Chapter 4

Development of an On-Chip Cooling Device for High-Heat-Flux Microprocessors

4.1 Introduction to Microfluidic Liquid-Cooling of Microprocessors

To further demonstrate the usefulness of a single-mask process capable of producing three-dimensionally independent microchannels, a device was fabricated to cool down hot microprocessors by liquid cooling on the backside of the chip. Microprocessor cooling has become an increasingly studied topic as the heat density of microprocessors will soon rise above the capabilities of current air cooled heat sink designs used for thermal management. The International Technology Roadmap of Semiconductors projected the heat density of single chips to reach 108W/cm^2 by 2018 [35]. Also, chip design has begun to be affected more by thermal constraints over physical component size constraints [36].

Heat is a problem for microprocessors because as temperature increases, the performance of the chip degrades. At chip temperatures greater than 85°C, the chip begins to fail due to electromigration [37]. High temperatures also result in higher power consumption for the chip due to increased subthreshold leakage current. Subthreshold leakage current is current that flows between the source and the drain of a transistor when the gate to source voltage is below the

threshold voltage. As transistors become smaller, the amount of subthreshold leakage current increases and high temperatures add to this problem [37].

In order to tackle this evolving problem, new methods are required that are compact and have a high heat removal capacity. Researchers have investigated liquid jet impingement cooling [38-41], thermoelectric methods [37], and low flow rate individual fluid droplet manipulation [42] for this purpose, but the most popular method to solve this problem is microfluidic cooling [43-46]. This method involves a cool fluid, usually water, flowing through microchannels on or touching the microprocessor backside. The smaller and closer the channels are to the active surface of the microprocessor, the less thermal resistance there is between the active surface and the cooling fluid, increasing the cooling capacity of the chip. Although several researchers have implemented microchannels directly on to the processor for this purpose, Colgan et al. have pointed out that the risks associated with on-chip channel fabrication are high due to the costliness of microprocessors [44]. With the accurate model developed for three-dimensionally independent microchannel fabrication using a single-mask, the risks during microchannel fabrication are reduced. These channels have the benefit that they can be etched in a single step and then sealed using buried channel technology, which only requires one additional step [47]. In addition, new cooling channel designs can be created with the gained three-dimensional design capability. An initial investigation of the performance of a microprocessor cooling device fabricated using the aforementioned method demonstrates the capability to cool the next generation of microprocessors.

4.2 Review of Chip-Cooling Literature

4.2.1 Microscale Heat Transfer

The idea of microchannel cooling was first presented by Tuckerman and Pease in 1981 in a study geared toward very-large-scale integrated (VLSI) circuit cooling. Their investigation

used an array of microchannels on a chip etched anisotropically to 300μm depth that were 50μm wide with spacing equivalent to the width of the channels. The substrate was 400μm thick, only 100μm deeper than the channel depth, so the thermal resistance of the chip was small, 1°C/W at the downstream portion of the channel. On the other side of the chip there was a 1cm^2 resistive heater. The convection coefficient, h, was found to be the most important quantity affecting sufficient heat transfer and it is related to the rate of heat removal, q, by

$$q = hA\left(T_s - T_{bulk}\right) \tag{9}$$

where A is the surface area of the fluid in contact with the heated surface, T_s is the temperature of the heated surface and T_{bulk} is the average temperature of the fluid at a given point. h is not constant over the length of a channel due to temperature variation along the channel length. In order to make this quantity large, Tuckerman and Pease determined that the aspect ratio of the channel, equal to the depth/width of the channel, should be as large as possible. This influences the heat transfer rate by increasing the amount of surface area in contact with the chip in relation to the hydraulic diameter, D_h, of the channel defined by

$$D_h = \frac{4A}{P} \tag{10}$$

where P is the perimeter of the channel. High aspect ratios increase h and also decrease the pressure drop over the channels because more channels can fit on the surface of a given chip, yielding a higher flow rate, Q, for a given differential pressure. Tuckerman and Pease experimentally demonstrated the ability to cool 790W/cm^2 from a chip with resulting inlet and outlet fluid temperatures of 23°C and 94°C, respectively.

Since Tuckerman and Pease's investigation into microchannel fluid flow for cooling applications, several studies have been performed to better understand the nature of heat transfer

at the microscale. One of the more basic differences observed is a different transition Reynolds number from laminar to turbulent flow, where the Reynolds number, Re, is defined by

$$\mathrm{Re} \equiv \frac{\rho u D_h}{\mu} \tag{11}$$

where ρ is the density of the fluid, u is the mean velocity of the fluid and μ is the viscosity of the fluid. The Reynolds number is a dimensionless quantity relating the inertial forces to the viscous forces in the fluid. For microflows, Re is generally very low (less than 2000). The Reynolds number for fully developed flow at the macroscale is around 2300, but Peng and Peterson observed it to be between 200 and 700 depending on the size of the channel [48]. As the hydraulic diameter decreases, the transition Reynolds number has been observed to decrease as well.

Fluid flow through microchannels has also been cited to differ from macroscale flow based on the amount of pressure required to produce a given flow rate in a channel. The Darcy friction factor, f, is defined by

$$f \equiv \frac{-(\Delta P/\Delta x) D_h}{\rho u^2/2} \tag{12}$$

where $\Delta P/\Delta x$ is the pressure drop per unit length. In fully developed macroscale laminar flow through a circular tube, the product of f and Re is a constant equal to 64 [49]. For other channel geometries on the macroscale, f*Re changes, but it is still a constant value. At the microscale, there have been several studies that have both proved and refuted that this relationship is still true in microchannel fluid flow.

Hegab et al. [50], Jaing [51], and Xu et al. [52] all observed experimentally that the conventional macroscale relationship for f still holds true in microscale flows. In addition, Toh et al. carried out a finite volume analysis and showed that as long as temperature is constant over the

channel, f*Re should match the fully developed models [53]. However, Peng and Peterson [48] and Qu [54] observed higher friction factors than could be explained by conventional theory. These differences were observed for all Reynolds numbers by Peng and Peterson, and only for Reynolds numbers greater than 500 by Qu. Steinke and Kandlikar tried to come up with an explanation for the difference in observed relationships by examining how each study collected their data and deduced a trend [55]. The major conclusion given by Steinke was that the studies that report a discrepancy from conventional friction factor relationships usually have not taken into account entrance and exit effects as well as developing lengths. This is true of both Peng and Peterson's study as well as Qu's study. The developing length, the length of channel flow required for the flow profile to become fully developed, $x_{fd,h}$, can be calculated by

$$x_{fd,h} = 0.05 \,\mathrm{Re}\, D_h \tag{13}$$

In microchannel flows used for cooling applications, the hydrodynamic entry length can be almost the entire length of the channel. Steinke and Kandlikar carried out an experiment of their own and found that theory predicted the friction factor well, with the exception of Re > 300 and for very low Reynolds numbers.

