of being displaced by actuators built into the lattice, it is important to understand their dynamic properties. By treating the flexural hinge as a beam, it is possible to model its stiffness using beam bending equations. The on-axis torsional stiffness is described by:

$$k_t = \frac{Ebh^3}{12L_{eff}}$$

Where $E$ is the modulus of elasticity (≈2.5 GPa), $b$ is the width of the hinge (2 mm), $h$ is the thickness of the flexible layer (0.25 mm), and $L_{eff}$ is the effective length of the hinge (0.216 mm). Since a castellated hinge geometry is used, the length of the flexing hinge is not well determined but can be approximated by taking the average of the minimum flexing gap and the maximum flexing gap. Given the geometry of our flexure hinges, the predicted torsional stiffness is 28.9 mN*mm/rad.

The on-axis stiffness of assembled mechanisms can be predicted for small displacements using conservation of energy:

$$\frac{1}{2}k_x\Delta x^2 = \sum \frac{1}{2}k_t\Delta\theta^2 = \frac{N}{2}k_t(\frac{\Delta x}{L})^2 \implies k_x = \frac{NK_t}{L^2}$$

where $N$ ($N = 4$) is the number of flexure hinges undergoing an equal deformation of $\Delta\theta$, which I approximate for small displacements as $\frac{\Delta x}{L}$, where $L$ is the spacing between hinges ($L = 2$ mm). Given this, the expression to predict $k_x$ from $k_t$ is simply $k_x = k_t$.

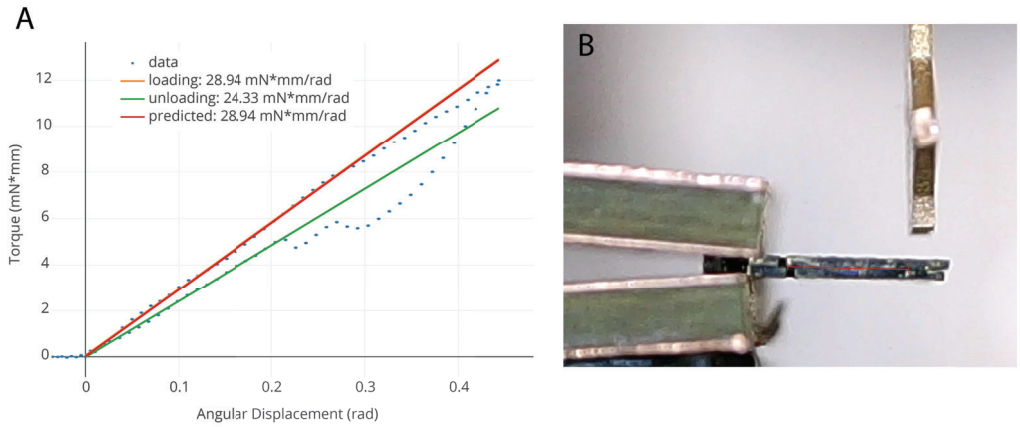



Figure 4-5: Stiffness measurement of a single flexure hinge. The flexure hinge exhibits some hysteresis such that the loading direction is stiffer than the unloading direction.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

I validated this estimation with measurements of the on- and off-axis stiffness of a single flexure hinge as well as an assembled parallelogram linkage (Figure 4-5). Measuring the torsional stiffness of the single hinge, I found the behavior to be slightly hysteretic such that the stiffness in the loading direction (28.9 mN*mm/rad) is almost 20% more than the stiffness it exhibits as it is unloaded (24.3 mN*mm/rad). Even with this hysteresis, these values closely match the analytical prediction for on-axis stiffness. The same flexure measures 9.7 N*mm/rad off-axis, which represents a two orders of magnitude difference in stiffness between on- and off-axis directions.

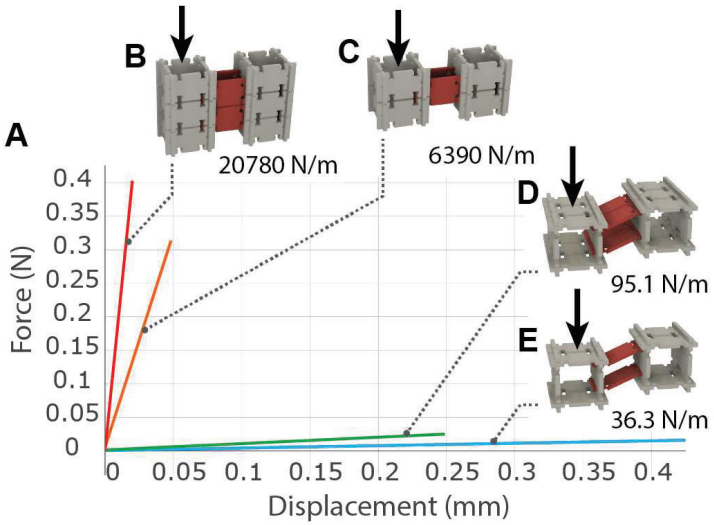



Figure 4-6: The flexures are measured with loads applied both on-axis (in the direction of motion) and off-axis. The flexures are much stiffer off-axis than on-axis and additional layers can be added to increase the anisotropy.

I performed the same test on an assembled parallelogram mechanism composed of two nodes connected by two double-hinged struts. I found the stiffness on-axis to be 36.3 N/m and off-axis to be 6390 N/m. By adding more layers to the parallelogram assembly, this off-axis stiffness can be increased even further. A two-layer parallelogram mechanism measures 95.1 N/m on-axis and 20780 N/m off-axis. These results are illustrated in Figure 4-6.

These results match the analytical predictions well. The measured stiffness of the single-layer parallelogram is approximately 20% more than that of our prediction. This difference is most likely explained by small geometrical differences and misalignment in the parts leading to over-constraint. When loading the assembled structure off axis, the stiffness of the fixturing becomes relatively significant. In order to remove this from the measurement, I measure the stiffness of the fixturing and then subtract out its contribution based on series springs.

## 4.4 Production Methods

The flexural parts are built using a laminate process that is often referred to as printed-circuit MEMS (PC-MEMS) or smart composite materials (SCM) [8]. This process is similar to how printed


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

circuit boards are constructed but rather than etching pre-laminated materials, this process involves the lamination of pre-machined layers of material using a thermoset adhesive.

In this case two rigid outer layers sandwich an inner flexible membrane. The outer layers feature cutouts to allow the inner flexible membrane to bend within these defined regions (but not elsewhere).

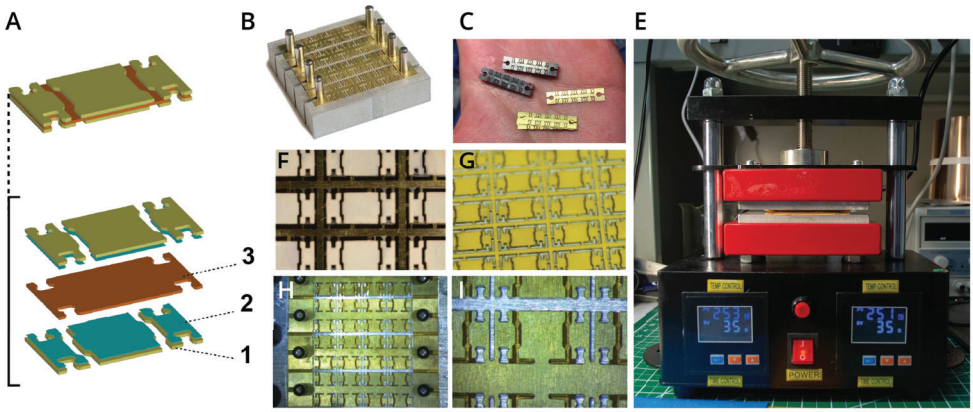



Figure 4-7: Part production methods for flexible parts. The flexible parts are made through a laminate process in which rigid metal layers (C) sandwich a flexible membrane (E) and are bonded together with the use of a thermoset adhesive (D). Alignment is accomplished using a jig with dowel pins (B) and results in an array of parts that are subsequently released.

**Rigid layer machining** The rigid layers are most efficiently machined using wire-EDM. I wire-cut the rigid outer layers from stacked sheets of 100 $\mu$m thick brass. I combine multiple parts onto a single central runner to enable the production of many parts with a single cut. In particular, I combine six parts into a single cutout and stack eight sheets of 100 $\mu$m brass, enabling the production of rigid layers for 24 parts with a single cut (Figure 4-7 C).

These rigid layers may also be cut using a micromachining laser (e.g. Oxford 532nm DPSS). This is sometimes attractive for small batches and test runs where time to go from the design to the first test article is more important than the production time per part.

**Flexible and adhesive layer machining** The flexible layer is laser micromachined (Oxford Systems) from 25.4 $\mu$m thick polyimide (Kapton) (Figure 4-7 G). After machining the rigid and flexible layers, they are subsequently bonded through the use of a 12.7 $\mu$m thick B-staged thermoset adhesive (Pyralux 1500). The adhesive is skim cut using the laser micromachining system (Oxford) such that it is cut through the adhesive but not the backing. Alignment holes are then cut through both the backing and the adhesive. The adhesive is weeded using tweezers to eliminate unwanted adhesive regions. This weeding process can be slightly arduous and is the likely the most labor intensive and least parallelizable part of the production process.

**Lamination** A number of different methods can be used to generate the heat and pressure required for lamination. A toaster oven with Kant-twist clamps is perhaps the simplest but lacks precision both in the temperature of the laminate and the pressure applied. A benchtop heatpress can control the temperature much more precisely but it can be easy to apply too much pressure (Figure 4-7


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

E). Replacing the crank and leadscrew of the upper press-plate with a simple weight is the simplest method to achieve more precise control over the pressure. This has another advantage which is that it allows for thermal expansion of the laminate materials while maintaining consistent pressure.

The adhesive is first tacked to the rigid brass layers using less heat and pressure than necessary for a full cure (120 °C for 15 mins). Alignment is achieved with the use of a wire-cut two-sided fixture that features matching dowel pin holes to those of the pre-machined layers (Figure 4-7 B). Once the adhesive is tacked to the brass layers, the backing can be removed and the laminate can go back in the heat press with the inner flexible layer inserted for the full cure (200 °C for 1 hour).

**Singulation** After lamination, the parts need to be released from the sheet. I used a micromachining laser to cut the parts from the rest of the sheet. Dowel pin registration holes are first cut in a piece of sacrificial material (such as brass). This provides a simple method of re-aligning the laminated sheet with respect to the laser machine's coordinate system. Subsequent production runs can be re-aligned to these drilled holes with the aid of registration crosses that are cut into the sacrificial substrate. A cutout toolpath is then generated and run, which trims the tabs connecting the parts to the scaffolding. The parts are then easily picked up and removed with tweezers. A quick (10 minute) immersion in an ultrasonic bath (with water) can be helpful to remove any debris generated by the laser cutting process.

This kind of lamination process is particularly attractive because the two-dimensionality makes it scalable. The process can be employed with small batch sizes of only a handful of parts for prototyping or can be scaled to much larger batch sizes for higher throughput production. Furthermore, the process is potentially amenable for implementation as a roll-to-roll process. In such a workflow webs of raw material may be continuously fed in, machined or stamped into shape, bonded, and exit as a roll of finished parts.

## 4.5 Discussion

While mechanisms made through the assembly of these laminated flexure parts can be highly performant, there are some aspects that need to be considered relative to more conventional mechanisms.

The thin flexure membranes can be prone to buckling if they are loaded in compression above their critical buckling load. This threshold can be calculated using the following equation:

$$F_{critical} = \frac{n\pi^2 EI}{L^2}$$

where $n$ is a factor based on the bounding constraints ($n = 4$ for fixed-fixed), the Young's modulus, $E$, for polyimide is approximately 2.5 GPa, and $I$, the moment area of inertia is $I = \frac{bh^3}{12}$. Since the rigid layers are castellated around the flexure hinge, there is not a clearly defined length but we can approximate it by taking the average of the maximum and minimum length between rigid boundaries. In this case, $L_{eff} \approx \frac{L_{max}+L_{min}}{2} = 0.216$ mm, $b$=1.92 mm, $h$=25 $\mu$m, and so $F_{critical} = 5.3$ N.

In order to avoid this buckling, a castellated hinge geometry can be used, which reduces the length of the unsupported flexible layer while still allowing a large range of motion [98]. This strategy, however, does contribute to reduced lifetime of the flexure hinge, which has been shown to



337
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

be a strong factor of the flexure hinge length [98]. In my testing, the earliest I have seen a flexural hinge fail is after 16 hours of testing at 20 Hz, or 1.16 million cycles (Chapter 8.1).

Another strategy to avoid buckling flexures is to use the principle of inversion. Flexures that are identified as having a predominately compressive load can be replaced with an inverted flexure part that converts that compressive load to a tensile load [99].

While these mechanisms can be stacked vertically to increase the off-axis stiffness, the scaling of off-axis stiffness with the total height of the flexure is less performant than with a monolithic construction. To measure this, I compared a finite element analysis of a series of monolithic flexures with the same conditions applied to a series of discrete flexures. I chose to study a the stack-up of a two-hinge flexure. Using the solid mechanics module of COMSOL Multiphysics, I applied a fixed constraint to one boundary and unit boundary load (1 N) to the opposite boundary. The distance between each layer is 1 mm and, in the discrete model, each layer is sized to be just less than this pitch (0.99 mm). The simulation is run at 20 intervals for a total flexure height ranging from 1 mm up to 20 mm. The results (Figure 4-8) show that the stiffness of the monolithic construction method increases more rapidly with each additional layer of material (slope = 0.134 N/$\mu$m/mm) than the discrete construction method (slope = 0.020 N/$\mu$m/mm).

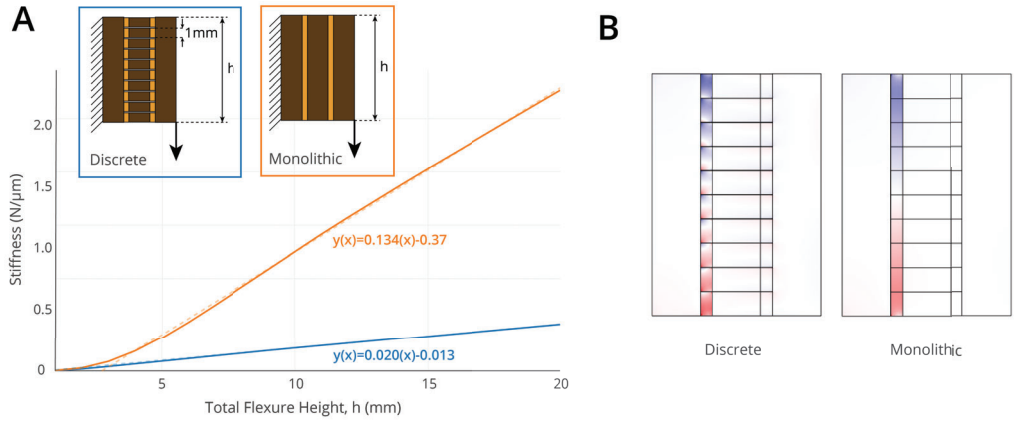



Figure 4-8: Simulation results comparing the off-axis stiffness scaling of monolithic and discretely assembled flexures. (A) Discretely assembled mechanisms have lower off-axis stiffness for an equivalent size compared with monolithic flexures and the different increases for increasing levels of discretization. (B) The strain distribution in the monolithic flexure is even and continuous while the discrete flexure is unable to evenly distribute forces evenly.

This difference in the scaling of off-axis stiffness can be explained by the fact that the discrete construction method is only able to transfer forces between layers at the bounds of the two-hinge flexure. Because the monolithic flexure is able to transfer forces through the entire region of material, strain is less concentrated in the flexure. This can be evidenced by comparing the volumetric strain distribution between these two simulations (Figure 4-8). The discrete construction method results in the top and bottom flexures being put in essentially pure tension and compression, respectively. Compressive forces within these flexural regions can be cause for concern as they can buckle the flexible membrane if they exceed the critical buckling threshold.

To perform better in this regard, future assembly systems could look at methods to join at


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

more regular intervals. It is likely challenging to connect the flexible region of one discrete flexure to another but entirely feasible to connect adjacent rigid regions on neighboring discrete flexures. Doing so would improve the scaling of off-axis stiffness significantly, at least in cases like this where the rigid region of the flexural part makes up a large proportion of overall part area. Another alternative would be to enable the assembly of parametrically sized parts. For example, if a particularly stiff flexure was needed to resist large off-axis loads, a tall flexure could be produced and assembled in place of many discrete ones.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 5

# Actuation

## 5.1 Actuation at the millimeter scale

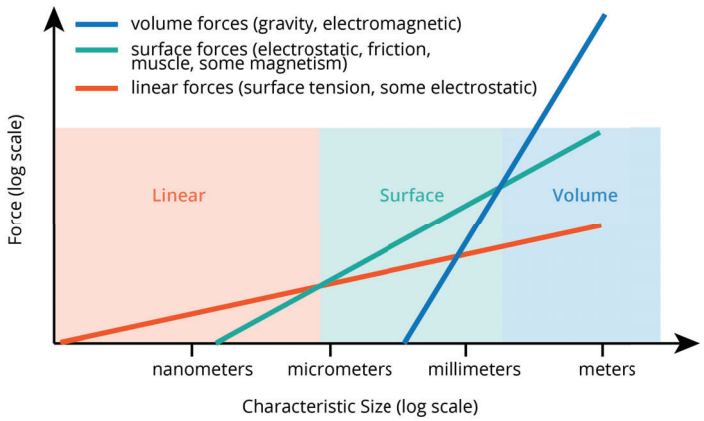



Figure 5-1: Derived from [87]. An overview of how forces scale across length-scales. At the millimeter scale the effects of volume forces, which dominate macro-scales, become less pronounced as surface and linear forces start to play a bigger role.

At the millimeter scale, a number of actuation techniques can be used to drive these mechanisms, including piezoelectric bending actuators [100], dielectric elastomer actuators [101], shape-memory alloy actuators [102], and electromagnetic actuators [103]. I performed analysis of a number of different actuation techniques in order to help down-select an appropriate actuator for integration in the millimeter scale assemblies. I will briefly review the high level analysis of these actuation techniques before exploring the detailed design of the electromagnetic Lorentz force actuators that I chose. In this analysis I assume that the actuator must be manufacturable using available direct write methods. Therefore the minimum feature size should 20 $\mu$m if it's to be made on the Oxford Laser and should be greater than 180 $\mu$m if it's to be made by wire-EDM.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Electrostatic** I looked both at electrostatic gap closing and lateral force actuators. Gap closing actuators are limited in displacement to 1/3 of the gap width due to the pull-in instability [104]. Due to this instability and the breakdown field strength of air, I found that this actuator would require on the order of 150 V to produce 5 mN of force with a maximum possible stroke of 20 $\mu$m. I deemed this force and stroke too limited to produce useful motion at the millimeter and centimeter scale. The lateral force electrostatic actuator was able to provide larger strokes but still provided insufficient force (Figure 5-2 B). For comparison, a Lorentz force actuator produced at the same scale is able to provide $O(100)\times$ the force and $O(100)\times$ the stroke of this kind of actuator.

**Piezoelectric** I also studied piezoelectric actuators, looking at unimorph and bimorph bending actuators as well as amplified stack actuators. A piezoelectric bimorph actuator at this scale is far more feasible than an electrostatic actuator and is able to provide forces in the neighborhood of 300 mN along with displacements of approximately 200 $\mu$m (Figure 5-2 C). A flexurally amplified piezoelectric stack actuator is able to provide even more force ($\approx$1 N) but with much more limited displacement (30 $\mu$m) (Figure 5-2 A).

**Electromagnetic** Of electromagnetic actuators I looked at a number of different configurations including Lorentz force actuation, two-phase linear stepper motors, as well as a more conventional iron-based linear motor. I ultimately chose the Lorentz force actuation for its manufacturing simplicity, its ability to achieve constant forces over long strokes, and the lack of any force between the forcer and mover other than the Lorentz force.

The two-phase linear stepper motor is composed of three part types: a magnet tile that is placed on the top surface of an assembly and two coil parts, each of which is geometrically shifted by a half phase with respect to the other (Figure 5-2 E). My analysis showed that this kind of actuator could produce force peaks of approximately 65 mN and could do so over an arbitrarily long stroke defined by the constraining linkage. However, a number of practical concerns arose with this design. The magnet tiles would not want to be positioned the way they need to be. This means that the assembly of the magnetic tiles would be complicated by forces trying to prevent that assembly from happening. Even once the tiles were in place, it would exert potentially non-trivial forces on the structure. The minimum useful constrained actuator built with this technique would require a large number of parts and so would limit the number of systems that I could prototype and test.

A linear motor using an iron core to shape the flux loop was also considered. This design is able to produce 200 mN of force with a 1 mm stroke. The stroke, in this case, is actually limited by the four bar linkage's approximation of a linear motion constraint. Unlike the Lorentz force actuator, the force this actuator produces is highly nonlinear over the stroke and means that it is likely only able to operate in a "bang-bang" modality, moving quickly from one side to the other. While the actual force producing components take up roughly the same volume as the Lorentz force actuator, the flexure required to constrain the motion is much larger because it must kept in tension to account for the attractive force between the forcer and mover.



341
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

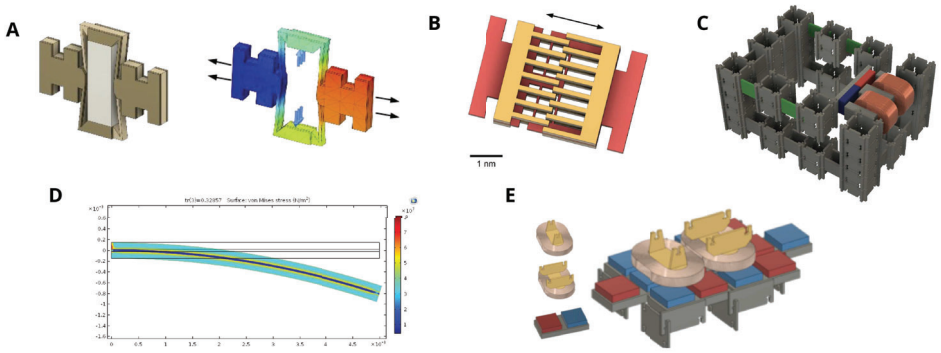



Figure 5-2: A number of actuation techniques were surveyed before Lorentz force actuation was chosen. (A) Piezoelectric stack actuator amplified by a flexural mechanism. (B) Lateral displacement electrostatic actuator. (C) Iron core electromagnetic actuator. (D) Modeling of a piezoelectric bending actuator. (E) An electromagnetic bipolar stepper motor concept.

## 5.2 Electromagnetic Lorentz Force Actuators

I ultimately decided to use Lorentz force actuators for these millimeter scale assemblies. The benefit of Lorentz force actuators over other electromagnetic actuators is that they can generate constant forces over large strokes and are capable of high-bandwidth operation. Lorentz force actuators are also appealing because there are no attractive forces between the forcer and the mover. This means that there is no cogging forces and output force is exactly proportional to input current.

The actuator components span two cells of the lattice geometry (Figure 5-3). This reduces the volumetric overhead associated with the integration of the coil and the magnetic core and means the force producing components can take up a larger portion of the overall volume. I designed the actuators to maximize their force over a 2 mm stroke. Like all other parts in this assembly system, the actuator components are designed to be vertically assemble-able and interface with the lattice using the same press-fit connections.

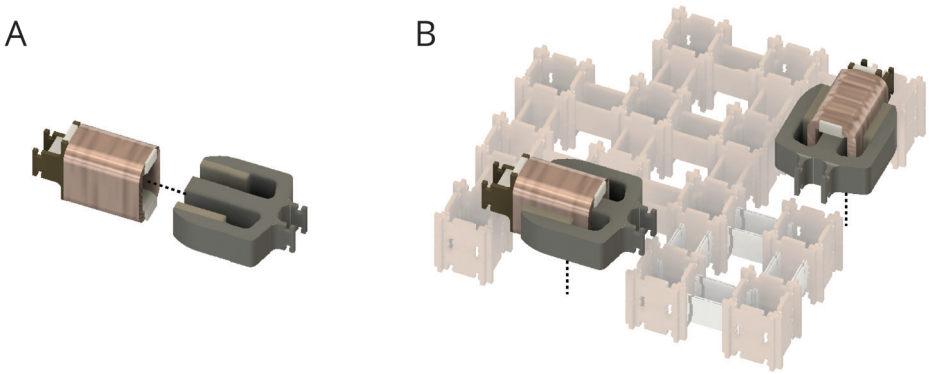



Figure 5-3: Diagram showing how the Lorentz force actuator fits into the assembled lattice. The actuator is composed of a coil part and a magnet part (A). It is assembled using the same press-fit connectors and takes up two cells of the lattice (B).


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

The actuator itself is composed of two parts: a coil part and a magnet part (Figure 5-3). The coil part consists of a frame around which a coil is wound. Two variants of the actuator have been developed. An earlier version with 80 turns of 34-AWG wire and, after refining the winding process, a later version with 320 turns of 40-AWG wire. The magnet part is composed of two neodymium magnets and a permeable core with the magnets oriented such that two opposing flux regions are created between the magnets and the central core. The coil is free to slide within the magnet part such that current flowing through the coil interacts with the magnetic flux in these gaps and produces a transverse force that is orthogonal to both the current and the flux. The detailed design and dimensioning of the actuator was based off of the modeling and experiments discussed in the next section.

In this section I will detail the modeling used to understand and predict the performance of this actuator, the experimental validation and characterization of the actuator, the production methods used to fabricate the actuator in small batches, and contextualize the performance of the actuator amongst other millimeter scale actuators in recently reported research.

### 5.2.1 Modeling

#### Magnetic Circuit Model

Voice coil actuators provide relatively high energy densities and produce constant forces over large strokes. In comparison with more traditional iron-core linear motors, voice coil actuators are less energy dense because of the necessarily large air gaps inherent in their design. They benefit, however, from having no inherent attraction between the forcer and mover and, as a result, have no cogging in their motion. This means that their force can be constant over a large stroke and their force varies linearly with the applied current. They are also more efficient, especially at high frequencies, because the iron is not cycled through its hysteresis loop. We can calculate their force from the standard Lorentz force equation for a current carrying wire in a magnetic field:

$$F = NIBL$$

where $N$ is the number of turns of the coil, $I$ is the current in the coil, $B$ is the flux density in the air-gap, and $L$ is the length (out-of-plane) of each coil turn that interacts with the magnetic field.

To calculate the flux density in the air gap we can use magnetic circuit analysis to get a first-order estimate and help establish intuition for the design space. We start with the assumption that there are no currents within the iron core so magnetomotive force sums to zero:

$$H_m t_m + H_g t_g = 0$$

and also assume an isotropic linear magnetic medium in the gap:

$$H_g = B_g/\mu_0$$

rearranging, we get,

$$H_m = -B_g t_g/(\mu_0 t_m)$$





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Then, assuming straight-line magnetization,

$$B_m = B_r + \mu_0 H_m$$

and flux in equals flux out,

$$A_m B_m - A_g B_g = 0$$

we get the flux in the gap to be (with an added loss coefficient $\eta$):

$$B_g = \frac{B_r \eta}{\frac{t_g}{t_m} + \frac{A_g}{A_m}}$$

where $B_r$ is the residual flux density of the magnet, $\eta$ is a loss coefficient which accounts for leakage flux, $t_g$ is the gap thickness, $t_m$ is the magnet thickness, $A_g$ is the area of the gap and $A_m$ is the area of the magnet. For these kinds of voice coil actuators in which the air gap is significant, the loss coefficient may be as high as 20-30%. Based on this analysis, estimations of the force can be made by plugging in the flux density of the air gap into the Lorentz force equation above. This is summarized in Table 5.1.

Table 5.1: A summary of the geometrical parameters and analytical force estimates for the 80-turn Lorentz force actuator.

| Parameter | Value |
|---|---|
| $t_g/t_m$ | 1.872 |
| $A_g/A_m$ | 1 |
| $B_r$ | 1.4 |
| $NI$ | $80 \times 0.6 = 48$ |
| $L$ | 2 mm |
| $F$, $\eta = 1.0$ | 46.8 mN |
| $F$, $\eta = 0.75$ | **35.1** mN |

**Numerical Model**

This model, however, over-predicts the force produced by the actuator since it neglects a number of losses inherent in the system. It assumes that all of the flux that leaves the magnet neatly re-enters the magnetic core without straying or folding back on itself. The model also neglects saturation of the iron. If the arms of the magnetic core are made insufficiently thick, the iron reaches the saturation region of the material's BH curve and limits flux density in the gap. This model also neglects eddy current losses in the iron, which become significant at high frequencies.

For a better estimation of the performance of this kind of actuator, I use a multiphysics modeling software like COMSOL. In COMSOL, I specify nonlinear magnetic materials to account for saturation, assign remnant flux densities to permanent magnets and calculate the force based on current flowing in regions of magnetic field (Figure 5-4).

I model only half of the voice coil actuator and use a magnetic insulation boundary condition to add a symmetric constraint. This forces all flux to be parallel at the mid-plane of the actuator and all currents to be perpendicular. While this assumption is not perfectly accurate, it is reasonable given the symmetry of this design.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

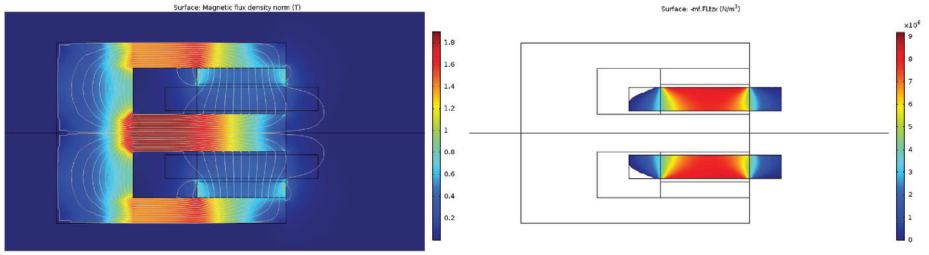



Figure 5-4: Magnetic field modeling of the magnet core and coil. (Left) Shows the flux density within the magnet core. (Right) Shows the force contribution within the coil cross-section.

There are many parameters that can be adjusted in both of these models that effect the resulting force prediction. Ultimately, however, we care about how well that simulation matches the physical hardware. Figure 5-5 shows the results of static testing. It's clear that the testing is in line with the experimental results. As expected, the analytical model over predicts the performance of the actuator and the numerical model comes closer to matching the experimental results.

### Dynamic Modeling

Understanding about how the actuator performs dynamically can be gained by looking at the impedance of the actuator coil. I model the electrical dynamics of the voice coil actuator as an LR circuit and perform analysis using an electrical simulation tool (Spice [105]). The time constant of this circuit gives us an indication of how quickly the voltage and current within the coil will respond to changing stimuli. For the PWM drive to be efficient, the frequency should be high enough that the drop in voltage and current in the coil during the off-periods is relatively small compared with the average value.

I estimate the inductance and resistance of the actuator based on its multiphysics model. COMSOL estimates the inductance of the 80 turn (34 AWG) coil to be nominally 15 $\mu$H with the magnetic core at its intermediate rest position. The resistance of the coil is estimated to be 1.4 $\Omega$.

$$\tau = L/R = 15\ \mu\text{H}/1.4\ \Omega = 10.7\ \mu\text{s}$$

$$1/(10.7\ \mu\text{s}) = 93\ \text{kHz}$$

What this shows is that to keep the voltage drop during of the off-cycle of the PWM pulse below 63% of the source voltage, the off-period of the PWM needs to be less than 10.7 $\mu$s. For a 50% duty cycle, this corresponds to 93 kHz.

This ripple current negatively contributes to coil heating, which ultimately limits the performance of the actuator. Motor heating is proportional to the square of the root mean squared current ($I_{RMS}$). This current can be calculated with the equation:

$$I_{RMS} = I_{avg} + \frac{I_{pk-pk,ripple}}{2\sqrt{3}}$$

I evaluate the RMS current and resulting temperature rise for three different PWM scenarios:

- At 31.25kHz and a 50% duty cycle, I see an average current of approximately 700mA with a




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

peak to peak ripple of 400mA. This equates to an RMS current of 700mA + 115mA = 815mA. This 115mA represents an extra 16% increase over the average current and becomes a 36% increase in power dissipated in the coil.

- At 1kHz and 50% duty cycle, I see an output current ripple that spans the full range of 1.4A. This equates to an RMS current of 700mA + 400mA = 1.1A. This extra 400mA represents an extra 57% increase in current and results in a 147% increase in dissipated power.

- At 250 kHz with a 50% duty cycle, we see an output current which spans between 740mA and 690mA (averaging 715mA with 50mA pk-pk ripple). The RMS current is then 715mA + 15mA = 730mA. This represents an increase in current of 2% and an increase in dissipated power of 4%.

Clearly it is beneficial in terms of power dissipation to drive these actuators with higher PWM frequencies and frequencies of at least 250 kHz appear to suitable for these low inductance actuators. This testing was performed on the 80-turn coils but applies just as readily to the actuators made with 320-turn coils. Because the 320-turn coils have higher inductance, they are slightly less susceptible to losses derived from ripple current but still see benefits from being driven with high frequency PWM.

### 5.2.2 Experimental Characterization

#### Static Characterization

During static testing, I varied the current through the actuator coils from 0 to 900 mA and measured the resulting output force. The results, pictured in Fig. 5-5, show that the force of the actuator is linear with current and closely matches both analytical and numerical simulations. Because the models do not account for static friction or irregularities in the winding of the coil, they tend to over-predict the performance of the actuator. The analytical model also neglects leakage flux, which is unrealistic given the size of the air gap in the actuator.

As the currents exceed 0.75 A, the results and the prediction diverge slightly. This is likely a result of high temperatures in the coil affecting the effective flux density supplied by the magnets. Given this, I take 600 mA as the actuator's maximum steady state operating current, at which it reaches a temperature of 67 °C and produces an output force of 42 mN. This output force is enough to lift 28 nodes of the lattice or 7 other actuator components, which is sufficient to produce useful motions and forces in discretely assembled machines with multiple degrees of freedom.

#### Thermal Concerns

The Lorentz force actuator is limited by its ability to dissipate heat that is built up in the coil. At a certain point the power dissipated by the coil causes a temperature increase in the windings that's high enough to damage the insulation of the magnet wire, causing the coil to short circuit, and the actuator to fail. For the polyamide bondable magnet wire that I use, this temperature is in the region of 200 °C.

Before this temperature there are other thermal concerns that should be evaluated. Permanent magnets are sensitive to high temperatures and can become permanently demagnetized when


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

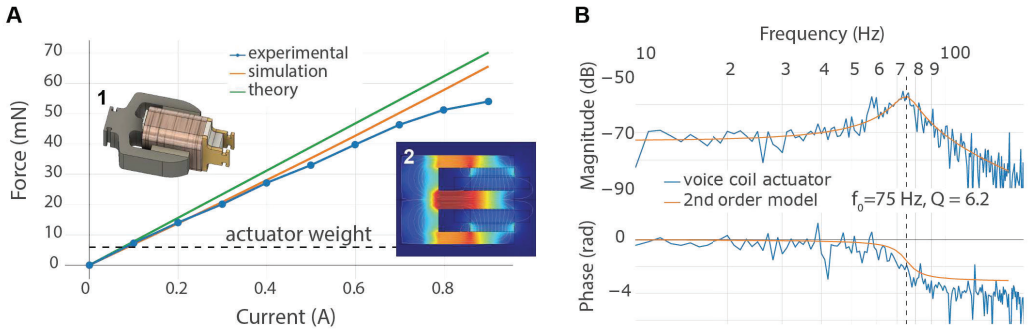



Figure 5-5: Actuator characterization. (A) Shows the results of static testing and comparisons with theoretical and numerical modeling. Current was progressively applied to the coils of the actuator component and the resulting force was measured by a milligram loadcell. (B) Shows the results of dynamic testing. A pseudo-random noise was used as the input to the actuator and the displacement was measured with a high-speed camera. Dividing the output response from the input signal resulted in the bode plot which shows the frequency response of the actuator and indicates it has a natural frequency of 75 Hz when constrained by an assembled parallelogram linkage.

operated improperly above their maximum temperature rating. This maximum temperature is dependent upon the geometry of the magnet and the surrounding flux guides, which can be represented by a permeance coefficient. In the case of this Lorentz force actuator, the permeance coefficient is defined with the following equation:

$$P_c = \frac{t_m}{t_g C_\phi}$$

where $t_m$ and $t_g$ are the thickness of the magnet and gap respectively and $C_\phi = A_{mag}/A_g$ which, in this case, is just 1. From Table 5.1, $t_g/t_m = 1.872$ so $P_c = 0.53$.

This permeance coefficient can also be described as the ratio of the flux density to the magnetic field at a given operating point. Using COMSOL, I looked at the average field and flux density in the permanent magnet. This method estimates a permeance coefficient closer to 0.81 which reflects a slightly more temperature resilient actuator.

Then, using the demagnetization curve, a load line can be overlaid starting at the origin with slope equal to the permeance coefficient (Figure 5-6). It's clear that above approximately 70 °C, the load line will be hitting the "knee" in the demagnetization curve and risk irreversibly reducing the residual flux density of the magnet.

To ensure that the actuators do not exceed this temperature, I use a thermal camera to measure approximate temperatures of the coil with various current levels. At 600 mA, the 34-AWG, 80 turn coil reaches 67 °C at steady state. At 800 mA, the temperature exceeds 96 °C when the current is applied for more than 30 seconds. So, this current is maintainable for short-bursts but not steady state.

While this design is acceptable for my application where I am operating in room temperature environments, it appears to be susceptible to demagnetization at high current operating points or in elevated temperature environments. A better design (in this respect) should have a thicker magnet per gap width. In brushless motors, permeance coefficients are often greater than 4 [107].

Even with this permeance coefficient, the current density that this actuator is capable of exceeds many macroscopic electromagnetic actuators. Dividing the drive current by the cross sectional area


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

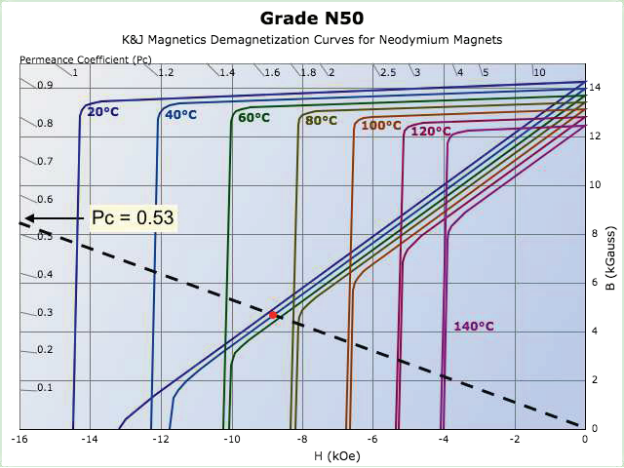



Figure 5-6: The demagnetization curve for N50 magnet from [106] with overlaid load line. Beyond 70 °C we risk hitting the knee of the curve and irreversibly reducing the residual flux density of the magnet.

of a single turn of the wire gives a measure of the current density. In this case the 600 mA operating current represents a current density of 24.2 A/mm$^2$, which is more than two times greater than the recommended maximum current density for macroscopic electromagnetic actuators [107]; this is possible because of the relative scaling of surface area and volume, which allows better heat transfer out of the coil [108].

Table 5.2: Measured Characteristics of the Actuator Component

| | |
|---|---|
| Coil Turns | 80 |
| Max. Continuous Current | 0.6 A |
| Max. Continuous Current Density | 24.2 A/mm$^2$ |
| Force (blocked mid-stroke) | 42 mN |
| Stroke (no load) | 2.1 mm |
| Resonance (incl. parallelogram linkage) | 75 Hz |
| Overall Mass | 511 mg |
| Energy Density | 89 mJ/kg |

### Dynamic Characterization

To measure the dynamic performance of the actuator, I supply a pseudo-random voltage across the actuator and use a high-speed camera to measure its response. The coil component of the actuator is rigidly fixed while the magnetic component is constrained by an assembled parallelogram flexure linkage, which approximates linear motion for small displacements. Dividing the frequency response of the output displacement by that of the input voltage, results in the transfer function that describes how the output relates to the input over a range of frequencies. This data, presented in Figure 5-5B, is described well by a second-order spring mass damper model with a 75 Hz natural frequency and a quality factor of 6.2. This model is useful in predicting the performance of the actuated system




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

over a range of frequencies and gives an estimation of the bandwidth of the actuator. The static and dynamic performance of the actuator component is summarized in Table 5.2.

I use an A-Star microcontroller (Pololu) to generate the actuation signals. A 16-bit LFSR is used to generate white noise which is output as PWM and amplified by an H-bridge (DRV8838). To measure the displacement I use the high-speed camera (Chronos), processing the data with motion tracking software (Tracker), and synchronize the data collection with the actuation signals.

Replacing the parallelogram linkage with a rigid strut and performing this same analysis gives an upper bound for the maximum bandwidth that this assembly framework is capable of supporting. The displacements of this rigid structure are two orders of magnitude smaller than that of the actuated mechanism and so a laser displacement sensor (MicroEpsilon) is used to measure the displacement over time. The actuator is controlled to perform a frequency sweep and the magnitude of the displacement at each frequency is recorded. The measurements, shown in Figure 5-7, show that the structure has a first resonant mode of 620 Hz. This is in close agreement to the simulation result of 692 Hz without any parameter tuning. This frequency is not a clearly defined upper bound, but instead represents a rule of thumb that structures assembled in this way should be usable up to near this frequency without worry about exciting unintended resonant modes.

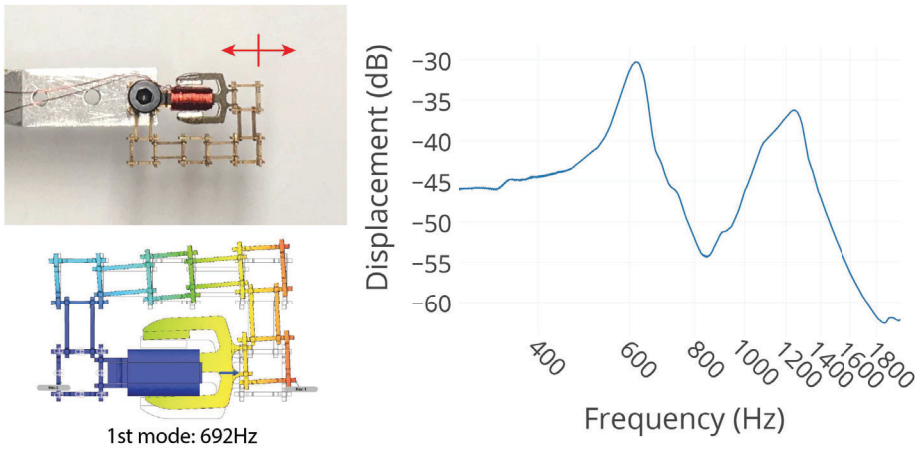



Figure 5-7: By using an actuator to excite a rigidly assembled structure it is possible to determine the resonant modes. The experimental results show a first resonant mode at 620 Hz, which closely follows the simulation result of 692 Hz.

### 5.2.3 Production

While the actuator components are more geometrically complex than the basic structural parts, they are still designed to be relatively easily mass produced and are made using a combination of direct-write processes including wire electrical discharge machining, laser micromachining, and CNC milling (Figure 5-8).

The actuator parts are composed of two part-types: the coil part and the magnetic part. The core of the mandrel that supports the coil is first machined on a Roland MDX-540 CNC milling machine. Conductive 254 $\mu$m thick phosphor-bronze layers are cut as a stack with the wire-EDM


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

and these are then bonded to the core using the same laminate process described above. I then wind the coil on the mandrel using a benchtop coil winding machine (Adams Maxwell). This process is surprisingly fast and I am able to wind a 320 turn coil in less than 60 seconds. Coils of this size are regularly produced in large quantities, relatively inexpensively using surface mount inductor coil production techniques [109].

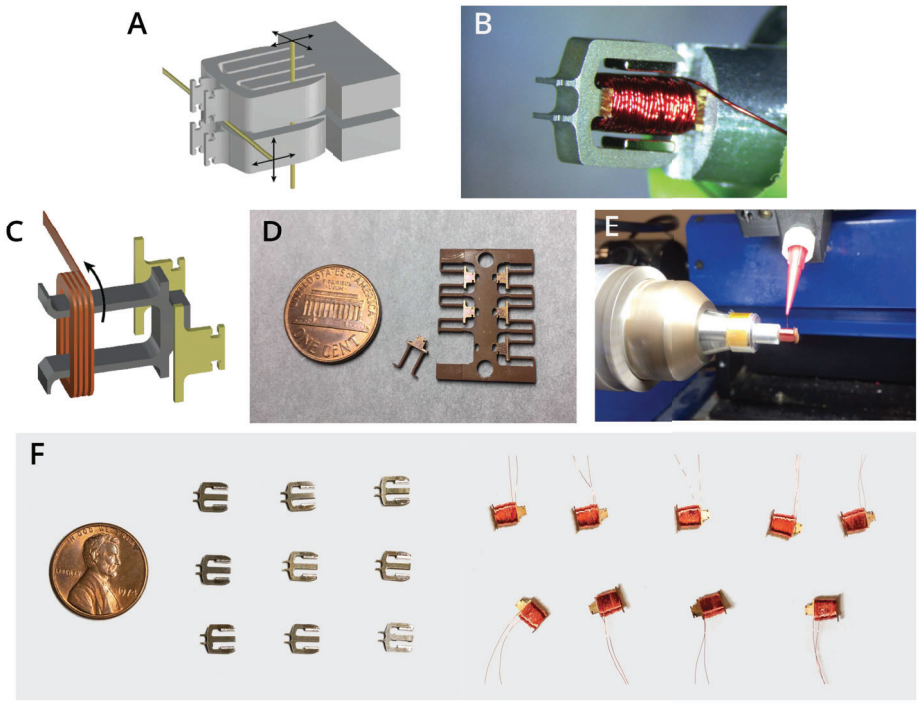



Figure 5-8: I developed a number of efficient methods to prototype these parts in small batches. The magnet part is cut in two orientations by wire-EDM (A,B). The coil part is produced in a two step process in which a mandrel is first made by laminating connectors onto a composite core and then wound in a coil winding machine (C,D,E). Using these techniques I produce 9 actuators in an afternoon (with the potential for many more).

The magnetic core is cut in batches from bar stock in the wire-EDM. Since the core features intricate geometry on multiple sides, I cut one profile, rotate the stock 90 degrees, cut another profile, and then cutoff all of the parts at once. Using this process enables the production of an array of nine magnetic cores in a little over an hour. Rare-earth permanent magnets (N50) are then fixed to the inside of the magnetic core arms with cyanoacrylate adhesive (Loctite 401). These production processes are illustrated in Figure 5-8.

### 5.2.4 Discussion

The bandwidth measured in this chapter is largely a result of the compliance of the mechanism elements. This bandwidth should be viewed as a design parameter that can be increased based on the requirements of task (up to the 620 Hz mode of the structure). This can be done, for example, by


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

simply stiffening the flexures using a thicker flexible membrane, shorter flexure length, or additional flexing elements.

The actuator component developed here compares well against other millimeter-scale actuators that have been recently reported in research. I compiled data regarding actuator force, mass, stroke, and bandwidth and plot the normalized actuator force (per weight) against maximum stroke (Fig. 5-9) for a number of piezoelectric, electrostatic, and miniature electromagnetic actuators. Wherever possible I include the mass of the whole actuator (stator and mover) as well as the mass of the motion constraint. I did not include shape memory alloy based actuators in this comparison because while they are capable of extremely high force densities, they were deemed to be too inefficient, too low speed, and too susceptible to ambient temperatures for the applications targeted in this work.

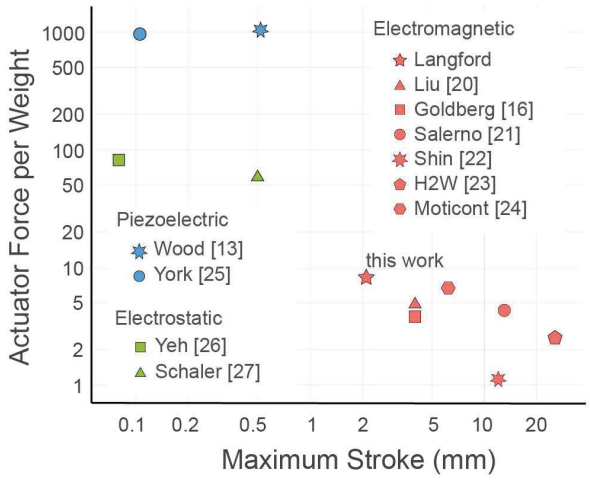



Figure 5-9: Comparison of normalized force versus stroke with other research and commercial millimeter-scale actuators. This actuator interpolates between existing small scale electromagnetic actuators and the force-dense but stroke-limited electrostatic and piezoelectric actuators.

The actuator component developed in this work has the highest normalized force (blocked force per weight) of any of the comparably sized electromagnetic actuators presented here [110] [103] [111] [112] [113] [114]. Piezoelectric [100] [115] and electrostatic [104] [116] actuators produce more force per mass but are more limited in their available stroke.

Even so, this actuator pails in comparison to biological muscles in terms of force density. Biological muscles have a specific actuation stress (pressure per mass) that is 25 times the maximum limit for an electromagnetic actuator [117]. On the other hand, biological muscles are relatively slow compared to these kinds of electromagnetic actuators with most working at a rate of up to approximately 7 Hz (with some exceptions such as small flying insects [118] [119]). Developing topologies and mechanisms that trade bandwidth for force density could be worth exploring in more depth. In chapter 6, an embodiment of this concept is developed as a walking motor. Section 11.2 does a more thorough analysis of this force density comparison.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 6

# Walking Motor

Given the structure, mechanism, and actuation capabilities developed and covered in previous chapters, it is possible to start assembling machines that integrate these functionalities to perform useful behaviors.

I built a Discretely Assembled Walking Motor (DAWM) to demonstrate the integration of structure, mechanism, and actuation. The DAWM system takes small cyclical steps to produce long range motion of a sliding or rotating element. This principle of locomotion is most commonly used with piezoelectric actuators [120] for applications such as nanometer-precision stages [121] and focusing motors in DSLR cameras [122]. Here, I use the voice coil actuator components developed for the assembled structures to enable the same kinds of motion with larger displacements and lower voltages.

## 6.1 Design

The walking motor is composed of five part types: structural nodes, rigid and dual-hinged struts, and magnetic and coil part types (Figure 6-1). The two actuator components are oriented perpendicularly to one another and the output of the actuators is coupled to a motor tip through multiple four-bar parallelogram linkages, which distribute the two degrees of freedom at the tip into a single degree of freedom at each actuator. As the motor tip is driven cyclically, it engages with a grooved sliding or rotating element. In this design, the motor tip consists of a 0.5 mm cylinder, which provides quasi-kinematic mating with the triangular grooves of the rotor. The triangular grooves are spaced 0.75 mm apart and correspond to a segment of the approximately circular trajectory of the motor tip. This geometry is designed to allow for the correction of motor-tip positioning errors within $\pm 0.15$ mm. The motor works with a variety of different surfaces including ones that are smooth; however, the grooved surface employed here provides the highest repeatability.

## 6.2 Characterization

To drive the actuator components within the walking motor, I use a custom control board which features an ATmega32u4 microcontroller and four DRV8838 H-bridges. A supply voltage of 1.4 V is provided to the board, which then modulates the current in the voice coil actuators with PWM. The




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

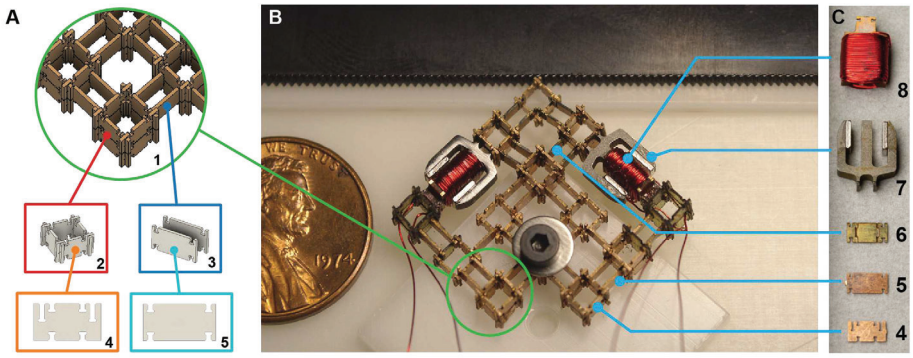



Figure 6-1: The walking motor is assembled from five part-types: rigid nodes and struts, flexible struts, magnet parts, and coil parts. Here, a single layer walking motor engages a linear slider.

current through the voice coils is monitored by a 0.1 Ω sense resistor and a benchtop oscilloscope (Tektronix MDO3024).

For the rotary motor application, the walking motor is attached to a 3D printed fixture (Sindoh) which enables the distance between the motor and the rotor to be controlled precisely using a fine-adjustment screw (Thorlabs). The grooved rotor is laser cut or wire-EDM cut and press-fit on a 608 skate bearing, which provides a low friction rotational degree-of-freedom. This experimental setup is pictured in Figure 6-2.

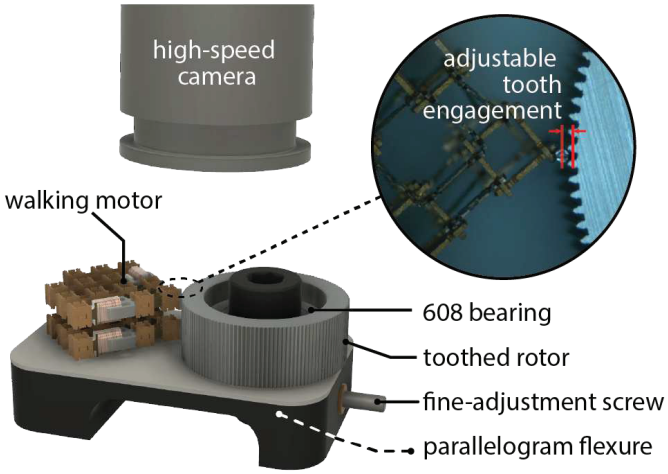



Figure 6-2: The experimental setup used to characterize the walking motor. A fine adjustment screw deflects a flexure to adjust the spacing between the motor and the toothed rotor.

### 6.2.1 Speed and Repeatability

Each actuator of the walking motor is driven with 90-degree phase-shifted sinusoidal or trapezoidal waveforms, corresponding to a roughly circular trajectory of the motor tip. I characterize the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

repeatability of the walking motor across a range of stepping frequencies, by taking 10 steps at each frequency and computing the average velocity at each. The frequency is swept up and down, increasing and decreasing the step rate a number of times, to get a sample size of six at each tested frequency.

To measure the speed of the rotor, a standard video camera is used to track a feature on the rotor to extract the position over time and measure the effective velocity. I use a high-speed camera (Chronos) to inspect tooth engagement with the rotor at high step-frequencies.

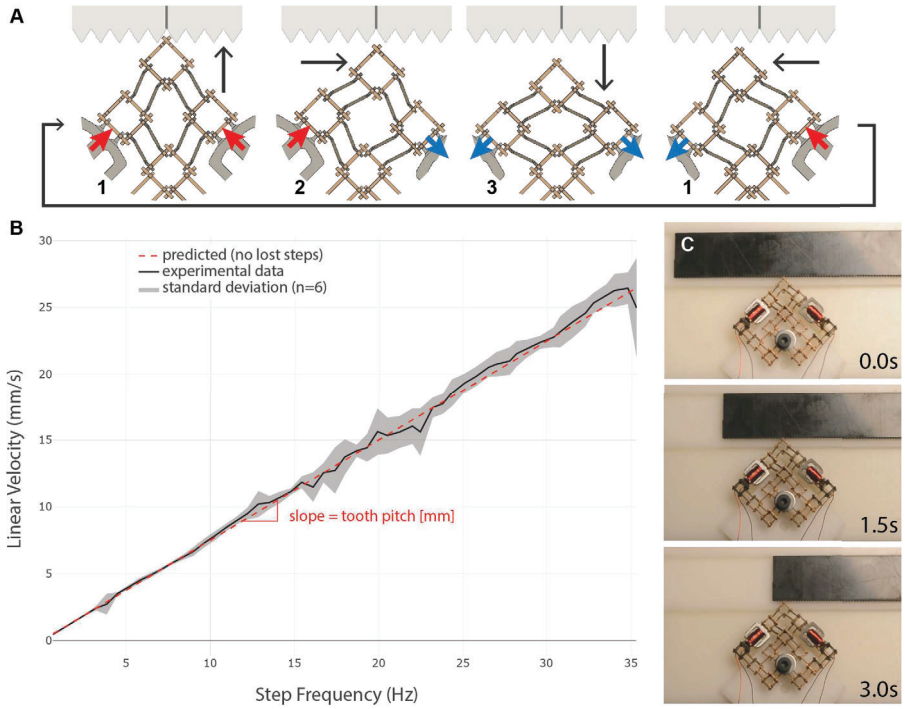



Figure 6-3: Speed and reliability testing done on a single-layer walking motor illustrates the series of steps during one actuation cycle (A) as well as the progression of a slider over three seconds (C) and the resulting velocity of the mover/rotor (B). The results show good agreement to the idealized relationship between frequency and velocity.

The results (Figure 6-3) show good agreement to the predicted performance. Below 12 Hz, the difference between measured and predicted velocities is negligible. Between 12 Hz and 34 Hz, there is more variability in the relationship between step frequency and velocity with the largest standard deviation being 25% of the mean. However, the overall slope remains consistent with our predictions. This indicates that at the higher step-rates the motor is just as likely to take a double-step as it is to miss a step. I hypothesize that step variability occurs because the motion amplitude increases as the step-rate approaches the resonant frequency of the walking motor and causes the motor to occasionally skip a tooth. Beyond 35 Hz the velocity drops dramatically as the steps become very erratic and intermittent, indicating a maximum open loop speed of approximately 25 mm/s.

This 25 mm/s maximum exceeds many commercially available piezo walking motors [121]. The corresponding 35 Hz maximum frequency is limited, in part, by the electronic drive circuitry. The


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

actuator uses 34-AWG wire, resulting in a low inductance coil (15 $\mu$H) that experiences current ripple losses when driven directly by an H-bridge with insufficiently high frequency PWM (1kHz, in this case). Decreasing the wire diameter to 40-AWG and increasing the PWM frequency to 31.25 kHz would reduce the current ripple by 88% and the resulting dissipation by 70%, potentially enabling actuation rates up to the 75 Hz bandwidth of the actuator-flexure combination.

The actuation rate could be further increased by stiffening the flexure degrees-of-freedom. These flexures were designed to be highly compliant to enable long strokes to suit a wide variety of applications. Given the electrical time constant of the actuator coils ($\tau$ = L/R = 100$\mu$s), actuation rates of several hundred cycles per second (up to the bandwidth of the structure) should be possible at the expense of the actuator stroke. While the DAWM system has not been tested to fatigue, prior research has shown that the flexural hinges used here can be made to survive millions of cycles before failure [98].

### 6.2.2 Force Additivity

In the case of a single walking motor, it is impossible to generate continuous force since the motor tip needs to disengage from the rack in order to move to the next tooth. However, using multiple walking motors in a stacked configuration, as depicted in Figure 6-4, enables continuous force production. In order to ensure that one tooth is always engaged with the rack, it is advantageous to be able to modulate the duty cycle of the motor-tip trajectory; that is to say, I control the time the tooth is in the engaged position relative to the disengaged position. I do this with a parametrically defined trapezoidal waveform, which is described by two parameters: on-time (engaged in rack) and off-time (disengaged from rack). Within a given walking-motor layer, the phase offset is always set to 90-degrees to ensure the trajectory of the motor tip is roughly circular. The phase between layers can be adjusted to modulate the degree to which the multiple walking motors are engaged with the rack simultaneously. At one extreme, in which there is no phase offset between walking motors, I expect the stacked walking motor to have the same behavior as a single walking motor but with twice the amount of force with each tooth engagement. At the other extreme, in which the two motors are out of phase with a phase offset of 180-degrees, I expect to see overlapping engagement of the motor tips and expect a more continuous force output from the stacked motor.

To characterize the effects of phasing multiple walking motors, I measured the blocked force of a stacked two-layer walking motor while driving at a 1 Hz step rate to ensure sufficient resolution to resolve the full force profile (Figure 6-4). To measure the blocked force of the motor, I use a milligram loadcell with a 24-bit ADC and amplifier (TI-ADS1231). The blocked force varies depending on where the motor engages the tooth of the rotor, which, in turn, is affected by where the motor and rotor are positioned relative to the load cell. To ensure consistent results, I take data wherever the peak force is maximum for a given trial. I determine this position by plotting the force output in real-time and slowly adjusting a micrometer to move the motor and rotor assembly relative to the loadcell.

When the two motors are driven in phase, their peak force is approximately twice (70 mN) that of the single motor (30 mN). When the two motors are driven out of phase, they produce a more uniform force which varies between 20 mN and 50 mN. The magnitudes of these forces are on par with the maximum blocked force of the individual actuators themselves. In the single motor case, 79% of the maximum blocked force of a single actuator is translated to the rotor at its peak force.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

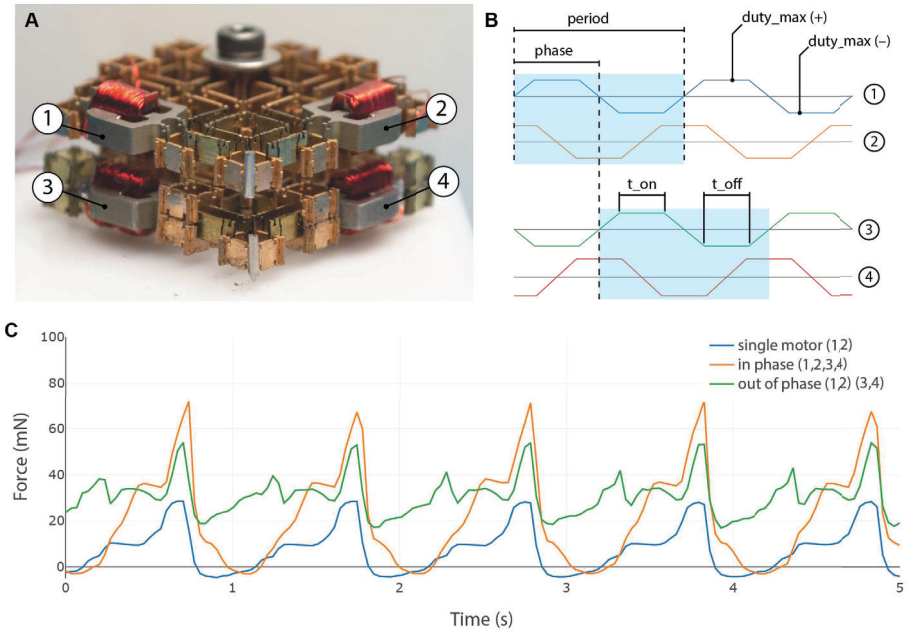



Figure 6-4: The force output of the walking motor (A) driven by a trapezoidal waveform (B) out-of-phase and in-phase (C). Continuous force production is possible when driving multiple walking motor layers out of phase with each other.

In the two-motor case, 92% of the maximum blocked force of two single-actuators is translated to the rotor at its peak force. This indicates that the assembled transmission is relatively lossless, at least in terms of static force transmission.

## 6.3 Discussion

This chapter has demonstrated the ability to assemble long range linear and rotary motion from individual discrete parts. Stacking two walking motors has been shown to both increase the force output and enable continuous force production. This could be extended with the use of even more walking motors to enable greater force production. A question remains about the possibility of using multiple walking motors to increase the speed of actuation. Doing so would likely require very low duty cycle strokes such that the motor tip is only engaged for a very short period of time. This is how biological muscles amplify the speed of relatively slow (≈5 Hz) individual molecular motors to the 1500 Hz effective stepping rate [123].

The efficiency of a walking or inchworm motor in most cases is not high. This can be evidenced by looking at the drive cycle and realizing that only a fraction of the time is the input energy converted directly into output work. (1) Current flows into locking actuator, producing electromagnetic force proportional to the current. This force balances with the reaction force of the flexural mechanism, and the motor tip displaces. (2) At some point the motor tip engages the rack of the moving element and the flexure displaces no further. (3) The drive coil now start producing force which goes into


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

displacing its own flexures and accelerating the moving mass. (4) Once it reaches its maximum stroke, the current in the locking actuator is reduced and eventually reversed to disengage from the rack. The drive coil is then also reversed and the cycle repeats. Of these four actuations only one (#3) goes into producing useful work.

Techniques exist to improve the efficiency of these actuators. First of all, these kinds of motors can operate much more efficiently if driven at resonance since the actuators are effectively used to pump energy into the dynamic system [122]. Rather than quasi-statically overcoming the flexure stiffness, the flexure stiffness is excited, keeping energy in the system. Other strategies such as using flexible drive elements have been shown to be capable of achieving 8% efficiency for electrostatic inchworm motors. However this strategy requires a shaped ratchet-pawl interface that restricts motion to a single, irreversible direction [124].

To constrain the long range motions, I use commercially available ball bearings and sliding surfaces for rotary and linear motion, respectively. The part set developed in this assembly framework is capable of making linear and rotary constraints, but would require custom part-types in order to interface with the motor-tip and so I leave that for future work.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 7

# Walkers

Once I was able to assemble the walking motor. I wondered if I could make something that walks and locomotes. To do this required the addition of a foot that can attach and detach from a surface. Electropermanent actuators are ideally suited for this kind of latching actuation on ferrous surfaces.

## 7.1 Electropermanent Feet

Electropermanent actuators are made by biasing a hard magnet with a soft one (Figure 7-1). A coil is wrapped around both (or, alternatively just the soft magnet). A pulse of current through the coil is able to flip the polarity of the soft magnet. This either confines the flux to circulate only between the magnets (when the polarities are mismatched), or allows the flux to flow out through the poles of the actuator to create attractive force with the permeable surface.

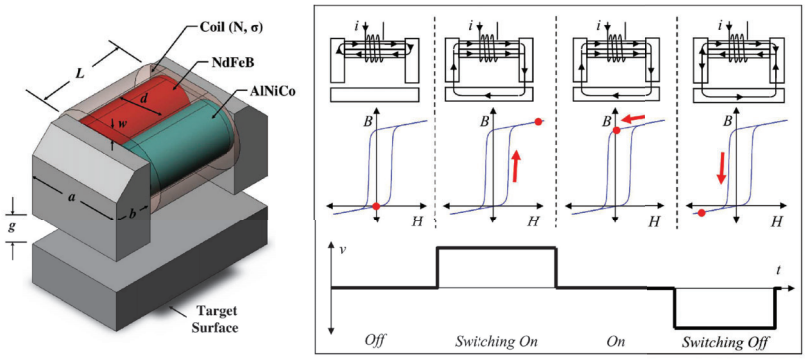



Figure 7-1: Reproduced from [58]. (Left) Construction of electropermanent magnet. (Right) Operation of the electropermanent magnet through the hysteresis loop of the BH curve.

Electropermanent actuators are better suited for this task than electromagnets for two primary reasons. The first is that below a frequency threshold electropermanent actuators are far more energy efficient. This is because they hold their state without any current. To attach to a surface a pulse of current is used to flip the soft magnet and then the actuator functions just like a permanent magnet would. The second reason is that the electropermanent magnet is a high reluctance magnetic




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

circuit by design. This is because the permeability of the hard and soft magnets is much closer to air than it is to iron and so their presence in the circuit is essentially a large air gap. This means that an air gap between the actuator and the surface contributes far less to a reduction in attractive force than it would in an equivalently sized electromagnet (Figure 7-2 A).

I had initially tested making a walker with electromagnets rather than electropermanent magnets and found that the attractive force was highly sensitive to very small air gaps between the foot and the surface. This manifested in the walking robot being very sensitive to the orientation of the wires that were providing power to the actuators. A walking robot using electropermanent feet does not suffer from this same sensitivity.

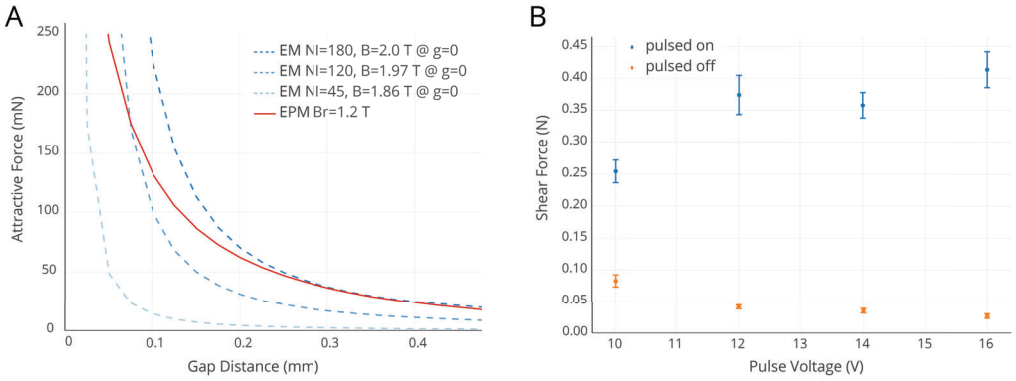



Figure 7-2: Electropermanent magnets are far less sensitive to large air-gaps and produce more attractive force than an equivalently sized electromagnet (A). The degree to which the electropermanent actuator is switched can be controlled (B). For this application, the 16 V actuation is beneficial for the biggest change in frictional force.

I tested the shear force exerted by the electropermanent actuator in its on and off state at a number of different voltage levels from 10 V to 16 V (Figure 7-2 B). The pulse duration was set to 300 $\mu$s. I found little difference in performance by increasing this pulse duration any further. I used a custom tabletop materials testing machine with a milligram load sensor to perform these tests as I anticipated very small ($<$1 N) forces.

The actuator is placed on a ground steel surface and a post is used to couple the actuator to the load cell. The steel piece is attached to the moving table of the testing machine. During the test the steel is advanced towards the loadcell and the maximum force seen by the loadcell is recorded. This provides measure of the dynamic friction between the ground steel surface and the actuator.

Testing the electropermanent feet at multiple different pulse voltages shows how the magnitude of the current pulse effects the performance of the foot. It's clear that a greater difference between the on and off state is achieved for higher voltages. In particular, it's clear that 16 V (the maximum voltage permitted by the H-bridge I'm using) should be used for optimal performance.

It is also apparent that the actuator does not switch fully off since shear force does not drop fully to zero in the off state. This could be an error in the construction leading to the soft magnet not fully switching polarity. For example, the winding may not be evenly distributed or leave some portion of the magnet exposed. While it would be ideal that the force in the off state drops perfectly to zero, the almost 20x increase in shear force between the on state (0.425 N) and off state (0.025


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

N) is entirely sufficient to provide attachment and detachment for an assembled walker.

