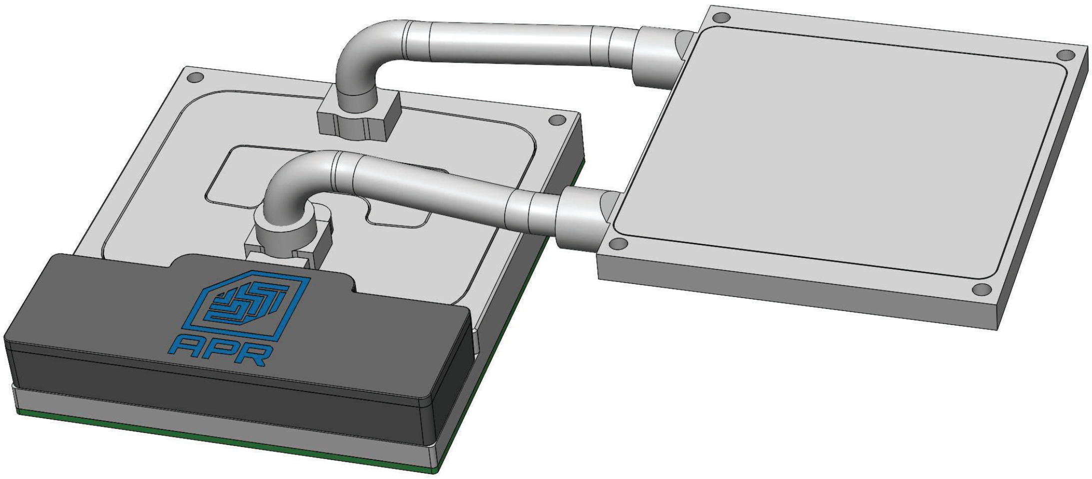
Effektiv mikro-vätskekylning av högprestanda-processorer för självkörande fordon (ChipCooling)

Publik rapport



Författare: Nils Sabelström, Elisabeth Söderlund

Datum: 2022-01-29

Projekt inom "Elektronik, mjukvara och kommunikation - FFI - 2019"

FFI Fordonsstrategisk Forskning och Innovation

Innehållsförteckning

1 Sammanfattning

Projektet har framgångsrikt demonstrerat att APR Technologies´ koncept för kylning av en fordons-ECU på 30W leder till en driftstemperatur <60°C på komponentnivå, och en förväntad ökning av livslängd ca 3 gånger (upp emot 15 år). Överskottsvärmen kan t.ex. ledas till bottenplattan eller existerande kyl-loop.

Den nya kylteknologin baseras på APR Technologies´ miniatyriserade pump och dielektrisk vätska. Tekniken är mycket effektiv, kompakt och helt tyst. Pumpen är konstruerad utan rörliga delar och integrerad i systemet, vilket leder till ett antal fördelar:

- ökad tillförlitlighet
- ökad livslängd
- ökad designfrihet
- reducerad energiförbrukning
- material-och kostnads-besparingar
- minskad mängd avfall
- minskat behov av service och/eller längre serviceintervall
- förbättrad hyttmiljö (svalare och tystare)

Två prototyper har tillverkats, varav den första enbart kyler den centrala enheten SoC (CPU, GPU etc), medan den andra har en större kyld area och även inkluderar omkringliggande komponenter (dvs hela ECUn). Den kylda ECUn är en på marknaden vanligt förekommande ECU och systemet kan lätt anpassas till liknande ECUer, vilket bör underlätta en framtida marknadsintroduktion.

Den mindre prototypen (SoC) erhöll en temperatur på 55°C. Temperaturen hos den större (ECU) landade på 58°C. Båda systemen uppfyller därmed projektmålet (<60°C). Den första prototypen utvärderades i temperaturintervallet 7-60°C, medan temperaturintervallet för den senare protypen utsträcktes ned till -41°C. TRL 5 har uppnåtts för APR´s teknologi i detta projekt.

Referenssystemet (fläktbaserat) resulterade i en temperatur på 62.5°C när den kördes på maxeffekt. Prototyp 2 (ECU) är 11% lättare och 22% mindre än referenskylaren.