Different effects have also been observed for heat transfer at the microscale as compared to the macroscale. In order to quantify the degree of heat transfer, a non-dimensional value called the Nusselt number is used. The Nusselt number, Nu, is defined by

$$Nu \equiv \frac{hD_h}{k} \tag{14}$$

where k is the thermal conductivity of the fluid. For laminar flow in large channels, this number is expected to be constant when the flow is fully developed, unaffected by Reynolds number,

Prandtl number, or axial location [49]. The Prandtl number is a non-dimensional fluid property relating the momentum and thermal diffusivities, defined as

$$\mathrm{Pr} \equiv \frac{C_p \mu}{k} \tag{15}$$

where C_p is the specific heat of the fluid. Studies by Wu [56] and Peng and Peterson [48] have shown that at Reynolds numbers less than 100 there is a significant relationship between Reynolds number and Nusselt number. Wu found that the Nusselt number increases linearly with increasing Reynolds number below this value. This trend is evident from Figure 28. Above 100, Nu begins to level off. Peng and Peterson observed a logarithmic relationship between Re and Nu as well as significant influence from the channel aspect ratio, the ratio of the hydraulic diameter to the channel pitch, and the Prandtl number. Qu et al. also performed an investigation in the laminar regime, with Re > 100, but did not note any increase of Nu with Re. However, in examining their data, Nu is almost constant with Re for D_h < 100, but for D_h > 100 there is a slight increase in Nu for 100 < Re < 500. It is also noteworthy that the predicted Nu from conventional theory was higher than the experimental values of Nu, but the values were corrected accurately after using a roughness-viscosity model to account for extra viscosity effects. For heat transfer in turbulent flows, special relationships have been developed and studies vary on the accuracy of the models. A particular correlation, the Gnielinski correlation, was used for prediction in two different studies. Adams et al. [57] found that the correlation under-predicted values of Nu, while Hegab et al. [50] found that the opposite was true.

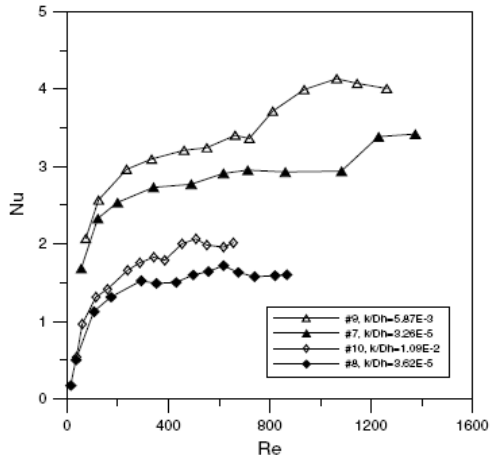

Figure 28: Nu vs. Re at low Reynolds numbers. A linear increase in Nu for Re < 100 is evident. Also, increased surface roughness (shown as hollow points) have increased heat transfer over smoother channels (solid points). Reproduced with permission from [56]. Copyright 2003, *Elsevier Science Ltd.*

There are still many aspects concerning heat transfer at the microscale that are not fully understood. It is believed that many of the apparently conflicting observations by researchers can be explained by taking into account effects that are normally neglected on the macroscale. One of these key areas is channel surface roughness. Wu et al. observed that in nearly identical channels with the only difference being surface roughness, both heat transfer and the pressure drop for a given Reynolds number were higher in the channel with higher roughness, citing a greater influence from roughness on the boundary sublayer of higher Reynolds number flows [56]. This observation is illustrated in Figure 28. Qu et al. noticed that surface roughness decreases the transitional Reynolds number for microchannel flow in their study of trapezoidal channels [54]. Also, Bergles et al. noted in their review paper that at small scales roughness must be accounted for in microchannel calculations in order for accurate predictions to result [58]. Due to the small scale of the channels, it is very hard to take into account all of the necessary conditions to produce consistent results and flow predictions, hence many different reports are produced on the nature of heat transfer at the microscale.

The major way that designers can influence the heat transfer properties of microfluidic channels is by the channel geometry. Several studies have been performed in order to optimize

microchannel geometry for a specific purpose. Renksizbulut and Niazmand performed a numerical analysis on trapezoidal channel geometry in order to determine the effects of trapezoidal dimensions on pressure drop and heat transfer [59]. They found that Nu increases when the ratio of the channel depth to the width of the bottom of the channel decreases. When the ratio is high, on the other hand, and the channel is shallow, creating a small angle between the surface of the substrate and the channel wall, the pressure gradient becomes very high in the developing region, but drops to a low value once the flow is fully developed. They actually observed a pressure overshoot in the corner of the channel in the developing region due to the high pressure gradient there. Zhimin and Fah investigated rectangular channels in three different cases: constant flow rate, constant pressure drop, and constant pumping power, to determine the effect of geometry on chip resistance [60]. They found that increasing the aspect ratio only decreases the chip resistance for a constant pressure differential or constant pumping power while little effect is seen in the constant flow rate case. However, increasing only the width increased the thermal resistance of the chip. Upadhye and Kandlikar verified that width decreases heat transfer capability in [61] using their models and they also verified that aspect ratio increases heat transfer and allows for more channels to be fabricated in a given area, decreasing the flow resistance. Peng and Peterson verified that increasing the aspect ratio increases the heat transfer for laminar flows, but they found that in turbulent flows, the aspect ratio of channel depth to width or width to depth should be 0.5 in order to maximize heat transfer capability [48]. For both turbulent and laminar flows, heat transfer was found by Peng and Peterson to increase with large ratios of D_h to the channel pitch. Only Bau addressed changing geometry over the axial length of the channel in order to improve heat transfer [62]. He found that there are advantages to decreasing the aspect ratio along the channel length and presented two different optimization schemes, one to minimize the maximum temperature over the chip and one to minimize temperature variation over the chip to less than 1°C.

Other than geometry, material properties are the other major aspect of design that can influence heat transfer performance. Wu et al. examined a hydrophilic substance, in their case silicon dioxide, in contact with the fluid and its effects on heat transfer [56]. They found that a hydrophilic coating increases the pressure drop through the channel, but it also increases the heat transfer rate. Some designers have chosen to modify the fluid by adding a coolant such as propylene glycol [63], but since the viscosity of water is significantly less, water is often used as a coolant. Other studies have also investigated two-phase flow for cooling [58, 64-66], but the complexity of these systems is much higher. The formation of a vapor phase increases the overall pressure required to drive the flow and the high pressures where the vapor forms can even push the flow upstream of the bubbles in the reverse direction [67]. The pressure fluctuation induced by vapor formation can also cause flow instability due to rapid bubble formation and collapse when that fluctuation is coupled to the flow rate [66]. Flow instability can cause problems when microscale heat transfer technology is to be implemented in a practical device, such as a microprocessor cooling chip.

4.2.2 Microprocessor Cooling Designs

Since Tuckerman and Pease's initial microchip cooling device, several groups have attempted their own improved designs in the hopes of realizing a commercial device. Pijnenburg et al. have demonstrated a device using conventional silicon processing technology to produce microchannels on the backside of a silicon chip [45]. They used the Bosch process to anisotropically etch silicon microchannels with 100μm width, 300μm depth, and 100μm spacing and then coated them with a layer of thermal SiO_2. These channels are wider than the channels used by Tuckerman and Pease. Fluidic connections were created through a Pyrex glass cover, which was glued to the silicon substrate, allowing water to flow into a reservoir, then through parallel channels on the chip. The chip was heated by a copper resistor plated on the opposite side of the silicon. Through Flowtherm CFD models and by experimental analysis they found

that the thermal resistance of the chip, R_{th}, decreases with increasing flow rate and increasing dissipated power. R_{th} is described by

$$R_{th} = \frac{T_h - T_{in}}{q} \tag{16}$$

where T_h is the average temperature of the heater, T_{in} is the bulk temperature of the fluid at the inlet, and q is the dissipated power from the device. However, the experimental results showed a lower thermal resistance than the Flowtherm model predicted for variable flow rates. Pijnenburg et al. demonstrated cooling of 370W/cm^2 with a maximum junction temperature of 120°C, 0.1L/min flow rate, and 15kPa pressure drop.

