Subatmosphärische Kühlung von Rechenzentren mit Wasser

Bodo Burandt[1*], Mathias Safarik[1],
Andreas Hantsch[2], Conrad Wächter[2]

[1] Institut für Luft- und Kältetechnik gGmbH, Bertolt-Brecht-Allee 20, 01309 Dresden
bodo.burandt@ilkdresden.de

[2] Cloud&Heat Technologies GmbH, Königsbrücker Str. 96, 01099 Dresden
andreas.hantsch@cloudandheat.com

* Korrespondenzautor

Kurzfassung

Rechenzentren haben einen hohen Elektroenergiebedarf, erzeugen Wärme und brauchen Kühlung. Sie wandeln 100 % des IT-Stromes in Wärme um. Die Leistungsdichte der Racks und der installierten Chips wächst stetig. Es wird von einer Leistungssteigerung von derzeit 350 W auf 1.000 W und mehr pro Chip ausgegangen. Die Wärmeabfuhr ist mit Luftkühlung nicht mehr zu realisieren. Eine direkte Flüssigkeitskühlung ist erforderlich.

Wasser ist ein idealer Wärmeträger mit nur einem Nachteil, es ist elektrisch leitend. Damit den Betreibern die Angst vor dem Kühlmedium Wasser genommen wird, erfolgt der Betrieb des Wasserkühlsystems bei einem Druck unter dem Umgebungsdruck, so dass bei einem Leck keine Flüssigkeit austritt. Es wird ein System der subatmosphärischen Kühlung inklusive einer Absicherung, dass kein Kühlmedium bei einem Leck austritt vorgestellt. Die Ergebnisse der praktischen Untersuchungen der subatmosphärischen Einphasenkühlung eines Racks sowie die Beherrschung von simulierten Lecks werden dargestellt.

Bei den hohen Energiedichten der Chips stößt auch die Einphasenflüssigkeitskühlung an seine Grenzen. Somit ist ein Übergang auf die subatmosphärische Zweiphasenkühlung logisch. Durch die Nutzung der Verdampfungsenthalpie wird deutlich weniger Wasser im Umlauf des Systems benötigt. Erste Ergebnisse der Tests der subatmosphärischen Zweiphasenflüssigkeitskühlung mit Wasser werden erläutert.

Stichwörter:
Flüssigkeitskühlung, Wasser, subatmosphärisch, Zweiphasenkühlung

1 Einleitung

Die Kapazitäten der RZ in Deutschland haben in den vergangenen 5 Jahren kontinuierlich zugenommen. Gemessen an der maximalen Stromaufnahme der installierten Hardware sind die Kapazitäten zwischen 2016 und 2021 um 30 % gestiegen. Es wird von einer weiteren Steigerung der RZ-Kapazitäten um voraussichtlich 20 % von 2021 bis zum Jahr 2025 ausgegangen [1]. Im Jahr 2014 betrug der Elektroenergiebedarf für RZ 1,8 % des Gesamtbedarfs in den USA [2]. Eine weitere Studie [3] verweist auf einen Anstieg um etwa 4 % von 2010 bis 2016, 24 % von 2005 bis 2019 und fast 90 % von 2000 bis 2005 und erwartet einen Anstieg um 4 % von 2014 bis 2020. Für das Jahr 2020 wurde für die USA 73 Mrd. kWh Energiebedarf für die RZ prognostiziert. Der Energiebedarf für Rechenzentren in Deutschland von 2010 bis 2022 ist im Bild 1 dargestellt. Für 2022 wird für die Rechenzentren von einem Energiebedarf von 17 Mrd. kWh/a ausgegangen [4].

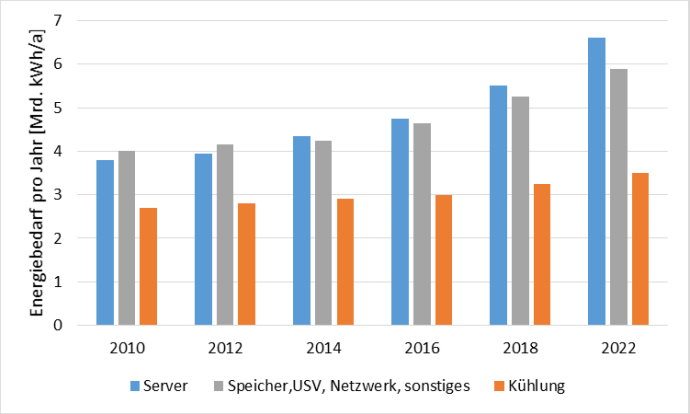

Bild 1: Energiebedarf für RZ in Deutschland (Werte aus [4])

Neben der Gesamtleistung der Rechenzentren nimmt auch die Energiedichte pro Rack und pro Chip zu. Ein durchschnittliches Rack hat eine Leistungsaufnahme von 7 kW, Hochleistungsracks haben bereits eine elektrische Anschlussleistung von bis zu 50 kW. Auch die Leistungsdichte der Chips nimmt deutlich zu. So hat z. Bsp. der Intel Core 2 Extreme QX6700 eine spezifische Leistung bezogen auf seine Fläche von 45 W/cm² und der AMD Athlon 64 X2 4600+ von 75 W/cm² [5]. Im Vergleich dazu liegt die spezifische Leistung eines Bügeleisens bei ca. 5 W/cm² und einer Kochplatte bei ca. 8 W/cm². Chip-Hersteller haben Produktpläne für CPUs und GPUs, die in den nächsten Jahren voraussichtlich 500 Watt pro Prozessor überschreiten werden [6].