I also measured the normal force exerted by this actuator. On a ground steel surface the normal force measures 2.23 N ±0.24 N (n=6). This is enough force to support the weight of over 400 actuator components (511 mg).

## 7.2 MOTILE – Modular Tiny Locomotion Element

The walker, MOTILE, is assembled from six part types: rigid struts, rigid nodes, 2-DoF struts, a magnetic part, a coil part, and electropermanent feet. A central Lorentz force actuator provides the motive force and the relative order of when the electropermanent feet are switched determines the direction of motion.

I tested the MOTILE on a number of different ferromagnetic surfaces including ones with a thin layer of lower friction material (Kapton). I ultimately found the best performance had the electropermanent feet in direct contact with a ferrous surface with a thickness greater than or equal to 1.5 mm. This ensures that the foot provides the full attractive force that it is capable of. I also found that adding a small plastic spring helped accentuate the difference in friction of the on and off states of the feet and was beneficial for consistent locomotion. These two springs (one per electropermanent foot) have the effect of reducing the friction between the unlatched foot and the surface without significantly effecting the friction force of the latched foot and the surface.

In addition to testing a range of different surface materials, I also tested the robot in a number of different orientations. Because the potential shear force (>0.4 N) and the normal force (2.2 N) of the feet are both much greater than the weight of the robot (≈0.02 N), the robot is able to walk up vertical walls as well as upside down on inverted surfaces (Figure 7-3).

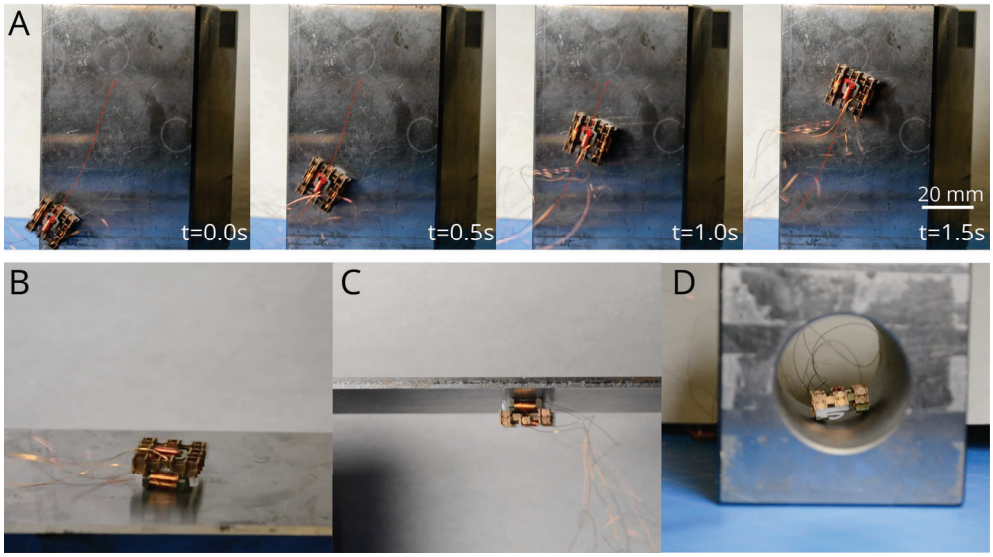



Figure 7-3: The walker is able to walk on surfaces of a variety of orientations including up vertical walls (A), on horizontal surfaces (B), upside down (C), and on non-flat surfaces (D).

I used motion tracking software (Tracker) to measure the speed of the MOTILE in a variety




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

of orientations. In analyzing the speed of the walker as it climbs a vertical ground steel block, the walker takes a trajectory that's roughly 60-degrees from horizontal. The MOTILE is driven at approximately 45 Hz. The speed is very consistent and measures approximately 36.2 mm/s going up and 44.2 mm/s going down. Doing the same on a flat surface, I measure an average speed of 37 mm/s over 5 trials.

It helps to translate these speeds to a "body-length per second" equivalent in order to compare this performance to other mobile climbing robots. Given its 15 mm body length, it climbs up the block at 2.4 BL/s and down at 2.95 BL/s. It moves on a flat surface at 2.46 BL/s. Comparing these values to others reported in recent literature reveals, surprisingly, that this little robot is among the fastest vertical climbing robots. That being said, there are certainly many caveats and limitations with this statement. For one, this walking "robot" is driven entirely off-board by external control electronics and as such carries less weight than robots that are untethered. Another severe limitation of this robot is the need for a ferromagnetic surface. However, this is certainly not the only robot to use magnetic forces for climbing and others have argued that many desirable applications for mobile climbing robots would be satisfied with the use of magnetic forces [125][126][127][128].

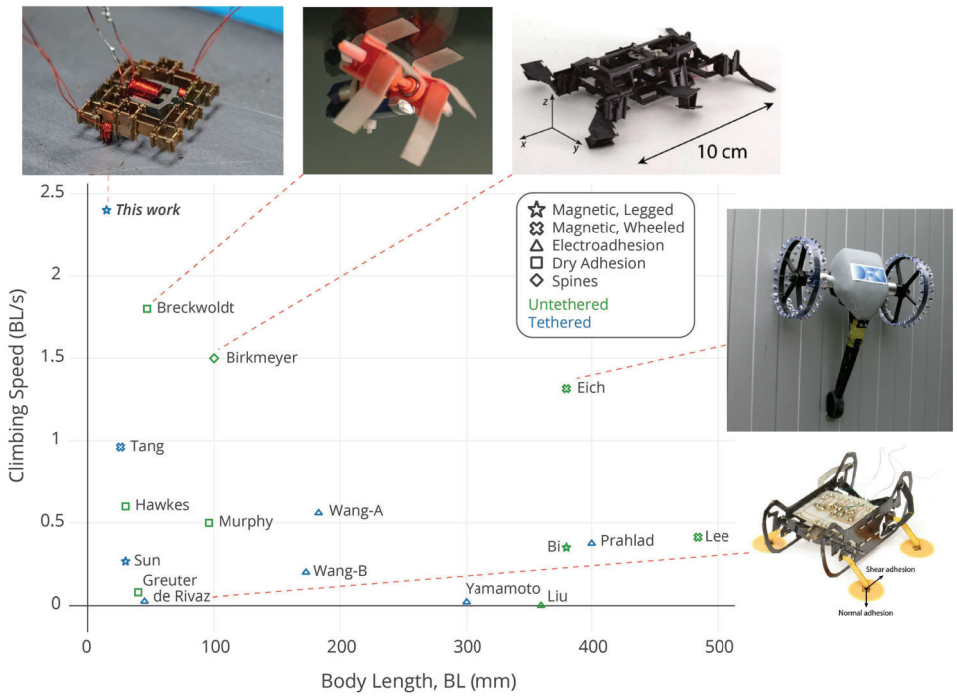



Figure 7-4: Comparing the climbing performance of the single degree-of-freedom MOTILE walker to other vertical climbing robots. This "robot" represents both the smallest and fastest (in body-lengths per second) of any robot in recently reported research but is not able to turn and doesn't contain onboard power or control.



361
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 7.3 Discussion

While I've demonstrated the assembly of a single degree-of-freedom walker, MOTILE, that can only move forward or backward, these modules can be composed to build positioning devices with the ability to move in multiple directions. Two MOTILE's placed side-by-side can be used to move forward and backward as well as turn left and right (Figure 7-5). Four MOTILE's arranged in a cross-shaped circular array can be used to translate in any direction as well as rotate about the plane.

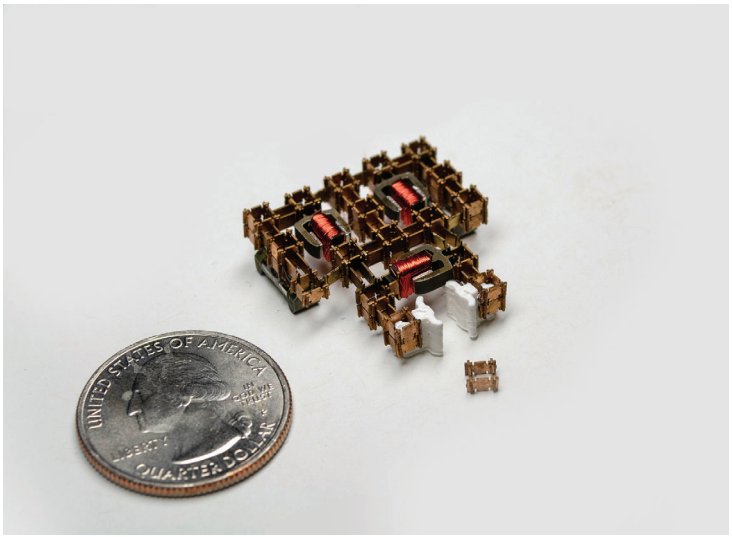



Figure 7-5: Two MOTILEs along with a gripper can be combined to form a mobile assembler that can pickup and place the same building block parts it is made of. This concept is developed further in Chapter 10.4

The motion of an individual MOTILE relies on a friction interface with the surface and so the exact size and direction of each step can vary from step to step. This results in motion that is sometimes not perfectly straight or that varies in speed (for example, climbing up versus down the steel block). It would be beneficial to eliminate these sources of uncertainty in order to build machines that can position precisely and repeatably without requiring an additional sensing mechanism. This could be accomplished with the use of a structured surface that attracts each step to a potential minimum. This structuring could be done with a physically grooved surface. I began testing versions of the feet and grooved surfaces that could be used in this way but encountered difficulties with the feet getting stuck in the grooves. While more complicated, one solution could be to include an additional actuator (or two) to raise and lower the feet with respect to one another. This would allow the disengaged foot to clear the top surface of the grooves before planting down in the next groove. Another potential way to structure the surface for repeatable open loop motion could be with a surface with spatially varied permeability. This strategy is what is used in linear stepper motor platens to give the motor its discrete steps. This could be fabricated relatively easily by incorporating permeable slats in a non-permeable surface and certainly seems worth exploring.

Beyond electropermanent feet, there are certainly other kinds of "feet" that could be used to


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

create locomotion. Directional friction pads, for example, would enable motion on arbitrary material flat surfaces. This strategy would only permit movement in a single direction (not bidirectionally). However, multiple walker modules could be combined to produce forward and turning motions.





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 8

# Positioning and Manipulation

Beyond the locomotion capabilities developed in Chapters 6 and 7, the ability to assemble systems that combine structure, mechanism, and actuation also enable the development of positioning and manipulation devices.

## 8.1 Positioning

Compliant flexures are often desirable in micro-positioning and scanning applications since they eliminate both backlash and friction, making open-loop positioning repeatable and reliable. These kind of systems are often powered by either electromagnetic and piezoelectric actuators. Positioning stages for precise high-speed tasks often use parallel configurations, where there is a closed kinematic chain, rather than serial configurations [129]. These tend to be higher-performance because the moving mass can be kept to a minimum [94].

A planar positioner amplifies the motion of two actuators to larger displacements of an end-effector tip. The positioner is designed such that the end-effector remains parallel to the base coordinate system through the whole range of motion. This is accomplished with the use of an extended four-bar linkage attached to one of the two arms. The linkage amplifies the motion of the individual actuators by 3.5 times, resulting in a peak to peak displacement of the end-effector of approximately 7 mm.

I conducted a frequency response measurement of the positioner system. Using a multi-functional oscilloscope (Digilent Analog Discovery 2), I measured the complex impedance of the the system from 1 Hz to 2 MHz. The results, pictured in Figure 8-1 B, clearly show a mechanical resonance around 10 Hz, representing the maximum bandwidth for controlled sub-resonant positioning. Near resonance, the displacement of the mechanism increases to 9 mm.

## 8.2 Manipulation

It is also possible to assemble mechanically actuated grippers and manipulators. I assembled a gripper that converts the linear displacement of two actuator components into a rotational motion of the gripper fingers (Figure 8-2). The $\pm 1$ mm displacement of the actuators is converted into a 4.25 mm stroke of each finger, resulting in a total 9.5 mm displacement between maximum (ungripped)





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

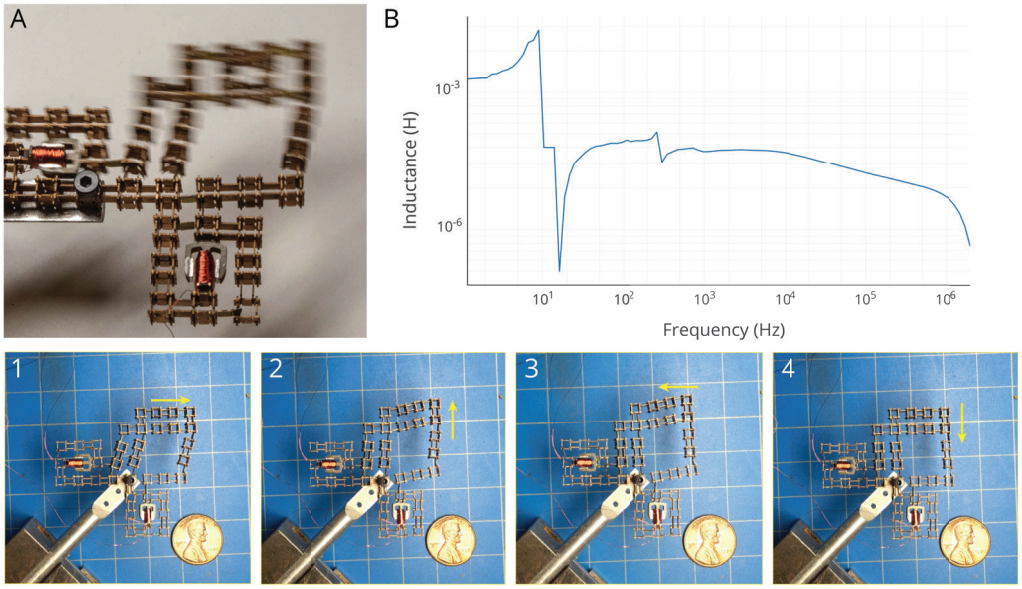



Figure 8-1: An assembled XY-positioning system (A) has a mechanical resonance around 10 Hz (B) and amplifies the motion of two actuators by 3.5 times (1-4).

and minimum (gripped) states. This enables the gripper to pick up, among other things, the parts that it is made of.

The two actuators can be coupled or decoupled. Coupling results in better open-loop coordination of the two gripper fingers. Decoupling could potentially allow for more advanced manipulations, enabling independent control of each finger.

With one extra geometric part-type, it's possible to extend this design from 2.5D to 3D. The additional part type serves to join pre-assembled orthogonal planes (Figure 8-2 C).

## 8.3 Discussion

**Positioner** The resonant frequency of the positioner is largely determined by the compliance of the flexure hinges and the moving mass of the end-effector. To increase the possible bandwidth of operation, the bronze parts could be replaced with lighter-weight parts made of fiberglass or carbon fiber, for example.

The positioner provides a good test case to evaluate the longevity of the flexure hinges since it represents a large off-axis load and requires large displacements. I excited the linkage at 20 Hz and let it run continuously for many hours. After 16 hours, an equivalent of 1.16 million cycles, one of the flexure hinges failed. This hinge was supporting a cantilever of 15 connected nodes, which represents a torsional load of approximately 0.223 N*mm divided amongst three hinges (75 mN*mm per hinge). This performance could be improved by simply building up more layers to better support the off-axis load.

Another possible way to improve the performance would be to use flexure hinges which a thicker flexible membrane. The current hinge design uses a 25.4 $\mu$m thick polyimide layer and is designed to


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

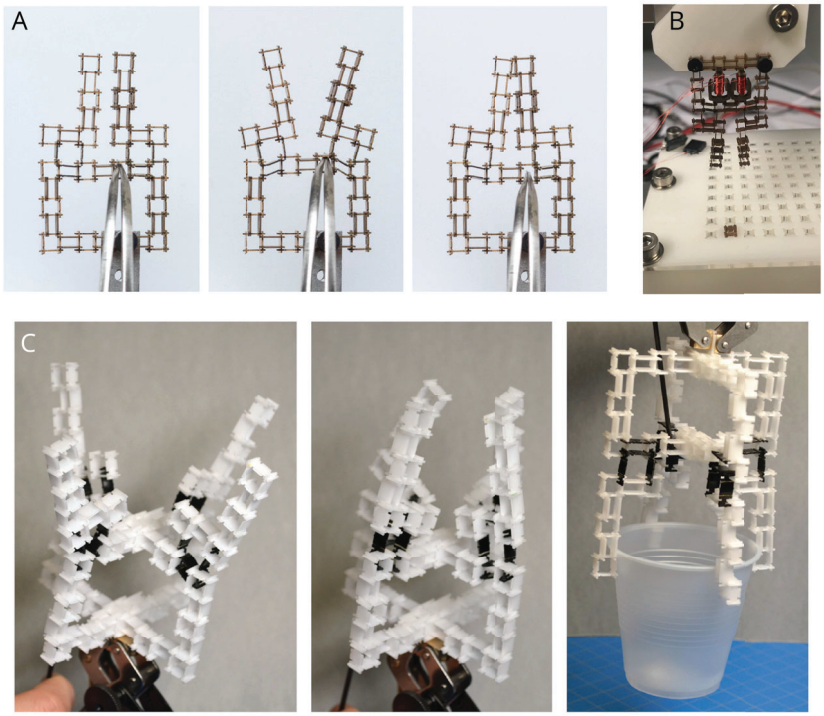



Figure 8-2: An assembled gripping mechanism (A) can be used to pickup and manipulate its own parts (B). The design can be extended to three-dimensions with an additional part-type (C).

be highly compliant. Substituting this layer with one that is 50.8 $\mu$m, or even thicker, may reduce the stroke of the mechanism slightly but would greatly increase both bandwidth and longevity.

**Gripper** In contrast with other "micro"-grippers that use piezoelectric actuation (e.g. [130]), this kind of electromagnetic gripper does not require significant mechanical amplification to achieve large displacements. While the gripper demonstrated is quite rudimentary, the building block approach potentially enables more extensible designs that incorporate additional degrees of freedom. Custom end-effector interfaces, made using separate processes like 3D printing, could also be incorporated to aid in the manipulation of certain kinds of parts.

**Manipulation Tools** Based on these positioning and manipulation primitives it should be possible to assemble more advanced manipulation tools that can be used to precisely arrange small parts. Prior work has shown the dexterity that is possible with flexure based linkages for translating, rotating, and assembling small parts [131] [132]. A concept using walking motors to provide large displacement actuation is developed in Figure 8-3. This concept uses a parallel kinematic chain to position and place parts on a lattice. The important functional parts of the machine are predominately assembled using the building block parts developed in this thesis. Concepts for assembling assembly tools are developed further in Chapter 10.4.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

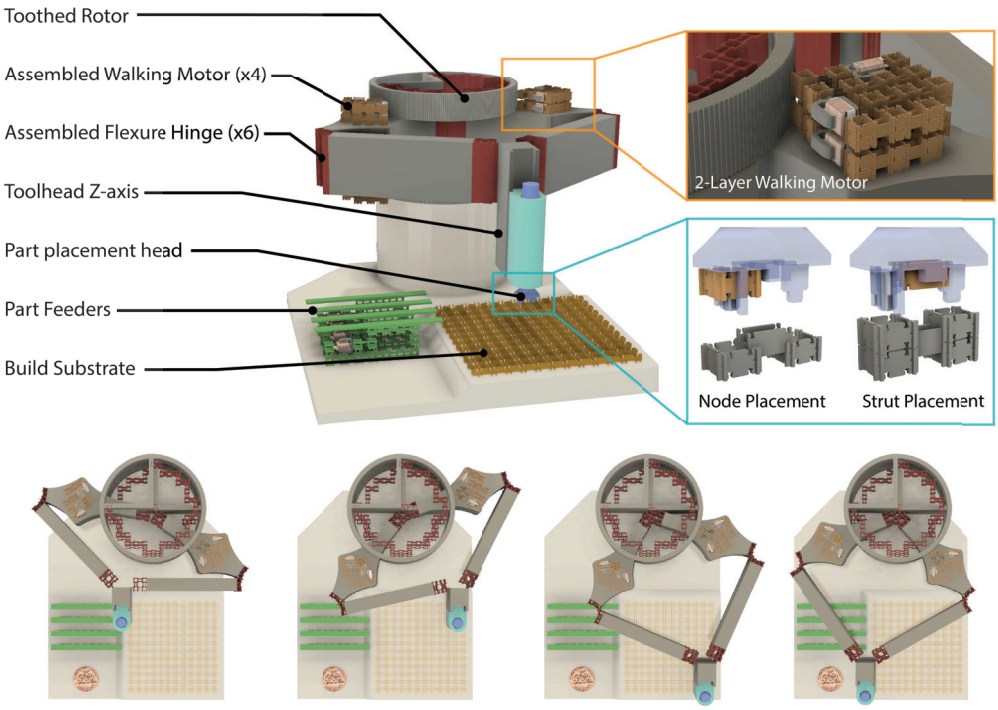



Figure 8-3: An assembler concept that uses walking motors to produce large displacements of a parallel kinematic chain to position and place parts in a lattice.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 9

# Circuitry

Up until now all of these machines and systems developed in this thesis have been powered and controlled by off-board electronics. However, because this assembly framework is agnostic to the part production methods used to make parts, it is possible to incorporate parts made using printed circuit board methods to incorporate circuitry into these structures. In this chapter I'll describe how routing, logic, and control systems can be assembled from a small set of parts to enable the integration of electronic functionality within these assembled machines.

## 9.1 Background

Research as far back as that 1950's has looked at ways to modularize and automate the assembly of electronics. Before printed circuit boards and integrated circuits had established themselves as the standard in the production of electronic circuitry, the US National Bureau of Standards had developed Project Tinkertoy, which sought to mechanize the assembly of electronic components [133] [134]. Project Tinkertoy was based around a set of modular ceramic wafers, each containing a single electronic functionality (such as resistance, capacitance, or a connector for a vacuum tube). These ceramic wafers could be produced in large quantities and patterned with these functionalities relatively efficiently. Then, a machine called the module assembler was able to pick from a feedstock of tiles, stack them, and then solder a set of wires on each side of the module to connect the wafers to one another and give the module its structure. Each of these modules housed the same functionality as a handful of discrete parts and simplified the assembly electronic assemblies by providing this common interface.

More recently, research efforts in digital fabrication have looked at integrating electronic functionality into additive manufacturing processes. Researchers have recently developed conductive ink formulations that enable the controlled deposition of highly-conductive traces [135] [136]. Still, these formulations are one to two orders of magnitude less conductive than bulk metal and often require a post-bake processing step to evaporate the solvent, which limits the substrate material choice. With these conductive inks, researchers have started to show that it is possible to print functional electronics. Using specially formulated anode and cathode inks, researchers are able to print lithium ion [137] and zinc-air [138] batteries. In another study, the conductive inks were conformally printed on 3D substrates to fabricate efficient antennas [139]. Ink-jet printing has also been used in similar


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

ways to deposit highly conductive silver traces to create electromechanical functionalities like an electrostatic motor [140]. While these printing techniques have shown promise, their integration into useful devices, which combine both electronic and mechanical functionality, have been lagging the research.

An alternative method to automate the assembly of electronics has looked at building block based approaches, in which electronic functionality is abstracted into individual building blocks and then assembled. These building blocks can range in complexity from high-level modules for modular cubesats [141] to simple raw-material parts [85] [61]. Between these extremes are architectures that employ parts that contain limited functionality such as resistance, capacitance, diodes, connectivity, or insulation [23]. In this work, MacCurdy et al., assembled functional electronic circuits from individual component building blocks. The building blocks themselves were made from printed circuit boards and commercially available press-fit connectors were used to join the blocks to each other. A 3D printer was modified with a part placement head and automated assembly was demonstrated through the assembly of an infrared remote control composed of 17 blocks. This building block based approach has proven the feasibility of assembling advanced electronic functionality from a limited part set and is the approach pursued and developed upon in this work.

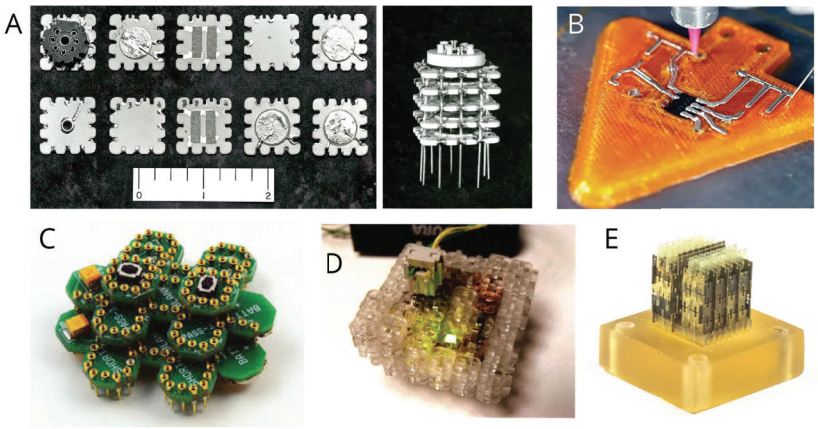



Figure 9-1: Prior work in automating the assembly and integration of electronics. In the 1950's Project Tinkertoy developed a standardized assembly framework for electronic building blocks (A) [133]. Recent work in digital fabrication and additive manufacturing have looked to integrate the printing of conductive traces (B) [142]. Building block approaches have shown the possibility of building advanced electronic functionalities from a small set of parts (C,D,E) [23] [85] [61].

## 9.2 Routing

The first capability that's necessary to integrate electronic functionality is the ability to route electrical traces. One way to do this within the assembly framework is with the introduction of seven part types to enable routing within the structural lattice. The seven part types include three kinds of node parts (conductive, insulating, and split-conductive) and four kinds of strut parts (conductive, insulating, split-conductive, and crossover) (Figure 9-2). These parts are likely not the minimal set of parts required to route, but have been found to enable the routing of most schematics in a


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

relatively space-efficient way.

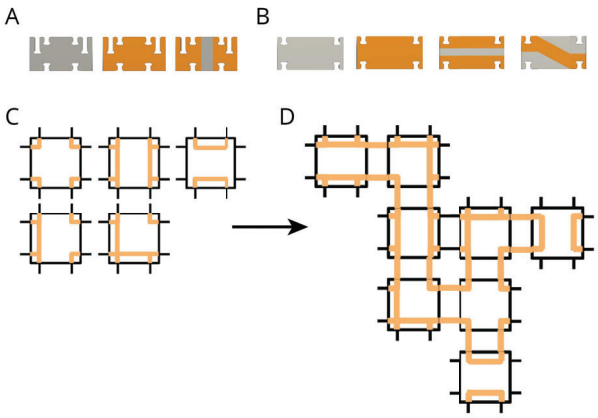



Figure 9-2: Circuit routing is accomplished with seven part types. Three node part-types (A) are assembled to form a variety of routing cells (C). These cells are then linked by variously conductive struts (B) to route traces within an assembled layer (D). The fourth strut type connects traces across layers of the assembly.

The three node parts enable the assembly of a variety of different "routing cells". These routing cells can then be oriented, arranged, and connected by various kinds of struts to enable the routing of traces within an assembled layer. This routing scheme allows each layer of the assembled structure to be treated like a two-layer circuit board. The fourth strut part-type, a crossover strut, acts as a "via" in this analogy, transferring signals between the top and bottom layers.

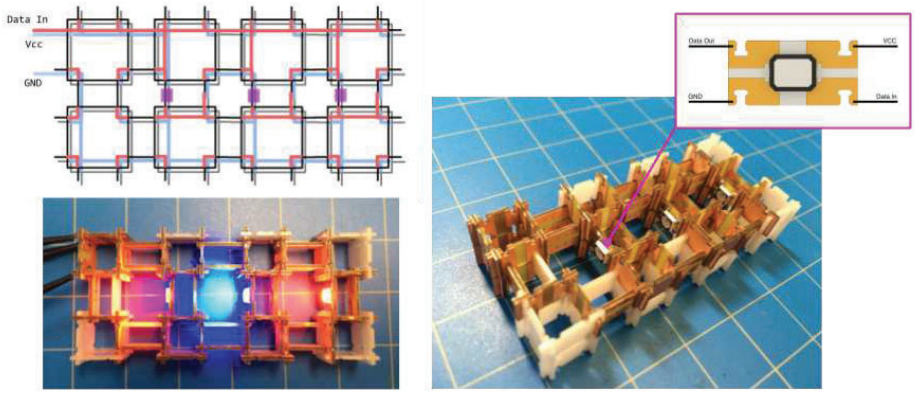



Figure 9-3: A series of serial-controlled RGB LED's (Neopixel) are routed and controlled within an assembled lattice. A serial data line snakes through the middle of the structure while a power and ground rail wraps around the outside.

Based on this routing methodology its possible to route various kinds of circuits within this assembly framework. As an example, I assembled a structure containing three Neopixel RGB LED's (Figure 9-3). These LED's take serial data input to control the output of the red, green, blue LED's. To control multiple LED's, their serial data inputs can be daisy-chained such that the serial data


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

output of one is connected to the input of the other. In the assembled structure, a serial data line is routed in a serpentine way through the assembly to connect all of the Neopixel LED's. The LED's also need power and ground connections and these are routed in a loop around the edge of the structure.

## 9.3 Logic

More complicated routing examples involve the creation of digital logic devices from logic gate building blocks.

Using various production methods, it's possible to embed an individual logic gate on an individual building block. Pictured in Figure 9-4, is a design of a logic block that embeds the smallest commercially packaged logic IC's (SOT1226) onto a strut, which can be assembled in the lattice architecture. Logic gates require a maximum of five input/output (I/O) pins: Vcc, GND, A, B, X (where A and B are two inputs and X is the logical output). In routing these building blocks, it's convenient if the Vcc and GND pins are symmetric and appear at both ends of the part. This number of input outputs (7) is compatible within the eight electrical connections available within a single strut-pair.

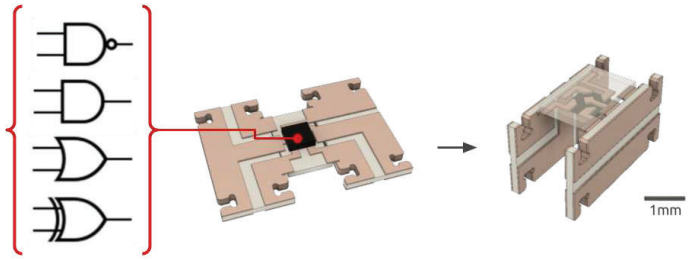



Figure 9-4: Building blocks can be made to incorporate the smallest commercially packaged logic gates.

With two NAND gate building blocks along with a few additional parts to route the required traces, it's possible to assemble an set-reset (SR) flip flop. A D-type flip flop augments an SR flip flop with two additional NAND gates such that the state can be latched and unlatched with a clock pulse. This is assembled with three additional layers to house the two extra NAND gates and an inverter. Furthermore, adding routing to connect Q-NOT to the D-input configures this flip flop to serve as a frequency divider with feedback.

The routing diagram to reproduce the desired circuit within the assembly framework is pictured in Figure 9-5. Power and ground are routed on inner loops of the structure such that they are accessible to all components in the system and so that the signal wires are accessible from the perimeter. Most of the routing is accomplished in a 2x2x3 grid of our lattice with the exception of the feedback loop connecting Q-NOT to D (required for serving as a frequency divider). With the D-type flip flop structures as building blocks, it's then possible to assemble higher-level assemblies like a ripple counter or frequency divider. This structure takes a clock pulse in and counts in binary on the Q outputs and provides an output square wave at one-sixteenth the frequency (Figure 9-5 D).


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

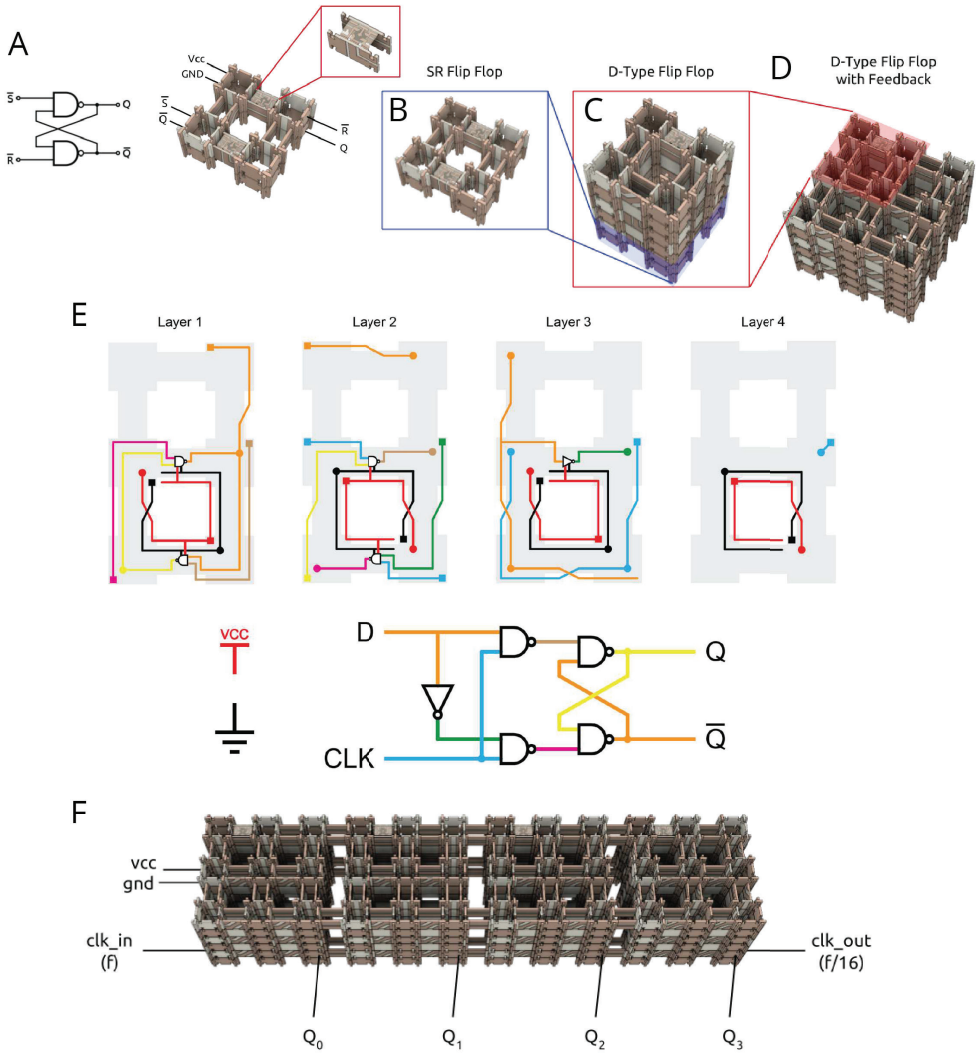



Figure 9-5: With logic gate building blocks a series of digital circuits can be assembled. An SR flip flop is assembled from two NAND gates (A,B) and then is composed to form a D-type flip flop (C). Feedback is added (D) and then used to construct a counter and clock divider (F). The routing for the D-type flip flop is shown (E).


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 9.4 Control

Beyond integrating logic gates on an individual building block, it's possible to embed even more functionality by taking advantage of the miniaturization of semiconductor products and incorporating whole processors and other devices (H-bridges, Op-Amps, FPGA's) directly on a building block part.

At the scale of the mechanical degrees of freedom in the assembly architecture it is possible to integrate more complex electrical behaviors. In particular, the smallest commercially available microcontrollers are able to fit within a single building block part within the same lattice. This both enables complex behaviors and control systems to be programmed into the parts as well as enables a high degree of flexibility, enabling serial communication, pulse width modulation (PWM), and analog to digital conversion, among other possibilities with a limited part set.

In adding these higher level parts to the assembly framework, higher density routing is sometimes necessary. The development of these higher level electronic part types went through a number of design revisions. I think it's instructive to enumerate them and provide explanations about what makes each attractive and why I ended up abandoning some directions in favor of others.

### 9.4.1 Control in struts

Embedding processors and amplifiers into individual struts enables the assembly of simple control systems for actuated degrees of freedom.

The first prototype of an assembled control system used an ATtiny10 microcontroller. This microcontroller has just four I/O pins, which is enough to demonstrate the feasibility of integrated a microcontroller into one of the parts of the assembly ecosystem. The details of how the part is fabricated is discussed in section 9.5.

The microcontroller is programmed to generate a triangle wave actuation signal. The building block is arranged in the structure such that the output of the microcontroller feeds into the input of an H-bridge building block. This component amplifies the actuation signals and passes them onto the actuator itself, producing motion. This enables the assembly of actuated degrees of freedom with only two external connections for power.

Beyond controlling a single degree of freedom with a single program, the I/O capabilities of this microcontroller and its integration in the lattice were ultimately found to be limiting. In particular, the bus-width in the lattice needed to be increased. The strut-pair model that works so well for mechanical degrees of freedom, is limiting when trying to assemble control systems that integrate and network multiple microcontroller and h-bridge building blocks.

### 9.4.2 Control between struts

In order to integrate networks of processors and components, I developed a design that enables higher density routing (Figure 9-7) while still leveraging the in-lattice electrical routing capabilities explored in Section 9.2. This meant positioning devices requiring more than eight I/O in locations spanning multiple lattice positions.

This next series of prototypes used ATtiny20 microcontrollers. This microcontroller is available in a wafer level chip scale package (WLCSP) that is approximately 1.5 mm x 1.5 mm and is the smallest off the shelf microcontroller at the time of this writing. It features 14 pins, which allows




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

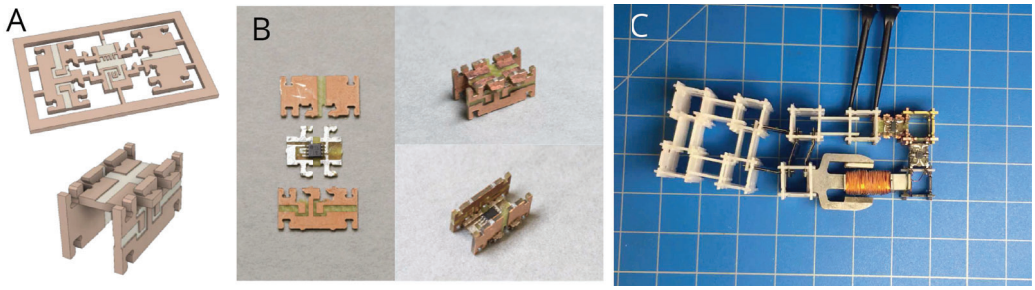



Figure 9-6: Embedding integrated circuit building blocks in struts enables the straightforward assembly of integrated controllers. A microcontroller within a building block (A,B) generates actuation signals that are amplified by an H-bridge building block to power an actuated degree of freedom (C).

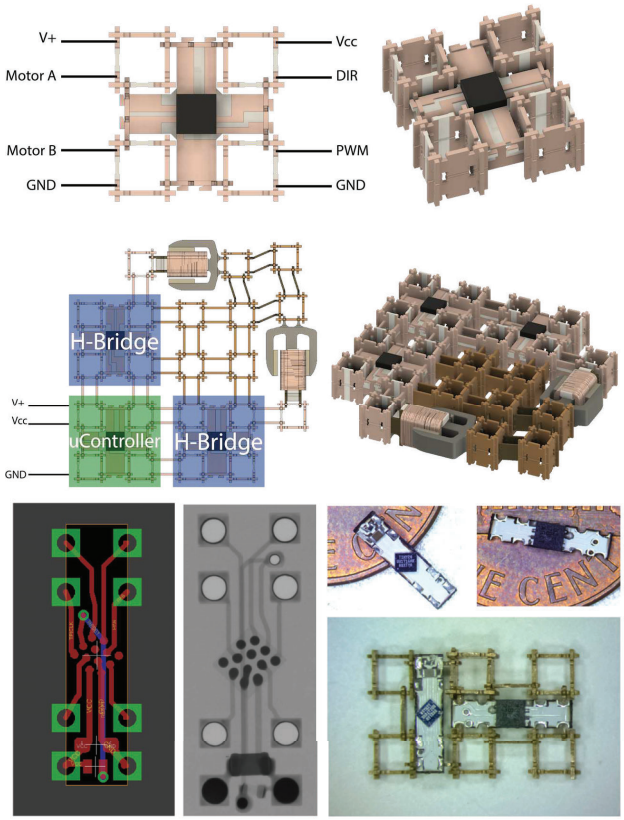



Figure 9-7: More advanced capabilities can be integrated by spanning multiple lattice sites. Wafer level chip scale package (WLCSP) microcontrollers and H-bridges are embedded on parts that interlock with the lattice.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

for more connectivity between adjacent parts. However, given the pitch of the ball grid array, it is not possible to route more than eight I/O pins from the board without resorting to more advanced PCB construction methods (e.g. plugged vias), which I was looking to avoid. That being said, only six I/O pins are necessary to enable the control of two degrees of freedom based on some input.

In this design the integrated circuit building blocks fill the negative space of the lattice. This is beneficial in term of spatial density. This design can be implemented in a variety of ways. For example, Figure 9-7 depicts a design where each microcontroller and H-bridge span four lattice locations while other variants are stick-shaped and span just two cells. These stick-shaped parts are still able to incorporate and the same number of I/O as the cross-shaped parts because ultimately the bottleneck in this configuration is the escape routing through the nodes of the lattice. While there are ideally 16 output connections coming from the four nodes, only eight of those are uniquely addressable. For this reason, the stick-shaped parts, which span two cells and still have eight possible I/O, are preferred.

### 9.4.3 Redefining struts for greater routing density

Having developed this first iteration of computation parts, I looked to enable even greater routing density through the modification of the part geometry. To overcome the routing constraints of the nodes and struts, I developed a laminated strut design that allows up to eight signals to be passed within each strut. The design also instantiates a new "strut-node" part type that merges two nodes and a strut. This merging greatly simplifies the routing of the electronic connections out of a high-density component like a microcontroller and reduces the overhead of the electrical connectors in the assembly.

The routing of signals in this framework is done with the use of variously patterned node parts. These instantiate the basic "routing cell" motifs developed in section 9.2 with capabilities such as 2-way pass-through, 90-degree turn, 3-way connection, 4-way connection, and no connection.

This design incorporates routing at a higher density than the other designs but retains a sparse lattice. The sparsity of the lattice is a free design parameter.

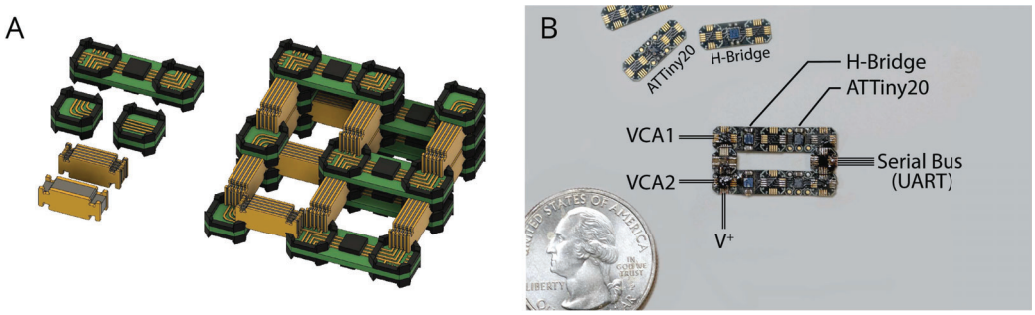



Figure 9-8: By adding higher density electrical connections between parts, more advanced functionality can be implemented (A). A two degree-of-freedom motion controller is assembled from individual microcontroller and H-bridge building blocks (B).


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### 9.4.4 Tiles – Decoupling mechanical and electrical routing

A four-connected lattice is much better suited for routing electrical degrees of freedom. This meant moving the electronic functionality into the nodes rather than the struts. This design breaks from prior designs in decoupling the electrical routing requirements from the mechanical construction of the lattice.

The tile parts feature an upgraded microcontroller processor, an ATSAMD10D14. This 32-bit ARM processor runs up to 48 MHz and has UART and PWM peripherals that can be mapped to pins in firmware. Four ports are defined, one for each side of the tile, each having two pins. The ports are configured either to function as a bidirectional UART or to output PWM on the two channels. As such, each tile has four electrical connections on each edge: Vcc, signal A, signal B, GND.

Using this assembly framework I assembled a two degree-of-freedom motion controller. A USB serial connection from a computer serves as the input. Commands are sent from a command-line terminal. The UART peripheral on the microcontroller interprets these commands and generates the specified actuation signals, which are sent to the H-bridge to be amplified and sent to the actuator.

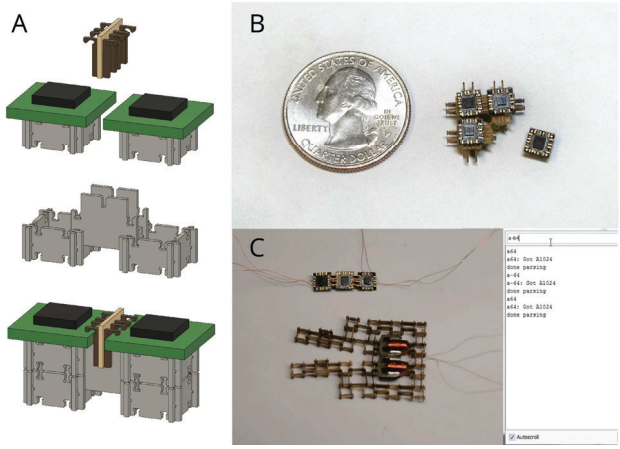



Figure 9-9: By decoupling the electrical connectivity from the mechanical lattice, higher density integration is possible with "tile" parts. These tiles are configured to accept commands from a computer terminal and control the actuated gripping mechanism (Bottom).

Routing of the electronic functionality with these tiles has different requirements than routing signals through the mechanical lattice. Instead, I find that two part-types enable planar routing with sufficient density. The two part types are:

- **Translator:** A translator connects the four edges of the tile in such a way that the polarity of the connections across the tile is not changed.
- **Rotator:** A rotator, in contrast, connects the four edges of the tile such that the polarity of the connections across the tile is reversed.

These two part-types are further configured by the presence or absence of a edge connector. For example, a rotator tile with edge connectors only on the North and East edge, acts as a 90-degree


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

connector, connecting the South neighbor to the West neighbor. The routing of a six degree of freedom controller is presented in Figure 9-10.

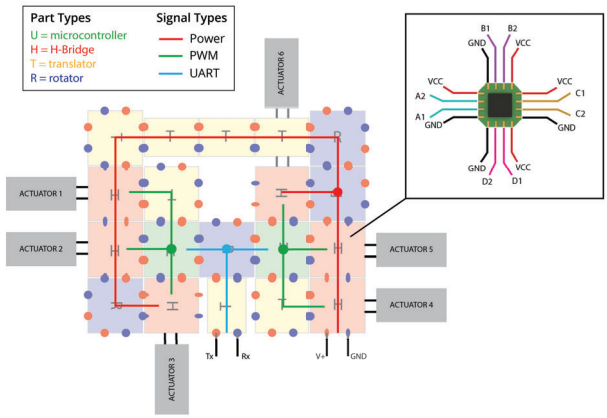



Figure 9-10: A six degree of freedom controller is routed using the tile parts. Red and blue circles on the edges represent the polarity of the signals.

## 9.5 Production Methods

The parts required to route circuitry and integrate computation in this assembly method can be produced using a number of different methods but printed circuit board processes tend to be well suited for most of these parts.

**Routing Parts** The seven part types required to route traces within the structure are just at the edge of commercially available PCB board-house capabilities (Figure 9-11 A). The parts are arrayed in a large panel. Fiducial markers and three holes are added to the frame for alignment of subsequent post-processing operations. The size and tolerance of the plated slots in the parts fall below what most manufacturers are capable of. However, a few manufacturers were able to meet the specifications with the use of laser-routing (Gold Phoenix and PCBUniverse). In addition, the printed circuit board manufacturers are unable to control the thickness of the boards within an acceptable tolerance for a press-fit connection. They can only hold $\pm 0.06$ mm on the overall thickness. Clearly a 0.31 mm part is not going to fit in a 0.25 mm slot and a 0.19 mm part is going to fall right out of the same slot. The actual specification varies from panel to panel and from production run to production run and so for small prototyping runs it can be a reasonable strategy to just pick panels that fall within the desired tolerances from a much larger batch. In one particular representative case, I measured the thickness of the parts to be an average of 0.288 mm with a high of 0.3 mm and a low of 0.280 mm. Because these dimensions are all greater than the desired thickness, I can polish the parts down to size. In order to bring the thickness down, I polished with combination of 2500 grit sandpaper and 1200 grit polishing compound between flat acrylic sheets to bring the thickness between 0.251 mm and 0.271 mm. These parts are then able to interlock with parts made using the wire-EDM without deforming them.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

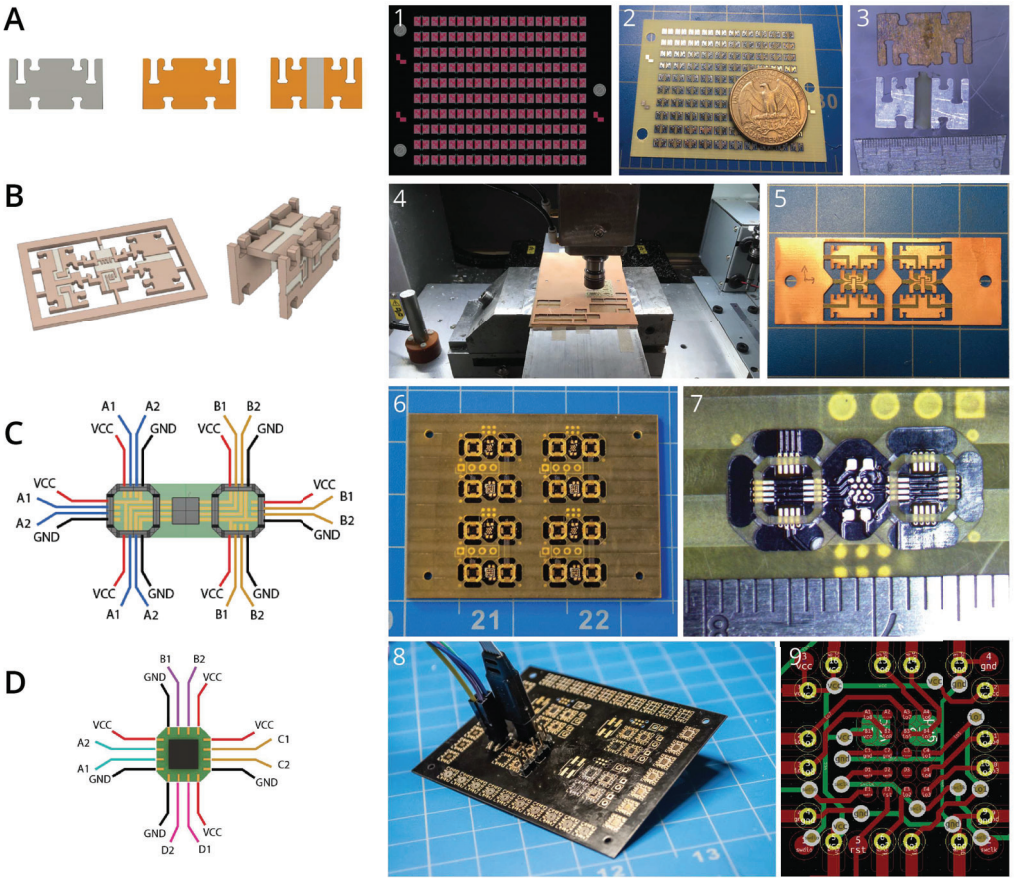



Figure 9-11: Circuit part production methods. (A) Production of routing parts is done using laser routed PCB processes. (B) Isolation routing of FR1 copper clad substrates enables the fabrication of three-dimensional struts with integrated circuits. (C) Strutnode parts are produced as a PCB and then laminated or overmolded with mechanical interlocking features. (D) Tile parts are also produced as a PCB and then affixed on top of a node part-type with epoxy.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Integrated Circuit Struts** For the larger-scale parts, in-house PCB fabrication methods can be used (Figure 9-11 B). In particular, isolation routing (milling) of copper clad FR1 can be used to rapidly fabricate building block parts with integrated circuitry to incorporate microprocessors and other integrated circuits. Three-dimensional struts are produced through the assembly of two dimensional parts. The parts are milled together as a panel. The panel features traces on both sides so alignment features (dowel pin holes) are used to align the panel for double-sided machining. Once the panel is machined on both sides, the tabs can be cut with a sharp knife to release the individual parts. At this point, the microcontroller or H-bridge is soldered onto the part. The microcontroller is programmed by attaching clips to the required pins, which have been broken out. Finally, the three-dimensional strut is assembled manually with pliers.

**Strutnode Parts** For parts that integrate off-the-shelf integrated circuits, printed circuit boards processes can be used in a way that is a little more natural but then augmented with lamination or molded features (Figure 9-11 C). The microcontrollers selected for integration into the strutnode building blocks use a wafer level chip scale packages (WLCSP) with micro-ball grid arrays ($\mu$BGA), which have a pitch between solder balls of 0.4 mm. It is possible to adequately route these parts on a two layer PCB using a 4-mil/4-mil trace/space routing design rule. Furthermore, via-in-pad (masked-plugged-vias) construction methods are not necessary in order to escape route the necessary functionality for these building block parts.

The boards are sourced from a commercial PCB fabricator (PCBWay). The thickness of the board is specified to be 0.6 mm. Gold plating is used on the contacts for improved definition, ease of soldering, and to enable wire-bonding, if required.

Once the boards have been fabricated, the components need to be soldered. A stencil makes the application of solder-paste relatively straightforward. Incorporating alignment features on both the circuit board and stencil can be helpful to establish the initial alignment. Once the paste is applied, the components are arranged on the board and the whole board is reflowed following the time and temperature protocol specified by the solder-paste. At this point, the functionality of the parts can be tested using the connectors neighboring the parts on the PCB panel.

At this point the boards are functional but have no way of mechanically interlocking with one another. To add the interlocking features I pursued two different strategies: lamination and molding.

The lamination process begins with the machining of top and bottom laminates that are to be bonded to each side of the PCB. Ultimately, I found that the lamination strategy was incompatible with the materials and geometries that were required given the required processing temperatures. Either the materials are too brittle to robustly support the feature sizes I require or their coefficient of thermal expansion ensures poor adhesion to the PCB substrate.

To mold the features on the PCB, I developed a vacuum-assisted molding workflow. The vacuum-assist is beneficial in ensuring that whole surface of the mold remains in tight contact with the PCB so that flashing is minimized and the PCB surfaces remain clean. A soft-tool (Oomoo 25) was cast from a machined laminate layer. Ports were added at both ends of the tool and this was then aligned to the PCB panel. A hard polymer resin was mixed and then pulled through a tube into the mold by a vacuum tube attached to the other port of the tool mold. This strategy was able to reliably produce the geometric features on top of the PCB surface. However, the soft tool deformed from the vacuum pressure and the resulting geometry differed significantly from the design intention.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Finally the parts are singulated from the panel by machining with a 1/32" endmill.

**Tile Parts** The tile parts, like the strutnode parts, are based around PCB fabrication processes. Fabrication of the tile parts proceeds in the same way as the strutnodes described above. However, instead of laminating or molding features on to the PCB surface, the tile parts are affixed on top of node parts. This is done as a final step once the tiles have been soldered and singulated from the panel. Individual tiles are arranged in an array of cutouts in a fixture. The fixture, with tiles firmly in place, is turned over, a drop of epoxy is dispensed on the bottom surface of each tile, and a node part is inserted into the alignment hole in the fixture to be attached to the tile.

Connectors for the tile parts are made from wire-cut bronze parts and laser micromachined fiberglass parts. These parts feature interlocking slots that are used to join four bronze connectors to a single fiberglass stabilizer. Press-fit forces are enough to rigidly affix this assembly without the need for any adhesive.

## 9.6 Discussion

Beyond the part types developed here, a number of other parts may be beneficial for the assembly of robotic systems and assemblers in particular. For one, power storage is an important capability that is still undeveloped. This could likely take the form of a battery or super capacitor part type. Zinc-air batteries have among the highest energy densities of battery technologies. These batteries are commonly used in hearing aids and are designed to provide energy for a low draw (a few milliamps) for a week or more. While they have great energy densities, zinc-air batteries are not rechargeable and expire with or without use in a few weeks time (after the tab is removed and air is let in). A size 10 battery (the smallest commonly available hearing aid battery) is 5.8 mm x 3.6 mm. Two of these could be stacked to fit on a single actuator sized building block that spans two cells of the lattice (taking up a volume roughly 5 mm x 9 mm x 4 mm). This would provide approximately 200 mAh of capacity, enough to power a single actuator, which draws 100 mA (the 320-turn coil), for two hours. The four actuators of the stacked walking motor would conservatively last for at least 30 minutes with a single battery part.

Other part types may be useful such as a wireless communication part. This could simply act as a relay to connect the UART input of an integrated processor with an external controller attached to a computer. This part could be useful for the coordination and control of multiple assemblers acting simultaneously. An ATBTLC1000 chip, for example, is a bluetooth system-on-chip (SoC) that is available in a 2 mm x 2 mm WLCS package [143].

A pain point in the development of these circuitry capabilities was in achieving higher density routing connections between lattice sites. The in-lattice routing is remarkably robust for basic connectivity between electronic functions but extending beyond the in-lattice routing has been error prone and delicate. It may make sense to explore alternatives to individual connectors for each signal. For example, anisotropic conductive adhesives ("Z-tape") that are only conductive in the Z-axis but not the XY axes could be used to very simply attach planar parts through a patterned substrate [144]. There are likely also ways to create similar functionality through microstructure [63].

Given the overhead and complication of physical connections it may be possible to replace some


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

portion of these physical connections altogether with non-contact connections. In [145] inductive communication is used between modules by taking advantage of the windings of the electropermanent latches used to connect between modules. While well suited for this specific case, this is unlikely to be a space efficient solution for other designs as inductors are difficult to produce at small scales. Another alternative may be to use optical connections between parts. For example, an infrared emitter and detector pair may be used to send and receive signals, respectively. This technique has the advantage of the possibility for communication over distances greater than one lattice pitch. This could potentially be an efficient means of connecting parts that are within line of sight but relatively far geographically. On-chip optical communication, for example, is being pursued as a means to send signals more efficiently over longer distances on integrated circuit boards [146].


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 10

# Automation

While assembling assemblers is the ultimate aim of this work, intermediate off-ramps of automated assembly using table-top platforms could enable the translation of digital designs into physical machines without manual intervention.

The standardization of the interface between parts simplifies assembly automation. Rather than requiring a several degree-of-freedom arm in order to position parts in arbitrary orientations, the parts are designed to be assembled vertically. This makes the assembly process much more similar to the pick and place assembly used to manufacture printed circuit boards than the complex arms used in automotive assembly.

Here, I describe the implementation of two automated assembly systems that leverage this simplified assembly framework for two different digital material geometries. The first assembles electronic digital materials detailed in prior work and the second expands upon this prior work to assemble systems that incorporate structure, mechanism, and actuation.

## 10.1 Electronic Digital Material Stapler Assembler

## 10.2 Design Overview

I first automated the assembly of the electronic digital materials developed in previous work [61]. The design of the assembler largely resembles that of conventional fabrication machines. However, the assembler differs from other fabrication tools in that it takes advantage of the inherent qualities of digital material structures. Unlike conventional machine tools, the assembler can correct positioning errors within a tolerance by registering with the digital material lattice, enabling the assembly of structures more accurate than the assembler itself. To enable the toolhead of the machine to self-align with the digital material structure, intentional compliance between build platform and the toolhead was added through the use of an XY-flexure mechanism located on the Y-axis. This flexure allows the alignment fingers of the stapler toolheads to register to the negative space in the lattice, correcting for positioning errors up to 0.5 mm in X- and Y-axes.

Each toolhead features a part placement mechanism and a magazine of parts. The magazine, which contains up to 100 parts, is made to be easily removable from the rest of the toolhead. This is achieved through the use of a repeatable coupling mechanism. The tapered magazine is forced up


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

against two dowel pins by a conical point setscrew, which interfaces with a conical hole on the front of the stapler magazine. This constrains all 6-degrees of freedom of the stapler magazine in such a way that it can be removed and reinserted midway through a build without any loss of precision.

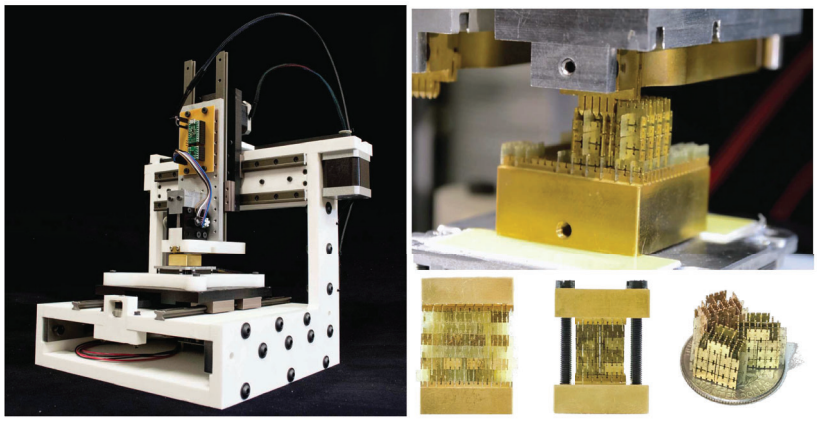


Figure 10-1: The stapler assembler builds electronic components from conductive and insulating parts.

### 10.2.1 Detailed Design

A 4-axis motion gantry is used to position two toolheads spatially with respect to the structure being built. The motion gantry is constructed from a steel reinforced HDPE frame. The frame is CNC milled to precisely position all holes and pockets. Precision ground steel is then bolted to the frame pieces to add rigidity and mass. The axes use polymer linear guideways for low friction motion in each axis. The axes are powered by high-torque NEMA–17 stepper motors and driven with GT2 timing belts. The motor and power transmission elements were specified to nominally provide a 1/16th microstep resolution of 5 $\mu$m and a maximum operating speed of 50 mm/s with a potential linear force delivery of 70 N in each axis. Backlash in the timing belts was measured to be between 125 $\mu$m and 250 $\mu$m in the X- and Y-axes but can be compensated for by approaching part placement locations from the same direction. Doing so, enables a positioning repeatability better than 25 $\mu$m in X- and Y-axes.

The part placement mechanism is driven by a 100:1 DC gear-motor through a worm gear. This mechanism produces roughly 32 N of downward force for part-insertion, which is almost three times the expected 12 N of force needed to insert a part. Using this mechanism, the maximum part deposition rate is 0.72 Hz.

### 10.2.2 Workflow and Characterization

The stapler assembler platform is able to assemble electronic components such as capacitors and inductors from conductive and insulating parts. The part placement speed for the assembler is approximately 0.2 Hz (or 1 part every 5 seconds), enabling a volumetric build rate of 80 $mm^3$/min. This is on par with the build speed of commercial 3D printers, which take a few hours to build a one


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

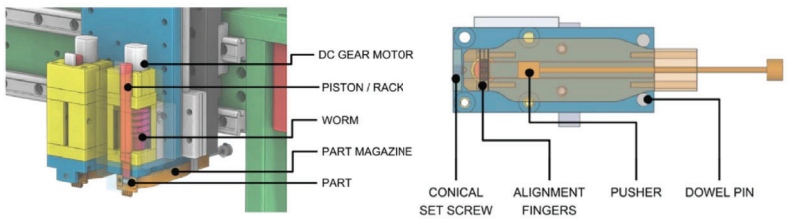



Figure 10-2: The stapler assembler features two toolheads (one for each part-type). The parts are deposited by a worm-rack mechanism powered by a DC gear motor. The part magazines are removable and interchangeable with quasi-kinematic coupling.

cubic inch structure. In the case of electronic functionality, this speed allows for the construction of two 5 pF capacitors in 20 minutes (including manual magazine refilling).

In order to generate designs and toolpaths for the assembler, a custom CAD/CAM workflow was developed and has been document in prior work [147]. The design tool enables the design, simulation, verification, toolpathing, and machine control within a single environment (Figure 10-3). This single environment greatly reduces the friction that is typically associated with converting a digital design into a physical object. Furthermore, design workflows like this lend support to the idea that assemblers can help lower barriers to the fabrication of integrated devices by empowering a single user to design, verify, and fabricate a device and go from building block parts to a functional device in an afternoon and on a desktop.

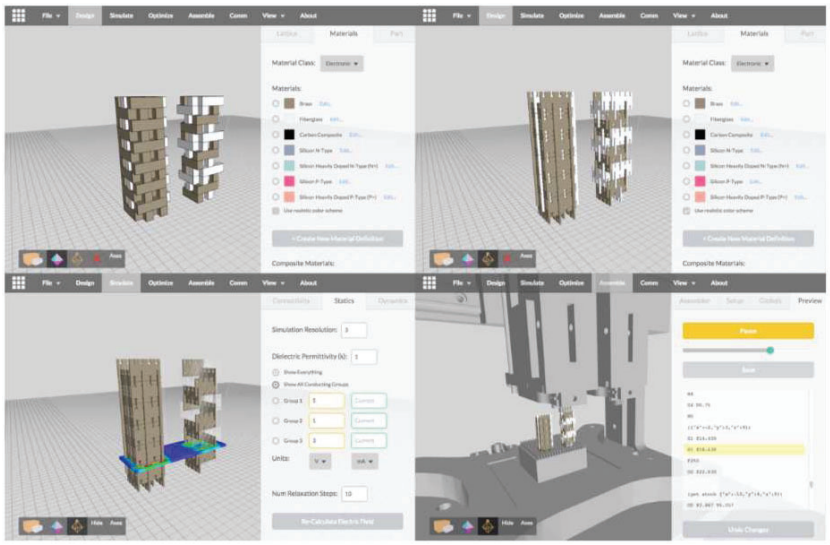


Figure 10-3: Reproduced from [147]. To control the assembler Ghassaei developed an integrated CAD/CAM workflow that enables design, simulation, toolpathing, and machine control within a single environment.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 10.3 FABBR – Functional Assembler of Building Block Robots

### 10.3.1 Design Overview

Rather than use a stapler-type toolhead, I redesigned the placement heads for pick and place type operation. While this necessitates an extra traversal for each part placement, it significantly reduces the complexity of the toolhead and allows for much easier accommodation of larger or unusual parts like actuators or electropermanent feet.

Each toolhead actively grips the part. While requiring an extra degree of freedom and a more complex design, this eliminates the careful tuning of the various combinations of attachment forces that need to be considered between the part, toolhead, substrate, and assembly. This active gripper design lets the assembly machine vary the grip strength from approximately nothing when un-gripped to an order of magnitude stronger than the passive gripping elements of the substrate.

Three toolheads are required to support the assembly of machines integrating structure, mechanism, and actuation (Figure 10-5). A node toolhead places node parts. A strut toolhead places a pair of struts which can be rigid or variously flexible. An actuator toolhead places a coil and magnet part simultaneously.

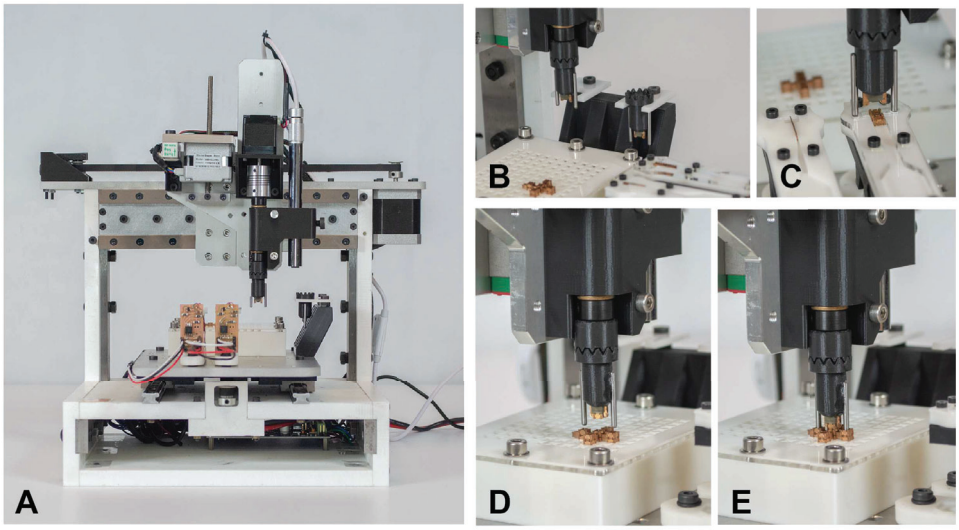



Figure 10-4: The upgraded assembler features an automatic tool-changer, a tool-mounted rotary axis, an on-tool microscope and a large 12x12 unit build substrate.

In developing the assembler I first attempted to automate the placement of individual (non-paired) struts. This proved challenging to do in a pick and place type operation since the size and aspect ratio of the struts necessitates precise and finely featured alignment jigs. The manipulation of a single strut is also challenging as they tend to act like playing cards; once stacked they are hard to unstack without active manipulation (like a feeder). I later decided to treat struts in pairs such that they are volumetric build elements. This allows the strut gripper design to much more closely resemble that of the node gripper and allows much larger features (and negative space) to be used for attachment and placement. In a future revision of the assembler it would certainly be


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

desirable to treat the struts as individual parts and could be accomplished by designing a an actuated stapler-type magazine that automatically separates and presents the next part to be placed.

The struts are pre-assembled using orthogonal press-fit plates that set their spacing. This required a slight redesign of the strut parts to introduce a slot for this extra connection. For the flexible struts it also required shifting the hinge degree-of-freedom towards the center of the part. In the case of a pair of two degree-of-freedom struts this reduces the spacing between hinges and results in less range of motion and more parasitic error when used as a linear motion constraint. These tradeoffs were seen as necessary in order to develop the first version of the assembler but shouldn't be seen as limiting for future revisions.

### 10.3.2 Detailed Design

I updated the design of the stapler assembler for use with the mechanical part types (Figure 10-4). This involved a number of important upgrades. First, the HDPE panels supporting the X-axis rails were replaced with aluminum. The extra stiffness provided by the aluminum plate allows for much better tensioning of the timing belt, resulting in significantly reduced backlash. The Z-axis timing belt drive was replaced with a leadscrew transmission to enable much higher placement forces to be exerted. The rotary degree of freedom was moved to the toolhead rather than the build platform, significantly reducing the required precision of the rotary axis since the toolhead is much smaller than the build platform (and so the Abbe error is significantly reduced). The upgraded assembler features three main enhancements: the active gripper toolheads, the build substrate, and the control system.