Testerna är utförda på ett sätt som lätt möjliggör för kunder att jämföra med existerande referenslösningar. Livslängden är direkt relaterad till temperatursvängningar under drift, och med en temperatur under 60°C medför att livslängden på huvudkretsen bör sträckas till 15 år [1].
Marknaden för ECUs ökar i takt med att ökat antal säkerhetskomponenter implementeras i fordon, och fordonen går mot högre grad av automation. Behovet av effektiv kylning ökar i samma takt. SAE har definierat fem nivåer av automation där nivå 5 är helt självkörande. En effektnivå på 30W kan sägas motsvara en SAE nivå 3 (conditional automation). I detta projekt har en effektnivå på 30W implementerats, vilket motsvarar en SAE nivå 3-4 (conditional to high automation). I ett kommande steg planerar vi ta fram en liknande lösning för 50W, vilket motsvarar en SAE nivå på 4-5 (high to full automation). Idag befinner sig ca 10% av fordonen på nivå 3. 100% självkörande har uppskattats kunna nås ca 2050, eller något senare [2].

I nästa steg strävar APR Technologies efter att en eller flera OEMs ska utvärdera lösningen i sin applikation och att lösningen därefter skall komma att designas in i produkter på marknaden.

2 Executive summary in English

The project has successfully demonstrated that APR Technologies' concept for cooling a vehicle ECU of 30W leads to an operating temperature <60°C at the component level, and an expected increase in service life of about 3 times (up to 15 years). The excess heat can e.g. be led to the bottom plate or existing cooling loop.

The new cooling technology is based on APR Technologies' miniaturized pump and dielectric fluid. The technology is very effective, compact and completely quiet. The pump, designed without moving parts, is integrated in the system, which leads to a number of advantages:

- increased reliability
- increased product lifetime
- increased design freedom
- reduced energy consumption
- material and cost savings
- reduced amount of waste
- reduced need for service and / or longer service intervals
- improved cab environment (cooler and quieter)

Two prototypes have been manufactured, of which the first only cools the central unit SoC (CPU, GPU etc.), while the second has a larger cooled area and also includes surrounding components (i.e. the entire ECU). The cooled ECU is an ECU commonly used on the market and the system can be easily adapted to similar ECUs, which should facilitate future market introduction.

The smaller prototype (SoC) obtained a temperature of 55°C. The temperature of the larger (ECU) landed at 58°C. Both systems thus meet the project target (<60°C). The first prototype was evaluated in the temperature range of 7-60°C, while the temperature range of the latter prototype was extended down to -41°C. TRL 5 has been achieved for APR´'s technology in this project.

The reference system (fan based) resulted in a temperature of 62.5°C when running at maximum power. Prototype 2 (ECU) is 11% lighter and 22% smaller than the reference cooling system.

The tests are performed in a way that easily enables customers to compare with existing reference solutions. The service life is directly related to temperature fluctuations during operation, and with a temperature below 60°C means that the service life of the main circuit should be extended to 15 years [1].

The market for ECUs is increasing as an increasing number of safety components are implemented in vehicles, and vehicles are moving towards a higher degree of automation. The need for efficient cooling is increasing at the same rate. SAE has defined five levels of automation where level 5 is completely self-driving. A power level of 30W can be said to correspond to an SAE level 3 (conditional automation). In this project a power level of 30W has been implemented, which corresponds to a SAE level of 3-4 (conditional to high automation). In the next step we plan to develop a similar solution for a 50W case, thus preparing for SAE level 4-5 (high to full automation). Today, about 10% of the vehicles are at level 3. 100% self-driving has been estimated to be reached around 2050, or slightly later [2].

In the near future, APR Technologies will strive for one or more OEMs to evaluate the solution in their application and for the solution to be designed into products on the market thereafter.

3 Background

Autonomous vehicles need high-end computer processing capability. The current market for such computing capability is dominated by consumer devices, with typical product lifetimes in the order of 3-5 years [3] [1], whereas the vehicle market needs components with lifetimes of 10-15 years. The need for advanced thermal management will be critical, especially within e-vehicles, and even more so in autonomous vehicles. It will be shown that the thermal management concept proposed by APR Technologies will enable such product lifetimes.

Development during the last few years have been extremely fast, where a transition from serial to parallel processors has occurred and several suppliers now have products in the processing device segment, for example Qualcomm, Nvidia and Intel. At the same time, the power consumption has been significantly lowered. speeding up the transition towards AD.