2 Kühlung von Rechenzentren – Stand der Technik

2.1 Luftkühlung

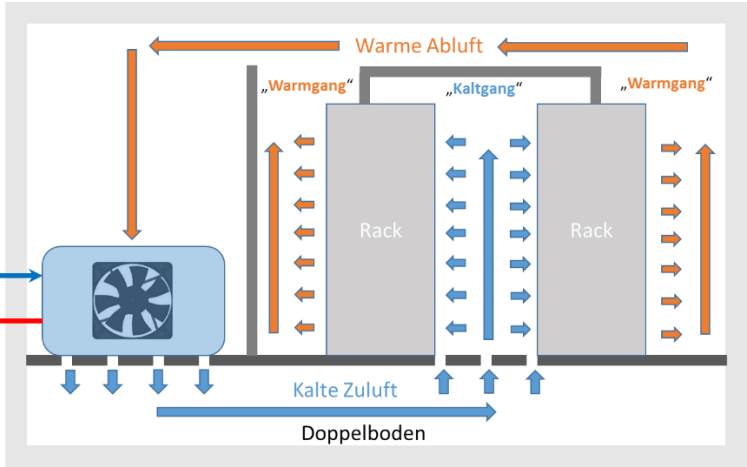

Bild 2: Aufbau Luftkühlung von Rechenzentren

Bei der Luftkühlung wird kalte Luft über einen Doppelboden zur Vorderseite der Racks geleitet. Die Luft durchströmt die Racks und führt deren Wärme auf der Rückseite ab. Die Luftführung ist durch eine entsprechende Einhausung so gestaltet, dass es zu keiner Vermischung von Warm- und Kaltluft kommen kann. Die erwärmte Luft strömt dann zu einer Computerraum-Klimaanlage wo sie dann mit kaltem Wasser abgekühlt wird (Bild 2). In typischen luftgekühlten Systemen wird die Zuluft auf Temperaturen von 12 bis 14 °C geregelt, die Rückluft erreicht die Luftkühler mit ca. 18 bis 22 °C [7] bzw. mit bis zu 35 °C [8]. Durch die Luftkühlung kann die Wärme auf Grund der Leistungssteigerung der Mikroprozessoren, die Wärmeflüsse von bis zu 100 W/cm^2 aufweisen, nicht mehr bei Einhaltung der zulässigen Betriebstemperatur abgeführt werden. In [9] wird für Luft eine maximale Wärmeabfuhrkapazität von 37 W/cm² für luftgekühlte Kühlkörper angegeben.

Weitere Möglichkeiten der Luftkühlung sind die Rack-Hintertür- und die Rack-Seitenwand-Kühlung. Hierbei werden wassergekühlte Luftwärmeübertrager in bzw. an den Rackschrank angeschlossen.

2.2 Flüssigkeitskühlung

Bei Flüssigkeiten sind die spez. Wärmekapazität 1.000 – 3.500-fach und die Wärmeleitfähigkeit 15 bis 25-fachgrößer als bei Luft.
Die Flüssigkeitskühlung wird in indirekte Flüssigkeitskühlung (Cold Plate / Direct-To-Chip) und direkte Flüssigkeitskühlung (Immersion) unterteilt.

Bei der indirekten Flüssigkeitskühlung sind flüssigkeitsdurchströmte Kühlkörper auf den Chips (CPU, GPU …) zur Abführung der Wärme angeordnet (Bild 3). Als Kühlmedium kann Wasser, dielektrische Flüssigkeiten und Kältemittel eingesetzt werden.

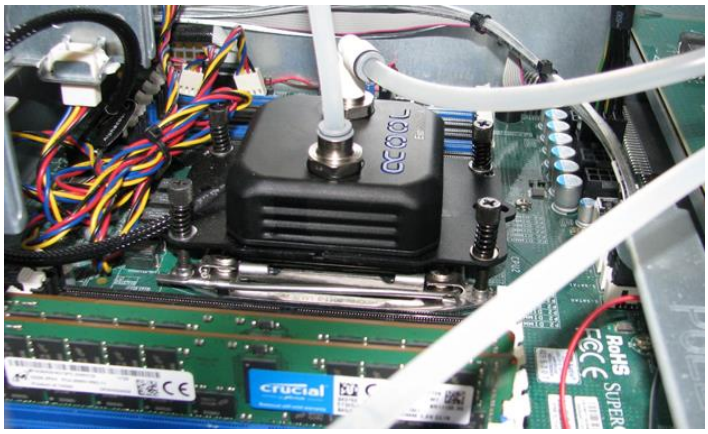

Bild 3: Kühlung GPU mittels flüssigkeitsdurchströmtem Kühlkörper

Bei der direkten Flüssigkeitskühlung wird der gesamte Server bzw. das gesamte Rack in eine dielektrische, wärmeleitende Flüssigkeit getaucht (Bild 4). Bei dieser Methode entfällt der Bedarf an Luftkühlung, da alle Komponenten in der Flüssigkeit eingetaucht sind. Die Wärme wird durch direkten Kontaktmit den Serverkomponenten an die Flüssigkeit übertragen. Oben wird die warme Flüssigkeit abgepumpt und über einen Wärmeübertrager zur Wärmeauskopplung geleitet und dann unten wieder gekühlt in den Tank zurückgeführt. Es gibt einphasige und zweiphasige Eintauchkühlsysteme. Als dielektrische Flüssigkeiten können Mineralöl, Flüssigkeiten auf Fluorkohlenstoffbasis und synthetische Flüssigkeiten eingesetzt werden. Nachteile der Immersion-Kühlung sind:

- Alle Komponenten der Server bzw. Racks müssen speziell für die Immersion-Kühlung geeignet sein.
- Ein größerer Aufwand ist beim Austausch von Komponenten erforderlich.
- Die eingesetzten großen Mengen der dielektrischen Flüssigkeiten sind hinsichtlich ihrer Umweltfreundlichkeit (Wassergefährdung, GWP-Wert) kritisch zu bewerten.
- Kosten der Flüssigkeit

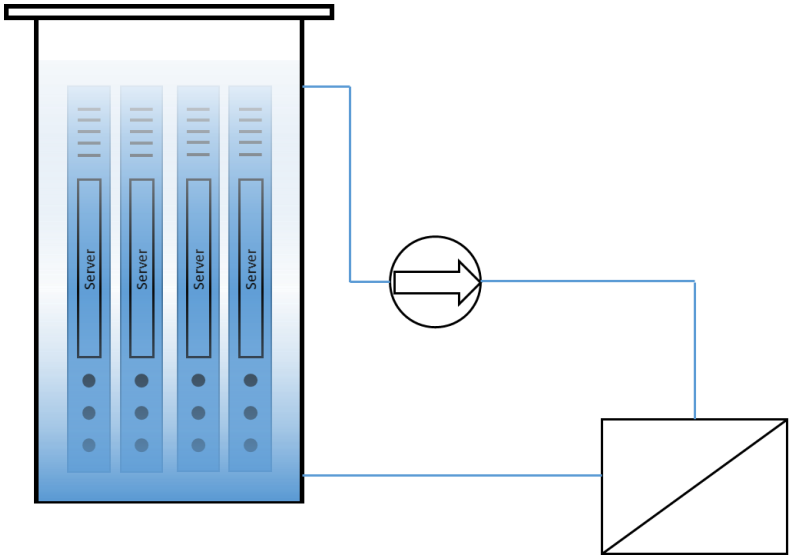

Bild 4: Immersion-Kühlung

Für das Kühlmedium sind folgende Eigenschaften vorteilhaft: hohe spez. Wärmekapazität, umweltfreundlich, Verdampfungstemperatur bei Umgebungsdruck größer 80 °C, elektrisch nicht leitend, preiswert, nicht korrosiv, nicht toxisch, nicht brennbar, hohe chemische und thermische Beständigkeit.
Von 3M™ werden für die Serverkühlung hier im speziellen für die Immersion-Kühlung High-Tech-Flüssigkeiten Novec™ der 7000-Reihe angeboten. Dieses sind elektrisch nicht leitende Flüssigkeiten.

In der Tabelle 1 sind die Fluideigenschaften der High-Tech-Flüssigkeiten NOVEC™ der 7000-Reihe von 3M™ im Vergleich zu Wasser zusammengestellt. Positive Eigenschaften wurden grün und negative Eigenschaften rot dargestellt.

Tabelle 1: High-Tech Flüssigkeit 3M™ Novec™ der 7000-Reihe im Vergleich zu Wasser

	NOVEC™ 7000	NOVEC™ 7100	NOVEC™ 7200	NOVEC™ 7300	Wasser
Chemische Formel		$C_4F_9OCH_3$	$C_4F_9OC_2H_5$	$C_6F_{13}OCH_3$	H_2O
Siedepunk @ 1 bar	34 °C	61 °C	76 °C	98 °C	100 °C
Dichte	1,4 kg/l	1,52 kg/l	1,43 kg/l	1,66 kg/l	1,0 kg/l
Kinematische Viskosität	0,32 mm²/s	0,38 mm²/s	0,43 mm²/s	0,71 mm²/s	1 mm²/s
Dielektrizitätskonstante	7,4	7,39	7,3	6,14	80
Verdampfungswärme	142 kJ/kg	111,6 kJ/kg	119 kJ/kg	101,7 kJ/kg	2.450 kJ/kg
Spez. Wärmekapazität	1,30 kJ/kg K	1,17 kJ/kg K	1,22 kJ/kg K	1,14 kJ/kg K	4,2 kJ/kg K
ODP	0	0	0	0	0
GWP	370	320	59	200	0
	nicht brennbar	nicht brennbar	nicht brennbar	nicht brennbar	nicht brennbar
	kein Gefahrstoff	kein Gefahrstoff	kein Gefahrstoff	kein Gefahrstoff	kein Gefahrstoff
	gering toxisch	gering toxisch	gering toxisch	gering toxisch	nicht toxisch
	elektrisch nicht leitend	elektrisch nicht leitend	elektrisch nicht leitend	elektrisch nicht leitend	elektrisch leitend

Für den Einsatz als Kühlmedium ist die Größe der spezifischen Wärmekapazität von großer Bedeutung. Hier weist Wasser einen deutlich größeren Wert mit einem Faktor von 3,2 bis 3,8 auf. Beim Einsatz eines NOVEC-Fluides müsste somit der Massenstrom um das gleiche Verhältnis angehoben werden. Wasser hat in unserem Einsatzzweck eine einzige negative Eigenschaft und das ist die elektrische Leitfähigkeit. Auch bei der Zweiphasenkühlung schneidet das NOVEC hinsichtlich der Verdampfungsenthalpie mit nur 80 - 140 kJ/kg im Vergleich zu Wasser mit 2.450 kJ/kg deutlich schlechter ab.

Das Problem der elektrischen Leitfähigkeit des Wassers tritt nur auf, wenn es aus dem geschlossenen Kühlkreislauf infolge Lecks austreten würde. Um diese Gefahr auszuschließen, wird das Kühlsystem subatmosphärisch betrieben.

