AFRL-RX-WP-TR-2020-0092

2.5D AND 3D FPGA-CENTRIC MICROFLUIDIC COOLING ARCHITECTURAL PLATFORMS FOR SUPERIOR COMPUTATIONAL THROUGHPUT

Muhannad Bakir and Yogendra Joshi
Georgia Institute of Technology (GIT)

Jon Goldman
Intel Programmable Solutions Group (Intel)

27 NOVEMBER 2019
Final Report

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

AIR FORCE RESEARCH LABORATORY
MATERIALS AND MANUFACTURING DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE OH 45433-7750
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 SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nationals.

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

AFRL-RX-WP-TR-2020-0092 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNATURE//________________
JOHN D. BLEVINS
Program Manager
Electronics and Sensors Branch
Manufacturing and Industrial Technologies Division
Materials and Manufacturing Directorate

//SIGNATURE//________________
ALAN P. ALBERT
Branch Chief
Electronics and Sensors Branch
Manufacturing and Industrial Technologies Division
Materials and Manufacturing 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.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Defense, Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it dows not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)
27 November 2019

2. REPORT TYPE
Final

3. DATES COVERED *(From – To)*
28 July 2016 – 27 November 2019

4. TITLE AND SUBTITLE
2.5D AND 3D FPGA-CENTRIC MICROFLUIDIC COOLING ARCHITECTURAL PLATFORMS FOR SUPERIOR COMPUTATIONAL THROUGHPUT

5a. CONTRACT NUMBER
FA8650-16-1-7674

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
61101E

5d. PROJECT NUMBER
DARPA

5e. TASK NUMBER

5f. WORK UNIT NUMBER
Y18J

6. AUTHOR(S)
Muhannad Bakir and Yogendra Joshi - Georgia Institute of Technology (GIT)
Jon Goldman - Intel Programmable Solutions Group (Intel)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Georgia Institute of Technology
505 Tenth St NW Code 0420
Atlanta GA 30332

Intel Programmable Solutions Group
2200 Mission College
Santa Clara CA 95052

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Materials and Manufacturing Directorate
WPAFB OH 45433-7750
Air Force Materiel Command
United States Air Force

Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington VA 22203-1714

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RXME

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RX-WP-TR-2020-0092

12. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.

13. SUPPLEMENTARY NOTES
Report contains color.

14. ABSTRACT ***(Maximum 200 words)***
During this program, Georgia Tech developed highly integrated microfluidic cooling technologies for 2.5D high-power density and high-performance electronics. Multiple fabrication technologies were developed and demonstrated. Lateral manifolds using 3D printing to enable ultra-compact inlet/outlet manifolds were also developed. Using a Stratix 10 FPGA as a platform, our proposed thermal technologies reduced the junction temperature of the FPGA by more than 32C when compared to air cooling. Moreover when compared to air cooling, the thermal coupling of the FPGA on the surrounding transceiver die temperatures was reduced by a factor of 10x to over 100x.

15. SUBJECT TERMS
DARPA report, 2.5D ICs, microfluidic cooling, high-performance, high-power ICs

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON *(Monitor)*
					John D. Blevins
a. REPORT	b. ABSTRACT	c. THIS PAGE	SAR	48	19b. TELEPHONE NUBER *(include area code)*
Unclassified	Unclassified	Unclassified			(937) 255-0223

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

TABLE OF CONTENTS

Section **Page**

LIST OF FIGURES and TABLES ii
List Of Figures ii
List Of Tables iii

1. SUMMARY 1

2. INTRODUCTION 2

3. KEY HIGHLIGHTS FROM THERMAL MODELING OF 2.5D FPGA ICs 3

4. FABRICATION OF SILICON MICROPIN-FINS COOLING STRUCTURES FOR AND HOT SPOT COOLERS 7

5. MICROFLUIDIC COOLING OF A 14NM 2.5D FPGA WITH 3D PRINTED LATERAL MANIFOLDS FOR HIGH DENSITY COMPUTING 9

6. MICROFLUIDIC COOLING USING MONOLITHIC COPPER MICROPIN-FINS 17

7. MONOLITHIC MICROPIN-FIN INTEGRATION INTO 2.5D STRATIX 10 FPGA AND TRANSCEIVER DICE 20

8. PRELIMINARY RELIABILITY STUDIES 24

9. DARPA ACCELERATOR 25

10. CONCLUSIONS 26

APPENDIX: DARPA ACCELERATOR FINAL FINDINGS 27
SUMMARY: High-level Take Away Learnings: 27
Statement Of Work Milestones 28
Appendix A – Competitor List 39
Appendix B – Customer Discovery Interviews 40

REFERENCES 41

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS 42

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

LIST OF FIGURES and TABLES

List Of Figures

Figure **Page**

Figure 1. Locations of Heaters with the Package 3
Figure 2. Manifold Design for 2.5D-SICs 4
Figure 3. Cooling Enhancement of Bridge-Wing Structure: (a) Manifold without Bridge- Wing Structure (b) Manifold with Bridge-Wing Structure 5
Figure 4. Thermal coupling between ICs in 2.5D Configuration. 6
Figure 5. Die spacing vs. thermal cross-talk vs. signaling........ 6
Figure 6. Heterogeneous silicon micropin-fins; micropin-fins with various dimensions across the chip to enable local heat removal enhancement at hot spots. 7
Figure 7. Schematic illustration of thermal setup developed to evaluate the performance of micropin-fins with localized hot-spot coolers........ 8
Figure 8. Junction temperature rise across the four heaters and the hotspot when: a) background and host spot power density is 100 W/cm^2, and (b) when the background power is increased to 250 W/cm^2. 8
Figure 9. Schematic of the microfluidic cooling 2.5D FPGA using 3D printed lateral manifolds and capping layer 9
Figure 10. Delidded S10 FPGA package showing the center FPGA surrounded by 4 transceivers using EMIB. Power dissipation of each die also shown. 11
Figure 11. Design and Fabrication of multi-diameter and multi-pitch micropin-fins to enables local optimization of hea removal across the FPGA ans transceiver dice. 12
Figure 12. 3D printed lateral inlet/outlets for the silicon micropin-fins. 13
Figure 13. Images of the delidded 2.5D FPGA package on the board (top) and images of the FPGA with microfluidic cooling using 3D printed lateral manifolds (bottom)........ 14
Figure 14. Average junction temperature vs. flow rate for all five dice on the package vs. flow rate........ 15
Figure 15. Pressure drop vs. flow rate for the micropin-fins. 15
Figure 16. Summary of thermal cross-talk between dice in the 2.5D FPGA when using air-cooling and microfluidic cooling. 16
Figure 17. Vision of locally electroplated copper micropin-fins and 3D printed lateral inlet/outlet manifolds........ 17
Figure 18. Schematic of the 3D copper micropin-fin fabrication process: cotton fibers (printing tip) are saturated with a copper electroplating solution and are used to locally fabricate the copper micropin-fins by applying a bias (between anode and substrate). 18
Figure 19. Developed system for copper micropin-fin fabrication on silicon chips and wafers using localized electroplating. 18
Figure 20. Colored SEM images of the fabricated copper micropin-fins on silicon chips........ 18
Figure 21. Pressure drop benefits from the 1mm tall copper micropin-fins. 19
Figure 22. Summary of all dimensions demonstrated with the copper micropin-fins. The pitch was approximately 1 mm. 19
Figure 23. Micropin-fin dimensions 20
Figure 24. Conceptual image of assembly process 20

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

Figure 25. Design details of frame and capping manifold 21
Figure 26. 3D printed frame and capping manifold 21
Figure 27. Silicon micropin-fins etched monolithically into the 2.5D FPGA/transceivers 22
Figure 28. SEM images of the etched micropin-fins monolithically into the 2.5D FPGA and transceivers: a) Micropin-fins over FPGA region, (b) Micropin-fins over XCVR region, (c) Heterogeneous micropin-fin design over XCVR and FPGA regions 23
Figure 29. Samples used for reliability studies: a) Image of the manifold after die attach, b) Rendered cross-section of the manifold showing fludic channels 24
Figure 30. Thermal cycling conditions and results 24

Figure A 1. Initial survey of the ost likely market segments 2016-2022 29
Figure A 2. At the end of initial market assesment, this is where we landed 30
Figure A 3. Components of the BoM 31
Figure A 4. Potential channels to Market 32
Figure A 5. a) NGC Opportunity/Specifications, b) Chip on small stack 33
Figure A 6. Refined Balue Chain Assessment 35
Figure A 7. Refined Product Hypothesis a) in this 1st embodiment, a true monolithically integrated solution and b) the 2nd embodiment 36
Figure A 8. Scaled BoM Estimate (100,000 units) 37

List Of Tables

Table 1. Mesh independent study. 5

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

1. SUMMARY

This report describes select key results from the DARPA project entitled: 2.5D and 3D FPGA-Centric Microfluidic Cooling Architectural Platforms for Superior Computational Throughput (Grant #: FA8650-16-1-7674). During this program, we developed microfluidic cooling technologies for 2.5D electronics, to improve performance, increase compute density, and increase energy efficiency through modeling and experiments. Focus was given to fabrication and integration strategies for microfluidic cooling based on a) silicon micropin-fin heterogeneous integration, b) silicon monolithic integration, and c) copper monolithic micropin-fins using localized electroplating. Lateral manifolds using 3D printing to enable ultra-compact inlet/outlet manifolds were also developed. Using a Stratix 10 FPGA as a platform, our proposed thermal technologies reduced the junction temperature of the FPGA by more than 32C when compared to air cooling. Moreover when compared to air cooling, the thermal coupling of the FPGA on the surrounding transceiver die temperatures was reduced by a factor of 10x to over 100x.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

2. INTRODUCTION

Today, designers of large-scale compute systems face multiple challenges in performance and energy scaling. If Moore's Law has not already ended, it is certainly in stark decline, as the fully-loaded cost per transistor in a 14nm or 10nm node is no longer half that of the prior generation. As a result, our technology "entitlement" of the past half-century has gone away, forcing us to find different technologies to replace chip-level integration.
From a workload execution perspective, system performance on a fixed number of instructions depends on just two factors: clock frequency, or cycles per second; and parallelism, or instructions per cycle. Growth in the former has been limited by power and power density: system clock rates of CPUs, ASICs, and accelerators are bounded because forced-air cooling with fans and heat sinks can dissipate only about 100 watts per square centimeter. In a similar vein, growth in the latter, whether task-, thread-, or instruction-level parallelism, will now be increasingly limited as mentioned above. The slowdown of Moore's law hampers our ability to continuously grow the number of execution cores, memories, and interconnect paths needed to increase parallelism.

Researchers in both academia and industry have proposed many solutions to these fundamental problems. The most promising directions aim to replace chip-centric scalability with package-level aggregation, using technologies broadly known as 2.5D and 3D integration. Yet power and power density remain a challenge: if it is high enough to limit single-chip performance, it will certainly constrain the scalability of multi-die packages. The aim of this program is to develop innovative microfluidic cooling technologies to enable silicon electronics with virtually no thermal constraints to enable superior performance electronics.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

3. KEY HIGHLIGHTS FROM THERMAL MODELING OF 2.5D FPGA ICs

During this program, significant thermal modeling was performed to help guide the design of the cooling hardware for 2.5D electronics. While the modeling is agnostic to the underlying chip technologies, in virtually all our modeling efforts, we assumed FPGA-CPU and FPGA-transceiver co-integration using 2.5D/3D technologies. In this section, we provide key highlights of the modeling.

We assume an FPGA die that is 25 mm×25 mm with four transceivers dice that are each 6 mm×6 mm. The FPGA is placed between two pairs of transceivers. The five chip arrangements, seen in Fig. 1, are considered in subsequent sections.

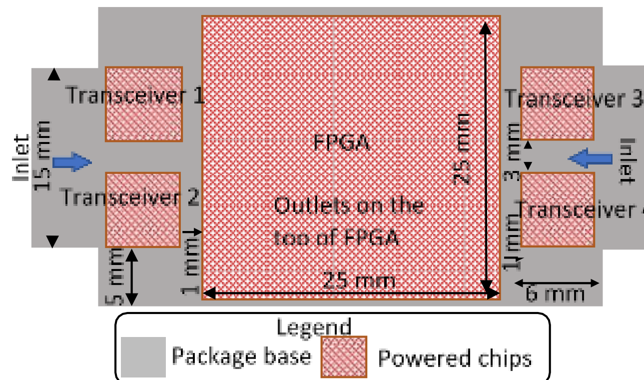


Figure 1. Locations of Heaters with the Package

To enable the parametric study that will be discussed in the next section, we build various models with the commercially available ANSYS Icepak 16.2 that uses FLUENT 16.2 solver (Icepak, 2013). The computational fluid dynamics and heat transfer (CFD/HT) solver is based on finite control volumes. Each full-scale simulation model includes 3 million nodes and accounts for inlets, outlets, and a micro-gap with five heaters and their cold plates in a package. Multi-level meshing with increased mesh is performed in the micro-gap including pin-fin structures. The boundary conditions are: Uniform inlet velocity at a flow rate of 5.4 cm^3/s, and inlet temperature at 55 ºC of coolant PAO (Polyalphaolefin) is applied for both inlets; free flow outlets with laminar flow condition; radiation is included in the model; ambient temperature is 55 ºC; Five conducting silicon chips with thickness of 0.6 mm that are located on the package of FR-4 have uniform volumetric power. Governing equations of fluidic mass, momentum and energy conservation are:

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0 \quad (1)$$

$$\rho\left(u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial z}\right)=-\frac{\partial P}{\partial x}+\mu\nabla^2 u \quad (2)$$

$$\rho\left(u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+w\frac{\partial v}{\partial z}\right)=-\frac{\partial P}{\partial y}+\mu\nabla^2 v \quad (3)$$

$$\rho\left(u\frac{\partial w}{\partial x}+v\frac{\partial w}{\partial y}+w\frac{\partial w}{\partial z}\right)=-\frac{\partial P}{\partial z}+\mu\nabla^2 w \quad (4)$$

$$\rho c_p\left(u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}+w\frac{\partial T}{\partial z}\right)=k\nabla^2 T \quad (5)$$