However, these novel processors, originally developed for laptops, gaming computers and in-door (office) environments (ambient temperature of approx. 20°C), still have a guaranteed life of 3-5 years [3] [1], whereas vehicle applications rather call for a guaranteed life of 10-15 years. The presently adopted cooling solutions, such as fans, are usually bulky, noisy, have limited a lifetime. Recent surveys show that buyers are increasingly concerned about the reliability of new car models [4].

In the automotive industry, a considerable effort is underway in the development of Advanced Driver Assistance Systems (ADAS) and Autonomous Driving (AD). These systems require a massive amount of sensory data from cameras, lidar, and radar etc. Efficient and low power solutions for removing the excessive heat in these units are of the uttermost importance. This is especially for electrical vehicles, where the power used to cool the units must be minimized. The units may in some cases be mounted in the vehicle cabin and the heat shall be transported to the external environment.

In this project the purpose was to demonstrate APR Technologies' concept idea for cooling of an ECU from Nvidia. The proposed solution is based on APR Technologies' proprietary pump.

Figure 1 - An Illustration of an envisioned Autonomous Driving experience

4 Purpose, research-related questions and method

The purpose was to demonstrate APR Technologies concept idea for cooling of heat-generating electronics, with focus on processors used in autonomous driving (AD), using APR's miniaturized pump and dielectric fluid. The technology is very effective, compact and completely quiet. The pump, designed without moving parts, is integrated in the system, which leads to a number of advantages:

- increased reliability
- increased product lifetime
- increased design freedom
- reduced energy consumption
- material and cost savings
- reduced amount of waste
- reduced need for service and / or longer service intervals
- improved cab environment (cooler and quieter)

The prototype design was divided into two stages. In the first stage a system was designed to cool only the main component (the SoC, System on Chip, containing both the CPU and GPU) of the ECU, to easily test design parameters. In the second stage, a chip level sized pump and cooling system was made small enough to be integrated into the aluminum casing of the ECU, that can be easily exchanged from the casing included in the ECU module as sold on the market.

5 Goal

The overall goal was to lower the temperature at the component level from 85°C or more, to <60°C, by improving the cooling capacity by a factor 3-5 [5], thereby increasing the lifetime of processors and surrounding components about 3 times, from about 3-5 to 10-15 years [3] [6], which is a desirable service life in the automotive industry.

Additional objectives:

- Cool a 30W chip to below 60°C, at an environment of 35°C
- Achieve TRL 5 for the solution
- Analyze the Freedom-to-Operate
- Formulate an IP strategy
- Perform a market analysis

6 Results and goal fulfillment

6.1 IP landscape

In this project Chalmers Industriteknik performed an analysis of the IP landscape. It was found out that the area of cooling of electronics and chips is a growing area, which had a peak around the year 2005-2006. It proves that the market is interested in this kind of solution and R&D activities are strong.

The country with the highest number of published patent applications is China. The growth observed in this country since 2006 is doubling every 4 years. There may be a market-driven reason for this behavior, but more likely there is a local incentive. Some time ago a patenting law in China changed and after that, the described trend started to be observable in all technology fields. In other countries, the patenting trend is rather stable, with just a slight growth over the last 10 years. After China, USA has the lead position, followed by applications filed in the PCT process, the EPO process, and Japan.

Furthermore, the analysis showed that Intel, IBM, and GlobalFoundries had the biggest overall portfolios. Other companies worth noting were NVIDIA, Qualcomm, Bosch, and Ericsson. Other players had comparably much smaller portfolios or had no patent applications at all in the patent classes referring to heat exchange and electronics. It is worth mentioning that patent application count in classes related to heat exchange was significantly lower than the classes related to electronics, which suggests that the latter companies are not active competitors and rather could be early adopters.

The number of identified relevant patent applications is very small compared to the total number of found patent applications. Although the idea of direct chip cooling is not novel, the technical challenges around it are complex, and not many solutions are fully developed and available commercially. It makes this niche rather interesting for further R&D development and gives a good possibility for patenting innovative solutions.

No directly impeding patents have been found during the patent landscape analysis. From the IP landscape studies a strategy was developed and will be integrated into the continued product development work.