Due to the complicated fluid entry system, and the fact that the wafer was sealed by a glass wafer and glue, the chip cooling demonstration by Pijnenburg et al. would be hard to implement into a commercial device. Dang et al. tackled this problem and created an on-chip cooling device with an integrated fluid entry system. The system uses anisotropically etched channels in silicon with a single inlet and outlet port for every three channels. The channels were etched 70μm deep and were 100μm in width with approximately the same spacing as channel width, producing one of the few low aspect ratio designs used in chip cooling studies. Channels were then sealed with a polymer overcoat after sacrificial layer deposition. Then, fluid ports were etched from the front-side of the wafer (the side with the electronic components) until they merged with the cooling channels fabricated on the wafer backside. These fluid ports then fit into entry ports on a circuit board when the chip is attached for packaging, as demonstrated in Figure 29. Not only did Dang et al. present a commercially viable on-chip cooling package, but they also demonstrated the ability to cool next generation microprocessors. They were able to remove 198W of heat using less than 100kPa resulting in a fluid temperature rise of less than 10°C.

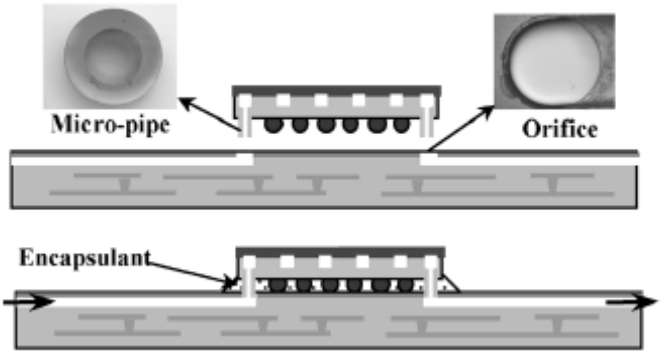

Figure 29: Flip-chip attachment of a microfluidic chip by Dang et al. Reproduced with permission from [35]. Copyright 2006, *IEEE*.

Other researchers have developed completely different cooling channel fabrication methods. Joo et al. demonstrated an IC compatible process whereby blocks of photoresist are patterned on a silicon substrate and then are covered with electroplated nickel [68]. Before the metal completely covers the photoresist block, the resist is removed and electroplating is used again to fully seal the channels. These channels are much smaller than the etched channels used by other groups, 10μm by 10μm channels were fabricated, but analysis has shown the ability of radially designed channels to cool 50W/cm^2 with air as the cooling fluid. Mo et al. also demonstrated a chip cooling method involving deposition of carbon nanotubes on a silicon layer and sealing with Pyrex glass [69, 70]. Nanotubes were used because of their high conductivity, 6000W/K/m, almost 50 times greater than that of silicon. However, they have not optimized their design and only a 6°C decrease in temperature was achieved as compared to an open channel with slightly less heat input and a higher flow rate of water.

As microprocessors evolve, the transistor density increases, but in the future this increase in density may not be constrained in only two dimensions. Future processors may incorporate 3-D circuit architectures. Koo et al. have addressed the cooling of these processor layers in a numerical study [64]. Their proposed design incorporates microfluidic channels running through the substrate in the dielectric layers between each circuit layer. Their control volume approach to solving the problem indicated that it is possible to cool 68W/cm^2 per layer while keeping the

maximum temperature of the chip below 85°C for a flow rate of 15mL/min and two-phase flow at sub-atmospheric pressure. They also found that placing the highest temperature regions, the logic areas, at the end of the channel helps keep pressure drop requirements low. Wei and Joshi used a similar idea, but for conventional 2-D processor architectures, stacking chips with microfluidic channels on them to form a liquid cooled heat sink [71]. Through modeling, Wei and Joshi determined that as more layers are added, the pressure drop for a given flow rate decreases because a larger flow area is available for the fluid to pass through. In terms of heat transfer, they discovered that for a fixed pressure drop or for fixed flow power, chip resistance decreases as more layers are added. However, for a constant flow rate, two layers produce a lower thermal resistance than one layer, but more than two layers in the heat sink results in a higher thermal resistance than that of a single chip layer.

Most of the aforementioned papers about microprocessor cooling used anisotropic fluidic channels that traveled in a straight path over the substrate. Kandlikar and Upadhye postulated in their paper that for practical removal of high rates of heat flux, different types of channels would have to be implemented [72]. The reason for this change is due to the high pressure requirement of a large number of small channels carrying a high flow rate of fluid. They suggested replacing many small channels with fewer larger channels containing staggered pins or other geometry to enhance heat transfer without creating large pressure losses. They also studied a split flow design where fluid enters through either end of a large channel and then flows transversely through channels that are half the length of the chip. For the staggered pin array, they found that pressure drop and mass flow rate decrease significantly for a given heat rate, and the split flow arrangement lowers the required pressure drop for a given heat load. These enhanced microchannels were able to cool upwards of 300W/cm^2 using around 35kPa in a simulation. Colgan et al. fabricated a liquid cooled heat sink with six adjacent 3mm long channels across the chip with staggered fins [44]. They showed the capability to cool a chip producing over

300W/cm^2 heat output using less than 35kPa. A simulation showed a maximum theoretical heat removal density of 460W/cm^2. A render of a section of the flow device is shown in Figure 30.

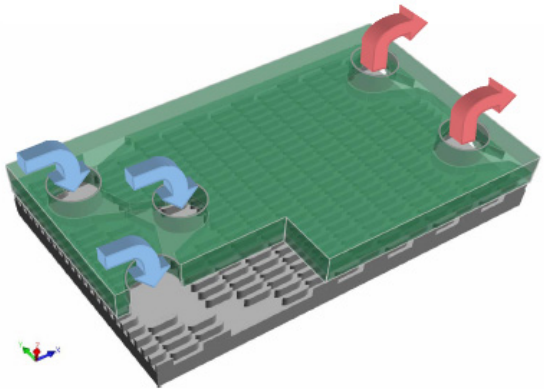

Figure 30: Microfluidic cooling device designed by Colgan et al. Reproduced with permission from [44]. Copyright 2005, *IEEE*.

The largest commercial product currently to come from this technology for desktop systems is produced by Cooligy, Inc. The company spawned from research performed at Stanford University, and the company has since been acquired by Emerson Network Power. Their system uses microchannels on the order of 30-100μm and aspect ratios of 5-15 to cool hot microchips [73]. They have shown the ability to cool an average of 250W/cm^2 and a peak heat flux of 360W/cm^2 while keeping the junction temperature at 81°C. 280mL/min coolant flow rate and less than 35kPa are required for this task.

4.3 Microprocessor Cooling Design and Setup

The purpose of this portion of the study was to prove that the newly proposed fabrication technique is suitable for creating microfluidic channels capable of cooling the next generation of microprocessors. The major drawback of this fabrication method for this application is the limitation in channel aspect ratios that are achievable. As is evident from Figure 23, 1:1 aspect ratio channels are only possible for very small channels. However, due to the desire to have as much surface area as possible in contact with the cooling fluid, and the fact that high pressure

flows are undesirable, it is not reasonable to have very small 1:1 aspect ratio channels for a microcooler design. Thus, larger channels have to be used.

The layout of the cooling chip examined here was designed similar to the chip used in Pijnenburg et al.'s study [45], with two large microfluidic reservoirs connected by several microchannels passing over the back of a heated chip. The first modification that had to be made to the design was the addition of pillars into the reservoir structure so that the reservoir would be able to withstand high fluid pressures without breaking after buried channel sealing. As a result, the reservoirs took the shape of a crisscross array of microchannels. The channels in the reservoir had to be large enough that a significant pressure loss did not occur over the reservoir, allowing equivalent pressure drop and, thus, equal flow rate through each channel. To achieve these goals, the channels had to be deep, and wide, but not wide enough to where mechanical stress would damage the sealing layer. The potential durability of the sealing layer on top of the channels was evaluated by the amount of wrinkles created by the inherent stress in the oxide/photoresist layer on the wafers used in the modeling investigation. The design with s = 2μm, l = 2μm, w = 10μm, c = 16, and t = 20 minutes achieves a depth of 137μm and a width of 327μm after etching with a small degree of wrinkles in the stressed oxide/photoresist layer. This pattern was chosen for the channels composing the reservoir.