**Gripper Toolhead** The gripper is actuated by a linear servo motor (Actuonix) that is housed on the Z-axis. The linear servo motor is composed of a small rotary DC motor and a small screw transmission that provides a 100:1 gear reduction. This enables the actuator to produce forces up to 35 N over a 20 mm stroke. The linear motor actuates the gripping mechanism through a thrust bearing that is magnetically coupled to the sliding part of the gripper toolhead. The thrust bearing is necessary to reduce the friction at the interface and allows the gripper to be rotated while actively gripping the part.

The sliding element of the gripper is used to actuate four gripping arms. The gripping arms are made of polyetherimide (Ultem). This material was chosen after a number of iterations because it machines very well, is relatively hard, but is not excessively stiff. This compliance allows a parallelogram flexural linkage to constrain the arms such that they can squeeze together to pick up a part while remaining stiff about the z-axis. A taper is used to couple the linear motion of the sliding element to the gripping action of the arms and provides mechanical advantage to amplify the gripping force.

The toolheads are able to be changed automatically during a build based on which part needs to be placed. The tool changing mechanism uses a magnet at the base of a tightly tolerance bore to attach and constrain the toolhead to the rotational axis. The toolheads are constrained and aligned about the rotational axis with a tab that interfaces with slot on the rotary axis. Picking up a tool simply involves moving to the appropriate X,Y location, driving the Z-axis down, and then moving sideways to pull the toolhead out of the flexural arms of the toolholder. Removing the tool is the same in reverse.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

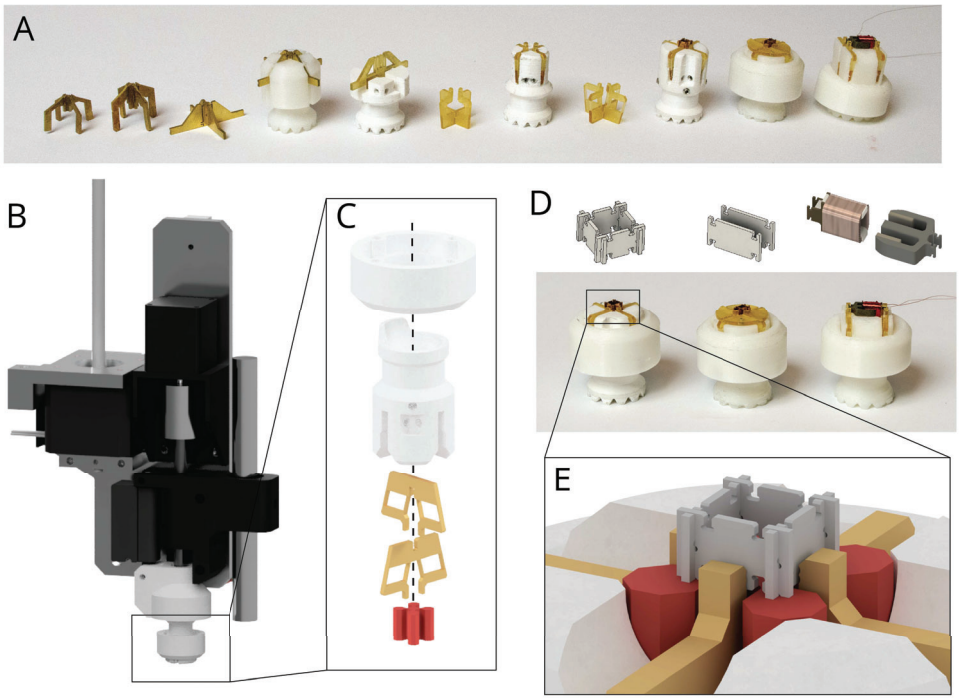



Figure 10-5: Development of the gripping toolhead required a number of iterations (A). The toolhead (B) integrates a linear servo to actuate the gripper (C) as well as a rotational degree of freedom. Three gripper toolheads are used and interchanged during assembly (D). The grippers feature machined Ultem flexures that grip the extents of the parts (E).

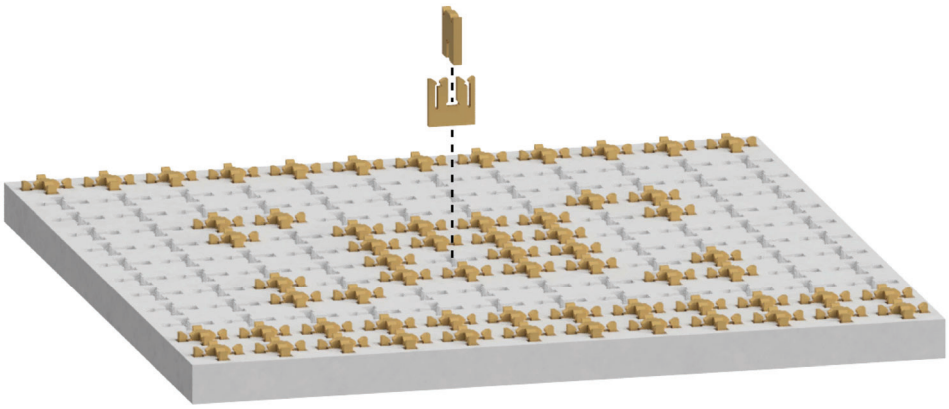


Figure 10-6: The build substrate is composed of individual flexural clips that are pressed into a plate with cutouts. The clips passively grip a specific part-type (node, strut, or actuator).




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Build Substrate** A build substrate is necessary to constrain the first layer of parts before they have been interlocked (Figure 10-6). For prototyping assembly, the build substrate is also used to position and constrain the feedstock of parts that are to be assembled. The substrate is composed of two parts: a flat plate with cutout features in X and Y axes spaced apart with the same pitch of the assembled lattice and individual clips for each part-type. The clips are designed to passively grip a certain part-type. They are cut using wire-EDM and are assembled to form an X-shape with flexurally preloaded tabs that hold a particular part in place. There are three different kinds of clips, one for nodes, one for struts, and one for actuators. The clips can be inserted or removed from the substrate in order to reconfigure the build area depending on what is being built.

Currently, when a build is finished, the assembly must be carefully removed from the bed of clips. A future version of the substrate may incorporate a single vertical degree of freedom mechanism that allows for all of the clips to be retracted simultaneously. This could simplify the removal process and reduce the chance of distorting the finished assembly.

**Control System** To run the assembler, I use an off the shelf controller (TinyG). This controller enables the simultaneous control of four stepper-motor driven axes (X,Y,Z,A) and supports G-code streaming and buffering to enable jerk-controlled motion. The additional functionality required for the active gripping of the toolheads and other peripheral functions is controlled through an external microcontroller (ATmega32u4) connected to the TinyG through the spindle and fan control ports.

The assembler is controlled by a browser-based custom developed software program. Communication with the TinyG happens through a USB-serial connection that communicates through a websocket to the browser. The browser console allows for command-line input to control the assembler but is also capable of running small macros or streaming whole G-code programs. For assembly automation, it's often beneficial to script the assembly tasks in javascript rather than writing a dedicated G-code program. Scripts have more flexibility to enable realtime control and automation (such as through the on-tool microscope camera) and programs that are able to call scripts and subroutines are more readable, modular, and reusable.

Each new toolhead needs to be calibrated for its X, Y, and Z coordinate offsets. To do this, the on-tool microscope is used to locate the assembler tool-axis directly over a building block part. The toolhead is lowered to just above the part, and is aligned to the part using fine-scale jog controls. The offset from the microscope center is recorded and stored in the javascript program. It is then recalled and used whenever the toolhead is active.

The geometry of the toolheads add some requirements to the order of assembly operations during a build. The actuator toolhead grips the actuator parts in a location that would interfere with other struts on that same layer. For this reason, the actuators must be the first objects placed after populating the node layer. Additionally, because of the vertical size of the actuators, the substrate node attachment point must be removed wherever an actuator is to be built on the first layer. Struts within the same layer can be placed in any order but it is often preferred to place the rigid struts first to ensure the node layer below is as rigid as possible before placing the flexible degrees of freedom.

### 10.3.3 Characterization

To demonstrate the capabilities of the assembler I assembled a walking motor. The motor is composed of 32 parts: 20 nodes, 6 rigid struts, 4 flexural struts, and 2 actuators. The assembly starts


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

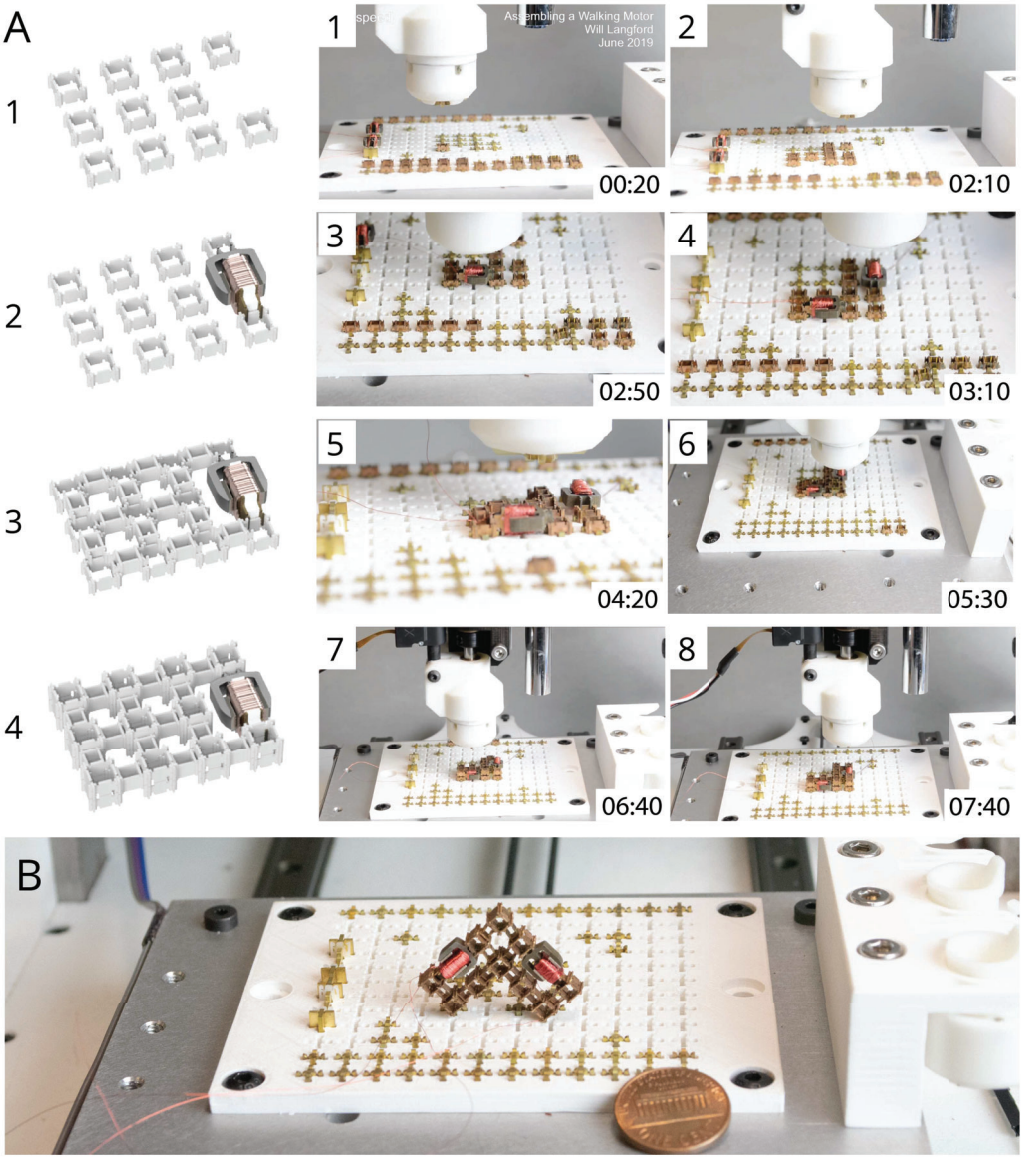



Figure 10-7: The assembly sequence (A) to build a walking motor (B). The assembly proceeds in stages with (1) the placement of nodes, then (2) the placement of actuators, then (3) the placement of rigid and flexible struts, and finally (4) the placement of the second node layer.



389
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

with a layer of 10 nodes. This is followed by the placement of the two actuators. The nodes are then connected by rigid and flexible struts and then finally capped off with another layer of nodes. The whole assembly process, including tool-changes, takes approximately 8 minutes (Figure 10-7). This is an equivalent rate of 15 seconds per part.

Much of the assembly time stems from the relative slowness of the actuated gripper and the z-axis. The z-axis is software limited to a maximum speed of 400 mm/min. Testing revealed that the z-axis is reliable up to approximately 800 mm/min, however, moving more slowly than this was found to be beneficial during early testing. Between picking and placing the assembler moves 13.5 mm above the pickup height. This move (both down and up), which happens four times per pick-and-place operation, accounts for 73.4% of the total time it takes to pick up a part and place it (Figure 10-8). This time could be drastically cut down by moving only the required distance for the picked part to clear the tallest point of the assembly (3 mm). This would cut the Z-move time from approximately 2.2 seconds to 0.48 seconds, and reduce the overall pick and place time by more than half, from 12 seconds to 5.12 seconds. The linear servo motor used in the gripping mechanism has a 100:1 internal gear reduction and so has a very high reflected inertia that must be overcome both to stop and start. As a result, a delay of one second after switching the gripper was found to be necessary for transients to settle to ensure that it is in the desired state before moving the toolhead. For a pick and place operation, this delay accounts for 16.6% of the time. Only 10% of the pick and place time is spend moving between XY locations.

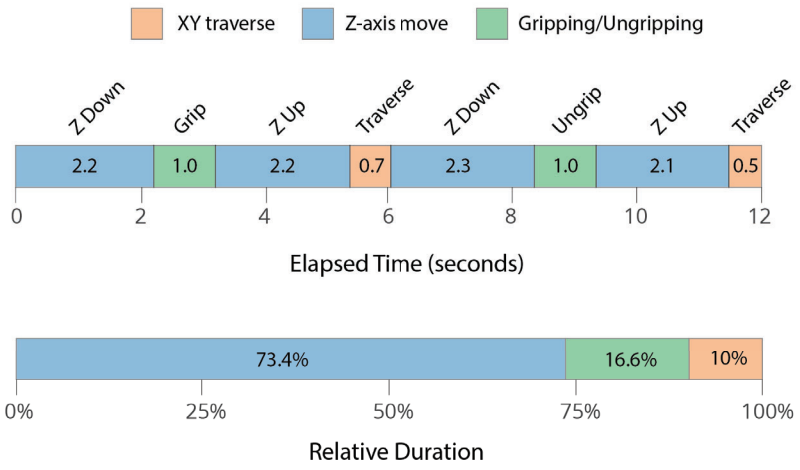



Figure 10-8: Analysis of the timeline during a single pick and place operation. Z-axis moves represent the majority of time spent.

### 10.3.4 Extensions

In the current embodiment of the assembler, the parts required for assembly must be pre-populated on the periphery of the build substrate. A future revision of the assembler would likely feature automated part feeders that each contain a particular part type and advance a new part as one


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

is picked up (Figure 10-9). This could be done using a variety of methods. I detailed the design of vibratory feeders for both nodes and individual struts. These feeders are capable of taking a randomly oriented bin of parts and feeding, orienting, and lining them up to be picked up at a given location. For node parts this involves a channel that is dimensioned to accept only a single node at a time in the preferred orientation. For individual strut parts it involves rotating the parts from their natural flat state into a vertically oriented channel.

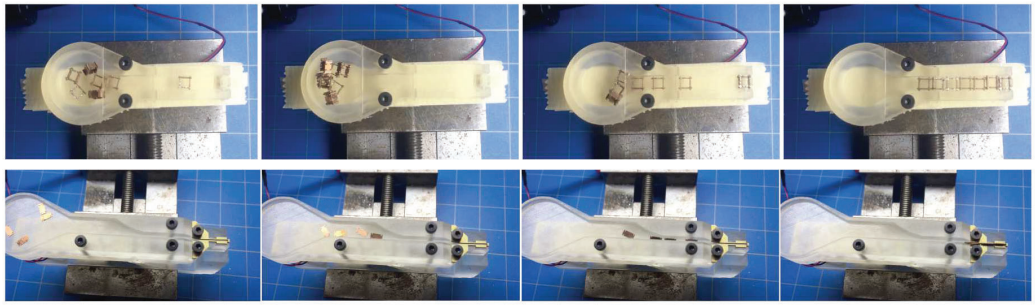


Figure 10-9: Vibratory part feeders can be used to align and feed parts to pickup locations.

Automated disassembly is desirable in addition to automated assembly. This would enable the rapid reconfiguration of systems and would mean that the parts wouldn't need to be single-use. While this is certainly possible to do in an automated way, the current toolheads and assembly machine do not support disassembly. This is because when pulling on a part on the topmost layer, there is no reason that the topmost part is removed and not a part on a lower layer with the same number of connections. Disassembly could be enabled with toolheads that apply downward force to neighboring parts while extracting the desired part. This could be accomplished with something as simple as a sprung ring that ensures that the removal force is directed at the specific part. Another way to accomplish disassembly would be use a different joint system. A joint system with an orthogonal clip for example, could enable a disassembly tool to unclip the part in question and then extract it.

## 10.4 Assembling Assemblers

Automated assembly systems like FABBR represent a way to go from a digital design file to an assembled machine. FABBR takes 8 minutes to assemble a walking motor composed of 32 parts. This is faster than an experienced human can manually assemble the same structure, but is far too slow if the goal is to assemble thousands, millions, or billions of parts.

Placing parts at the current rate, a billion element volume (a cube with one thousand parts on side) would take on the order of a human lifetime (80 years) to build. Even if the part deposition speed was sped up to 10 kHz (the speed of inkjet printing), it would take a full day to build the billion element volume.

Clearly it is important to find more parallel ways to assemble parts. Naturally, placing whole layers of the volume at once could be desirable. However, as the number of simultaneously assembled connections in a digital material is increased, there tends to be a corresponding increase in the error rate and assembly force [61]. As discussed in section 1.2.3, biology is able to scale manufacturing


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

throughput with the size of the manufacturing problem. It is able to do this because ribosomes can assemble ribosomes.

Inspired by the biological approach, this work has focused on parallelizing assembly by assembling assemblers. The range of functionality that can be assembled from the building blocks presented in this work lays the ground work for the assembly of machines that can move and manipulate their own parts. Assemblers need capabilities including locomotion, part manipulation, and integrated control.

- **Locomotion** While locomotion could be done in an unstructured environment with a conventionally designed walking robot, an assembling assembler needs motion to be aligned to a lattice or grid. This can be done, for example, with a grooved surface that registers imprecise local motions such that repeated steps are globally precise as long as each step falls within a certain tolerance. The walker demonstrated in Chapter 7 represents a basic step towards this locomotion capability. Walkers that are able to locomote in multiple directions may be made by combining two single degree-of-freedom walkers at a right angle to one another.

- **Part Manipulation** The assembled assemblers need to be able to pick up and place parts. This may be accomplished either with electropermanent latches or with mechanical grippers, as is demonstrated in Chapter 8.2.

- **Integrated Control** Each assembler will consist of a handful of degrees of freedom (≈7). In order to coordinate their actions, each assembler needs processors and control circuits on board in order to receive high level movement commands and translate them into controlled actuations.

In addition to these capabilities, the assemblers need to rely on a tailored environment. Just as ribosomes do not operate in a vacuum but rely on a supportive environment to deliver feedstock parts and energy, the assemblers built from these parts will need a supportive and structured environment in which to operate.

Between conventional pick-and-place style assemblers and assembled assemblers, are a family of machine configurations that may serve as stepping towards "true" self replicating machines. The concept pictured in Figure 10-10 depicts one of these. Mobile assemblers locomote on a lattice aligned grid in two axes (X and Y). This grid could be constructed in a number of different ways. The grid could be physically grooved to allow shaped feet to interlock. Or the grid could be defined based on spatially patterned permeability. For example, a permeable lattice with the pitch spacing of a single walker step would be physically flat but still allow the assembler to magnetically align itself on each step.

The assemblers are themselves built from the building block parts. They feature a gripper that is able to pickup parts and place them in the desired location. A conventional Z-axis serves two purposes: (1) it supplies the required force to press-fit an arranged part into place and (2) it moves up as the next layer needs to be placed.

Assembly would proceed as follows: (1) An assembler translates to a part pickup location on the lattice substrate. (2) The assembler picks up the part and translates to the target part location. (3) The part is placed, interfacing with grooves on the substrate to align to the build front. (4) The assembler moves out of the build envelope to pickup another part. (5) Once the assembler has exited


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

the build envelope, the Z-axis moves down to attach the placed part to the in-progress assembly. This process then repeats. Multiple assemblers could act simultaneously, each arranging a part on the substrate to be placed.

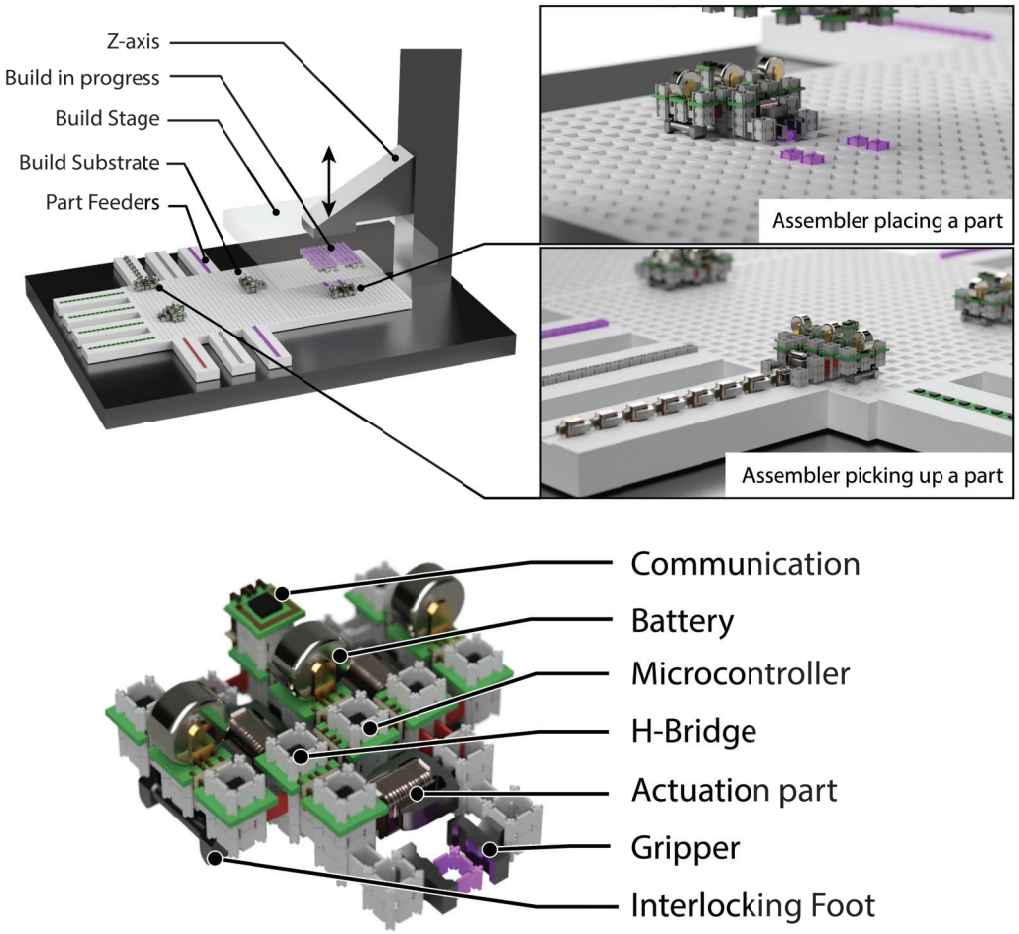



Figure 10-10: A stepping stone concept towards full self-replicating assembly. Mobile assemblers (bottom), assembled from building block parts, are able to locomote on a lattice-aligned grid, pick up and manipulate parts. Parts can be placed in parallel with multiple assemblers working simultaneously with the z-axis advanced as needed.

This concept encompasses the ability to assemble assemblers. However, it is limited in terms of throughput since the motion of the Z-axis needs to be coordinated with the displacement of the assemblers from the build envelope. To unlock the throughput benefits associated with assembling assemblers, the assemblers themselves need to be capable to placing parts without an external force generating structure. In another embodiment, the assemblers could feature a yet-undeveloped capability of a high-force actuator. Even with an actuator capable to generating large-forces, the 2 N normal strength of the electropermanent feet would be unable to support the 50 N loads required to assemble a node. Based on this, it is clear that the part geometry and joint system may need to




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

be adapted in support of assembled assemblers.

Beyond the relatively conventional layout of the concept depicted in Figure 10-10, with a few additional degrees-of-freedom, the mobile assemblers could enable more unconventional machine tool configurations such as the concept depicted in Figure 10-11. In this concept, assemblers have an additional body hinge that allows the assembler to locomote on arbitrarily oriented surfaces and transition between horizontal locomotion and vertical locomotion. This could enable assemblers to operate simultaneously on many build fronts of a single object.

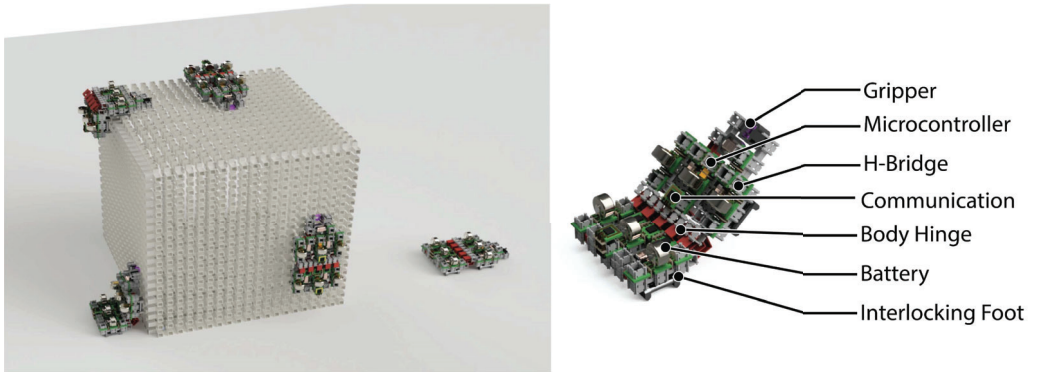



Figure 10-11: An extra degree-of-freedom hinge-joint could enable the assemblers to work on other surface orientations and allow for three-dimensional structures to be build without a conventional z-axis.

### 10.4.1 Current Limitations

A major question that has come up in this assembling assemblers work is, How can the assemblers generate the required force to attach a part? In the current embodiment of this work, force required to locomote (and the force that is generated by the actuator components) (50 mN) is three orders of magnitude less than the force required to insert a part (50 N).

One strategy to generate this force is could be to use force amplifying linkages. The maximum mechanical advantage created by a lever-arm linkage is often limited to on the order of 10:1 given spatial constraints. Generating a 1000:1 reduction would require three of these 10:1 stages and, as a result, seems impractical. Alternatively, a toggle-action linkage can generate considerable amounts of mechanical advantage but only over a very short stroke. These linkages may be built directly on the assemblers themselves, but space constraints may mean that it makes more sense to assemble linkages between mobile assemblers. In this envisioned concept a part may be placed by an assembler and then pressed into place by a pair of mobile assemblers that manipulate a toggle-action linkage between them.

Another strategy is to obviate the need to massive force amplification entirely by designing the parts to assemble with very low forces. As discussed in Section 2.3, there are a number of different joint systems that require a minimal amount of force to attach parts.

As is discussed in Chapter 7, this kind of mobile assembler concept hinges upon reliable open-loop positioning. An important next step for this work should be the development of a structured surface that constrains each step of the mobile assembler to a grid. This could enable precise open-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

loop positioning over an arbitrarily sized build envelope and remove the usual inverse relationship between build envelope and precision that is common in conventional gantry-based machine tools [148].





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 11

# Scaling

While I have focused development at the millimeter-scale, other application domains may be possible by scaling the constituent parts and assemblies. For example, with smaller micro-scale building block parts, applications requiring finer-grain resolution can be explored such as ingestible medical robots [149] and dexterous microsurgical tools [150]. On the other hand, larger parts could be used to assemble architectural scale system that are not only structural but also functional.

In this chapter I will look at some of the opportunities and challenges in scaling this approach both down and up in length scale. I'll look at the methods to produce parts at a range of different sizes, the physics that determines the scaling of actuation, and the potential for the assembly process to be scaled both down and up.

## 11.1 Part Production

Because the parts used in this assembly method are two-dimensional and have a relatively simple geometric shape, they can be made using a range of different processes, in a range of different materials, and at a range of different length scales. As such, the choice of length-scale should be seen as a parameter that can be adjusted based on the application at hand.

For prototyping, I often started with centimeter-scale parts to begin exploring how certain part types could be composed to build useful mechanisms and machines. This allows for rapid fabrication and iteration using commercial lasercutters. The parts at this scale are 0.8 mm thick and just over 1 cm in their longest dimension. Acetal plastic (Delrin) works well at this scale for the purpose of prototyping. More robust parts can also be made by wire-cutting metal.

The millimeter parts, which are the parts predominately used in this work, are made using the wire-EDM with the 100 $\mu$m wire, which cuts with about a 150 $\mu$m kerf. The parts are 3.5 mm in the their longest dimension. Laser micromachining can also be used at this scale but it is much slower than wire-EDM and only useful for very rapid prototyping of one or a few parts.

Using laser micromachining allows for the fabrication of even smaller parts. Using a 532 nm diode-pumped solid state laser (Oxford Lasers Inc.), which has a kerf of about 20 $\mu$m, it is possible to make parts that are 75 $\mu$m thick and are 1 mm long in their longest dimension.

Even below this scale, a femtosecond machining laser (Pharos) is able to approach micron feature sizes. This is possible because the pulse lengths are so short that the peak power is high enough to




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

break atomic bonds without ablating the material, reducing or eliminating the heat affected zone [151].

Beyond scaling down in size, it is also worth thinking about how part production can be scaled up in throughput. Doing so for the basic rigid part types is relatively straightforward as the simple two dimensional geometry can be stamped with a single tool and die. To mass produce the flexural part types, roll to roll processes seem attractive. A web of the constituent materials (bronze and polyimide) would enter the machine and be tape cast with an adhesive film, stamped, and then pressed between heated rollers, and then singulated by another stamping press. Circuitry parts can leverage PCB fabrication capabilities but streamline the processes to only include the necessary steps. Actuators are likely the most difficult to produce in very large quantities. Coils and mandrels can likely use be made in roll-to-roll fashion with progressive stamping, bonding, and then automated coil winding. The magnetic cores may require a redesign in order to simplify their production at scale to eliminate the fine feature sizes in multiple orientations.

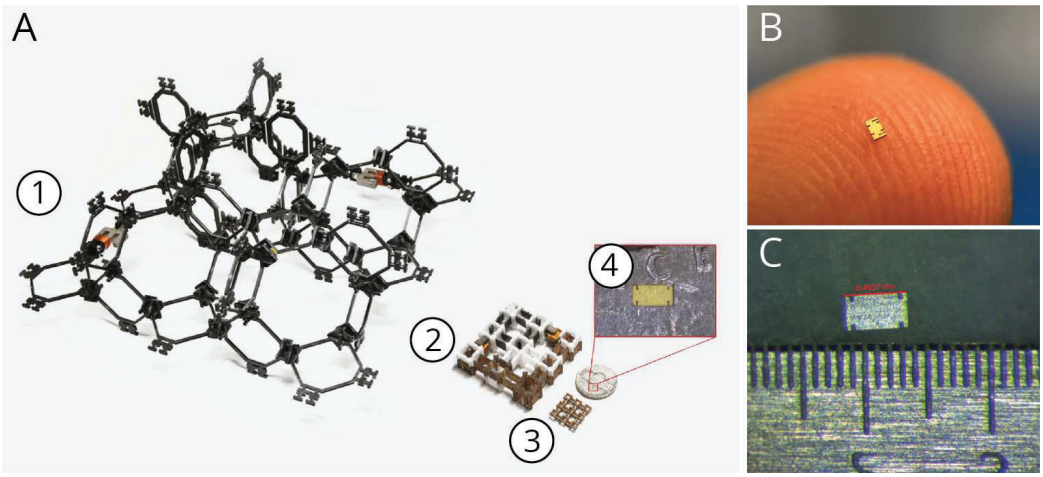



Figure 11-1: This assembly method is possible at a wide range of length-scales from decimeter (a) to centimeter (b) millimeter (c) and sub-millimeter (d).

## 11.2 Actuation

The discrete assembly approach is relatively scale agnostic. Colleagues and I have assembled walking motors, for example, at scales ranging from decimeters to millimeters (Figure 11-1). At the larger scale, Copplestone used a Kelvin lattice topology for its sparsity and ability to tile three-dimensional space. The same principles developed in this work at the millimeter scale are applicable at this larger scale, enabling structure, mechanism, and actuation to be combined.

Actuation is generally a challenge to scale since it often involves the most intensive fabrication and relies on a variety of physics working together (e.g. electricity and magnetism). It is possible to project the performance of the actuated assemblies developed in this work to other length scales through a proportionality scaling analysis (Table 11.1). The scale variable, s, is used to represent how a particular quantity scales, as is detailed in [108]. In this case, I assume that the actuation




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

current can scale with a constant temperature rise ($s^1$), rather than constant current density ($s^2$), because surface area scales favorably with respect to volume. Based on this assumption, both power and force (as well as their respective densities) scale favorably to smaller length-scales. Furthermore, bandwidth also scales favorably, proportionally increasing for every shrink in size. Efficiency, however, suffers at smaller length scales and decreases proportionally with scale.

Table 11.1: Scaling Laws for Voice Coil Actuation

| | | | |
|---|---|---|---|
| Mass | $s^3$ | Power | $s^2$ |
| Current | $s^1$ | Power Density | $s^{-1}$ |
| Force | $s^2$ | Force Density | $s^{-1}$ |
| Bandwidth | $s^{-1}$ | Efficiency | $s^1$ |
| Energy Output | $s^3$ | Energy Density | $s^0$ |

This points to the possibility of scaling this assembly approach down in length-scale. At some-point, however, the increase in bandwidth, force, and power density will be outweighed by the decrease in efficiency and will warrant a different type of actuation (for example, piezoelectric or electrostatic).

It can be instructive to look at the force density of an array of actuators to look at how these actuators stack up against larger commercial off-the-shelf (COTS) options. A representative commercial voice coil actuator [113] (25 mm diameter, 16.5 mm length, 58 g mass, 2.2 N blocked force) has a force density of 38 N/kg or 0.272 N/mm$^3$. For comparison, densely packing the Lorentz force actuator components and neglecting any required wire routing or mechanical constraints results in a force density of approximately 82 N/kg or 0.114 N/mm$^3$. Clearly the actuators developed in this work are much less force dense per volume but per mass, are more than two times stronger.

Ultimately the force per mass metric is likely more significant than the force per volume metric because, in the target applications, what usually matters is the ability to accelerate a mass, not overcome viscous forces. Furthermore, these actuators being much less dense than the COTS option is likely beneficial in terms of heat exchange with the environment. Since efficiency decreases as the actuators are scaled down, it becomes more and more important to efficiently remove heat from the coils.

This force density metric points to the possibility of replacing larger macroscopic actuators with arrays of these actuator components. By tiling the actuators in parallel, their force output can be increased. Furthermore, it is possible to arrange actuators in series as well, to increase displacement. This approach of assembling a large scale actuation from individual "micro-actuation" building blocks is likely not the most power dense or space-efficient because of the overhead of connectors between them, but enables the rapid production and extensibility of robotic systems.

Building up larger actuations from individual building blocks is exactly how biological muscles work. In fact, biological muscles demonstrate the feasibility of assembling very high force density actuators from individual building blocks. Muscle generates a specific actuation stress (N/m$^2$/kg) of approximately 2000 kPa/kg [15], which is about 25 times the specific actuation stress possible with any kind of Lorentz force actuator.

Biological muscle force scales with the cross-sectional area of the muscle ($s^2$). They represent a remarkable design that is able to produce large macroscopic forces and strains based on the motion of billions of microscopic parts [15]. There are many different kinds of biological muscles. Skeletal


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

muscles, those muscles that we have conscious control over, are powered by the molecular motor protein, myosin II. Each of these molecular motors takes discrete steps approximately 5-10 nm in length along actin filaments, generating on the order of 5 pN of force. These filaments are arranged into segments called sarcomeres that expand and contract based on the motion of these individual proteins. These sarcomeres, each of which is a little over 2 $\mu$m in length, are the basic units of the higher order muscle fibers. Muscle fibers, each of which is a single cell, can span the full length of the muscle so a 20 cm long bicep, for example, may contain as many as $(20 * 10^{-2})/(2 * 10^{-6}) = 10^5$ sarcomeres. Each sarcomere can expand and contract at a rate of approximately 10 $\mu$/s, with each of the 300 myosin heads per thick filament taking roughly 5 steps per second [152]. This remarkable feat serves as inspiration for how it might be possible to assemble fungible and extensible actuators from individual building blocks.

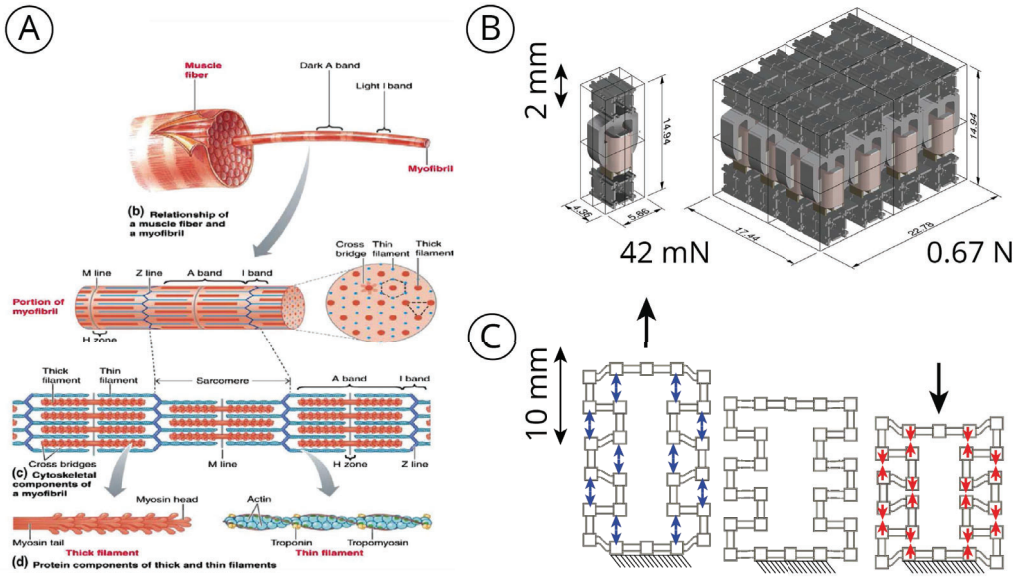



Figure 11-2: Biological muscles create macroscopic forces and displacements based on the motion of billions of microscopic parts arranged in hierarchical levels (A) [153]. The Lorentz force actuation components can be arranged in parallel to increase the force output (B) or in series to increase the displacement (C).

## 11.3 Assembly

Assembly at smaller length scales is often complicated by decreased positioning accuracy and difficulties in part manipulation because of the scaling of surface forces relative to inertial ones. I expect the discrete assembly method to be more amenable to high throughput assembly at small scales for a number of reasons.

- **Error Correction:** With the use of compliance in the toolhead, the assembler can take advantage of the interlocking nature of the parts. The shape of the interlocking slots enables



399
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

a degree of error correction and prior work has shown that the degree of positioning error tolerance can be as high as 39% of the part spacing [22].

- **Well-Defined Adhesion Forces:** Conventional micro-manipulation and micro-grasping often struggles with controlling the adhesion forces between the part and the tool and, in particular, it can be difficult to release parts at small scales because of surface forces. In this assembly framework the part-part interlocking forces can be made orders of magnitude stronger than the part-tool gripping force. Furthermore, with an actively driven gripping mechanism, the part-tool adhesion force can be modulated by more than an order of magnitude, enabling parts to be both picked from and placed into passive flexural bonding sites without the uncertainty of "sticking" parts.
- **Environment Agnostic:** Unlike many other micro-assembly strategies, which rely on liquid environments to suspend and manipulate parts, this one does not impose any constraints on the environment.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 12

# Modeling

Commercial design and simulation tools struggle to model these kinds of discretely assembled structures that integrate a wide variety of functionality. While some multiphysics tools can model an actuated mechanism and structure (e.g. COMSOL), even these tools do not scale easily to encompass the dynamic range (maximum extents per feature size) of discretely assembled functional structures. Furthermore, no commercial tool exists to design control systems and circuitry embedded within the mechanical degrees of freedom that they control.

Instead, my colleagues and I decided to build our own tools to support end-to-end workflows for discrete assembly. One of these tools, DMDesign, developed by Amanda Ghassaei spans design, simulation, toolpathing, and machine control for electronic digital materials [147] [22]. Another tool, AMOEBA (also developed by Amanda Ghassaei), grew out of DMDesign to encompass a greater range of functionality including mechanisms, actuation, and control systems [147]. This tool enables users to build up integrated systems one building block at a time but also supports hierarchical representations, allowing users to abstract modular motifs that can be reused across a range of designs.

In this section I'll overview some approaches that have been used to model these kinds of integrated structures. I specifically focus on compliant mechanisms since the large displacement analysis of flexible structures offers the biggest challenge in modeling these systems efficiently. Furthermore, I detail a design tool developed to model and optimize discretely assembled compliant mechanisms.

## 12.1 Modeling Approaches

A number of different methods can be used for the simulation of mechanisms. Conventional mechanisms, which are typically made up of rigid links and revolute or prismatic degrees of freedom, are often simulated using rigid body models. While these are often analytical for mechanisms with only a handful of links and joints, numerical methods also exist. Automated analysis of arbitrary rigid body mechanisms is often complicated by the need to find closed kinematic loops to express the equations of motion. This approach is very accurate and efficient when the user has some knowledge of the mechanism being solved, but can problematic with mechanisms of changing and various topologies. Singularities, for example, are a classic problem in kinematic analysis in which the mechanism locks up and has multiple equally possible trajectories [154]. Compliant mechanisms, on the other hand,




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

are often modeled using finite element methods. For mechanisms undergoing large displacements, this necessitates nonlinear finite element solvers [155]. A third approach, which is discussed by Jin et al., sits between these two approaches. Rather than modeling motion through the deflection of static elements, the authors describe a pseudo rigid body model in which virtual torsion springs are attached between each rigid bar. The problem of solving for the nonlinear kinematics of an arbitrary mechanism is then posed as an optimization which seeks to minimize the potential energy stored in the virtual springs [156]. Aviles et al. details a similar approach and does a more thorough treatment of the formulation with Lagrange multipliers to efficiently solve the static equilibrium problem [157].

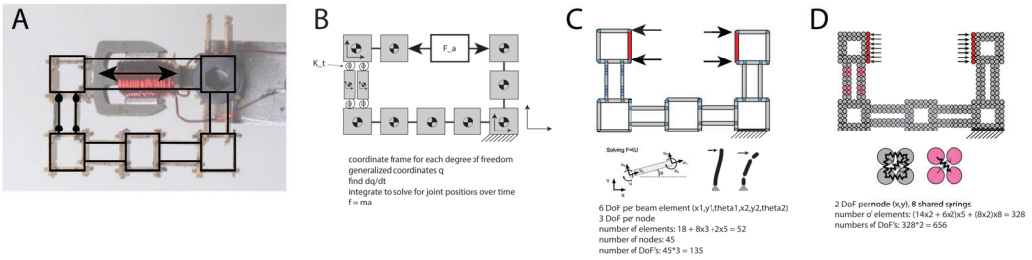



Figure 12-1: A number of modeling approaches can be used to model an actuated mechanism (A) including pseudo-rigid body models (B), plane-frame models (C), and discrete element models (D).

## 12.2 Optimization Approaches

Prior work in the topological optimization of compliant mechanisms has largely focused on two approaches: continuum-based methods and ground structure-based methods. Sigmund shows that the formulation of the compliant mechanism problem using these frameworks is very similar to the problem of optimizing for light, stiff structures. However, rather than simply minimizing compliance, the objective is to minimize compliance while effectively translating an input force or displacement into desired output work. In this case, Sigmund outlines an objective function based on mechanical advantage between the input and output force subject to constraints including the volume of material and the minimum and maximum densities each pixel can represent [158].

For the optimization of discrete-valued structures which aren't amenable to relaxing a continuum problem, heuristic methods like genetic algorithms are often used. Jin et al., for example, demonstrate a genetic algorithm implementation to find suitable mechanisms given bounding constraints for their pseudo rigid body models (in this case, a displacement inverting mechanism) [156].

## 12.3 Compliant Mechanism Simulation and Optimization

From a structural optimization perspective, discretely assembled mechanisms offer a unique advantage over conventionally fabricated devices. In contrast with prior topology optimization work, which has largely tried to approximate continuum materials with discrete voxels or ground structure elements, the discrete assembly approach enables a discrete-valued optimization of truly discrete materials; in essence, the physical parts are the finite elements themselves. I developed a nonlinear


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

finite element plane-frame solver capable of quickly simulating discrete mechanism problems. This solver is integrated in a browser based interface with interactive controls to aid in exploration of the discrete design space. Since the discrete design space is not well suited to traditional gradient based optimization methods, I developed an interactive heuristic for sequentially replacing parts to maximize displacement in a particular direction.

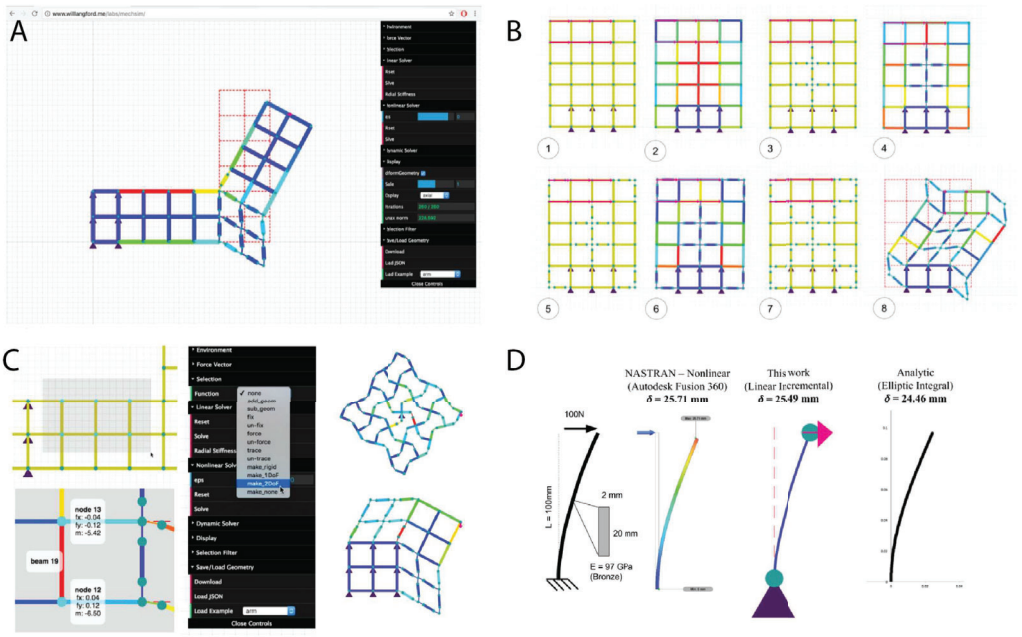



Figure 12-2: A browser based compliant mechanism design and simulation tool (A) can be used to interactively optimize the design of compliant mechanisms (B). Interactive features allow the visualization of forces and moments in the structure to inform the design (C). The simulation is physically meaningful, matching closely results from legacy finite element solvers and analytic solutions.

This tool is focused around planar mechanisms and so is purely two-dimensional. As such, each each member has six degrees of freedom: two translational degrees of freedom and a rotational degree of freedom at each node. The design tool allows users to visualize the stiffness of the mechanism in all axes with the help of a radial stiffness plot. This helps give the user an intuition about the design space by allowing them to, for example, substitute rigid members with flexural parts and note the change in stiffness of the mechanism.

The two main computational expenses in solving the large-displacement problem are in assembling the system stiffness matrix and solving the system. The system stiffness matrix is inherently sparse and symmetric. This is especially certain in my case because of the topology of the discretely assembled mechanisms: there are no beam elements that span outside of their local neighborhood. This trait can be exploited to more efficiently assemble and solve the system by using sparse matrix data representations.

This software tool represents a design study of a simulation and optimization tool for compliant mechanisms. Future work should focus on extensions from 2D to 2.5D to 3D, integration of more



403
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

physics (e.g. electricity and magnetism), and optimization methods that allow users to specify desired outcomes and have the tool place individual parts.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 13

# Applications and Extensions

Based on the work demonstrated in this thesis a number of applications and extensions of the work are currently being pursued.

## 13.1 Discrete Integrated Circuit Electronics (DICE)

One of extensions of the work being pursued is the assembly of reconfigurable high-performance computing. Much like the state of robot design and fabrication, the development of integrated circuits today often means choosing between the cost, expense, and long turn-time of an application specific integrated circuit (ASIC) or bulky and less performant (but quicker to develop) integration of existing integrated circuits (IC's). An assembly based approach for integrating heterogeneous "chiplets" could enable the configuration and re-configuration of compute hardware tailored to the computational problem to be solved (Figure 13-1). Potential benefits of this approach include reducing the development turn-time of a novel compute solution, rapidly reconfiguring systems to reflect changing workloads, on-demand production, and increasing system integration across the existing boundaries between chips, packages, boards, and interconnect.

A number of techniques have been proposed for ways to flexibly and densely integrate heterogenous chiplets. In comparison with passive interposer approaches [159] [160], discrete assembly allows for reconfiguration and on-demand fabrication such that the routing between chips is not baked into hardware, but rather, can be incrementally built, modified, and extended upon. In comparison with approaches that use active interposers, or "switch fabrics" [161], discrete assembly offers the possibility of reconfiguration without the power and spatial overhead of crosspoint switches and FPGA's.

By scaling the size and complexity of the component parts, it is possible to balance the tradeoffs between the overhead of the connections and the refinement of the discretization. A number of components could be integrated within this framework including CPU's, GPU's, FPGA's, memory, analog, and input-output parts. Each of these component types come in a wide variety of capabilities and sizes. Looking at CPU's for example, the smallest commercial off the shelf (COTS) packaged microcontroller is the ATtiny20-UUR at just 1.4 mm by 1.55 mm. This 12-pin part is an 8-bit single cycle per instruction microcontroller that operates up to 12 MHz. One step up from that in complexity and in similarly small packages ($<$ 2 mm x 2 mm) are microcontrollers like the Kinetis


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

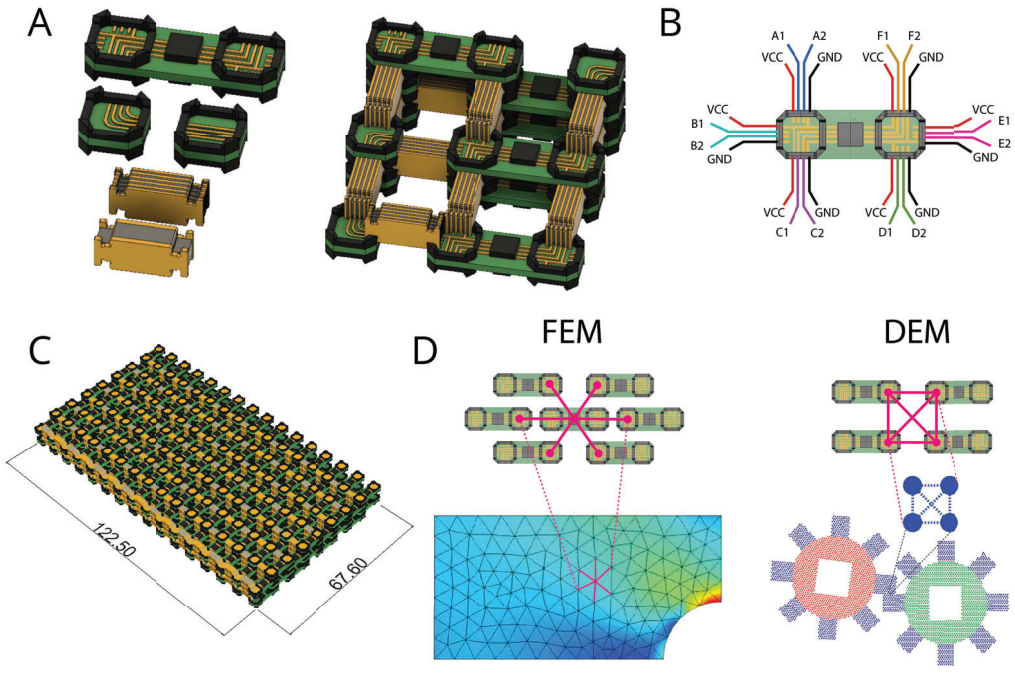



Figure 13-1: DICE application overview. (A) Discrete parts with embedded integrated circuits including microprocessors, FPGA's, analog peripherals, and memory. (B) The parts interlock with neighboring parts and have a number of input and output ports that can be connected to form a range of different topologies. (C) 256 microcontroller building blocks match the operations per second of a pentium processor. Configuring and reconfiguring to match the structure of the problem allows for potential speed ups from parallelization. (D) Reconfiguration could enable the physical architecture of the computer to mirror the computational structure of the problem being solved. This is particularly evident in problems like finite element analysis (FEA) and discrete element methods (DEM).

KL03 and Maxim MAX32660, an ARM Cortex-M0+ and ARM Cortex-M4F, respectively. These 32-bit processors are capable of running up to 96 MHz but have a 2 or 3 stage instruction pipeline. The Maxim part has an on-chip floating point unit which enables just 1-3 cycles per FLOP.

It is instructive to look at some very back-of-the-envelope numbers to get a sense of how many of these smaller CPU's are necessary to match the specifications for current state of the art computer processors. I take the 14-nm Intel i7-8700 Coffee Lake processor as the current state of the art. The base clock speed of 3.7GHz sells for $360, consumes 95 W of power, and takes up a footprint of 37.5 mm x 37.5 mm x 4.4 mm. To achieve a similar instruction throughput with the 40-nm Maxim MAX32660 (neglecting bandwidth and latency constraints), one would need about 40 processors (3.7 GHz / 96 MHz). Mounting those processors on DICE parts with a 50% spatial overhead in each linear dimension results in an assembly that is approximately one quarter of the volume (12 mm x 12 mm x 4 mm) and one quarter of the cost ($80) of the monolithic i7 processor. Furthermore, the power consumption of 40 MAX32660's (15 mW each) is less than 1 percent of the 95 W consumed by the Intel processor. While this number neglects real-world considerations such as available


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

bandwidth between processors and the intricacies of sharing memory, it represents the aim of this approach, which is to demonstrate the feasibility of assembling high performance computing from more basic constituent parts in a way that enables incremental extensibility and reconfiguration based on computing requirements.

A key driver for heterogeneous integration is the ability to combine parts made using dissimilar processes. For example, it is often beneficial to keep analog circuitry more spaced out while taking advantage of the ability to shrink the digital circuitry by using a more advanced process node. For this reason, DICE components will not just include microcontroller processors and other digital circuitry, but also analog-to-digital converters, GPIO expanders, and wireless components. While the aim of this work would be to eventually demonstrate integration of discrete chiplets designed specifically for this kind of discrete assembly, it is possible to demonstrate the feasibility of this approach using commercially available off the shelf components (Table 13.1).

| | | | | | |
|---|---|---|---|---|---|
| Microcontroller | MAX32660 | 96 MHz | 1.6 mm x 1.6 mm | 14 IO | $1.5 |
| FPGA | MACHXO2 | <400 MHz | 2.5 mm x 2.5 mm | 18 IO | $5.2 |
| Op-Amp | MAX4292 | 500 kHz GBP | 1.5 mm x 1.5 mm | – | $1.6 |
| ADC | MAX19777 | 3 Msps | 0.9 mm x 1.4 mm | 2-ch | $0.9 |
| Memory | M95128 | 200 MHz | 1.2 mm x 1.2 mm | – | $0.45 |

Table 13.1: DICE parts that are being considered as starting points for the assembly and reconfiguration of high-performance computing systems.

This assembly methodology potentially opens up novel computer architectures that are not practical with existing methods. For example, the ability to assemble three-dimensional volumes of individual processors could enable the realization of much more parallel [162] [163] and defect tolerant [164] [165] computer architectures. Furthermore, because this discrete assembly method enables reconfiguration, the architecture can be reshaped based on the problem to be solved. Many compute-bound problems have a structure that can be exploited. Multiphysics simulations using finite element analysis or discrete element methods, for example, are governed by local interactions between neighboring vertices or nodes. By structuring the physical hardware to match this computational structure, the computation can be done almost entirely in parallel.

The hypothesis is that assembly of heterogeneous integrated circuits can be radically simplified by standardizing the the assembly interface between parts and allowing their interlocking geometry to dictate the precision of their placement. Evaluating the assembly automation of DICE parts will not so much prioritize throughput as it will the time it takes to go from idea to physical embodiment. Discretely assembling integrated electronics should enable users to go from idea to design to physical construction in hours, not months. In addition to this, the error-tolerance of the discrete assembly approach is critical to be able to scale down to the smallest pitch-densities that are desired (tens of microns).

## 13.2 Relaxing Constraints

This thesis details an approach to robotic construction that increases flexibility and lowers barriers to integrating functionality. This approach was largely developed within the confines of a number of strict constraints. These include, for example, the use of reversible mechanical connectors between


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

parts, the conformity to a regular rectangular lattice, and the limited part-set (with few exceptions). Implementing this approach relies on a relatively significant infrastructure. A cheap and accessible supply of building block parts is necessary. Additionally, an automated assembly tool is needed in order to realize the on-demand and direct-write fabrication possibilities. Not everyone reading this will have access to these kinds of materials and tools. As such, I think it is helpful to enumerate and elaborate on a number of possible relaxations of this work that make this work more generally implementable without the infrastructure requirements.

**A modular design tool.** In one instantiation, a design tool featuring discrete functional parts enables the design, simulation, and verification of a machine or system that incorporates structure, mechanism, and actuation. From this design, various physical instantiations of the machine could be "compiled." An assembler could assemble the system from the building blocks described in this work. But there are other ways the device could be compiled including as monolithic or semi-monolithic digitally fabricated parts.

**Mass-produced Actuator Components** The actuator components developed in this work are the most fabrication intensive and require non-standardized processes to develop and fabricate. Getting a version of these actuators fabricated in mass quantities would not be a significant challenge using conventional manufacturing processes. These discrete actuator components could then enable the extensible assembly of robotic systems.

**Monolithic compliant mechanisms.** The design could be converted into a monolithic, single material representation that could be additively manufactured or laser cut (Figure 13-2). I actually employed this approach in testing some of my mechanism ideas. It allows for very rapid prototyping of structures and mechanisms that exhibit the same qualitative behaviors of the discretely assembled counterparts. This approach of monolithic construction doesn't enable the same kind of heterogeneity as does the discrete assembly method and the mechanisms don't have the same degree of anisotropy, but it makes accessible the modular approach developed in this thesis.

**A ready supply of node parts.** In another envisioned relaxation of this work, a mass produced node part type enables manual assembly integrated systems from laser-cut, 3D printed, or printed circuit board parts. This node part type could function as a "universal" connector that mechanically joins two dimensional parts made using these various processes.

In this concept, the nodes serve a purely mechanical role in connecting parts. Electrical connections would be made using PCB panels that interface directly with the nodes and other parts (Figure 13-3).

Given this supply of node parts, the construction of custom integrated robotic systems would begin with users designing their integrated system using modular building blocks in a digital design tool. The system could be simulated and verified for operation. Compilation of the design into a physical artifact could synthesize the design into elements that can be manufactured using commonly available tools (at least for a robotics lab) such as 3D printers, laser cutters, and PCB circuit designs. These parts would be cut, and then manually integrated through the use of the node connectors. Furthermore, a free parameter exists to determine the level of discretization. The design tool could output individual building block parts to be laser-cut, for example, or it could output a whole layer


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

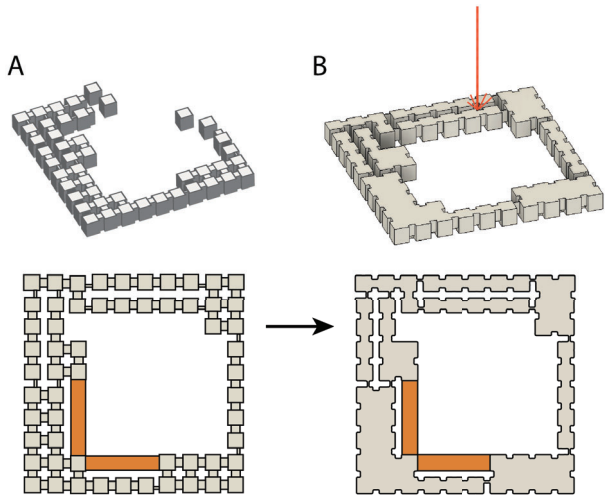



Figure 13-2: One possible relaxation of this approach involves compiling the discrete design (A) into a monolithic construction (B) that can be fabricated using 3D printers or laser-cutters.

of the device that features cutouts that interface with the node parts. This strategy allows for the integration of parts made using a variety of different processes while eliminating the need to assemble many individual parts where there is no need (Figure 13-3).

The nodes could serve purely mechanical role in connecting parts these various kinds of parts, or could be made using more advanced processes to incorporate electrical routing. The manufacturing of these more advanced node parts would likely be done in a similar way to how commercial electronic connectors are made, using stamped contacts and over-molded liquid crystal plastic bodies, for example. To encompass all of the possible routing degrees of freedom, a family of node parts would be necessary. Each of these having a particular routing arrangement such as passthrough, four-way, three-way, and no-connection.

These electrical nodes would enable the discretization of the electronic parts and allow connections to be made through layers of the assembly rather than remaining within a single plane.

**Access to digital fabrication tools.** At a larger scale than the millimeter parts developed in this work, these node parts could even be 3D printed, enabling very rapid prototyping of these kinds of assembled systems with commonly available digital fabrication tools. Custom node-parts could be developed to interface with off the shelf actuators and sensors.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

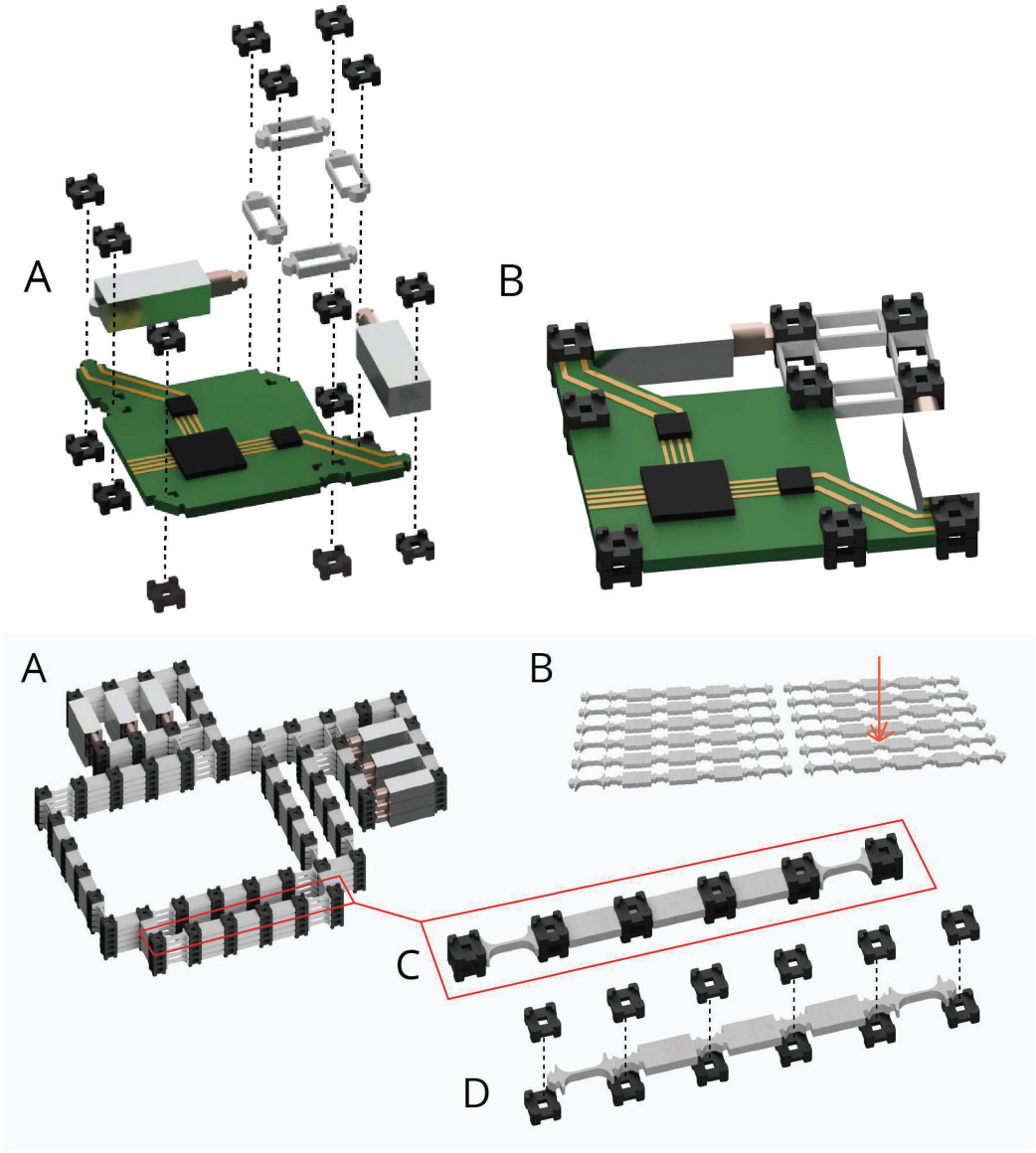



Figure 13-3: Another possible relaxation of this work enables the assembly of integrated machines from lasercut, 3D printed, and PCB parts using mass-produced nodes to build a walking motor (Top). These assemblies need not discretize at the finest level everywhere. The assembly can be broken down into sub-parts that still feature cutouts to interface with node parts to build a 2D positioner from semi-monolithic parts (Bottom).


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 14

# Conclusion

As we look to make devices that are more integrated and that combine more functionality through digital fabrication, current research has largely avoided assembly, preferring to try to integrate various functionalities within a single monolithic process. While doing so has extended what is possible and enabled new kinds of integrated robotic systems, I argue that it is ultimately limited by inherent constraints of each process and that no single monolithic process will be able to span the full range of desired functionality.

Instead, the approach introduced in this thesis represents a unified approach to robotic construction and a step towards on-demand robot fabrication. Rather than integrating functionality in a single process, it looks to enable integration through the standardization of assembly interfaces. Parts, with these assembly interfaces, can be produced using a wide variety of separate and dedicated processes, often enabling parts that are higher-performance and more functional than they could be in a monolithic process. This includes millimeter scale actuation components that have among the highest force density of any recently reported electromagnetic actuator as well as anisotropic compliant mechanisms that are more than two orders of magnitude stiffer off-axis as they are in the desired direction of motion.

Based on the capabilities developed in this work (structure, mechanism, actuation, circuitry, and control) a number of machines have been assembled that integrate these in various ways. A walking motor and walker, MOTILE, demonstrated two methods to convert small cyclic steps of discrete parts into long range motion. While this work has demonstrated the assembly of bidirectional linear motion, these systems can be thought of as modular primitives from which more powerful and complex degrees of freedom can be assembled.

The assembly process has been automated for structure, mechanism, and actuation part types, enabling a physical machine to be "compiled" directly from digital designs. The assembly platform has been shown to be able to take advantage of the interlocking joints of the discrete parts to build assemblies that are more precise than the machine itself. The speed of assembly currently exceeds that of manual assembly but requires time and effort to layout the feedstock of parts. Further developing the part production pipeline, going from raw materials to building block parts at a known pickup location on the assembler, would unlock the full potential of automated assembly.

This work interpolates between the field of modular robotics and digital fabrication, developing a set of relatively simple parts that can assemble relatively complex machines. In doing so, it lays



411
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

the groundwork for a viable path towards self-replicating machines by assembling assemblers. In this work, I demonstrated the basic capabilities that are necessary to do this including locomotion, part manipulation, and integrated control. In order to assemble assemblers, however, a few more challenges need to be overcome. First and foremost, control needs to be integrated within the same assembly as structure, mechanism, and actuation. Design studies and experiments have shown that this is nearly possible but higher-density electrical connectors between tiles need to be made more robust and reliable. Second, locomotion needs to register to a lattice. Currently, the exact step-size of the walker varies slightly from step to step. Assembling assemblers will critically rely on reliable open-loop positioning of machines by registering each step to a grid or lattice. Finally, the disparity between the force generated by the actuation components and the force required to place a part needs to be overcome. This likely warrants a redesign of the joints between parts to enable separate tuning of assembly and disassembly forces.

In the same way that computers and digital technology rest on the ability of a small family of gates to evaluate arbitrary logical expressions [166], this work points to a small family of building blocks that enable the construction of arbitrary robotic capabilities. While the design and fabrication of robots today often requires the integration of many diverse and custom parts, by standardizing the part set through the development of discrete assembly workflows, it is possible to take advantage of the same benefits inherent in computational universality and enable more flexible, inexpensive, and rapid design of robotic systems. The work developed here shows the utility of even a limited part set and the opportunity for future work to refine the part set to encompass a full range of robotic capabilities.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Appendix A

# Lessons Learned

One of my favorite chapters in Ara Knaian's thesis is his chapter on lessons learned [58]. This chapter is very much inspired by that.

Here are some things that I've learned along the way:

- **Balance thinking with doing.** Even if you know an experiment is not going to work, you learn a lot in the process and often spark new ideas or directions.
- **Document early and often.** This means taking pictures and videos even when things aren't working perfectly. Often it's these work-in-progress photos/videos, which can feel like a waste of time in the moment, that end up becoming the things that encapsulate a whole project.
- **Don't be afraid to pivot when things aren't working.** I spent a while trying to build a transistor out of press-fit parts with very little progress. Eventually I decided to just put a transistor on a part and that spiraled into this thesis.
- **But also, don't give up an idea just because it gets hard.** It can be very tempting to switch gears to another project or problem when something gets hard. Sometimes this is good to do, but sometimes you end up bouncing between projects, not really making progress on any one in particular. Balancing these is a continual process.
- **Balance process development with end results.** It can be easy to get sucked into developing better processes and tools to make the thing you want to make faster, better, and cheaper. Developing better processes and workflows is great and makes results more reproducible and consistent. But spending too much time on it can detract from the whole point of developing the process in the first place.
- **Measure.** Test assumptions with experiments. Some of the most exciting moments of this research have been in performing an experiment only to find out that I was completely wrong about something.
- **Don't forget about books.** With the wealth of freely available research on the internet it can be easy to disregard books. But papers often dance over the core concepts and relationships in a particular area of research. Books represent a more curated set of knowledge that can help you establish a base of core principles from which you can project forward.