6.2 Design & testing of prototypes

The cooling system was designed to be compatible with a prominent ECU made specifically for AI calculations already introduced to the market, the Nvidia Jetson AGX Xavier.
This ECU is available both as an ECU module to add into a present control system in a car, and as a development kit with a reference cooling solution. The prototype was made using the development kit, to enable easy comparison for prospective customers, who are sure to have internal test results that can be compared to the reference cooler. While the prototypes within the scope of this project are made specifically for this ECU, the system is designed to be easily adjustable to other similar ECU designs of any brand.

The cooling system in full contains a heat exchanger at the heat generating device that is connected through a liquid loop to a cold side heat exchanger. This cold side should in the end user case should be connected to an already present high performance cooling system, such as battery cooling in electric vehicles or motor cooling system in an internal combustion engine car.

6.2.1 Prototype 1

For prototype one, the system was designed for modularity and expandability, to enable testing with different sizes of the proprietary pump technology developed by APR Technologies AB. In this stage, only the heat exchanger on the heat generating side was included, cooling only the SoC, using a chiller type heat exchanger on the cold side of the loop.
Other heat generating PCB components were cooled using an aluminum heatsink. The prototype is shown in Figure 2.

From the testing, an optimal design was selected, and this system was used for the performance data displayed in the results section of this report.

Figure 2 – Prototype 1, as assembled on the development kit.

6.2.2 Prototype 2

In the second prototype, the proprietary pump and the entire cooling system was condensed into a total size comparable to the ECU module aluminium casing. This system not only cools the main chip but the entire ECU. The finished prototype is shown in Figure 3.

6.2.3 Test results and final specification

Thermal tests were conducted using a publicly available load software, the *jetson_benchmarks* project's *unet-segmentation* benchmark , which was found to produce the highest junction temperatures and most stable power output during testing.

In Table 1 below, the results of the evaluated prototypes and the reference design cooler are summarized.

Note that since prototype 1 is considered a development step for prototype 2, the weight requirements were disregarded.