For the actual cooling channels, the aspect ratio was desired to be as close as possible to 1 while having a rather large depth. The etch duration for the desired channel geometry also had to be 20 minutes because this was the same etch duration that had already been chosen for the reservoir channels. For the cooling channels, the pattern s = 3μm, l = 2μm, w = 10μm, c = 15, and t = 20 was chosen to produce a channel depth of 124μm and a width of 316μm. These channels were designed on the chip surface in a meander pattern with two turns so that a more even temperature distribution over the wafer would result. The spacing between each channel was designed to be 30μm. Figure 31 is an overhead view of the channels.

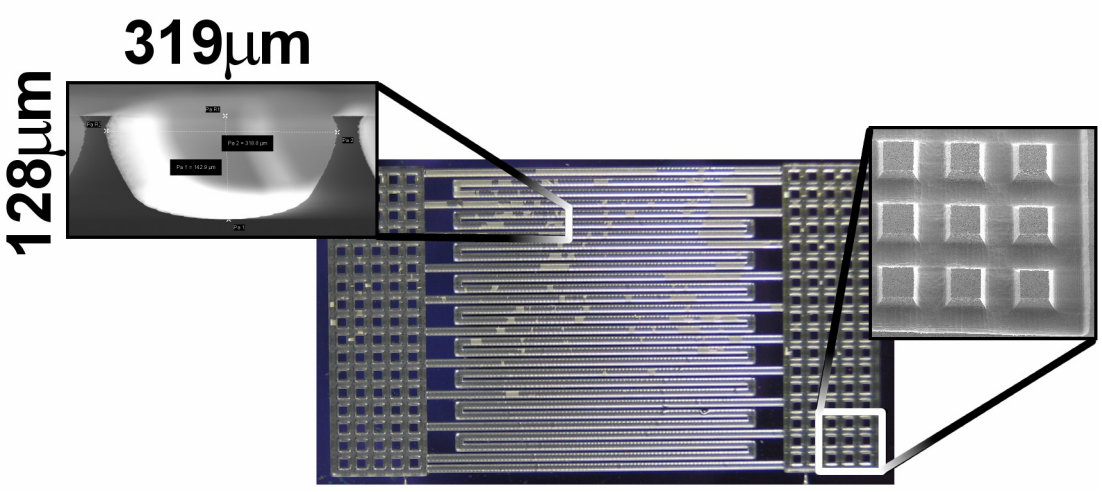

Figure 31: Overhead view of the meander channel cooling chip. The reservoirs have pillars to distribute stress on the sealing oxide from the fluid. Entry ports are etched from the opposite side into the middle of the reservoirs.

A diagram of the process flow for the cooling chips is shown in Figure 32. The mask was transferred onto the wafer and the channels were etched using the same method as the wafers that were used in the Alcatel modeling (Figure 32c), with the etching parameters of Table 5. The etching step had to be carried out for 40 minutes as opposed to the 20 minutes that the model called for in order to achieve the correct dimensions. The reason for this is due to etch reduction by microloading. The local exposed wafer density of the mask for the cooling channels is 18.6%, more than twice the local exposed substrate density of the wafer used for modeling (8.7%). However, at this longer etch duration, both the reservoir and the cooling channels were accurate to less than 5% of the predictions. This finding suggests that the etch duration given by the model will have to be modified by a scale factor dependent on the exposed wafer density. With the time adjustment, the global geometry will result in the initial desired channel dimensions.

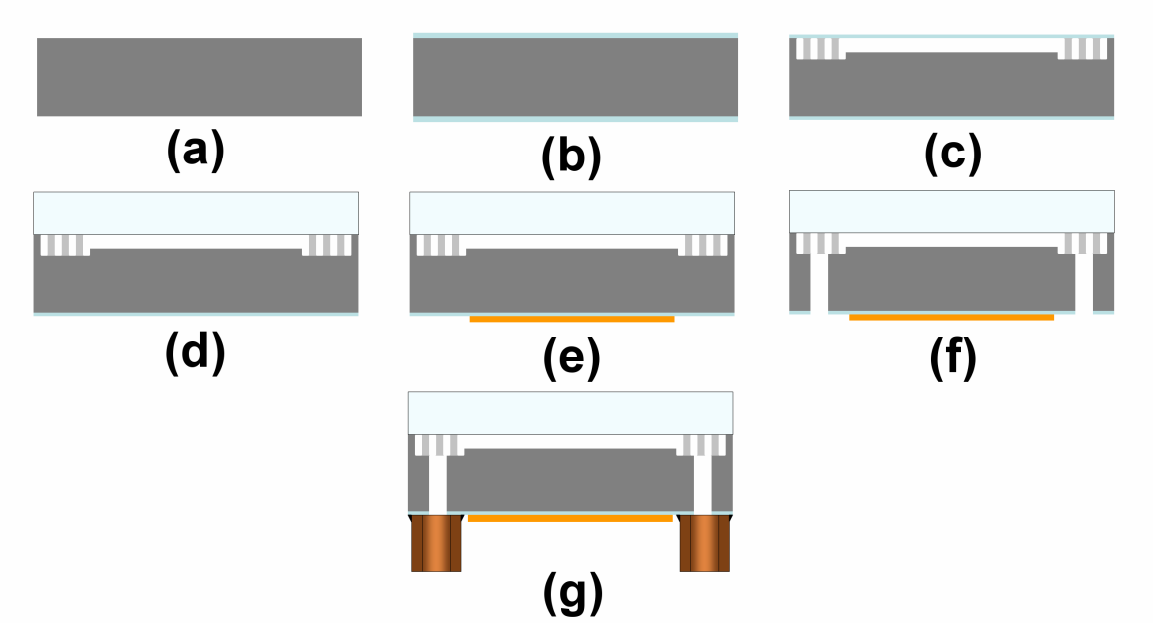

Figure 32: Process flow of the cooling chip. (a) A bare silicon wafer with two polished surfaces. (b) Thermal oxidation. (c) Etching of microfluidic channels. (d) Anodic bonding to a Pyrex wafer. (e) Lift-off patterning of a heater/RTD layer of Ti/Pt. (f) Anisotropic etching of entry ports. (g) Gluing of PEEK ports for connection to fluid source.

After etching, the silicon substrate was anodically bonded to a 600μm thick Pyrex wafer to seal the channels (Figure 32d). While the channels were designed for buried channel sealing with silicon oxynitride, this initial investigation used Pyrex glass to ensure adequate sealing under a wide range of pressures. After sealing, the backside of the wafer was covered with a layer of S1813 photoresist spun at 2000rpm for 30 seconds. The photoresist was then patterned to expose areas for metal to be deposited on the wafer, forming resistive heating elements used to simulate the heat output of a microprocessor (Figure 32e). The design for the metal layer is shown in Figure 33. The mask is designed to produce four heaters over the wafer and eight resistance temperature detectors (RTDs). An RTD changes its resistance with temperature, so the temperature of the RTD can be determined by measuring its resistance. The heaters were arranged to simulate the heat distribution on a modern single core processor, such as the Pentium IV; however, for this study the heat density was kept constant over the entire wafer as a proof-of-

concept. The metals that were deposited to form the heaters and RTDs were 400Å Ti followed by 2600Å Pt. Platinum was used because its resistance changes linearly with temperature through a wide range of temperatures, making it the perfect material for an RTD. The lift-off technique was used to pattern the evaporated metal which covered a total surface area of $1cm^2$.