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Additionally, throughout my research I've discovered some tools that have been enormously helpful in building these small scale systems:

- **Parallel Jaw Pliers.** I think I owe my PhD to this tool. Manual assembly without this is fraught with the possibility of damaging parts by applying force in the wrong direction. It's also great for grabbing pins and bolt-heads without marring them.
- **Micro-vises** like the Proxxon 24260 are essential for holding and manipulating small parts.
- **Pin-vises** can be used not just for holding small cylindrical parts, but also work great for precisely holding and using reamers, drills, and small taps.
- **Micro Spatulas** are more precise than syringe tips for applying very small amounts of adhesive.
- **EZ-Hook** electrical connectors are good for grabbing small wires and traces but contribute a non-trivial amount of resistance to the circuit.
- A **micro-ruler** on a stick is great for calibrating microscope measurements.
- **Gelpak** on a glass plate makes a great fixturing solution for laser micro-machining as well as all kinds of precise soldering and assembly work.
- I ended up using the **Analog Discovery 2** a great deal for things ranging from debugging serial bus issues, closing the loop on microcode, and even frequency response measurements.
- **Masking tape and superglue** makes for quick and reliable machining fixturing.
- Choosing the right endmill for the material makes a huge difference. Check out **single-flute super-O mills** for "grabby" plastics like HDPE. **Diamond-cut mills** work great on composites. **Straight-flute endmills** leave a nice finish on both the top and bottom sides of plywood. Avoid chatter by using the shortest, stubbiest endmill that you can for the job.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Bibliography


[1] S. Tibbits, "From Automated to Autonomous Assembly," *Architectural Design*, vol. 87, no. 4, pp. 6–15, 2017.

[2] J. Paik, "Soft robot design methodology for 'push-button' manufacturing," *Nature Reviews Materials*, pp. 1–3, 2018.

[3] Z. Butler and A. Rizzi, "Distributed and Cellular Robots," in *Springer Hanbook of Robotics*, pp. 911–920, 2008.

[4] T. Hornyak, "Be afraid: DARPA unveils Terminator-like Atlas robot." https://www.cnet.com/news/be-afraid-darpa-unveils-terminator-like-atlas-robot/, jul.

[5] "PETMAN (Protection Ensemble Test Mannequin) Humanoid Military Robot." https://www.army-technology.com/projects/petman/.

[6] Massachusetts Institute of Technology, "MIT Backflipping Mini Cheetah." https://youtu.be/xNeZWP5Mx9s, 2019.

[7] H. W. Park, P. M. Wensing, and S. Kim, "High-speed bounding with the MIT Cheetah 2: Control design and experiments," *International Journal of Robotics Research*, vol. 36, no. 2, pp. 167–192, 2017.

[8] J. P. Whitney, P. S. Sreetharan, K. Y. Ma, and R. J. Wood, "Pop-up book MEMS," *Journal of Micromechanics and Microengineering*, vol. 21, no. 11, p. 115021, 2011.

[9] C. Semini, M. Baker, K. Laxman, V. Chandan, T. Maruthiram, R. Morgan, M. Frigerio, V. Barasuol, D. G. Caldwell, and G. Rey, "A Brief Overview of a Novel, Highly-Integrated Hydraulic Servo Actuator with Additive-Manufactured Titanium Body," *IROS Workshop 2016*, pp. 4–7, 2016.

[10] M. Wehner, R. L. Truby, D. J. Fitzgerald, B. Mosadegh, G. M. Whitesides, J. A. Lewis, and R. J. Wood, "An integrated design and fabrication strategy for entirely soft, autonomous robots," *Nature*, vol. 536, p. 451, aug 2016.

[11] D. Barry, "Chromosome and Kinetochore." https://www.wehi.edu.au/wehi-tv/chromosome-and-kinetochore, 2013.

[12] A. J. Roberts, T. Kon, P. J. Knight, K. Sutoh, and S. A. Burgess, "Functions and mechanics of dynein motor proteins," *Nature Reviews Molecular Cell Biology*, vol. 14, no. 11, pp. 713–726, 2013.






DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[13] J. Lin and D. Nicastro, "Asymmetric distribution and spatial switching of dynein activity generates ciliary motility," *Science*, vol. 360, no. 6387, 2018.

[14] DNA Learning Center, "Transcription and Tranlation: Translation (Advanced)." https://www.dnalc.org/resources/3d/16-translation-advanced.html, 2010.

[15] R. Milo and R. Phillips, "Cell Biology by the Numbers," 2016.

[16] Particle Sciences, "Protein Structure," *Technical Brief 2009 vol. 8*, vol. 8, pp. 1–2, 2009.

[17] BioNinja, "Amino Acids." https://ib.bioninja.com.au/standard-level/topic-2-molecular-biology/24-proteins/amino-acids.html.

[18] G. A. Popescu, T. Mahale, and N. A. Gershenfeld, "Digital materials for digital printing," *NIP & Digital Fabrication Conference*, pp. 1–4, 2006.

[19] J. D. Hiller and H. Lipson, "Fully Recyclable Multi-Material Printing," *Solid Freeform Fabrication Symposium (SFF'09)*, 2009.

[20] K. C. Cheung and N. Gershenfeld, "Reversibly Assembled Cellular Composite Materials," *Science*, vol. 341, no. September, pp. 1219–1221, 2013.

[21] B. Jenett, S. Calisch, D. Cellucci, N. Cramer, N. Gershenfeld, S. Swei, and K. C. Cheung, "Digital Morphing Wing: Active Wing Shaping Concept Using Composite Lattice-Based Cellular Structures," *Soft Robotics*, 2017.

[22] W. Langford, A. Ghassaei, and N. Gershenfeld, "Automated Assembly of Electronic Digital Materials," in *ASME MSEC*, (Blacksburg, VA), pp. 1–10, 2016.

[23] R. MacCurdy, a. McNicoll, and H. Lipson, "Bitblox: Printable digital materials for electromechanical machines," *The International Journal of Robotics Research*, jul 2014.

[24] C. D. Onal, M. T. Tolley, R. J. Wood, and D. Rus, "Origami-Inspired Printed Robots," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 5, pp. 2214–2221, 2015.

[25] A. Ion, J. Frohnhofen, L. Wall, R. Kovacs, M. Alistar, J. Lindsay, P. Lopes, H.-T. Chen, and P. Baudisch, "Metamaterial Mechanisms," pp. 529–539, 2016.

[26] N. Cramer, D. Cellucci, K. C. Cheung, S. Swei, and M. Teodorescu, "Design and Testing of FERVOR: FlexiblE and Reconfigurable Voxel-based Robot," *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2730–2735, 2017.

[27] D. M. Aukes, B. Goldberg, M. R. Cutkosky, and R. J. Wood, "An Analytic Framework for Developing Inherently-Manufacturable Pop-up Laminate Devices,"

[28] A. T. Baisch and R. J. Wood, "Pop-up assembly of a quadrupedal ambulatory MicroRobot," *IEEE International Conference on Intelligent Robots and Systems*, pp. 1518–1524, 2013.

[29] M. Schaffner, J. A. Faber, L. Pianegonda, P. A. Rühs, F. Coulter, and A. R. Studart, "3D printing of robotic soft actuators with programmable bioinspired architectures," *Nature Communications*, vol. 9, no. 1, 2018.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[30] R. F. Shepherd, F. Ilievski, W. Choi, S. A. Morin, A. A. Stokes, A. D. Mazzeo, X. Chen, M. Wang, and G. M. Whitesides, "Multigait soft robot," *Proceedings of the National Academy of Sciences*, 2011.

[31] T. Ranzani, S. Russo, N. W. Bartlett, M. Wehner, and R. J. Wood, "Increasing the Dimensionality of Soft Microstructures through Injection-Induced Self-Folding," *Advanced Materials*, vol. 1802739, p. 1802739, 2018.

[32] E. MacDonald and R. Wicker, "Multiprocess 3D printing for increasing component functionality," *Science*, vol. 353, no. 6307, 2016.

[33] J. Hiller, "Digital Materials: Voxel Design, Rapid Assembly, Structural Properties, and Design Methods," no. May, 2011.

[34] W. Langford, A. Ghassaei, N. Gershenfeld, and B. Jenett, "Hierarchical assembly of a self-replicating spacecraft," in *IEEE Aerospace Conference Proceedings2*, 2017.

[35] A. W. Burks, "Von Neumann's Self-Reproducing Automata," Tech. Rep. June, The University of Michigan, 1969.

[36] W. M. Stevens, *Self-Replication, Construction and Computation.* Doctor of philosophy, University of Kent, 2009.

[37] A. Ellery, "Are Self-Replicating Machines Feasible?," *Journal of Spacecraft and Rockets*, vol. 53, no. 2, pp. 317–327, 2016.

[38] V. Zykov, E. Mytilinaios, M. Desnoyer, and H. Lipson, "Evolved and Designed Self-Reproducing Modular Robotics," *IEEE Transactions on Robotics*, vol. 23, pp. 308–319, apr 2007.

[39] M. Yim, W.-M. Shen, B. Salemi, D. Rus, M. Moll, H. Lipson, and E. Klavins, "Modular Self-reconfigurable Robot Systems: Challenges and Opportunities for the Future," *IEEE Robotics & Automation Magazine*, vol. 14, no. 1, pp. 43–52, 2007.

[40] A. N. Knaian, K. C. Cheung, M. B. Lobovsky, A. J. Oines, P. Schmidt-Neilsen, and N. A. Gershenfeld, "The Milli-Motein: A self-folding chain of programmable matter with a one centimeter module pitch," *IEEE International Conference on Intelligent Robots and Systems*, pp. 1447–1453, 2012.

[41] C. H. Belke and J. Paik, "Mori: A Modular Origami Robot," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 5, pp. 2153–2164, 2017.

[42] K. Gilpin and D. Rus, "Modular Robot Systems From Self-Assembly to Self- Disassembly," *IEEE Robotics & Automation Magazine*, pp. 38–55, 2010.

[43] A. S. Howe, "Self-assembling modular robotic structures," *IEEE Robotics and Automation Magazine*, vol. 14, no. 4, pp. 26–33, 2007.

[44] T. Toth-Fejel, R. A. Freitas, and M. S. Moses, "Modeling Kinematic Cellular Automata Final Report," tech. rep., NASA NIAC, 2004.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[45] G. Chirikjian, "An architecture for self-replicating lunar factories," *NIAC Phase I*, 2004.

[46] M. S. Moses, H. Ma, K. C. Wolfe, and G. S. Chirikjian, "An architecture for universal construction via modular robotic components," *Robotics and Autonomous Systems*, vol. 62, no. 7, pp. 945–965, 2014.

[47] T. V. Karpinets, D. J. Greenwood, C. E. Sams, and J. T. Ammons, "RNA: Protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis," *BMC Biology*, vol. 4, pp. 1–10, 2006.

[48] Nanoscribe, "Https://www.nanoscribe.de/en/," tech. rep.

[49] A. M. Pickering and K. J. A. Davies, "Degradation of damaged proteins: The main function of the 20S proteasome," *Progress in Molecular Biology and Translational Science*, vol. 109, pp. 227–248, 2012.

[50] M. Sprinzl, "Correction of errors during protein synthesis," *Angewandte Chemie - International Edition*, vol. 48, no. 21, pp. 3738–3739, 2009.

[51] L. Eshun-Wilson, R. Zhang, D. Portran, M. V. Nachury, D. B. Toso, T. Löhr, M. Vendruscolo, M. Bonomi, J. S. Fraser, and E. Nogales, "Effects of $\alpha$-tubulin acetylation on microtubule structure and stability," *Proceedings of the National Academy of Sciences*, vol. 116, no. 21, pp. 10366–10371, 2019.

[52] J. Pelletier and S. Jun, "The Bacterial Nucleoid," vol. 1624, pp. 311–322, 2017.

[53] P. Mooney, T. Sulerud, J. F. Pelletier, M. R. Dilsaver, M. Tomschik, C. Geisler, and J. C. Gatlin, "Tau-based fluorescent protein fusions to visualize microtubules," *Cytoskeleton*, vol. 74, no. 6, pp. 221–232, 2017.

[54] M. C. Good, "Scaffold Proteins : Hubs for Controlling the Flow of Cellular Information," vol. 680, no. 2011, pp. 680–687, 2012.

[55] D. Rus and M. T. Tolley, "Design, fabrication and control of soft robots," *Nature*, vol. 521, no. 7553, pp. 467–475, 2015.

[56] D. Rus and M. T. Tolley, "Design, fabrication and control of origami robots," *Nature Reviews Materials*, pp. 1–12, 2018.

[57] J. Tocchio, "Exploded Views – Canon AE-1." https://www.casualphotophile.com/2015/03/26/exploded-views-canon-ae-1/, 2015.

[58] A. N. Knaian, *Electropermanent Magnetic Connectors and Actuators : Devices and Their Application in Programmable Matter*. Doctor of philosophy, Massachusetts Institute of Technology, 2010.

[59] Cirris, "Connector Life Cycles." https://www.cirris.com/learning-center/product-articles/other-products/227-connector-life-cycles.

[60] P. V. Dijk, "Critical Aspects of Electrical Connector Contacts," tech. rep.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[61] W. K. Langford, *Electronic Digital Materials*. Master of science, Massachusetts Institute of Technology, 2014.

[62] O. Tessmann, "Topological Interlocking Assemblies," *Ecaade 30*, vol. 2, no. Lynn 2009, pp. 201–210, 2011.

[63] A. G. Gillies and R. S. Fearing, "A micromolded connector for reconfigurable millirobots," *Journal of Micromechanics and Microengineering*, vol. 20, p. 105011, oct 2010.

[64] B. Jenett, N. Gershenfeld, and P. Guerrier, "Building Block-Based Assembly of Scalable Metallic Lattices," *Proceedings of the 13th Manufacturing Science and Engineering Conference*, pp. 1–6, 2018.

[65] B. Jenett, D. Cellucci, C. Gregg, and K. Cheung, "Meso-scale digital materials: modular, reconfigurable, lattice- based structures," *Proceedings of the 2016 Manufacturing Science and Engineering Conference*, pp. 1–11, 2016.

[66] G. Trinh, G. Copplestone, M. O'Connor, S. Hu, S. Nowak, K. Cheung, B. Jenett, and D. Cellucci, "Robotically assembled aerospace structures: Digital material assembly using a gantry-type assembler," *IEEE Aerospace Conference Proceedings*, pp. 1–7, 2017.

[67] A. Hsu, A. Wong-Foy, B. McCoy, C. Cowan, J. Marlow, B. Chavez, T. Kobayashi, D. Shockey, and R. Pelrine, "Application of micro-robots for building carbon fiber trusses," *2016 International Conference on Manipulation, Automation and Robotics at Small Scales, MARSS 2016*, pp. 1–6, 2016.

[68] DuPont, "DuPont ™ Pyralux ® LF Sheet Adhesive," tech. rep.

[69] 3M, "3M ™ XYZ / Isotropic Electrically Conductive Adhesive Transfer Tape 9709," tech. rep., 2011.

[70] J. Neubert and H. Lipson, "Soldercubes: a self-soldering self-reconfiguring modular robot system," *Autonomous Robots*, jul 2015.

[71] H. Lipson, "Rapid Assemblers for Voxel-Based VLSI Robotics," tech. rep., 2015.

[72] A. Miyachi America, "GENERAL PRINCIPLES Fundamentals of Small Parts Resistance Welding,"

[73] T. A. Schaedler, A. J. Jacobsen, A. Torrents, A. E. Sorensen, J. Lian, J. R. Greer, L. Valdevit, and W. B. Carter, "Ultralight metallic microlattices," *Science*, vol. 334, no. 6058, pp. 962–965, 2011.

[74] NTopology, "nTopology." https://ntopology.com/.

[75] Autodesk, "Autodesk Within." http://www.withinlab.com/.

[76] Renishaw, "Metal Additive Manufacturing (3D printing)." https://www.renishaw.com/en/metal-3d-printing–32084.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[77] N. B. Cramer, D. W. Cellucci, O. B. Formoso, C. E. Gregg, B. E. Jenett, J. H. Kim, M. Lendraitis, S. S. Swei, G. T. Trinh, K. V. Trinh, and K. C. Cheung, "Elastic shape morphing of ultralight structures by programmable assembly," *Smart Materials and Structures*, vol. 28, no. 5, 2019.

[78] M. Mirkhalaf, A. Sunesara, B. Ashrafi, and F. Barthelat, "Toughness by segmentation: Fabrication, testing and micromechanics of architectured ceramic panels for impact applications," *International Journal of Solids and Structures*, vol. 0, pp. 1–14, 2018.

[79] Y. Feng, T. Siegmund, E. Habtour, and J. Riddick, "Impact mechanics of topologically interlocked material assemblies," *International Journal of Impact Engineering*, vol. 75, no. July 2016, pp. 140–149, 2015.

[80] L. Djumas, A. Molotnikov, G. P. Simon, and Y. Estrin, "Enhanced Mechanical Performance of Bio-Inspired Hybrid Structures Utilising Topological Interlocking Geometry," *Scientific Reports*, vol. 6, no. January, pp. 1–10, 2016.

[81] X. Zheng, H. Lee, T. H. Weisgraber, M. Shusteff, J. DeOtte, E. B. Duoss, J. D. Kuntz, M. M. Biener, Q. Ge, J. a. Jackson, S. O. Kucheyev, N. X. Fang, and C. M. Spadaccini, "Ultralight, ultrastiff mechanical metamaterials.," *Science (New York, N.Y.)*, vol. 344, pp. 1373–7, jun 2014.

[82] K. Cheung, *Digital cellular solids: reconfigurable composite materials*. Doctor of philosophy, Massachusetts Institute of Technology, 2012.

[83] J. D. Hiller, J. Miller, and H. Lipson, "Microbricks for Three-Dimensional Reconfigurable Modular Microsystems," *Journal of Microelectromechanical Systems*, vol. 20, pp. 1089–1097, oct 2011.

[84] J. Hiller and H. Lipson, "Methods of Parallel Voxel Manipulation for 3D Digital Printing," in *Proceedings of the 18th solid freeform fabrication symposium.*, p. 12, 2007.

[85] J. Ward, *Additive assembly of digital materials*. Master of science in media arts and sciences, Massachusetts Institute of Technology, 2010.

[86] J. Hiller and H. Lipson, "Design and analysis of digital materials for physical 3D voxel printing," *Rapid Prototyping Journal*, vol. 15, no. 2, pp. 137–149, 2009.

[87] R. J. Wood, "The challenge of manufacturing between macro and micro," *American Scientist*, vol. 102, no. 2, pp. 124–131, 2014.

[88] C. Conde and A. Cáceres, "Microtubule assembly, organization and dynamics in axons and dendrites.," *Nature reviews. Neuroscience*, vol. 10, pp. 319–32, may 2009.

[89] S. P. Magleby, "Library of Compliant Mechanisms," in *Handbook of Compliant Mechanisms*, pp. 147–153, 2013.

[90] N. Meisel, A. Gaynor, C. Williams, and J. Guest, "Multiple-material topology optimization of compliant mechanisms created via polyjet 3D printing," in *24th International SFF Symposium - An Additive Manufacturing Conference, SFF 2013*, 2013.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[91] F. B, "Caress of the Gaze: A Gaze Actuated 3D Printed Body Architecture," *Proceedings of Posthuman Frontiers: Acadia 2016*, pp. 352–361, 2016.

[92] J. Gafford, Y. Ding, A. Harris, A. Moser, and C. Walsh, "Shape Deposition Manufacturing of a Soft , Atraumatic , Deployable Surgical Grasper," no. 1, pp. 4–5.

[93] R. J. Wood, S. Avadhanula, R. Sahai, E. Steltz, and R. S. Fearing, "Microrobot Design Using Fiber Reinforced Composites," *Journal of Mechanical Design*, vol. 130, no. 5, p. 052304, 2008.

[94] H. McClintock, F. Z. Temel, N. Doshi, J.-s. Koh, and R. J. Wood, "The milliDelta: A high-bandwidth, high-precision, millimeter-scale Delta robot," *Science Robotics*, vol. 3, no. 14, p. eaar3018, 2018.

[95] A. M. Hoover and R. S. Fearing, "Analysis of off-axis performance of compliant mechanisms with applications to mobile millirobot design," *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009*, pp. 2770–2776, 2009.

[96] B. P. Trease, Y.-M. Moon, and S. Kota, "Design of Large-Displacement Compliant Joints," *Journal of Mechanical Design*, vol. 127, no. 4, p. 788, 2005.

[97] R. M. Panas and J. B. Hopkins, "Eliminating Underconstraint in Double Parallelogram Flexure Mechanisms," *Journal of Mechanical Design*, vol. 137, no. 9, p. 092301, 2015.

[98] R. Malka, A. L. Desbiens, Y. Chen, and R. J. Wood, "Principles of microscale flexure hinge design for enhanced endurance," *IEEE International Conference on Intelligent Robots and Systems*, no. Iros, pp. 2879–2885, 2014.

[99] A. E. Guerinot, S. P. Magleby, L. L. Howell, and R. H. Todd, "Compliant Joint Design Principles for High Compressive Load Situations," *Journal of Mechanical Design*, vol. 127, no. 4, p. 774, 2005.

[100] R. Wood, E. Steltz, and R. Fearing, "Optimal energy density piezoelectric bending actuators," *Sensors and Actuators A: Physical*, vol. 119, no. October 2004, pp. 476–488, 2005.

[101] A. P. Gerratt, B. Balakrisnan, I. Penskiy, and S. Bergbreiter, "Dielectric elastomer actuators fabricated using a micro-molding process," *Smart Materials and Structures*, vol. 23, p. 055004, may 2014.

[102] N. J. Kohut, A. M. Hoover, K. Y. Ma, S. S. Baek, and R. S. Fearing, "MEDIC: A legged millirobot utilizing novel obstacle traversal," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 802–808, 2011.

[103] B. Goldberg, M. Karpelson, O. Ozcan, and R. J. Wood, "Planar fabrication of a mesoscale voice coil actuator," *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6319–6325, 2014.

[104] R. Yeh, S. Hollar, and K. S. J. Pister, "Single mask, large force, and large displacement electrostatic linear inchworm motors," *Journal of Microelectromechanical Systems*, vol. 11, no. 4, pp. 330–336, 2002.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[105] Linear Technologies, "LTSpice." https://www.analog.com/en/design-center/design-tools-and-calculators/ltspice-simulator.html.

[106] K&J Magnetics Inc., "Temperature and Neodymium Magnets." https://www.kjmagnetics.com/blog.asp?p=temperature-and-neodymium-magnets, 2016.

[107] D. C. Hanselman, *Brushless permanent magnet motor design*. 2006.

[108] W. S. N. Trimmer, "Microrobots and Micromechanical Systems," *Sensors and Actuators*, vol. 19, no. 3, pp. 267–287, 1989.

[109] Ingrid West Machinery Ltd., "Fully Automatic Hot Air Coil Winding Machine." http://coilwindingmachines.eu/automatic_winding_machines/hot_air_coil_winding_machine.html.

[110] C.-S. Liu and P. D. Lin, "Miniaturized auto-focusing VCM actuator with zero holding current," *Optics Express*, vol. 17, no. 12, p. 9754, 2009.

[111] M. Salerno, A. Firouzeh, and J. Paik, "A Low Profile Electromagnetic Actuator Design and Model for an Origami Parallel Platform," *Journal of Mechanisms and Robotics*, vol. 9, no. 4, p. 041005, 2017.

[112] B. H. Shin and S. Y. Lee, "Micro mobile robots using electromagnetic oscillatory actuator," *Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics*, pp. 575–580, 2012.

[113] H2W Technologies, "H2W NCC10-15-027-1RC." https://www.h2wtech.com/product/voice-coil-actuators/NCC10-15-027-1RC.

[114] Moticont, "LVCM-010-013-01." http://moticont.com/voice-coil-motor.htm.

[115] P. A. York, S. Member, R. J. Wood, and S. Member, "A geometrically-amplified in-plane piezoelectric actuator for mesoscale robotic systems," pp. 1263–1268, 2017.

[116] E. W. Schaler, L. Jiang, C. Lee, and R. S. Fearing, "Bidirectional, Thin-Film Repulsive-/Attractive-Force Electrostatic Actuators for a Crawling Milli-Robot," *MARSS 2018 (to appear)*, pp. 1–8, 2018.

[117] J. E. Huber, N. a. Fleck, and M. F. Ashby, "The selection of mechanical actuators based on performance indices," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 453, no. 1965, pp. 2185–2205, 1997.

[118] M. S. Tu, "Cardiac-like behavior of an insect flight muscle," *Journal of Experimental Biology*, vol. 207, no. 14, pp. 2455–2464, 2004.

[119] H. Iwamoto, "Structure, function and evolution of insect flight muscle," *Biophysics*, vol. 7, pp. 21–28, 2011.

[120] R. Merry, R. Van De Molengraft, and M. Steinbuch, "Modeling of a walking piezo actuator," *Sensors and Actuators, A: Physical*, 2010.

[121] Physikinstrumente, "Piezo Walk Piezo Motors." https://www.physikinstrumente.com/, 2018.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[122] K. Spanner and B. Koc, "Piezoelectric Motors, an Overview," *Actuators*, vol. 5, no. 1, p. 6, 2016.

[123] J. B. Reese, S. L. Bober, M. B. Daly, C. Program, F. Chase, B. Sciences, P. Care, F. Chase, S. Behavior, and B. Neoplasms, "HHS Public Access," vol. 123, no. 24, pp. 4757–4763, 2018.

[124] I. Penskiy and S. Bergbreiter, "Optimized electrostatic inchworm motors using a flexible driving arm," *Journal of Micromechanics and Microengineering*, vol. 23, no. 1, 2013.

[125] Z. Bi, Y. Guan, S. Chen, H. Zhu, and H. Zhang, "A miniature biped wall-climbing robot for inspection of magnetic metal surfaces," *2012 IEEE International Conference on Robotics and Biomimetics, ROBIO 2012 - Conference Digest*, pp. 324–329, 2012.

[126] X. Tang, D. Zhang, Z. Li, and J. Chen, "An omni-directional wall-climbing microrobot with magnetic wheels directly integrated with electromagnetic micromotors," *International Journal of Advanced Robotic Systems*, vol. 9, 2012.

[127] M. Eich and T. V??gele, "Design and control of a lightweight magnetic climbing robot for vessel inspection," *2011 19th Mediterranean Conference on Control and Automation, MED 2011*, no. June, pp. 1200–1205, 2011.

[128] G. Lee, G. Wu, J. Kim, and T. Seo, "High-payload climbing and transitioning by compliant locomotion with magnetic adhesion," *Robotics and Autonomous Systems*, vol. 60, no. 10, pp. 1308–1316, 2012.

[129] J. Pinskier, B. Shirinzadeh, and L. Clark, "Design & optimization of a compact, large amplification XY flexure-mechanism," *International Conference on Manipulation, Automation and Robotics at Small Scales, MARSS 2017 - Proceedings*, 2017.

[130] M. N. M. Zubir, B. Shirinzadeh, Y. Tian, M. Nashrul, M. N. M. Zubir, B. Shirinzadeh, and Y. Tian, "Development of a novel flexure-based microgripper for high precision micro-object manipulation," *Sensors and Actuators A: Physical*, vol. 150, pp. 257–266, mar 2009.

[131] A. M. Hoover, S. Avadhanula, R. E. Groff, and R. S. Fearing, "A rapidly prototyped 2-axis positioning stage for microassembly using large displacement compliant mechanisms," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2006, pp. 289–295, 2006.

[132] R. S. Fearing, R. Sahai, and A. Hoover, "Rapidly Prototyping Millirobots using Toolkits and Microassembly," tech. rep.

[133] R. L. Henry, "Project Tinkertoy," *Journal of the Franklin Institute*, vol. 256, no. 6, pp. 598–599, 2003.

[134] R. K. Jurgen, "Whatever happened to project tinkertoy?," *IEEE Spectrum*, vol. 24, no. 5, pp. 20–21, 1987.

[135] A. Russo, B. Y. Ahn, J. J. Adams, E. B. Duoss, J. T. Bernhard, and J. A. Lewis, "Pen-on-Paper Flexible Electronics," *Advanced Materials*, vol. 23, no. 30, pp. 3426–3430, 2011.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[136] D. Zhao, T. Liu, J. G. Park, M. Zhang, J.-m. M. Chen, B. Wang, J. Gyu, M. Zhang, J.-m. M. Chen, and B. Wang, "Conductivity enhancement of aerosol-jet printed electronics by using silver nanoparticles ink with carbon nanotubes," *Microelectronic Engineering*, vol. 96, pp. 71–75, 2012.

[137] K. Sun, T. S. Wei, B. Y. Ahn, J. Y. Seo, S. J. Dillon, and J. a. Lewis, "3D printing of interdigitated Li-ion microbattery architectures," *Advanced Materials*, vol. 25, no. 33, pp. 4539–4543, 2013.

[138] E. Malone, K. Rasa, D. Cohen, T. Isaacson, H. Lashley, and H. Lipson, "Freeform fabrication of zinc-air batteries and electromechanical assemblies," *Rapid Prototyping Journal*, vol. 10, no. 1, pp. 58–69, 2004.

[139] J. J. Adams, E. B. Duoss, T. F. Malkowski, M. J. Motala, B. Y. Ahn, R. G. Nuzzo, J. T. Bernhard, and J. a. Lewis, "Conformal printing of electrically small antennas on three-dimensional surfaces.," *Advanced materials (Deerfield Beach, Fla.)*, vol. 23, pp. 1335–40, mar 2011.

[140] S. Fuller, E. Wilhelm, and J. Jacobson, "Ink-jet printed nanoparticle microelectromechanical systems," *Journal of Microelectromechanical Systems*, vol. 11, no. 1, pp. 54–60, 2002.

[141] G. T. Trinh, D. Cellucci, W. Langford, S. Im, A. G. Luna, and K. C. Cheung, "Modular Rapidly Manufactured Small Satellite ( MRMSS )," no. January, pp. 1–11, 2015.

[142] Voxel8, "Voxel8." http://www.voxel8.co/.

[143] Microchip, "ATBTLC1000." https://www.microchip.com/wwwproducts/en/ATBTLC1000.

[144] V. R. Marinov, O. Swenson, Y. Atanasov, and N. Schneck, "Laser-assisted ultrathin bare die packaging: a route to a new class of microelectronic devices," *Spie*, vol. 8608, pp. 86080L–86080L–19, 2013.

[145] K. Gilpin, A. Knaian, and D. Rus, "Robot pebbles: One centimeter modules for programmable matter through self-disassembly," *2010 IEEE International Conference on Robotics and Automation*, pp. 2485–2492, may 2010.

[146] M. Haurylau, N. A. Nelson, D. H. Albonesi, E. G. Friedman, P. M. Fauchet, G. Chen, H. Chen, and J. Zhang, "On-chip optical interconnect roadmap: Challenges and critical directions," *IEEE Journal on Selected Topics in Quantum Electronics*, vol. 12, no. 6, pp. 1699–1704, 2006.

[147] A. P. Ghassaei, *Rapid Design and Simulation of Functional Digital Materials by in partial fulfillment of the requirements for the degree of at the by.* Master of science, Massachusetts Institute of Technology, 2016.

[148] M. Carney and B. Jenett, "Relative Robots: Scaling AUtomated Assembly of Discrete Cellular Lattices," *Proceedings of the 2016 Manufacturing Science and Engineering Conference*, 2016.

[149] M. Sitti, *Mobile Micro-Robotics*. Cambridge: MIT Press, 2017.

[150] J. B. Gafford, S. Member, S. B. Kesner, R. J. Wood, and C. J. Walsh, "Force-Sensing Surgical Grasper Enabled by Pop-Up Book MEMS," pp. 2552–2558, 2013.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[151] J. Meijer, K. Du, a. Gillner, D. Hoffmann, V. Kovalenko, T. Masuzawa, a. Ostendorf, R. Poprawe, and W. Schulz, "Laser Machining by short and ultrashort pulses, state of the art and new opportunities in the age of the photons," *CIRP Annals - Manufacturing Technology*, vol. 51, no. 2, pp. 531–550, 2002.

[152] L. Harvey, B. Arnold, Z. S Lawrence, l. M. Pau, B. David, and D. James, *Molecular Cell Biology. 4th edition.* 2000.

[153] W. R. Frontera and J. Ochala, "Skeletal Muscle: A Brief Review of Structure and Function," *Behavior Genetics*, vol. 45, no. 2, pp. 183–195, 2015.

[154] L. Cao, A. T. Dolovich, A. L. Schwab, J. L. Herder, and W. C. Zhang, "Toward a Unified Design Approach for Both Compliant Mechanisms and Rigid-Body Mechanisms: Module Optimization," *Journal of Mechanical Design*, vol. 137, no. 12, p. 122301, 2015.

[155] W. C. O. Mara, R. B. Herring, and L. P. Hunt, "Handbook of Compliant Mechanisms,"

[156] M. Jin, X. Zhang, B. Zhu, and X. Z. Jin, Mohui, "A numerical method for static analysis of pseudo-rigid-body model of compliant mechanisms.," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 228, no. 228, pp. 3170–3177, 2014.

[157] R. Avilés, M. Goizalde Ajuria, M. Victoria Hormaza, A. Hernández, F. Elements, I. N. Analysis, and A. N. D. Design, "A procedure based on finite elements for the solution of nonlinear problems in the kinematic analysis of mechanisms," *Finite Elements in Analysis and Design*, vol. 22, no. 4, pp. 305–327, 1996.

[158] O. Sigmund, "On the Design of Compliant Mechanisms Using Topology Optimization," *Mechanics Based Design of Structures and Machines*, vol. 25, no. 4, pp. 493–524, 1997.

[159] T. Lu, C. Ortega, J. Kulick, G. H. Bernstein, S. Ardisson, R. Engelhardt, P. Way, and P. O. Box, "Rapid SoC Prototyping Utilizing Quilt Packaging Technology For Modular Functional IC Partitioning," *2016 International Symposium on Rapid System Prototyping (RSP)*, pp. 79–85, 2016.

[160] J. H. Lau, *Fan-Out Wafer-Level Packaging.* Springer Nature Singapore, 2018.

[161] ZGlue, "zGlue." https://www.zglue.com/.

[162] A. Olofsson, T. Nordström, and Z. Ul-Abdin, "Kickstarting high-performance energy-efficient manycore architectures with Epiphany," *Conference Record - Asilomar Conference on Signals, Systems and Computers*, vol. 2015-April, no. May, pp. 1719–1726, 2015.

[163] A. Olofsson, "Epiphany-V: A 1024 processor 64-bit RISC System-On-Chip," pp. 1–15, 2016.

[164] C. E. Leiserson, Z. S. Abuhamdeh, D. C. Douglas, C. R. Feynman, M. N. Ganmukhi, J. V. Hill, W. D. Hillis, B. C. Kuszmaul, M. A. Pierre, D. S. Wells, M. C. Wong-Chan, S. W. Yang, and R. Zak, "The network architecture of the connection machine CM-5," *Journal of Parallel and Distributed Computing*, vol. 33, no. 2, pp. 145–158, 1996.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[165] J. R. Heath, P. J. Kuekes, G. S. Snider, and R. S. Williams, "A Defect-Tolerant Computer Architecture : Opportunities for Nanotechnology," vol. 280, no. June, 1998.

[166] C. E. Shannon, "A symbolic analysis of relay and switching circuits," 1938.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Laser Direct-Write Fabrication of MEMS

by

Prashant Patil

S.M., Massachusetts Institute of Technology (2013)
M.Tech., Engineering Physics & Nanoscience
Indian Institute of Technology, Bombay (2011)
B.Tech., Engineering Physics
Indian Institute of Technology, Bombay (2011)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2019

© Massachusetts Institute of Technology 2019. All rights reserved.

Author .....
Program in Media Arts and Sciences
School of Architecture and Planning
August 9, 2019

Certified by .....
Neil Gershenfeld
Director, Center for Bits and Atoms
Thesis Supervisor

Accepted by .....
Tod Machover
Academic Head, Program in Media Arts and Sciences



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Laser Direct-Write Fabrication of MEMS**

by

Prashant Patil

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
on August 9, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

## Abstract

Micro-electromechanical systems (MEMS) have many applications in healthcare, consumer electronics, and automobile industry. Unfortunately, the development of novel MEMS is significantly hindered by the limitations of the state-of-the-art MEMS microfabrication processes such as high cost of equipment ownership, long development time, and limited choice of fabrication material selection and integration. Recent developments in alternate MEMS fabrication processes such as PCB-MEMS, laminate MEMS, pop-up book MEMS, and soft-MEMS have reduced fabrication cost, increased material choice, and facilitated material integration. However, MEMS fabricated using these methods have a large lateral feature size, and low aspect ratio as compared to MEMS produced utilizing conventional deep reactive ion etching (DRIE) microfabrication process. Moreover, fabricating MEMS with six degrees of freedom (DOF) free-standing microstructures using these processes is challenging. Finally, the choice of fabrication material is relatively limited, and each material requires a separate manufacturing process.

This thesis presents a novel MEMS fabrication process called multi-lamina assembly of laser micromachined laminates (MALL), which can fabricate MEMS comparable to DRIE, enable creating free-standing microstructures with six degrees of freedom, and further expand the choice of fabrication material. Moreover, the proposed approach offers a single microfabrication method to process a wide range of materials. A novel microfabrication process called laser-assisted material phase-change and expulsion (LAMPE) micromachining is developed. Using this process, the fabrication of high aspect ratio structures with lateral features as small as $10\mu m$, and aspect ratio as large as 10:1 is demonstrated in metals, silicon, and diamond. Previously, such high aspect ratio and small lateral feature structures could be fabricated in silicon alone using the deep reactive ion etching process.

The LAMPE micromachining process is used to manufacture individual layers of a MEMS. Subsequently, the micromachined laminates are stack assembled and bonded to construct MEMS devices. Using the MALL process, fabrication of six degrees of freedom free-standing structures as thin as $10\mu m$ is demonstrated. In addition, the gap between the free-standing structure and the substrate can be as small as $12.5\mu m$. The utility of the MALL process is demonstrated by fabricating three MEMS. First, an electrostatic comb-drive actuator is fabricated using copper as the structural material. The distance





DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

between the comb-drive fingers is $10\mu m$, and the thickness of the fingers is $100\mu m$. This is the first demonstration of using metal to fabricate comb-drive structure with such small lateral feature and high aspect ratio. Second, a MEM relay for high-current switching application is demonstrated. The current carrying capacity of the MEM relay is higher than $100mA$. Finally, the development of high-aspect-ratio diamond rotors for enhancing the resolution of magic-angle spinning nuclear magnetic resonance spectroscopy (MAS-NMR) is presented. This is the first demonstration of micromachining such ultra-deep (5 mm), and ultra-high aspect ratio (10:1) holes in the diamond.

The MALL process can manufacture MEMS comparable to the conventional DRIE microfabrication process. Moreover, the manufacturing cost per device in the MALL is less than DRIE. However, DRIE offers a high part production rate than MALL. The part production rate in MALL can be matched with DRIE using multiple laser sources. For matching the part production rate, the investment required to purchase a laser micromachining tool with multiple lasers is comparable to the cost of a DRIE tool. Thus, an equal investment in MALL and DRIE results in an equal part production rate. The MALL process significantly reduces the time required for material integration, process development, and design iteration. As a result, the MEMS device development time is reduced from many months (in DRIE) to a day. The MALL process empowers rapid testing of new MEMS concepts and theory. Moreover, MALL can be used to fabricate one-of-a-kind MEMS devices and used for low-volume production, where initial high investment can not be justified.

The MALL process enables greater material selection and integration, rapid development, and integrated packing, thereby empowering a new paradigm in MEMS design, functionality, and application. The tools and material cost of MALL fabrication can be as low as $\$25,000$, which is affordable to a wider scientific community. The low capital investment and use of low-cost materials enable MEMS fabrication for masses and can expedite the development of novel MEMS.

Thesis Supervisor: Neil Gershenfeld
Title: Director, Center for Bits and Atoms


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Laser Direct-Write Fabrication of MEMS**

by

Prashant Patil

This dissertation has been reviewed and approved by the following committee members:

Carol Livermore . . . . . .

Associate Professor of Mechanical Engineering

Mechanical and Industrial Engineering

Northeastern University

Jeffrey Lang . . . . . .

Professor of Electrical Engineering

Electrical Engineering and Computer Science

Massachusetts Institute of Technology



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## Acknowledgments

I would like to express my deepest gratitude to my advisor, Neil Gershenfeld. Without your guidance and encouragement, this work would not have been possible. Thank you for putting your time and energy in mentoring me, providing space and resources, and fostering my creativity. The Truth is Empowering. Neil, thank you for providing honest constructive criticism and advises. For shaking my self-satisfying assumptions, raising my bar for achievement, and from a place of care, pushing me beyond my comfort zone. You taught me to think big, ask difficult questions, tackle hard problems, and push the envelope of what is possible. Thank you for believing in me.

I would also like to extend my deepest gratitude to my thesis committee - Carol Livermore and Jeffrey Lang - for their unwavering guidance. I am extremely grateful to be mentored by you. Thank you for taking the time to periodically meet with me, helping solve my problems, providing invaluable insights, and sharing your extensive knowledge. The completion of my dissertation would not have been possible without your guidance.

I also wish to thank Joe Jacobson, Shuguang Zhang, Ed Boyden, and Robert Griffin for their suggestions and collaborations.

CBA is a special place. It is a perfect fit for many misfits like myself. Neil, thank you for creating such a wonderful place and making me part of it. I will certainly miss this place.

Special thanks to Nadya for being a dear friend over many years. Meeting you at the MIT-India workshop in India was certainly the most life-changing moment of my life, and has led to this fascinating journey. My early days in the US were difficult, but you made them much easier.

I cannot leave CBA without mentioning Joe, who has been there from the beginning and done many behind-the-scene work which has made this thesis possible. Joe, thank you for everything, for many life advises, creative nicknames, and many interesting conversations. I will miss it. I would also like to extend my sincere thanks to CBA staff James, Kara, Tom, John, and Ryan. Thank you for operating everything smoothly, creating a sense of community, and sharing many lunches together. Thanks should also go to Jamie. Thank you Jamie, for being a dear friend, and for many proofreading sessions (and for secretly correcting my English).

I am extremely grateful to my cohort-mates Will, Sam, and Noah. Thank you for many brainstorming sessions, encouragement, and training sessions. Thank you for being generous with your time and entertaining my never-ending CAD/CAM/FAB questions. Many thanks to Ben for entertaining conversations, and letting me be Yoko's work dad (read fun dad).

I'd like to thank my dear friend, Filippos. Thank you for being there during my high and low, and for introducing me to Greek culture. I will miss our entertaining conversations about research, academics, and Greece. I'd like to thank Patritsia for helping me with my presentation and documents. Thank you for all the encouragement, appreciation, and friendship. I'd like thank Charles for building the lab together, many conversations, road trips, and friendship over the year. I'd like to thank Sarah Hovsepian, you were the first friend made in the US. Thanks for friendship over the year.

Thank you, Sherry and Neil, for introducing me to the amazing world for fablab.

I wish to thank CBA family - Amira, Erik, Jake, Sara, James P, Eric W., Matt C., Matt K., Soma, Amanda, Grace, Kenny, Rehmi, Kimin Jun, Manu Prakash, and Ara. I



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

had a great pleasure working with you, learning from you, and hanging out with you. Many thanks extended family at molecular machine - Pranam, Lisa, and Rui. I also wish to thank Dan and Salima for collaborating on the diamond rotor project. Thanks should also go to Bianca and Sunny for helping with building the laser micromachining system.

I would like to extend my sincere thanks to Amna, Joseph, Anirudh, and Kshitij. You have made my stay at MIT extra special. I will forever cherish those MIT days (and night outs). Many thanks to the design innovation workshop team, Dolly, Paul Anand, Anna, and Misha. I'd like to thank Rajesh for making me part of his workshop series, for many years of friendship, and mentorship.

Many thanks to Pragun, Dhairya, Udayan, Shrishti, Harshit, Harpreet, Santosh, Pranav, Neil, Tanuj, Vishal, Jitendra and Deepak. Many thanks to JNV, Betul friends Apurve, Hemant, Lalchand, Dinesh, Kumudini, Kepin, Surendra, Shivendra, Kamlesh, and all my classmates. Special thanks to folks at MIT Yoga club.

I am also grateful to my previous mentors Prof. Subhabhrata Dhar, Prof. S.S. Major, and Dr. So-Hye Cho. And many thanks to my previous lab mates at IIT Bombay - Jitendra, and Hari.

Special thanks to little furry friends Rumba, Eddy, Yoko, Lola, Penelope for bringing joy in my life.

I'd like to thank my high-school teacher Shamim Turki ma'am for admitting me to Jawahar Navodaya Vidyalaya, Betul. There is an old saying, "If you change the beginning of the story, you change the whole story". Ma'am, you have changed my story.

Thank you Nupur, for our special friendship, for encouraging me to be my best, and always supporting me. Thank you for being with me in this journey; your presence made every part of it wonderful.

Finally, I'd like to thank my parents - Mom and Dad, you have made so many sacrifices for me. Your unconditional love and support has made me who I am today. Many thanks to my brother Rinku, bhaiya you have been my first science teacher, thank you for infecting me with your tinkering habits. My sister Pranjali for pampering me when I was a baby, and my sister Pranoti for many life lessons.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Contents

**Appendices** **9**

**1 Introduction** **16**

1.1 Motivation . . . . . 16

1.2 Background . . . . . 18

1.2.1 PCB-MEMS . . . . . 18

1.2.2 Laminate MEMS . . . . . 18

1.2.3 Pop-up Book MEMS . . . . . 20

1.2.4 Limitations of Laser Micromachining . . . . . 20

1.3 MALL . . . . . 22

1.3.1 LAMPE Micromachining . . . . . 22

1.3.2 Electro-deburring . . . . . 24

1.3.3 Multi-Lamina Assembly . . . . . 24

1.4 Comparison with Prior Work . . . . . 25

1.4.1 High Aspect Ratio & Small Lateral Feature . . . . . 26

1.4.2 Freestanding Structures with Six Degrees of Freedom . . . . . 27

**2 Theory of Pulsed Laser Ablation** **30**

2.1 Pulsed Laser Ablation . . . . . 30

2.2 Laser-Micromachining System . . . . . 32

2.2.1 Laser Source . . . . . 33



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

2.2.2 Beam Delivery & Focusing Optics . . . . . 36
2.2.3 Motion System . . . . . 39
2.3 Summary . . . . . 40

**3 Laser-assisted Material Phase-change & Expulsion 41**
3.1 Introduction . . . . . 41
3.2 LAMPE . . . . . 43
3.2.1 Difference between PLA and LAMPE Micromachining . . . . . 45
3.3 Feature Size in LAMPE Micromachining . . . . . 46
3.3.1 Experiment & Results . . . . . 47
3.3.2 Analytical Model . . . . . 48
3.3.3 Numerical Simulation . . . . . 52
3.4 LAMPE High-Aspect-Ratio Micromachining . . . . . 55
3.5 Electro-deburring of LAMPE Micromachined Parts . . . . . 58
3.6 LAMPE Micromachining of Silicon . . . . . 60
3.6.1 Introduction . . . . . 60
3.6.2 Numerical Simulation . . . . . 65
3.6.3 Results & Discussion . . . . . 67
3.7 Conclusions . . . . . 70

**4 MALL MEMS Fabrication 72**
4.1 Multi-Lamina Assembly . . . . . 73
4.2 Comb-Drive Actuator . . . . . 75
4.2.1 Design . . . . . 75
4.2.2 Fabrication . . . . . 78
4.2.3 Results & Discussion . . . . . 80
4.2.4 Time Required for First Prototype . . . . . 83
4.3 Micro-Electromechanical Relay . . . . . 86
4.3.1 Design . . . . . 86


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

4.3.2 Fabrication . . . 91
4.3.3 Results & Discussion . . . 94
4.4 Diamond Rotor . . . 99
4.4.1 Introduction . . . 99
4.4.2 Diamond Rotor . . . 100
4.4.3 Laser Micromachining of Diamond . . . 101
4.4.4 Theoretical Explanation . . . 101
4.5 Conclusions . . . 104

**5 Discussion & Conclusion** **106**

5.1 Advantages of MALL . . . 106
5.1.1 Greater Choice of Fabrication Materials . . . 106
5.1.2 Fabrication with Incompatible Materials . . . 107
5.1.3 Low-Cost Fabrication . . . 107
5.1.4 Rapid Development . . . 110
5.1.5 Integrated Packaging . . . 112
5.2 Limitations of MALL . . . 112
5.2.1 Inadequate for Fabricating 2.5D Structures . . . 112
5.2.2 Limit on Minimum Thickness of Microstructures . . . 113
5.2.3 Limit on the Gap Between Laminates . . . 116
5.2.4 Poor Alignment Accuracy . . . 117
5.2.5 Heat Affected Zone . . . 117
5.3 Volume Manufacturing . . . 117
5.4 MEMS for Masses . . . 119
5.5 Thesis Contributions . . . 122
5.6 Conclusion . . . 123
5.7 Recommendations for Future Work . . . 124


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**A** **126**

A.1 Manufacturing Cost Estimates for DRIE and MALL . . . . . 126

A.1.1 Cost of Manufacturing Using DRIE . . . . . 127

A.1.2 Cost of Manufacturing Using MALL . . . . . 128

**B** **131**

B.1 Equipment Required for MALL Fabrication . . . . . 131

B.2 MEMS Design . . . . . 133

B.3 Fabrication . . . . . 133

B.3.1 Laser Micromachining . . . . . 133

B.3.2 Electro-deburring . . . . . 134

B.3.3 Lamination Process . . . . . 135

B.4 MEMS Device Testing . . . . . 136

B.5 Things That did Not work . . . . . 136



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# List of Figures


1-1 Literature review of alternate MEMS fabrication methods. . . . . 19

1-2 Literature review of laser micromachining . . . . . 21

1-3 Schematic diagram of MALL fabrication process. . . . . 23

1-4 Schematic diagram of multi-lamina assembly. . . . . 25

1-5 Schematic diagram of surface micromachining process. . . . . 27

2-1 Schematic diagram of pulsed laser ablation (PLA) process. . . . . 31

2-2 The plot of the average power , $P_{avg}$ with respect to the pulse frequency . . . 34

2-3 The plot of pulse energy, $E_p$ with respect to the pulse frequency. . . . . 34

2-4 The plot of peak power (left) and pulse intensity (right )with respect to the pulse frequency. . . . . 35

2-5 Intensity profile of the focused laser beam. . . . . 38

3-1 Schematic diagram of the LAMPE micromachining process. . . . . 44

3-2 Comparison of lateral features produced in PLA and LAMPE micromachining 46

3-3 Minimum feature size achievable using LAMPE micromachining. . . . . 48

3-4 Electro-deburring of LAMPE micromachined crater. . . . . 49

3-5 Pulse exposure with varying pitch distance to confirm material melting & expulsion. . . . . 53

3-6 Simulation of LAMPE micromachining of copper . . . . . 56

3-7 High-aspect-ratio microfabrication using LAMPE micromachining . . . . . 57

3-8 SEM image showing resolidified melt ejects. . . . . 58






DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

3-9 Polarization curve of copper in phosphoric acid. . . . . 59
3-10 Electro-deburring of LAMPE micromachined parts. . . . . 61
3-11 High aspect ratio interdigitated fingers fabricated using LAMPE micromachining. . . . . 62
3-12 Pulsed laser ablation (PLA) of silicon. . . . . 63
3-13 The cross-section image showing laser-assisted oxidation of silicon. . . . . 65
3-14 The optical microscope & energy dispersive x-ray spectroscopy (EDS) image confirming oxidation of silicon . . . . . 66
3-15 Simulation of LAMPE micromachining of silicon. . . . . 68
3-16 Example of high-aspect-ratio microstructures fabricated using LAMPE micromachining of silicon. . . . . 69

4-1 Alignment accuracy of multi-lamina assembly. . . . . 74
4-2 Schematic diagram of comb-drive actuator. . . . . 76
4-3 Fabrication of comb-drive device layer using LAMPE micromachining of copper. 78
4-4 Electro-deburring of LAMPE micromachined comb-drive structure. . . . . 79
4-5 Layers of the comb-drive actuator fabricated using LAMPE micromachining . 80
4-6 Assembled and bonded layers of the comb-drive actuator . . . . . 81
4-7 SEM image of the fabricated comb-drive actuator. . . . . 81
4-8 Features of the fabricated comb-drive actuator. . . . . 82
4-9 Displacement measurement of comb-drive actuator fingers. . . . . 83
4-10 Schematic diagram of the micro-electromechanical (MEM) relay. . . . . 87
4-11 The bottom view of the actuator electrode showing movable contact and the insulating polyimide layer. . . . . 88
4-12 Schematic diagram showing design parameters of the the MEM relay. . . . . 89
4-13 Optical microscope image of the fabricated actuator electrode of MEM relay. 92
4-14 SEM image of the fabricated substrate layer of MEM relay. . . . . 93
4-15 Fabricated MEM relay. . . . . 95


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

4-16 The source-drain current-voltage measurement at various gate voltage. . . . . 96
4-17 The SEM images showing the operation of the MEM relay. . . . . 96
4-18 The current $I_D$ vs. time plot of MEM relay demonstrating current switching. 97
4-19 Damaged polyimide insulating layer due to the dielectric break down of air . . 98
4-20 Maximum theoretical MAS frequency for various diameter rotors. . . . . 100
4-21 Schematic diagram showing the dimensions of the diamond rotor . . . . . 101
4-22 Laser micromachined inner and outer diameter of diamond rotor.. . . . . 102
4-23 The transmittance and absorption coefficient of CVD grown diamond. . . . . 103
4-24 The optical penetration depth of CVD diamond. . . . . 104

5-1 Cost of fabricating the comb-drive actuator using DRIE and MALL process. . 108
5-2 Cost of fabricating MEM relay using DRIE and MALL process. . . . . 110
5-3 Fabrication of 2.5D structure in silicon using laser-micromachining followed by anisotropic etch . . . . . 113
5-4 Fabrication of 2.5D structures using micro-milling. . . . . 114
5-5 Manipulation and laser micromachining of ultra-thin metal foils using a rigid frame. . . . . 115
5-6 The $12.5\mu m$ gap between the freestanding structure and the substrate fabricated using $12,5\mu m$ thick adhesive layer. . . . . 116
5-7 Part production rate of MEM relay fabrication using DRIE and MALL process. 119

B-1 Broken comb-drive actuators . . . . . 138
B-2 More broken comb-drive actuators . . . . . 139
B-3 MEM relay designs tested in this work . . . . . 141
B-4 Broken MEM relays . . . . . 143
B-5 Failed attempt to fabricate comb-drive in nitride wafer . . . . . 144




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 1

# Introduction

## 1.1 Motivation

Micro-electromechanical systems (MEMS) have many applications in fields like healthcare, consumer electronics, automobile, and energy. However, the development of novel MEMS devices has been significantly hindered by several limitations of the lithography-based microfabrication processes used to manufacture MEMS. First, these processes require access to expensive microfabrication tools such as photolithography systems, mask aligners, and deposition and etching tools. Moreover, these processes have a lengthy process development time. The high cost of equipment ownership and long development time associated with the lithography-based microfabrication processes makes them unsuitable for research and development, and low-volume MEMS production.

Second, the fabrication material is largely restricted to silicon and related materials such as silicon oxide and silicon nitride. The restriction in the choice of fabrication material limits the variety of MEM systems that can be designed and fabricated. For example, magnetic MEMS can potentially have distinct advantages over electrostatic/piezoelectric actuators such as high strength, polarity, and long actuation distance. However, fabricating magnetic MEMS is challenging due to the incompatibility of magnetic materials with lithography-based MEMS fabrication processes [1]. Furthermore, the performance of MEMS devices



440
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

is compromised due to the restricted choice of fabrication materials. For example, even though spring steel is an excellent choice for making high-performance micro-springs for MEMS, generally silicon is used because spring steel is incompatible with lithography-based MEMS fabrication processes.

Finally, during MEMS design and fabrication, integrating various materials and corresponding fabrication process is often the most difficult task [2]. In the lithography-based fabrication process, MEM systems are fabricated in monolithic fashion by successive deposition and patterning layers of materials. Each material layer is deposited and patterned using a unique process, which must be compatible with the previously deposited materials. However, integrating these processes is difficult due to material incompatibilities, thermal constraints (for example, polymeric parts cannot be subjected to high-temperature processing step such as LPCVD silicon deposition), and mechanical stability of interim structures. In summary, the conventional lithography-based MEMS fabrication processes are costly, offer a limited choice of fabrication materials, and the integration of various materials is challenging.

In this thesis, I present a novel approach for MEMS fabrication to overcome the above mentioned limitations. This approach uses multi-lamina assembly of laser micromachined laminates to fabricate MEMS and consists of two steps. First, the individual microstructure laminates are fabricated using direct-write laser-micromachining process. Next, the fabricated microstructure laminates are stack assembled and bonded to fabricate MEM systems. The presented MALL MEMS fabrication process enables greater material section and integration, low-cost manufacturing, rapid development, and integrated packaging.

Section 1.2 reviews the current work on laminate assembly and laser-micromachining. Section 1.3 presents the MALL process and its application in fabricating MEMS. Finally, section 1.4 compares the MALL process with prior work.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 1.2 Background

In recent years, there has been a considerable interest in developing alternative fabrication methods to manufacture MEM systems. These fabrication methods and their limitations are discussed in the following sections:

### 1.2.1 PCB-MEMS

PCB-MEMS utilizes printed circuit board (PCB) fabrication technology to fabricate MEM systems. Figure1-1(a), (b), and (c) shows mesoscale RF MEMS switches fabricated using PCB technology [3][4][5]. The functionalities and performance of these PCB-MEMS have been far from the MEMS fabricated using conventional lithography-based processes. This is due to several limitations of the PCB fabrication technology. First, the minimum feature size produced using PCB fabrication technology is larger than the feature size achievable in lithography-based microfabrication processes. Moreover, the aspect ratio of the fabricated structures is less than the aspect ratio achievable in conventional microfabrication processes such as deep reactive ion etching (DRIE). Second, the fabrication material is limited to materials available for PCB manufacturing such as copper foils, FR4, and certain polymers [6]. Finally, the alignment accuracy achievable in PCB technology is an order larger than the alignment accuracy required for silicon-based MEMS [3].

### 1.2.2 Laminate MEMS

Another method that is used to fabricate MEMS is laminate MEMS. In the laminate MEMS fabrication process, first, individual layers of the MEM systems are fabricated using various post semiconductor manufacturing (PSM) processes, and then the fabricated layers are laminated to manufacture MEMS. Since individual layers are fabricated separately, the lamination process enables the fabrication of MEMS from heterogeneous materials with incompatible fabrication processes. Some examples of laminate MEMS that are reported in the literature are optical waveguide switch [7](figure1-1(d)), magnetically actuated micro-




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

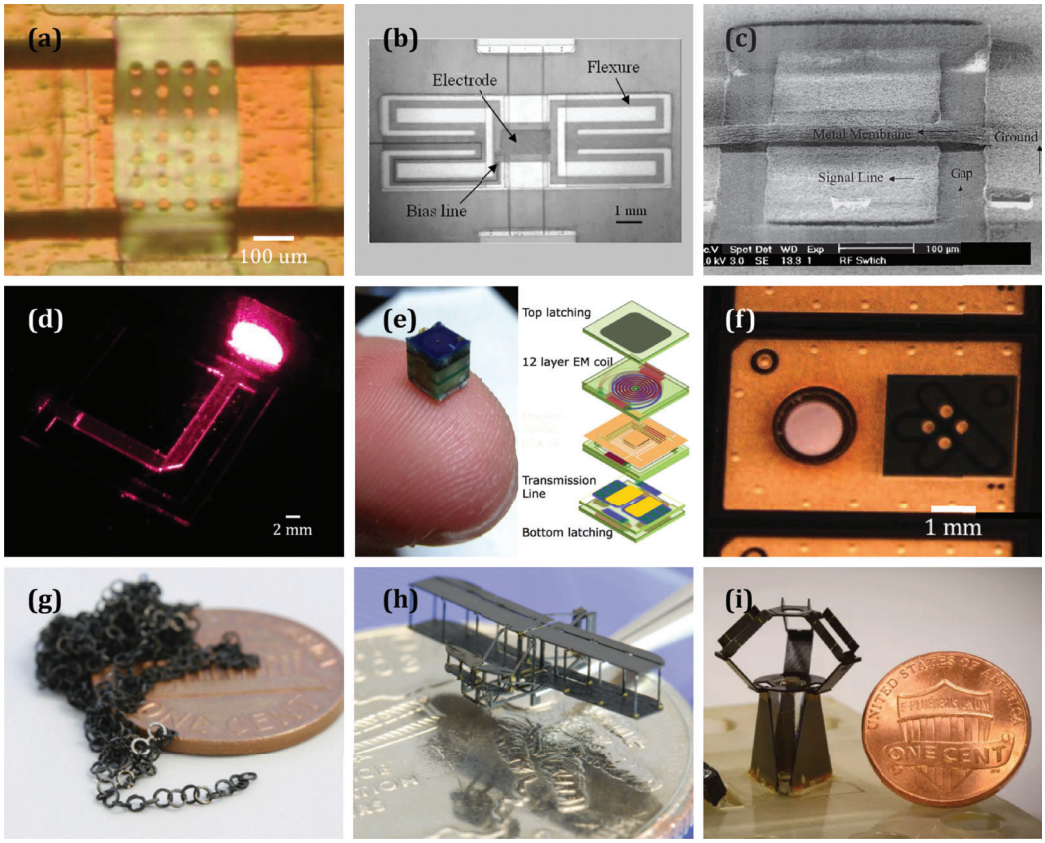



**Figure 1-1:** (a), (b) & (c) Mesoscale RF MEMS switches fabricated using PCB manufacturing technology [3][4][5]. (d) Optical waveguide switch fabricated using laminate MEMS technology [7]. (e) Laminate MEMS magnetically actuated micro-switch [8]. The device measures $3.2mm \times 3.2mm$. (f) MEMS Microphone fabricated using laminate MEMS [9]. (g) & (h) Meso-scale micro-structures fabricated using Pop-Up MEMS [9] . (i) Millimeter-scale robots fabricated using Pop-up MEMS [10]




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

switch [8] (figure1-1(e)), and laminate microphone [11] (figure 1-1(f)).

The laminate MEMS utilize many fabrication processes common to conventional microfabrication such as photolithography, wet etching, and deep reactive ion etching (DRIE). As a result, the laminate MEMS fabrication processes have similar limitations as the conventional MEMS fabricator processes, mainly high cost, long development time, and limited choice of fabrication materials.

### 1.2.3 Pop-up Book MEMS

Another method that has been reported to fabricate MEMS is pop-up book MEMS [9]. In this method, three-dimensional MEMS are fabricated by first, bulk micromachining rigid and flexible layers, then, laminating them, and finally folding the multi-layered rigid-flex laminate to build MEM systems. Figure 1-1(g) and (h) show reported mesoscale microstructures [9], and figure 1-1(i) shows fabricated millimeter-scale robots [10] using pop-up MEMS process. However, the pop-up book MEM systems are an order larger than the MEMS that can be fabricated using lithography-based microfabrication technology.

### 1.2.4 Limitations of Laser Micromachining

Although laser-micromachining has been around for many years, its application in fabricating MEMS has been limited [12][13][14][15] due to several reasons. First, manufacturing MEMS requires fabricating microstructures with small lateral feature-size (typically $10\mu m$) and large aspect ratio (typically 10:1 or higher). On the other hand, laser micromachining produces structures with larger lateral-features and low aspect ratio. As a result, the application of laser-micromachining has been largely limited to patterning surfaces [13] (figure 1-2(a)), drilling holes (figure 1-2(b)) [16][17], fabricating polymer micro-trenches (figure 1-2(c)) [18], and making millimeter-scale structures (figure 1-2(d), (e) & (f)) [19][18][13].

Second, the laser micromachined parts suffer from micro-burrs and debris around the machined parts. These imperfections are of the order of tens of microns, making these laser-


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

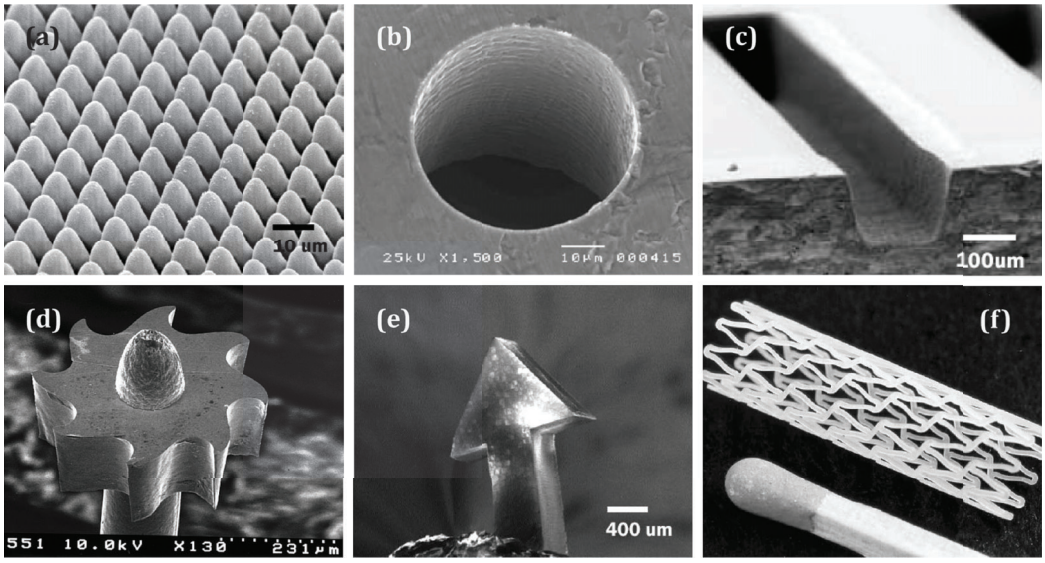



**Figure 1-2:** a) Surface patterning using an excimer laser [13]. (b) A 52$\mu m$ diameter hole drilled through stainless steel using a copper vapor laser (CVL) [16][17]. (c) Micro-trench machined into PMMA using femtosecond laser [18]. (d) Nickel micro-turbine motor (outside diameter is 470$\mu m$ and height is 150$\mu m$) fabricated using Laser-LIGA [19]. (e) Silicon wafer micromachined using nanosecond laser [13]. (f) A medical stent micromachined from a biodegradable polymer using a femtosecond laser [18].




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

micromachined parts unsuitable for MEMS fabrication. Eliminating these micro-burrs from delicate MEMS microstructure parts is challenging. Methods like sonication and physical scrub are not practical for MEMS structures due to their fragile nature. Similarly, etching process often results in dissolving the MEM structures along with the burrs because the size of the burrs is of the same order as the size of MEMS parts. Due to this reason, the laser-micromachining process has been largely used to fabricate rigid parts such as patterned surfaces, holes and trenches, where physical scrub or sonication can be utilized to remove burrs.

## 1.3 MALL

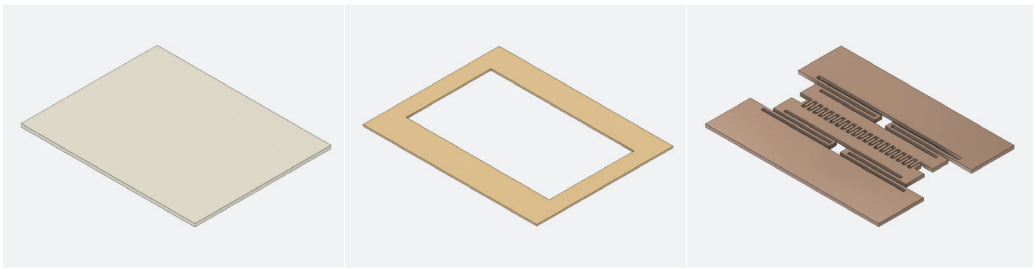
This thesis presents a novel approach to fabricate MEMS using multi-lamina assembly of laser-micromachined laminates (MALL). First, a novel microfabrication method called laser-assisted material phase-change and expulsion (LAMPE) micromachining is developed and used to manufacture small lateral-feature and high aspect ratio microstructure laminates. Second, these fabricated microstructure laminates are stack assembled to construct MEM systems. Figure 1-3 shows the schematic diagram of the MALL fabrication process.

### 1.3.1 LAMPE Micromachining

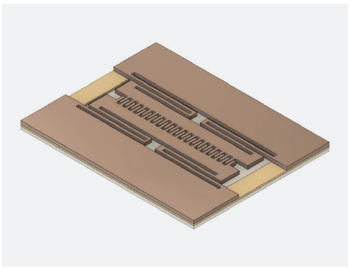
This thesis develops a novel approach to laser micromachine material sheets to create high aspect ratio structures significantly exceeding the optical penetration depth and thermal diffusion length of the material. The proposed approach utilizes laser-assisted material phase-change and expulsion (LAMPE) process to achieve the high aspect ratio while still maintaining the small lateral features. In this process, low energy laser power is used to first, induce a phase-change in the material, and then expel the altered material through the rear of the sheet utilizing pressure generated by vapor baubles formed at the laser-material interface due to the partial vaporization of material.

The proposed LAMPE micromachining can be used for a wide range of materials such


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.



(a)



(b)

**Figure 1-3:** Schematic diagram of the MALL fabrication process. It consists of two steps: (a) fabricating the individual layer of MEM system using LAMPE micromachining, and (b) Multi-lamina assembly of LAMPE micromachined parts to fabricate MEM system.

as metals, silicon, and diamond. In LAMPE micromachining of metal sheets, first, the longitudinal profile of the laser beam is used to melt (i.e., induce phase-change) the material, and then, the liquid-phase metal is expelled through the rear of the sheet; as a result, creating features as small as the diffraction-limited spot size and as deep as the Rayleigh length of the focused beam. Such small features-size and high aspect ratio are not achievable using conventional laser micromachining utilizing vaporization to remove the material.

Similarly, in LAMPE micromachining of silicon, low energy laser pulses are used to oxidize the silicon (i.e. induce phase-change). The newly formed silicon oxide transmits the laser (due to its low absorption coefficient), causing further oxidation of silicon beneath it. As a result, enabling oxidation far beyond the optical penetration depth of silicon. The formed silicon oxide is granular and expelled through the rear of the wafer utilizing pressure generated by vapor baubles formed at the laser-material interface due to the partial vaporization of material.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Likewise, in LAMPE micromachining of diamond, first, the laser is used to convert the diamond into graphite (i.e. induce phase-change), and then the newly formed graphite is removed by oxidizing the graphite into carbon dioxide. Using LAMPE micromachining of diamond, a very high aspect ratio machining of diamond is reported.