3 Subatmosphärische Kühlung mit Wasser

3.1 Prüfstand für subatmosphärische Kühlung eines Racks

Für die Untersuchungen der subatmosphärischen Kühlung mit Wasser wurde ein Test-Rack mit 4 Servern und 3 zusätzlichen Loadbanks (elektrische Heizer) aufgebaut (Bild 5). Die 4 Server konnten mit 4 Stress-Level betrieben werden (Bild 6). Die Loadbanks mit einer Gesamtleistung von 20 kW haben eine stufenlose Regelung.

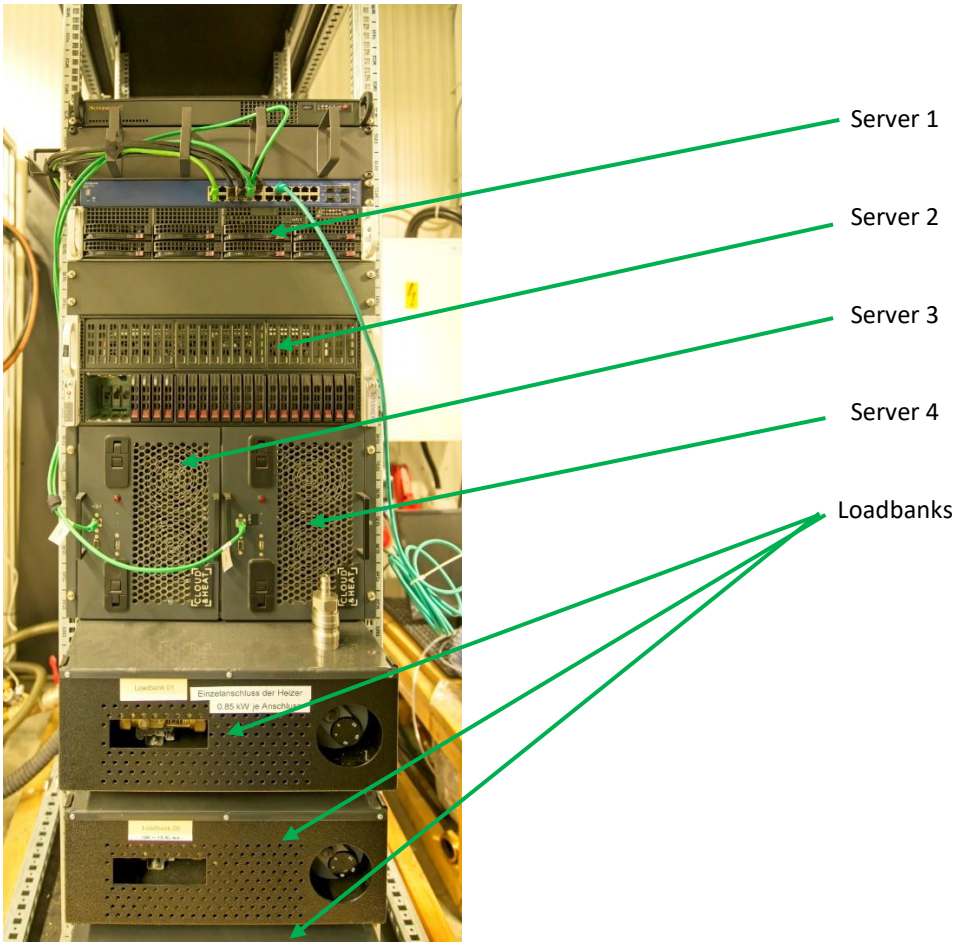

Bild 5: Test-Rack mit Server und Loadbanks

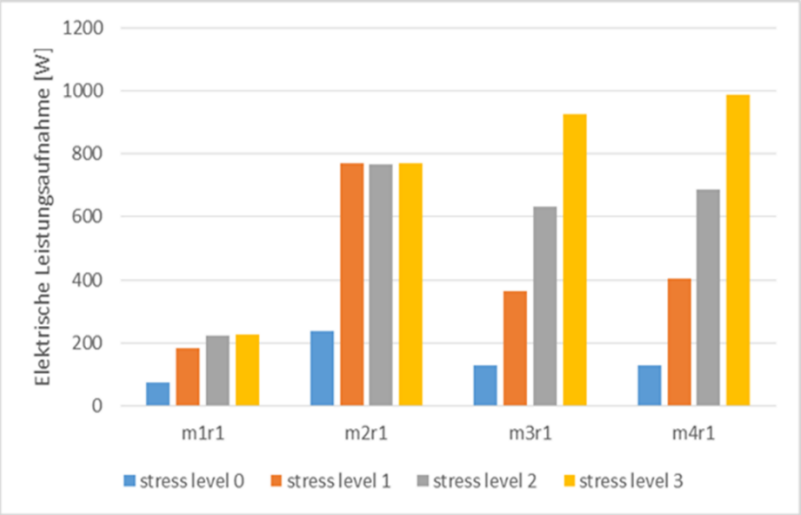

Bild 6: Leistung der Server bei unterschiedlichen Stress-Levels

Die Auslegung der subatmosphärischen Wasserkühlung erfolgte somit für eine Kühlleistung von 25 kW.

3.2 Leckageuntersuchungen

Als erstes wurden Untersuchungen bei unterschiedlich großen Leckagen durchgeführt. Ziel der Untersuchungen war es, nachzuweisen, das bei Lecks kein Wasser aus der Anlage austritt. Hierzu wurden an verschiedenen Stellen des Kreislaufes Lecks unterschiedlicher Größe (Ø 0,5 mm bis Ø 1,5 mm) eingebaut. Als erstes wurde der Druckanstieg mit und ohne Vakuumpumpe (VP) gemessen (Bild 7).