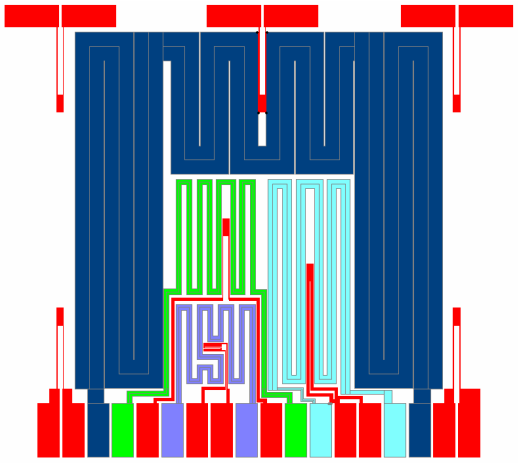

Figure 33: Mask design for the metal layer. Pads for wire bonding are located on the edges of the chip connected to RTDs (shown in red) and heaters (four in total).

After deposition and lift-off of the Ti/Pt metal layer, the fluid entry ports were etched into the substrate (Figure 32f). Circular entry ports 400μm in diameter were anisotropically etched using the Bosch process above the center of the fluid reservoirs from the front-side (metal-side) of the wafer until the entry ports and the cooling reservoirs were completely joined. Polyetherether ketone (PEEK) tubes with a 4.3mm inner diameter were then glued onto the chip to allow fluidic connection to the channels (Figure 32g). The entire chip was then glued and wire-bonded to a custom-made printed circuit board (PCB) to provide electrical connections.

The completed chip was then placed into a fluidic network powered by a container of 20°C deionized (DI) water pressurized by a cylinder of compressed nitrogen gas. In the open-loop system shown schematically and in a photograph in Figure 34 and 35, respectively, the fluid passes from the pressurized container, through the chip, and then to a digital scale where the display is monitored by a digital camera taking pictures every 1/15 seconds. A differential

pressure transducer measures the pressure difference on either side of the chip and a thermocouple placed in the outlet stream measures the outlet temperature of the water. The resistance heaters on the chip are connected to independent voltage sources to control the heat input and the RTDs are attached to an ohmmeter to track the resistance changes.

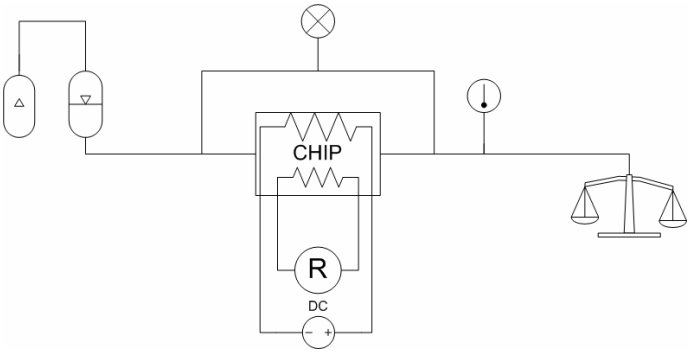

Figure 34: Schematic of the fluid and electrical system.

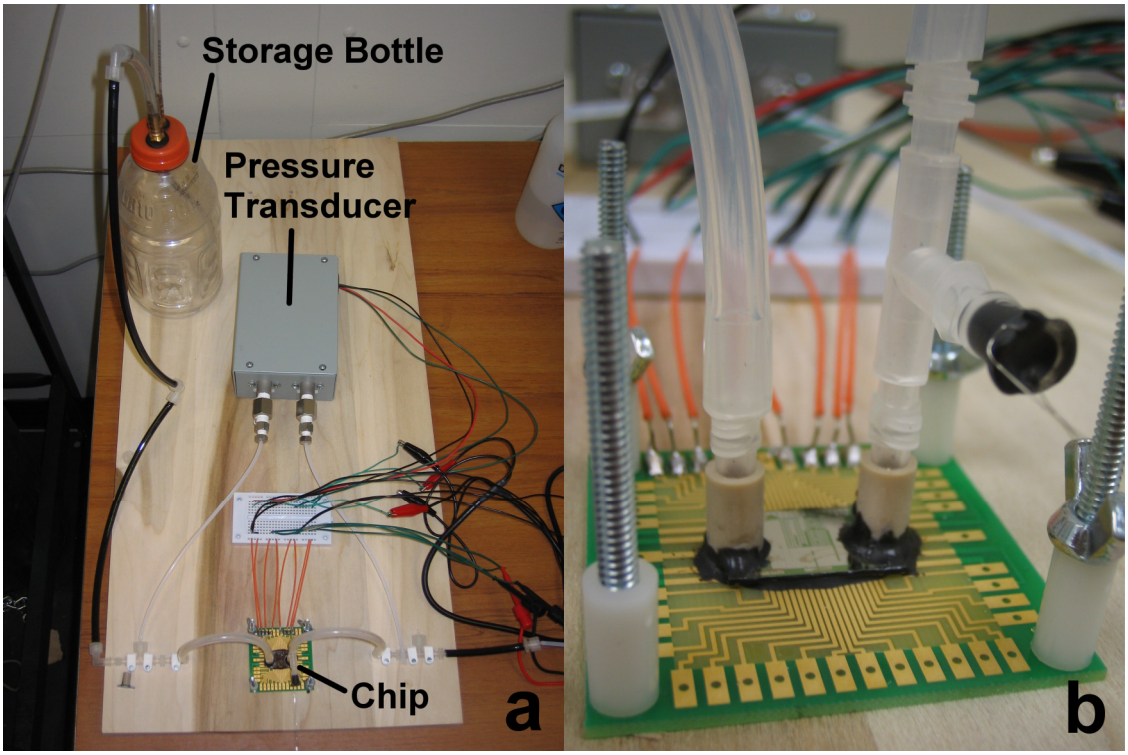

Figure 35: (a) Photograph of the cooling system (b) Photograph of the cooling chip. The thermocouple for the outlet stream is connected through the black plug.

After the fluidic network was set up, the fluid supply container was pressurized, forcing water to flow through the system. The friction factor, equation 12, and the Reynolds number, equation 11, were calculated for different values of differential pressure over the chip in order to examine the relationship between f and Re for the device. Next, different values of heat density were applied to the chip to examine the effectiveness of the chip for cooling, as measured by the thermal resistance, R_{th}, of the chip calculated by equation 16. The amount of heat into the system, q, is calculated by

$$q = VI = I^2 R = V^2 / R \qquad (17)$$

where V is the voltage applied to a resistor having a resistance, R, thereby inducing a current, I, through the resistor. The goal of this experiment is to remove 100W/cm^2 of heat flux from the system while keeping the maximum temperature of the chip below 85°C, the critical temperature before failure occurs in microprocessors. Several different power densities were applied to the chip up to 100W/cm^2 to determine the cooling behavior of the chip.

In order to compare the results to theory, a computational fluid dynamics (CFD) model of the entire chip was created and the same tests that were run on the chip were simulated. The CFD model was different from the actual chip in that it did not include the pillars in the reservoir due to meshing constraints and no natural convection boundaries were included; the chip was completely insulated. Also, during the simulation, the physical properties of the fluid were kept at a constant value calculated at the average of the inlet and outlet fluid temperatures and the cooling channels were approximated as rectangles with the same depth and width dimensions as the actual cooling channels.