Using the LAMPE micromachining process, lateral-features as small as $10\mu m$ and aspect ratio as large as 10:1 can be fabricated. These features are comparable to features produced using conventional lithography-based microfabrication processes such as deep reactive ion etching (DRIE). The LAMPE micromachining process is described in chapter 3.

### 1.3.2 Electro-deburring

For metals, the expelled liquid-metal resolidifies at the edges of the LAMPE micromachined part. These resolidified liquid-metal burrs are removed using an electro-deburring process developed in this thesis. The process utilizes high current density at sharp burrs and edges to selectively electrochemically etch burrs without significantly affecting the dimensions of the LAMPE micromachined part. It must be noted that for MEM systems that do not require high aspect ratio structures with small lateral features, traditional pulsed laser ablation (PLA) micromachining in conjunction with the electro-deburring process can be used for MALL MEMS fabrication. The electro-deburring process is described in section 3.5.

### 1.3.3 Multi-Lamina Assembly

The next step in the MALL process is to assemble the laser-micromachined layers. Mechanical alignment of laminates using dowel pins and alignment holes is widely used in precision macro-fabrication, and employed to align and bond microstructure laminates. Figure 1-4 shows the schematic diagram of the alignment-pin and alignment-hole assembly process. Using this method, alignment accuracy exceeding $2.5\mu m$ is demonstrated. The multi-lamina assembly process is described in section 4.1.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

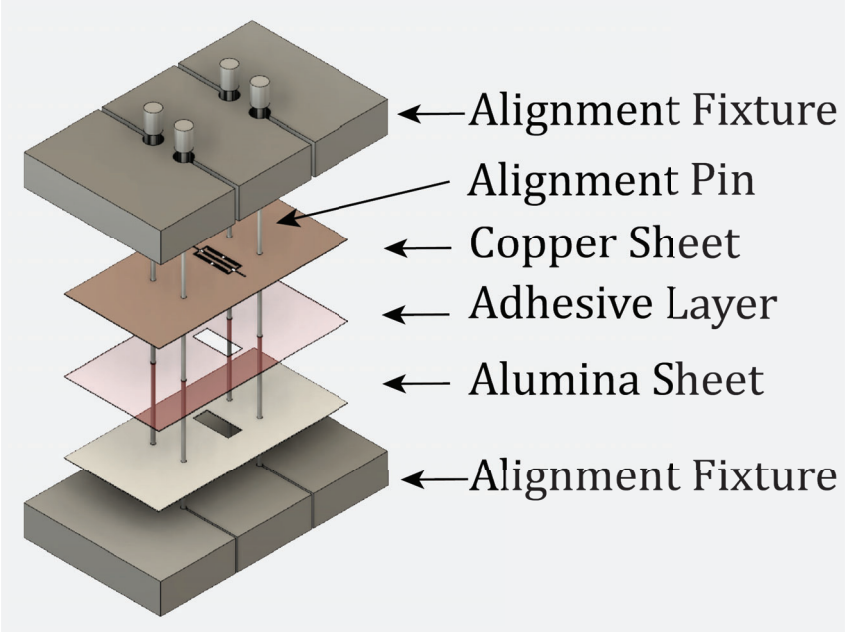



**Figure 1-4:** Schematic diagram of multi-lamina assembly.

## 1.4 Comparison with Prior Work

Table 1.1 provides a thorough comparison of MALL fabrication with other alternate MEMS fabrication processes. As can be seen, DRIE is desirable for small lateral features and high aspect ratio, however, falls short on the choice of fabrication material. After DRIE, the LIGA process is desirable for small lateral features and high aspect ratio, however, fall short on creating free-standing microstructures and require expensive x-ray source.

The MALL process presented in this thesis uses the similar lamination process that is used in PCB-MEMS, laminate MEMS, and Pop-up book MEMS. However, there are two major differences between the MALL process and other laminate PCB-MEMS, laminate MEMS, and pop-up book MEMS processes described in section 1.2. These differences are discussed in the following sections.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### 1.4.1 High Aspect Ratio & Small Lateral Feature

In the conventional MEMS fabrication process, deep reactive ion etching (DRIE) is used to fabricate structures with high aspect ratio and small lateral features. Typically, structures with the aspect ratio as large as 10:1 and lateral features as small as $5 - 10\mu m$ can be micromachined using DRIE [20]. The high aspect ratio machining is required to fabricate MEMS structures that can move in the direction parallel to the MEMS device plane such as interdigitated comb fingers and beam springs. These structures are common in many MEMS such as comb-drive actuators [21], accelerometers [22], resonators [23], and electromechanical filters [24].

In the PCB-MEMS, individual layers of the MEM system are fabricated using photolithography, followed by wet etching. However, the wet etching process is isotropic, thereby resulting in low aspect ratio. Similarly, in laminate MEMS, the individual layers are fabricated using various processes such as stamping, laser cutting, and electroforming [6]. These processes produce structures with larger features and limited aspect ratio than DRIE. Hence, the performance of these MEMS has been far from MEMS produced using conventional microfabrication process utilizing DRIE [4][25][26][27][11]. Moreover, the low aspect ratio structures have restricted movement in the direction of the device plane, and the motion is constrained to the direction normal to the device. As a result, the application of PCB-MEMS and laminate MEMS have been limited to MEM systems in which the structures move perpendicular to lamination plane, such as MEMS RF switch [4] and microphones [11].

In the MALL process, individual layers of the MEM systems are fabricated using LAMPE micromachining. The LAMPE micromachining process can fabricate microstructure laminates with aspect ratio as large as 10:1 and features as small as $10\mu m$. Such high aspect ratio while maintaining the small lateral features is not achievable in PCB-MEMS, laminate MEMS, and pop-up book MEMS fabrication processes. In fact, traditionally, such high aspect ratio and small lateral features were attainable exclusively using deep reactive ion etching (DRIE) [28]. The small features and large aspect ratio


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

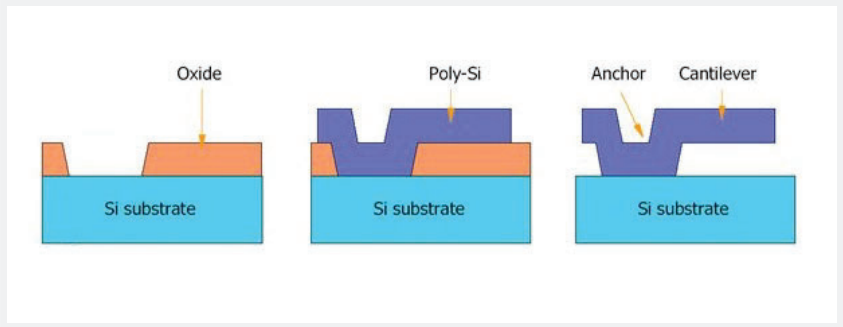



**Figure 1-5:** Schematic diagram of surface micromachining process. (a) First, a sacrificial layer is deposited and patterned. (b) Subsequently, structural layer is deposited on top of the sacrificial layer. (c) Finally, the sacrificial layer is removed to create free-standing microstructures.

structures fabricated using LAMPE micromachining are comparable to structures produced in DRIE. Hence, the MALL fabrication process can manufacture MEMS that are comparable to MEMS fabricated using DRIE.

### 1.4.2 Freestanding Structures with Six Degrees of Freedom

In conventional MEMS fabrication, surface micromachining [48][49][50] is used to fabricate free-standing microstructures. Figure 1-5 shows the schematic diagram of the surface micromachining process. First, a sacrificial layer is deposited on a substrate and patterned using photolithography and etching. Next, a structural layer is deposited and patterned on top of the sacrificial layer. Finally, the sacrificial layer is removed to create free-standing microstructures. The deposition process, such as chemical vapor deposition (CVD) used to deposit the sacrificial layer and structural layer, allows a high degree of control over the thickness of the deposited material and the gap created between structure and the substrate.

Fabricating similar thin freestanding structures in the lamination process is challenging. The difficulty is due to two reasons. First, as the thickness of the laminate is reduced, handling and micromachining these ultra-thin layers becomes difficult. Second, the freestanding structures are created using a spacer layer, which is patterned and sandwiched between the layer containing freestanding structure and the substrate. The




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

thickness of the spacer layer determines the gap between the freestanding structure and the substrate. As this gap and the size of freestanding structures is reduced, manipulating and patterning the ultra-thin spacer layer gets challenging. As a result, the applications of PCB-MEMS and laminate MEMS have been largely limited to either fabricating MEMS with no freestanding structures such as microfluidic devices [51][52] and optical MEMS [7] or MEMS with large gap between the freestanding structure and substrate such as such as MEMS microphone and electromagnetic switch [4][11][8][25][27].

The laser micromachining process used in MALL allows effortless patterning of B-staged epoxy adhesive layers. By using patterned adhesive layers as a spacer layer, tack bonding process, and high alignment accuracy, this thesis demonstrates fabricating freestanding structures with a gap between the structures as small as $12.5\mu m$ (as shown in figure 5-6). Similarly, by bonding ultra-thin film to rigid frame for manipulation and performing LAMPE micromachining, this thesis demonstrates fabricating freestanding structures as thin as $10\mu m$ (as shown in figure 5-5). These microstructures can be fabricated with small features and satisfactory aspect ratio to allow six degrees of freedom (DOF) movement.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| Parameter | Material Choice | Minimum Lateral Features | Maximum Aspect Ratio | Minimum Layer Thickness | Minimum Gab Between Layers | Allow Sacrificial Layer? | Comments |
|---|---|---|---|---|---|---|---|
| **DRIE** | Silicon | $0.374\mu m$ [29], $10-15\mu m$ [30] | 107:1 [29], $\sim 40:1$, [30] | $< 1\mu m$ | $< 1\mu m$ | Yes | High cost |
| **PCB-MEMS / Laminate MEMS** | Metals, FR4, polymers [32][33] | $10\mu m$ [34] , $25\mu m$ [35][36] | Quantitative data not available [6][37] | $\sim 50\mu m$[38] | $20-30\mu m$ [39], $87\mu m$ [40], $70\mu m$ [41] | Yes | Low aspect ratio due to isotropic wet etching of metals |
| **Pop-up Book MEMS** | Metals, FR4, and polymers | few microns [42], $\sim 20\mu m$ [43] | NA | NA | NA | NA | Produces Meso-scale MEMS |
| **Direct-write Printing** | Colloidal metal and ceramic particles | $20\mu m$ [44], $100\mu m$ [45], $5\mu m$ [46] | 400 [46] | $20\mu m$ [44], $100\mu m$ [45], $5\mu m$ [46] | NA | NA | Very slow process [46] |
| **LIGA** | Metals that can be electroplated | $7\mu m$ [47] | 50:1 [47] | $\sim 1\mu m$ | NA | Yes | Require expensive x-ray source |
| **MALL** | Metals, Silicon, Diamond, and Polymers | $5-10\mu m$ | 10:1 | $12.5\mu m$ | $12.5\mu m$ | Potentially yes but not demonstrated in this thesis | Rapid development, low-cost fabrication |

**Table 1.1:** Comparison of various MEMS fabrication processes.

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 2

# Theory of Pulsed Laser Ablation

This chapter describes the theory of pulsed laser ablation (PLA) and its application in micromachining. The theoretical understanding of PLA is essential in estimating the limits on feature size and aspect ratio achievable in PLA micromachining, causes of these limitations, and potential solutions. Section 2.1 describes the analytical model of the pulsed laser ablation process and its application in laser micromachining. Section 2.2 characterizes the laser micromachining system used in this work.

## 2.1 Pulsed Laser Ablation

Figure 2-1 shows the schematic diagram of pulsed laser ablation process. When a laser beam hits a material target, it is absorbed by the free electron gas of the material through electromagnetic interaction, thereby increasing the temperature of the electron gas. Then, the energy of the electron gas is transferred to the lattice through the electron-phonon coupling, causing an increase in temperature of the material [53].

The duration between laser absorption and heating of the material is in the order of a few picoseconds. As a result, the electron-lattice interaction can be ignored for laser ablation utilizing long pulse duration in the order of nanoseconds, and the temperature distribution can be modeled using the classical heat diffusion equation. The absorbed laser heats the




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

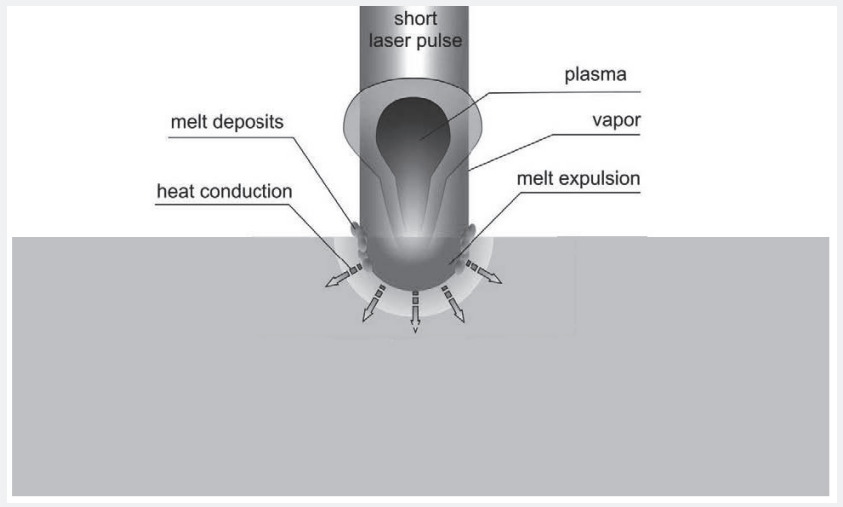



**Figure 2-1:** Schematic diagram of pulsed laser ablation (PLA) process. [54]

material and increases its temperature. When the temperature reaches the boiling point, the material vaporizes and gets removed from the target material.

The absorbed laser energy per unit volume and per unit time, $Q$ is given by the expression [55]:

$$Q(x, y, z) = (1 - R)I(x, y, t)f(x) \tag{2.1}$$

where $R$ is reflective of the material surface, $I(x, y, t)$ is the intensity of the laser beam, and $f(z)$ is the attenuation function due to laser beam absorption, given by the equation [55]:

$$f(z) = \alpha(T(z))exp\left[ - \int_0^z \alpha(T(z'))dz' \right] \tag{2.2}$$

where $\alpha(T(z))$ is the temperature-dependent absorption coefficient. For pulsed ablation of metals studied in this work, it is assumed that the value of $\alpha$ is temperature independent. As a result, the above equation reduces to:

$$f(z) = \alpha e^{-\alpha z} \tag{2.3}$$




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

Substituting the expression for $f(z)$ in equation 2.1, we get the expression for the laser heat source as:

$$Q(x, y, z) = (1 - R)\alpha I(x, y, t)e^{-\alpha z} \tag{2.4}$$

Here, the characteristic length, known as optical penetration depth, $l_\alpha$ is defined as the distance at which the intensity is $1/e$ times the maximum intensity. Therefore,

$$l_\alpha = \frac{1}{\alpha} \tag{2.5}$$

Note that in the derivation of equation 2.4, it is assumed that the laser intensity is constant in the direction of propagation of the laser. This assumption is valid in situations where the optical penetration depth is smaller than the Rayleigh length of the laser beam. This assumption holds for the laser micromachining performed in this work.

The heat conduction in the material can be modeled using the heat diffusion equation [55]:

$$\rho c_p \frac{\partial T}{\partial t} = \rho c_p D \nabla^2 T + Q\delta(\vec{x}, t) \tag{2.6}$$

This heat conduction leads to an increase in temperature of the material. Once the temperature reaches boiling point, the material gets vaporized and removed from the target, resulting in micromachining. Next section describes the laser micromachining system used in this work.

## 2.2 Laser-Micromachining System

This work uses a commercially available laser micromachining system from Oxford Lasers Ltd [56]. The system consists of three major components: 1) Laser source, 2) Focusing optics, and 3) Motion system. These components are discussed in the following sections.



456
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### 2.2.1 Laser Source

The laser source is an important component of a laser micromachining system and determines the feature size and type of materials that can be processed. There are two types of lasers that are employed in laser micromachining: pulsed laser and continuous-wave (cw) laser. The pulsed laser has two advantages over continuous-wave laser. First, the short laser pulse enables fine control over the energy dumped into the target, which is essential for machining small features. Second, the peak power of the individual pulse is several orders of magnitude higher than the average power of the laser. As a result, the pulsed laser can ablate the material even with much lower average power.

The laser source used in this work is a Q-switched frequency-doubled Nd:YAG diode-pumped solid-state (DPSS) laser. The wavelength of the laser is $532nm$ and the pulse duration is $20ns$. The mode of the laser beam is $TEM00$ and the beam quality $M^2 < 1.2$.

The average power, $P_{avg}$ of the laser is between $2 - 6.5W$ and depends upon the pulse frequency. Figure 2-2 shows the variation of measured average power, $P_{avg}$ with respect to the laser pulse frequency, $F_p$. The average power at $5KHz$ pulse repetition rate is $2.8W$.

The pulse energy $E_p$ for individual pulses can be calculated by dividing the average measured power by pulse frequency, $F_p$.

$$E_p(J) = \frac{P_{avg}\ (W)}{F_p\ (Hz)} \tag{2.7}$$

Figure 2-3 shows the variation of pulse energy, $E_p$ with respect to pulse frequency, $F_p$.The pulse energy at $5KHz$ pulse repetition rate is $561\mu J$.

The peak power (or pulse power) can be calculated by dividing the pulse energy with pulse duration ($20ns$ for the laser used).

$$Peak\ Power\ (W) = \frac{E_p\ (J)}{Pulse\ Duration\ (s)} \tag{2.8}$$

Figure 2-4 shows the variation of the peak power with respect to pulse frequency. It can be seen that the peak power is in the order of kilowatts even though the average power is



457
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

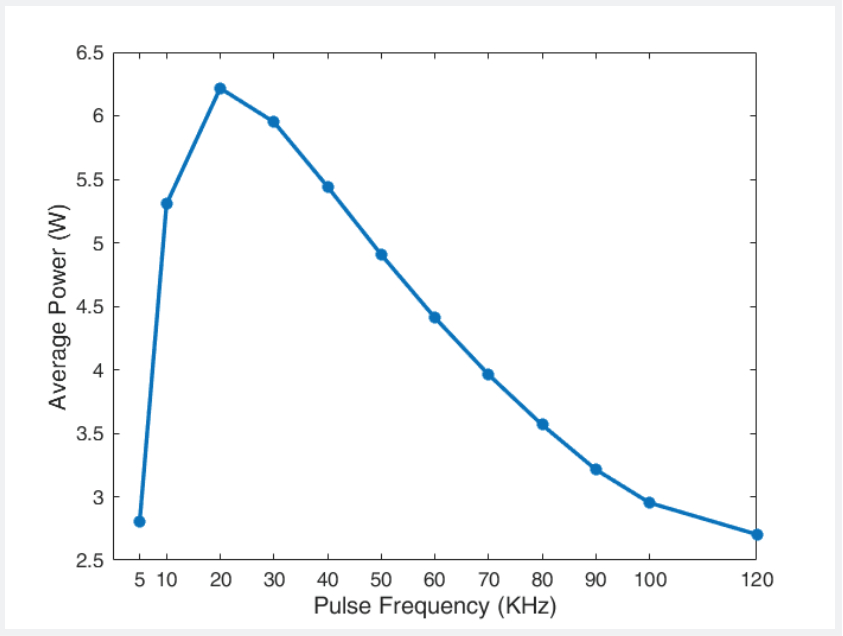



**Figure 2-2:** The plot of the average power , $P_{avg}$ with respect to the pulse frequency

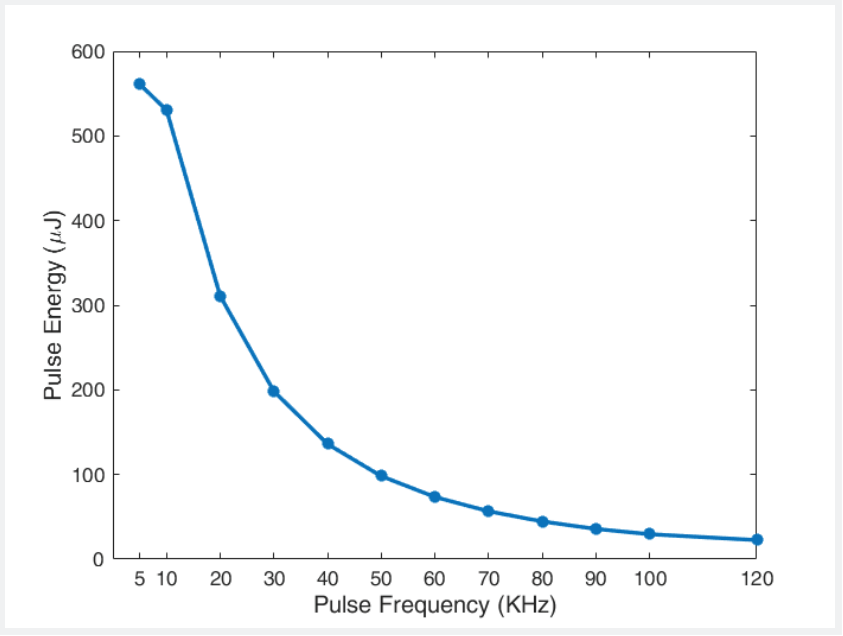



**Figure 2-3:** The plot of pulse energy, $E_p$ with respect to the pulse frequency.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

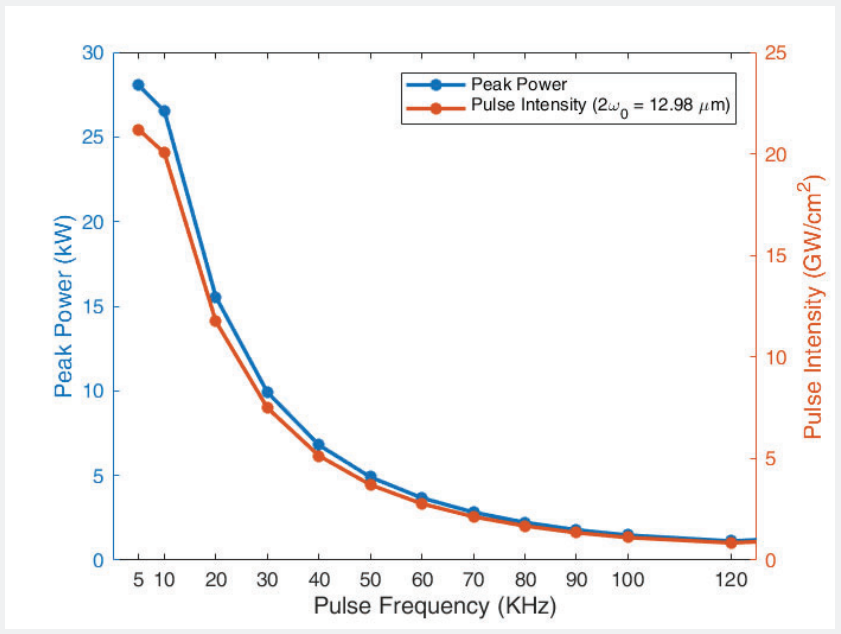



**Figure 2-4:** The plot of peak power (left) and pulse intensity (right )with respect to the pulse frequency.

less than $6.5W$. The peak power of the laser pulses is responsible for the ablation of the material. The peak power at $5KHz$ pulse repetition rate is $28kW$. Table 2.1 summarizes the values of the parameters calculated for the laser used in this work.

| **Laser Parameter** | **Value** |
|---|---|
| Wavelength, $\lambda$ | $532nm$ |
| Average Power, $P_{avg}$ | $2.8W$ |
| Pulse Frequency, $F_p$ | $5KHz$ |
| Pulse Duration, $\tau_l$ | $\approx 20ns$ |
| Pulse Energy, $E_p$ | $561\mu J$ |
| Peak Power, $P_{pk}$ | $28kW$ |

**Table 2.1:** The values of various parameters of laser source.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### 2.2.2 Beam Delivery & Focusing Optics

The laser beam is delivered to the target material using a series of mirrors, lenses, and other optical elements. First, the laser beam emitted from the source is expanded using a beam expander. The diameter of the laser beam after the beam expansion is $10mm$. The beam is then allowed to pass through a variable attenuator, which is used to attenuate the laser beam and control the energy dumped into the target. Finally, the laser beam is focused onto the target using either a lens or a microscope objective.

#### Diameter of Focused Beam

For a $TEM_{00}$ mode laser focused using a thin lens, the intensity profile at the focal plane has a Gaussian distribution and given by the equation [55]:

$$I(r) = I_0 exp\left( - \frac{2r^2}{\omega_0^2} \right) \tag{2.9}$$

where $\omega_0$ is the beam spot radius, defined as the distance at which the intensity $I(\omega_0) = I_0/e^2 = 0.135I_0$ (2.2.2). The Gaussian intensity distribution has no obvious boundaries to give it a characteristic dimension. However, the distance between the points where the intensity has fallen to $1/e^2 = 0.135$ of the maximum value is generally considered as the diameter of the focused beam and contains 86% of the total power of the laser beam [57].

The value of the diameter of beam spot size, $2\omega_0$ is given by the equation [57]:

$$2\omega_0 \ (Beam\ Spot\ Diameter) = 2.44\frac{f\lambda}{d} \tag{2.10}$$

where $\lambda$ is the wavelength of laser beam, $d$ is the beam waist, and $f$ is the focal length of the lens.

For the laser used in this work, the wavelength $\lambda = 532nm$, beam waist (or aperture diameter) $d = 10mm$, and focal length $f = 100mm$. The calculated focused spot size $2\omega_0 = 12.98\mu m$. The focused beam spot size $2\omega_0$ is an important parameter in determining the minimum achievable feature size in laser-micromachining.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

In addition to lens, microscope objectives can also be used for focusing the beam. Generally, an infinity-corrected objective is desirable as it enables simultaneous laser micromachining and imaging, and easy insertion of additional optical components. In an infinity-corrected objective, the light effectively travels parallel to the optical axis. As a result, additional optical components such as optical filters, polarizer and beamsplitters can be inserted between the tube lens. To create an image with an infinity-corrected objective, a tube lens must be used to focus the image.

The diffraction-limited spot size of a laser beam focused using a microscope objective is given by the equation:

$$2\omega_0 = \frac{1.22 * \lambda}{NA} \tag{2.11}$$

where $\lambda$ = wavelength of light and $NA$ is the numerical aperture.

**Intensity of Focused Beam**

The average pulse intensity of the focused laser beam is given by the expression:

$$Average\ Pulse\ Intensity, I_p\ (\frac{W}{cm^2}) = \frac{Peak\ Power\ (W)}{Focused\ Beam\ Spot\ Size\ (cm^2)} \tag{2.12}$$

The peak power of the laser pulse at $5KHz$ is $28W$ and the diameter of the focused beam spot is $\omega = 12.98\mu m$; therefore, the pulse intensity is $21.2GW/cm^2$. The right y-axis of figure 2-4 shows the variation of pulse intensity $I_p$ with respect to the pulse frequency.

The total power of the laser beam is given by the equation [55][58][59]:

$$P = 2\pi \int_0^\infty rI(r)dr = \frac{\pi\omega_0^2 I_0}{2} \tag{2.13}$$

$$\implies I_0 = \frac{2P}{\pi\omega_0^2} \tag{2.14}$$




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

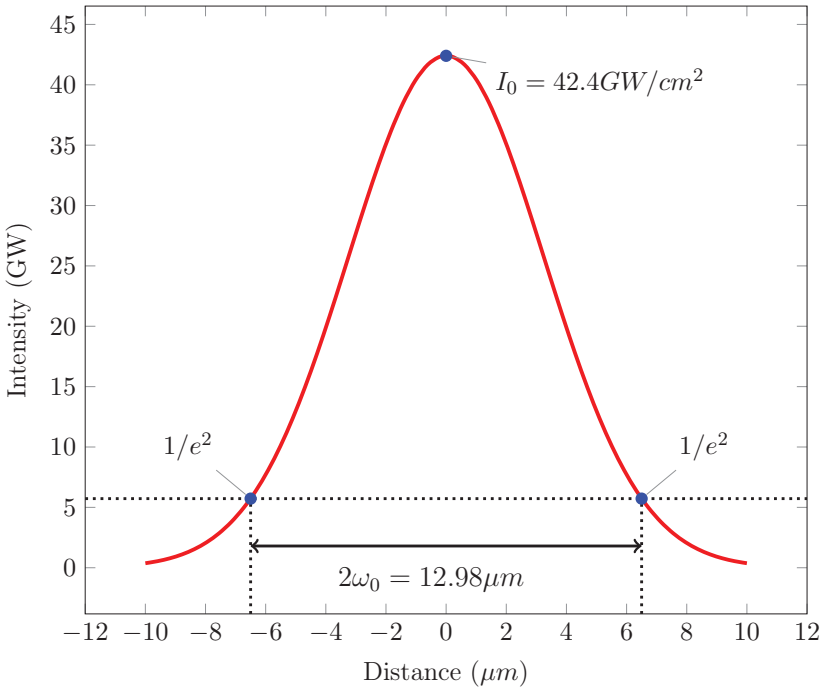



**Figure 2-5:** Intensity profile of focused laser beam. The beam spot diameter is $I_0 = 21.2MW/cm^2$ and the maximum intensity $I_0 = 21.2MW/cm^2$.

As can be seen, the intensity of the central peak of a Gaussian beam is twice the average intensity of the laser pulse. The peak power of the laser at $5KHz$ pulse frequency is $P = 28KW$ (Table 2.1). Substituting this value in equation 2.14 along with $2\omega_0 = 12.98\mu m$ (2.2.2), we get $I_0 = 2I_p = 42.4GW/cm^2$. Figure 2-5 shows the intensity profile of the laser beam.

### Rayleigh length

The radius of the laser beam at a distance $z$ from the focal plane is given by the equation:

$$\omega(z) = \omega_0\sqrt{1 + \left(\frac{z}{z_R}\right)^2} \tag{2.15}$$

where $z_R$ is the Rayleigh length and defined as the distance at which the radius of the beam



462
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

is $\omega_0\sqrt{2}$, and given by the expression:

$$z_R = \frac{\pi\omega_0^2}{\lambda} \tag{2.16}$$

For the laser used in this work, the Rayleigh length, $Z_R = 850\mu m$. The region $|z| \leq z_R$ is known as the Rayleigh range.

### 2.2.3 Motion System

A 2-axis stage is used to move the target with respect to the laser to perform laser micromachining. For each axis, the resolution is $0.250\mu m$, repeatability is $0.750\mu m$, and accuracy is $+/-2\mu m$. The maximum attainable speed is $250mm/s$, and the maximum attainable acceleration/deceleration is $10,000mm/sec^2$. It must be noted that the practical limit on maximum attainable speed for micromachining MEMS parts is restricted to $1mm/s$ or less due to their small size and limited acceleration/deceleration of the stage.

The feature size in laser micromachining process depends upon the laser energy dumped into the target. This dumped energy is controlled by changing the pulse energy and the overlap between the successive laser pulse exposure. For the system used in this work, the focused beam spot size is $12\mu m$, and the minimum pulse repetition rate is $5KHz$, as a result, the target is set to $1mm/s$ speed to ensure a pulse overlap of 20%. However, it must be noted that due to the small size of the MEM systems and limitation on the maximum attainable acceleration, the actual speed while machining sub-millimeter part is less than $1mm/sec$, resulting in more energy dumped into the target than desired.

Using a galvanometer scanner is desirable instead of using a 2-axis stage. The galvanometer can provide much higher feed rate, thereby enabling fabricating small features by allowing fine control over the energy dumped into the target. Moreover, the high feed rate can increase the micromachining rate. A typical 2-axis galvanometer mirror can operate at $100Hz$ frequency [60]. Assuming, the scan distance of $1mm$, a feed rate of $100mm/s$ can be easily achieved.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 2.3 Summary

In summary, this chapter lays the foundation for the theory of laser micromachining, which is used in next chapter to estimate the theoretical limit on the feature size of LAMPE micromachining process. The value of laser parameters such as focused beam spot size, pulse energy, and peak power determines the mode of laser micromachining, and must be determined. The laser pulse energy above a threshold value results in pulsed laser ablation, while the pulse energy below this threshold results in LAMPE micromachining. The threshold value depends upon the material property, beam spot size, and laser pulse duration. This chapter describes the experimental procedure and calculations used to measure and derive these laser parameters for the system used in this thesis. This parameters are summarized in table 2.1.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 3

# Laser-assisted Material Phase-change & Expulsion

## 3.1 Introduction

One motivation of this thesis is to expand the material library for fabricating MEMS, especially enabling fabrication using metals because they offer a wide range of mechanical and electrical properties. Moreover, these properties can be tuned by the alloying process. The ability to fabricate MEM systems using metals will empower new paradigms in MEMS design. Moreover, the metal foils used for MEMS fabrication are low-cost and easy to fabricate using roll forming process [61]. Unfortunately, the existing wet and dry etching processes are isotropic and not suitable for fabricating high aspect ratio microstructures required for making MEMS.

At a macro-scale, lasers are widely used for cutting metal sheets. In this process, laser is used to melt the metal, and a high-pressure air jet is used to expel the melted material. However, high-pressure jets cannot be employed at a micro-scale due to the delicate nature of the laser micromachined structures. Due to this reason, laser micromachining utilizes vaporization to remove the material. However, the vaporization process results in low aspect ratio micromachining.



465
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

The low aspect ratio in pulsed laser ablation micromachining utilizing vaporization is due to several reasons. First, during the ablation process, the vaporized material creates a plasma above the ablated region (figure 2-1). This plasma absorbs the laser, thereby reducing the energy available for ablation. As the depth of ablation crater increases, the laser absorption by plasma becomes significant, and laser energy available to the target is no longer sufficient to process the material. This phenomenon limits the maximum depth of micromachining achievable using the PLA laser-micromachining process.

Second, the aspect ratio depends upon the optical penetration depth of the material. For metals, the absorption coefficient $\alpha > 10^5 cm^{-1}$ [55], and the optical penetration depth, $1/\alpha$ is in the order of a few nanometer. Therefore, the laser energy is confined to a very thin layer on the surface, and the laser acts as a surface heat source. Due to isotropic heat conduction, the surface heating of the metal results in an equal temperature gradient in all directions, thereby resulting in approximately 1:1 aspect ratio of ablation profile. However, it must be noted that for laser with Gaussian intensity distribution, the depth of the ablation profile is larger than the width due to high intensity at the center of the Gaussian beam.

Third, the material vaporization also results in large melt-pool around the ablated region (figure 2-1). This melted material is shifted by the high pressure escaping vapors, thereby increasing the width of the ablation region.

This thesis presents a novel approach to laser micromachine material sheets to create high aspect ratio structures significantly exceeding the optical penetration depth and thermal diffusion length of the material. In the proposed approach, instead of using the vaporization to remove the material, the metal is melted, and the liquid-metal is expelled through the rear of the sheet utilizing pressure generated by vapor baubles formed at the laser-material interface due to the partial vaporization of the material. The expelled material resolidifies at the periphery of the micromachined part. This resolidified material is removed using the developed electro-deburring process. The proposed approach, referred to as laser-assisted material phase-change and expulsion micromachining or LAMPE micromachining can be used to fabricate high aspect ratio structures with small lateral features.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 3.2 LAMPE

Figure 3-1 shows the schematic diagram of the LAMPE micromachining process. This process is similar to laser cutting of thick sheets of metals. In the laser cutting of metals, the laser is used to melt the metal (not vaporize) and a high-pressure air jet is used to expel the melted material. However, in the LAMPE micromachining process, the metal is ejected by the pressure generated from the vapor bubbles formed at the laser-metal junction.

A careful setting of laser power is required to achieve laser-assisted material-phase change and expulsion process. The laser power below it would not result in partial vaporization of the material, and no expulsion will occur. The laser power above it will result in the creation of large melt-pool and complete vaporization, thereby increasing the feature size. The value of laser power required for LAMPE micromachining can be estimated using Eq. (3.6). It is empirically shown that the LAMPE micromachining process can result in lateral feature size as small as the diffraction-limited spot size of the beam and the aspect ratio as long as the Rayleigh length. It must be noted that the LAMPE micromachining process is limited to fabricating high aspect ratio structure only in sheets.

The LAMPE micromachining process has several advantages over laser ablation micromachining. First, the laser power required for LAMPE micromachining is significantly lower than the power required for laser micromachining using vaporization. The low power requirement is because the enthalpy of melting is less than the enthalpy of vaporization. Moreover, the absence of vapor plasma (figure 3-1) results in no laser absorption and the entire laser power is available for material processing. Second, the thermal and mechanical shock exerted on the material during melting is less the thermal and mechanical shock exerted on the material during vaporization. As a result, the heat-affected zone (HAZ) in LAMPE micromachining is smaller than HAZ resulting in vaporization.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

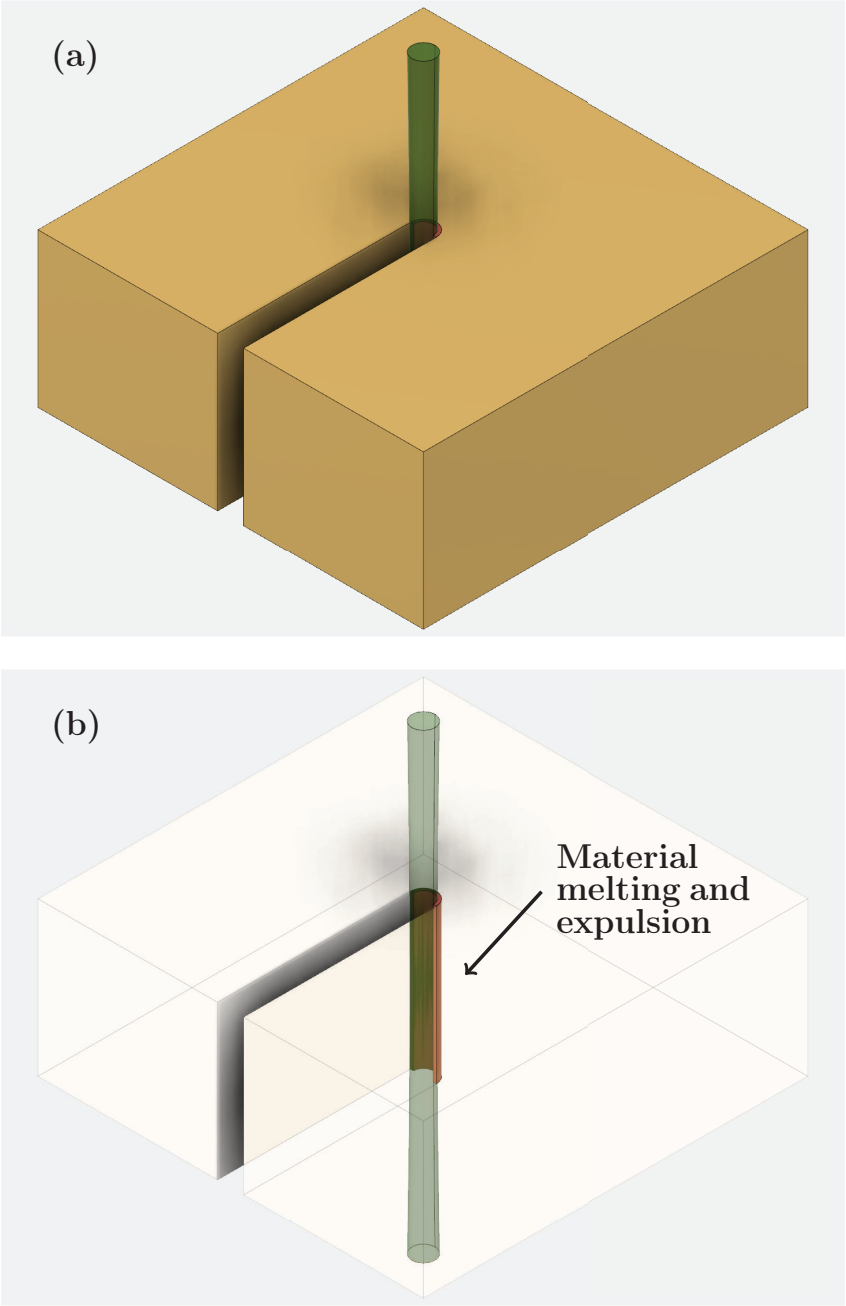



**Figure 3-1:** (a) Schematic diagram of LAMPE micromachining process illustrating fabrication of high aspect ratio structures. (b) Image highlighting the melt region of the sheet. The melted material is expelled downwards.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### 3.2.1 Difference between PLA and LAMPE Micromachining

The differences between conventional laser micromachining utilizing pulsed laser ablation and LAMPE micromachining process are as follows:

1. In conventional pulsed laser ablation (PLA) micromachining, the material is removed using vaporization, and the mass of the target is reduced after laser micromachining process. However, in the LAMPE micromachining process, the metal is melted, and the liquid-metal is expelled through the rear of the sheet using the pressure generated by vapor baubles formed at the laser-material interface. This expelled liquid-metal resolidifies at the edges of the micromachined part. Thus, the mass of the target remains unchanged. In LAMPE micromachining, the material is removed using the electro-deburring process.

2. The laser fluence required in PLA micromachining is higher than the laser fluence required in LAMPE micromachining process because the enthalpy of vaporization is higher than the enthalpy of melting. The low fluence requirement for LAMPE micromachining considerably reduces the cost of a laser source.

3. The conventional PLA laser micromachining process can be used to fabricate 2.5D structures by raster scanning overlapping pulse irradiation spots [62][63]. However, the LAMPE micromachining process is limited to micromachining material sheets.

4. The PLA micromachining produces lateral features larger than diffraction-limited spot size because of the formation of large melt-pool around the ablation region. However, as shown in section 3.4, the LAMPE micromachining can produce lateral features as small as the diffraction-limited spot size. Figure 3-2 shows the lateral feature size created using PLA and LAMPE micromachining process. As can be seen, the minimum feature size produced in PLA micromachining is $28\mu m$ and in LAMPE micromachining is $8\mu m$.

5. In the conventional laser micromachining, the vaporized material creates a plasma



469
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

above the ablation region. This plasma absorbs the laser beam, thereby reducing the energy available for vaporization and limits the achievable depth of micromachining. In contrast, no vapor plasma is formed in LAMPE micromachining, and the aspect ratio is only limited by the Rayleigh length of the laser beam.

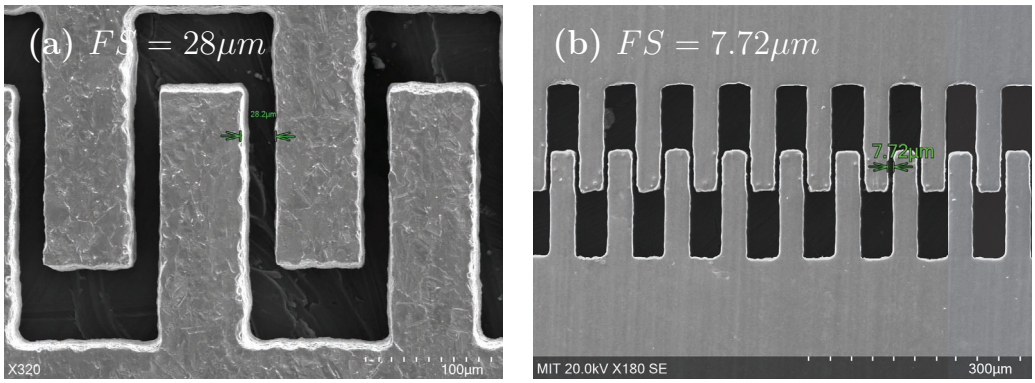



**Figure 3-2:** Comparison of lateral feature size (FS) produced in PLA and LAMPE micromachining. Note that these parts are cleaned using the electro-deburring process, which is discussed in section 3.5. (a) Lateral features produced in pulsed laser ablation (PLA) micromachining. (b) Lateral features produced in LAMPE micromachining.

## 3.3 Feature Size in LAMPE Micromachining

The minimum achievable feature size depends upon the focused beam spot size and laser pulse duration. In a conventional pulsed ablation process, the material is removed through vaporization, resulting in large melt-pool, plasma plum, and thermal shock on the material lattice. As a result, the minimum feature size is much larger than the theoretical limit [64][65][66][62]. However, using the laser-assisted material phase-change and expulsion (LAMPE) micromachining, feature size comparable to the theoretical limit can be achieved. Moreover, using LAMPE micromachining, much higher aspect ratio can be achieved than previously possible.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| Laser Parameter | Value |
|---|---|
| Wavelength, $\lambda$ | $532nm$ |
| Average Power, $P_{avg}$ | $28mW$ |
| Pulse Frequency, $F_p$ | $5KHz$ |
| Pulse Duration, $\tau_l$ | $\approx 20ns$ |
| Reflectivity of Copper, $R$ | 0.43 |
| Pulse Energy, $E_p$ | $3.2\mu J$ |
| Peak Power, $P_{pk}$ | $280W$ |
| Beam Spot size, $2\omega_0$ | $1\mu m$ |
| Rayleigh Length, $Z_R$ | $1.5\mu m$ |
| Maximum Intensity, $I_0$ of Focused Beam | $71.5GW/cm^2$ |

**Table 3.1:** Laser parameters for LAMPE micromachining.

### 3.3.1 Experiment & Results

Figure 3-3 shows the minimum achievable feature size of $5\mu m$ using LAMPE micromachining process. This feature is created by irradiating a single pulse of laser with wavelength $532nm$ and pulse duration $20ns$ on a copper target. The pulse energy is $5.6\mu J$, which is set by reducing the original pulse energy of $561\mu J$ to its 1 % value using an external attenuator. The reflectivity of the copper target, $R = 0.47$; therefore, the effective energy available for material processing is $3.2\mu J$. The laser beam is focused using a microscope objective with numerical aperture $NA = 0.65$. The focused beam spot diameter is $1\mu m$, giving the maximum intensity at the center of Gaussian distribution as $I_0 = 71.5GW/cm^2$. Table 3.1 summarizes the laser conditions used for LAMPE micromachining process.

The small feature size is possible in the LAMPE micromachining because the laser is used to melt and expel the material, thereby eliminating the drawbacks of vaporization present in the pulsed ablation process. The expelled liquid metal resolidifies at the rim of the hole and can be removed using an electro-deburring process developed in this work (section 3.5). Figure 3-4 shows the LAMPE fabricated hole before and after removing the resolidified metal.

The next section describes the analytical model for the LAMPE micromachining process. This model is similar to the pulse ablation model described in reference [55], with



471
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

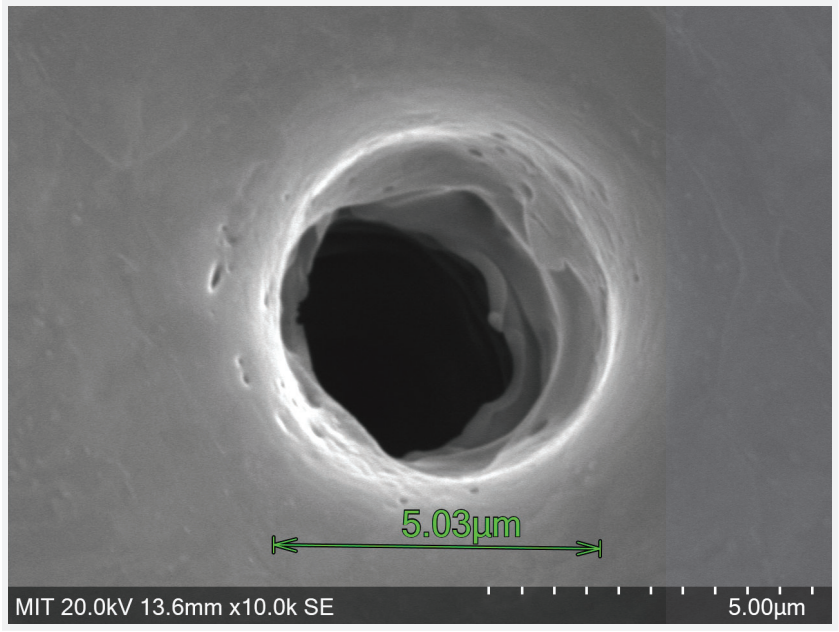



**Figure 3-3:** A crater fabricated using LAMPE micromachining to demonstrate the minimum achievable feature size. The diameter of the hole is $5\mu m$.

two major distinctions. First, the enthalpy of melting is used instead of enthalpy of vaporization. Second, as there is no vaporization, the model excludes plasma formation, laser absorption due to plasma, expansion of liquid and vapor, and hydrodynamic expulsion of melted material.

### 3.3.2 Analytical Model

This section describes the theoretical model to estimate the minimum achievable feature size in LAMPE micromachining and compares it with experiment results. As discussed earlier, the LAMPE micromachining is performed on a copper target using a single pulse of laser with pulse duration 20 nanosecond and wavelength $532nm$.

The optical penetration of the laser beam inside material is given by the Beer-Lambert law (eq: 3.1). Here, we can define a characteristic length, the optical penetration depth




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

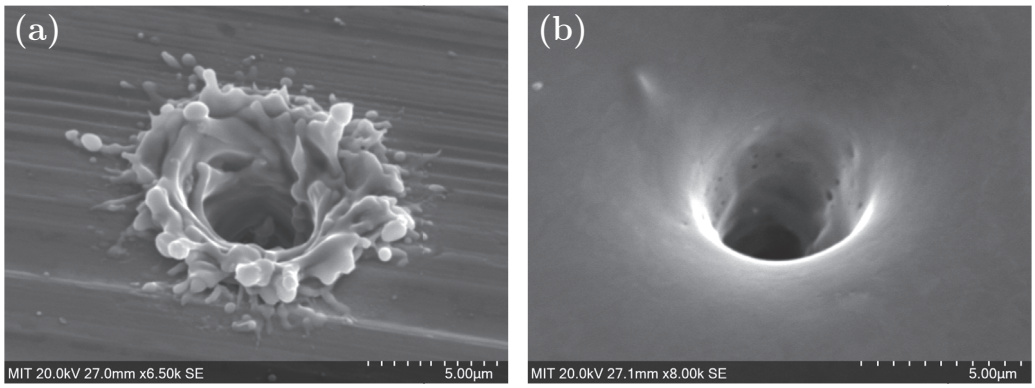



**Figure 3-4:** Electro-deburring of LAMPE micromachined crater. (a) Before electro-deburring. (b) After electro-deburring.

$l_\alpha = 1/\alpha$, which is a measure of how deep the laser penetrates inside the material.

$$I(z) = I_0 e^{-\alpha z} \tag{3.1}$$

where $I_0$ is the laser intensity at the surface, and $\alpha$ is the absorption coefficient of the material.

For copper, the absorption coefficient at $532nm$ is $\alpha_{Cu} = 7.14 \times 10^5 cm^{-1}$ [55], and the optical penetration depth is $14nm$. As a result, the laser is absorbed only at the surface and acts as a surface heat source. The absorbed heat further propagates to the bulk of the material through thermal conduction and can be modeled using the heat diffusion equation:

$$\rho c_p \frac{\partial T}{\partial t} + \rho c_p D \nabla^2 T = 0 \tag{3.2}$$

where $\rho$ is the mass density of the material, $c_p$ is the heat capacity of the material, $T$ is the temperature, and $D$ is the thermal diffusivity of the material.

For the experiment described in section 3.3, the diameter of the focused spot, $\omega_0 = 1\mu m$ and the optical penetration depth $l_\alpha = 14nm$ is much smaller than the dimensions of the target (for the copper target, $length, a = 10mm$, $width, b = 10mm$, and $thickness, t = 100\mu m$) i.e.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

$$\omega_0, l_\alpha \ll a, b, t$$

As a result, the laser surface source can be modeled as a *point source in space* and the fundamental solution of (3.2), describing the evolution of temperature distribution with time, is given by expression [55]:

$$T(\vec{x}, t) \approx \frac{Q}{\rho c_p (4\pi D t)^{\frac{3}{2}}} exp\left( - \frac{|\vec{x}|^2}{4Dt} \right) \tag{3.3}$$

The solution above depicts 1/e decay of the temperature in space. For laser irradiation with pulse duration $\tau_l$, we can define a thermal diffusivity length $l_T$, at which the temperature reaches 1/e times the maximum value. Therefore, at $|\vec{x}| = l_T$

$$-\frac{l_T^2}{4D\tau_l} = 1 \tag{3.4}$$

which gives the expression for thermal diffusion length as:

$$l_T = 2\sqrt{D\tau_l} \tag{3.5}$$

The thermal diffusion length is a measure of minimum feature size achievable using the LAMPE process. The thermal diffusivity $D$ for copper is $1.14 cm^2/s$. When the copper target is processed with a single laser pulse of duration $\tau_l = 20ns$, the thermal diffusivity length is $3.02\mu m$. Since the heat dissipates isotropically in all directions, the achievable feature size is $2 \times l_T + Beam\ Diameter = 7\mu m$. As shown in 3-3, the feature size is less than the theoretical value. This discrepancy could be due to ignoring the Gaussian intensity distribution of the focused laser beam.

It must be noted that the expression for thermal diffusion length $l_T$ given by (3.5) is applicable only for a point source in space. For all other cases, the thermal diffusion length $l'_T$, defined as $l'_T = T(0)(1/e)$, depends upon the particular boundary conditions of the problem and may differ significantly from equation (3.5).


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Threshold Fluence for LAMPE**

The threshold fluence can be estimated by calculating the energy required to melt the volume of the material covered by the thermal diffusion length. This volume, $V_{l_T}$ can be approximated as a hemisphere of radius $l_T$. If we ignore the heat loss due to surface reflection, convention, radiation, and hydrodynamic material expulsion, the threshold pulse energy is given by the expression:

$$E_{th_m} = \rho V_{l_T} c_p (T_m - T_{amb}) + \rho V_{l_T} \Delta H_m \tag{3.6}$$

where $\rho$ is mass density, $V_{l_T}$ is LAMPE micromachined volume, $c_p$ is specific heat capacity, $T_m$ is melting temperature, $T_{amb}$ is ambient temperature, and $\Delta H_m$ is latent heat of fusion (melting).

For copper, $\rho = 8.94 g/cm^3$, $c_p = 0.47 J/gK$ at $1000K$, $T_m = 1357K$, and $\Delta H_m = 200 J/g$. The ambient temperature is set to $T_{amb} = 298K$ and the volume $V_{l_T} = 5.76 \times 10^{-11} cm^3$. Substituting these values in equation 3.6, we get:

$$E_{th_m} = 0.36 \mu J \tag{3.7}$$

The laser beam is focused using a microscope objective with numerical aperture $NA = 0.65$ giving the focused beam spot diameter of $1 \mu m$. By diving the pulse energy $E_{th_m}$ by beam spot area, we get the threshold fluence $F_{th_m} = 4.6 \mu J/\mu m^2$. For comparison, an approximate estimation of pulse energy required for vaporization can be calculated by substituting $\Delta H_m$ with $\Delta H_v = 4700 J/g$, which gives $E_{th_v} = 3.04 \mu J$. Note that the actual pulse energy required for vaporization is much higher than this estimate due to the absorption of the laser by plasma plume [55].

The pulse energy for LAMPE micromachining, $E_{LAMPE}$ is greater than $E_{th_m}$ and less than $E_{th_v}$, but closer to $E_{th_v}$. The exact value can be determined empirically by performing a series of experiments of varying pulse energy between $E_{th_m}$ and $E_{th_v}$. As described in section 3.3 for LAMPE micromachining, the pulse energy $E_p = 3.2 \mu J$ is selected such




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

that it is greater than $E_{th_m} = 0.36\mu J$ to ensure melting of the metal, and comparable to $E_{th_v} = 3.04\mu J$ to ensure partial vaporization to generate recoil force to expel the melted material. It must be noted that the actual value $E_{th_v}$ is higher than $3.04\mu J$ because of heat loss due to convection, radiation, laser absorption by plasma, hydrodynamic material expulsion.

When processing new materials, I suggest using the calculations shown above to estimate the required pulse energy $E_{LAMPE}$ and fluence $F_{LAMPE}$ for LAMPE micromachining and use that as a starting point for further process development.

It must be noted that the material is largely melted (and only partially vaporized). This hypothesis can be validated empirically by performing a series of dot array machining with varying pitch distance. Figure 3-5 shows the micromachined dot array at pitch distance of $20\mu m$, $10\mu m$, and $5\mu m$. In figure 3-5(a), the machined dots are far apart, and the melted metal is forced to expel upward, creating a crater. As the dot patterns overlap, the pulsed laser ablation process involving vaporization to remove material results in a recess or $2.5D$ micromachining [63]. However, the LAMPE micromachining process results in resolidified slag as shown in figure 3-5(c). The slag is formed due to the mere shifting of the melted material and resolidification when dots overlap. These findings confirm that in the LAMPE micromachining process, there is minimal vaporization, and the micromachining process is dominated by liquid-metal expulsion.

### 3.3.3 Numerical Simulation

This section describes finite elements analysis (FEA) of LAMPE micromachining to estimate the minimum achievable feature size and compares it with the analytical model (section 3.3.2) and experiment results (section 3.3.1).

For finite element analysis, COMSOL multiphysics software is used [67]. The three-dimensional heat diffusion equation is solved to determine the temperature distribution in the material. Since the optical penetration depth of the laser is small ($14nm$), the source term $Q$ of the heat diffusion equation is set to $Q = 0$, and the laser heating is modeled as a




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

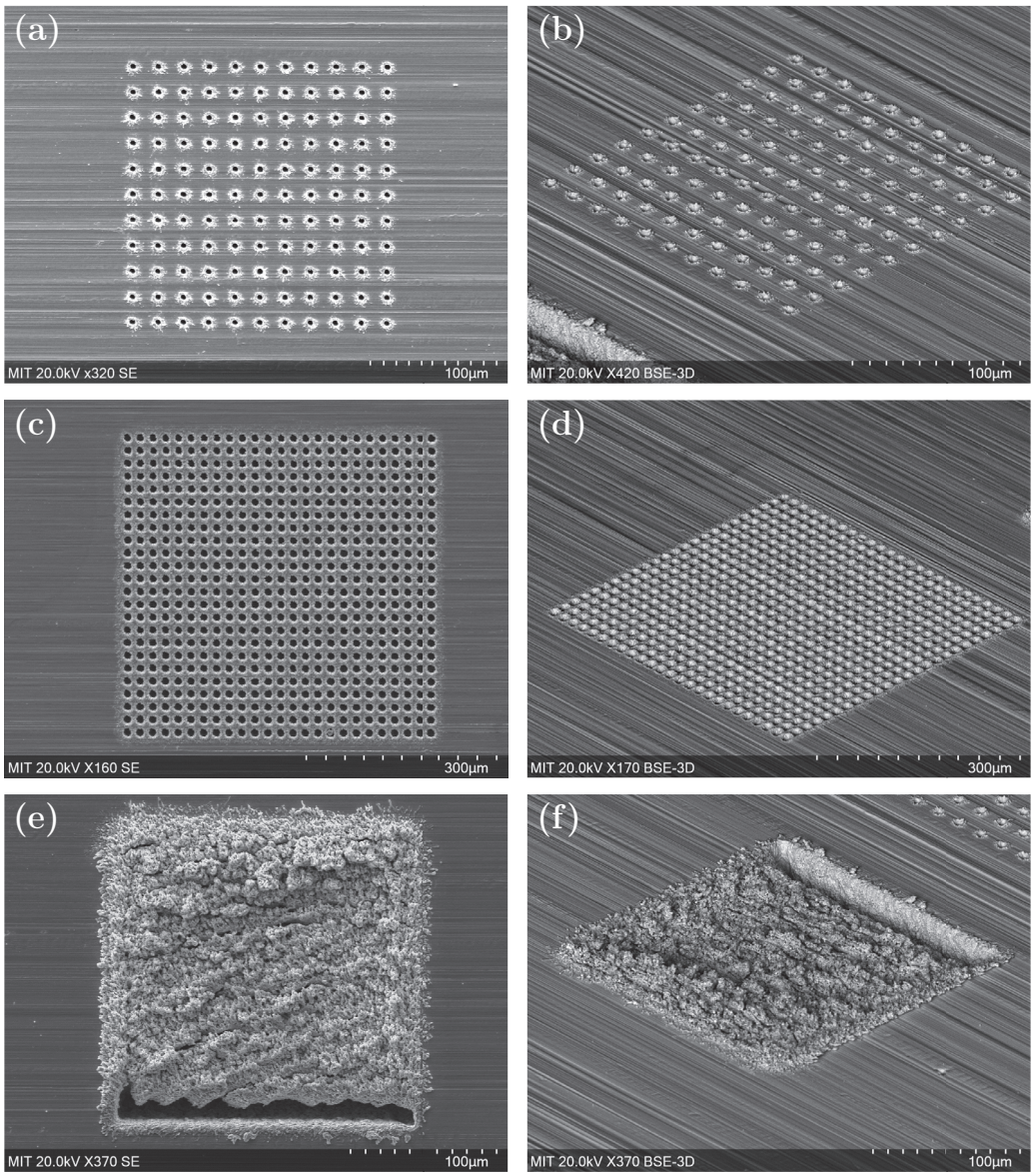



**Figure 3-5:** Pulse exposure with varying pitch distance to confirm material melting & expulsion. (a) & (b) Pitch = $20\mu m$, (c) & (d) Pitch = $10\mu m$, and (e) & (f) Pitch = $5\mu m$.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

surface heat source with Gaussian intensity distribution. Thus the intensity distribution is given by the expression:

$$I(r) = I_0 exp\Big( - \frac{r^2}{\omega_0^2} \Big) \tag{3.8}$$

where maximum intensity, $I_0 = 71.5 GW/cm^2$ and the radius of the focused beam, $\omega_0 = 0.5 \mu m$ (Table 3.1)

The laser-assisted material phase-change and expulsion (LAMPE) is modeled by adding a phase-change (i.e., melting) heat flux thermal boundary condition to the heat diffusion equation and removing the melted material from the modeling domain. The heat flux due to melting can be modeled using the expression [68]:

$$q_m = h_m (T - T_m) \tag{3.9}$$

where $q_m$ is heat flux due to material melting, $T_m$ is the melting temperature and $h_m$ is the temperature-dependent heat transfer coefficient.

The value of $h_m$ is zero below the melting temperature, and increases linearly above the melting temperature. In COMSOL, $h_m$ is modeled as a ramp-function and the slope is set such that the temperature of the material does not significantly increase above the melting temperature. Finally, the material expulsion is modeled by assigning a velocity to the solid boundary. This velocity is given by the expression:

$$v_m = \frac{q_a}{\rho \Delta H_m} \tag{3.10}$$

where $v_m$ is the material melting velocity, $\rho$ is the material density, and $\Delta H_m$ is enthalpy of fusion.

The heat diffusion equation is solved for the duration of the laser pulse ($20 ns$). Figure


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

3-6 shows the LAMPE micromachining profile of copper using a single pulse of laser. The LAMPE process is isotropic, and the minimum achievable feature size is of the order of $5\mu m$. This results agree with the analytical estimates ( 3.3.2) and the experimental findings (3.3.1). The next section describes fabricating high aspect ratio (HAR) microstructures using the LAMPE micromachining process.

## 3.4 LAMPE High-Aspect-Ratio Micromachining

We saw in the previous section that short laser pulses can be utilized to melt, partially evaporate, and expel the material. The analytical model, along with process development, can be used to empirically estimate the pulse duration and threshold pulse energy $E_{LAMPE}$ required for achieving laser-assisted melting and expulsion.

The high aspect ratio machining is achieved by setting the pulse energy to $E_{th}$ to ensure melting and expulsion, and using the laser to gradually melt the vertical face of the laminate (shown in figure 3-1(b), and expel it downwards. The expelled material resolidifies at the rear of the sheet, as shown in figure 3-8. This process is similar to laser-cutting at macro-scale where metal is melted and expelled using the high-pressure air jet. However, in the LAMPE process, the material is expelled using the pressure generated by vapor baubles formed at the laser-material interface due to the partial vaporization of the material.

Figure 3-7 shows the $10\mu m$ wide micro-slights fabricated using LAMPE micromachining on a $100\mu m$ thick copper sheet. The high aspect ratio (HAR) structure is fabricated using the pulse energy of $56\mu J$, which is set by reducing the original pulse energy of $561\mu J$ to its 10% value using an external attenuator. The reflectivity of the copper target, $R = 0.47$; therefore, the effective energy available for material processing is $32\mu J$. The laser beam is focused using a $100mm$ focal length lens and the focused beam spot diameter is $12.98\mu m$. Table 3.2 summarizes the laser conditions used for the LAMPE micromachining process.

In the LAMPE micromachining process, the expelled metal resolidifies at the edges of the micromachined part creating micro-burrs. Figure 3-8 shows the SEM image of the backside




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

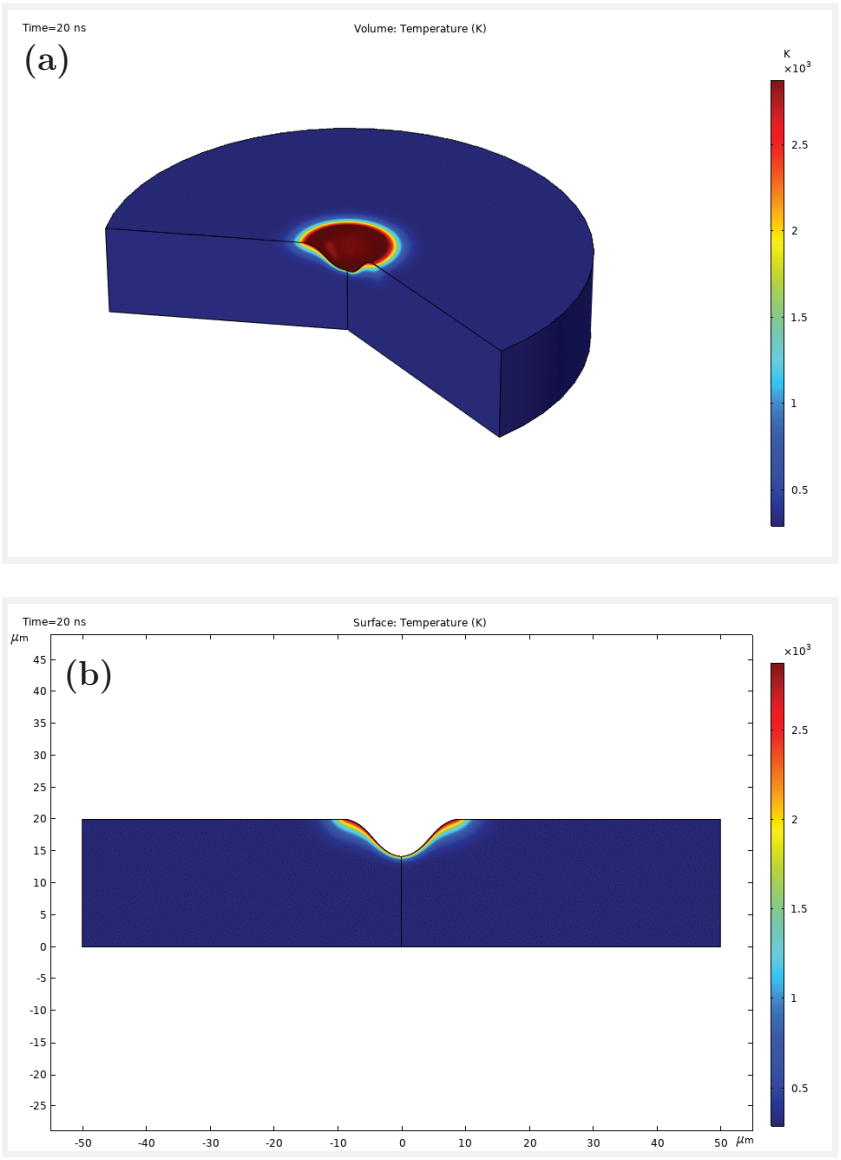



**Figure 3-6:** Simulation of LAMPE micromachining of copper. (a) Isometric view of ablated profile, and (b) Side view of the ablated profile showing crater width and depth.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| Laser Parameter | Value |
|---|---|
| Wavelength, $\lambda$ | $532nm$ |
| Average Power, $P_{avg}$ | $280mW$ |
| Pulse Frequency, $F_p$ | $5KHz$ |
| Pulse Duration, $\tau_l$ | $\approx 20ns$ |
| Reflectivity of Copper, $R$ | 0.43 |
| Pulse Energy, $E_p$ | $32\mu J$ |
| Peak Power, $P_{pk}$ | $2.8kW$ |
| Beam Spot size, $2\omega_0$ | $12.98\mu m$ |
| Rayleigh Length, $Z_R$ | $850\mu m$ |

**Table 3.2:** Laser parameters for LAMPE micromachining high aspect ratio (HAR) features.

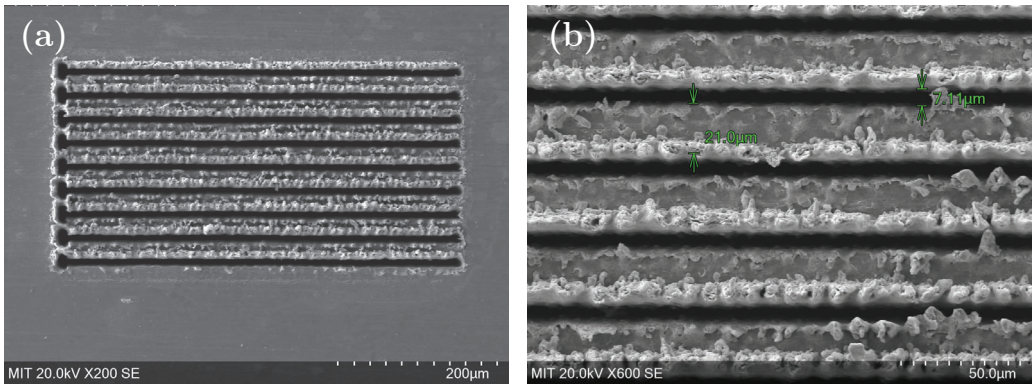



**Figure 3-7:** High-aspect-ratio microfabrication using LAMPE micromachining. (a) $10\mu m$ wide micro-slits fabricated on $100\mu m$ thick copper sheet using LAMPE micromachining. (b) Magnified image showing the dimension of the micro-slit.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

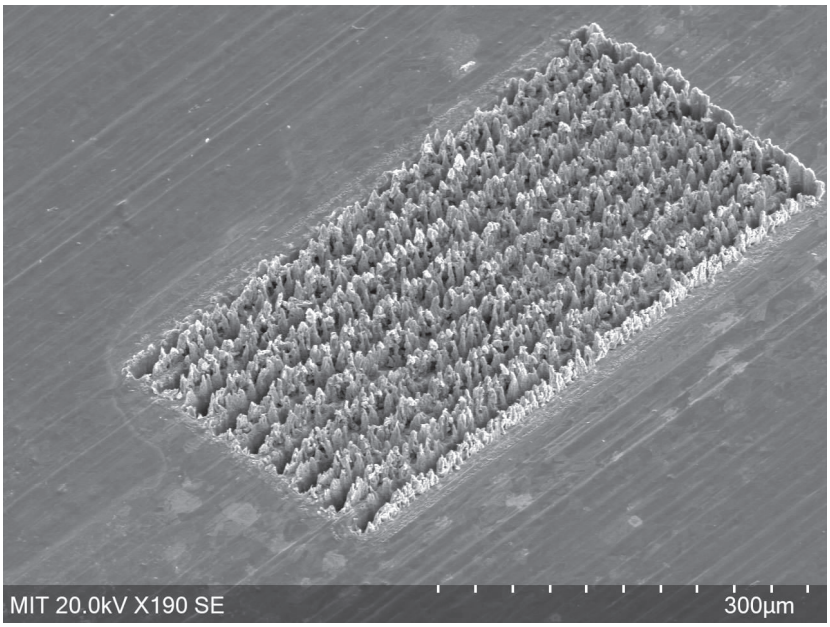



**Figure 3-8:** SEM image of the backside of the LAMPE micromachined copper sheet showing resolidified melt ejects.

of the LAMPE micromachined copper sheet, displaying the micro-burrs. These micro-burrs must be removed from the laser micromachined part to make them suitable for MEMS fabrication. The next section describes the developed electro-deburring process to remove these micro-burrs.