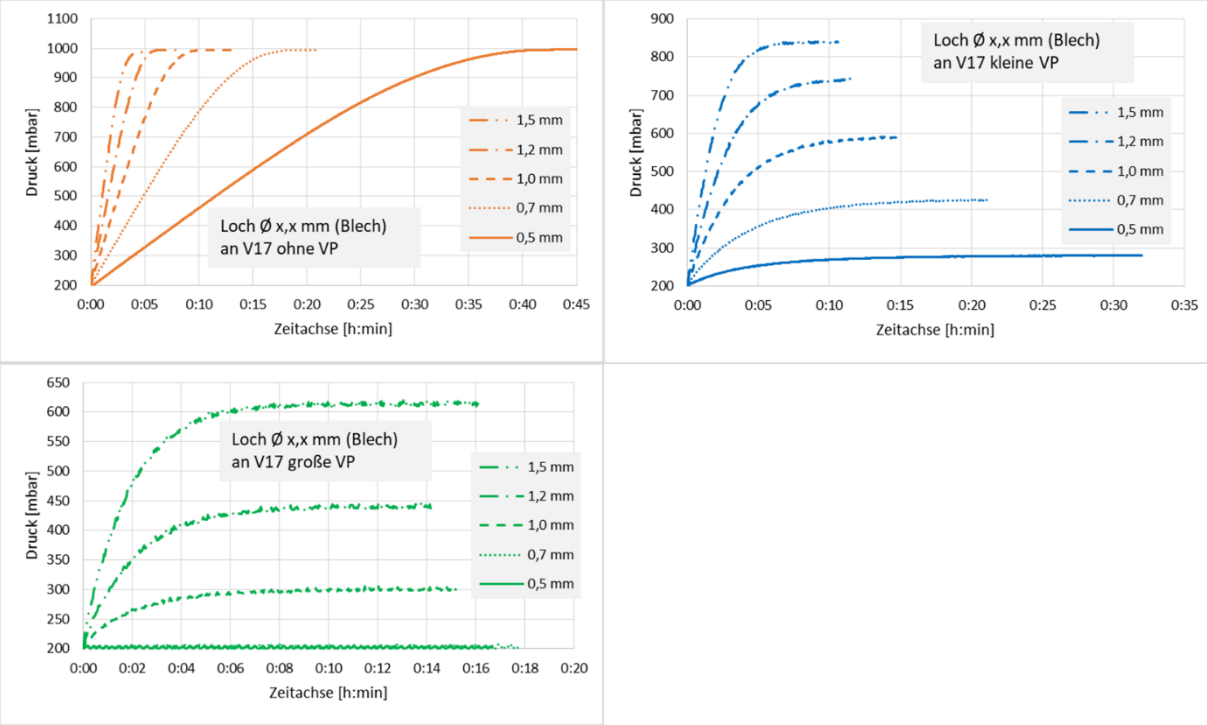

Bild 7: Zeitlicher Druckverlauf – kleine VP (3 l/min @ 400 mbar); große VP (30 l/min @ 400 mbar)

Das Leck V17 befindet sich im Kühlwasservorlauf am Eintritt in das Rack. Jeder Versuch begann mit einem Anfangsabsolutdruck von 200 mbar. Nach weniger als 5 min liegt bei einer Leckgröße größer Ø 1,2 mm bereits ein

Druckausgleich mit der Umgebung vor. Durch Einsatz der Vakuumpumpen wird der Druckanstieg flacher und ein Druckausgleich kann verhindert werden.