The thermal diffusion equation in the silicon solid domain is:

$$\frac{\partial}{\partial x}\left(k_{si}\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_{si}\frac{\partial T}{\partial y}\right)+\frac{\partial}{\partial z}\left(k_{si}\frac{\partial T}{\partial z}\right)=0 \quad (6)$$

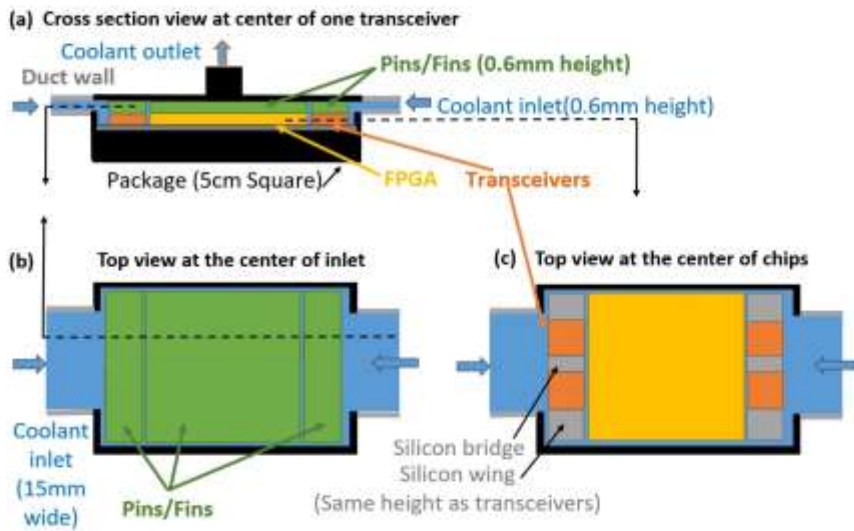


Figure 2. Manifold Design for 2.5D-SICs

Cooling performance of micropin-fin enhanced silicon bridges and wings has been studied with the same power map and flow rate at 5.4 cm^3/s, as shown in Fig. 3. PAO (Polyalphaolefin), a dielectric coolant employed in avionics applications has been employed. Coolant at 55 ºC is pumped into both manifolds from two inlets at both sides and discharged through two square outlets of 6 mm×6 mm at the top of FPGA. In the cooling manifold without silicon bridge and wings, shown as Fig. 3 (a), straight fins of thickness 0.18 mm and pitch 0.18 mm are directly located on four transceivers. Cylindrical pins of radius 0.2 mm and pitch 1.8 mm are located on the FPGA. Fig. 3 (a) shows the temperature field of five heaters. Maximum temperatures of the four transceivers reach 179.6 ºC. Pressure drop is 4850 N/m^2. In contrast, by implementing silicon bridge-wing structures with straight fins of the same dimension as on transceivers, the temperatures of transceivers are significantly reduced to 105 ºC, indicated in Fig. 3 (b).

Although the transceiver temperatures are still higher than the desired 85 ºC and pressure drop increases to 19,937 N/m^2, the cooling enhancement has been demonstrated. The transceiver temperatures could be further alleviated by optimizing the fin structures on the transceivers, and

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

the FPGA, and inlet and outlet arrangements.

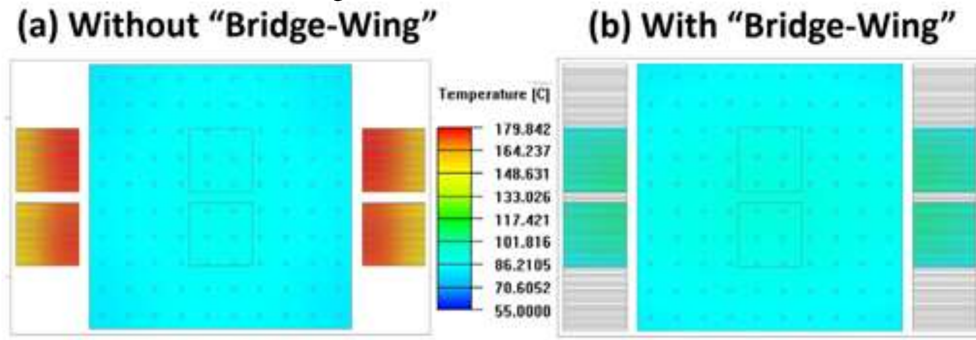


Figure 3. Cooling Enhancement of Bridge-Wing Structure: (a) Manifold without Bridge-Wing Structure (b) Manifold with Bridge-Wing Structure

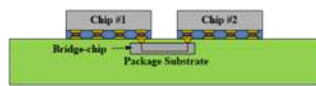
Mesh independent test has been performed as shown in Table 1 from both thermal and hydraulic perspectives. Compared with other cases, mesh case 2 with 3.74 million nodes are elected, as maximum temperatures, center location temperature, pressure drop, and velocity are relatively stable with lower grid resource consuming.

Table 1. Mesh independent study.

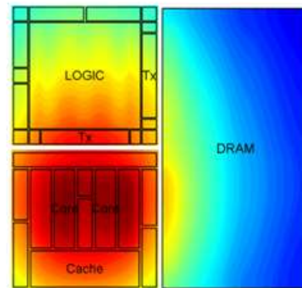
Mesh Cases	1	2 (Chosen)	3	4
Nodes Numbers (Million)	2.96	3.74	3.98	7.54
Max Temperature of FPGA (^oC)	86.3	85.3	85.4	84.9
Temperature at FPGA Center (^oC)	85.0	84.0	83.6	83.2
Max Temperatures of Transceivers (^oC)	89.1	89.0	90.2	89.3
Temperatures at Transceiver Center (^oC)	88.4	88.6	89.8	88.9
Pressure Drop (N/m^2)	44647	45880	45600	46064
Average Coolant Velocity on the Top of FPGA (m/s)	3.0	3.2	3.2	3.3

Moreover, we performed thermal coupling modeling between adjacent dice in 2.5D using finite-volume-method (implemented in Matlab). Using an air-cooling configuration with an integrated heat spreader over the ICs, it is observed that due to the large power densities and close proximity between dice in a 2.5D platform, there is strong thermal coupling between the high power dice (CPU, FPGA), and the low power dice (memory) (Figure 4). Most of the thermal coupling is due to the integrated heat spreader atop the ICs. In order to counter this effect, one might postulate that increasing the distance between ICs is advantageous from a cooling perspective. As shown in Figure 5, while increasing the distance between ICs does reduce the thermal coupling, it significantly increases interconnect latency and energy per bit between the ICs. As such, increasing the distance between ICs is very counterproductive to the electrical performance of the system. As shown later, when microfluidic cooling is used, we are able to virtually eliminate thermal cross-talk while keeping dice in very close proximity (and thus, low latency and low energy signaling).

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

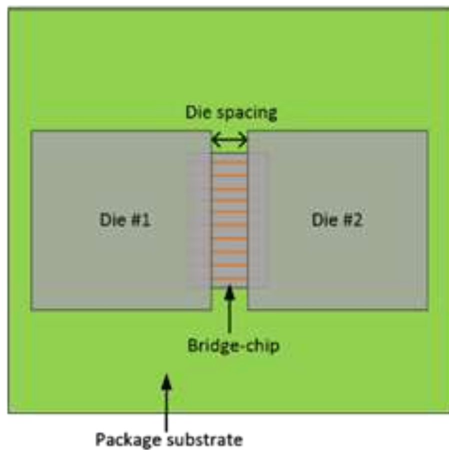


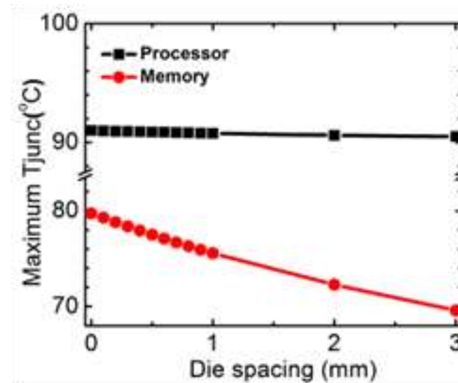
Bridge-chip



104.92 °C

Figure 4. Thermal coupling between ICs in 2.5D Configuration.





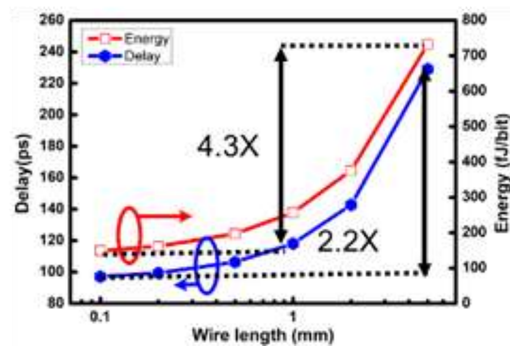


Figure 5. Die spacing vs. thermal cross-talk vs. signaling.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

4. FABRICATION OF SILICON MICROPIN-FINS COOLING STRUCTURES FOR AND HOT SPOT COOLERS

During this program, significant effort was devoted to silicon microfabrication of the micro-fluidic heat sink technologies. In particular, emphasis was placed on forming micropin-fins heat sinks with a wide range of dimensions (diameters and heights) to meet a wide range of thermal resistance and pressure drop values, as shown in Figure 6. Specifically, knowing that power dissipation is not uniform across a chip, we sought to develop methods by which we form regions of enhanced heat transfer using dense micro-pin fin arrays as shown in Figure 6; this work extends efforts developed in ICECool APPS program. Moreover, we extended processes from ICECool APPs in which silicon micropin-fins can be co-integrated with micro-gaps that provide very large heat transfer to enable hot spot cooling, as shown in Figure 7. To test these structures, the test-bed shown in Figure 7 was fabricated and characterized. Figure 8 illustrates junction temperature rise across the four heaters and the hotspot when (a) background and hot spot power density is 100 W/cm^2, and (b) when the hotspot power is increased to 250 W/cm^2. As we see in the figure, when the hot-spot power density more than doubles, the increase in the junction temperature is not significant. The overall experimental setup plays a critical role in the reported numbers and optimization is left for future work as our focus was to apply these advanced thermal concepts and technologies to the 2.5D FPGA system discussed in the next section.

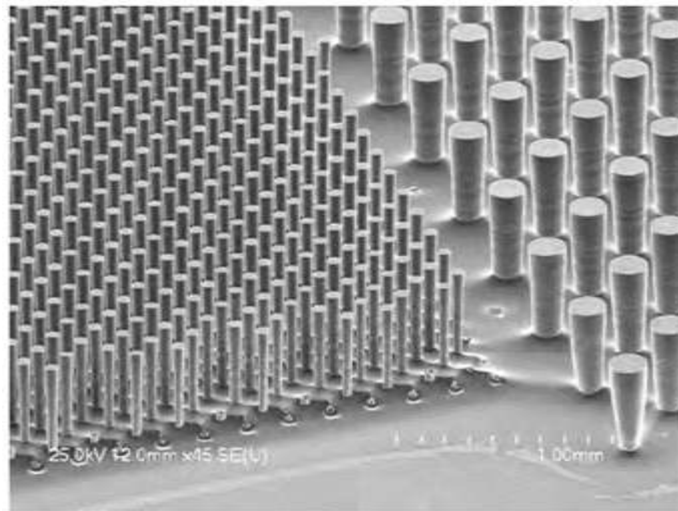


Figure 6. Heterogeneous silicon micropin-fins; micropin-fins with various dimensions across the chip to enable local heat removal enhancement at hot spots.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

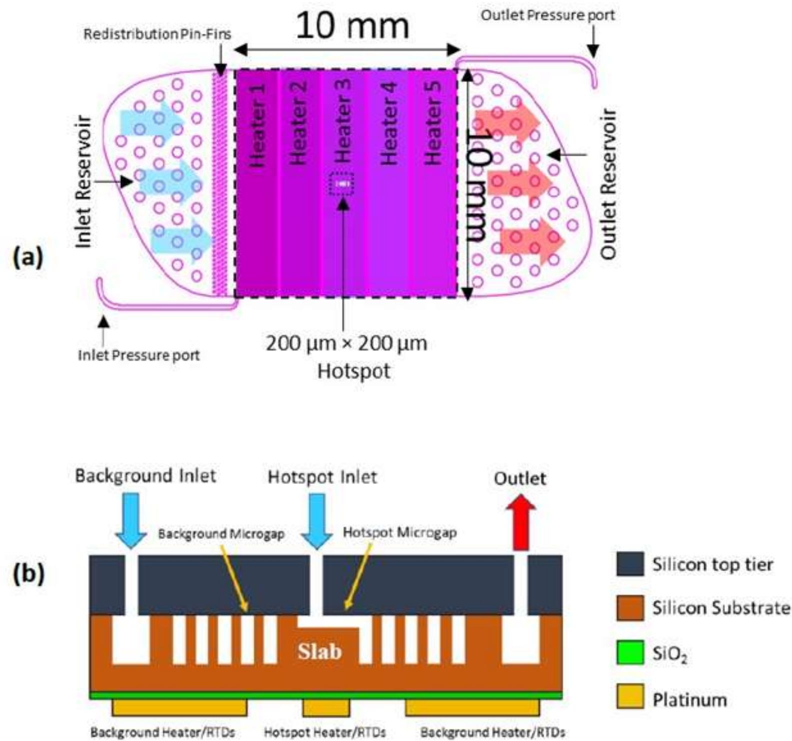


Figure 7. Schematic illustration of thermal setup developed to evaluate the performance of micropin-fins with localized hot-spot coolers.