## 3.5 Electro-deburring of LAMPE Micromachined Parts

The LAMPE micromachining process can create small lateral features with high aspect ratio. However, the micromachined parts suffer from resolidified micro-burrs. These burrs pose two difficulties in making MEMS. First, in a typical MEM system, microstructure parts are very close to each other (typically as close as $5\mu m - 10\mu m$), and often held at a different electric potential. For example, in an electrostatic comb-drive actuator, the comb fingers are kept close to each other to increase the electrostatic force of attraction and a bias voltage is applied between the fingers. The protruding burrs can create electric contact between these




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

fingers. Second, the sharp metal protrusion can generate field-emitted electrons resulting in the dielectric breakdown of air, thereby causing damage to the MEM system [69].

Removing the burrs in MEMS microstructures is challenging due to two reasons. First, macro-scale deburring processes such as physical scrub and abrasive jet can not be utilized for MEMS structures due to their small size and delicate nature. Second, at macro-scale, etching process can etch the burrs without significantly affecting the dimensions of the parts. However, in MEMS parts, the burrs and the features are of the same order, and these methods result in significant etching of the parts.

This thesis develops a procedure to remove micro-burrs without affecting the MEMS microstructures. The process setup is similar to electropolishing process and utilizes the high-voltage regime. Figure 3-9 shows the polarization curve for copper electropolishing in phosphoric acid. As can be seen in figure 3-9, the current first increases linearly followed by a plateau and then increase linearly again. In the plateau region and beyond, the current density is uniform across the surface of the part due to the saturation of mass transport species [70], resulting in saturation of etch rate.

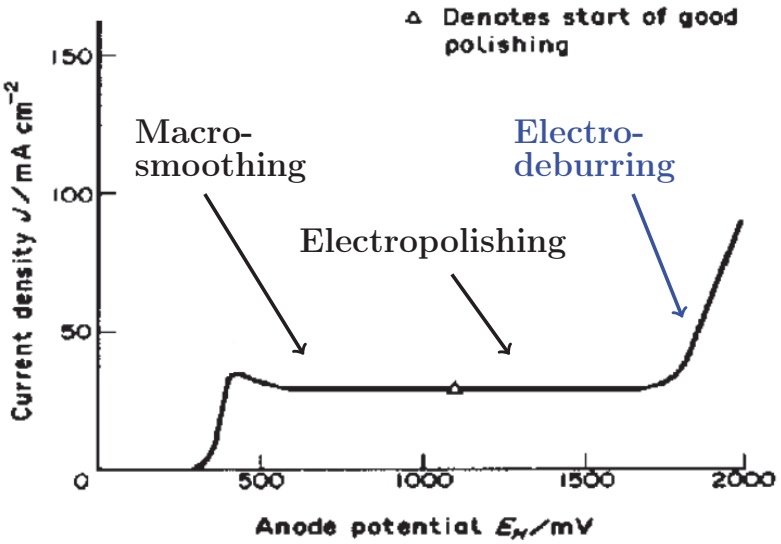



**Figure 3-9:** Polarization curve of copper in phosphoric acid [70]




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

The burrs produced in LAMPE micromachining process have two interesting properties; first, these burrs have sharp features and large surface area, and second, they are located at the edges of the micromachined parts. The current density is higher near the burrs due to two reasons: first, the burrs have sharp features, and second, they are located at the edges where the current density is higher. As a result, the electric field and the available mass transport spices are high at the burrs, resulting in a high etch rate. When the cell voltage is increased beyond the electropolishing voltage, the etch rate at the burr increases significantly; however, the etch rate at the other area remains the same due to saturation in the etch rate. Using this phenomenon, the burrs can be selectively etched without significantly affecting the dimensions of the micro-parts.

Figure 3-10 shows performing electro-deburring of copper in $14M$ phosphoric acid at $5V$ for 30 seconds. For certain application, electro-deburring can be followed by electroplating of gold to avoid surface oxidation.

Figure 3-11 shows the high aspect ratio interdigitated fingers fabricated using the LAMPE micromachining. Such small features and high aspect ratio were previously achievable only using deep reactive ion etching (DRIE).

It must be noted that for MEMS that does not require small lateral-features and high aspect ratio, traditional pulsed laser ablation (PLA) micromachining in conjunction with the electro-deburring process can be used for MALL MEMS fabrication.

## 3.6 LAMPE Micromachining of Silicon

### 3.6.1 Introduction

Silicon is widely used as a mechanical material for fabricating MEM systems [71]. Conventionally deep reactive ion etching (DRIE) of silicon is used to fabricate MEMS [72]. However, DRIE tools are expensive and require expensive etching gases. The LAMPE micromachining is an attractive alternative for manufacturing high aspect ratio (HAR) microstructure silicon laminates. Moreover, these LAMPE micromachined silicon




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

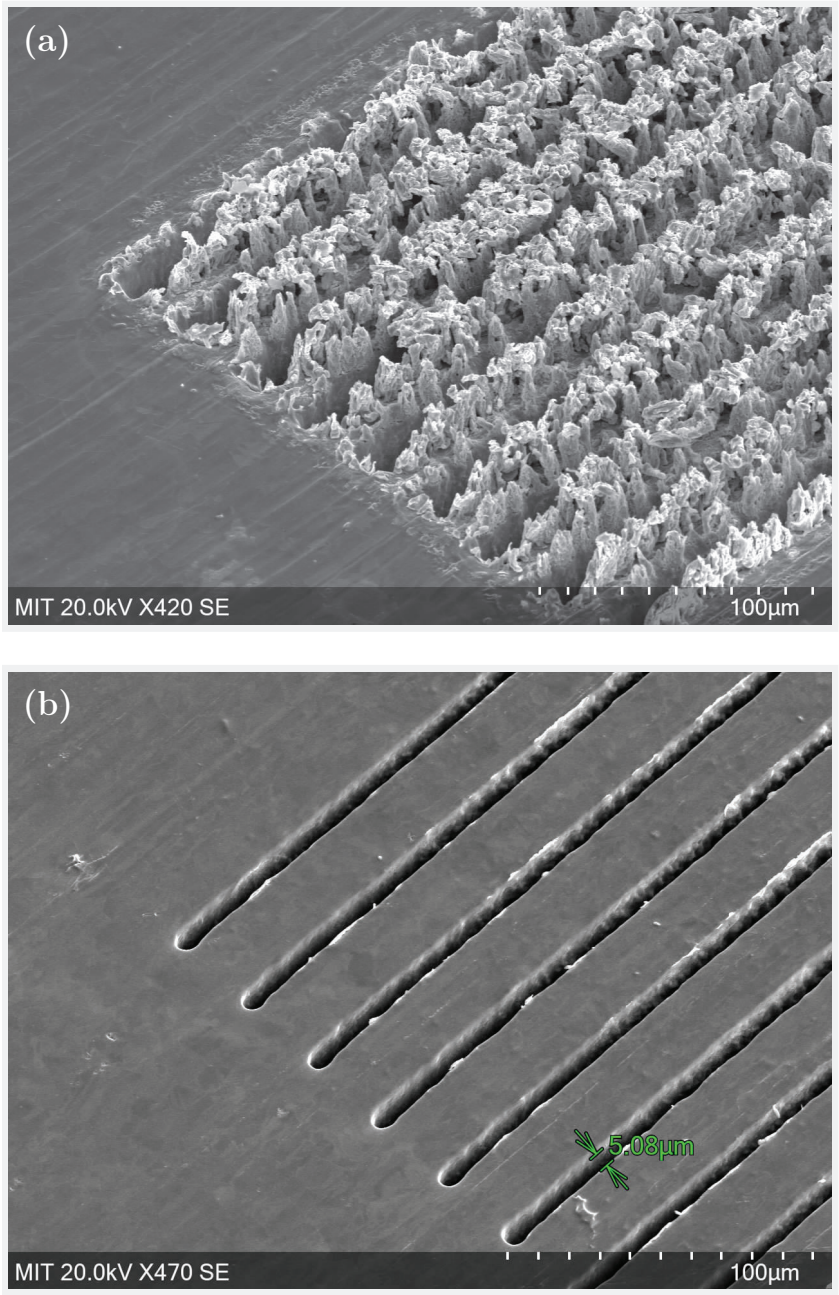



**Figure 3-10:** Electro-deburring of LAMPE micromachined parts to remove micro burrs. (a) Before electro-deburring. (b) After electro-deburring.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

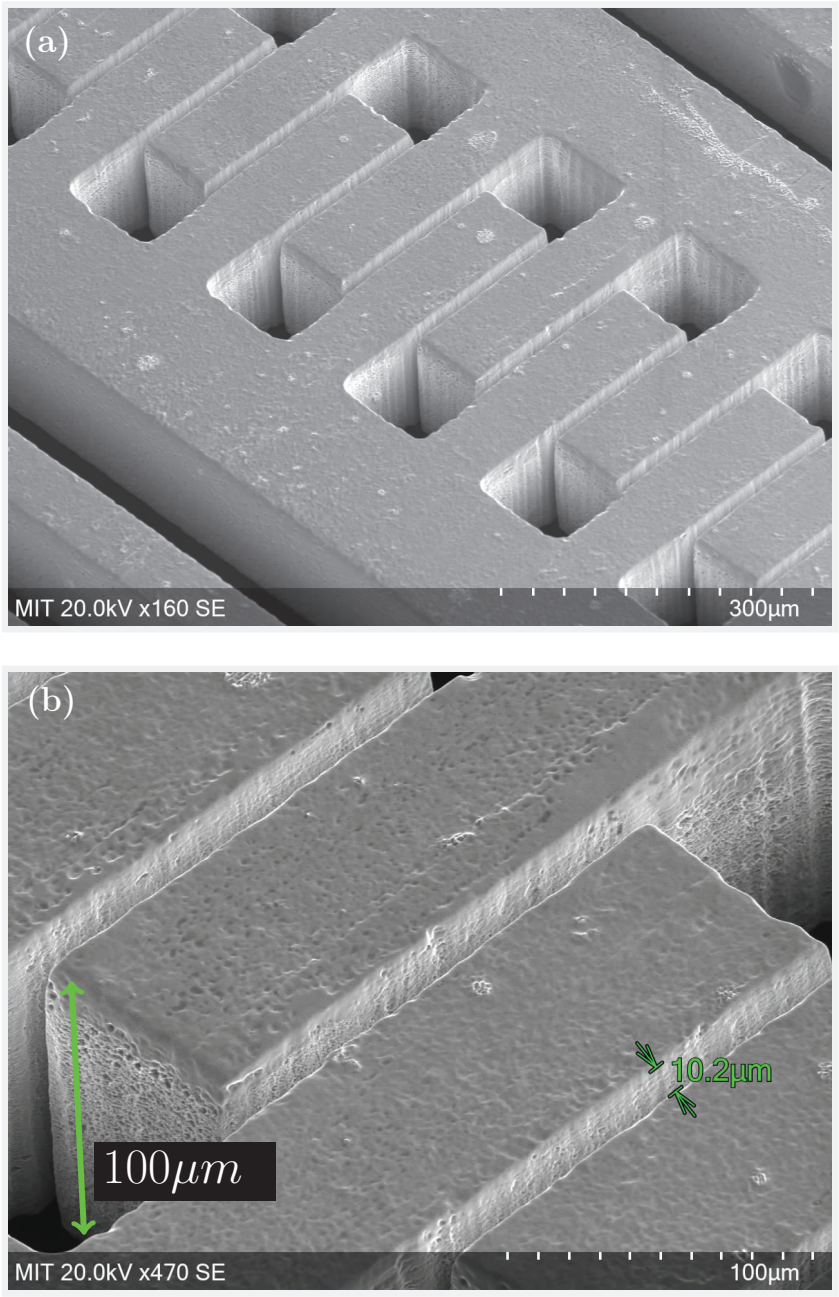



**Figure 3-11:** High aspect ratio interdigitated fingers fabricated using the LAMPE micromachining. (a) Low-magnification image showing the interdigitated fingers. (b) Magnified image of the fingers showing 10:1 aspect ratio. The lateral feature size is $10\mu m$.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

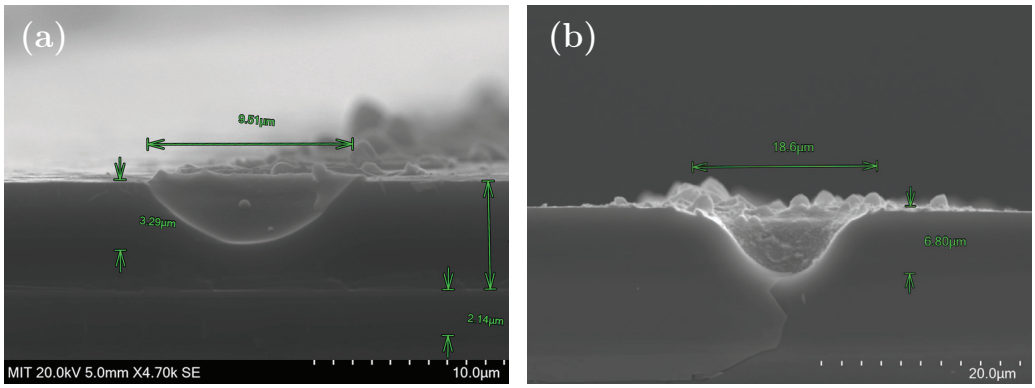



**Figure 3-12:** Pulsed laser ablation (PLA) of silicon using $20ns$ long laser pulses. (a) Ablation profile created by a $5.6\mu J$ pulse. The diameter of the crater is $9.51\mu m$ and it is $3.29\mu m$ deep. (b) Ablation profile created by a $56\mu J$ pulse. The diameter of the crater is $18.6\mu m$ and the depth is $6.8\mu m$.

laminates in conjunction with the MALL process, can be used to make MEM systems.

The absorption coefficient of silicon at $532nm$ is $\alpha_{Si} = 9 \times 10^3 cm^{-1}$ [55]; therefore, the optical penetration depth is $l_\alpha = 1\mu m$. Due to the shallow penetration depth, the laser can be modeled as a surface heat source, and the temperature distribution can be determined by solving the heat diffusion equation. Figure 3-12 shows the finite element analysis (FEA) of heat distribution in the pulsed laser ablation process.

When the temperature reaches the boiling point of the silicon, the material is vaporized, forming an ablation creator. Figure 3-12 (a) & (b) show the ablation profile created by a single laser pulse of energy, $5.6\mu J$ and $56\mu J$, respectively. The ablation profile shown in figure 3-12 agrees with the features reported in literature [73][74][75][76][77][78][79]. The aspect ratio is limited due to small optical penetration depth and small heat diffusion length. Although multi-pulse ablation can be used to increase the aspect ratio, the saturation depth is quickly reached because of the inefficient removal of material due to narrow space and the absorption of the laser by vapor plasma. Due to the large lateral features and low aspect ratio, the application of laser-micromachining silicon has been largely limited to drilling holes and wafer dicing.

Karnakis et. al. [80] reported micromachining silicon with small lateral features in the




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

order of $10 - 20\mu m$ and high aspect ratio exceeding the optical penetration depth. The high aspect ratio has been attributed to a homogeneous bubble nucleation phase explosion mechanism [80][81]. This thesis shows further reducing the lateral features down to diffraction-limited spot size and increase in aspect ratio using the LAMPE micromachining process. In the LAMPE process, first, the silicon is converted into silicon oxide using laser (i.e. induce phase-change), and next, the silicon oxide is expelled from the rear of the wafer using the pressure generated from vapor baubles formed at the laser-silicon interface due to partial vaporization of silicon.

When silicon is irradiated with low power laser in the oxygen environment, it converts to silicon oxide without resulting in any ablation. The absorption coefficient of $SiO_2$ is $\alpha < 1cm^{-1}$ [55]; therefore, the optical penetration depth is $l_\alpha > 1cm$. Since the absorption coefficient of silicon oxide is low, the newly formed silicon oxide transmits the large percentage of the laser energy, which further oxidizes the silicon beneath. In this manner, a high aspect ratio structure much longer than the optical penetration depth of silicon can be oxidized. Figure 3-13 shows the laser-assisted oxidation of silicon. As can be seen, the oxidation length far exceeds the optical penetration depth. The oxidation of the silicon can be verified by using an optical microscope and energy-dispersive x-ray spectroscopy (EDS) image as shown in figure 3-14

The successive oxidation of silicon takes place until the bottom of the wafer is reached. The oxidized silicon is in granular form and loosely bound to the silicon wafer. As a result, the small recoil pressure generated by partially vaporized silicon is sufficient to expel the granular silicon oxide from the bottom of the wafer, resulting in micromachining.

The maximum achievable aspect ratio is limited by two factors. First, the oxidized silicon exists in granular form and the effective absorption coefficient of granular oxide is larger than bulk silicon oxide due to light scattering. Second, the silicon oxide is expelled only when the oxidation is reached to the bottom of the wafer. As a result, for thick wafers, the attenuation could be significant to stop oxidation before the rear of the wafer is reached, thereby resulting in no micromachining. The aspect ratio in LAMPE micromachining could




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

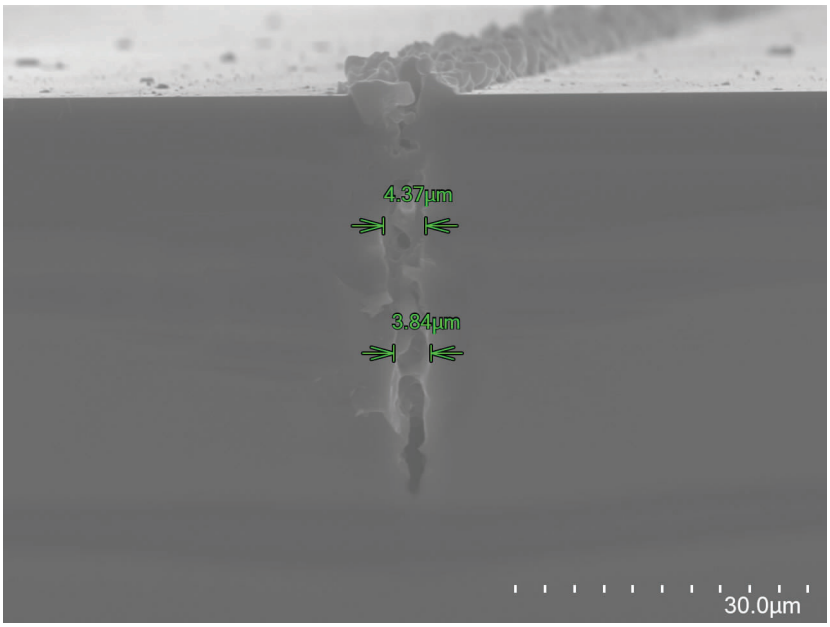



**Figure 3-13:** The cross-section image showing laser assisted oxidation of silicon. As can be seen, the laser assisted oxidation can significantly increase the "effective optical penetration depth".

be further increased by performing the micromachining in HF vapor. The HF vapor can facilitate the selective removal of silicon oxide by forming gaseous $SiF_4$.

### 3.6.2 Numerical Simulation

The time evolution of the increase in "effective optical penetration depth" due to oxidation of silicon can be simulated using finite element analysis. The intensity of the laser beam is reduced to 10% of the intensity using an external attenuator, and the reduced peak intensity is $I_0 = 4.2GW$ (see figure 2-5). The radius of the focused beam is $\omega_0 = 6.48\mu m$. The Gaussian intensity profile of the beam is given by the equation:

$$I(r) = (1 - R)4.2GWexp\left( - \frac{2r^2}{6.48\mu m^2} \right) \tag{3.11}$$

Beer-Lambert law is used to model the absorption of the laser. The phase-change from silicon to silicon oxide is modeled by changing the absorption coefficient of the material




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

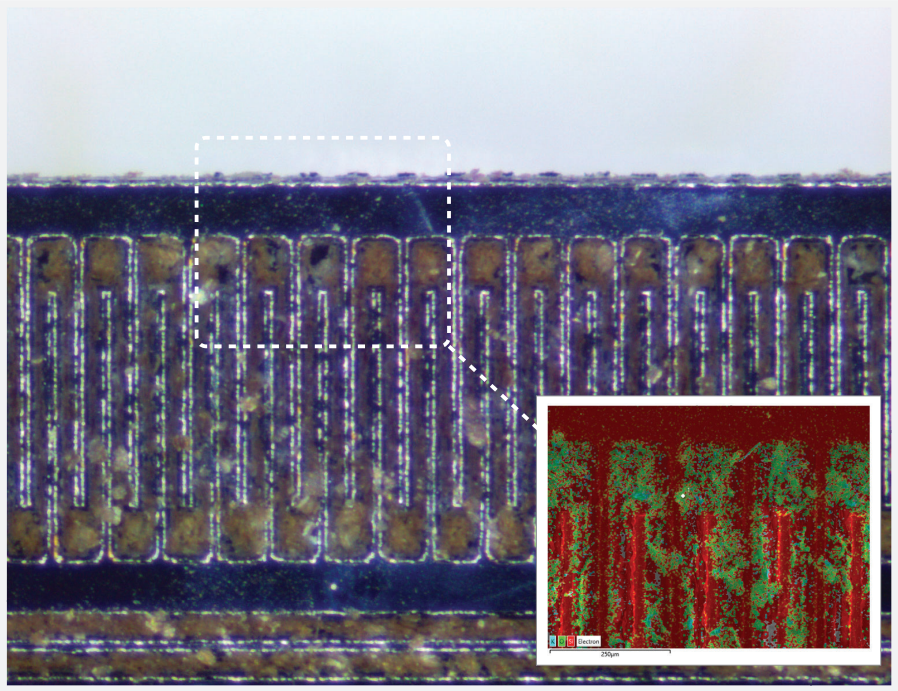


**Figure 3-14:** The optical microscope image confirming the oxidized silicon. The energy dispersive x-ray spectroscopy (EDS) image shown in the inset further confirms the present of silicon oxide.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

after phase-change and by introducing a heat flux due to oxidation. Figure 3-15(a) shows the temperature profile in silicon during conventional pulse ablation. Figure 3-15(b) shows the distribution of the temperature in silicon when silicon oxidation and change in absorption coefficient is taken into account. It can be seen in figure 3-15(b) that the oxidation of silicon significantly increases the "effective optical penetration depth". Note that the expulsion of the material is not simulated due limitation of the FEA package.

### 3.6.3 Results & Discussion

Figure 3-16 shows the comb micro-structure fabricated using LAMPE micromachining of silicon. The LAMPE micromachined silicon structures reported in this thesis are comparable to the structures fabricated using deep reactive ion etching (DRIE) [82].

The laser fluence required for the LAMPE micromachining process is much lower than the pulse ablation micromachining. As a result, the described process significantly reduces the power requirement of laser source and could be an attractive method to prototype silicon MEM systems. It must be noted that since the material is expelled from the rear of the substrate, the LAMPE micromachining is limited to fabricating lamina microstructures.

#### LAMPE Micromachining Rate

The machining rate in LAMPE micromachining process is given by the expression:

$$LAMPE\ Micromachining\ Rate = \frac{\eta \omega_0 t f}{N} \tag{3.12}$$

where $\eta = distance\ between\ the\ neighboring\ irradiation\ spot/laser\ beam\ diamter$ and generally set to $\eta = 0.25$, $\omega_0$ is the diameter of focused laser beam, $t$ is the thickness of the wafer, $f$ is the micromachining feed rate, and $N$ is the number of passes required to completely remove the material. For the LAMPE micromachining of silicon described above, the wafer thickness is $t = 50\mu m$, and the laser beam spot diameter is $\omega_0 = 10\mu m$, and it is empirically determined that $\eta = 0.25$, $f = 60\ mm/min$, and $N = 10$. Substituting these




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

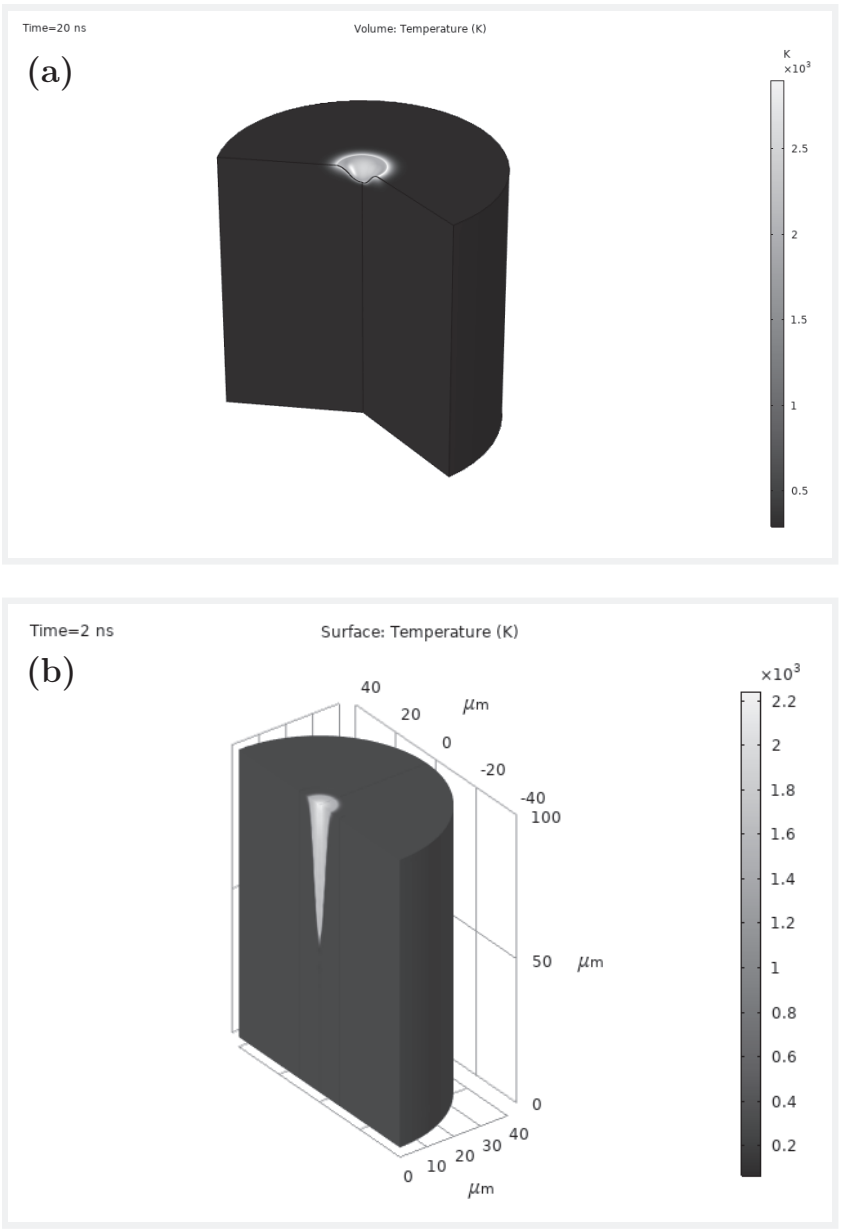



**Figure 3-15:** (a) Simulation of pulsed laser ablation of silicon, and (b) Simulation of LAMPE micromachining of silicon.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

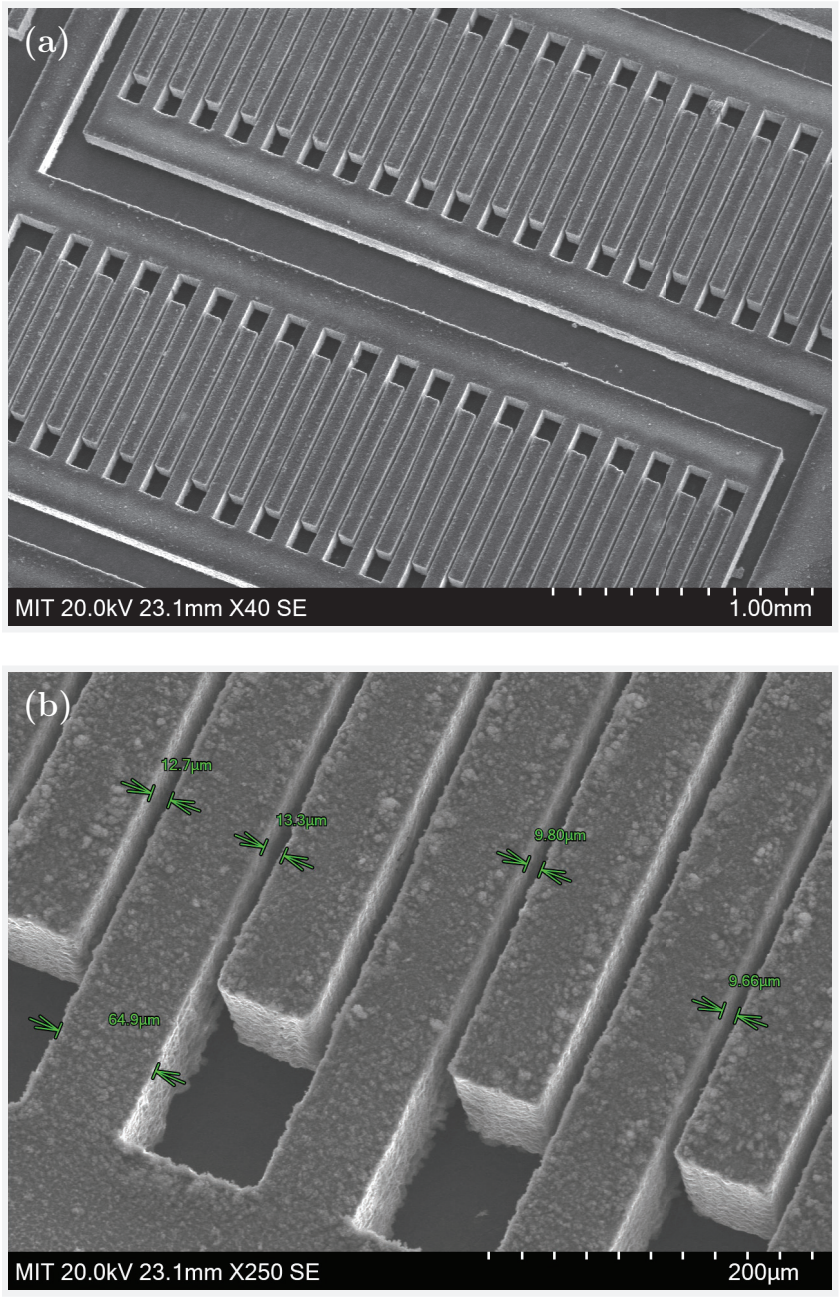



**Figure 3-16:** Examples of high-aspect-ratio microstructures fabricated using LAMPE micromachining of silicon. (a) Low-magnification image showing fabricated interdigitated finger structures. (b) High-magnification image showing the distance between the fingers.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

values in (3.12), we get $LAMPE\ Micromachining\ Rate = 7.5 \times 10^5 \mu m^3/min.$

## 3.7 Conclusions

This chapter concludes that laser micromachining, utilizing the LAMPE process, can fabricate high aspect ratio structures with lateral features as small as the diffraction-limited spot size of the laser. Previously, fabricating such high aspect ratio structures with small lateral features was not possible using any fabrication process (including the pulse laser ablation micromachining).

In the LAMPE process, the laser is used to melt and shift the material, and then the shifted material is removed utilizing the electro-deburring process. Using this process, high aspect ratio structures are fabricated by utilizing the laser to melt the vertical face of the sheet and expelling it downward (shown in figure 3-1(b)), and then removing the expelled material using electro-deburring. Using the LAMPE process, I have shown fabricating 10:1 aspect ratio structures with lateral feature size of $10\mu m$.

The LAMPE micromachining process can also be used to fabricate high aspect ratio (10:1) structure in silicon with lateral features as small as $10\mu m$. In the LAMPE micromachining of silicon, the laser is used to turn silicon into granular silicon oxide. The silicon oxide is then removed by the pressure generated by vapor bubbles formed at the laser-silicon interface. It must be noted that no deburring is required for silicon as the oxide is in the powder phase and gets expelled from the silicon substrate. Although similar features can be produced using deep reactive ion etching (DRIE), the LAMPE micromachining of silicon enables low-cost fabrication and rapid prototyping.

The LAMPE process can be implemented by using the appropriate laser power and pulse duration. The appropriate laser power required for LAMPE micromachining can be calculated using the procedure described in section 3.3.2. Similarly, the laser pulse duration can be estimated using the eq. (3.5), for a desired lateral-feature size. However, it must be noted that some degree of process development is required to achieve the minimum feature




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

for a given aspect ratio.

The main advantage of the LAMPE process is that it enables high aspect ratio machining that was not achievable previously. Moreover, the laser power required for LAMPE micromachining is less than the laser power required for pulsed laser ablation micromachining. The main disadvantage of the LAMPE process is that it only allows micromachining material laminates. In other words, the LAMPE process can not be used to fabricate 2.5D structures.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 4

# MALL MEMS Fabrication

The MALL MEMS fabrication process consists of two steps. First, the individual layers of the MEMS are fabricated using laser micromachining. Second, these layers are stack assembled and bonded to construct the MEM systems. In the previous chapter, I showed fabricating microstructure laminates using the LAMPE micromachining process. In this chapter, I will show assembling these microstructure layers to fabricate the MEM systems. Further, I will demonstrate fabricating a comb-drive actuator and a MEM relay using the MALL fabrication method. Section 4.1 describes the multi-lamina assembly process used to align and bond the laminate layers.

Section 4.2 describes fabricating a comb-drive actuator using copper as a structural material. The comb-drive actuator is selected because the interdigitated comb finger structure is an essential building block in many MEM systems such as optical shutters [21], micro-grippers [83], microengines [84], accelerometer [22], resonators [23], and electromechanical filters [24]. The ability to fabricate interdigitated comb finger structures using copper can significantly reduce the cost of these MEM systems. Moreover, fabricating comb fingers is challenging because it requires high aspect ratio structures with small lateral features. Previously, comb fingers with comparable features and aspect ratio could be fabricated from silicon alone using the lithography-based deep reactive ion etching (DRIE) process [28]. However, in this thesis, I will demonstrate fabricating these




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

structures using the LAMPE micromachining process.

Section 4.3 describes fabricating a MEM relay using the MALL fabrication process. These MEM relays could be used for high-current switching applications. Finally, section 4.4 shows fabricating a high-aspect-ratio diamond rotor for enhancing the resolution of magic-angle spinning nuclear magnetic resonance spectroscopy (MAS-NMR).

## 4.1 Multi-Lamina Assembly

The second step in the MALL process is to stack assemble the LAMPE micromachined laminates to fabricate MEM systems. For many MEMS, the alignment accuracy better than a few micrometers is desirable. Mechanical alignment of laminates using dowel pins and alignment holes is widely used in precision macro-fabrication. This method is employed to align and bond microstructure laminates. The alignment accuracy exceeding $2.5\mu m$ is demonstrated using the method.

Figure 1-4 shows the schematic diagram of the alignment-pin and alignment-hole assembly process. The precise hand-assembly is carried out under an eyepiece-less stereo microscope with long working distance and large depth of focus [85]. The microscope uses multi-lenticular technology to provide true depth perception, which aids in hand-to-eye coordination; as a result, facilitating the micro-assembly process.

Four precision-ground alignment pins with $508\mu m$ diameter are used for alignment. The clearance between the alignment pin and alignment hole determines the alignment accuracy and must be minimized. For a given alignment pin, first, a series of holes is micromachined (figure 4-1(a)). Subsequently, the alignment pins are inserted and the clearance between the pin and the hole is measured (figure 4-1(b)). Finally, laser micromachining setting for the hole that gave the minimum clearance (figure 4-1(c)) is used to micromachine alignment holes.

To characterize the alignment accuracy, two layers with alignment marks are assembled and bonded. The bonded layers are illuminated from the rear, and the alignment marks




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

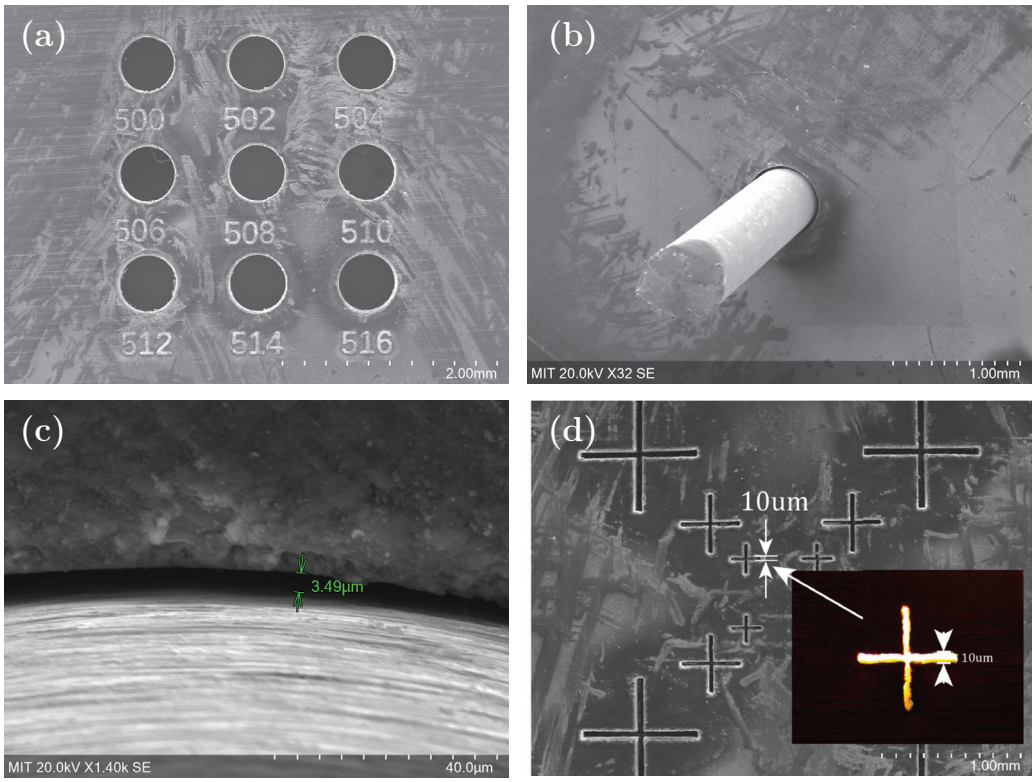



**Figure 4-1:** (a) Micro-assembly of discrete lamina sheets using alignment pins. (b) Alignment pin inserted inside the alignment hole. (c) The clearance between the alignment pin and the alignment-hole. (d) SEM image of aligned alignment marks, showing alignment accuracy down to $2\mu m$ . The inset shows the black-lit optical image of an alignment of marks.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

are studied using an optical microscope. The inset of figure 4-1(d) shows back-illuminated light transmitting through the two aligned marks. The width of the slit is $10\mu m$, and the overlap between the two aligned markers is better than 75%. As a result, it can be concluded that the alignment accuracy is better than $2.5\mu m$. The alignment accuracy can be further improved by creating passive alignment features in laminates [86][87][88].

The aligned layers are bonded using Dupont FR1500 and 3M thermal bonding film 583 adhesive sheets. For adhesion, the Dupont FR1500 is cured at $185°C$ for 5-20 minutes, depending upon the device design. Similarly, the 3M thermal bonding film 583 can be bonded by using pressure, applying solvent or heating at $120°C$. In some devices, it is desirable to first tack bond the adhesive sheet to one layer, and then attach another layer and perform complete cure for lamination. For tack bonding, the FR1500 sheet is cured at $120°C$ for 10 minutes, and 3M thermal bonding film 583 is pressure or solvent activated.

## 4.2 Comb-Drive Actuator

### 4.2.1 Design

Figure 4-2 shows the design of a comb-drive actuator. It consists of two sets of comb fingers. The width of each finger is $30\mu m$, and the length is $150\mu m$. The distance between the two sets of fingers is $10\mu m$. One of the finger structure is fixed while the other finger structure is connected to a compliant suspension spring. The suspension spring uses a folder-flexure design to minimize the side-pull between the fingers and maximize the lateral displacement [89]. The length of the beam is $1.66mm$, and the width is $20\mu m$. The thickness of the fingers and suspension spring is $100\mu m$. Lastly, the device is made of copper, and a $100\mu m$ thick aluminum oxide ceramic sheet is used as the insulating substrate. Table 4.1 lists the design parameters of the comb-drive actuator.

The spring constant of the folded-flexure design in the lateral direction is given by the




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

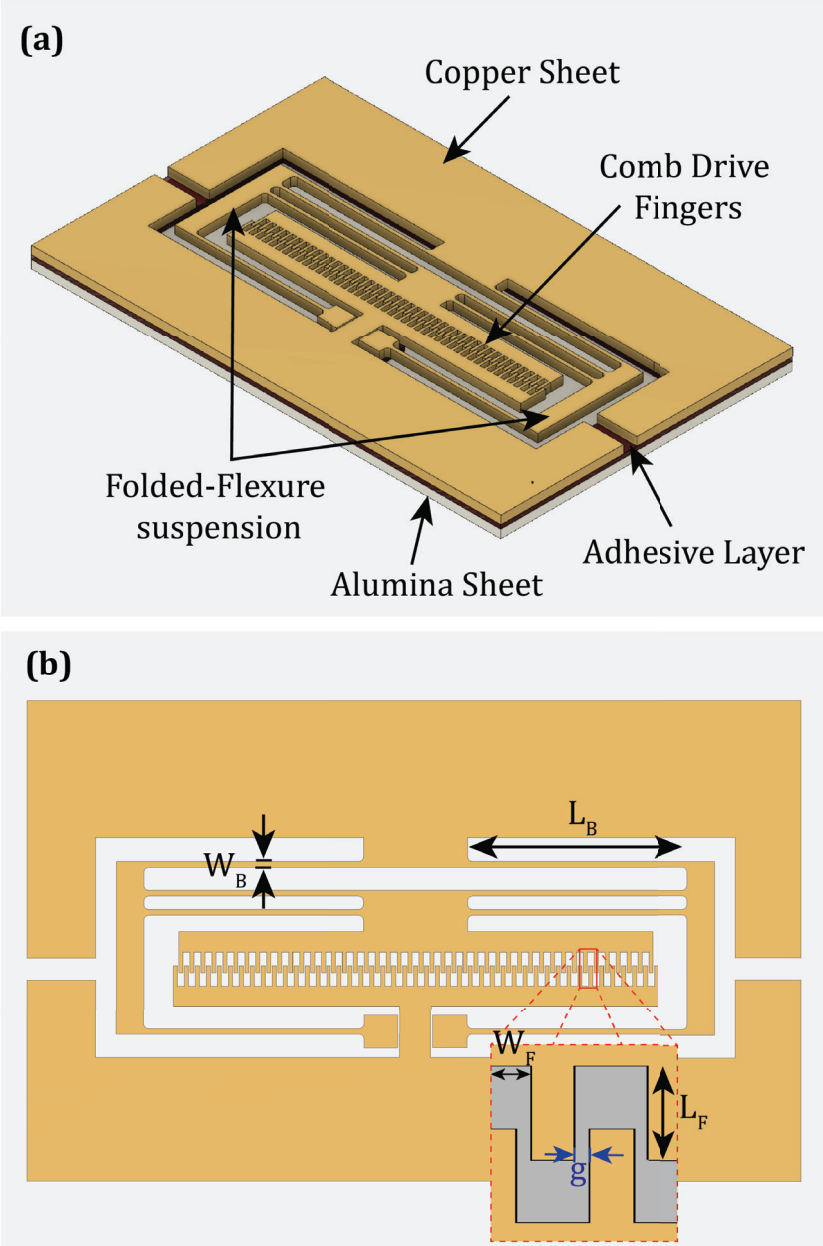



**Figure 4-2:** (a) Schematic diagram of the comb-drive actuator. (b) Diagram showing design parameters . These parameters are listed in Table 4.1.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| **Design Parameters** | **Value** |
|---|---|
| Young's Modulus, $E$ | $128GPa$ (Copper) |
| Beam Thickness, $h$ | $100\mu m$ |
| Beam Length, $L$ | $1.66mm$ |
| Beam Width, $b$ | $20\mu m$ |
| Spring Constant, $k$ | $44.8N/m$ |
| Comb Finger Length, $L_F$ | $150\mu m$ |
| Comb Finger Width, $W_F$ | $30\mu m$ |
| Comb Finger Gap, $g$ | $10\mu m$ |
| Number of Fingers, $n$ | 46 |
| Applied voltage and corresponding displacement | @280V, d = $7\mu m$ |

**Table 4.1:** Values of the design parameters for comb-drive actuator design.

expression [89]:

$$k = \frac{2Ehb^3}{L^3} \tag{4.1}$$

where $E$ is Young's modulus, $h$ is the thickness of beam spring, $b$ is the width of the beam spring, and $L$ is the length of the beam spring. Substituting the values of $E$, $h$, $b$ and $L$ from table 4.1, we get the spring constant of the folded-flexure spring as $44.8N/m$.

The electrostatic tangential force, $F$ between the fingers is given by the expression [89]:

$$F = \frac{n\epsilon_0 hV^2}{g} \tag{4.2}$$

At equilibrium, the electrostatic force is balanced by the spring force; as a result, the equilibrium displacement at voltage $V$ is given by the expression:

$$y = \frac{n\epsilon_0 hV^2}{kg} \tag{4.3}$$

The calculated value of displacement at $V = 280V$ is $7\mu m$.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

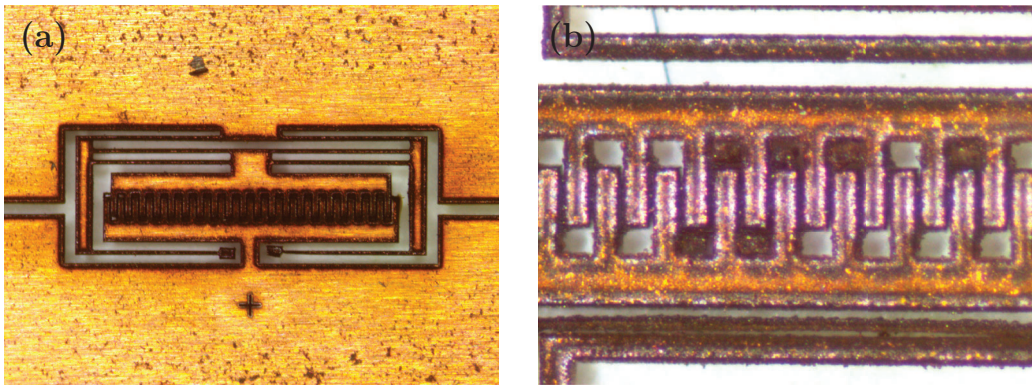



**Figure 4-3:** Fabrication of comb-drive device layer using LAMPE micromachining of copper: (a) Optical microscope image showing micro-burrs present in LAMPE micromachined part. (b) Magnified image showing micro-burrs present between the fingers. These burrs result in short-circuits between the two sets of fingers.

### 4.2.2 Fabrication

The fabrication of the comb-drive actuator consists of two steps. The first step is to fabricate the individual layer of the device using the LAMPE micromachining process. Figure 4-3(a) shows the optical microscope image of the LAMPE micromachined copper sheet containing the comb-drive device structure. As can be seen in figure 4-3(a), the micromachined part suffers from carbon deposits and micro-burrs. The micro-burrs result in electrical contact between the two sets of comb fingers (4-3(b)) and must be removed.

The micro-burrs are removed using the electro-deburring process described in section 3.5. The electro-deburring process is carried out in phosphoric acid, and the LAMPE micromachined part is used as the anode while another copper sheet is used as a cathode. The electro-deburring is carried out at $5V$ for 30 seconds. Figure 4-4(b) shows the SEM image of the comb-drive structure after removing the micro-burrs.

A $100\mu m$ thick aluminum oxide sheet is micromachined and used as an insulating substrate. Prior to micromachining the aluminum oxide substrate, a $50\mu m$ thick stage-B epoxy adhesive film is bonded to it. This adhesive layer is later used to bond the copper device layer with the aluminum oxide substrate layer. Figure 4-5 shows fabricated layers of the comb-drive actuator. For every layer,alignment holes are machined, in addition to the




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

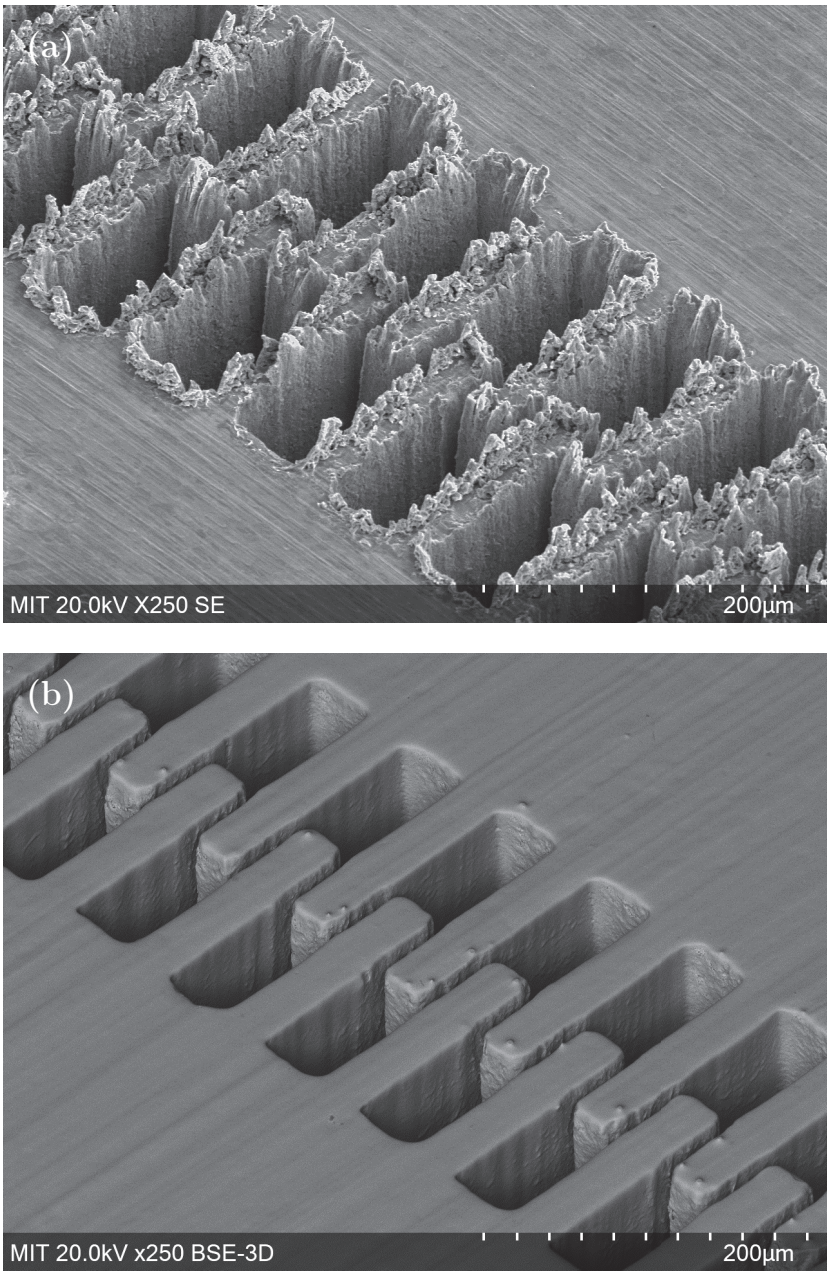



**Figure 4-4:** Electro-deburring of LAMPE micromachined comb-drive structure to remove the micro-burrs between the fingers. (a) Before electro-deburring. (b) After electro-deburring showing clean comb-drive fingers.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

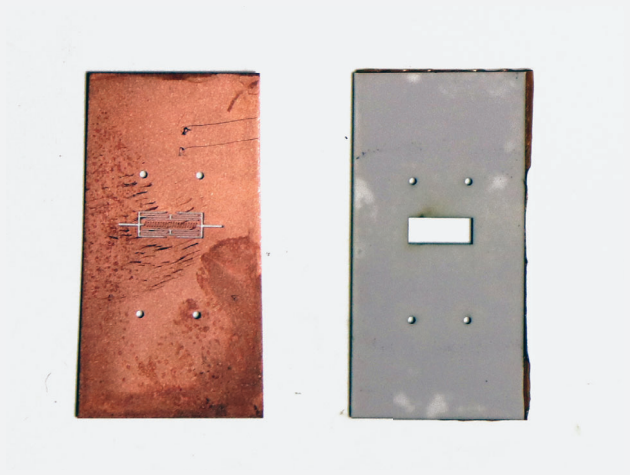


**Figure 4-5:** Layers of the comb-drive actuator fabricated using LAMPE micromachining. The copper layer is used for the device, and the aluminum oxide layer is used for the substrate layer.

device structure. These alignment holes are subsequently used for multi-lamina assembly.

The second step in the MALL MEMS fabrication method is the multi-lamina assembly of laser-micromachined laminates. This assembly process is described in section 4.1. Once the layers are bonded (figure 4-6(a)), the device is cut out from the laminate structure using laser cutting (figure 4-6(b))

### 4.2.3 Results & Discussion

Figure 4-7 shows the fabricated comb-drive actuator. The gap between the comb-drive fingers is $10\mu m$, and the thickness of the beam spring is $20\mu m$ (figure 4-8). The gap between the fingers can be further reduced by reducing the thickness of the copper layer.

For measuring the displacement of the comb-drive actuator, a bias voltage is applied between the two sets of comb-drive fingers and the displacement is observed with an optical microscope. As the bias voltage is gradually increased, small movement in the fingers is observed. At approximately $280V$, the two sets of fingers make electrical connections, resulting in oscillation (the oscillation is observed because when the fingers make electrical contact, the electrostatic force is inactive and fingers are pulled apart by


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

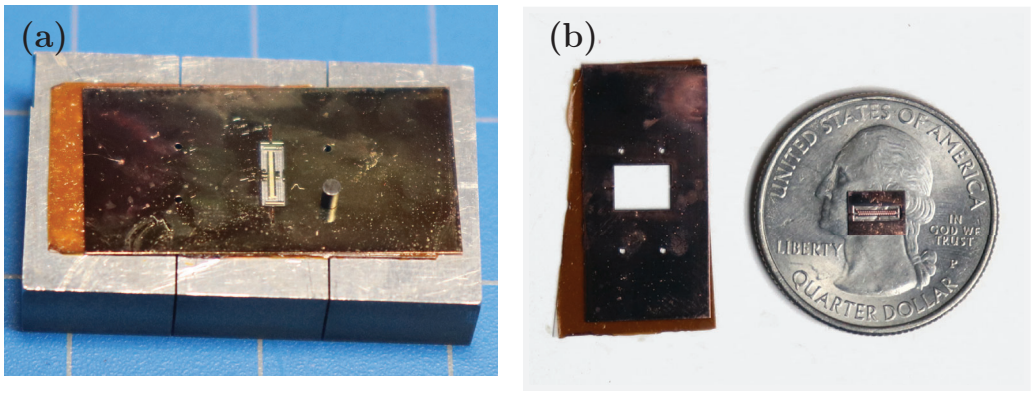



**Figure 4-6:** (a) Assembled and bonded layers of comb-drive actuator. (b) The device cut-out from the laminated layers using laser cutting.

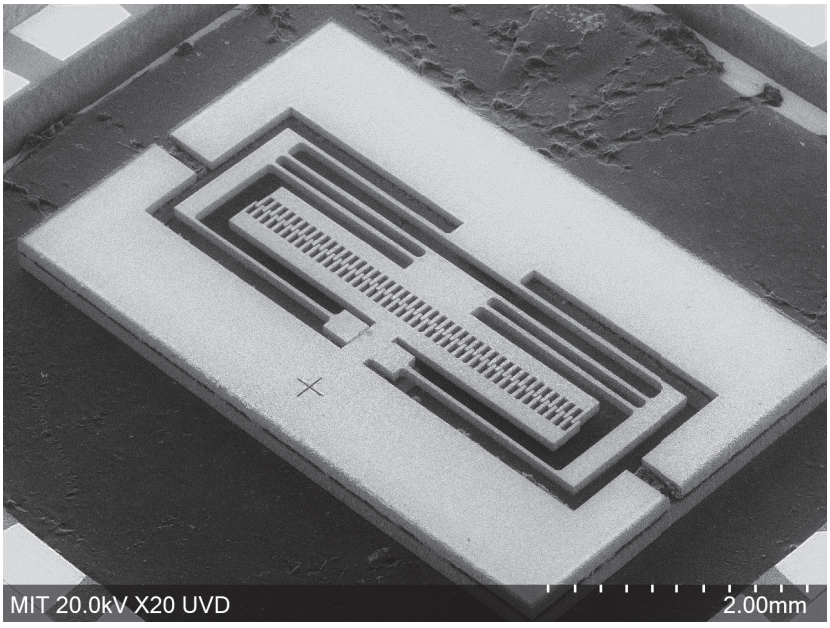



**Figure 4-7:** SEM image of the fabricated comb-drive actuator.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

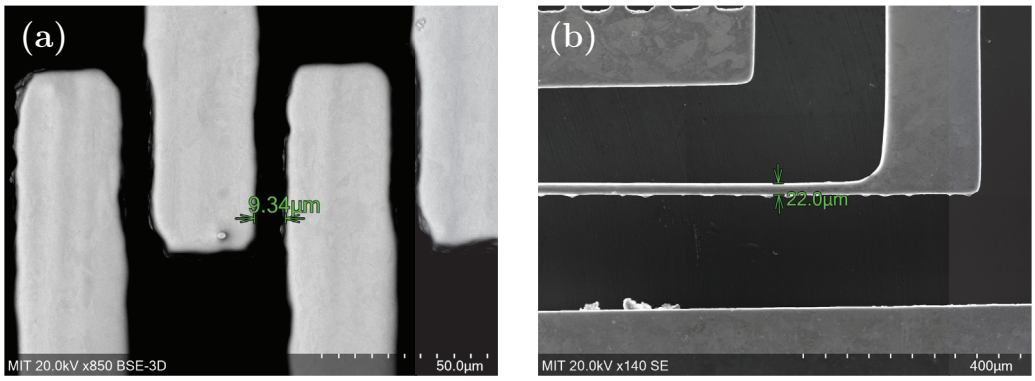



**Figure 4-8:** Features of the fabricated comb-drive actuator. (a) The distance between the electrostatic-comb fingers is $10\mu m$. (b) The width of the micro-beam is $20\mu m$.

the beam spring. However, immediately after the fingers are disconnected, the electrostatic force gets activated, pulling back the fingers together).

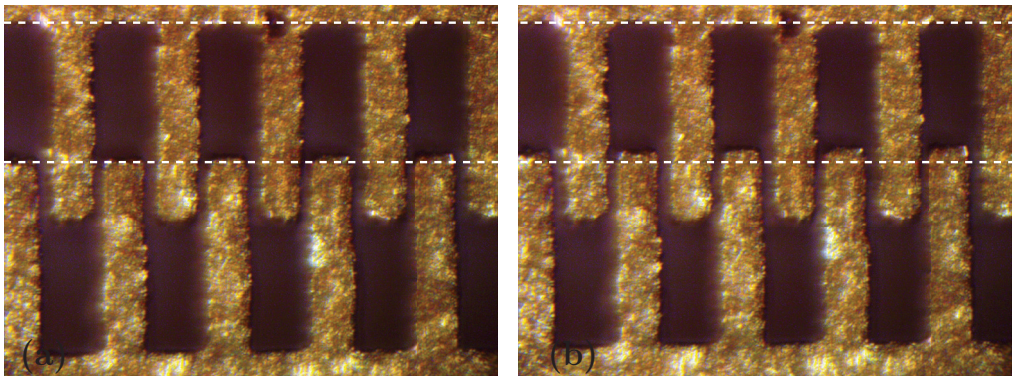
Figure 4-9(a) shows a small displacement of the comb-drive finger just before the oscillation begins, and the bias voltage is $280 \pm 10V$. As can be seen, this small distance is difficult to measure. Figure 4-9(b) shows the digitally magnified image of the fingers. The distance between the fingers is $10\mu m$ (figure 4-8), using this distance as the reference, it can be estimated that the displacement of the finger is approximately $4 - 6\mu m$. However, precise measurement inside a scanning electron microscope is required to accurately measure the displacement of fingers.

As can be seen in figure 4-8(a), the surface of the comb-drive fingers is not smooth and small protrusions, and recess exist at the surface. These protrusion and recess can increase or decrease the distance between the fingers, thereby changing the finger displacement, $d$. The size of these protrusion or recess are on average $1\mu m$, resulting in the error bounds in the gap, $g = 9.34 \pm 2\mu m$. This error bound in the gap results in the error bound in displacement equal to $d = 6.7 - 8.8\mu m$. These estimates are close to the measured value of $4 - 6\mu m$. The small discrepancy could be due to incorrect measurement of displacement derived from figure 4-9.

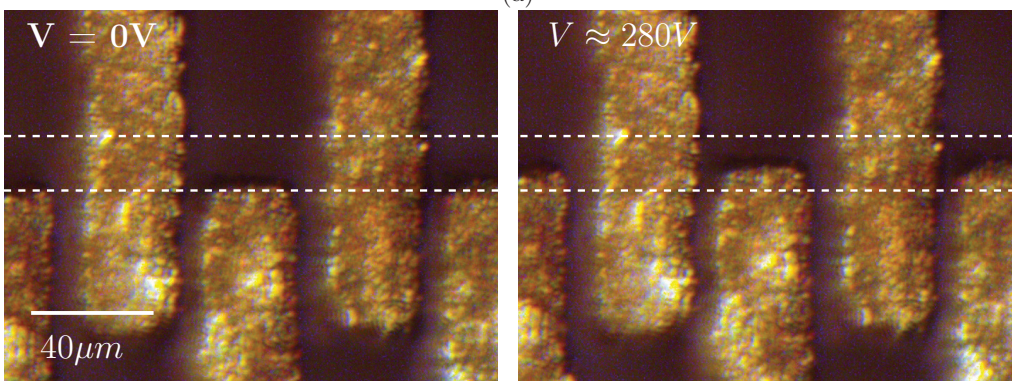
The displacement of the comb-drive fingers can be increased by increasing the length of




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.




(a)




(b)

**Figure 4-9:** Displacement measurement of comb-drive actuator fingers. (a) Small displacement of bottom comb-drive fingers observed at $280 \pm 10V$. (b) Digitally magnified image showing the displacement of fingers. The displacement is approximately $4m - 6\mu m$

the beam springs (or reducing the spring constant). Moreover, a vapor-deposited parylene coating could be used to avoid the electrical connection between the fingers.

### 4.2.4 Time Required for First Prototype

The MALL MEMS fabrication process significantly shortens the time require for fabricating the first prototype. This fact is illustrated by using the example of the comb-drive actuator fabricated in this section. The first step in the concept to creation cycle is to design the device and select the desired fabrication material. Once the design and the material is finalized, the next step is to determine the laser micromachining parameters required to




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

micromachine the material sheet of a given thickness.

For determining process parameters, a micromachining pattern containing horizontal and vertical lines of varying width is used as a test pattern. Multiple instances of this test pattern are micromachined at various laser power, laser frequency, feed rate, and pulse duration (in this work, the pulse duration is fixed to 20 ns). These micromachined patterns are inspected in a scanning electron microscope, and the pattern with smallest feature is determined and the laser micromachining setting corresponding to this pattern is used for device micromachining. The laser micromachining process development time for a new material or sheet thickness is about 6 hours. The determined laser micromachining parameter is used fabricate the comb-drive structure. Depending upon the design, this process can take anywhere between few minutes to one hour. For this estimate, it is assumed that it will take one hour to micromachine the whole device.

The next step is to perform electro-deburring. For electro-deburring, the information about the required electrolytic solution can be derived from here [90]. The laser micromachined part is used as the anode, and another piece of same metal is used as the cathode. The information about the electropolishing voltage can also be derived from reference [90]. As discussed in section 3.5, the electro-deburring voltage must be greater than the electropolishing voltage to ensure selective etching burrs. Some degree of process development may be required for new material, but any voltage above electropolishing voltage results in electro-deburring. Often the easiest option to increase the voltage until the bubbles start to appear at the anode, and perform electro-deburring at this voltage. Generally 30 seconds are enough to completely remove the burrs. For this estimate, it is assumed that the electropolishing process will take total of one hour.

The final step is to align and laminate the individual device layer. This thesis uses Dupont FR1500 and 3M thermal bonding film 583. These adhesive films are very well studied and do not require any process development. Therefore, the bonding can be performed in a total of one hour (30 minutes of handling and 30 minutes of the adhesive curing cycle).

The total time required in fabricating the first prototype is approximately 10 hours. It




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

must be noted that if the fabrication material remains the same, testing new MEMS device takes approximately 4 hours.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 4.3 Micro-Electromechanical Relay

### 4.3.1 Design

The design of the MEM relay is shown in figure 4-10(a). It consists of a movable actuator electrode anchored to the substrate with two cantilever beams. Below the actuator electrode, there are three fixed electrodes. These three electrodes are gate, input, and the output contacts (figure 4-10(b)). The square base of the actuator electrode and the square gate electrode below it form an electrostatic parallel-plate actuator. When a voltage is applied between the actuator electrode and the gate electrode, the actuator electrode is pulled-down due to electrostatic attraction.

A moving electrical contact is attached to the actuator electrode via $12\mu m$ thick dielectric layer as shown in the inset of figure 4-10. Figure 4-11 shows the bottom view of the actuator electrode and displays the movable contact as well as the dielectric layer. The dielectric layer acts as an insulator between the actuator electrode and the movable contact (figure 4-10(a)), as well as between the actuator and the gate electrode.

The movable contact, input contact, and output contact are separated by a small air gap, as shown in the inset of figure 4-10(a). The input and output contacts are fixed while the movable contact is free to move. When a voltage is applied between the actuator electrode and the gate electrode, the actuator electrode is pulled-down towards the gate. As a result, the movable contact makes an electrical connection simultaneously with input and output contacts, completing the electrical connection between the input and output electrode and turning ON the relay. When the voltage is removed, the actuator electrode returns to its original position due to the restoring force of the beam springs.