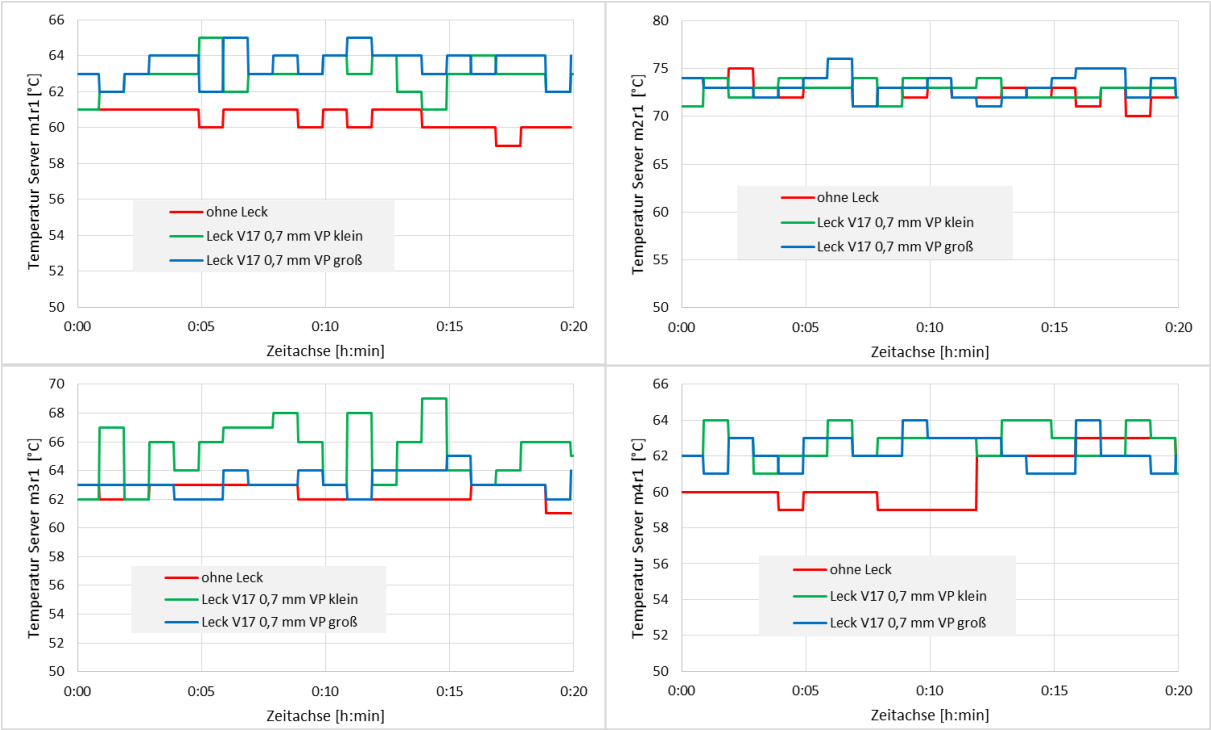

Bild 8: Temperatur Server mit und ohne Leck

Im zweiten Schritt wurde der Einfluss eines Lecks auf die Serverkühlung untersucht. Die Server wurden mit dem Stress-Level 4 betrieben. Bei einem Leck ergibt sich eine geringe Erhöhung der Servertemperatur (außer Server 2), da jetzt neben dem Kühlwasser auch Luft mit unterschiedlicher Verteilung durch die Kühlkörper strömt (Bild 8). Somit ist ein sicherer Betrieb der Server auch bei kleinen Lecks, die durch die Vakuumpumpe ausgeglichen werden können, möglich.

3.3 Einphasenkühlung

Die Leistungsmessungen der subatmosphärischen Kühlung erfolgten bei Absolutdrücken zwischen 200 und 1.000 mbar. Der Unterdruck im System konnte problemlos über die angeschlossene kleine VP eingestellt werden. Die Ergebnisse zeigen, dass der Absolutdruck keinen Einfluss auf die Kühlung der elektronischen Bauteile hat.

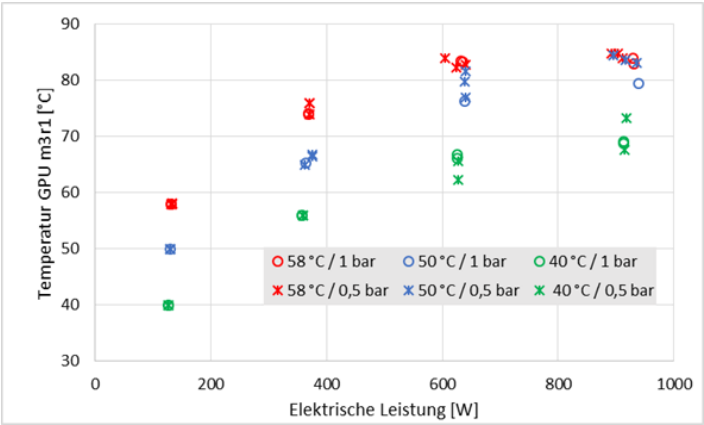

Bild 9: Temperatur GPU Server 3 als Funktion des Anlagendruckes, der Kühlwassertemperatur und der elektrischen Leistungsaufnahme des Servers

3.4 Zweiphasenkühlung

Für die Untersuchung der Zweiphasenkühlung wurden zwei Kühlkörper mit einer elektrischen Heizung (Chip-Simulation) gekoppelt. Die Heizfläche hat eine Größe von 25 cm² und ist zwischen 0 und 500 W stufenlos regelbar. Beide Kühlkörper werden parallel durchströmt (Bild 10).