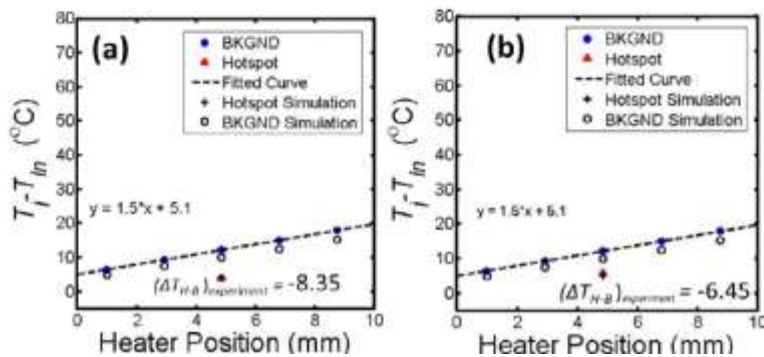


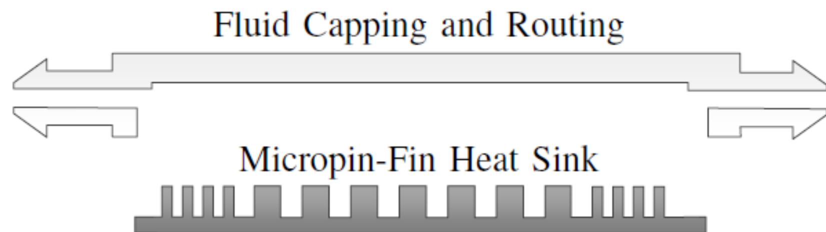
Figure 8. Junction temperature rise across the four heaters and the hotspot when: a) background and host spot power density is 100 W/cm^2, and (b) when the background power is increased to 250 W/cm^2.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

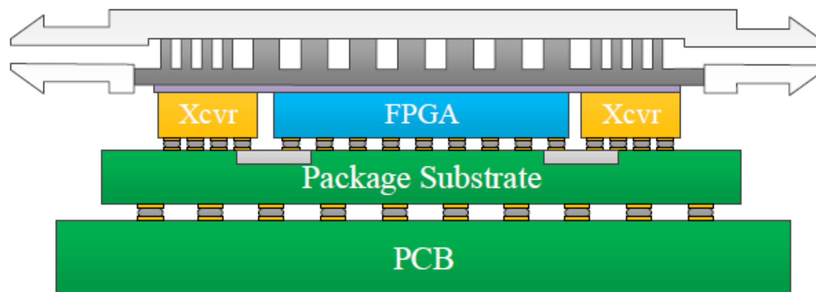
5. MICROFLUIDIC COOLING OF A 14NM 2.5D FPGA WITH 3D PRINTED LATERAL MANIFOLDS FOR HIGH DENSITY COMPUTING

In this work, a silicon micropin-fin heat sink is designed and fabricated to cool a Stratix 10 GX field-programmable gate array (FPGA) consisting of five heterogeneous dice.

The micropin-fin geometry is locally varied to match the heat flux of the underlying dice. The silicon micropin-fin heat sink is embedded in a 3D-printed plastic piece which seals the top of the silicon micropin-fins and connects the heat sink to inlet and outlet tubing. A conceptual cross-sectional diagram of the heat sink concept can be seen in Figure 9. The experiments in this work were carried out using a Stratix 10 engineering silicon (ES) development kit from Intel. The board carries a Stratix 10 GX FPGA which is a 2.5D device consisting of a 14 nm FPGA core die surrounded by four transceiver dice, connected through Intel's embedded multi-die interconnect bridge (EMIB). A photo of a delidded package can be seen in Fig. 10. Each transceiver tile (die) contains 24 transmitters and receivers, for a total of 96 in the package. Each receiver requires a dedicated reference clock connected directly through the package to the transceiver tile on which it resides. The Stratix 10 ES development board only has reference clocks connected to three of the four transceiver tiles, so only three of the four transceiver tiles are used in this work.



(a) Two parts of the microfluidic heat sink



(b) Microfluidic heat sink assembled on dice

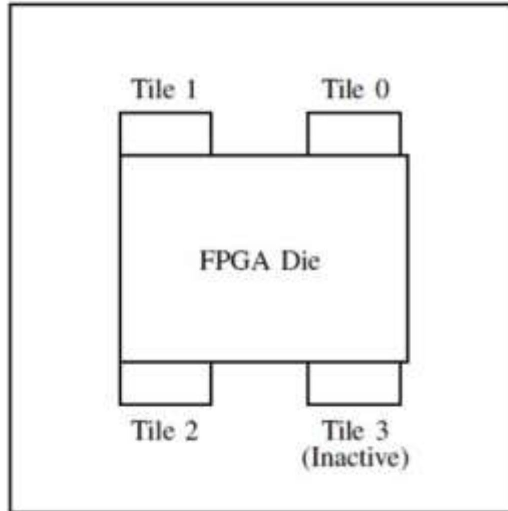
Figure 9. Schematic of the microfluidic cooling 2.5D FPGA using 3D printed lateral manifolds and capping layer

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

The benchmark program used in this work was designed to mimic a high power use case of the FPGA. It consists of a portion which dissipates power on the FPGA, and a portion which dissipates power on the transceivers. The core of the design consists of a streaming fast Fourier transform (FFT) block followed by six first-in, first-out (FIFO) buffers operating on random inputs hard coded into the FPGA. This design is implemented on the FPGA through a combination of programmable logic blocks and digital signal processing (DSP) blocks. Much of the computational throughput as well as power dissipation comes from the FPGA's DSP blocks, which perform arithmetic operations on floating point operands. The FFT core was replicated 160 times across the FPGA and clocked at 475 MHz. The clock (and power) could be raised higher, but the voltage regulator modules (VRMs) on the board are only designed to supply a maximum of 100A of current on the VCC rail, causing instability when the FPGA was clocked at significantly higher frequencies. Even at 475 MHz, the FPGA used well over 100A on the VCC rail and air was therefore blown over the VRMs to prevent them from overheating during experiments. In addition to this FFT design, 72 transceiver channels were utilized to dissipate power on three of the four transceiver tiles. Each of the 72 transceiver channels were programmed to run in enhanced physical coding sublayer (PCS) mode with serial loopback enabled. This design was modified from a publicly available design from the Intel FPGA Wiki. The maximum data rate within the Intel transceiver intellectual property (IP) for the GX series of Stratix 10 FPGAs is 16 Gbit/s per channel, but the transceiver clocks were increased during runtime to overclock the transceivers to a data rate of 22 Gbit/s. The future Stratix 10 TX FPGA will be available with up to 56 Gbit/s of bandwidth per channel and will likely dissipate even more power.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

(a) Delidded Stratix 10 FPGA



(b) Stratix 10 FPGA die layout

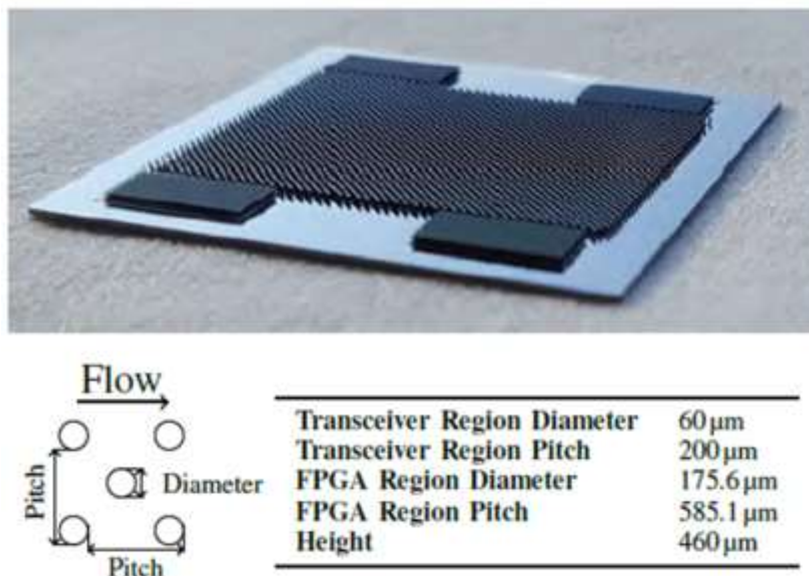
Die	Power	Area	Power Density
FPGA	110 W	5.8 cm^2	19 W/cm^2
Transceiver Tile 0	21 W	0.38 cm^2	56 W/cm^2
Transceiver Tile 1	21 W	0.38 cm^2	56 W/cm^2
Transceiver Tile 2	21 W	0.38 cm^2	56 W/cm^2
Inactive Tile 3	0.72 W	0.38 cm^2	1.9 W/cm^2

Figure 10. Delidded S10 FPGA package showing the center FPGA surrounded by 4 transceivers using EMIB. Power dissipation of each die also shown.

One temperature sensor on each die is read using a Nios II soft processor on the FPGA which feeds these numbers back to an attached computer for logging. Voltage and current on the power rails of the board were measured through on-board sensors which interface to a board test system

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

(BTS) interface that comes with the development board. The complete microfluidic heat sink assembly consists of two parts: the etched silicon heat sink, through which heat is transferred to the fluid, and a 3D printed plastic enclosure which encases the silicon insert and routes fluid between the inlet/outlet tubes and the silicon micropin-fins. A photograph of the silicon micropin-fin heat sink can be seen in Fig. 11. The heat sink was fabricated through Bosch process etching of a 900 µm thick silicon wafer to a depth of approximately 460 µm.



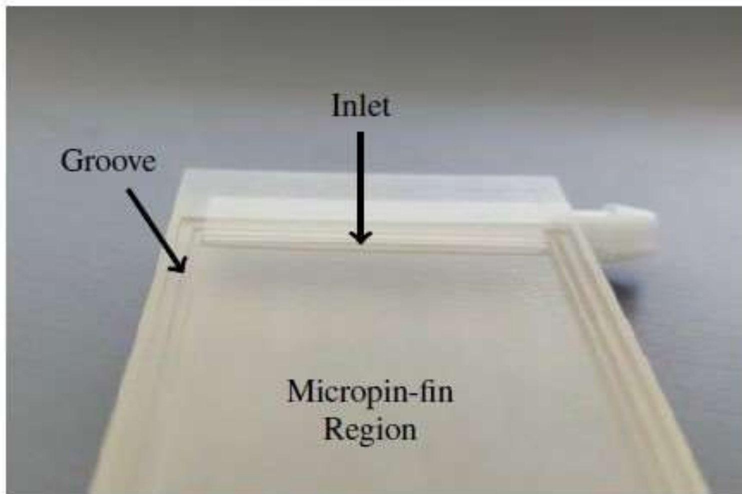
Transceiver Region Diameter	60 µm
Transceiver Region Pitch	200 µm
FPGA Region Diameter	175.6 µm
FPGA Region Pitch	585.1 µm
Height	460 µm

Figure 11. Design and Fabrication of multi-diameter and multi-pitch micropin-fins to enables local optimization of hea removal across the FPGA ans transceiver dice.

The 3D printed enclosure can be seen in Fig. 12. A recess with a nominal depth of 450 µm exists for the silicon heat sink. A further recess with an additional nominal depth of 450 µm exists to enclose the micropin-fins. The height of the micropin fins was made to be approximately 10 µm taller than this recess to reduce the likelihood of a gap existing between the tops of the micropin-fins and the plastic. Since the micropin-fins were designed to be taller than the cavity, the edges of the die were not expected to touch the edges of the plastic, but this gap was filled with epoxy. Epoxy was dispensed with a syringe along the edges of the cavity before inserting the silicon die. A groove was added between the edges where epoxy was applied and the micropin-fin region, so that epoxy which was pushed out during assembly would fill this groove before clogging the micropin-fins. After assembling the complete microfluidic heat sink, MasterGel Pro thermal interface material (TIM), which has a thermal conductivity of 8 W/(m K), was applied to the back sides of the five Stratix 10 dice and the heat sink was mounted on top. Pressure was applied using a custom 3D printed mounting bracket. The edges of the heat sink nearest to the inlet and outlet were visually aligned to the edges of the package, while the other two edges were aligned to the edges of the mounting bracket, which was designed to be situated in the center of the package and have the same width as the heat sink enclosure. Images of the mounted heat sink can be seen in Figure 13.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

(a) Top view of 3D printed enclosure



(b) Close up of enclosure features

Figure 12. 3D printed lateral inlet/outlets for the silicon micropin-fins.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

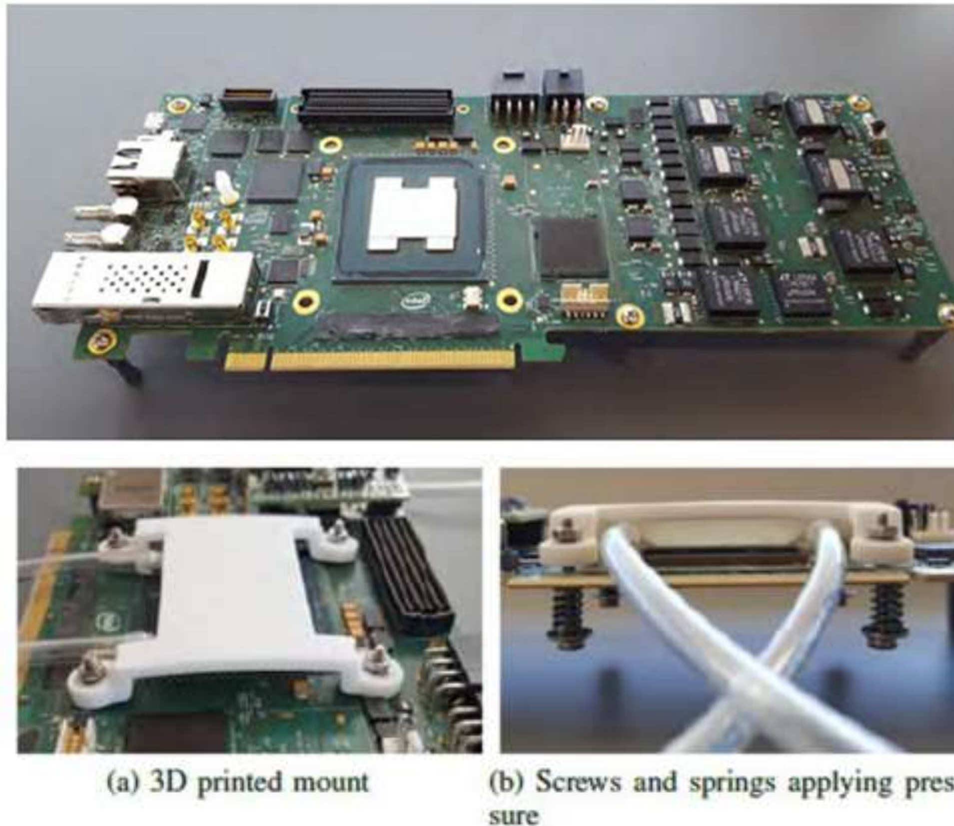


Figure 13. Images of the delidded 2.5D FPGA package on the board (top) and images of the FPGA with microfluidic cooling using 3D printed lateral manifolds (bottom).