4.4 Microprocessor Cooling Results and Analysis

Before heat was applied to the system, the characteristics of the flow were determined by measuring the value of f and Re for different applied system pressures. In fully developed laminar flow, the value of f*Re will be a constant value, dependent only on the geometry of the channel. In the setup used in this experiment, this trend was not observed. Figure 36 contains a plot of f*Re vs. Re for the meander channels. The key measurement in determining the value of f, $\Delta P/\Delta x$, was found by subtracting the pressure loss from the location of the differential pressure transducer port to the end of the chip reservoir from the total measured differential pressure over the chip and dividing this value by the total length of the channel, 32mm. The pressure loss from the differential pressure transducer port to the end of the chip reservoir was determined by flowing fluid through only this portion of the chip into atmospheric pressure. Although the observed results do not seem to follow traditional macroscale theory, the reason for this is likely due to the fact that entrance and exit losses, as well as the development length were not taken into account during the calculation for f. The hydrodynamic entry length, $x_{fd,h}$, ranges from less than 1% of the total channel length, at a Reynolds number of 30, to 7.5% at a Reynolds number of 270. Table 9 contains this information in detail. Besides at the very lowest Re measurement, the increasing value of f could be attributed to the elongation of the entry length, which has a higher differential pressure per unit length than the fully developed region. The deviation from this trend at the lowest Re measurement was also observed by Steinke and Kandlikar in their experiment carefully designed to measure the relationship of f vs. Re in [55], suggesting influence from a factor other than the development length.

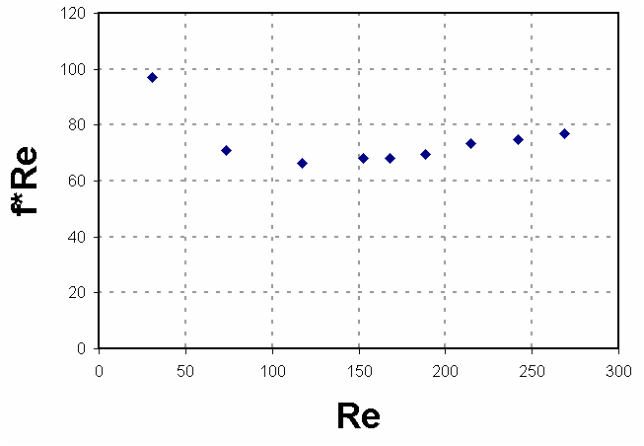

Figure 36: f*Re vs. Re for the cooling chip. The change in the value as Re increases is thought to be due to entrance, exit, and development effects in addition to others.

Table 9: f, f*Re, and $x_{fd,h}$ vs. Re.

Re	f	f*Re	$x_{fd,h}$, μm	$x_{fd,h}$, %
30.7	3.16	97.0	273	0.85
73.6	0.96	71.0	655	2.05
117.5	0.57	66.5	1046	3.27
152.5	0.45	68.2	1357	4.24
167.9	0.41	68.2	1494	4.67
188.6	0.37	69.6	1678	5.24
214.7	0.34	73.3	1910	5.97
241.8	0.31	74.8	2152	6.72
268.6	0.29	77.0	2390	7.47

After the friction factor versus Reynolds number relationship had been found with zero heat flux into the chip, the chip was heated to determine the cooling capabilities of the microfluidic channel design. Based on simulation, the chip was run at a constant 750μL/s flow rate, the lowest practical flow rate while keeping the chip safely below 85°C at 100W/cm^2 input power (78°C). The power to the chip ranged linearly from 10W/cm^2 to 100W/cm^2 in 10W/cm^2 increments. The maximum temperature of the chip was measured using a thermocouple attached to the location on the chip of the maximum temperature, which was near the channel outlet. Figure 37 shows the maximum chip temperature, T_{max}, vs. the power density input to the chip for the simulation and the experiment. The trend for both cases is linear, with the experimental

results showing a lower slope of the trend. It is understandable that the simulation would over-predict T_{max}, as the simulated chip was assumed to be fully insulated and surface roughness was not taken into account. SEM pictures of the surface roughness in channels fabricated using this method are shown in Figure 38. The surface roughness is approximately 1μm on both the sides and the bottom of the trench with the roughness pattern on the sides being more random (Figure 38b) and the pattern on the bottom being determined by the mask geometry (Figure 38a). Surface roughness will increase heat transfer because it affects the viscous sublayer of the flow near the walls. Evidence of increased heat transfer in the experiment as compared to simulation is further shown by the agreement of the outlet temperature of the fluid for all cases, as shown in Figure 39. The fluid outlet temperature, T_{out}, is related to the amount of heat into the chip by

$$q = \dot{m} C_p \left(T_{out} - T_{in} \right) \tag{18}$$

where $\dot{m}$ is the mass flow rate of water into the chip. Since T_{out} was the same for each case, this indicates that in the experiment almost all of the heat input was removed by the water while keeping the chip at a lower maximum temperature due to increased heat transfer. Figure 40 shows a contour map of the temperature distribution on the front and back sides of the simulated chip.

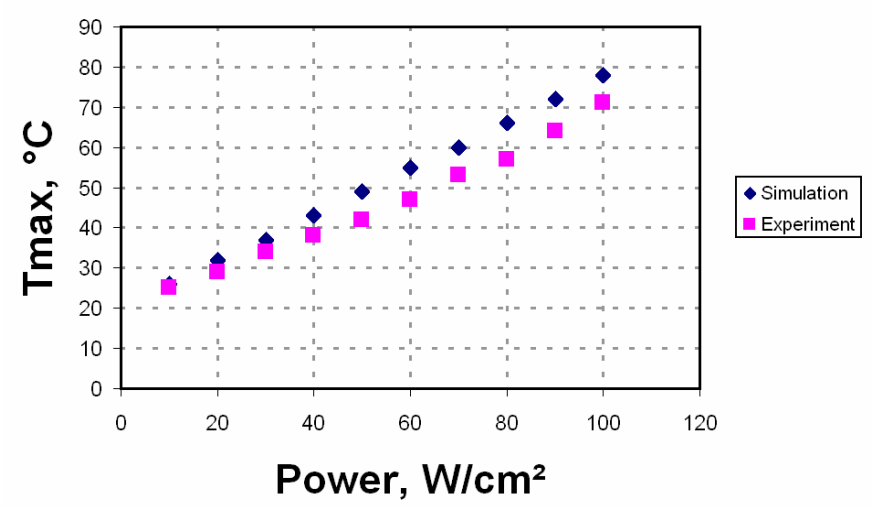

Figure 37: T_{max} vs. power density. The experiment shows enhanced heat transfer compared to the simulation.

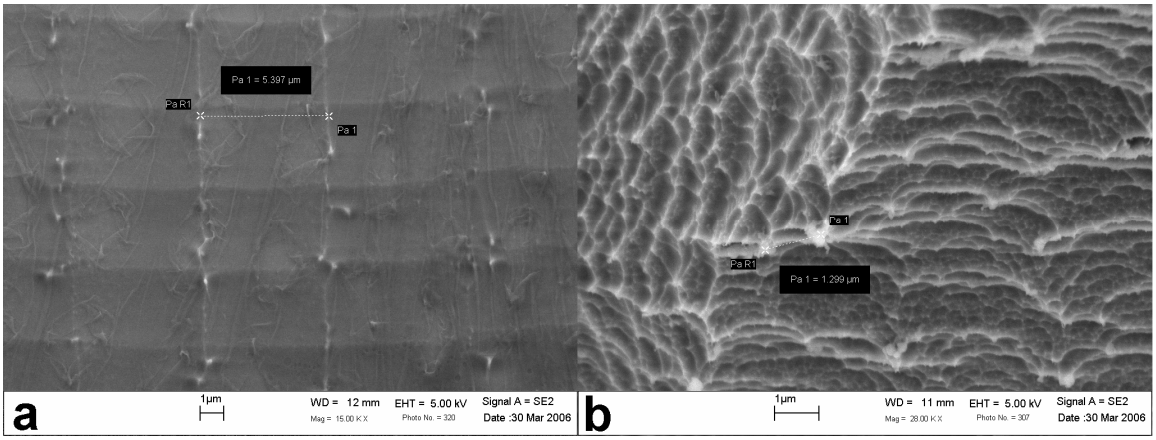

Figure 38: Surface roughness on the bottom and sides of trenches. The surface roughness in both cases is ~1µm.