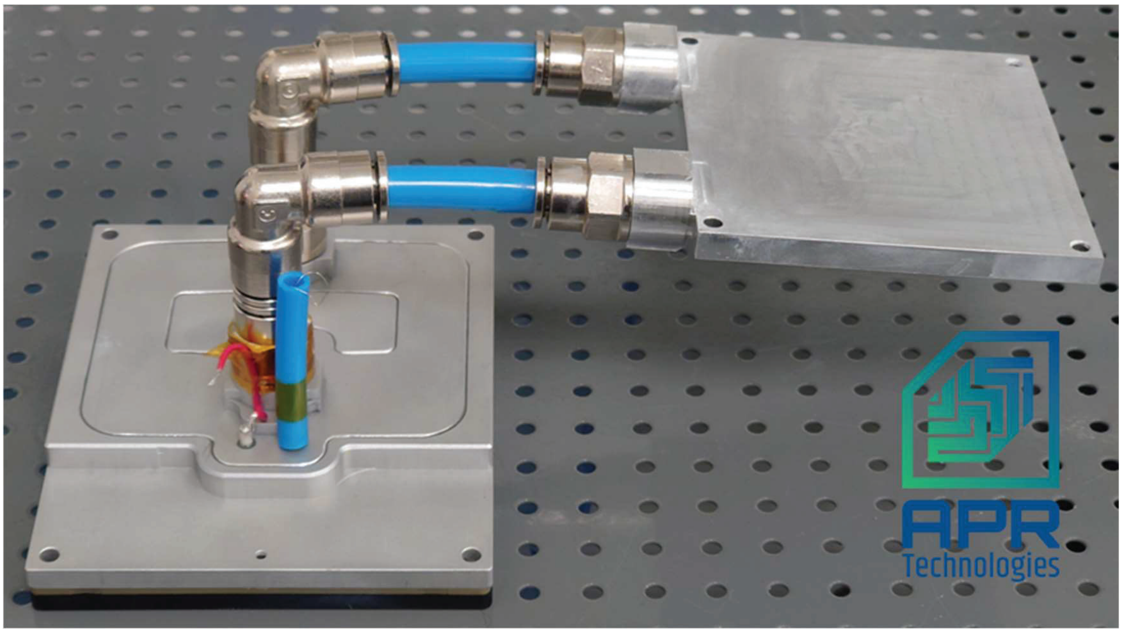


Figure 3 – Prototype 2 mounted on the ECU module

It can be seen that prototype 2 is both smaller and lighter than the reference system, while providing adequate cooling performance even without any moving parts. A small performance decrease can be observed between prototype 1 and 2, where not only the main SoC component is cooled, but the entire ECU.
The system was shown to work in the full range of -41°C to 60°C.
Note that while the reference cooler is close to the target, the fan was forced to run at full speed, in intended operation the hotspot is around 85°C. The fan was forced to full speed for ease of comparison for prospective partners.
At the testing temperature 35°C, The ECU had a power output that was slightly above the specified 30W.

Table 1 – Comparison chart between the evaluated systems and target specification.

Parameter	Target Specification	Reference cooler	Prototype 1	Prototype 2
Weight	< 350g	350g	574g	312g
Volume	< 260ml	260ml	-	187ml + tubing 15ml
SoC Hotspot Temperature Tair = 35°C, 30.5W	< 60°C Tair = 35°C, 30W	62.5°C (100% Fan duty)	55°C	58°C
Temperature range	-40°C to 35°C	-20°C to 57°C *1	7°C to 60°C	-41°C to 60°C
No moving parts	yes	No	yes	Yes

*1 Low temperature is assumed as a common value, high temperature from cooling performance.

From these results, the required specification of the system was met, and the resulting low temperatures should translate to a prolonged product life expectancy of 10-15years.

6.3 Product requirements & market for automotive ECUs

A requirement specification and analysis of the automotive ECU market has been conducted and provided by Veoneer.

In the automotive industry, a considerable effort is underway in the development of Advanced Driver Assistance Systems (ADAS) and Autonomous Driving (AD). These systems require a massive amount of sensory data from cameras, lidar, and radar etc. to achieve a sufficient safety level. At the same time, the industry is moving from a distributed architecture with a central gateway towards what is called a Consolidated Zonal Architecture, increasing the power needed per processing unit.

Typical suppliers of processing devices are Qualcomm, Nvidia and Intel. Some vehicle OEMs, like Tesla, have their in-house developed central vehicle units although this currently only does the FSD (Full Self-Driving) application.

The future needed power consumption of these consolidated processing units can be estimated by the needed processing operation throughput, and the efficiency of the processor. This efficiency is usually compared using the unit TOPS/W (Tera Operations Per Second and Watt), which ranges from 0.1 to 5 TOPS/W. Where more advanced automotive features will use the higher end of around 5 TOPS/W. The throughput of the needed processing in TOPS has been anticipated depending on the SAE level (The level of automation from L1 to L5), and is summarized in Table 2 below:

Table 2 – Expected heat output of future ECUs for each SAE level.

SAE level	Required TOPS	Corresponding Power
L1	-	-
L2	2	< 20W
L3	10	< 30W
L4	50	< 50W
L5	250	50W -

With increasing SAE level, higher end processing will be used, and as can be seen in Fig. 3, a large segment will be using processing units in the range of 20-50W. In the Figure 4, the transition to 100% L5 is projected to take until at the earliest around 2050 [2].

Thus, the market need for cooling solutions in this power range is expected to be considerably large and the need is projected to last for a long time.