The assembled heat sink and board were tested in an open loop system with deionized water as a coolant. Temperature measurements were made at the inlet, outlet, and in the surrounding ambient air using K-type thermocouples. Flow rates were measured with a Kobold rotameter and Omega electronic flow meter, both calibrated through repeated filling of a known volume of fluid. Calibration took place with deionized water at ~21.2 ^{o}C, which was within 2.4 ^{o}C of all fluid inlet temperatures used for testing. Die temperatures were measured using on-die temperature diodes with temperature measurement IP integrated into the FPGA design. An initial experiment was conducted at the lowest flow rate to find the time necessary for the temperatures to reach steady state. Little systematic variation was observed after the first measurement. Nonetheless, temperatures were allowed to stabilize for one minute prior to taking all measurements with the microfluidic heat sink. A similar experiment was conducted with the air cooled heat sink, which took significantly longer to stabilize. Data was then taken for approximately one minute at a rate of approximately four data points per second. The averages of die temperatures over these one-minute periods are reported in this section. The maximum standard deviation of any die temperatures across any of these periods was 0.87 ^{o}C. The temperature of the FPGA die as well as the three active transceiver dice can be seen as a function of flow rate in Fig. 14. Pressure drop vs. flow rate is shown in Figure 15. The FPGA die temperature measurement is the lowest of the four dice, while Tile 2 has the highest temperature due to its close proximity to the outlet. The temperature difference between Tile 2 and the other

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

tiles decreases as flow rate is increased and the fluid temperature at the outlet drops. These results agree reasonably well with those from simulations. The measured FPGA temperatures were approximately 3 °C to 4 °C higher than the simulated temperatures of the region above the FPGA and the measured transceiver temperatures were approximately 6 °C to 8 °C higher than the simulated temperatures of the regions above the transceivers. It is expected that the die temperatures are higher than those at the base of the heat sink (where temperature was simulated) due to the heat conduction through the TIM. Reducing the TIM bond line thickness may be an area for further temperature improvement.

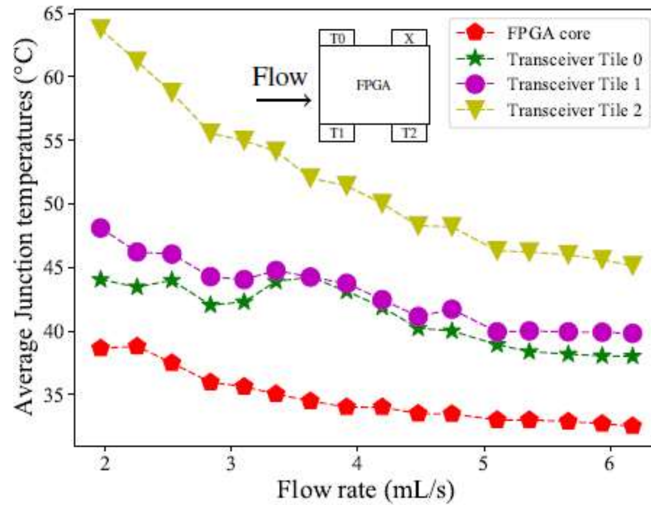


Figure 14. Average junction temperature vs. flow rate for all five dice on the package vs. flow rate.

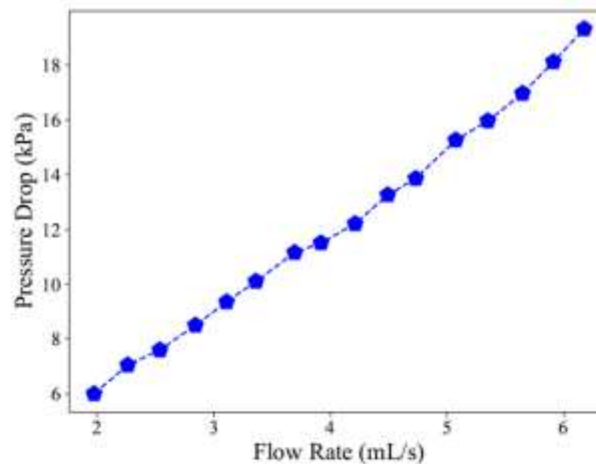


Figure 15. Pressure drop vs. flow rate for the micropin-fins.

Multiple experiments were performed to benchmark both the microfluidic cooling solution and the air cooled solution. For example, an additional experiment was performed in which the power on the FPGA die was modulated by varying the clock frequency to the FFT computational blocks. Under air-cooling, the temperature of the FPGA die increases with increasing power, with a slope of approximately 0.46 °C/W. However, it can also be seen from the data that was

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

collected that the slopes of the transceiver die temperatures are all approximately equal to the slope of the FPGA die temperature. From the lowest FPGA die power of 75.6 W, corresponding to an FFT clock frequency of 300 MHz, to the highest FPGA die power of 113 W, corresponding to 475 MHz, the temperature rose 17.0 ^{o}C on the FPGA die while the temperatures of transceiver tiles 0, 1,and 2 rose by 18.0 ^{o}C, 15.9 ^{o}C, and 16.8 ^{o}C, respectively. This indicates strong thermal coupling between adjacent dice (in air cooling due to the heat spreader), where the conditions on the FPGA die have a large effect on the temperatures of the surrounding dice. The power of the FPGA die was similarly varied with the microfluidic cooled heat sink with a flow rate of 6.18 mL/s and an inlet temperature of ~19.5 ^{o}C. The FPGA die temperature increased with increasing power, with a slope of approximately 0.057 ^{o}C/W, which is significantly lower than the slope seen with the air cooled heat sink because of the comparatively lower thermal resistance of the microfluidic heat sink. While the temperature of the transceiver tiles with air cooling closely followed the temperature of the FPGA die, the temperatures of transceiver tiles 0 and 1 are nearly constant with microfluidic cooling as the FPGA die power changes from 77.5 W to 117.4 W.

The average Tile 2 temperature measurement increases by 1.7 ^{o}C because it is located close to the outlet and likely receives more fluid which has been warmed by the FPGA die. Transceiver powers were also varied with the microfluidic heat sink with a flow rate of 6.19 mL/s and inlet temperature of ~19.2 ^{o}C. While transceiver temperatures dropped by 10.1 ^{o}C to 14.6 ^{o}C when total transceiver power was changed from 63.34 W to 22.61 W, the average temperature measurement on the FPGA die only changed by 0.13 ^{o}C. The effects of FPGA die power on the steady state temperatures of all four dice with both the air cooled heat sink and microfluidic cooled heat sink are summarized in Table II. The effect of the FPGA power on the surrounding transceiver die temperatures was reduced by a factor of 10x to over 100x when compared with the air cooled heat sink. This significant reduction in thermal coupling between adjacent dice is likely a result of the significantly reduced thickness of the microfluidic heat sink. Heat is rapidly transferred to the fluid and extracted, whereas heat must conduct through a large shared thermal mass before being extracted by the air cooled heat sink. This also has an effect on the thermal time constant of the two systems. It was observed that the time necessary for temperature to reach a steady state with the air cooled heat sink was approximately 10 min to 15 min, but only 1 s to 3 s with the microfluidic heat sink.

Die	Air	Microfluidic
FPGA	0.46 °C/W	0.057 °C/W
Transciever Tile 0	0.48 °C/W	0.003 °C/W
Transciever Tile 1	0.43 °C/W	0.004 °C/W
Transciever Tile 2	0.45 °C/W	0.045 °C/W

Figure 16. Summary of thermal cross-talk between dice in the 2.5D FPGA when using air-cooling and microfluidic cooling.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

6. MICROFLUIDIC COOLING USING MONOLITHIC COPPER MICROPIN-FINS

During this program, we explored two options for monolithic microfluidic cooling. In this section, we describe the option of copper micropin-fins on the back side of chips using localized electroplating, and in the next section (Section V), we will report our results on monolithic silicon micropin-fin integration on the dice. Figure 17 illustrates the vision: copper micropin-fins on the back side of the silicon chips, encapsulated by 3D printed lateral manifolds (i.e., the same 3D printed lateral manifolds in the previous section). The potential benefits of this approach include: 1) copper has better thermal conductivity than silicon and thus improving micro-pin fin efficiency and heat removal, 2) the fabricated copper micropin-fins can be made very tall in height thereby reducing pressure drop (due to larger hydraulic diameter) and improving heat removal (due to larger effective surface area), and 3) eliminate the need for direct silicon etching, which is a “more fabrication risky process” due to the limited number of 2.5D Stratix 10 FPGA chips/boards.

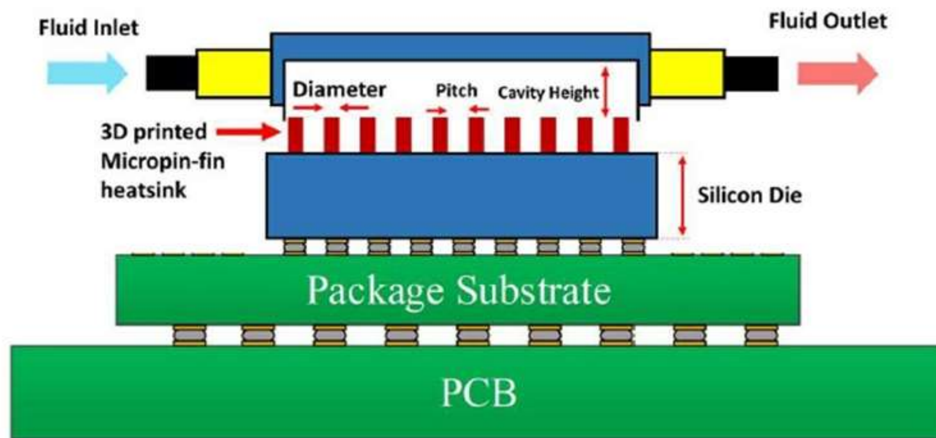


Figure 17. Vision of locally electroplated copper micropin-fins and 3D printed lateral inlet/outlet manifolds.

The approach we undertook to enable this effort is based on localized electroplating, which occurs by saturating a cotton fibers tip with a copper electroplating solution and applying a bias between the tip and the chip, as shown in Figure 18. While the bias is applied, the electrolyte saturated tip is slowly moved upward in a very controlled manner; by doing so, one can begin to create copper micropin-fins. Figure 19 illustrates the developed system to enable localized copper electroplating with integrated electronics to control the mechanical motion of the printing tip, which is critical to the success of this approach. Colored SEM images of the copper micropin-fins is shown in Figure 20, along with all dimensions experimentally demonstrated. The printed copper micropin-fins were greater than 1 mm in height, which can significantly reduce pressure drop (Figure 21). These devices were fabricated and packaged to enable microfluidic cooling testing; pressure drop data was collected while thermal data was more challenging to collect as the thin-film heaters were damaged during testing. While the demonstrated dimensions are promising for microfluidic cooling (as their dimensions are comparable to some copper cold plate solutions), future work should focus on methods to scale down the copper micropin-fin diameter and pitch.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

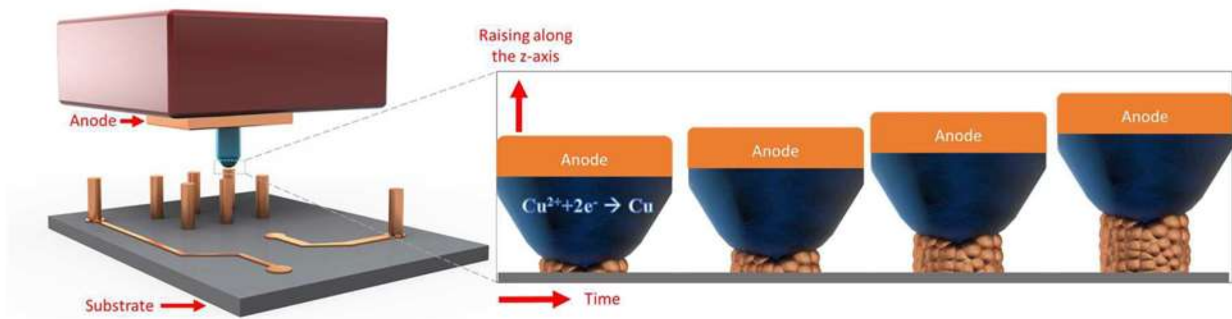


Figure 18. Schematic of the 3D copper micropin-fin fabrication process: cotton fibers (printing tip) are saturated with a copper electroplating solution and are used to locally fabricate the copper micropin-fins by applying a bias (between anode and substrate).

In Figure 18, the bias is applied while simultaneously slowly moving the printing tip away from the substrate resulting in copper micropin fin pillars.

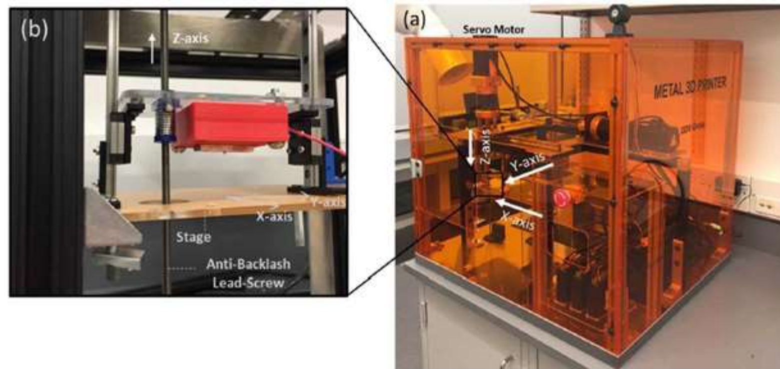


Figure 19. Developed system for copper micropin-fin fabrication on silicon chips and wafers using localized electroplating.