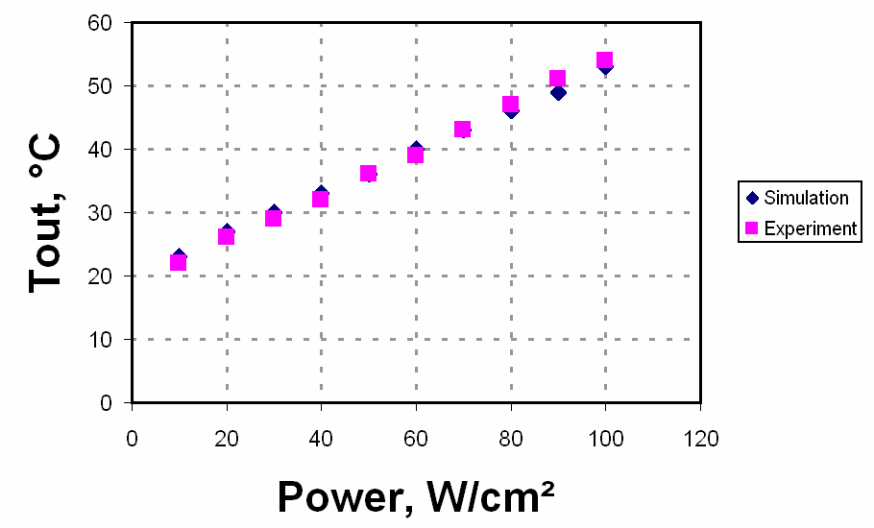

Figure 39: T_{out} vs. power density. Good agreement results between simulation and experimental results for the temperature of the water at the exit of the chip, demonstrating that the water was responsible for removing almost all of the heat.

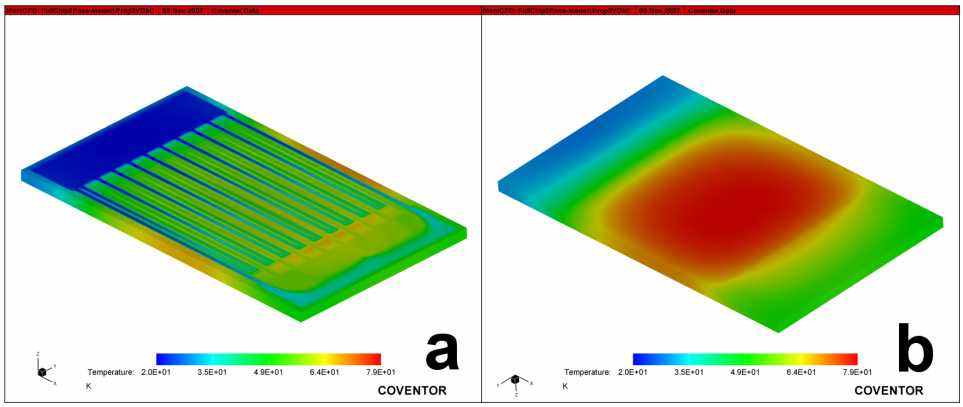

Figure 40: Temperature contour plot for the simulated cooling chip. The input power is 100W/cm^2. The red and blue areas are hot and cold, respectively. (a) The back-side of the chip with the outlet temperature of 53°C. (b) The heated front-side of the chip with a maximum temperature of 78°C.

Both the simulation and the experimental measurements do indicate that the meander channel design fabricated with the technique described in this thesis is capable of cooling in excess of 100W/cm^2 from a chip while maintaining the maximum temperature of the chip below its maximum allowable value, 85°C. According to the linear trend of the experimental data, at an identical flow rate of 750μL/s, 130W/cm^2 heat density can be removed from the chip while

keeping the chip just below 85°C. In this setup, 24mW is required in order to run the fluid at this rate, calculated by the product of pressure drop and flow rate. Compared to a chip with one large channel as wide as the heated area, simulation has shown that the meander channel chip used in this study is able to keep the chip 18°C cooler at a heat flux of 100W/cm^2 for the same flow rate.

The heat transfer capability of the chip can be further analyzed by calculating the thermal resistance of the chip. Thermal resistance is calculated by equation 16 and it represents the resistance to heat flow from the heated surface of the chip to the fluid. Figure 41 contains a plot of the thermal resistance vs. the power density input for the experiment and simulation. Both the simulated and the experimental results did not present much variation of R_{th} as the heat flux changed, but the experimental value for thermal resistance, ≈0.4°C/W, was lower than the simulation, ≈0.5°C/W. Again, the fact that the simulation had fully insulated boundary conditions and that surface roughness was not accounted for can explain the lower thermal resistance in the actual chip. The thermal resistance is the best measure of the overall heat transfer ability for this design, rather than *h* or Nu, due to the meander channel pattern. Usually, *h* or Nu is used to quantify the heat transfer ability of the chip, but since *h* is dependent on the location along the channel, an average value of *h* has to be found. Figure 42 shows the variation of *h* along the channel length measured from a simulated case with 100W/cm^2 input power. The two spikes in the data, the first is much smaller than the second, are located at the turns in the channel. It can easily be seen that there is not an appropriate average value of *h* for the channel geometry, and thus, no appropriate value for the Nusselt number.

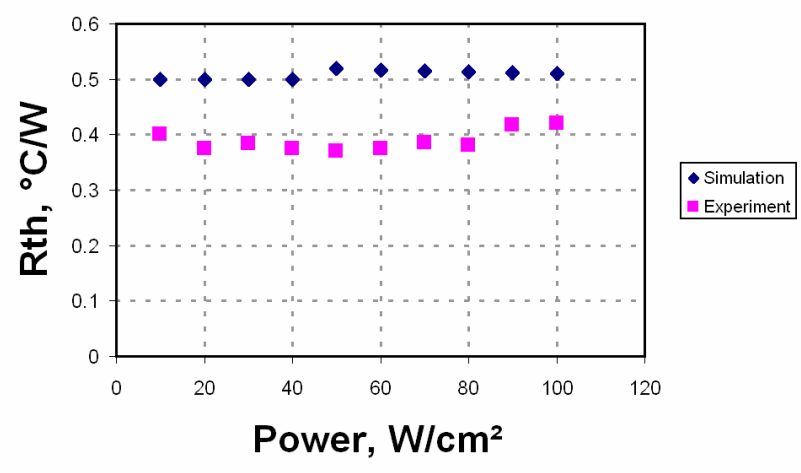

Figure 41: R_{th} vs. power density. R_{th} was fairly constant for all values of input power.

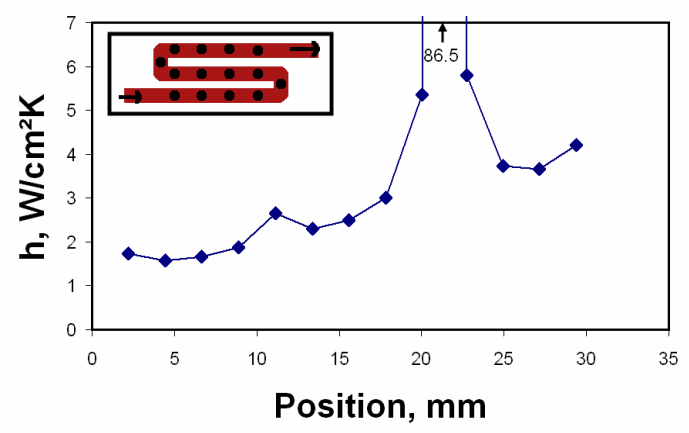

Figure 42: h vs. position over the meander channel. The heat flux is 100W/cm^2. The maximum value of h is 86.5W/cm^{2}K at the second channel turn shown in the sub-diagram. There is no appropriate average value for h over the channel.