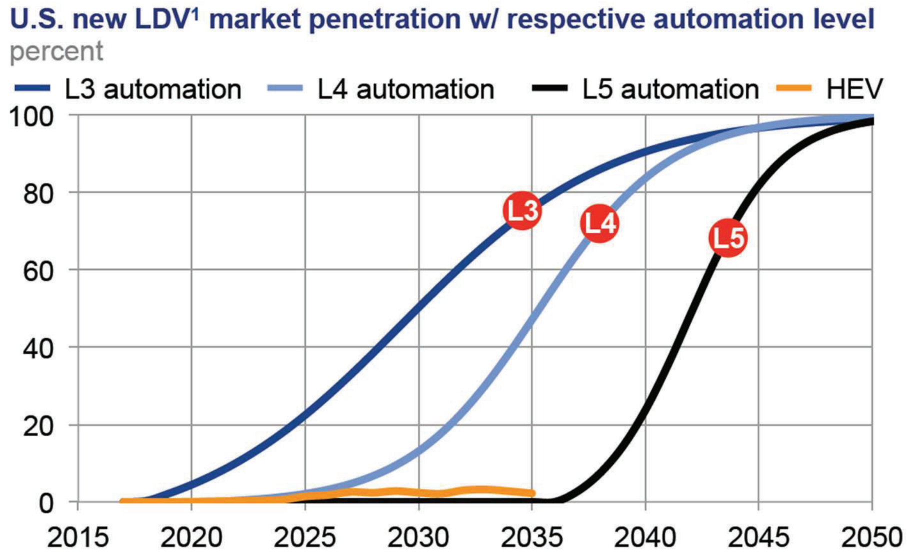


Figure 4 – Expected time to market penetration for SAE level L3 to L5 [2]

6.4 Analysis of production technical aspects and commercialization

As produced using prototyping production methods, the system has a high production cost. However, the system is designed with automation and high production rate assembly techniques in mind, such as laser weld sealing of the different parts, and automated robot assembly of the proprietary pumps. This pump technology should with the right investments and necessary research be possible to produce at costs comparable even to air-cooled systems that use industrially safe and reliable fans.

Presently APR Technologies AB is planning to build a pilot line for production of cost-efficient pumps for electronics applications. It will be able to produce 40 000 pump units per year by early 2023.

7 Dissemination and publishing

7.1 Dissemination of knowledge and results

Hur har/planeras projektresultatet att användas och spridas?	Markera med X	Kommentar
Öka kunskapen inom området	x	
Föras vidare till andra avancerade tekniska utvecklingsprojekt	x	
Föras vidare till produktutvecklingsprojekt	x	
Introduceras på marknaden	x	
Användas i utredningar/regelverk/ tillståndsärenden/ politiska beslut		

The dissemination will mainly occur through this report, APR´s home page, press releases, publications, conference contributions and market activities.

7.2 Publications

No publications have been made so far but are planned in the near future.

8 Conclusions and future studies

Within this project, it was concluded that:

- The produced prototype meets the specification:
    - Chip hotspot was < 60°C in a 35°C environment for powers up to SAE L3 ECUs
    - Usable temperature range of -41°C to +60°C
    - Lighter and smaller than the cooler reference design
    - Has no moving parts
- The prototype has been developed to a TRL level of 5
- The produced prototype is adaptable to other ECUs and is design for manufacturability
- Our proprietary pump technology IP is novel, and no directly impeding patents have been found during the patent landscape analysis
- There is a large market need for this type of technology

There are numerous points of future studies to conduct. While the prototype met specification, there is always room for improvement, and while the system is designed to be easily adapter to other ECUs and cheaper manufacturing techniques, deeper analysis would be needed as a final step before entering mass production.

Some possible future studies are listed below:

- Develop and evaluate a similar 50W case (adopted for SAE level 5)
- Performance optimization studies, to reach SAE L5 segment ECUs
- Testing to achieve higher TRL level
- Making and testing customized systems for end-user specific ECUs
- Mass production cost and required investment analysis

9 Participating parties and contact persons

Party	Contact person	Email	Logo
APR Technologies AB	Elisabeth Söderlund	elisabeth.soderlund@aprtec.com	
Veoneer Sweden AB	Olof Eriksson	olof.eriksson@veoneer.com	
Stiftelsen Chalmers Industriteknik	Camilla Johansson	camilla.johansson@cit.chalmers.se	

References

[1] A. Webber, "Calculating Useful Lifetimes of Embedded Processors," Texas Instruments Application Report SPRABX4A, October 2017.

[2] "Connected and Automated Driving Systems," [Online]. Available: http://smarttransport.solutions/2018/05/29/expected-timeline/. [Accessed 27 January 2022].

[3] K. Boughton, "The Truth About Processor "Degradation"," Anandtech, 5 march 2008. [Online]. Available: https://www.anandtech.com/show/2468/6.

[4] "U.S. Auto Avoider Study Finds," *J.D.Power,* 2016.

[5] "Handbook for Robustness Validation of Automotive Electrical/Electronic Modules," in *SAE J1211*.

[6] E. Sperling, ""Are Chips Getting More Reliable?"," *Semiconductor Engineering,* January 2016.

[7] "Jetson_benchmarks Project," [Online]. Available: https://github.com/NVIDIA-AI-IOT/jetson_benchmarks.