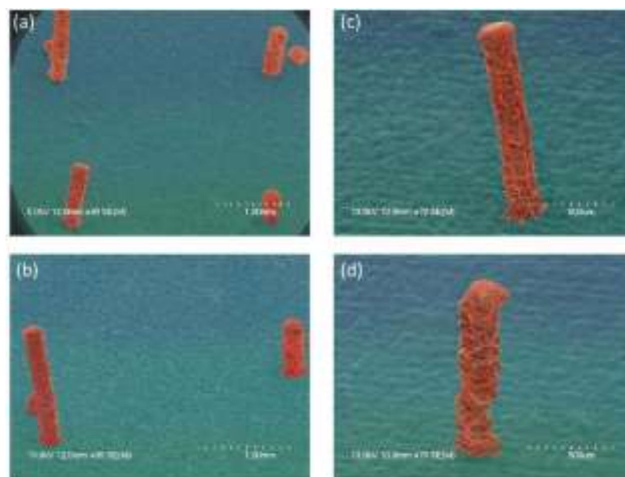


Figure 20. Colored SEM images of the fabricated copper micropin-fins on silicon chips.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

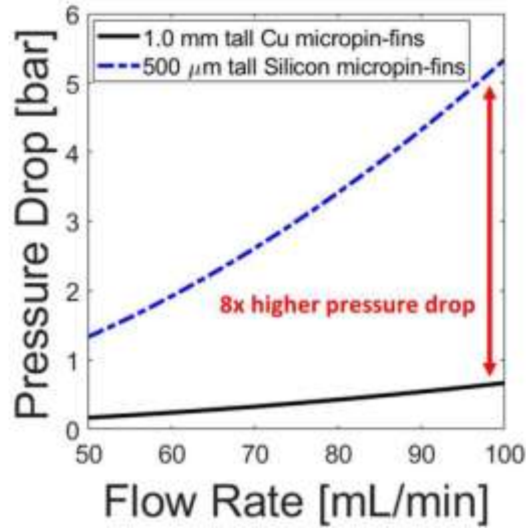


Figure 21. Pressure drop benefits from the 1mm tall copper micropin-fins.

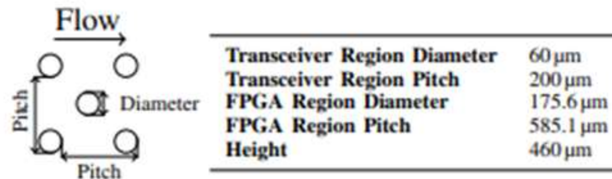
Sample#	Number of Micropillars	Average Diameter	Standard Deviation	Average Tip Diameter
1	10	253.67 μm	26.95 μm	213 μm
2	9	228.55 μm	38.43 μm	175 μm
3	9	203.93 μm	16.39 μm	148 μm
4	11	362.73 μm	8.29 μm	320 μm

Figure 22. Summary of all dimensions demonstrated with the copper micropin-fins. The pitch was approximately 1 mm.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

7. MONOLITHIC MICROPIN-FIN INTEGRATION INTO 2.5D STRATIX 10 FPGA AND TRANSCEIVER DICE

In the second year of the program, we began to receive more Stratix 10 FPGAs from Intel, which allowed us to experiment with the 'more high-risk fabrication process' of silicon micropin-fin monolithic integration into the 2.5D FPGA and transceivers. To deal with the heterogeneous power densities of the transceiver and the FPGA dice, micropin-fins of different densities were ***etched directly*** on the transceiver and FPGA regions, as shown in Fig. 23. To ensure leak-proof fluid delivery to all the dice, 3D printed manifolds that conform to the profiles of individual dice are used, similar to the ones discussed in prior sections. The manifold was designed as a two-part solution (Fig. 24), with a frame attached to the package with epoxy, and a capping manifold with fluidic routing channels and ports to deliver fluid to each dice. The capping manifold was designed to snap on to pre-designed grooves on the attached frame. The capping manifold was designed to have fluid routing cavities to direct fluid to different dice as well as individual ports to deliver fluid to different dice. Custom silicone gaskets were inserted between the delivery ports and etched silicon.



Transceiver Region Diameter	60 µm
Transceiver Region Pitch	200 µm
FPGA Region Diameter	175.6 µm
FPGA Region Pitch	585.1 µm
Height	460 µm

Figure 23. Micropin-fin dimensions

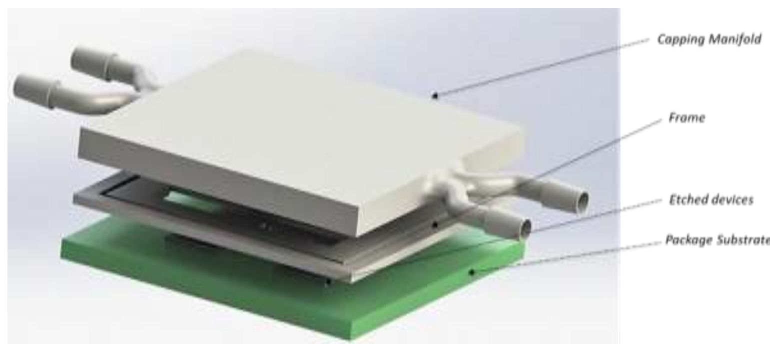


Figure 24. Conceptual image of assembly process

The micropin-fin densities were scaled to the ratio of power densities of the dice to achieve a more uniform temperature profile across the package and minimize thermal coupling between the dice. Another design technique employed to minimize thermal coupling is to split the fluid paths between the low power FPGA die and the higher power transceiver dice to prevent the heated fluid from high power areas heating up the lower power areas. Detailed design of the manifold and the frame is shown in Fig. 25.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

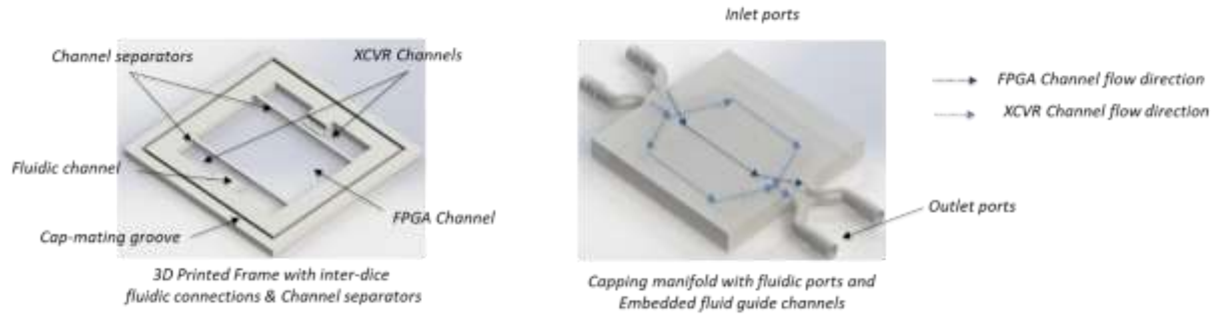


Figure 25. Design details of frame and capping manifold

To test the concept, the 3D printed manifold was designed using Solidworks 2018 and printed with 3D Systems Visijet M3 Crystal material jetting polymer from 3D systems, using a Projet 3510HD industrial 3D printer. Fig. 26 shows the photographs of the 3D printed frame and manifold. An optimized BOSCH silicon etching process was used to etch pinfins of different dimensions on the FPGA and transceiver regions simultaneously. The process was first tested on a de-lidded electrically non-functional Stratix-10 FPGA package from Intel. The process was optimized to pattern and etch the silicon, without damaging the organic substrate, while accounting for the height difference of various dice and the package region. Fig. 27 shows the etched mechanical package, and Fig 28 shows close-up SEMs of the etched micropin-fins. A mean micropin-fin height of 378.337µm was etched, with no visible damage to the organic substrate.

At the time of writing this report, we are in the process of making a final 2.5D FPGA sample with monolithic microfluidic integration and benchmarking its electrical performance in a similar fashion to the experiments noted in Section III. We will update the DARPA team on the results of the experiments.

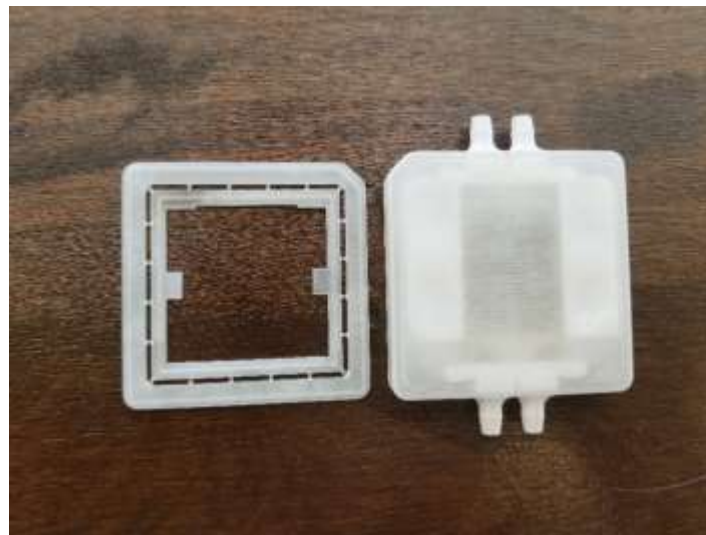


Figure 26. 3D printed frame and capping manifold

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

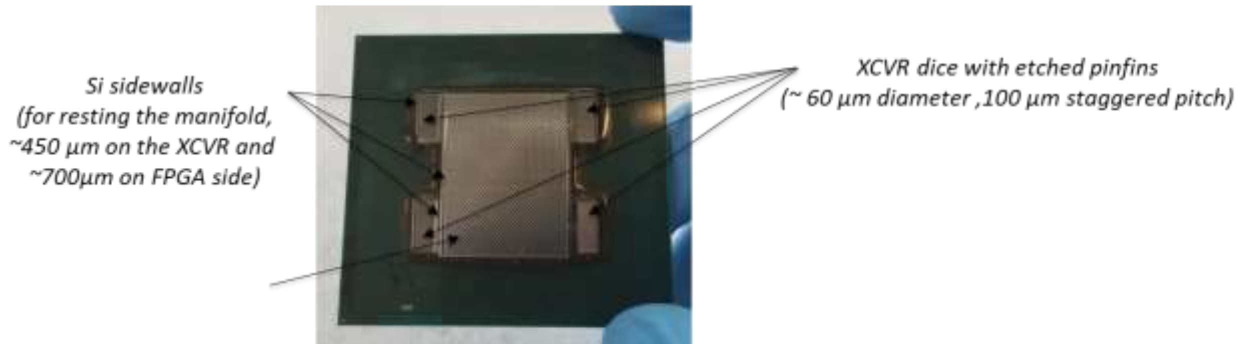
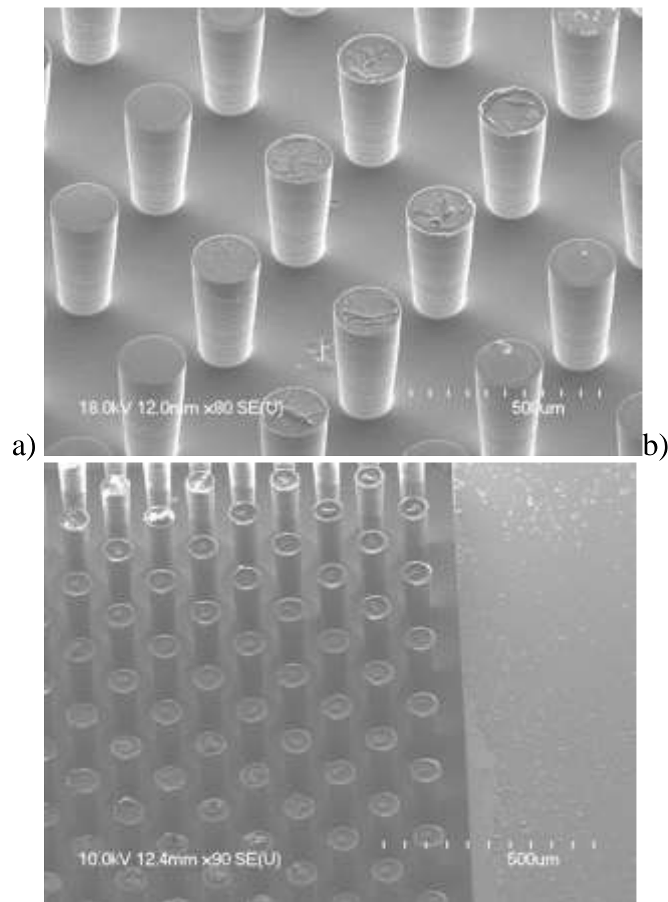


Figure 27. Silicon micropin-fins etched monolithically into the 2.5D FPGA/transceivers



DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

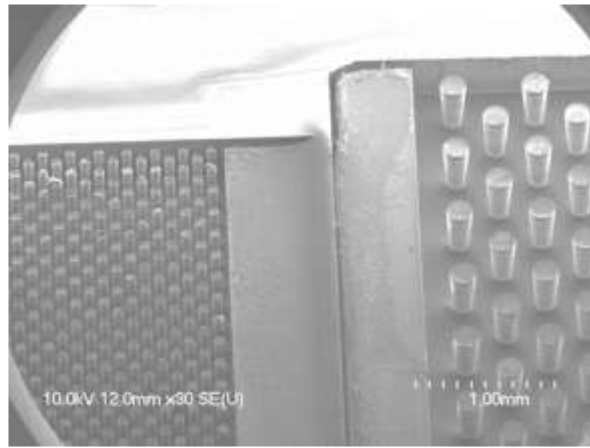
c) 

Figure 28. SEM images of the etched micropin-fins monolithically into the 2.5D FPGA and transceivers: a) Micropin-fins over FPGA region, (b) Micropin-fins over XCVR region, (c) Heterogeneous micropin-fin design over XCVR and FPGA regions

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

8. PRELIMINARY RELIABILITY STUDIES

In this experiment, we performed thermal cycling tests of a 2.5D testbed with monolithic microfluidic cooling. The system contains three dice with micropin-fin heatsinks, encased with a 3D printed microfluidic manifold for fluid delivery (Fig. 29). The testbed was designed to emulate a CPU-FPGA-Memory multi-die assembly, which is a relevant use- case for data-center applications. The CPU die and FPGA die were designed to have 1 cm ×1.5 cm lateral dimensions while the memory die has a 1 cm×2 cm area. Each die has 200μm DRIE etched micropin-fin region atop and serpentine platinum coils as heat sources on the bottom side.