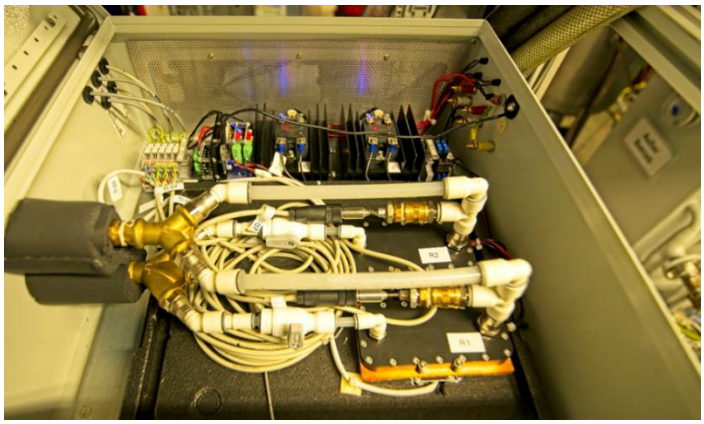

Bild 10: Versuchsaufbau Zweiphasenkühlung – parallele Kühlkörper

Die Betriebsweise der Zweiphasenkühlung mit Wasser soll so erfolgen, das im Kühlkörper eine Teilverdampfung der Flüssigkeit erfolgt, so dass am Austritt ein Flüssigkeits-Dampf-Gemisch vorliegt. Für den Kühlkreislauf wurden zwei Varianten vorgesehen. Im Bild 11 ist der prinzipielle Aufbau einer Zweiphasen-Rackkühlung zu sehen. Eine Flüssigkeitspumpe fördert das Kühlwasser zum Rack. Im Rack wird das Kühlwasser an die einzelnen Server, die mit parallelen und/oder in Reihe geschalteten Chips bestückt sind, verteilt. Bei der Variante 1 strömt das aus dem Kühlkörper kommende Flüssigkeits-Dampf-Gemisch über die Sammelleitung des Racks in einen Kondensator. Dort wird der Dampf kondensiert und die Flüssigkeit ggf. unterkühlt. Bei der Variante 2 erfolgt im Sammler des Racks eine Trennung von Dampf und Flüssigkeit. Der Dampf wird oben am Sammler in den Kondensator abgeführt und die nicht verdampfte Flüssigkeit wird in einen außen liegenden Sammler geführt. In den Sammler läuft dann auch das Kondensat aus dem Kondensator. Anschließend wird die Flüssigkeit in einem Wärmeübertrager abgekühlt und anschließend dem Rack wieder zugeführt. Beide Varianten können mit dem aufgebauten Prüfstand (Bild 10 und Bild 12) getestet werden. Die ersten Ergebnisse sind im Bild 13 und im Bild 14 zusammengefasst. Es wurde die Ein- und Zweiphasenkühlung untersucht.