#### Design Parameters

Figure 4-12 shows the design parameters of the MEM relay. The length of the beam spring is $L_B = 1mm$, and the width is $W_B = 20\mu m$. The length, $L_A$ and width, $W_A$ of the square base of the actuator electrode is $1mm$ , and width of the movable contact, $W_C$ is $500\mu m$.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

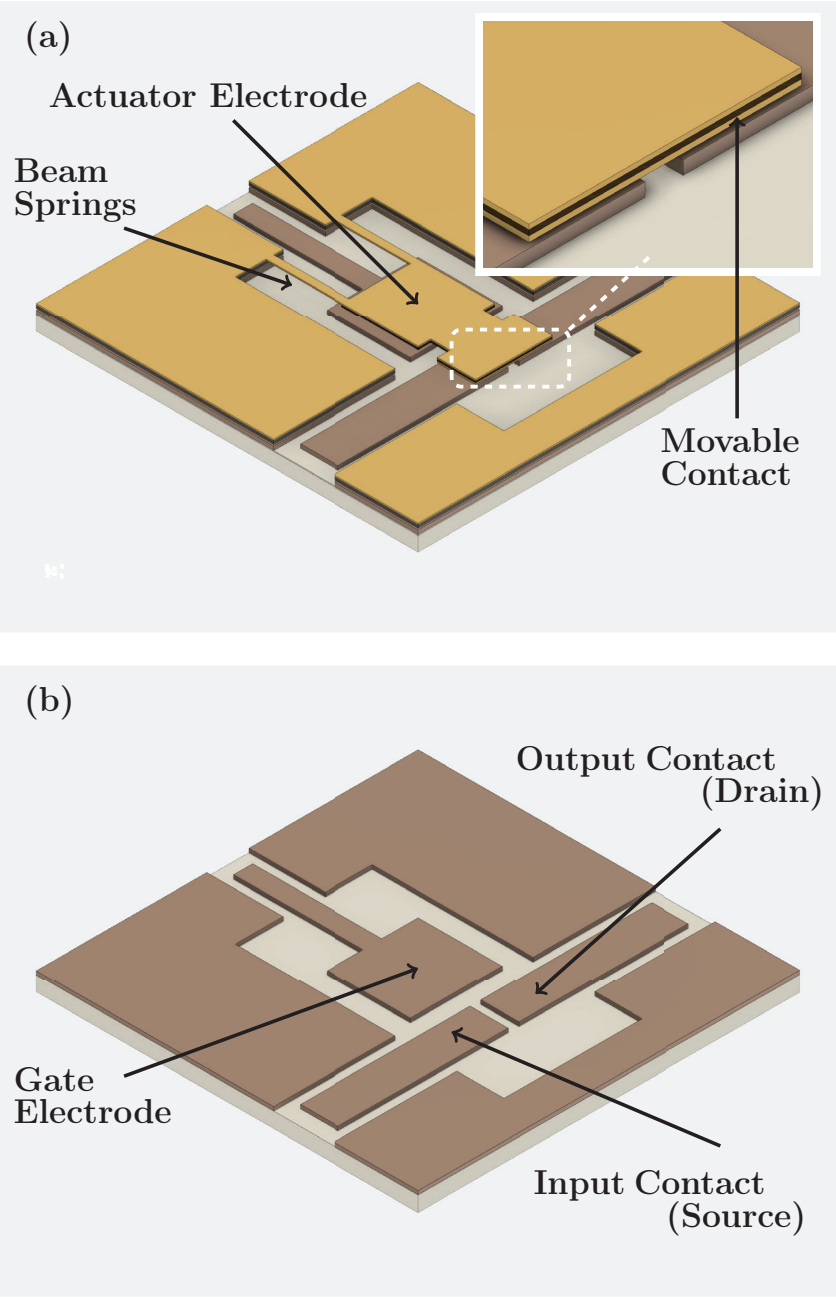



**Figure 4-10:** Schematic diagram of the micro-electromechanical (MEM) relay. It consists of two layers: (a) Top layer containing actuator electrode, beam springs, and movable contact (inset) (b) Bottom substrate layer containing gate, input, and output electrodes.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

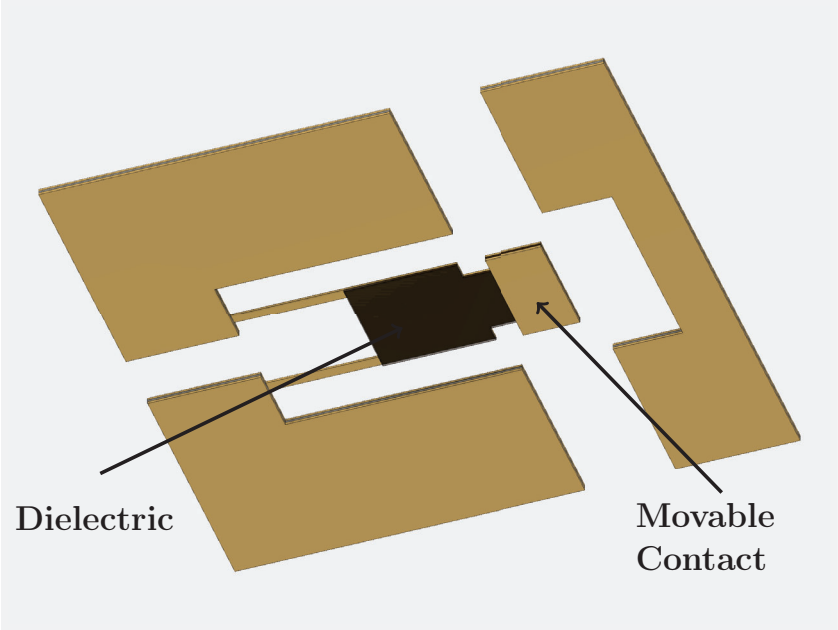



**Figure 4-11:** The bottom view of the actuator electrode showing movable contact and the insulating polyimide layer. The polyimide layer is used as an insulating layer between the actuator electrode and movable contact, and between the actuator electrode and the gate electrode.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

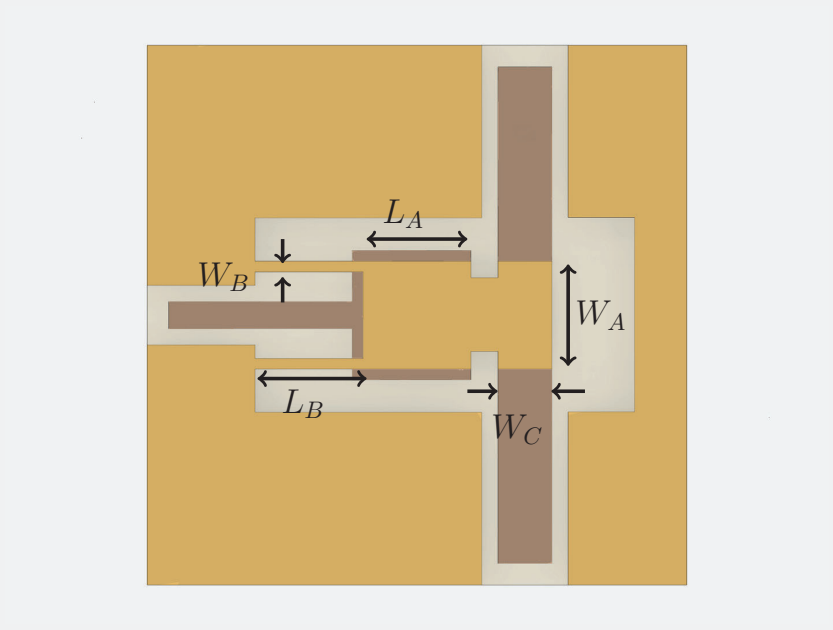



**Figure 4-12:** Schematic diagram showing design parameters of the the MEM relay. These parameters are listed in Table 4.2

The actuator electrode and movable contact are fabricated by LAMPE micromachining $18\mu m$ copper sheet, and the dielectric layer between the two sheets is $12\mu m$ thick polyimide. The distance between the movable contact and input/output contact is $25\mu m$. The distance between the gate electrode and actuator electrode is $55\mu m$, which is the sum of polyimide thickness ($12\mu m$), movable contact thickness ($18\mu m$), and the distance between movable contact and input/output contacts ($25\mu m$). Table 4.2 summarizes the design parameters of the MEM relay.

### Analytical Model

The spring constant of a cantilever beam is given by the equation [91]:

$$k = \frac{3EI}{{L_B}^3} \tag{4.4}$$

where $I = W_B h^3/12$, $h$ is the thickness, $a$ is the width of the beam, $l$ is the length of the




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| **Design Parameters** | **Value** |
|---|---|
| Young's Modulus of Copper, $E$ | $128GPa$ |
| Beam Length, $L_B$ | $1000\mu m$ |
| Beam Width, $W_B$ | $20\mu m$ |
| Beam Thickness, $h$ | $18\mu m$ |
| Actuator Length, $L_A$ | $1000\mu m$ |
| Actuator Width, $W_A$ | $1000\mu m$ |
| Gap between Actuator and Gate, $d$ | $55\mu m$ |
| Pull-in Voltage, $V_{PI}$ | **256V** |

**Table 4.2:** Values of the design parameters for MEM relay.

beam, and $E$ is Young's modulus of the material.

For the given design, two beam springs are connected in parallel, therefore, the effective spring constant

$$k_{eff} = 2k \tag{4.5}$$

Substituting the beam spring design parameters (table 4.2 ) in equation 4.4 and equation 4.5, we get the effective spring constant $k_{eff} = 7.5\mu N/\mu m$.

The actuator electrode can be modeled as a parallel-plate capacitor actuator with one plate fixed while the other plate connected to a beam spring with spring constant $k_{eff}$. When a voltage $V$ is applied between the two plates, the attractive force between the plates is balanced by the restoring force of the beam spring. If $d$ is the distance between the plate at zero bias and $x$ is the distance at bias $V$, then the force on the actuator plate is given by the equation [91]:

$$F_e = \frac{1}{2}\frac{\epsilon_0 A}{(d-x)^2}V^2 \tag{4.6}$$

where $\epsilon_0$ is the permittivity of free space and $A$ is the area of the parallel plates of the capacitor. The electrostatic force $F_e$ is balanced by the restoring force of beam spring, and




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

the displacement at a particular bias voltage can be calculated by solving the equation:

$$\frac{1}{2}\frac{\epsilon A}{(d-x)^2}V^2 - xk_{eff} = 0 \tag{4.7}$$

The equation 4.7 can be used to derive the expression for pull-in voltage [91] and given by the expression:

$$V_P = \sqrt{\frac{8k_{eff}d^3}{27\epsilon_0 A}} \tag{4.8}$$

By substituting the values of design parameters for the actuator electrode from Table 4.2, we get the pull-in voltage $V_P = 204$

### 4.3.2 Fabrication

The MALL fabrication of the MEM relay consists of three steps. First, the fabrication of the actuator electrode, which contains one electrode of the parallel-plate actuator and movable contact. Second, fabrication of the substrate layer, which contains the gate, input, and the output electrode. Third, stack-assembly of actuator electrode and substrate layer to make the MEM relay.

The actuator electrode itself consists of three layers. The top layer is a $18\mu m$ thick copper sheet patterned to form two beam springs, an actuator-plate, and a post for movable contact. The middle layer is a $12\mu m$ thick polyimide layer and covers the actuator-plate and post for movable contact. The middle layer is used to isolate the actuator electrode from the gate electrode and movable contact.

The bottom layer is an $18\mu m$ thick copper layer consisting of a movable contact (figure 4-11). The actuator electrode is fabricated by, first laser micromachining individual layers, and then stack-assembling and bonding them. Figure 4-13 shows the fabricated actuator electrode and the inset shows (actuator electrode)-polyimide-(movable contact) laminate structure.

The substrate layer is fabricated from copper-clad printed circuit board (PCB). The




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

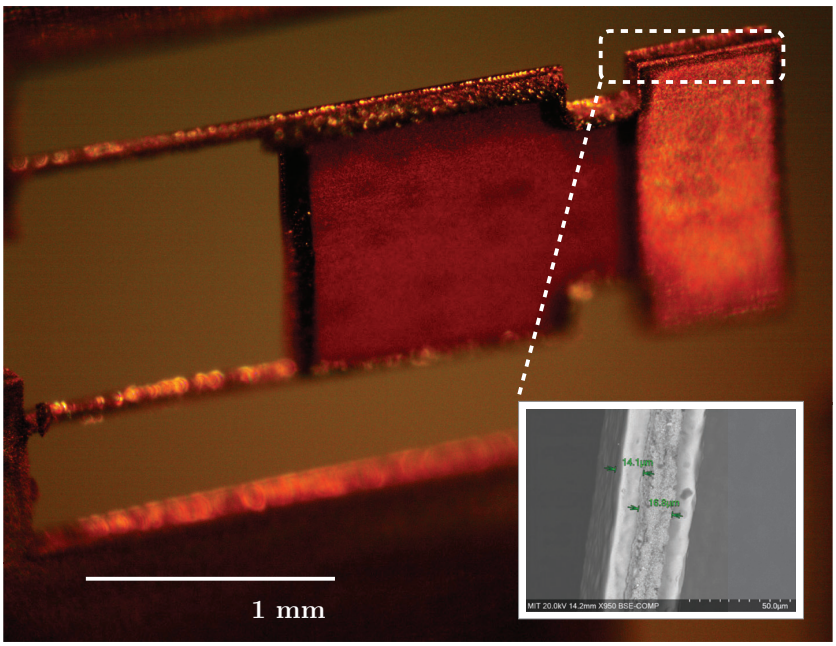



**Figure 4-13:** Optical microscope image of the fabricated actuator electrode showing the polyimide insulating layer and movable contact. The inset shows the copper-polyimide-copper laminate structure.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

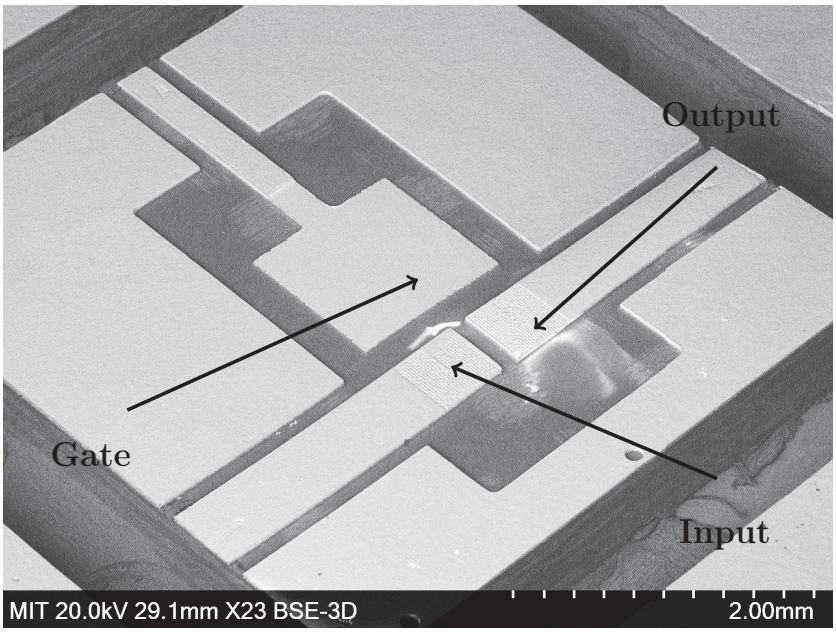



**Figure 4-14:** SEM image of the fabricated substrate layer showing gate, input, and output electrodes.

copper layer is $35\mu m$ thick, and the FR4 support layer is $0.7mm$ thick. The gate, input, and output electrodes are patterned by micro-milling the top copper layer using a $125\mu m$ diameter end-mill. The minimum achievable feature size is $127\mu m$, which is sufficient for this application. The micro-milling process results in micro-burrs at the edges of the electrode. These burrs are removed using the electro-deburring process described in section 3.5. Lastly, four alignment holes are drilled using $1.55mm$ drill bits. The fabricated substrate layer is shown in figure 4-14.

Finally, the actuator electrode is aligned and bonded over the substrate layer using the multi-lamina assembly process described in section 4.1. For alignment, dowel pins of diameter $1.5mm$ are used, which provides sufficient alignment accuracy required for the MEM relay. A $25\mu m$ thick adhesive layer is used to bond the two layers and create the desired gap between the movable contact and input/output contacts. Figure 4-15 shows the fabricated MEM relay. It can be seen in the inset image of figure 4-15 that the gap between the movable contact and input/output contact is $27\mu m$, which is close to the thickness of




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

the adhesive layer.

### 4.3.3 Results & Discussion

The fabricated MEM relay is tested by measuring the current between the input and output electrode (i.e., the source-drain current, $I_D$) at various gate voltage $V_G$. An HP4156 semiconductor parameter analyzer equipped with four source-measure units (SMU) is used to take the measurement. The actuator electrode is connected to electrical ground, and voltage $V_G$ is applied to the gate electrode. For measuring the source-drain current $I_D$, the input electrode (source) is connected to ground, and the voltage $V_{DS}$ is applied to the output (drain) electrode.

Figure 4-16 shows $I_D$ vs. $V_{DS}$ measurement at different gate voltage $V_G$. The $V_{DS}$ is varied from $0V$ to $10V$, and the corresponding $I_D$ is measured and plotted. For the gate voltage $V_G$ less than $220V$, there is an air gap between the movable contact and input/output contact (figure 4-17(a)) and the current $I_D = 0$. When the voltage $V_G > 240V$, the movable contact makes the electrical connection with the input and output contact (figure 4-17(b)) and the current, $I_D$ flows from input to the output electrode. The observed current saturation is due to the maximum source-current limit ($I_{max} = 100mA$) of the measuring instrument.

To observe the switching action of MEM relay, a voltage $V_G = 240V$ is manually applied, and the corresponding current $I_D$ is measured and plotted with respect to time. Figure 4-18 shows the current vs. time plot when the relay is turned ON and OFF repeatedly. As stated earlier, the square wave nature of the current profile is due to the limit on maximum current that can be sourced from the measuring instrument.

An insulating coating between the actuator and gate electrode is essential to avoid short-circuit between the two electrodes. One of the goals of this work is to demonstrate the assembly of a MEM relay from conducting and insulating parts; as a result, a polyimide film is used as an insulating layer instead of using vapor-deposited dielectric material such as parylene.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

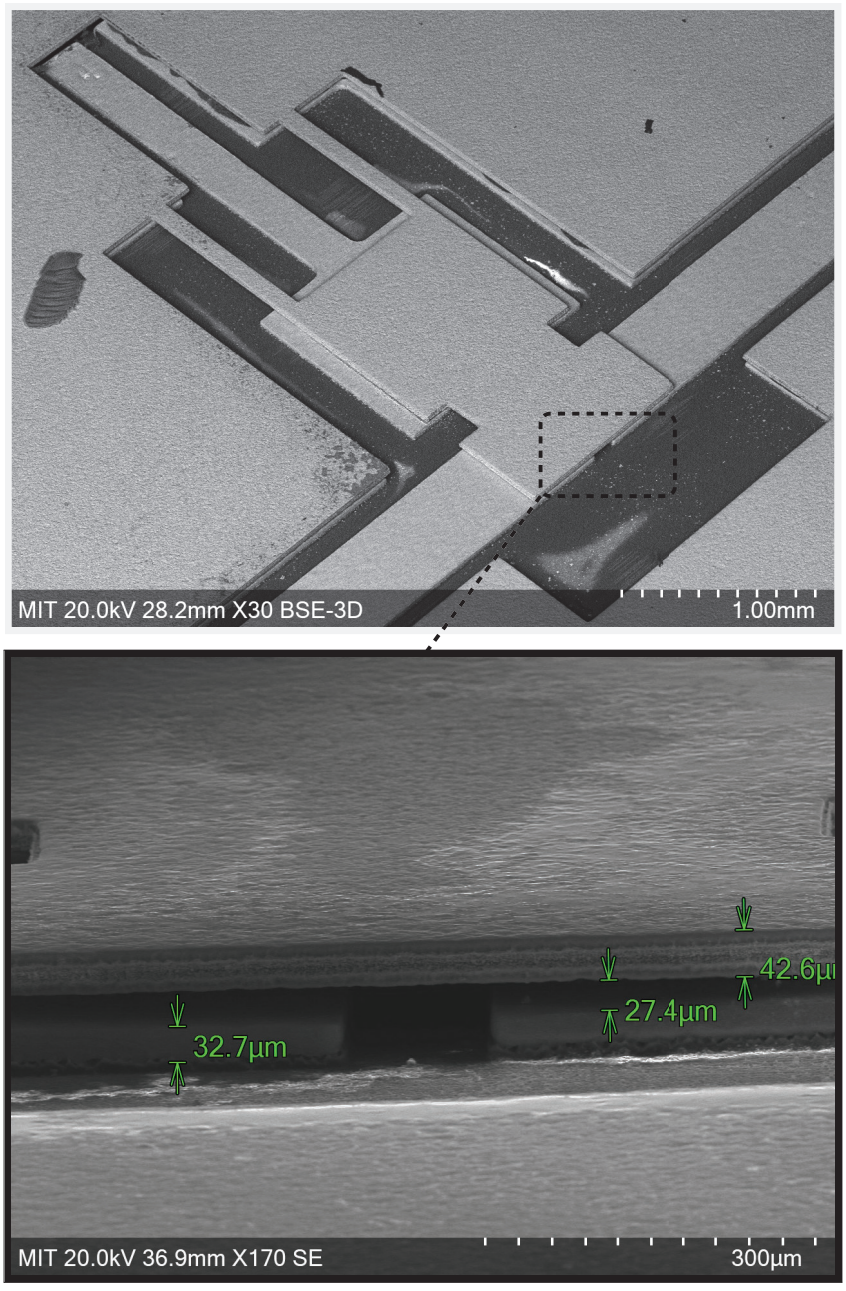



**Figure 4-15:** The SEM image of the assembled MEM relay. The inset shows the distance between the movable contact and input & output contacts, which is 27.4$\mu m$

.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

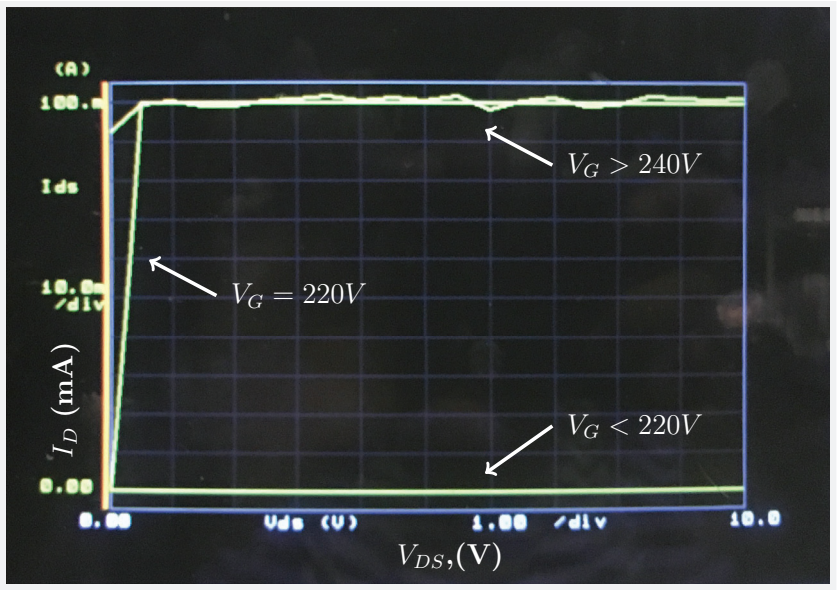



**Figure 4-16:** The source-drain current-voltage ($I_D$ vs. $V_{DS}$) measurement at various $V_G$ gate voltage .

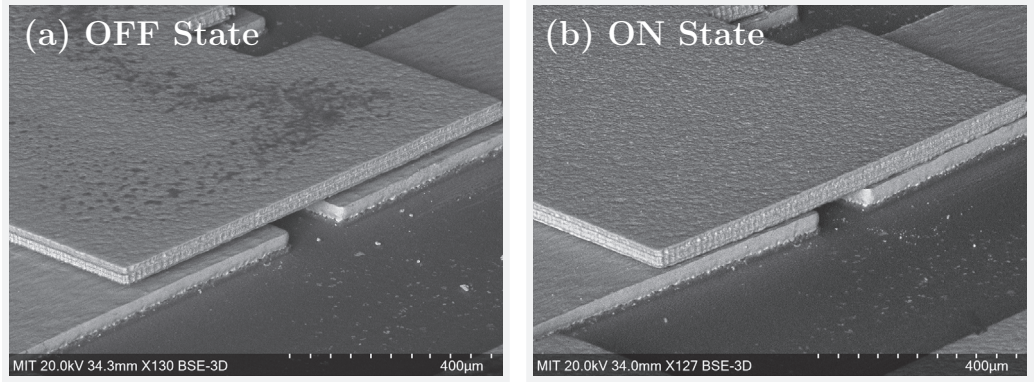



**Figure 4-17:** The SEM images showing the operation of MEM relay. (a) Small air-gap between the movable contact and input & output contact when the relay is off. (b) The formation of electrical contact between the movable electrode and input & output contact when the relay is turned on.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

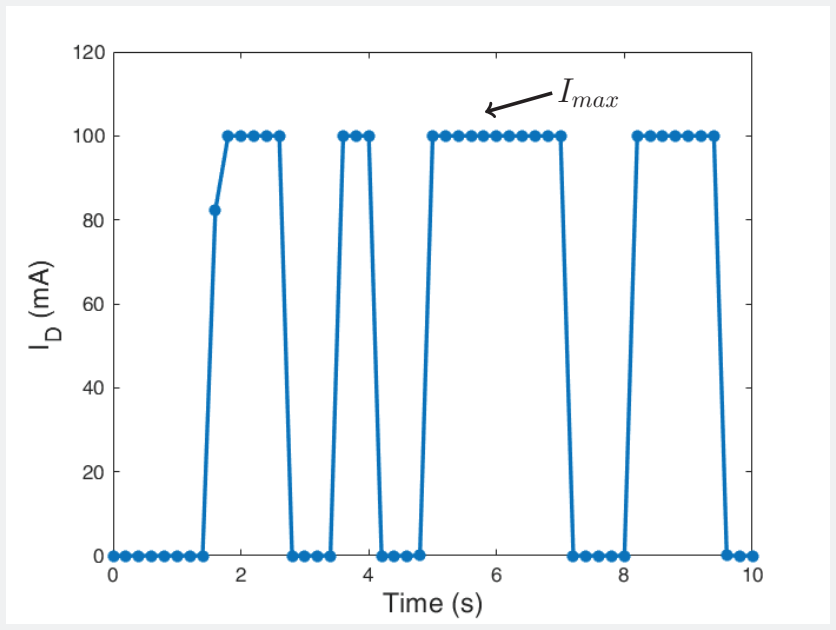



**Figure 4-18:** The current $I_D$ vs. time plot of MEM relay demonstrating current switching. The current $I_D$ is measured by applying a constant voltage $V_{GS} = 10V$ between the input and output electrode and manually cycling the gate voltage $V_G = 240V$ on and off.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

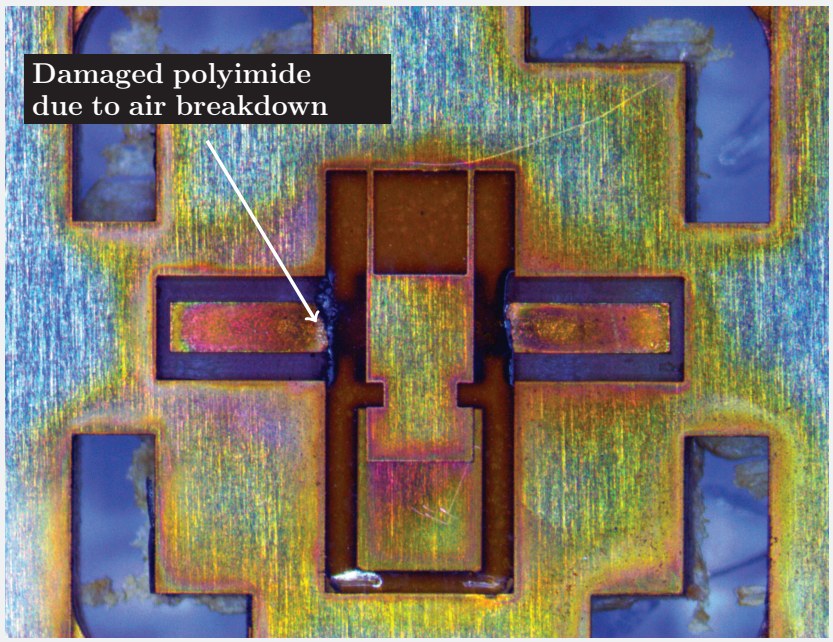



**Figure 4-19:** Damaged polyimide insulating layer due to the dielectric break down of air.

Unfortunately, the polyimide film results in three major complications. First, the thickness of the polyimide film adds an additional $12\mu m$ between the actuator and gate electrode, thereby increasing the pull-in voltage. Second, the polyimide film does not provide a conformal coating around the gate electrode. As a result, at high voltage, avalanche breakdown of gas molecules can occur [92][93], resulting in the destruction of the relay. This phenomenon is evident in the relay design, where the polyimide film is attached to the gate electrode, as shown in figure 4-19. As can be seen, the polyimide layer is burned due to the dielectric breakdown of air.

Third, the polyimide layer used in MEM relay fabrication causes stiction of the actuator electrode with the gate electrode. The stiction problem in MEM systems is widely known [94] and can be minimized by applying an anti-stiction coating [95][96]. In summary, using a vapor-deposited dielectric coating such as parylene coating can drastically improve the performance of MEM relay by reducing the pull-in voltage, increasing the maximum safe operating voltage, and minimizing stiction.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## 4.4 Diamond Rotor

### 4.4.1 Introduction

Amyloid fibrils are self-assembled filamentous structures associated with more than fifty diseases, including Alzheimer's and Parkinson's [97]. Determination of the atomic-resolution 3D structure of amyloid fibrils is essential to develop a cure for these diseases [98]. However, obtaining atomic-resolution structural information using classical methods such as electron microscopy, atomic force microscopy, x-ray crystallography, and solution-state NMR spectroscopy is challenging as amyloid fibrils are inherently noncrystalline, insoluble and lack 3D order.

Recent advances in magic angle spinning (MAS) [99][100] solid-state NMR spectroscopy has enabled determination of the 3D structure of amyloid fibrils [101][102]. In MAS, a sample is rotated around an axis tilted at 57.74° with respect to the static magnetic field, resulting in the averaging of anisotropy of nuclear interaction, which increases the resolution of NMR spectroscopy. The resolution of MAS NMR increases with increasing spinning frequency [103]. Unfortunately, the spinning frequency of state-of-the-art MAS probes or rotors is limited to $110KHz - 140KHz$ [104] [105]; as a result, the resolution of MAS NMR is limited.

The maximum spinning frequency of MAS rotors depends upon the fluidic properties of surrounding gas, and the speed at the circumference of the rotor must be less than the speed of sound in this gas [106][107][108]. Figure 4-20 shows the calculated maximum theoretical limit of MAS frequency of rotors with various diameters currently employed in MAS NMR. The theoretical limit is calculated using the value of the speed of sound in helium at various temperatures [109]. As can be seen, the theoretical limit on the maximum frequency increases with decreasing the diameter of the rotor. For the $0.7mm$ MAS rotor, smallest diameter rotor currently employed, the maximum theoretical limit on spinning frequency is $454.7KHz$. However, in practice, the spinning frequency is limited by the tensile strength of the zirconia material used for fabricating rotors. The maximum spinning




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

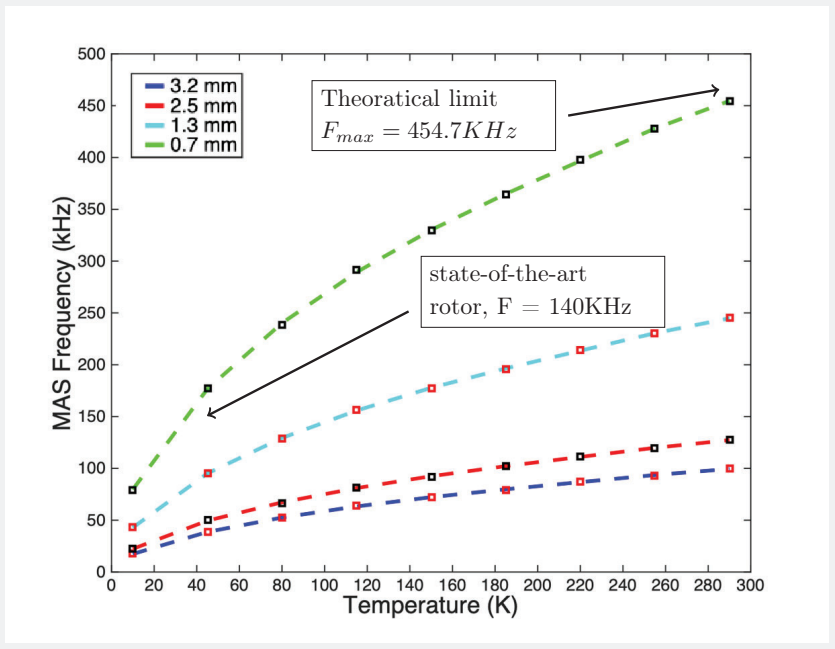



**Figure 4-20:** Maximum theoretical MAS frequency for various diameter rotors. As can be seen, the $0.7mm$ diameter rotors have the highest MAS spinning frequency.

frequency of state-of-the-art zirconia ($ZrO2$) rotor is approximately $126KHz$ [104][105] (figure 4-20).

### 4.4.2 Diamond Rotor

Diamond is the hardest material, and the tensile strength as high as $5190MPa$ has been reported for chemical vapor deposition (CVD) grown diamonds [110]. As a result, diamond is an excellent choice for making MAS rotors and expected to spin at a higher frequency than state-of-the-art zirconia rotors. Moreover, diamond is transparent to microwaves; as a result, ideal for dynamic nuclear polarization (DNP) experiments, which is used in conjunction with NMR. However, machining diamond is difficult because of its hardness. Although laser cutting is employed to cut diamond sheets, the process does not produce the desired aspect-ratio required for fabricating a diamond rotor. This thesis shows high aspect ratio laser micromachining of diamond and its application in making the diamond rotor for MAS NMR.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

**Figure 4-21:** Schematic diagram showing the dimensions of the MAS diamond rotor probe.

The employed laser micromachining method is an extension of LAMPE micromachining discussed in chapter 3 and discussed next.

### 4.4.3 Laser Micromachining of Diamond

Figure 4-21 shows the schematic diagram of the diamond rotor. The inner diameter of the rotor is $500\mu m$, and the outer diameter is $700\mu m$. The total length of the rotor is $5mm$.

The chemical vapor deposition (CVD) grown diamond sheets are obtained from Element Six (www.e6.com). The sheet is cut into small square logs of dimension $1mm \times 1mm \times 5mm$, and the rotors are fabricated by machining these logs. For laser micromachining, the log is mounted vertically, and the laser micromachining is performed from one end to the other. The focus is gradually moved until the complete hole is drilled. Figure 4-22 shows the micromachined inner diameter and the outer diameter of the diamond rotor.

### 4.4.4 Theoretical Explanation

The high aspect ratio depends upon the optical penetration depth. The optical penetration depth of the CVD diamond is measured using transmission spectroscopy and described next.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

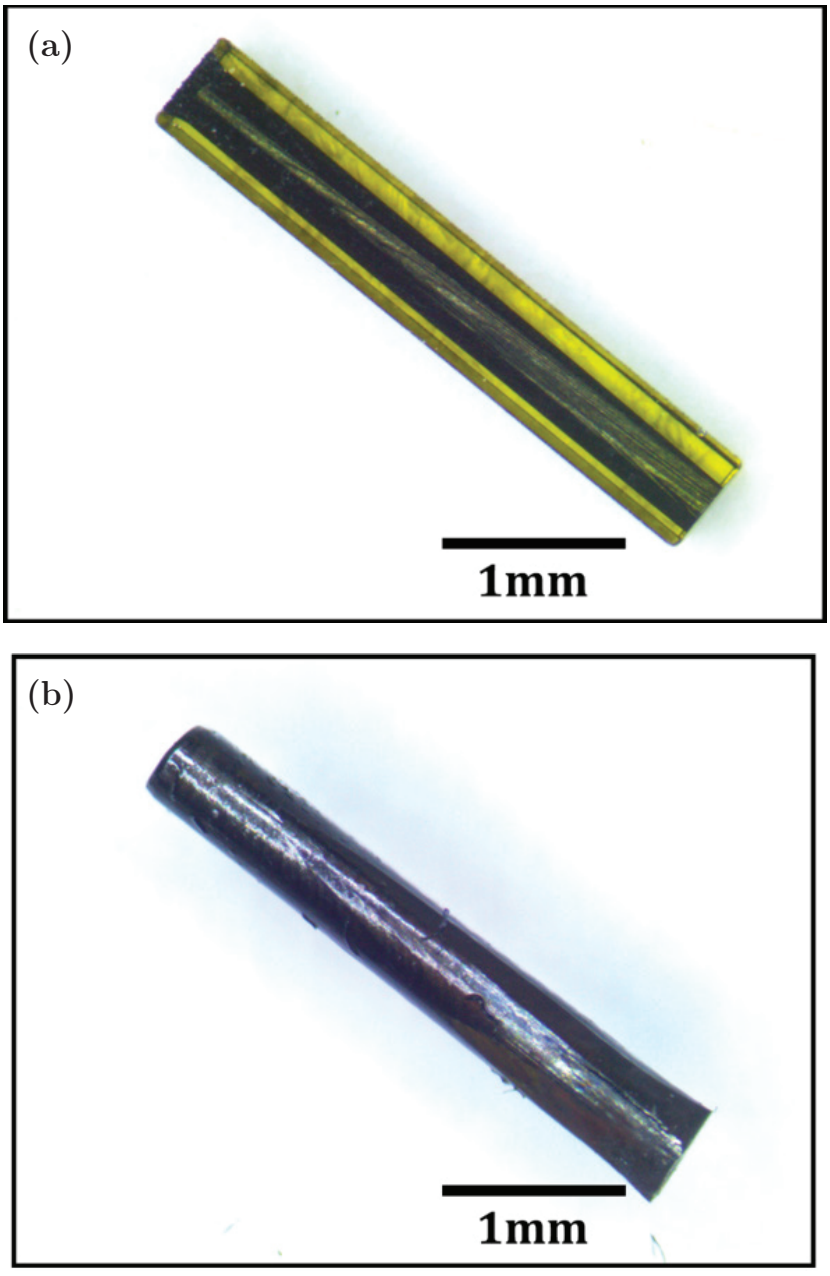



**Figure 4-22:** Laser micromachining of diamond rotor (a) Inner diameter. (b) Outer diameter.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

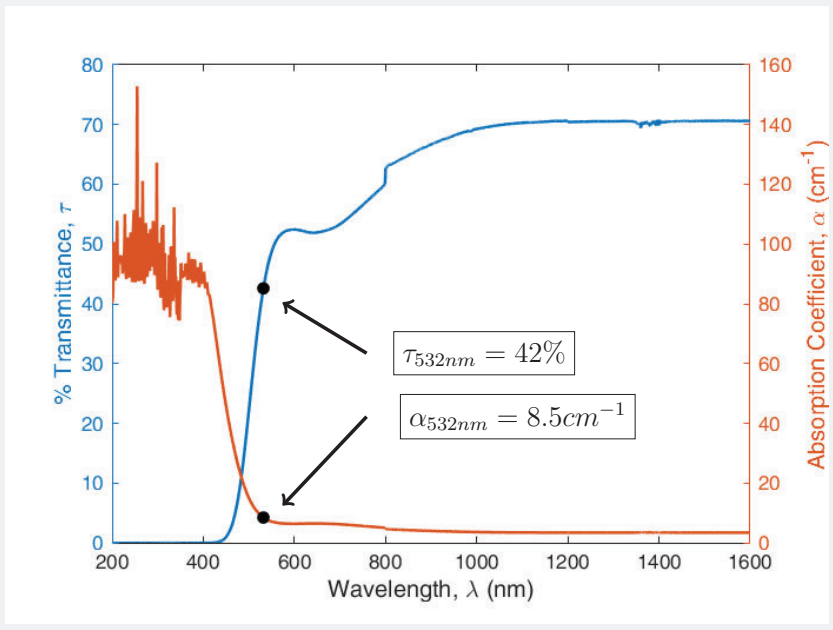



**Figure 4-23:** The transmittance and absorption coefficient of CVD diamond. At $532nm$ the transmittance, $\tau = 42\%$ and the absorption coefficient $\alpha = 8.5cm^{-1}$

### Optical Penetration Depth

The absorption coefficient $\alpha$ is calculated from the transmission spectrum. Figure 4-23 shows the transmittance spectra and the absorption coefficient plot of the CVD diamond. The transmittance $\tau = 42\%$ and absorption coefficient $\alpha = 8.5cm^{-1}$ at $532nm$.

Figure 4-24 shows the variation of optical penetration depth, $l_\alpha = 1/\alpha$, with respect to the wavelength. As can be seen, at $532nm$, the absorption is sufficiently low to allow large optical penetration depth but large enough to cause laser absorption and heating. Once the laser is absorbed by the diamond, it causes a phase change resulting in graphitization of the diamond [111] [112]. The absorption coefficient of graphite is much larger than the CVD diamond, resulting in a sudden increase in the laser absorption and increase in temperature. Finally, the graphite is oxidized into $CO_2$ in the presence of atmospheric oxygen. Although this work uses $532nm$ wavelength for diamond micromachining, the author recommends using wavelength around $1028nm$ to take advantage of a further


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

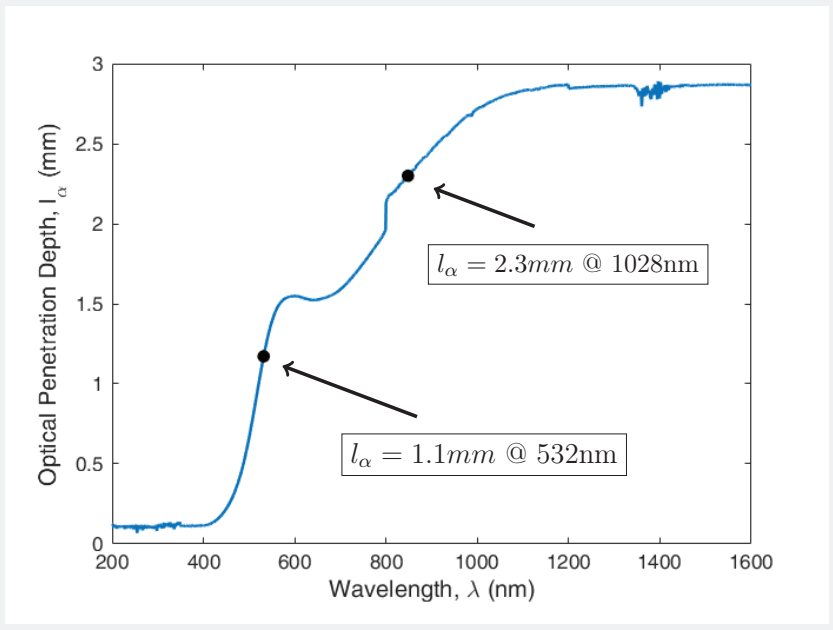



**Figure 4-24:** The optical penetration depth of CVD diamond at $532nm$ is $1.1mm$.

increase in optical penetration depth while still maintaining sufficient absorption (figure 4-24).

## 4.5 Conclusions

In conclusion, the MALL process can be used to manufacture functional MEMS devices using metals. These MEMS devices have lateral features and aspect ratio comparable to devices fabricated using DRIE. Moreover, using MALL, MEMS devices can be rapidly developed and fabricated at low cost.

In DRIE, process development is required for fabricating a new MEMS. The process parameters are optimized by performing a series of etch cycle and varying the etching parameters. Each etch cycle requires the loading and unloading of the wafer, and often the process is done manually. As a result, DRIE process development takes substantial time. However, in the MALL process, various parameters such as laser power, pulse frequency,




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

feed rate, and pulse duration can be tested in a single cycle, thereby considerably shortening the process development time.

In DRIE, the etch rate depends upon the device area due to the loading effect [113][72]. As a result, for each MEMS design, a new set of DRIE process parameters must be determined, which significantly hinders design iteration and increases device development time. However, in the MALL process, the laser micromachining rate is independent of the device area (or design); as a result, many MEMS designs can be tested without requiring any process development.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Chapter 5

# Discussion & Conclusion

## 5.1 Advantages of MALL

The MALL fabrication process offers several advantages over conventional lithography-based MEMS fabrication processes such as greater material selection and integration, low cost of fabrication, rapid development, and integrated packaging. These advantages are discussed in detail in the following sections.

### 5.1.1 Greater Choice of Fabrication Materials

The fabrication materials used in the lithography-based MEMS manufacturing process is largely limited to silicon. On the other hand, the MALL process enables fabrication with a wide range of materials such as metals, ceramics, polymers, and composite materials. The ability to fabricate MEMS using a wide range of materials greatly enhances the functionality and performance of these systems. For example, materials with interesting properties such as magnetic, piezoelectric, pyroelectric, thermoelectric, and superconducting materials can be employed to design novel MEM systems.

The application of metals in fabricating MEM systems is especially of great interest. Metals offer a wide range of mechanical and electrical properties, and these properties can be tuned using the alloying process. Moreover, due to the roll forming process, the cost and




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

energy requirement for making metal foils is much less than that of silicon wafers.

### 5.1.2 Fabrication with Incompatible Materials

In conventional microfabrication process, MEMS are fabricated layer-by-layer by successive deposition and patterning of material layers. Each material layer is deposited and patterned using a unique process, and this process must be compatible with the previously deposited materials. The integration of these processes is often difficult due to material incompatibilities and thermal constraints. However, in the MALL fabrication process, the individual layers of MEMS are fabricated separately, and stack-assembled and bonded to make a system. As a result, materials with incompatible fabrication process can be bonded together, in any order, to manufacture MEMS.

### 5.1.3 Low-Cost Fabrication

The high cost of MEMS fabrication is due to two reasons. First, the large capital investment in microfabrication tools such as exposure system, mask aligners, chemical vapor deposition (CVD), and deep reactive ion etching (DRIE) systems. Second, the high cost of raw materials and chemicals such as photoresists, developers, etching solutions and gases, and silicon wafers. In contrast, the MALL process requires a laser micromachining system and electrochemical etching tools, which have a much low cost of ownership than the traditional microfabrication tools. Similarly, the raw materials used in MALL fabrication are metal foils, polymers, and ceramics. These materials are much more cost-effective than silicon wafers, utilized in conventional MEMS fabrication process.

The DRIE can produce multiple parts in parallel; as a result, after the initial investment, equipment cost shared per device is significantly reduced. In fact, the DRIE equipment cost per device for fabricating the comb-drive mentioned in section 4.2 is approximately \$0.09 (detailed calculations for cost estimates are presented in appendix A.1). However, the material cost per device in DRIE remains substantial. For the comb-drive actuator, the material cost per device is \$0.51 (Appendix A.1), which is about 85% of the total device




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

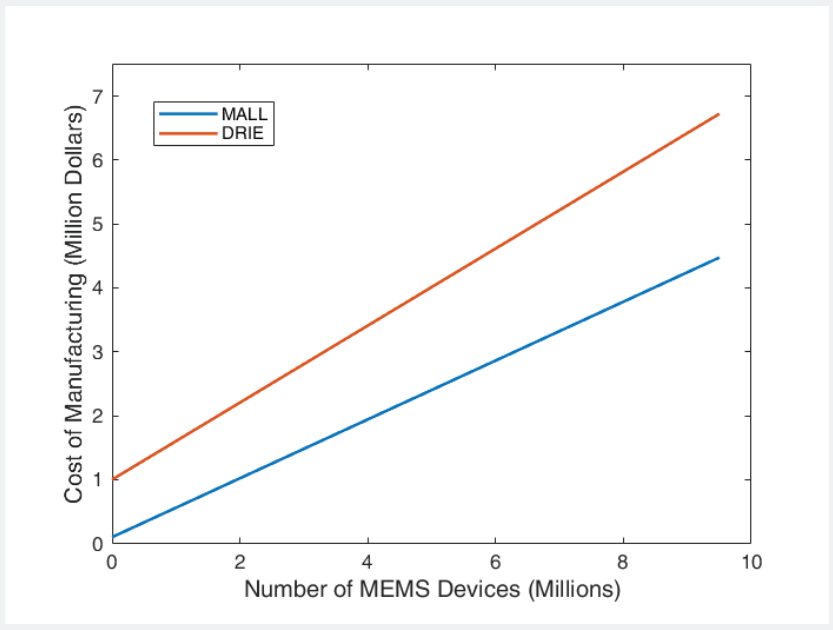



**Figure 5-1:** Cost of fabricating the comb-drive actuator using DRIE and MALL process.

cost.

In the MALL fabrication process, even though the initial cost of a laser micromachining system is low, the equipment cost per device is significantly high due to the low part production rate. For the comb-drive actuator discussed in section 4.2, the equipment cost per device is \$0.46 (appendix A.1). However, the material cost per device is significantly low, about \$0.0015 per device (appendix A.1). As a result, the total cost per device for both DRIE and MALL is comparable. Table 5.1 summarizes the manufacturing time and cost associated with fabricating the comb-drive actuator using DRIE and MALL process.

Although the cost per device for both the processes is comparable, the manufacturing cost for MALL is less due to low initial capital investment. Figure 5-1 compares the manufacturing cost of DRIE and MALL process. As can be seen, for a given number of fabricated MEMS, the corresponding manufacturing cost using the MALL process is $\$0.9M$ less than the cost associated with DRIE.

In MALL process, individual layers are fabricated using LAMPE micromachining, and


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| Parameter (per device) | DRIE | MALL |
|---|---|---|
| Manufacturing time | $0.2min$ | $12min$ |
| Equipment cost | \$0.09 | \$0.46 |
| Material cost | \$0.51 | \$0.0015 |
| Total manufacturing cost | \$0.51 | \$0.4615 |

**Table 5.1:** Time and cost of fabricating the comb-drive actuator using DRIE and MALL fabrication process. The detailed calculations for cost estimations are presented in appendix A.1

| Parameter (per device) | DRIE | MALL |
|---|---|---|
| Manufacturing time | 4.8 seconds | 30 seconds |
| Equipment cost | \$0.03 | \$0.02 |
| Material cost | \$0.61 | \$0.0015 |
| Total manufacturing cost | \$0.64 | \$0.0215 |

**Table 5.2:** Time and cost of fabricating MEM relay using DRIE and MALL fabrication process.

the rate of machining increases with a reduction in the sheet thickness. As a result, the MALL process is especially advantageous for MEMS using thin laminates such as the relay discussed in section 4.3. The thin laminate for MEM relay is fabricated by removing the excess material using LAMPE micromachining. The boundary of the excess region is $60mm$ long, and the feed rate of $120mm/min$ is used for micromachining. As a result, one relay can be fabricated in 30 seconds. The cost of a laser micromachining system is $\$100,000$, and assuming the lifespan of 5 years, the cost per device comes out to be \$0.02 (including the material cost).

Fabricating the same MEM relay using DRIE costs \$0.64 per device, 95% of which comes from material cost. The cost of one wafer is \$220 [114], and each wafer can accommodate 360 devices; therefore, the material cost per device is \$0.61. The remaining \$0.03 comes from equipment cost per device. Table 5.2 summarizes the cost per device for MEM relay using DRIE and MALL, and figure 5-2 compares the manufacturing cost for these processes.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

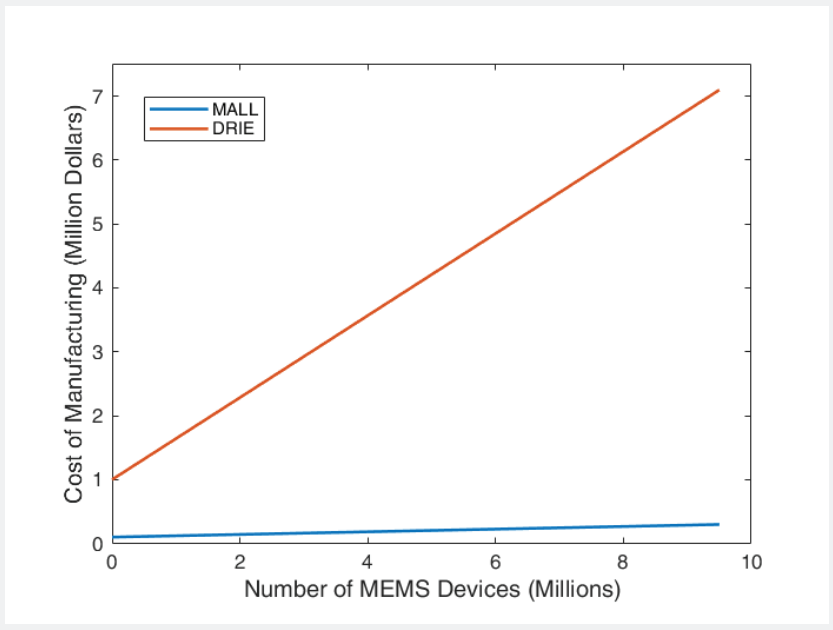



**Figure 5-2:** Cost of fabricating MEM relay using DRIE and MALL process.

### 5.1.4 Rapid Development

#### Rapid Material Integration

In a conventional lithography-based fabrication process, MEM systems are fabricated layer-by-layer. Fabricating each layer involves a series deposition and patterning steps such as chemical vapor deposition (CVD), spin-coating photoresist, resist exposure and development, and wet or dry etching. The fabrication process for each layer must be optimized and should be compatible with the previously deposited materials. This step-by-step process optimization and integration is extremely time-consuming. Moreover, in conventional MEMS fabrication, MEM chips and packages are fabricated separately and integrated at a later stage. This integration of chips and packages further increases the development time.

In the MALL fabrication process, individual layers can be manufactured independently with minimal process development and effortlessly bonded to make MEM systems; as a


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

result, significantly reducing the device development time. Moreover, laser micromachining system used in MALL can exist side by side with other macro-fabrication equipment (such as milling, molding, and embossing), and utilizes similar computer-aided design (CAD) and computer aided manufacturing (CAM) tools, thereby enabling integrated design of MEMS and packaging, simultaneous fabrication, and rapid integration.

In conventional MEMS fabrication, the three-dimensional MEMS design is converted into a series of two-dimensional lithography masks, which are subsequently used for fabrication. However, in MALL, the laser micromachining toolpath can be directly exported from the integrated CAD/CAM design tool. This integration of design, manufacturing, and finite element analysis (FEA) of MEMS in one environment enables the rapid iteration of design, simulation, and prototyping.

**Rapid Process Development**

In DRIE, some degree of process development is required to etch the desired MEMS structure. The etching process parameters are determined by performing a series of etch cycles and varying the etching parameters. Each etch cycle requires the loading and unloading of wafer, and often the process is done manually. As a result, the process development take substantial time.

However, in the MALL process, various laser micromachining parameters such as laser power, pulse frequency, feed rate, and pulse duration can be tested in a single cycle by performing a series of micromachining tests. As a result, considerably shortening the process development time. The laser micromachining process development time for new material or new sheet thickness is about 6 hours.

**Rapid Design Iteration**

In MALL, the micromachining rate of PLA or LAMPE micromachining is independent of the device area (or design). As a result, many designs can be tested without requiring any process development. However, in DRIE, the etch rate depends upon the device area due to


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

the loading effect [113][72]. As a result, for each MEMS design, a new set of DRIE process parameters must be determined, which significantly hinders design iteration and increases device development time.

### 5.1.5 Integrated Packaging

The packaging of MEMS is essential for protecting the system from a harsh environment. Moreover, in certain MEM systems, the packaging also acts as an interface between the MEMS system and the external environment. The microfabrication processes used to fabricate MEMS chips are incompatible with the macro-fabrication processes that are used to fabricate MEMS packages. As a result, the MEMS die, and the package are fabricated independently and then integrated later. This integration process is challenging and increases development time and cost. Often, the cost of MEMS packaging is higher than the cost of fabricating MEMS chips.

In MALL, the fabrication process and materials are compatible with the macro-fabrication processes. As a result, the MEMS layers and the packaging layers can be fabricated simultaneously. Moreover, these layers can be laminated in a single step, thereby enabling the fabrication of MEM systems with integrated packaging. However, the lamination process uses polymer adhesive films for bonding the packaging layers. These polymer films are permeable to water molecules [115][116]. Thus, the MALL integrated packaging can not be used for package MEMS requiring a hermetic seal.

## 5.2 Limitations of MALL

### 5.2.1 Inadequate for Fabricating 2.5D Structures

The conventional microfabrication utilizes deep reactive ion etching (DRIE) and anisotropic wet etching processes [82][117] to fabricate MEMS structures. These processes allow control over etch depth, and therefore, enable the fabrication of two-and-a-half dimension ($2.5D$) structures. However, the LAMPE micromachining process developed in


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

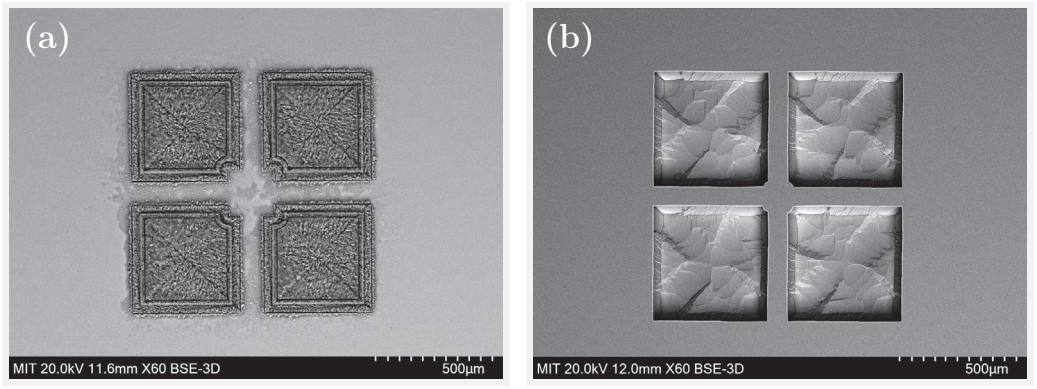



**Figure 5-3:** Fabrication of 2.5D structure in silicon using laser-micromachining: (a) Silicon nitride mask layer patterned using laser-micromachining. (b) $KOH$ anisotropic etch to fabricate 2.5D structures.

this thesis is restricted to micromachining laminates.

One approach to overcome this limitation is to use $2.5D$ structures fabricated in polymers using the conventional pulsed ablation laser-micromachining process. Another approach is to integrate the laser-micromachining with silicon anisotropic etching process. In this approach, first, a silicon nitride mask layer is removed using laser-micromachining (figure 5-3(a)), and then, potassium hydroxide etching [2] is performed to create $2.5D$ structures (figure 5-3(b)). The third approach is to use micro-milling followed by the electro-deburring process (section 3.5) to create $2.5D$ structures. The micro-milling offers vertical features as small as $5\mu m$. However, the minimum achievable feature in the lateral direction is limited to $50\mu m$. figure 5-4(b) shows micro-milling $2.5D$ recess with a varying depth.

### 5.2.2 Limit on Minimum Thickness of Microstructures

The conventional microfabrication process utilizes surface micromachining [48][49][50] to fabricate thin free-standing microstructures. Figure 1-5 shows the schematic diagram of the surface micromachining process. First, a sacrificial layer is deposited on a substrate and patterned using photolithography and etching. Next, the structural layer is deposited and patterned on top of the sacrificial layer. Finally, the sacrificial layer is removed to create free-standing microstructures. The deposition process, such as chemical vapor deposition



537
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

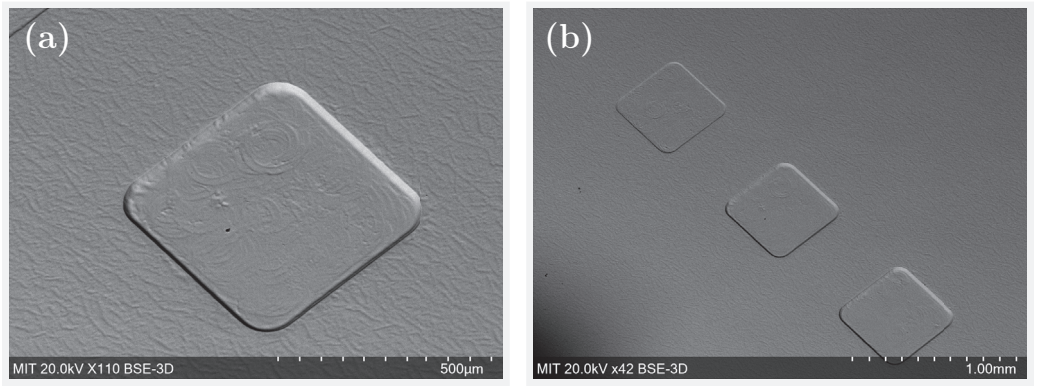



**Figure 5-4:** Fabrication of 2.5D structures using micro-milling. (a) Square recess fabricated using micro-milling followed by electro-deburring. Feature size as small as $100\mu m$ can be fabricated using this method. (b) The micro-milling method provides a high degree of control over the recess depth. The figure shows fabricated recess with depth $5\mu m$, $10\mu m$, and $15\mu m$.

(CVD) used to deposit the sacrificial layer or structural layer, allows a high degree of control over the thickness of the deposited material. As a result, free-standing structures with a small thickness (tens of nanometers) can be fabricated (such structures are desirable for applications such as high-speed resonators [24][118]).

However, in the MALL process, MEMS structures are fabricated with thin sheets of materials and manually assembled or roll-to-roll laminated. There is a practical limit on the thickness of sheets that can be manually handled or laminated using the roll-to-roll process. Typically, foils with thickness greater than or equal to $20\mu m$ can be used for manual assembly and roll-to-roll lamination process. As the thickness of the laminate is reduced, handling of these ultra-thin laminates becomes difficult. For a MEM device requiring sheets with a thickness between $20\mu m$ to $5\mu m$, a rigid or flexible frame can be utilized. Figure 5-5 shows a $10\mu m$ thin microstructure supported by a frame. For fabricating this structure, first, a window is micromachined in $200\mu m$ thick silicon substrate. Next, a $10\mu m$ thick silver foil is bonded to the silicon substrate. Finally, the device structure is laser micromachined on the bonded silver foil. For MEM systems requiring metal foils with thickness in the order of $5\mu m$ to $1\mu m$, the desired metal can be electrodeposited over water or solvent-soluble films, and such laminates can be used to fabricate MEMS.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

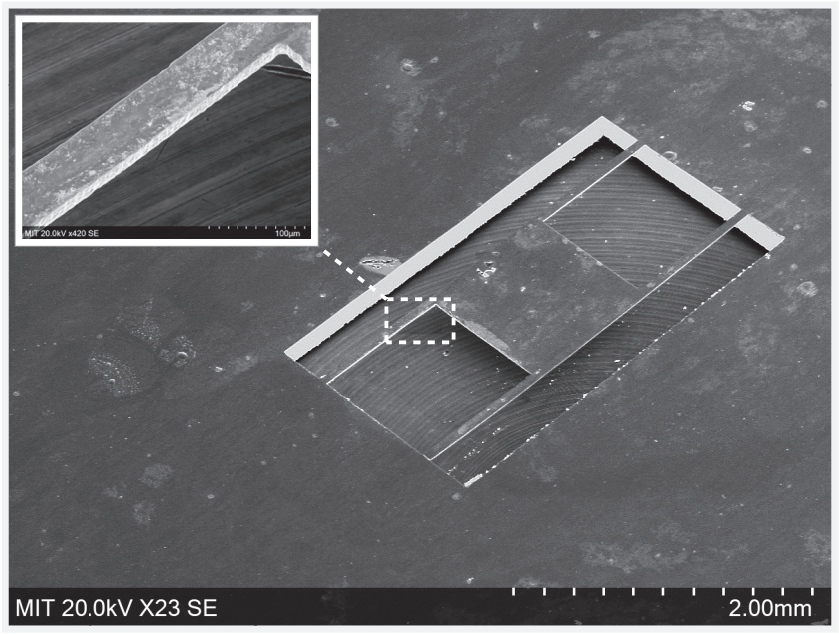



**Figure 5-5:** Manipulation and laser micromachining of ultra-thin metal foils using a rigid frame. First, a window is micromachined in the rigid frame. Next, a $10\mu m$ thin metal foil is laminated on this frame. Finally, laser micromachining is performed to create microstructure.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

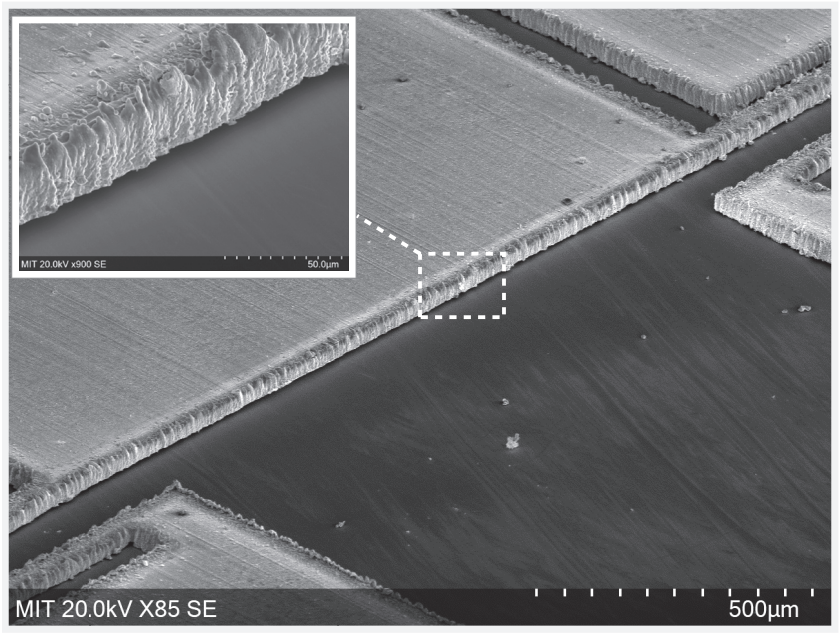



**Figure 5-6:** The 12.5$\mu m$ gap between the freestanding structure and the substrate fabricated using a 12.5$\mu m$ thick adhesive layer.

### 5.2.3 Limit on the Gap Between Laminates

As we saw in section 5.2.2, the gap between the MEMS structure and the substrate can be controlled by controlling the thickness of the deposited sacrificial layer (figure 1-5). As a result, a gap as small as few tens of nanometers can be fabricated using the surface micromachining process. These small gaps are desirable for certain MEMS devices such as electrostatic relays, where a small gap results in low pull-in voltage [119][120].

However, in the MALL process, the gap between the free-standing structure and substrate depends upon the thickness of the adhesive layer. This work demonstrates that a gap as small as 12.5$\mu m$ can be fabricated using a 12.5$\mu m$ thick adhesive layer. Figure 5-6 shows a cantilever fabricated with a 12.5$\mu m$ gap between the structure and substrate. One approach to further reducing the gap can be to spin-coat the adhesive layer directly on the laminate sheet.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

### 5.2.4 Poor Alignment Accuracy

In the conventional microfabrication process, MEMS layers can be fabricated with sub-micron alignment accuracy [121][122]. However, the alignment accuracy in the MALL process is limited to $2.5\mu m$ (figure 1-4). The alignment accuracy can be further improved by utilizing kinematic coupling [123] and passive mechanical alignment features [124].

### 5.2.5 Heat Affected Zone

The laser micro-machined parts suffer from the heat-affected zone (HAZ) [125]. Although the LAMPE micromachining minimizes HAZ by circumventing vaporization and plasma formation, the heat-affected zone can not be completely eliminated. This thesis relies on the scanning electron microscope pictures (such as figure 3-4 and figure 3-11) to estimate the extent of HAZ, which is found to be approximately $1\mu m$. However, additional metrology techniques must be employed to accurately estimate the heat-affected zone.

The heating of the material can lead to a change in material properties. For MEMS with a characteristic length smaller than the heat-affected zone, this change in material property can significantly affect the performance of a system. The heat-affected zone can be minimized in some materials, which have large optical penetration depth, by utilizing ultrashort pulsed laser micromachining [126].

## 5.3 Volume Manufacturing

In deep reactive ion etching (DRIE), many devices can be fabricated in parallel, resulting in a high part production rate. However, the laser direct-write micromachining process used in MALL is a serial process and MEMS are fabricated one device at a time. As a result, the part production rate in MALL process is low. Figure 5-7 compares the part production rate of DRIE and MALL. As can be seen, the MALL process is suitable for low-volume production and requires less initial investments. However, for investments greater than one million dollars, the part production rate of DRIE process increases significantly.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

The part production rate in MALL process can be increased by implementing the following modifications:

1. Using a roll-to-roll lamination process to automate the assembly.

2. Integrating laser-micromachining and roll-to-roll lamination process by delivering the laser over the laminates using galvanometer scanner mirrors. In this system, individual layers of a device can be fabricated simultaneously, and laminated to manufacturer MEM systems. Using laser-micromachining with the roll-to-roll lamination process also eliminates the need for layer alignment, as the patterns can be spatially offset to ensure accurate alignment.

3. Using multiple laser beams to fabricate many devices in parallel. These multiple beams can be either from different laser sources or split from a single high power laser beam.

4. Using high power laser and other precision manufacturing methods such as embossing, micro-milling, knife-cutting, wire-EDM to machine large-features at a high machining rate and reserving the low-etch rate laser-micromachining for small-features.

5. Integrating MEMS packaging with the roll-to-roll lamination process by including package layers in the lamination process. These package layers can be fabricated using embossing, micro-milling, and other conventional fabrication tools

In conventional MEMS fabrication, upgrading from a 150 mm wafer facility to a 300 mm wafer facility is a significant cost. However, in MALL fabrication, adding additional lasers for parallel manufacturing is easy and the cost is incremental. After the initial investment in the laser micromachining system, for every additional \$20,000, an additional laser can be attached to the system for parallel manufacturing (see section **??**). Assuming that a laser micromachining system is $\$100,000$, and a DRIE system is $1M$. For an equal investment of $\$1M$ in MALL, a laser micromachining system with 46 lasers can be built, which can produce 46 devices in parallel. Using such system, for the comb-drive actuator discussed in


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

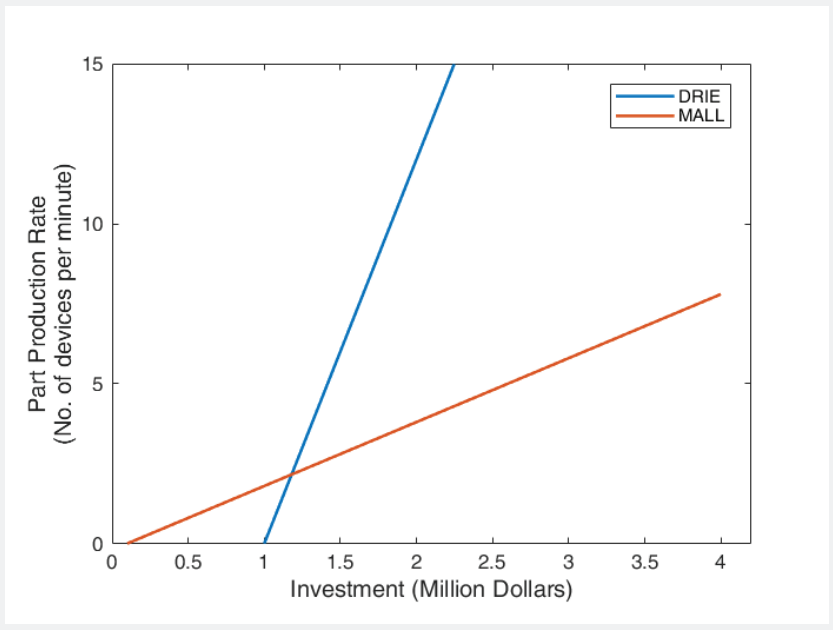



**Figure 5-7:** Part production rate of MEM relay fabrication using DRIE and MALL process.

section section 4.2, 46 devices can be manufactured in $12min$ or 0.26 devices per minute. Thus, the part production rate in MALL can be matched with DRIE.

## 5.4 MEMS for Masses

Micro-electromechanical systems have many potential applications in healthcare, automobile, and consumer electronics. Unfortunately, due to the high cost of equipment ownership, research in MEMS has been primarily limited to industries and universities with microfabrication facilities.

A typical microfabrication facility requires tools such as spin coaters, mask aligners and exposure systems, chemical vapor deposition (CVD) systems for material deposition, and etching tools such as deep reactive ion etching (DRIE). These tools are not only expensive but also require specialized buildings to house them. For example, the photolithography process should be performed in a cleanroom environment to minimize interference of dust


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

particles in exposure process. Similarly, the DRIE process requires specialized gas fittings for etching gases and safety alarms for leak detection. In addition, the chemicals used in lithography-based microfabrication process are expensive, hazardous, and require special storage and disposal.

In recent years, there has been considerable interest in developing table-top microfabrication tools for low-cost fabrication and rapid development of MEMS. For example, focused ion beam (FIB) milling [127][128] is extensively used as a rapid-prototyping tool for microfabrication. However, FIB systems cost approximately a million dollars and have high maintenance cost. Moreover, the gallium filaments used in FIB is consumed rapidly during operation, resulting in high operating cost.

Another approach that is gaining popularity as a table-top microfabrication tool is 3D micro-additive manufacturing [129]. This approach utilizes stereolithography [130] and two-photon polymerization [131] to fabricate MEMS. These processes have the advantage of offering true three-dimensional microfabrication. However, the fabrication material is limited to certain photo-curable polymers. Moreover, the print time is often much longer. Furthermore, for stereolithography, the print resolution is larger than the features required in many MEMS devices, and the build volume of the two-photon polymerization is small for many MEM systems.