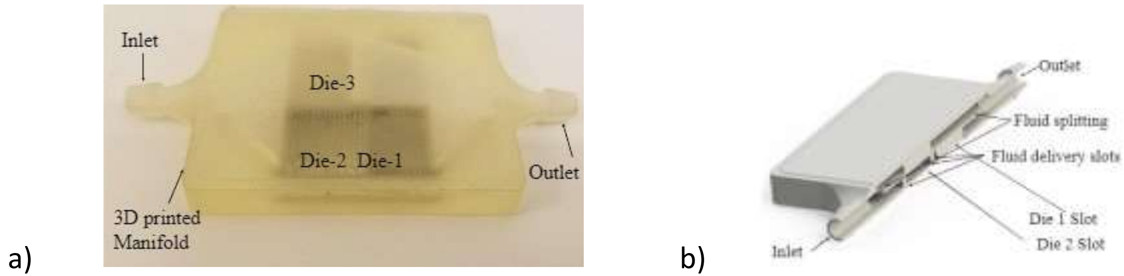
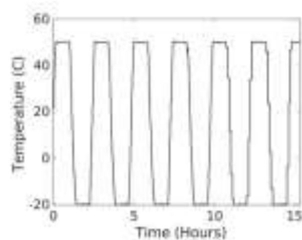


Figure 29. Samples used for reliability studies: a) Image of the manifold after die attach, b) Rendered cross-section of the manifold showing fludic channels

The manifold was designed to be a single inlet-single outlet solution with no inter- die fluidic connections. The input and output ports are 0.11 inches in diameter. The inlet port was split into 3 unequal channels inside the manifold, each feeding an individual die. The outlet port was also designed with the same symmetry. This split was done to a first order scaling for the power dissipation for each die. The structure was designed using Solidworks 2018 and printed with 3D Systems Visijet M3 Crystal material jetting polymer from 3D systems, using a Projet 3510HD industrial 3D printer.

Fluidic measurements were done in an open-loop measurement system, and the pressure-flow-rate relation of the assembly was obtained. The device was then subjected to thermal cycling between -20 °C and 50 °C in an Espec SH241 Temperature and Humidity chamber, controlled by a LabView program over a GPIB interface. The upper limit of temperature was chosen as 50 °C for this demonstration as the Visijet M3 Crystal material jetting polymer used for this demonstration has a softening point of 56 °C. Results for pressure-drop before and after thermal cycling are shown in Fig. 30 and suggest the heat sink was not impacted by the thermal cycling. Future work includes more thermal cycling tests, and doing so with the liquid cooled FPGA.

Thermal Cycling Profile   Fluidic Measurement Results



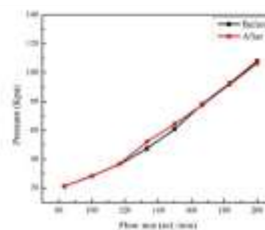


Figure 30. Thermal cycling conditions and results

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

9. DARPA ACCELERATOR

Towards the end of this program, DARPA awarded us a new task to explore potential technology transition of the research. As of the writing of this report (Nov. 12, 2019), the DARPA accelerator effort continues and thus, a separate report was submitted on Jan 2020. The complete final report is shown in Appendix A.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

10. CONCLUSIONS

In this program, our focus was on developing microfluidic cooling technologies for 2.5D electronics, to improve performance, increase compute density, and increase energy efficiency. To accomplish these goals, we performed both ANSYS thermal modeling, and built on prior developed fundamental finite-volume method modeling (supported by NSF) to benchmark various cooling scenarios. We explored multiple technology options for the efficient integration of microfluidic pin-fin heat sinks with electronics using a) silicon micropin-fin heterogeneous integration, b) silicon monolithic integration, and c) copper monolithic micropin-fins using localized electroplating. Moreover, we developed lateral manifolds using 3D printing to enable ultra-compact inlet/outlet manifolds; this approach is compatible with all three micropin fin heat sink integrations noted above and provides efficient fluidic delivery at extremely compact form factors. A key factor in the research was the limited availability of the new Stratix 10 boards during the research period; as such, in discussions with DARPA we opted to demonstrate our initial electrical measurements using the silicon micropin fin heterogeneous integration approach. At the time of writing this report, we continue to do final experiments on the 2.5D FPGA silicon using monolithic micropin-fins.

This program funded a number of graduate students at various levels of effort/time and provided them with a very unique interdisciplinary research experience covering electrical engineering (FPGA programming, electrical modeling), mechanical engineering (cooling and thermal modeling), and chemical engineering (microfabrication in the cleanroom).

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

APPENDIX: DARPA ACCELERATOR FINAL FINDINGS

DARPA Accelerator ICECOOL 2.5-D/3-D Microfluidic Cooling Final Report
POC: Jonathan Goldman, Principal, Georgia Tech VentureLab
Team: Professor Muhannad Bakir, Professor Yogendra Joshi, Mr. Jonathan Goldman

SUMMARY: High-level Take Away Learnings:

1) Product-Market Fit: Next generation, high performance computing desperately needs new cooling solutions; improving the state-of-the-art will not provide sufficient cooling needs to address coming performance requirements. The industry is already accommodating these limitations through methods like frequency and voltage throttling; and allowing for 'dark silicon' during certain kinds of high-stress compute operations. This is being keenly felt in the HPC market segment. Unfortunately, this reduces computing performance significantly especially as industry continues to move towards 2.5D and 3D integration. Radical thinking is needed, making microfluidic cooling a must- have solution to industry.

2) Monolithic vs. Attached Cooling: We assumed that etching and directly bonding a separate silicon wafer with microfluidic cooling onto the backside of the processor would be the least intrusive and easiest minimum viable product (MVP) for inserting the technology into the industry. *However, we were surprised to find that market thought leaders want to adopt the monolithic integration approach*: etching these cooling structures directly onto the backside of the silicon chip/wafer. Positive feedback was received from Google, Microsoft, Cerebras, and Northrup Grumman; however, work needs to be done to enable the supply-chain to deliver the integrated solution. This is the biggest obstacle we ran into and one that has slowed down market adoption. Unless this is addressed, it is difficult to foresee monolithic microfluidic cooling being adopted.

3) Solution Design/Integration: Industry does not know how to make the integrated silicon microfluidic package and how to connect to it fluidically. The challenge is that implementing this approach into the design integration, fabrication, and packaging supply-chain is going to take leadership from industry standards bodies, such as SEMI. As we move towards an era wherein "the silicon becomes the cooling package," integration issues must be addressed. This means delivering an integrated solution that includes the fluidic 'balance of system:' the capping layer, quick connect/disconnect fluidic connections, pumping, and electrical co-design. Work must be done to address issues surrounding standards, solution design, materials selection, reliability, and testing/burn- in. While there is significant excitement in the industry about what the ICECool program has accomplished, there appears to be a need for a focused program addressing how to push monolithic microfluidic cooling into the supply chain and how to address the packaging and co-design issues; in other words, the solution must be much more than the silicon heat sink.

4) Manufacturing/Supply-Chain Integration: The solution lies at the intersection of the fab, the OSAT and the customer. In order to address this supply chain question, we spoke to

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

GlobalFoundries about their ability to absorb such a process; we spoke with the CEO, CTO, and director of 3D integration and packaging. Their response was that they alone could not do it as this technology would need an innovative co-packaging solution and would require an OSAT partner such as Amkor. We spoke to high-level executives at Amkor, and their feedback was that they are not equipped to handle such radical technologies because they do not commonly work with silicon processes. There is need for research and development investments into the packaging of silicon monolithic microfluidic cooling solutions. Fabs such as Intel and GlobalFoundries and OSATs such as Amkor and ASE Group will not drive this technology forward without a "Prime" demanding it. ***We are pursuing prototype demonstrations with Microsoft and Cerebras, at this time, that would move this technology towards potential transition to the appropriate parties in the supply-chain.***

5) Cost: Some have speculated the cost would be barrier to entry, but our analysis seems to suggest otherwise. Our initial estimates are the solution would cost between $0.20-0.25/W BEFORE scaling, which is below the current direct liquid cooling cost of $0.275/W.

Statement Of Work Milestones

1.1 Application Identification: The initial markets identified: Military (radar), Data center/HPC, Telecom (RF), Power electronics and LEDs were cataloged from a size and CAGR point-of-view and shown in tab labelled "Market" of the attached spreadsheet.

1.2 Initial Competitive Analysis: The major competitors in the liquid cooling space on the market today are cold-door heat exchange units, including Motivair Corp.[1] and OptiCool Technologies[2]. For more challenging heat fluxes, liquid cold-block solutions, such as Asetek[3], CoolIT[4], and CoolerMaster[5] are found in HPC and gaming applications. Oil immersion system providers, including Green Revolution Cooling[6] have the advantage of using a dielectric fluid, but are messier to deal with when charging and servicing in a data center environment. These systems are favored in hot climates, such as in HPC racks built for onsite use in the oil & gas industry.[7] Competitive technologies such as JetCool[8] and IMEC (Belgium) are using jet impingement cooling to use the advantages of phase-change cooling to promote the promise of high rates of heat transfer.[9] A technology that combines microjets and microchannels has demonstrated 1000W/cm^2 in the lab at Purdue, but has not been realized in a product.[10] A list of competitors are included in Appendix A at the end of this document.

1.3 Early Market Analysis: Figure 1 shows an initial survey of the most likely market segments.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

Segment	Base Yr/$	Future Yr/$	CAGR (%)
Military (Radar)	**2017/$337.28M**	**2023/$992.38M**	**19.71**
Power Devices	**2015/$10.57B**	**2022/$31.26B**	**15.4***
Data Ctr/HPC	**2018/$12.9B**	**2022/$19.6B**	**11.02**
RF PAs	**2018/$21.4B**	**2023/$30.6B**	**7.4**
High Brightness LEDs	**2017/$13.8B**	**2024/$27.28B**	**10.20**

*Between 2016 & 2022

Figure A 1. Initial survey of the ost likely market segments 2016-2022

Based upon sheer market size, it appeared that power devices (SiC and GaN HBTs), RF power amplifiers and high-brightness LEDs would be the most interesting markets. However, there were several factors that caused us to rule out these segments as prospects for our technology. While Telecom (RF), power electronics and LEDs all exhibit moderate to high power densities, the following three reasons were found for these segments being reticent to adopt a new cooling solution:

1. Power electronics (wide-bandgap GaN and SiC HBTs), typically carrying higher currents, have larger packages with ample contact pads. Today, these packages use water- ethylene-glycol (WEG) loops or oil cooling and are unlikely to change. While packaging formats are a consideration, a slightly larger package for slightly more liquid or higher pump rates are easier to implement than a completely new approach[11];
2. High-power RF devices live inside of sealed, vented containers on cell towers and cannot tolerate liquid pump failures which could result from freezing temperatures. This makes adoption by military much more difficult.[12], and
3. LEDs are generally very price sensitive and can rely on large luminaire surface area to spread heat and function with only convective air cooling.

This left us to consider HPC and data center applications and the smaller market for military radar.

1.4 Potential Customer Interviews: For interviews around the need for novel cooling solutions in military radar, we engaged in many discussions with Raytheon (Jason Milne) and Northrop Grumman (Girish Upadhya, et. al.). Some of the challenges presented related to the need for completely leak-free fluid connections for field hardware. While these kinds of connectors are available, miniaturized versions to service amplifiers at the device scale are only now entering the market and not yet qualified for military applications (e.g.: Staubli[13] & CEJN[14]). Nonetheless, we were drawn into further discussions by RF supplier as they saw the value of being able to integrate their chips with 2.5D packages.[15] A complete list of Customer Discovery interviews is given in Appendix B.

For early interviews in the HPC space, we were able to speak with Mythic Semiconductor[16],

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

Global Foundries (GF) and Google[17]. All three saw the growing trend in 2.5-D packages. In the case of GF, there was a question as to how they would integrate these newer cooling approaches into the supply chain. They believe the OSATs may need to integrate this solution.[18,19] At the end of our initial market assessment, this is where we landed (Figure 2):

Segment	TDP (W)	W/cm2	Tj(°C)
Military (Radar)	**125***	**500***	**<100***
Power Devices	**660-12,500**	**22-47**	**125-175**
Data Ctr/HPC	**120-300**	**30-40**	**90-105**
RF/Telecom	**125-850**	**152-235**	**150-225**
LEDs	**19-30**	**21-163**	**100-245**

*Data from Northrop Grumman

Figure A 2. At the end of initial market assesment, this is where we landed

1.5 Initial Product Hypothesis: Initially, we assumed the business would include several elements:

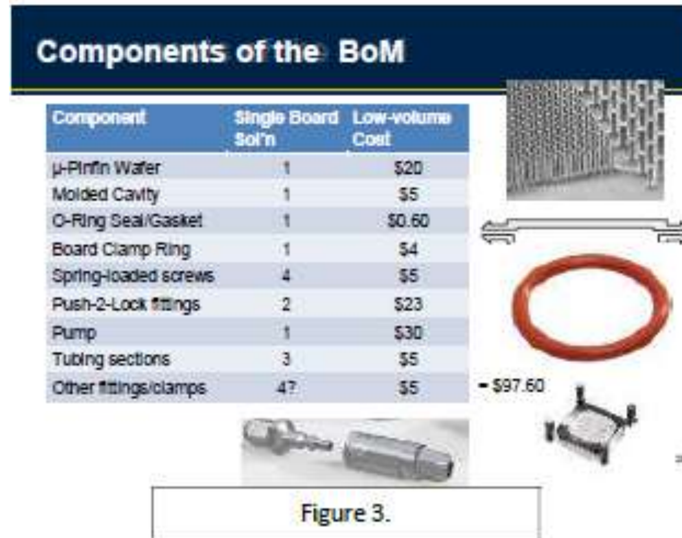
1. The custom modeling and design for each chip's particular micropillar pattern would be done in-house and provided to the fab so the patterns could be etched on the backside.
2. The in-house design and manufacture of the microfluidic housing / capping layer that would encase the top of the chip package.
3. The fluid connectors, tubing, pump and zero-leak fittings that would enable a server or blade to connect to a liquid manifold similar to that which are provided today at the rear of standard computing rack equipment.