The flow characteristics through the chip at varying input heat flux values were also examined. It had already been determined that as Reynolds number increases, there is some variance in the value of f*Re most likely due to entrance, exit, and developing flow effects. When f*Re was examined for the same flow rate of 750µL/s with increasing heat flux to the chip, f*Re was found to increase linearly, shown in Figure 43. However, since the entrance and exit effects are constant when the flow rate is the same, another mechanism is responsible for this change. In this case, the reason for the increase in f*Re is due to decreased viscosity of the water

when temperature increases. The average temperature of the fluid, found by averaging the inlet and outlet temperatures of the water, was used to calculate the average density and viscosity of the water, both of which are independent variables in equations 11 and 12 for the Reynolds number and friction factor, respectively. The simulation, on the other hand, predicted an almost constant friction factor close to the experimental values as heat flux to the chip increased. However, the simulation was carried out using constant fluid properties over the whole fluid. In both cases, constant water properties were assumed, but this was only actually true for the simulation, which returned a constant value for f*Re, which is in agreement with macroscale theory. Thus, it is likely that the observed variation of the product of the friction factor and the Reynolds number is due to the change in fluid properties from inlet to outlet and the inability of conventional equations to capture this change.

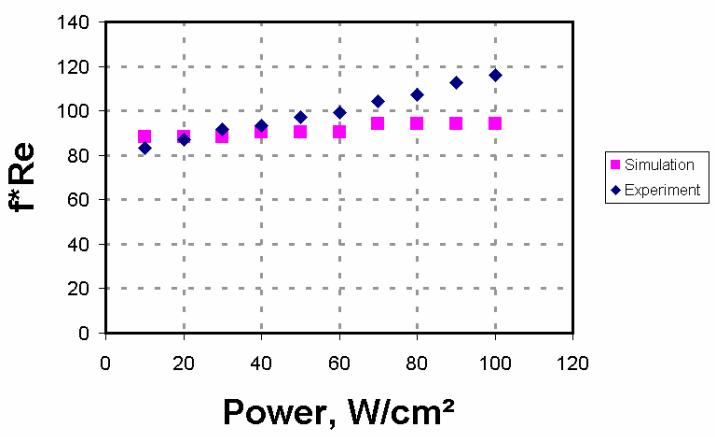

Figure 43: f*Re vs. power density. The increase in f*Re for the experiment is due to the temperature variation of the fluid over the chip.

When the exact values of f and Re are compared between the simulation and the experimental cases, a difference in behavior is observed. Figure 44 contains a plot of these values versus the heat flux into the system. The simulated case was run for four different sets of water physical properties, depending on the average temperature of the fluid, hence, the step changes that are shown are more likely even and smooth changes. Regardless, the simulation shows an

increase in Reynolds number and a decrease in *f* as the heat flux, and thus the temperature, of the system increases. However, the experimental results only show a noticeable increase in Re and not *f*. The effect of changing temperature on *f* is much less in the experimental case because the viscosity of the fluid is still high at the inlet where the temperature of the fluid is 20°C, even though it decreases at the end of the channel where the fluid is 54°C. In the simulation, the viscosity is assumed to be constant for the fluid over the whole chip, so the pressure drop reduces by approximately 0.5kPa/°C of the average water temperature, which is physically inaccurate. In addition, the simulation under-predicts the value of *f* compared to what was actually observed. The reasons for this difference in values can be linked back to the same reasons that the heat transfer results were slightly different. The simulation also predicted a higher Reynolds number than the experimental results indicated.

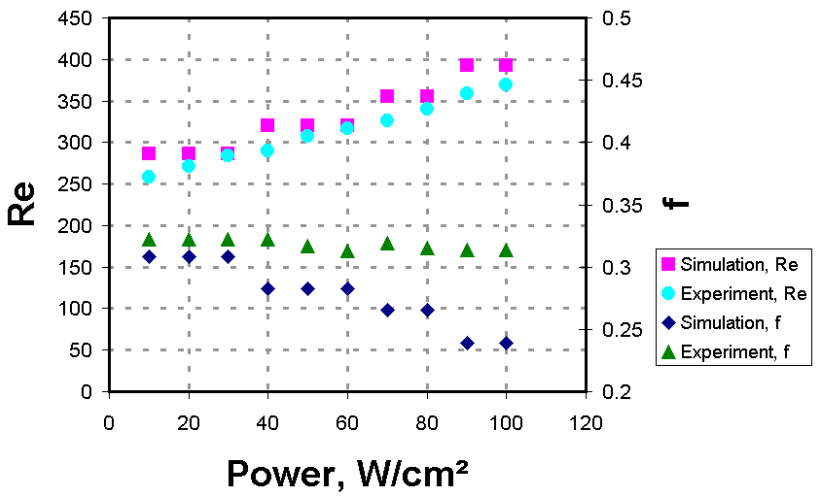

Figure 44: Re and *f* vs. power density. The simulation was run for four different fluid property values according to the average temperature of the fluid; thus, the step changes observed are more likely even, smooth trends.

Examining the results of the experiment compared to the simulation, the results generally agree when the assumptions that were made in each case are taken into account. The major sources for inaccuracy between the two data sets are simplified boundary conditions, zero surface

roughness, and constant fluid properties over the chip in the simulated channels. If the simulation was run with variable fluid properties over the wafer, more similar results to the experimental case are expected to result, especially in terms of *f*. The results of heat transfer showed very good agreement between both sets of data. Based on the data collected, there is no concrete evidence to suggest any difference in the behavior of the fluid in the macroscale and the microchannels used in this study.

When addressed from the standpoint of performance, the chip was proved to be capable of cooling the next generation of microprocessors, demonstrated by the potential to cool 130W/cm^2 with a flow rate of 750μL/s, or 45mL/min. Compared to results presented by Pijnenburg et al., the work which this study's cooling design was inspired by, the isotropically fabricated channels used here show less cooling ability and degraded flow performance. However, these results are understandable considering that this study used a meandered channel and the channels used by Pijnenburg et al. passed straight from one reservoir to the other across the heated surface, reducing the pressure drop over the channels. Their channels were also high aspect ratio channels (3:1), giving enhanced performance as previously described in the chip cooling literature review. A summary of the results of each study are given in Table 10. Although the channel aspect ratio is not as flexible using the RIE-lag-based etch process, improved performance will result if the channels pass straight over the heated surface, producing much more similar results to the work by Pijnenburg et al.

Despite the limitations in achievable channel aspect ratio, a cooling chip designed and fabricated using predictable three-dimensionally independent fabrication as demonstrated in this thesis are more than sufficient to keep microprocessors and other high-flux microchips cool for the next 10 years. However, there is still a great amount of room for performance enhancement without changing the fabrication process at all. Modification and optimization of the channel geometry will result in an even greater amount of heat transfer than was presented here, providing

the potential for a quick, low-cost option for heat management at the microscale for an even longer period of time.

Table 10: Cooling Comparison between Pinenburg et al. and Gantz.

	Pijnenburg et al.	Gantz
	@ Q = 45mL/min	
R_{th} (°C/W)	0.33	0.4
$\Delta P/\Delta x$ (kPa/mm)	0.67	1.74
	Published results @ heat flux = 100W/cm^2	
Q (mL/min)	89	45
$\Delta P/\Delta x$ (kPa/mm)	1	1.74
R_{th} (°C/W)	0.285	0.4
T_{out} (°C)	38	54
T_h (°C)	50	62
Power (mW)	16	24