As can be seen above, the major obstacle in engaging a larger scientific community in MEMS research is the substantial capital investment required to procure microfabrication tools. In addition, the high cost of equipment maintenance and consumables such as silicon wafer, photoresists, developers, and etchants further hinder the access of MEMS fabrication to a large portion of the scientific community.

In contrast, the MALL process utilizes a laser micromachining system, which has three advantages. First, single equipment replaces a set of tools, mainly spin coater, mask aligner, exposure system, and DRIE system, required for micromachining. Second, the cost of laser micromachining system is significantly less than conventional microfabrication tools. Third, the consumables are limited to electropolishing solution and metal foils. These materials


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

cost \$0.0015 per device, significantly lower than of silicon wafers used in conventional MEMS fabrication.

This thesis uses a commercially available laser micromachining system from Oxford Lasers Ltd [56], which has a high cost of ownership. However, the laser micromachining systems can be built at a much lower cost and used to fabricate MEM systems. The major cost in the Oxford Laser micromachining system comes from the 2-axis motion system and the laser source. The high cost of the 2-axis stage is due to its long-travel range of $100mm$. However, for fabricating MEMS, a travel range of $10mm$ is sufficient. As a result, the expensive 2-axis stage can be replaced with 2-axis galvanometer mirrors (costs around \$2500 [60]) or a piezo linear stage (costs around \$750 per axis [132]). A $5W$ pulsed laser system costs about $\$15,000$ (D. Roland, personal communication, November 6, 2018). The recent development in rapid prototyping of rapid prototyping machines [133][134] has considerably reduced the complexity and cost of machine building. Assuming another \$2500 for focusing optics and machine building, the total cost of the system comes out to be $\$20,000$.

As can be seen above, the major cost in a laser micromachining system comes from the laser source. The cost of the laser source depends upon its average power. A high power laser requires a high-intensity pumping source, resulting in increase in the cost of the laser. For example, the diode-pumped laser used in this work uses an array of light-emitting diode (LED) to pump the laser medium. These diode-arrays require specialized high-current power supply and cooling system, which are not only expensive but also have a large footprint.

The LAMPE micromachining process requires average power in the range of $280mW$ (Table 3.2), thereby needing a laser with significantly low power. The low power lasers require less intense pumping source; as a result, cost less and have a small footprint. The low-cost laser system, in conjunction with low-cost machine building can be used to develop low-cost table-top MEMS manufacturing systems.

The table-top MEMS manufacturing can reduce the cost of MEMS fabrication and enable rapid prototyping, thereby democratizing MEMS fabrication and accelerating the


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

development of novel micro-electromechanical systems.

## 5.5 Thesis Contributions

The key contributions of this thesis are as follows:

1. The LAMPE micromachining process to fabricate high aspect ratio (as large as 10:1) structures with diffraction-limited lateral features. The LAMPE micromachining can be used to process a wide range of materials, including metals, silicon, and diamond. In LAMPE micromachining of metals, laser is used to melt (i.e., induce phase-change) the metal, and then the liquid-metal is expelled by the pressure generated from vapor baubles formed at the laser-metal interface. Similarly, in LAMPE micromachining of silicon, the laser is used to oxidize the silicon (i.e. induce phase-change), which transmits the laser (due to low absorption coefficient), causing further oxidation of silicon beneath. The formed silicon oxide is granular and expelled by pressure generated from vapor baubles formed at the laser-material interface due to partial vaporization of silicon. Likewise, in LAMPE micromachining of diamond, laser is used to convert the diamond into graphite (i.e., induce phase-change), and then the newly formed graphite is oxidized into carbon dioxide in the presence of environmental oxygen.

2. The electro-deburring process to remove the burrs present in laser micromachined parts, thereby enabling their application in MEMS parts.

3. The MALL MEMS fabrication process, which can potentially enable greater material selection and integration, greater design flexibility, easier packaging, low-cost manufacturing, and rapid development of micro-electromechanical systems.

4. An electrostatic comb-drive actuator fabricated from metal. Previously, it was assumed that these parts with such small features and high aspect ratio could be fabricated from silicon alone, using the lithography-based deep reactive ion etching (DRIE) process.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

5. MEM relays for high-current switching application. These relays can be assembled from the elementary conducting and insulating parts and could be used as the logic organ for Von Neumann universal constructor [135].

6. Demonstrating high-aspect-ratio machining of diamond and its application in fabricating diamond rotors for magic-angle spinning nuclear magnetic resonance (MAS-NMR) spectroscopy.

## 5.6 Conclusion

In conclusion, the MALL MEMS fabrication process presented in this thesis has the potential to replace the conventional microfabrication process used for MEMS manufacturing. The MALL process enables MEMS fabrication with a wide range of material, thereby empowering a new paradigm in MEMS design, functionality, and applications. Moreover, the manufacturing cost per device in MALL fabrication is low due to using low-cost fabrication materials such as metal foils. In contrast, in conventional microfabrication, the fabrication material is largely limited to silicon, which restricts the design, functionality, and application of MEMS, and increases the manufacturing cost per device.

The development time and cost in MALL is less due to the independent development of device layers and their effortless integration. The MALL process is well suited for one-of-a-kind and low-demand MEMS market because of the low cost of development. In contrast, creating new MEMS devices in conventional microfabrication requires developing and integrating various material deposition/etching processes, which increases the development time and cost. Thus, even though the cost of manufacturing per device is low, the total cost per device is large. Due to this limitation, the commercialization of MEMS products has been largely limited to markets in which the huge demand can compensate for the high cost of development.

The tools and material cost of MALL fabrication is approximately $\$25,000$, which is


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

affordable to a wider scientific community. This democratization of MEMS fabrication can expedite the development of novel MEMS systems. Traditionally, the research and development of novel MEMS devices have been largely limited to a few industries and universities, due to the high cost of equipment ownership.

## 5.7 Recommendations for Future Work

The recommendations for future work are listed below:

1. This work relies on optical microscope image and X-ray dispersion spectroscopy (EDS) analysis to conclude oxidation of silicon during LAMPE micromachining. The laser-assisted oxidation of silicon can be further confirmed by performing LAMPE micromachining in an inert environment. My hypothesis is that in an inert environment, the LAMPE micromachining of silicon should not result in any material removal, due to the absence of oxidation.

2. To further increase the aspect ratio in LAMPE micromachining of silicon, the I recommend performing LAMPE micromachining in $HF$ vapor environment. My hypothesis is that the $HF$ vapor will react with the newly formed $SiO_2$ to form colorless gaseous silicon tetrafluoride $SiF_4$, which would be removed efficiently. As a result, the depth of the micromachining can be further increased by gradually moving the focus downwards as the material is removed.

3. An x-ray diffraction measurement is recommended to verify the graphitization of diamond. Moreover, the oxidation of graphite into $CO_2$ in the oxygen environment can be verified by performing LAMPE micromachining in an inert environment. My hypothesis is that in the inert environment, the LAMPE micromachining of diamond would not lead to any material removal due to the absence of oxidation of graphite.

4. This work uses a 3-axis stage to move the machining target with respect to the laser. However, using a galvanometer scanner is recommended to increase the LAMPE


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

micromachining rate.

5. For the MEM relay discussed in section 4.3, it is recommended to apply a conformal coating of a dielectric material such as parylene on the gate electrode to isolate it from the actuator electrode. The application of parylene coating is expected to reduce the pull-in voltage and increase the maximum safe operating voltage.

6. For materials with low absorption coefficient such as silicon and diamond, it is recommended to use ultra-short laser (in the order of femtoseconds to picoseconds) for LAMPE micromachining. The ultra-short laser is expected induce phase-change while minimizing the heat-affected zone (HAZ). However, it must be noted that the LAMPE micromachining of metals requires long pulses in the order of nanoseconds. In metals, the optical penetration depth is very small (a few nanometer) and the metal is melted through heat conduction. As a result, long pulses in the order of nanoseconds are required to dump sufficient energy to induce the melting and expulsion.

7. For machining films with thickness less than $10\mu m$, it is recommended to use ultra-short pulsed lasers irrespective of the absorption coefficient of the material. In ultra-short pulsed laser micromachining, the micromachining depth per pulse is equal to the optical penetration depth of the material. The high pulse energy and repetition rate ( $100KHz$) lead to sufficient number of pulse irradiation per spot, resulting in machining depth in the order $10\mu m$.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Appendix A

## A.1 Manufacturing Cost Estimates for DRIE and MALL

The cost of producing one device is given by the expression:

$$Cost\ per\ device = Material\ cost\ per\ device + Equipment\ cost\ per\ device \tag{A.1}$$

The material consumable cost can be calculated by adding the cost of wafer, etching gas, and electricity etc. Due to lack of reliable information about the etching gas and electricity consumption, only wafer cost is included in the presented calculation.

The material cost per device is given by the expression:

$$Material\ cost\ per\ device = \frac{Cost\ of\ wafer \times Area\ of\ device}{Area\ of\ wafer} \tag{A.2}$$

The equipment cost per part can be calculated by dividing the equipment cost with total number of parts produced in the lifespan of the equipment and given by the expression:

$$Equipment\ cost\ per\ device = \frac{Cost\ of\ Equipment}{Device\ production\ rate \times Equipment\ lifetime} \tag{A.3}$$


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

The total number of parts produces is given by the expression:

$$Number\ of\ parts\ = Parts\ produced\ per\ dollar \times Total\ investment \tag{A.4}$$

The *Cost per device* and *Device production rate* depends upon the design of the device. The calculations presented here are for the comb-drive actuator discussed in section 4.2. The assumptions and values of different variables used for calculations are listed in Table A.1

### A.1.1 Cost of Manufacturing Using DRIE

#### Part Production Rate

It is assumed that the device is fabricated on SOI wafer and DRIE is used to etch the comb-drive structures. The etch rate is $1\mu m/min$ and the device thickness is $100\mu m$. Therefore, it will take $100min$ to process a wafer (neglecting wafer loading/unloading time).

The length of the comb-drive actuator is $6.5mm$ and the width is $4.5mm$. Assuming $150mm$ diameter SOI wafers are used for production, and $500\mu m$ margin is kept around the device for dicing the wafer, the total number of devices produced per wafer is 428. As a result, number of devices produced per minute or production rate is 4.28 devices per minute.

$$DRIE\ Part\ Production\ Rate = 4.28\ devices/min$$

#### Manufacturing Cost

Assuming that the lifespan of DRIE system is 5 years or 2628000 minutes, the total number of the devices that can be produced during the lifespan of the system is 11 million parts. If the cost of the equipment is $\$1M$, the equipment consumable cost per device is \$0.09

Assuming SOI wafer is used for fabrication and each wafer costs \$220 [114]. Since 428 devices can be fabricated per wafer, the material cost per device is $\$220/428 = \$0.51$. The total cost of fabrication per device using DRIE is \$0.6, and 6.94 devices can be produced with one dollar. As can be seen, 85% of the total cost is due to material cost.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

$$DRIE\ Cost\ per\ Device = \$0.51$$

### A.1.2 Cost of Manufacturing Using MALL

#### Part Production Rate

The laser micromachining process is a serial process and the part production rate depends upon the actual design of the device. Moreover, depending upon the design, laser bulk micromachining (LBM) can be employed to rapidly remove the material from the region with large features, while LAMPE micromachining can be employed for region with small features and high aspect ratio. As a result, unlike DRIE, the part production rate in MALL depends upon the actual design of the MEMS device.

To calculate the MALL manufacturing cost, the comb-drive design discussed in section 4.2 is used. The device is fabricated by first removing the excess material around the beam springs and interdigitated fingers using LBM, followed by LAMPE micro-machining interdigitated fingers. The etch rate of the laser bulk micromachining is given by the equation:

$$E_{LBM} = \eta A_{\phi} A_h F$$

where $\eta$ is .25, $A_{\phi}$ is diameter of ablation profile, $A_h$ is depth of the ablation profile, and $F$ is feed rate. The expression for feed rate is given by the equation:

$$F \leq \eta A_{\phi} F_p$$

For laser bulk micromachining copper in fabricating comb-drive actuator, $\eta = 0.25$, $A_{\phi} = 25\mu m$, $A_h = 25\mu m$, and $F_p = 10KHz$. As a result,

$F_{max} = 100mm/s$

The practical limit on achieving feed-rate depends upon the acceleration/deceleration rating and inertia of the motion system. The laser micromachining system used in this


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

work has acceleration/deceleration rating of $10,000 mm/s^2$ and large inertia; as a result, the maximum achievable feed-rate for MEMS length-scale distance is of the order of $1 mm/sec$. However, using galvanometer mirrors, much higher feed rate can be achieved. A typical 2-axis galvanometer mirror can operate at $100 Hz$ frequency [60]. Assuming, the scan distance of $1mm$, a feed rate of $100 mm/s$ can be easily achieved. For the presented calculation, a feed rate of $50 mm/sec$ is used, to account for acceleration/deceleration limits of galvanometer mirrors, resulting in material removal rate of $E_{LBM} = 0.29 mm^3/min$. For fabricating the comb-drive actuator, about 18% of the device volume is removed, which takes approximately 1.8 minutes.

Finally, the comb-drive fingers are fabricated using LAMPE micromachining. The machining length is $24.6mm$ and cutting feed rate is $2.4 mm/min$ (or $40 \mu m/s$); therefore it requires $10.2 min$ to LAMPE micromachine the comb-drive actuator. The total time required to fabricate a comb-drive actuator in MALL process is 12 minutes. The part production rate is:

$$MALL\ Part\ Production\ Rate = 0.08\ devices/min$$

**Manufacturing Cost**

Assuming the lifespan of laser micromachining system is 5 years or 2628000 minutes, the total number of the devices that can be produced during the lifespan of the system is 0.22 million parts. Assuming the cost of the laser micromachining is $\$100,000$, the equipment cost per device comes out to be \$0.46.

The material for MALL fabrication is $100 \mu m$ thick copper foil. A roll of copper foil $6^{"}$ wide & $268 ft$ long cost \$449 [136] and can accommodate 297040 devices; therefore, the material cost for each device is \$0.0015. The total cost of device fabrication is:

$$MALL\ Cost\ per\ Device = \$0.4615$$


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

| Variable | Value |
|---|---|
| **DRIE** | |
| Cost of DRIE | \$1M |
| Lifetime of DRIE | 5 years |
| Yield | 100% |
| DRIE etch rate | $1\mu m/min$ |
| Wafer loading/unloading time | $10min$ |
| Cost of SOI wafer [114] | \$220 |
| Diameter of SOI wafer | $150mm$ |
| Comb-drive area | $6.5mm \times 4.5mm$ |
| Comb-drive thickness | $100\mu m$ |
| Dice-saw margin | $500\mu m$ each side |
| **MALL** | |
| Cost of laser micromachining system | $\$100,000$ |
| Lifetime of laser micromachining system | 5 years |
| Cost of copper foil ($6" \times 268ft$ )[136] | \$449 |
| Comb-drive critical machining length (for LAMPE Micromachining) | $24.4mm$ |
| Percentage of total volume removed using LBM micromachining | 18% |

**Table A.1:** The values of variables used for calculating manufacturing cost.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Appendix B

Here, I describe the MALL fabrication process in detail.

## B.1 Equipment Required for MALL Fabrication

Below is the list of tools required for MALL MEMS fabrication:

1. A laser micromachining system. This thesis uses a laser micromachining system from Oxford Laser Ltd [56]. (I am interested in building a low-cost DIY laser micromachining system in future. Please check http://mtm.cba.mit.edu/ for updates.)

2. Depending upon the feature size of the device, a scanning electron microscope or a high resolution optical microscope is required to inspect the MEMS device during fabrication. A low magnification stereo microscope is also useful for quick inspection.

3. A chemical hood and a voltage source is required to perform electro-deburring process.

4. A critical point dryer may be required. In my experience, for MEMS fabricated with metals, dipping the device in isopropanol (IPA) and drying on a hot plate is sufficient. However, for MEMS fabricated with silicon and containing small features, a critical point dryer is desirable.

5. A hot plate is useful for quickly drying the wafers.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

6. A toaster oven with a temperature controller is required for certain adhesive that need curing at higher temperature. Particularly, the Dupont FR1500 adhesive used in MEM relay, require baking at 185 degree centigrade. Other adhesive that is used in this work is 3M thermal bonding film 583, which can be either solvent activated or pressure activated, and does not requires baking.

7. A vacuum tweezer is desirable to facilitate manipulating the delicate sheets after laser micromachining process.

8. Depending upon the MEMS device, addition tools may be required for device testing. For example, in this thesis, the MEM relay is tested using a semiconductor parameter analyzer. A probe station is also required for electrical testing of the device.

9. For certain MEMS devices, a table-top precision milling machine can be useful. The bottom layer of MEM relay discussed in section 4.3 is fabricated using a table-top precision milling machine.

Below is the list of materials required for MALL fabrication:

1. The main material used in MALL MEMS fabrication is metal foils. These metal foils can be sourced from https://www.mcmaster.com/.

2. The main adhesive layers used in this work are Dupont FR1500, 3M thermal bonding film 583, and scotch double sided tape. These adhesives can be sourced from there corresponding vendors.

3. Electrolyte for electro-deburring process. For copper, either copper sulfate or phosphoric acid can be used.

4. Dowel pins or gauge pins for using them as alignment pins.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

## B.2 MEMS Design

In this thesis, I use Fusion 360 for designing MEMS. The software offers the advantage of integrating the computer aided design (CAD) and computer aided manufacturing (CAM), and is well suited for MALL MEMS design. Since the MALL process is a laminate process, it is advantageous to design the MEMS system in form of layers. Once the design is finished, the g-code for laser toolpath can be exported using the CAM interface of the Fusion 360 program. This g-code program is used for performing the laser micromachining.

Depending upon the MEMS device design it is advisable to export two different laser toolpath. The first toolpath is for the part of the region where small feature sizes are not required and high power laser can be used to remove the material at high micromachining rate. The second toolpath is for parts of the MEMS design which have small features, and require LAMPE micromachining process.

## B.3 Fabrication

### B.3.1 Laser Micromachining

In this thesis, the laser micromachining is performed using a commercially available laser micromachining system from Oxford Laser Ltd [56]. Following are steps for

S1: The metal foil is clamped to the stage using a clip. For metal foils with thickness up to $20\mu m$, tweezers can be used for manipulations. However, for thickness below $20\mu m$, it is desirable to attach the metal foil with to a rigid frame for easy manipulation.

S2: Selecting the laser source: For metals, the laser micromachining process is generally independent of the laser wavelength. This is due to the fact that the absorption of laser by the metal is low in the UV, Vis, and IR range. The pulse duration of the laser is essential to get the small feature size. Use equation (3.5) to determine the required pulse width for a desired feature size. Often the laser pulse width is fixed, and the




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

user only has control over the pulse power. The laser system used in this work has fixed pulse duration of 20 $ns$.

S3: Selecting the focusing optics: Depending upon the feature size, the desired lens or objective is installed. The laser micromachining system is equipped with a digital camera for sample inspection. Next, the sample is brought into the focal plane of the lens by inspecting in the imaging system.

S4: Selecting laser power and process development: Use equation (3.6) to determine the required laser power for the laser micromachining. A series of power tests in the vicinity of this calculated laser power can be used to determine the exact value of the power which gives the minimum feature size. The laser micromachining should be performed at this value. It must be noted that, the feature size also depends upon the degree to which the sample is at the focal plane. Usually, before performing the power test, I perform a series of focus tests to determine the best focus. Sometimes a second pass might be required if the laser micromachined part does not fall off during the laser micromachining process.

### B.3.2 Electro-deburring

The steps for performing electro-deburring are as follows:

S1: Selecting the electrolyte: The choice of electrolyte depends upon the deburring metal. Since the eletro-deburring process is similar to electropolishing and electroplating process, the choice of electrolyte is same as of that used for electropolishing or electroplating. There is a great body of work on available electrolytes for various metals. In this work, copper is used for fabricating MEMS. For copper, 14M phosphoric acid is used as the electrolytic solution.

S2: Electro-deburring setup: The laser micromachined copper part is connected to anode and another copper sheet is connected to the cathode. Both electrodes are dipped inside 4M phosphoric acid solution.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

S3: Selecting the electro-deburring voltage: Setting the right voltage is essential to remove burrs. As can be seen in figure 3-9, the electro-deburring voltage must be greater than the electropolishing temperature to ensure the removal of the burrs. The electropolishing process is well studied and the value of the electropolishing voltage for a given metal and electrolyte can be found in the literature. The electro-deburring voltage should be above this value. In this work, the electro-deburring of laser micromachined parts is performed at $5V$ for 30 seconds. Generally, the electro-deburring voltage is a little less than the votlage at which the bubbles starts to form at the anode.

S4: For some MEMS devices, a gold electroplating may be desirable to prevent oxidation of the metal.

S5: After the electro-deburring, the laser micromachined part can be rinsed in DI water followed by isopropanol rinse. The cleaned laser micromachining part can be either air dried at room temperature or at 85 degree centigrade inside an oven.

### B.3.3 Lamination Process

The steps for lamination are as follows:

S1: For the lamination process, the adhesive layer is first tack bonded to the substrate layer. For tack bonding the Dupont FR1500 adhesive layer, the curing is performed at $120°C$.

S2: Once the adhesive is tack bonded to the substrate, the backing layer is removed and the top device layer is attached. Alignment pins are used for the alignment. For devices requiring high alignment accuracy, alignment pins with diameter of $500\mu m$ is used to improve the alignment accuracy. However, for the device which does not require a high alignment accuracy, thicker alignment pins (for example, $1.5mm$ diameter dowel pins) can be used to simplify the assembly process.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

S3: Once aligned, the laminate structure is sandwiched between a custom-made alignment fixture and clamped using a paper clip and put inside the oven for full curing. The temperature of the oven is set to $185°C$, and the baking is performed for $5-20$ minutes depending upon the device and adhesive layer thickness.

S4: After the lamination process, the MEMS device can be cut-out from the laminate structure using laser cutting. However, this step is not required during the device development.

## B.4 MEMS Device Testing

The actual steps for testing the device depend upon the actual design of the device. Here, I outline the general steps which are common to many MEMS devices:

S1: After the device fabrication, it is advised to inspect the device in an optical microscope to check the integrity of the device.

S2: Care must be taken to handle the MEMS devices, as they are very fragile. Do not blow air to dry or remove dust particles, it will break the device.

S3: For electrical testing, use probe station probes to make the electrical connection. A wire-bonder can be used as well. In my experience, for devices made out of copper, using gold ball-bonding at $100°C$ works very well.

## B.5 Things That did Not work

Here, I list all the failed experiments and things that did not work:

1. First time observed high diffraction-limited features while trying to machine $50\mu m$ silicon wafer. Unfortunately, the success was very short lived, the wafer was very fragile and the device broke. Figure B-1(a) shows the the broken comb-drive structure.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

2. Decided to use brass as it is less fragile and cheaper. However, the part had much carbon deposited as shown in figure B-1(b).

3. Tried all sorts of way to remove the carbon and burrs with no success. Tried physical scrubs, but the MEMS parts were very fragile and would deform. Tried sonication but it won't remove the burrs. Tried prolonged sonication, and broke many devices. Figure B-1(c) shows once such device. Concluded sonication can not be used to remove the burrs.

4. Tried many adhesive layers with no success. The Dupont FR1500 would stick but the bonding was not sufficient and the device layer will come off as shown in figure B-1(d). It turned out that I wasn't baking it at right temperature. The right baking temperature is $185°C$.

5. Decided to make comb-drive device in a nitride wafer by patterning the top layer and then performing a potassium hydroxide etch. Figure B-5(a) shows the device after laser micromachining. Unfortunately, the device did not survive the etch step (figure B-5(b))

6. I tried making 2.5D structure in silicon using laser micromachining, however, the surface roughness was very high as shown in figure

7. I tried to use normal pulsed laser ablation micromachining to fabricate comb-drive actuators. Despite of spending months, it did not work. Either the distance between the comb-drive was too large ($\sim 30 = 40\mu m$) or there was some material left between the comb-drive fingers (figure B-2(a)) which resulted in a short-circuit between the fingers. Any amount of electro-deburring would not remove this left-over material. The LAMPE micromachining ensures complete removal of material, while maintaining the small lateral features.

8. After using LAMPE micromachining, I was able to fabricate a device, in which, there was no short-circuit between the comb-drive. I tested the device and the fingers moved.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

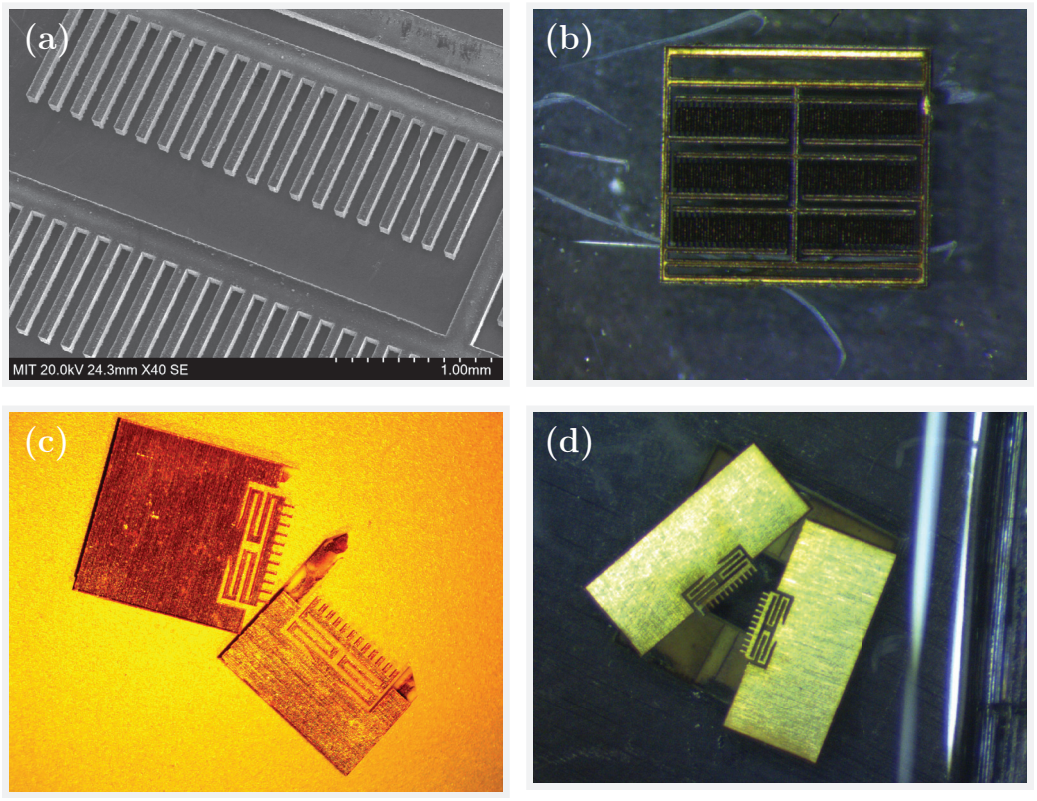



**Figure B-1:** (a) Broken comb-drive structure device fabricated in $50\mu m$ silicon wafer. (b) Comb-drive in $250\mu m$ brass. The micromachined part had much carbon deposits. (c) Broken device after 5 minutes of sonication to try to remove burrs. The sonication was carried in a brass cleaner, which is a mild enchant. (d) Broken device due to insufficient adhesive of the device layer to the substrate layer.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

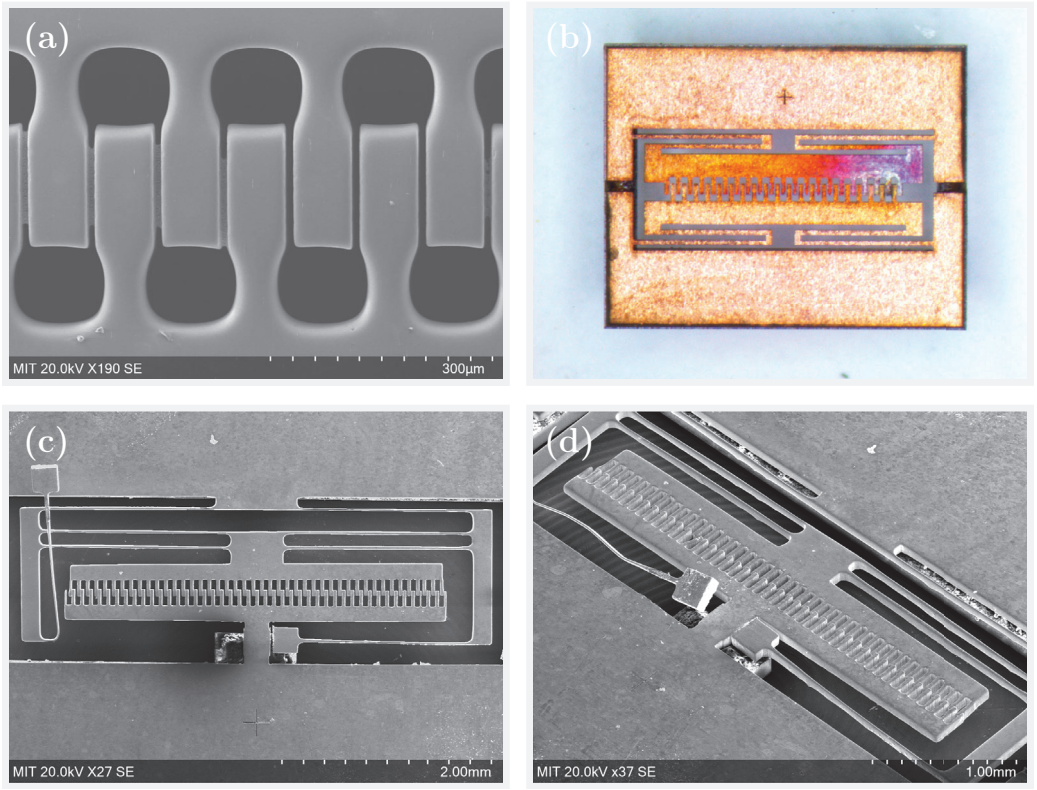



**Figure B-2:** (a) Material left between the comb-drive fingers when normal pulsed laser ablation micromachining is used. (b) Comb-drive actuator after the short-circuit due to side-pull. (c) & (d) Broken device because the beam-spring will come off from the post. It required increasing the pressure during the lamination and baking process to make sure the beam spring is securely bonded to the post.

However, as usual the success for short lived. The design of the beam spring was not suitable for the comb-drive, and could not resist the side-pull. After a certain voltage, the fingers got connected resulting in short circuit. Figure B-2(b) shows the first device that worked and then resulted in short circuit.

9. Changed the design and used folded-flexure springs to minimize the side-pull. However, the spring won't stick to the post as shown in figure B-2(c). It required increasing the pressure during laminate assembly and baking process to make the beam spring stick to the post.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

10. I tried multiple device design for MEM relays. Figure B-3(a) shows the cantilever design. This was fabricated by laminating the top cantilever laminate on a silicon substrate. I decided not pursue this direction because it required silicon substrate, however, I wanted to show device demonstration with metal. Moreover, it required depositing gold electrodes on the substrate. Furthermore, I was worried about the stiction problem.

11. I also tried $10\mu m$ thick silver layer attached to $200\mu m$ thick silicon frame for the actuator electrode. Figure B-3(b) shows the fabricated device. I decided to not pursue it for the same reasons as above. It used silicon wafer, and required deposited metal electrodes on the bottom substrate.

12. Another design that I tried is shown in figure B-3(c). It did not work because the cantilever was making electrical contact with the gate and output electrode simultaneously. The MEM relay shown in figure B-3(d) has similar design and failed due to the same reason.

13. I spent a large amount of time trying to solve the problem of actuator electrode making electrical connection with the gate electrode. First, I tried reducing the height of the gate electrode below the input/output electrode. I was hoping that the actuator electrode will make contact with the input/output and will be blocked to further move and make connection with the gate electrode (figure B-4(a)). However, I was wrong. The actuator electrode plate is thin and can easily bend to cause electrical connection with the gate electrode.

14. Next, I tried bringing the input/output electrode closer to the center to minimize the bending of the cantilever plate (figure B-4(b)). However, it was not sufficient, and the actuator electrode was still making the electrical connection with the gate electrode.

15. I also tried attaching a stop layer on the cantilever as shown in figure B-4(c). However, when the voltage was applied, it made the contact with gate electrode first, thereby,


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

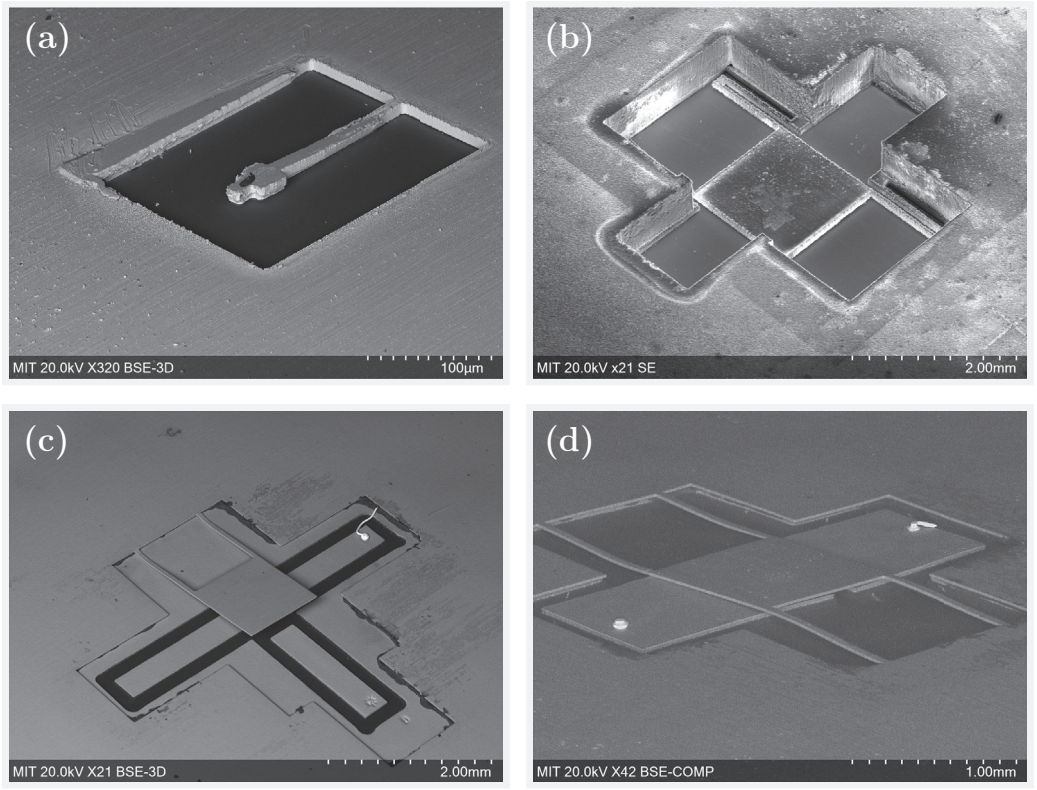



**Figure B-3:** (a) MEM Relay with cantilever design. I decided to not pursue this direction as it required deposited electrodes on the bottom silicon substrate. Moreover, I was worried about the stiction. (b) MEM relay design using $200\mu m$ thick silicon wafer as the frame. I decided to not pursue it because I wanted to show silicon-less MEMS fabrication. Moreover, it required depositing gold electron on the bottom substrate. (c) MEM relay with cantilever design. This design failed because the cantilever was making electrical connection with both the gate electrode and drain electron simultaneously. (d) Another MEM relay design which failed due to similar reason.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

hindering the connection of actuator contact with the input/output contact.

16. I tried $25\mu m$ thick layer of corona dope (MG chemicals) to isolate the actuator electrode with the gate electrode B-4(d). To make the electrical contact with the gate electrode, the corona dope layer is removed from the edges as shown in the figure B-4(d). It worked very well. However, the pull-in voltage was between $500 - 600V$. Finally, I settled with the design discussed in section 4.3.


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

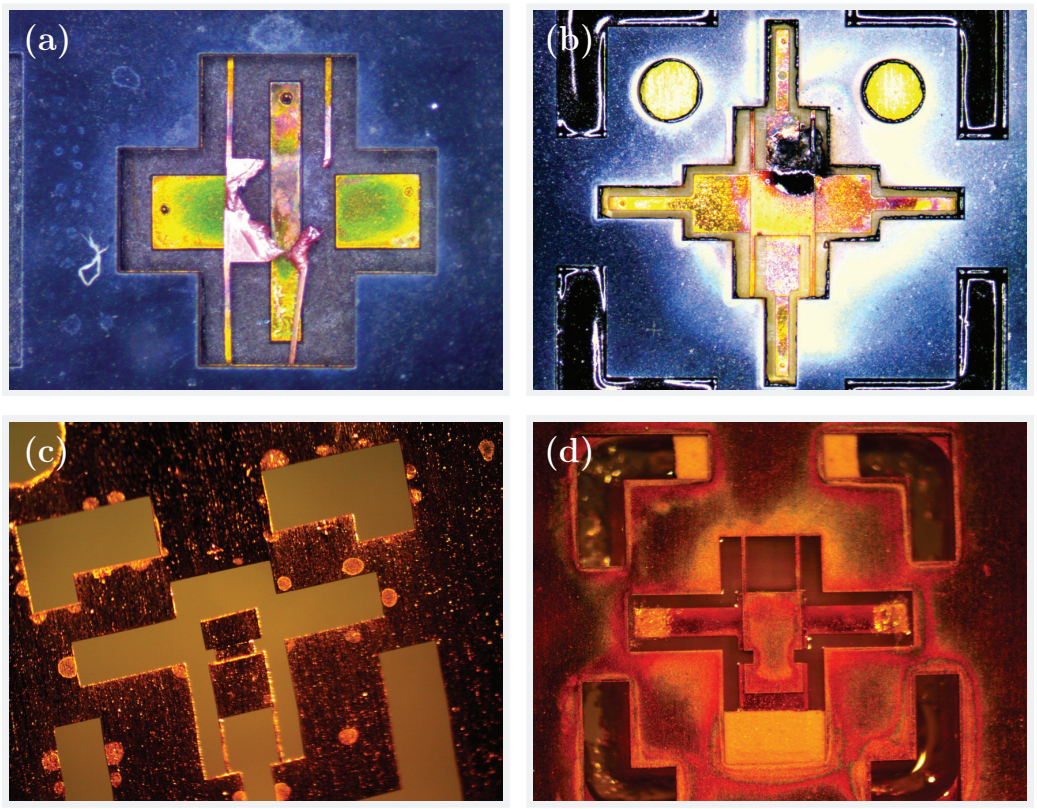



**Figure B-4:** (a) MEM relay design which broke due to the short circuit between the actuator electrode and the gate electrode (b) Another MEM relay design which broke due to similar reasons. (c) MEM relay design in which a post was attached to the cantilever to isolate actuator and the gate electrode. Unfortunately, the post made contact with electrode, thereby, hindering the connection between movable contact and the input/ouput contact (d) MEM relay design using a $25\mu m$ thick corona dope (MG chemicals) layer to isolate gate and actuator electrode. The device worked, however, the pull-in voltage was between $500 - 600V$ due to large thickness of the insulating layer.




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

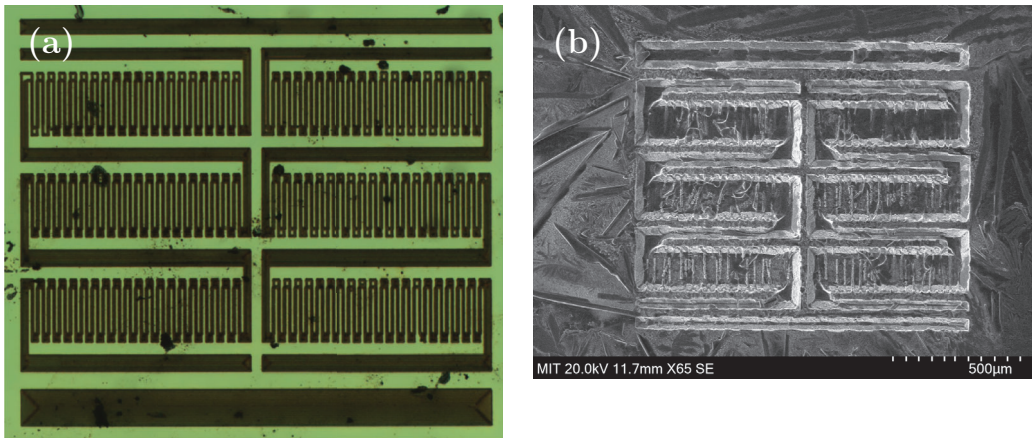



**Figure B-5:** (a) Comb-drive structure Laser micromachined on nitride wafer (b) Comb-drive device after potassium hydroxide etch




DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# Bibliography

[1] D. Niarchos, "Magnetic MEMS: key issues and some applications," *Sensors and Actuators, A: Physical*, vol. 109, no. 1-2, pp. 166–173, 2003.

[2] S. D. Senturia, *Microsystem Design*. Springer Science & Business Media, 2007.

[3] R. Ramadoss, S. Lee, Y. C. Lee, V. M. Bright, and K. C. Gupta, "Fabrication, Assembly, and Testing of RF MEMS Capacitive Switches Using Flexible Printed Circuit Technology," *IEEE Transactions on Advanced Packaging*, vol. 26, no. 3, pp. 248–254, 2003.

[4] M. Weber B S and L. C. Kretly, "PCB-MEMS RF switch: Parametric analysis and design guide lines," *SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference Proceedings*, pp. 141–144, 2007.

[5] H. P. Chang, J. Qian, B. A. Cetiner, F. De Flaviis, M. Bachman, and G. P. Li, "RF MEMS switches fabricated on microwave-laminate printed circuit boards," *IEEE Electron Device Letters*, vol. 24, no. 4, pp. 227–229, 2003.

[6] M. A. Mapar and M. Omidian, "Laminate MEMS for Heterogeneous Integrated Systems," *MRS Online Proceedings Library Archive*, vol. 1427, 2012.

[7] J. Tsai, A. Y. Zhang, G. P. Li, and M. Bachman, "A laminate cantilever waveguide optical switch," *Proceedings - Electronic Components and Technology Conference*, pp. 203–207, 2012.



569
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[8] M. Bachman and G.-P. Li, "MEMS in laminates and package substrates," *2013 IEEE International Symposium on Advanced Packaging Materials*, vol. 1, no. Paper 32, pp. 121–125, 2013.

[9] J. P. Whitney, P. S. Sreetharan, K. Y. Ma, and R. J. Wood, "Pop-up book MEMS," *Journal of Micromechanics and Microengineering*, vol. 21, no. 11, 2011.

[10] H. McClintock, F. Z. Temel, N. Doshi, J.-S. Koh, and R. J. Wood, "The milliDelta: A high-bandwidth, high-precision, millimeter-scale Delta robot," *Science Robotics*, vol. 3, no. 14, p. eaar3018, 2018.

[11] M. Bachman and G. P. Li, "Laminates for MEMS and BioMEMS," *14th International Conference on Electronic Materials and Packaging, EMAP 2012*, no. figure 2, 2012.

[12] A. S. Holmes, "Laser processes for MEMS manufacture," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 4426, no. 43, pp. 203–209, 2002.

[13] M. C. Gower, "Laser micromachining for manufacturing MEMS devices," *MEMS Components and Applications for Industry, Automobiles, Aerospace, and Communication*, vol. 4559, no. October 2001, pp. 53–59, 2001.

[14] M. El-Bandrawy and M. Gupta, "Femtosecond laser micromachining of silicon for MEMS," *Proc. SPIE 4977, Photon Processing in Microelectronics and Photonics II*, vol. 4977, no. October 2003, pp. 219–225, 2003.

[15] C. Abbott, R. M. Allott, B. Bann, K. L. Boehlen, M. C. Gower, P. T. Rumsby, I. Stassen Boehlen, and N. Sykes, "New techniques for laser micromachining MEMS devices," *High-Power Laser Ablation IV*, vol. 4760, no. September 2002, pp. 281–288, 2002.

[16] M. R. H. Knowles, G. Rutterford, D. Karnakis, and A. Ferguson, "Micro-machining of metals, ceramics and polymers using nanosecond lasers," *International Journal of Advanced Manufacturing Technology*, vol. 33, no. 1-2, pp. 95–102, 2007.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[17] H. K. Tönshoff, C. Momma, A. Ostendorf, S. Nolte, and G. Kamlage, "Microdrilling of metals with ultrashort laser pulses," *Journal of Laser Applications*, vol. 12, no. 1, pp. 23–27, 2000.

[18] N. H. Rizvi, "Femtosecond Laser Micromachining: Current Status and Applications," *Optics Communications*, vol. 114, no. 1-2, pp. 5660–7, 2002.

[19] R. A. Lawes, A. S. Holmes, and F. N. Goodall, "The formation of moulds for 3D microstructures using excimer laser ablation," *Microsystem Technologies*, vol. 3, no. 1, pp. 17–19, 1996.

[20] J. Yeom, Y. Wu, J. C. Selby, and M. A. Shannon, "Maximum achievable aspect ratio in deep reactive ion etching of silicon due to aspect ratio dependent transport and the microloading effect," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, vol. 23, no. 6, p. 2319, 2005.

[21] P. Jaecklin, C. Linder, N. F. D. Rooij, and R. Vuilleumier, "Optical Microshutters and Torsional Micromirrors for Light Modulator Arrays," *Cycle*, pp. 124–127, 2007.

[22] K. C. Yung, S. M. Mei, and T. M. Yue, "Rapid prototyping of polymer-based MEMS devices using UV YAG laser," *Journal of Micromechanics and Microengineering*, vol. 14, no. 12, pp. 1682–1686, 2004.

[23] W. C. Tang, T. C. H. Nguyen, and R. T. Howe, "Laterally driven polysilicon resonant microstructures," *Micromechanics and MEMS: Classic and Seminal Papers to 1990*, vol. 20, pp. 187–193, 1997.

[24] L. Lin, C.-C. Nguyen, R. Howe, and A. Pisano, "Microelectromechanical filters for signal processing," *Proceedings IEEE Micro Electro Mechanical Systems*, vol. 7, no. 3, pp. 226–231, 1992.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[25] J. N. Palasagaram and R. Ramadoss, "MEMS-capacitive pressure sensor fabricated using printed-circuit-processing techniques," *IEEE Sensors Journal*, vol. 6, no. 6, pp. 1374–1375, 2006.

[26] B. J. Janney, J. J. Christilda, and S. S. Mary, "MEMS Accelerometer Fabricated Using Printed Circuit Processing Techniques," in *IEEE International Symposium on Industrial Electronics*, pp. 3250–3254, 2007.

[27] M. Bachman, S. J. Kim, G.-P. Li, Y. Zhang, and M. Wang, "Latched microswitches in laminates for high power 0–6.5 GHz applications," *Electronics Letters*, vol. 50, no. 14, pp. 1010–1012, 2014.

[28] J. Henri, G. Han, B. Meint de, E. Miko, and F. Jan, "A survey on the reactive ion etching of silicon in microtechnology," *Journal of Micromechanics and Microengineering*, vol. 6, no. 1, p. 14, 1996.

[29] F. Marty, L. Rousseau, B. Saadany, B. Mercier, O. Français, Y. Mita, and T. Bourouina, "Advanced etching of silicon based on deep reactive ion etching for silicon high aspect ratio microstructures and three-dimensional micro-and nanostructures," *Microelectronics journal*, vol. 36, no. 7, pp. 673–677, 2005.

[30] A. Ayon, R. Bayt, and K. Breuer, "Deep reactive ion etching: a promising technology for micro-and nanosatellites," *Smart materials and structures*, vol. 10, no. 6, p. 1135, 2001.

[31] K. Fujimoto, N. Maeda, H. Kitada, Y. Kim, A. Kawai, K. Arai, T. Nakamura, K. Suzuki, and T. Ohba, "Development of multi-stack process on wafer-on-wafer (wow)," in *2010 IEEE CPMT Symposium Japan*, pp. 1–4, IEEE, 2010.

[32] A. Luque, J. M. Soto, F. Perdigones, C. Aracil, and J. M. Quero, "Electroosmotic impulsion device for integration in PCB-MEMS," *Proceedings of the 2013 Spanish Conference on Electron Devices, CDE 2013*, pp. 119–122, 2013.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[33] M. Bachman and G. P. Li, "MEMS in laminates," *Proceedings - Electronic Components and Technology Conference*, pp. 262–267, 2011.

[34] M. Bachman and G. Li, "Integrated mems in package," *Circuit World*, vol. 38, no. 4, pp. 184–192, 2012.

[35] Y. Shim, J. Ruan, Z. Wu, and M. Rais-Zadeh, "An integrated rf mems tunable filter," in *2012 IEEE 25th International Conference on Micro Electro Mechanical Systems (MEMS)*, pp. 15–18, IEEE, 2012.

[36] P. Goel, "Mems non-silicon fabrication technologies," *Sensors & Transducers*, vol. 139, no. 4, p. 1, 2012.

[37] M. Bachman and G.-P. Li, "Laminates for mems and biomems," in *2012 14th International Conference on Electronic Materials and Packaging (EMAP)*, pp. 1–5, IEEE, 2012.

[38] M. Maddela, R. Ramadoss, and R. Lempkowski, "A mems-based tunable coplanar patch antenna fabricated using pcb processing techniques," *Journal of Micromechanics and Microengineering*, vol. 17, no. 4, p. 812, 2007.

[39] A. Luque, G. Flores, F. Perdigones, D. Medina, J. Garcia, and J. Quero, "Single axis accelerometer fabricated using printed circuit board techniques and laser ablation," *Sensors and Actuators A: Physical*, vol. 192, pp. 119–123, 2013.

[40] J. Rogers, R. Ramadoss, P. Ozmun, and R. Dean, "A microelectromechanical accelerometer fabricated using printed circuit processing techniques," *Journal of Micromechanics and Microengineering*, vol. 18, no. 1, p. 015013, 2007.

[41] A. Cardenas-Valencia, J. Dlutowski, S. Knighton, C. Biver, J. Bumgarner, and L. Langebrake, "Aluminum-anode, silicon-based micro-cells for powering expendable mems and lab-on-a-chip devices," *Sensors and Actuators B: Chemical*, vol. 122, no. 1, pp. 328–336, 2007.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[42] J. P. Whitney, P. S. Sreetharan, K. Y. Ma, and R. J. Wood, "Pop-up book MEMS," *Journal of Micromechanics and Microengineering*, vol. 21, no. 11, 2011.

[43] P. A. York, N. T. Jafferis, and R. J. Wood, "Meso scale flextensional piezoelectric actuators," *Smart Materials and Structures*, vol. 27, no. 1, p. 015008, 2017.

[44] J. Lessing, A. C. Glavan, S. B. Walker, C. Keplinger, J. A. Lewis, and G. M. Whitesides, "Inkjet printing of conductive inks with high lateral resolution on omniphobic "rf paper" for paper-based electronics and mems," *Advanced Materials*, vol. 26, no. 27, pp. 4677–4682, 2014.

[45] J. A. Lewis, "Direct-write assembly of ceramics from colloidal inks," *Current Opinion in Solid State and Materials Science*, vol. 6, no. 3, pp. 245–250, 2002.

[46] C. W. Visser, R. Pohl, C. Sun, G.-W. Römer, B. Huis in 't Veld, and D. Lohse, "Toward 3d printing of pure metals by laser-induced forward transfer," *Advanced materials*, vol. 27, no. 27, pp. 4087–4092, 2015.

[47] C. G. K. Malek, "Su8 resist for low-cost x-ray patterning of high-resolution, high-aspect-ratio mems," *Microelectronics Journal*, vol. 33, no. 1-2, pp. 101–105, 2002.

[48] R. T. Howe, "Surface micromachining for microsensors and microactuators," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, vol. 6, no. 6, p. 1809, 1988.

[49] J. W. Gardner, O. O. Awadelkarim, and V. K. Varadan, *Microsensors, MEMS, and Smart Devices*. Wiley New York, 2001.

[50] J. G. Korvink, *MEMS : A Practical Guide to Design , Analysis and Applications*. Springer Science \& Business Media, 2010.

[51] L. L. Wu, S. Babikian, G. P. Li, and M. Bachman, "Microfluidic printed circuit boards," *Proceedings - Electronic Components and Technology Conference*, pp. 1576–1581, 2011.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[52] B. H. Weigl, R. Bardell, T. Schulte, F. Battrell, and J. Hayenga, "Design and rapid prototyping of thin-film laminate-based microfluidic devices," *Biomedical Microdevices*, vol. 3, no. 4, pp. 267–274, 2001.

[53] B.N.Chichkov, C.Momma, S.Nolte, F. von Alvensleben, and A.Tiinnermann, "Femtosecond, picosecond and nanosecond laser ablation of solids," *Appl. Phys*, vol. 63, pp. 109–115, 1996.

[54] K. H. Leitz, B. Redlingshöer, Y. Reg, A. Otto, and M. Schmidt, "Metal ablation with short and ultrashort laser pulses," *Physics Procedia*, vol. 12, no. PART 2, pp. 230–238, 2011.

[55] D. Bauerle, *Laser Processing and Chemistry*. Springer Science \& Business Media, third ed., 2013.

[56] *Laser micromachining*, accessed August 10, 2019. `https://www.oxfordlasers.com/laser-micromachining/`.

[57] W. M. Steen and J. Mazumder, *Laser Material Porcessing*. springer science \& business media, 2010.

[58] R. Paschotta *et al.*, *Encyclopedia of laser physics and technology*, vol. 1. Wiley Online Library, 2008.

[59] *Optical Intensity*, accessed August 8, 2019. `https://www.rp-photonics.com/optical_intensity.html`.

[60] *Small Beam Diameter Scanning Galvo Mirror Systems*, accessed August 2, 2019. `https://www.thorlabs.com/thorproduct.cfm?partnumber=GVS202`.

[61] W. F. Hosford and R. M. Caddell, *Metal forming: mechanics and metallurgy*. Cambridge University Press, 2011.

[62] D. Karnakis, G. Rutterford, M. Knowles, T. Dobrev, P. Petkov, and S. Dimov, "High quality laser milling of ceramics, dielectrics and metals using nanosecond



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

and picosecond lasers," in *Photon Processing in Microelectronics and Photonics V*, vol. 6106, p. 610604, 2006.

[63] P. Heyl, T. Olschewski, and R. W. Wijnaendts, "Manufacturing of 3D structures for micro-tools using laser ablation," *Microelectronic Engineering*, vol. 57-58, pp. 775–780, 2001.

[64] J. Meijer, "Laser beam machining (LBM), state of the art and new opportunities," *Journal of Materials Processing Technology*, vol. 149, no. 1-3, pp. 2–17, 2004.

[65] D. Breitling, A. Ruf, and F. Dausinger, "Fundamental aspects in machining of metals with short and ultrashort laser pulses," *Photon Processing in Microelectronics and Photonics III*, vol. 5339, no. July 2004, p. 49, 2004.

[66] A. C. Forsman, P. S. Banks, M. D. Perry, E. M. Campbell, A. L. Dodell, and M. S. Armas, "Double-pulse machining as a technique for the enhancement of material removal rates in laser machining of metals," *Journal of Applied Physics*, vol. 98, no. 3, 2005.

[67] R. W. Pryor, *Multiphysics modeling using COMSOL: a first principles approach*. Jones & Bartlett Publishers, 2009.

[68] *Modeling thermal ablation for material removal*, accessed August 8, 2019. `https://www.comsol.com/blogs/modeling-thermal-ablation-for-material-removal/`.

[69] D. B. Go and A. Venkattraman, "Microscale gas breakdown: Ion-enhanced field emission and the modified Paschen's curve," *Journal of Physics D: Applied Physics*, vol. 47, no. 50, 2014.

[70] D. Landolt, "Fundamental Aspects of Electropolishing," *Ekcuochimiru Aore*, vol. 12, 1987.

[71] K. E. Petersen, "Silicon as a mechanical material," in *Proceedings of the IEEE*, vol. 70, pp. 420–457, IEEE, 1982.



576
DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[72] B. Wu, A. Kumar, and S. Pamarthy, "High aspect ratio silicon etch: A review," *Journal of Applied Physics*, vol. 108, no. 5, 2010.

[73] F. Brandi, N. Burdet, R. Carzino, and A. Diaspro, "Very large spot size effect in nanosecond laser drilling efficiency of silicon," *Optics Express*, vol. 18, no. 22, p. 23488, 2010.

[74] C. Molpeceres, S. Lauzurica, J. L. Ocãa, J. J. Gandía, L. Urbina, and J. Cárabe, "Microprocessing of ITO and a-Si thin films using ns laser sources," *Journal of Micromechanics and Microengineering*, vol. 15, no. 6, pp. 1271–1278, 2005.

[75] J. Kaspar, A. Luft, S. Nolte, M. Will, and E. Beyer, "Laser helical drilling of silicon wafers with ns to fs pulses: Scanning electron microscopy and transmission electron microscopy characterization of drilled through-holes," *Journal of Laser Applications*, vol. 18, no. 2, pp. 85–92, 2006.

[76] S. T. Hendow and S. A. Shakir, "Structuring materials with nanosecond laser pulses," *Optics Express*, vol. 18, no. 10, p. 10188, 2010.

[77] D. J. Lim, H. Ki, and J. Mazumder, "Mass removal modes in the laser ablation of silicon by a Q-switched diode-pumped solid-state laser (DPSSL)," *Journal of Physics D: Applied Physics*, vol. 39, no. 12, pp. 2624–2635, 2006.

[78] Y. H. Lee and K. J. Choi, "Analysis of silicon via hole drilling for wafer level chip stacking by UV laser," *International Journal of Precision Engineering and Manufacturing*, vol. 11, no. 4, pp. 501–507, 2010.

[79] H. Herfurth, R. Patwa, T. Lauterborn, S. Heinemann, and H. Pantsar, "Micromachining with tailored nanosecond pulses," *Photonics North 2007*, vol. 6796, no. October 2007, p. 67961G, 2007.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[80] D. Karnakis, G. Rutterford, and M. Knowles, "High power dpss laser micromachining of silicon and stainless steel," in *Proceedings of the Third International WLT-Conference on Lasers in Manufacturing*, no. June, 2005.

[81] J. H. Yoo, S. H. Jeong, X. L. Mao, R. Greif, and R. E. Russo, "Evidence for phase-explosion and generation of large particles during high power nanosecond laser ablation of silicon," *Applied Physics Letters*, vol. 76, no. 6, pp. 783–785, 2000.

[82] B. Wu, A. Kumar, and S. Pamarthy, "High aspect ratio silicon etch: A review," *Journal of Applied Physics*, vol. 108, no. 5, 2010.

[83] F. Beyeler, A. Neild, and S. Oberti, "Monolithically Fabricated Microgripper With Integrated Force Sensor for Manipulating Microobjects and Biological Cells Aligned in an Ultrasonic Field," *Journal of Microelectromechanical Systems*, vol. 16, no. 1, pp. 7–15, 2007.

[84] E. J. Garcia and J. J. Sniegowski, "Surface Micromachined Microengine as the Driver for Electromechanical Gears," *Eurosensors IX*, pp. 365–368, 1995.

[85] *Lynx Evo Eyepiece-less Stereo Microscope*, accessed August 10, 2019. `http://www.visioneng.us/products/stereo-microscopes/lynx-evo-dynascope-inspection/`.

[86] X. Xie and C. Livermore, "Passively self-aligned assembly of compact barrel hinges for high-performance, out-of-plane mems actuators," *Proceedings of the IEEE International Conference on Micro Electro Mechanical Systems (MEMS)*, pp. 813–816, 2017.

[87] X. Xie and C. Livermore, "A pivot-hinged, multilayer SU-8 micro motion amplifier assembled by a self-aligned approach," *Proceedings of the IEEE International Conference on Micro Electro Mechanical Systems (MEMS)*, vol. 2016-Febru, no. January, pp. 75–78, 2016.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[88] S. H. Lee, F. Niklaus, J. J. McMahon, J. Yu, R. J. Kumar, H.-f. Li, R. J. Gutmann, T. S. Cale, and J.-Q. Lu, "Fine Keyed Alignment and Bonding for Wafer-Level 3D ICs," *MRS Proc.*, vol. 914, pp. F10–05, 2006.

[89] M. Elwenspoek, R. Legtenberg, and a. W. Groeneveld, "Comb-drive actuators for large displacements," *Journal of Micromechanics and Microengineering*, vol. 6, no. 3, pp. 320–329, 1999.

[90] L. J. Durney, *Graham's Electroplating Engineering Handbook*. Springer Science & Business Media, 1984.

[91] V. Kaajakari, *Practical MEMS*. Small Gear Publishing, 2009.

[92] F. W. Strong, J. L. Skinner, P. M. Dentinger, and N. C. Tien, "Electrical breakdown across micron scale gaps in MEMS structures," *Reliability, Packaging, Testing, and Characterization of MEMS/MOEMS V*, vol. 6111, p. 611103, 2006.

[93] M. Klas, S. Matejík, B. Radjenović, and M. Radmilović-Radjenović, "Experimental and theoretical studies of the breakdown voltage characteristics at micrometre separations in air," *Epl*, vol. 95, no. 3, 2011.

[94] A. Hariri, J. W. Zu, and R. Ben Mrad, "Modeling of dry stiction in micro electromechanical systems (MEMS)," *Journal of Micromechanics and Microengineering*, vol. 16, no. 7, pp. 1195–1206, 2006.

[95] Y. X. Zhuang, O. Hansen, T. Knieling, C. Wang, P. Rombach, W. Lang, W. Benecke, M. Kehlenbeck, and J. Koblitz, "Vapor-phase self-assembled monolayers for anti-stiction applications in MEMS," *Journal of Microelectromechanical Systems*, vol. 16, no. 6, pp. 1451–1460, 2007.

[96] W. R. Ashurst, C. Carraro, and R. Maboudian, "Vapor phase anti-stiction coatings for MEMS," *IEEE Transactions on Device and Materials Reliability*, vol. 3, no. 4, pp. 173–178, 2003.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[97] T. P. Knowles, M. Vendruscolo, and C. M. Dobson, "The amyloid state and its association with protein misfolding diseases," *Nature Reviews Molecular Cell Biology*, vol. 15, no. 6, pp. 384–396, 2014.

[98] H. Jang, J. Zheng, R. Lal, and R. Nussinov, "New structures help the modeling of toxic amyloidß ion channels," *Trends in Biochemical Sciences*, vol. 33, no. 2, pp. 91–100, 2008.

[99] R. Andrew, ER and Bradbury, A and Eades, "Nuclear magnetic resonance spectra from a crystal rotated at high speed," *Nature*, vol. 182, no. 4650, p. 1658, 1958.

[100] M. Cohen, R. P. Feynman, and L. J. Lowe, "Physical Review Letters Free Induction Decays of Rotating Solids," *Phys. Rev. Lett*, vol. 2, no. 7, p. 71, 1957.

[101] C. P. Jaroniec, C. E. MacPhee, V. S. Bajaj, M. T. McMahon, C. M. Dobson, and R. G. Griffin, "High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 3, pp. 711–716, 2004.

[102] R. Tycko, "Solid State NMR Studies of Amyloid Fibril Structure Robert," *Annual review of physical chemistry*, vol. 62, no. 1, pp. 279—-299, 2011.

[103] A. Böckmann, M. Ernst, and B. H. Meier, "Spinning proteins, the faster, the better?," *Journal of Magnetic Resonance*, vol. 253, pp. 71–79, 2015.

[104] T. Polenova1, R. Gupta, and A. Goldbourt, "Magic Angle Spinning NMR Spectroscopy: A Versatile Technique for Structural and Dynamic Analysis of Solid-Phase Systems," vol. 25, no. 3, pp. 289–313, 2016.

[105] S. Penzel, A. Oss, M. L. Org, A. Samoson, A. Böckmann, M. Ernst, and B. H. Meier, "Spinning faster: protein NMR at MAS frequencies up to 126 kHz," *Journal of Biomolecular NMR*, vol. 73, no. 1-2, pp. 19–29, 2019.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[106] E. Bouleau, P. Saint-Bonnet, F. Mentink-Vigier, H. Takahashi, J. F. Jacquot, M. Bardet, F. Aussenac, A. Purea, F. Engelke, S. Hediger, D. Lee, and G. De Paëpe, "Pushing NMR sensitivity limits using dynamic nuclear polarization with closed-loop cryogenic helium sample spinning," *Chemical Science*, vol. 6, no. 12, pp. 6806–6812, 2015.

[107] W. H. Potter, "Apparatus to rotate samples rapidly at temperatures less than 2 K in high transverse magnetic fields," *Review of Scientific Instruments*, vol. 42, no. 5, pp. 618–625, 1971.

[108] S. Wegner, *Solid State NMR News and Updates*, accessed August 10, 2019. `https://www.bruker.com/fileadmin/user_upload/8-PDF-Docs/MagneticResonance/Events_NMR/UM_DE18/Solid_State_NMR_Workshop_2.pdf`.

[109] J. Jensen, R. G. Stewart, W. Tuttle, and H. Brechna, *Brookhaven national laboratory selected cryogenic data notebook: sections I-IX*, vol. 1. Brookhaven National Laboratory, 1980.

[110] D. Olson, G. Reynolds, G. Virshup, F. Friedlander, B. James, and L. Partain, "Tensile strength of synthetic chemical- vapor-deposited diamond," *Journal of Applied Physics*, vol. 5177, no. August, 1995.

[111] A. G. Hamlir, B. J. Roberts, U. Kingdom, and A. Energy, "February 20, 1960," *Nature*, vol. 185, no. 4712, p. 522, 1960.

[112] A. D. Vita, G. Galli, A. Canningt, and R. Car, "Diamond-To-Graphite Transitions," vol. 379, no. February, pp. 523–526, 1996.

[113] C. Hedlund, H. Blom, and S. Berg, "Microloading effect in reactive ion etching," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, vol. 12, no. 4, pp. 1962–1965, 1994.

[114] *University wafer*, accessed August 10, 2019. `https://order.universitywafer.com`.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[115] P. Ramm, J. J.-Q. Lu, and M. M. Taklo, *Handbook of wafer bonding*. John Wiley & Sons, 2011.

[116] S. Beeby, G. Ensel, N. M. White, and M. Kraft, *MEMS mechanical sensors*. Artech House, 2004.

[117] J. Henri, G. Han, B. Meint de, E. Miko, and F. Jan, "A survey on the reactive ion etching of silicon in microtechnology," *Journal of Micromechanics and Microengineering*, vol. 6, no. 1, p. 14, 1996.

[118] M. W. Judy and R. T. Howe, "Polysilicon Hollow Lateral Resonators," in *[1993] Proceedings IEEE Micro Electro Mechanical Systems*, pp. 265—-271, IEEE, 1993.

[119] Y.-H. Yoon, Y. Jin, C.-K. Kim, and S. Hong, "4-Terminal MEMS relay with an extremely low contact resistance employing a novel one-contact design," in *2017 19th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*, pp. 906–909, IEEE, 2017.

[120] R. Nathanael, V. Pott, H. Kam, J. Jeon, E. Alon, and T. J. K. Liu, "Four-terminal-relay body-biasing schemes for complementary logic circuits," *IEEE Electron Device Letters*, vol. 31, no. 8, pp. 890–892, 2010.

[121] S. Farrens, *Wafer level 3-D ICs process technology*. Springer Science \& Business Media, 2009.

[122] S. H. Lee, K. N. Chen, and J. J. Q. Lu, "Wafer-to-wafer alignment for three-dimensional integration: A review," *Journal of Microelectromechanical Systems*, vol. 20, no. 4, pp. 885–898, 2011.

[123] B. T. Tung, N. Watanabe, F. Kato, K. Kikuchi, and M. Aoyagi, "Flip-chip bonding alignment accuracy enhancement using self-aligned interconnection elements to realize low-temperature construction of ultrafine-pitch copper bump interconnections," *Proceedings - Electronic Components and Technology Conference*, pp. 62–67, 2014.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[124] A. H. Slocum and A. C. Weber, "Precision passive mechanical alignment of wafers," *Journal of Microelectromechanical Systems*, vol. 12, no. 6, pp. 826–834, 2003.

[125] M. Stafe, A. Marcu, and N. Puscas, *Pulsed Laser Ablation of Solids: Basics, Theory and Applications*, vol. 53. Springer Science \& Business Media, 2013.

[126] K. Sugioka and Y. Cheng, "Femtosecond laser three-dimensional micro-and nanofabrication," *Applied Physics Reviews*, vol. 1, no. 4, 2014.

[127] A. A. Tseng, "Recent developments in micromilling using focused ion beam technology," *Journal of micromechanics and microengineering*, vol. 14, no. 4, p. R15, 2004.

[128] K. Watanabe, M. Kinoshita, T. Mine, M. Morishita, K. Fujisaki, R. Matsui, M. Sagawa, S. Machida, H. Oba, Y. Sugiyama, *et al.*, "Plasma ion-beam 3d printing: A novel method for rapid fabrication of customized mems sensors," in *2018 IEEE Micro Electro Mechanical Systems (MEMS)*, pp. 459–462, IEEE, 2018.

[129] M. Vaezi, H. Seitz, and S. Yang, "A review on 3D micro-additive manufacturing technologies," *International Journal of Advanced Manufacturing Technology*, vol. 67, no. 5-8, pp. 1721–1754, 2013.

[130] S. Maruo and K. Ikuta, "Submicron stereolithography for the production of freely movable mechanisms by using single-photon polymerization," *Sensors and Actuators, A: Physical*, vol. 100, no. 1, pp. 70–76, 2002.

[131] S. Maruo and K. Ikuta, "Movable microstructures made by two-photon three-dimensional microfabrication," in *MHS'99. Proceedings of 1999 International Symposium on Micromechatronics and Human Science (Cat. No. 99TH8478)*, pp. 173–178, IEEE, 1999.

[132] *Agilis Piezo Motor Driven Linear Stages*, accessed August 2, 2019. `https://www.newport.com/f/agilis-piezo-motor-linear-stages`.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

[133] N. N. M. Peek, *Making machines that make: object-oriented hardware meets object-oriented software*. PhD thesis, Massachusetts Institute of Technology, 2016.

[134] *Machine That Makes*, accessed August 2, 2019. `http://mtm.cba.mit.edu/`.

[135] J. T. Schwartz, J. von Neumann, and A. W. Burks, "Theory of Self-Reproducing Automata," 2006.

[136] *Nimrod copper*, accessed August 10, 2019. `https://www.nimrodcopper.com/shop/4-mil-copper-foils-(-004)-13250`.



DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

# LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

| ACRONYM | DESCRIPTION |
|---|---|
| AQFP | Adiabatic Quantum Flux Parametrons |
| CSP | Chip-Scale Packages |
| DICE | Discrete Integrated Circuit Electronics |
| EDS | Electronic Design Automation |
| NGMM | Next Generation Microelectronics Manufacturing |
| SMD | Surface Mount Devices |
| VMD | Volume Mount Devices |


DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.