While we assumed it would be necessary to design and build #s 1 & 2 above in-house, we assumed the other pieces would be purchased and integrated by whomever was building the server. This would be done internally by the 'hyperscaler' data centers: AWS, Azure, Facebook, etc., or integrated by service providers that build racks for HPC and co-location data centers such as Dell, Penguin Computing, CoolIT, etc.

1.6 Basic Cost Model: The 'fluidic wafers' would be produced by the fab and bonded to the CPU/GPU wafers before shipping to the OSAT for dicing and packaging. We engaged with BRIDG, an 8" (200mm) specialty foundry in Orlando, FL to assess costs for development and scaling. Given the etch process disclosed, they committed to being able to hit a volume of 120 wafers/hr; with future expansion capable of handling 300mm wafers. At a yield of approximately 20 chips per wafer, they estimated the cost per chip would be roughly $20. This cost would scale down further for 300 mm wafers. In addition to this cost element, the other components of the product solution would include the injection-molded fluidic housing, o-ring (to seal the fluidic housing to the chip package), sealing frame (to provide the sealing force between the fluidic housing and the chip), tubing and fittings, a low-profile pump and a set of specialty zero-leak, push-to-connect fittings. We engaged with suppliers such as Staubli and CEJN, who manufacture these 1 mm ID fittings and received volume pricing estimates of $43/set for 316L SS and $23/set for Ni-plated brass.[20] The low-volume BoM estimate is shown in Figure 3. As proposed, the

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

BoM delivers a ‘hot- swap’ fluidic server blade that would connect to fluid supply/return manifolds at the rear of the server/HPC rack.



Component	Single Board Sol'n	Low-volume Cost
μ-Pinfin Wafer	1	$20
Molded Cavity	1	$5
O-Ring Seal/Gasket	1	$0.60
Board Clamp Ring	1	$4
Spring-loaded screws	4	$5
Push-2-Lock fittings	2	$23
Pump	1	$30
Tubing sections	3	$5
Other fittings/clamps	4?	$5

Figure A 3. Components of the BoM

Hypothesis Go-to-Market Strategy: The original strategy revolved around the notion that each chip (CPU/GPU) would have its own thermal profile, including hot spots. This would require each chip to have its own micro-pinfin layout in order to optimize heat removal. Thus, it was originally thought that the design of these structures would be coupled to the manufacture and delivery of the entire thermal soulution by a sstartup. However, unlike liquid thermal solution providers today (e.g.: Asetek and CoolIT) who can integrate their solutions at the server level after the chip is designed, made and packaged, our solution must integrate into the chip fab supply chain. So, while we thought a NewCo could deliver turn-key thermal solutions from design to fab to solution implementation, we had to rethink the delivery of the solution based upon interviews and engagement with the supply chain (Figure 4).

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

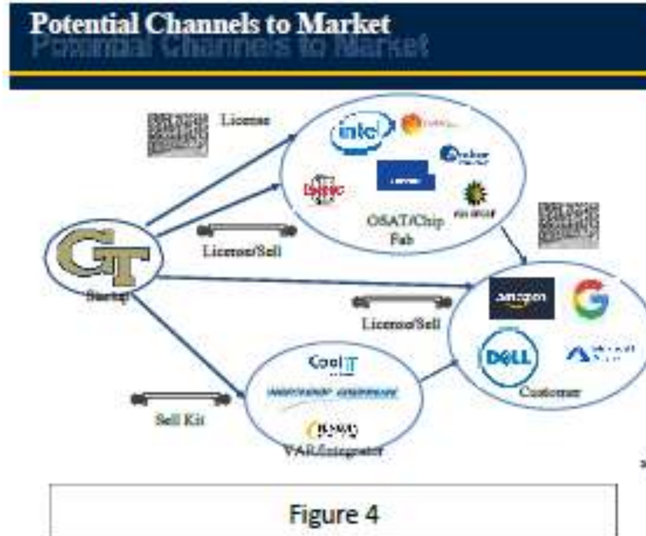


Figure A 4. Potential channels to Market

2.1. Customer Engagement Plan: Based upon initial feedback from the early stages of our Customer Discovery, we decided to focus on 2 main customer segments: 1) high-power amplifiers (HEMTs) for military radars, and 2) data center/HPC cooling of CPUs, GPUs and specialized processors for AI/ML applications. These markets are expected to grow at CAGRs of 19.71% and 11.02%, respectively.

Nearly 60 separate discussion were held throughout the period of the contract. Interviews were conducted through a combination of trade shows and their associated follow-ups (specifically, IEEE's 69th Electronic Components and Technology Conference (ECTC), held May 28-31 in Las Vegas, and Supercomputing 2019 (SC19), held November 18-21 in Denver, Colorado) and a series of phone discussions with industry leaders. Discussions around radar applications were conducted over several conference calls with a team at Northrop Grumman (Linthicum, MD). A listing of dates, interviewees, titles and summaries is shown the tab labelled "Interviews" of the attached spreadsheet. The main takeaways from these interviews are:

- Radar electrical engineers can design systems that outstrip the thermal solutions typically applied. ("Yes, we are thermally limited."[21]) However, as severe as these design requirements appear, there does not seem to be an urgency for delivering this next generation of array performance.
- Liquid cooling is becoming more widely accepted in both the hyperscale data centers and in high-performance computing (HPC), particularly where AI-focused hardware platforms drive density and heat.
- The limits of traditional liquid cooling are becoming evident and there's a desire for improved approaches that can enable either nex-gen compute infrastructure OR drive existing platforms harder.
- Suppliers of this HPC compute infrastructure will accept new, more capable cooling products if offered, but the products must be inserted at the correct point in the value chain.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

- Because our solution essentially replaces the CPU/GPU package lid with our integrated microfluidic solution we essentially turn the package into the cooling solution.
- OSATs like ASE Group and Amkor both stated that they could deliver this kind of integrated fluidic package, but this solution must be driven by the "Primes" (AWS, Microsoft, Google, etc.) and justified with favorable economics by the fab (Intel, GF, TSMC, etc.).[22,23]

2.1.1 Feedback on Radar Application: We held 4 conversation with Northrop Grumman Systems (NGS) around their application. They provided a generic spec for a 10 x 10 array of chips on LTCC substrates (Figure 5a). Unfortunately, although the hotspot was very small (0.001 in^2), the 10,000W/cm^2 spec made for an extremely large localized heat flux that violated their maximum junction temperature limitation. We met the chip base temperature spec of 100 °C at 2000 W/cm^2, but not 10,000 W/cm^2 (Figure 5b).

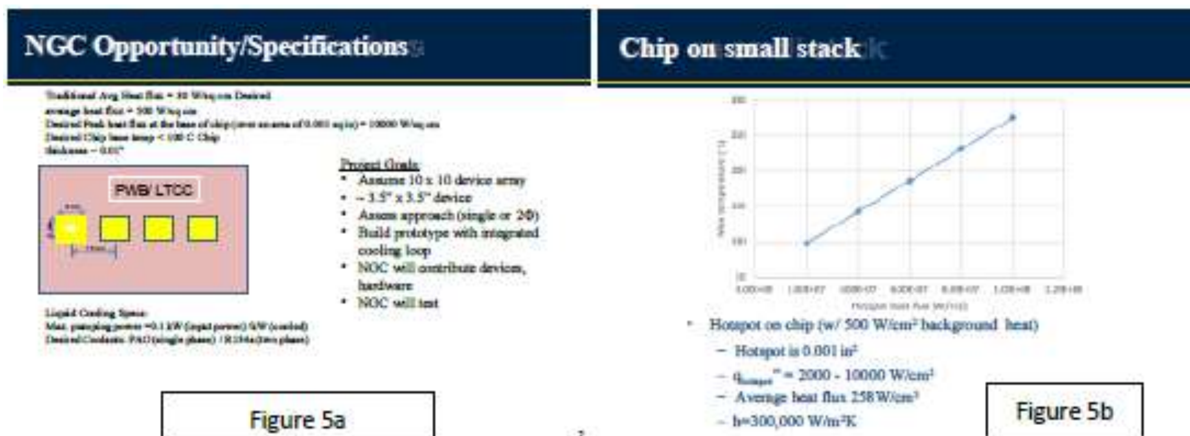


Figure A 5. a) NGC Opportunity/Specifications, b) Chip on small stack

2.1.2. Feedback on HPC/AI Application: We held approximately 35 interviews between 2 trade shows (ECTC and SC2019). While liquid cooling solutions are being implemented in various forms (immersion, pool boiling and liquid cold block), there was ample evidence of 3 trends that spoke to the need for more effective liquid cooling:

- The current trajectory of CPU/GPU power density will begin to saturate the capabilities of direct liquid cooling in the next few years. "Right now, Altera can't give us more than we can handle. We can handle 400 W. But we can't handle 800W."[24]
- Liquid cold block solutions, such as those from Asetek, CoolIT, etc., separately address CPU, GPU and memory modules. As more chips use 2.5-D and 3-D integration, thermal cross talk may become more serious. One example we encountered was with Altera's Stratix X FPGA, where some of the transceiver modules are becoming too hot.[25]
- As rack density increases, servers are pushing up against raised-floor data center load limits. All the copper being used, and all the liquid required for circulation is adding significant weight to the racks. Standard racks are 1200mm deep and standard servers are

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

1000mm deep, leaving only 200 mm at the rear for Power Distribution Units (PDUs) and liquid flow and return lines. For high-density (HPC) racks, the challenges of fitting enough PDU equipment and achieving the required tubing bend-radius within that same 200 mm footprint are significant. One expert with 40 years of data center design and implementation experience indicated raised-floor load limits are (approximately) 350 kg/m^2. Since a rack is 0.72m^2, this would place a per-rack load limit of 252 kg (554 lbs). He has seen specifications for 90 kW/rack, and all the additional copper, cold-doors and PDUs can push weights beyond the load limit before ever coming close to 90 kW.[26] A more efficient cooling method would reduce the fluid weight. An increase in larger SoP die utilizing an integrated 2.5-D cooling solution could reduce the amount of copper required (our solution require none), thus realizing both weight and cost savings.

2.2 Application Identification for Entry Market: Based upon our interviews, it became evident that the HPC market has the potential for the most immediate adoption. As discussed below, we now understand the product insertion strategy into the supply chain, and are now engaged with 2 potential customers with a defined need: Cerebras Systems and Microsoft's Olympus server platform. As chip densities approach 500-1000 W/cm^2, the attractiveness of our solution becomes compelling. While we do not see power electronics for EV applications exceeding their requirements for water-ethylene-glycol (WEG) loops, it is possible that HPC applications for on- board EV applications (processing vision and LIDAR information) may push those cards to consider a monolithic cooling approach, but we do not yet have sufficient evidence to know when this segment might materialize.

2.3 Refined Competitive Analysis: We received detailed cost information when engaging with CoolIT at ECTC. Their CDU rack costs $60,000 and has enough pumping power to handle up to 10 racks, so this element will add $6K/rack. Copper-based chip and DIMM plates will cost about $8K/rack. Liquid manifold will add another $8K/rack for a total of $22K/rack. CoolIT claims this will handle up to 80kW, which nets to $0.275/W.[27] This is below the $0.60/W cost target driven by Intel as previously given to us by the former CEO of immersion cooling provider Green Revolution Cooling.[28] Thus, our target must be below this cost. Given our BoM of < $100 (Figure 3), and therefore a healthy margin at a price of $200, it is anticipated that a product that could handle a TDP of 1000W would therefore cost about $0.20/W.

2.4 Refined Value Chain and Markey Analysis: The nature of the BoM for our solution consists of 2 primary types: 1) those elements that are customized to the particular chip or MCM: the micropillar array and fluidic housing or molded cavity (items 1 & 2 in Figure 3), and 2) those that are fairly invariant across customers: the tubes, fittings, clamps, brackets and pump (items 3 – 9 in Figure 3). While items 1 and 2 will have to be licensed, since these represent the core of the innovation, a NewCo may need to be retained to do the initial design work and may represent on-going non-recurring engineering revenue opportunities. This is not uncommon for disruptive innovations: the industry must be taught how to use the technology since the intended licensees: chip fabs and OSATs may not have the requisite thermal design and fluidics competencies internally. BoM Item #1 will have to be done by the fab, whether captive (Intel, AMD, Samsung) or outsourced (TSMC, Global Foundries, etc.). BoM #2 would be

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

molded by the OSATs since the cooling solution is now essentially integrated with the package. The rest of the BoM (balance of system) would be spec'd components purchased by end customer. This end customer would come in four varieties:

1. hyperscaler data center (e.g.: Amazon, Google, Microsoft, Facebook);
2. HPC solution providers (e.g.: Cray, DellEMC, Fujitsu, HPE, etc.) that builds clusters for their customers colo data center space;
3. VARs or integrators that build special-purpose HPC clusters (Penguin Computing, etc.), and
4. Cooling solution providers that integrate solutions on behalf of their data center customers.

These supply-chain relationships can be seen in Figure 6.

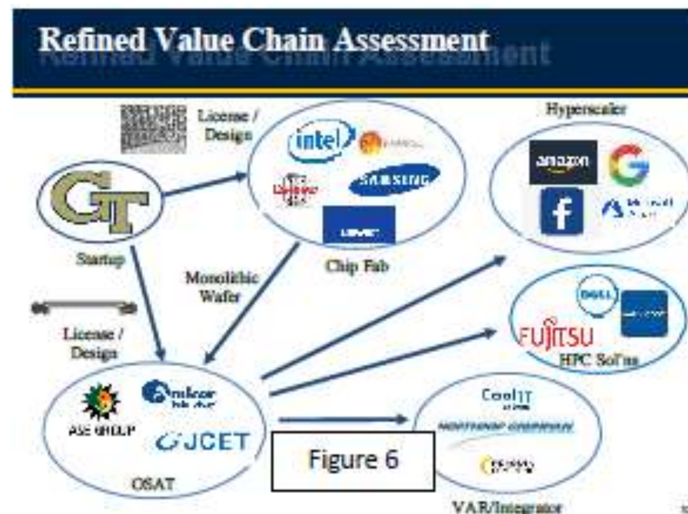


Figure A 6. Refined Balue Chain Assessment

2.5 Refined Product Hypothesis: There are essentially 2 potential embodiments of the product, depending upon where and how these microfluidic cooling structures are created.

1. In this 1st embodiment, a true monolithically integrated solution (Figure 7a) would require the backside (topside) of the wafer to be etched with the micro-pillar structures to dramatically enhance surface area. A fluidic housing would then seal to the lower half of the chip package (properly underfilled) to enable a closed fluidic loop. This fluid loop would be pumped with a small form-factor pump that would be located at the rear of the server and connect to the external supply and return through small zero-leak connectors that are commercially available, which would enable 'plug-and-play' functionality. Although there are potential issues associated with ensuring that processed wafers will have the required structural robustness to accommodate the topside etching process, we heard from several potential users (Google, Cerebras & Microsoft) that the processing challenges were tractable. They strongly felt the monolithic integration approach was much more compelling than the 2nd embodiment as the preferred Go-To-Market strategy.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

2. The 2nd embodiment (Figure 7b) would include a 2nd, blank, Si wafer in which the micro pillars are etched, then bonded to the smooth topside of a standard CPU or GPU. Each CPU/GPU chip design would require its own custom micro pillar wafer to be bonded to it; and would require its own custom fluidic housing. Specifications such as pinfin diameter, height, density will be custom to each chip design and can include variations therein to accommodate features such as hotspots across the CPU/GPU. This oxide or copper bond will achieve virtually identical results with little to no disruption in the current manufacturing sequence. The rest of the Bill-of-Materials would remain the same. This is an MVP that could be delivered to the market with less technical risk; capable of handling many-100sW of TDP with a pump that could deliver flow rates at pressure drops that would be a function of the individual chips micropillar density and layout. Although more straightforward from an integration point-of-view, the customers we spoke with did not feel there would be enough of an advantage in this approach to justify the additional complexity.

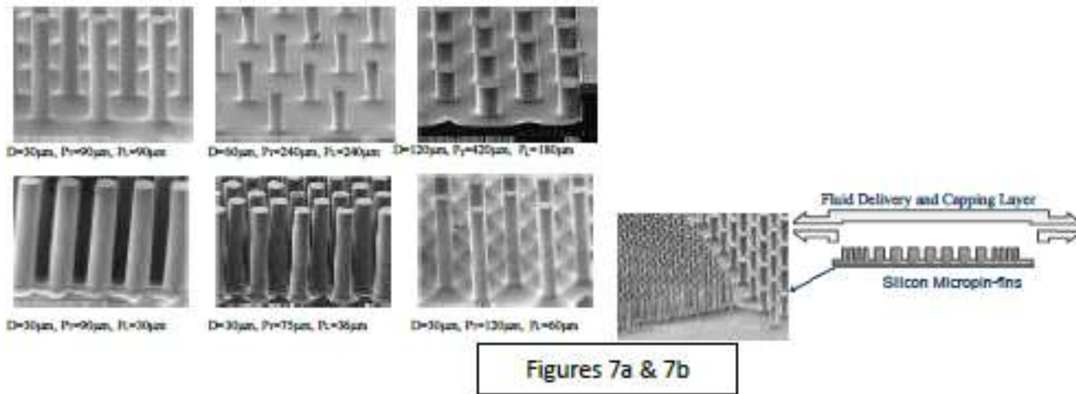


Figure A 7. Refined Product Hypothesis a) in this 1st embodiment, a true monolithically integrated solution and b) the 2nd embodiment

2.6 Refined Cost Model: Using reasonable projects and data collected from vendors that make equipment and components for the liquid cooling space, we estimated the cost for our solution at a ‘scaled’ level (100,000 units per annum production, or greater). Although the cost of the etched wafer is estimated to be cut in half ($20 to $10), there is significant contribution from the price declines of all the other components, demonstrating the impact of the supply-chain. The estimate cost of the BoM is $55.40, netting to a cost of $0.11/W (Figure 8).

2.7 Updated IP Securement Strategy: Much of the technology generated for this approach has come out of the ICECOOL Program, and as such, much of it has been already published. There has been a fair bit of know-how developed to properly process these wafers, though the etching approach is based off of a modified Bosch etch process. We do believe that there is some potential for IP around the integrated fluidic housing that must seal to the chip. Initially, some designs could be patented for polymer packaging solutions, but ultimately, we seek to develop Si-based packaging solutions that would more directly integrate with standard fab processing.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

2.8 Updated Go-to-Market Strategy: In the case of an integrated microfluidic cooling solution, the barrier to adoption revolves around ***who would own the INTEGRATION*** of this solution. As previously described, whether we are talking about the 1st Product Embodiment (monolithic integration) or the 2nd Product Embodiment (bonded fluidic wafer), the fabrication of that wafer must be done at the chip fab level. This could either be done by a captive fab (Intel, Samsung) or by an outsourced fab (TSMC, Global Foundries, BRIDG, etc.). If we were to first focus on the 1st Product Embodiment as the rational MVP for initial market entry, then NewCo would license the technology to the fab, and would do the design and simulation work to deliver the necessary information about required micro-pillar layout to ensure the effectiveness of the thermal design. The fab would process the backside of the wafers, which would then be packaged to send out to the OSAT for dicing and packaging. As such, NewCo would receive on-going NRE from the fabs for these engagements for every new chip design.

Component	Single Board Sol'n	High-volume Cost Estimate
µ-Pinfin Wafer	1	$10
Molded Cavity	1	$2
O-Ring Seal/Gasket	1	$0.40
Board Clamp Ring	1	$1.50
Spring-loaded screws	4	$3.50
Push-2-Lock fittings	2	$13
Pump	1	$22
Tubing sections	3	$2
Other fittings/clamps	4?	$1

Figure A 8. Scaled BoM Estimate (100,000 units)

It is the OSAT that would have the responsibility for creating the package that would accommodate the now-modified CPU/GPU chip and encapsulating it in a housing of the appropriate design (Figure 6). NewCo would license OSATs (Amkor, ASE Group, JCET, etc.) to be able to design and build housings that accommodate the fluidically-enable chips. Here too, NewCo could receive on-going NRE for these fluidic design and modeling engagements for every new fluidic package design until these competencies were established internally by the OSATs.

The package would then be sold to OEMs building HPC boards and servers (e.g.: Supermicro, Dell, Google, Penguin, etc.) and integrated with the correct board frame, fluidic connectors, tubing, pump, etc. to enable a complete solution. NewCo would support this supply chain by generating the optimized designed for the fluidic wafers based upon CPU/GPU core layouts, while enabling the downstream players with licensing to use and integrate the necessary

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

components. It is anticipated a small team of 3 or 4 thermal designers would be sufficient to handle the stream of custom designs, and another 3 resources would be required to engage with customers to drive new solutions as well as business resources to drive and manage licensing deals.

3.0 Transition Planning (Go to Market Strategy)

3.1 Go/No-Go Decision: Given the nature of the supply-chain integration challenge, we must conclude that there is a 'No-Go' decision for a startup at this time. The basic technology works, and it appears the initial preferred embodiment is the monolithic approach ("1st Embodiment" described in Section 2.5). However, work must be done to develop and deliver a standardized package integration scheme that can seal to the backside of the silicon wafer and easily integrate with standard zero-leak, push-to-lock fluidic connectors to enable easy package encapsulation and rapid solution/system integration. It is anticipated that our pending collaborations with Microsoft and Cerebras will lay the foundation for these necessary elements and allow the supply-chain to learn how to use these elements in delivering an entire solution. As noted in Section 2.8 above, a NewCo would launch and be positioned to transfer the technology while teaching, through licensing and NRE engagements, the supply-chain how to effectively use the technology.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

Appendix A – Competitor List

Company	Model	HQ	Max Power	Technology
Advanced Cooling Technologies	ICE-Lok	Lancaster, PA		Pumped 2-phase cooling
Microcool		Decatur, AL		Cu cold plates, 2-phase & thermo siphons
Boyd Baknor		Mississauga, Ontario		Serpentine cold plates and vapor chambers
LiquidCool Solutions	SCS, SGS, LSS, LSS-RT	Rochester, MN		Single-phase dielectric immersion
Green Revolution Cooling				Liquid Dielectric Immersion
Huawei Liquid Cooling				Liquid Cold Plate
Fujitsu	Cool-Central			
Asetek	Rack CDU D2C	Denmark		Liquid Cold Plate
Motivair Corp	ChilledDoor Rack Cooling System	Amherst, NY	75 kW	Active Rear door heat exchanger
OptiCool Technologies	Cool Door System	Rochester, NY		Low-pressure, refrigerant-based cooled door
CoolIT Systems	Direct Contact Liquid Cooling (DCLC)	Calgary, Alberta		Passive and split-flow coldplates
CoolCentric	Rear Door Heat Exchanger (RDHx)	Raleigh, NC	33 kW	Liquid Heat Exchanger
CPU/GPU Coolers:				
Cooler Master	MasterLiquid ML120R	New Taipei City, Taiwa	1 kW	Coppercold-plate (Aluminum radiator!)
Corsair	Hydro Series™ H150i	Fremont, CA	1 kW	Copper cold-plate
EKWB	EK-FB ASRock X470 Gaming K4 RGB Monoblock - Nickel	Komenda, Slovenia	1 kW	Copper and Aluminum cold-plates (Aluminum-only solutions in separate product line called Fluid Gaming (https://www.ekfluidgaming.com)
Enermax	Liqtech TR4	Taoyuan, Taiwan	500+W	Copper cold-plate
EVGA	Hydro Copper Waterblock for GTX 1080 Ti FTW3 400-HC-5699-B1	Brea, CA	1 kW	Copper cold-plate
NZXT	Kraken X72	City of Industry, CA	1 kW	Copper cold-plate
Thermaltake	Pacific V-GTX 1080Ti	Taipei City, Taiwan	500+W	Copper cold-plate
Zalman	CNPS9900 MAX	Gyeonggi-do, Korea	300W	Copper cold-plate
In the lab:				
IMEC	Impingement Cooling	Belgium		Impingement cooling
École Polytechnique Fédérale de Lausanne		Switzerland		2-phase microchannel cooling using refrigerant HFO1234ze
UIUC-Sandia		Urbana-Champaign		Enhanced jumping droplet condensation
Purdue-Mudawar/Sung		Lafayette, IN		Microjets & microchannels
JetCool		Littleton, MA		Impingement Cooling
Partners/Channel				
Wave Computing		Campbell, CA		Chris Nicol, CTO
PSSC Labs		Lake Forest, CA		Paul Green, Gov't Sales
Quanta Cloud Technology		San Jose, CA & Taiwan		
ScaleMP		Ft. Lee, NJ		
Iron Systems		Fremont, CA		Robert Keith, CTO
Penguin Computing		Fremont, CA		Philip Pokorny, CTO

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

Appendix B – Customer Discovery Interviews

Legend:

M = Face-to-face meeting

P = Phone interview

ECTC = IEEE's 69th Electronic Components & Technology Conference, May 28-31, Las Vegas, NV

SC2019 = Supercomputing 2019 (SC19), November 18-21, Denver, CO

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

REFERENCES

[1] www.chilleddoor.com
[2] www.opticooltechnologies.com
[3] www.asetek.com
[4] www.coolitsystems.com
[5] www.coolermaster.com
[6] www.grcooling.com
[7] Phuc Nguyen, IT Technical Support, 2CRSI, Supercomputing 2019, Denver, CO
[8] www.jetcool.com
[9] http://www.eenewseurope.com/news/package-level-3d-printed-chip-cooler-beats-all-air-alternatives
[10] https://www.sciencedaily.com/releases/2008/07/080701180141.htm
[11] Brij Singh, Sr. Staff Engineer - Advanced Technology, John Deere, Personal communications, December 21, 2018
[12] Jason Milne, Raytheon, Personal communications, September 28, 2018
[13] https://www.staubli.com/en-us/connectors/
[14] https://www.cejn.com/en-us/
[15] Kevin Anderson, Director, Integration Technology Research, et. al., Qorvo, Personal communications, April 1, 2019
[16] Ty Garibay, VP of Engineering, Mythic Semiconductor, Personal communications, April 22, 2019
[17] Xiaojin Wei, Hardware Engineer, Google, Personal communications, February 7, 2019
[18] Daniel Berger & Gary Patton, Global Foundries, Personal communications, April 29, 2019
[19] Daniel Berger, Senior Director, Global Foundries, Personal communications, May 22, 2019
[20] Nicolas Monnier, Product Manager, Staubli Corporation, Personal communications, June 12, 2019,
[21] Girish Upadhya, Senior Advisory Engineer, Northrop Grumman, Personal communications, April 22, 2019
[22] Mark Gerber, Sr. Director, Engineering Marketing & Technical Promotion, ASE Group, Personal communication at ECTC, May 29, 2019
[23] Curtis Zwenger, Vice President, Research and Development, Amkor, Personal communications, June 12, 2019
[24] Jeff Milrod, President, BittWare, Personal communications, November 19, 2019
[25] ibid
[26] Jim Hearndon, Enterprise Technologist, DellEMC, Personal communications, November 19, 2019
[27] Jason Zeller, Product Marketing Manager, CoolIT, Personal communications, May 30, 2019
[28] Christiaan Best, CEO, CTO & Founder, Green Revolution Cooling, Personal Communications, August 21, 2018.

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

AFRL	Air Force Research Laboratory
BTS	Board Test System
CFD/HT	Computational Fluid Dynamics And Heat Transfer
DARPA	Defense Advanced Research Projects Agency
DSP	Digital Signal Processing
EMIB	Embedded Multi-die Interconnect Bridge
ES	Engineering Silicon
FFT	Fast Fourier Transform
FIFO	First In First Out
FPGA	Field-Programmable Gate Array
IP	Intellectual Property
PAO	Polyalphaolefin
PCS	Physical Coding Sublayer
RXME	Electronics and Sensors Branch, Manufacturing and Industrial Technologies Division, Materials and Manufacturing Directorate of AFRL
TIMM	Thermal Interface Material
VRM	Voltage Regulator Modules
WPAFB	Wright Patterson Air Force Base

DISTRIBUTION STATEMENT A. Approved for pubic release.
Distribution is unlimited.