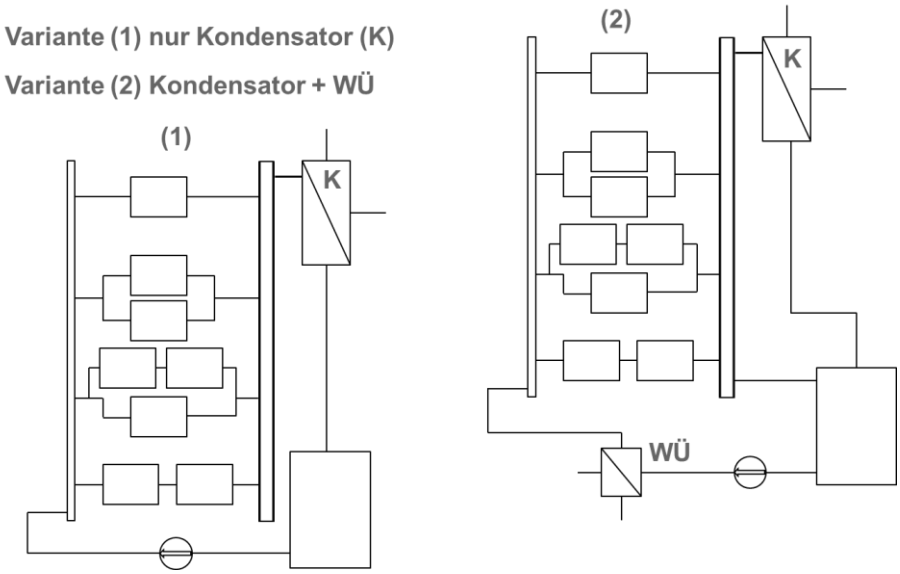

Bild 11: Varianten Kühlkreislauf Zweiphasenkühlung

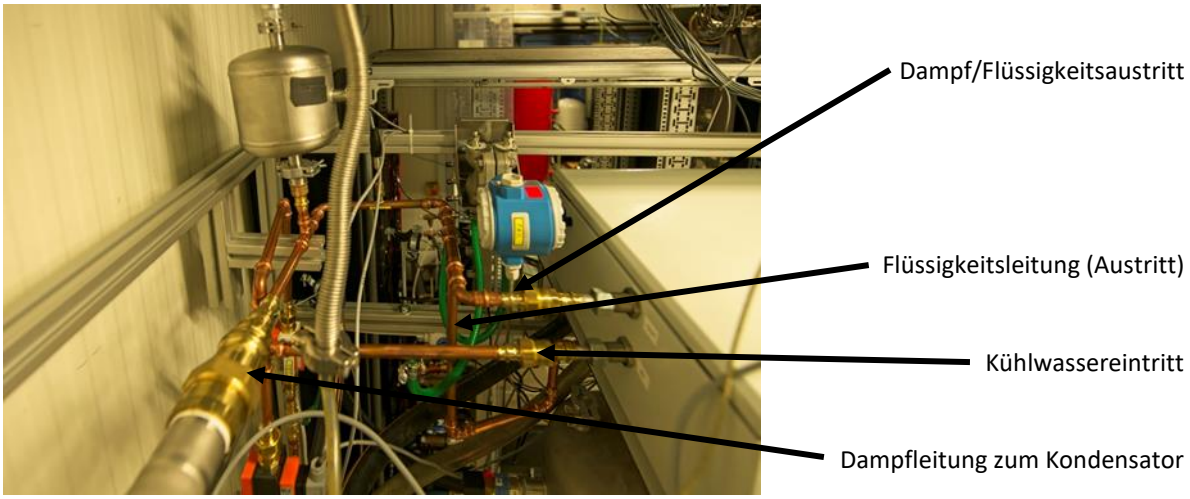

Bild 12: Ausschnitt aus Kühlkreislauf Zweiphasenkühlung

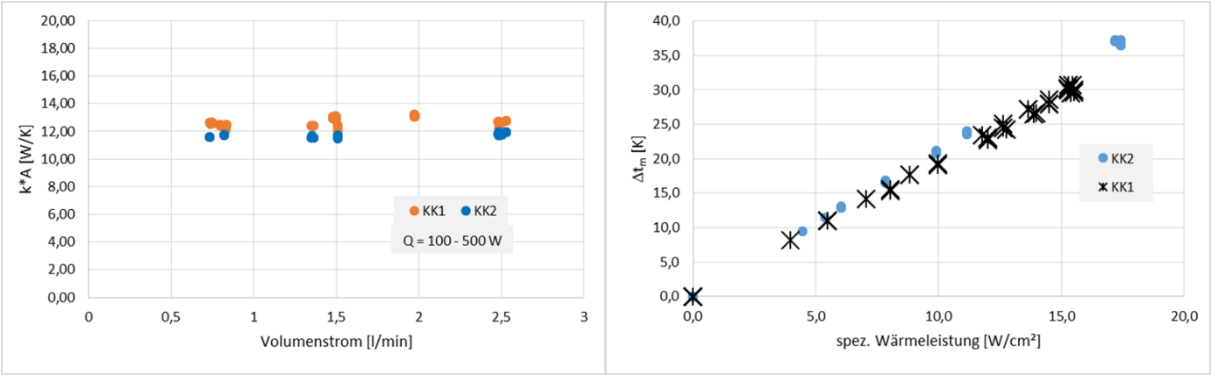

Bild 13: Ergebnisse Einphasenkühlung

Mit dem Prüfstand kann der Kühlprozess als subatmosphärische Ein- und Zweiphasenkühlung sicher realisiert werden. Die Ergebnisse der Einphasenkühlung (Bild 13 links) zeigt überraschender Weise, dass durch eine Variation des Volumenstromes keine Änderungen des k*A-Wertes auftreten. Auch wurden relativ hohe mittlere Temperaturdifferenzen von 30 K bei einer spezifischen Wärmeleistung von 16 W/cm² gemessen. Die mittlere Temperaturdifferenz ist nur abhängig von der spezifischen Wärmeleistung.

$$\dot{Q} = k * A * \Delta t_m$$

$$k * A = \frac{\dot{Q}}{\Delta t_m}$$

$$\Delta t_m = \frac{\left(t_{KüW,A} - t_H\right) - \left(t_{KüW,E} - t_H\right)}{ln\left(\frac{t_{KüW,A} - t_H}{t_{KüW,E} - t_H}\right)}$$

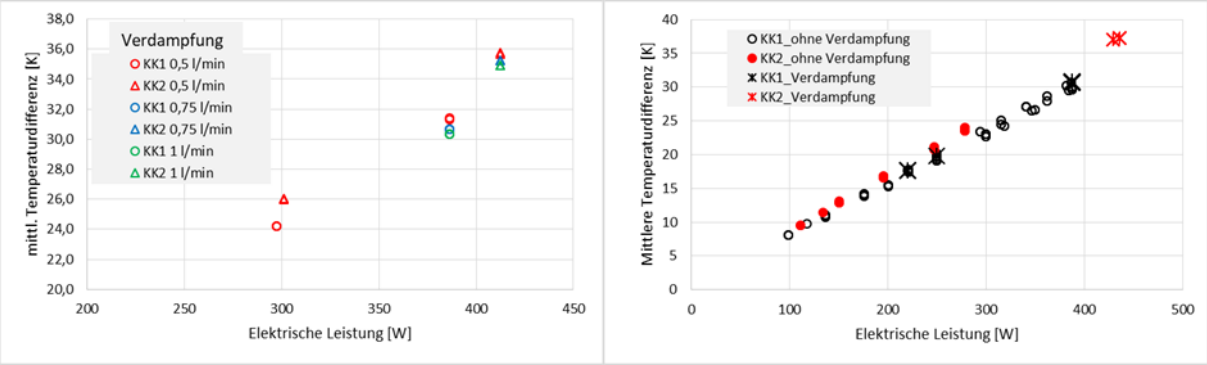

Bild 14: Ergebnisse Zweiphasenkühlung und Vergleich mit Einphasenkühlung

Gleiche Ergebnisse werden bei der Zweiphasenkühlung (Bild 14 links) erzielt. Die mittlere Temperaturdifferenz ist unabhängig vom Kühlwasservolumenstrom. Ein Vergleich von Ein- und Zweiphasenkühlung ist im Bild 14 rechts dargestellt. Wider Erwarten konnten keine Unterschiede in der mittleren Temperaturdifferenz zwischen Ein-und Zweiphasenkühlung ermittelt werden. Nach ersten Erkenntnissen ist dies auf einen großen Kontaktwiderstand zwischen Chip-Simulator und Kühlkörper zurückzuführen. Weitere Untersuchungen sollen diesen Einfluss und das Potenzial der Technologie aufzeichnen.

4 Zusammenfassung

- Flüssigkeitskühlung von Chips im Unterdruck wurde erfolgreich getestet
- Simulierte Lecks führten nicht zum Flüssigkeitsaustritt
- Auch die Zweiphasenkühlung konnte mit dem System sicher realisiert werden
- Eine Verbesserung des Wärmeübergangs durch die Zweiphasenkühlung konnte nicht gemessen werden, da der Wärmewiderstand zwischen Wärmequelle und Kühlkörper zu groß war
- Anpassung und Auslegung der Kühlkörper waren nicht Bestandteil dieser Untersuchungen

5 Danksagung

Das Projekt wird im Rahmen des Förderprogramms des Bundesministeriums für Wirtschaft und Klimaschutz „FuE-Förderung gemeinnütziger externer Industrieforschungseinrichtungen – Innovationskompetenz“ INNO-KOM Modul "Marktorientierte Forschung und Entwicklung“ unter dem Förderkennzeichen 48MF200067 gefördert.

Gefördert durch:

Bundesministerium
für Wirtschaft
und Klimaschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

6 Literaturverzeichnis

[1] Ralph Hintermann u.a.: Bitkom-Studie: Rechenzentren in Deutschland, Aktuelle Marktentwicklungen, Stand 2022; Berlin Borderstep Institut

[2] S. A. Nada u.a.: Experimental investigation of hydrothermal characteristics of data center servers` liquid cooling system for different flow configurations and geometric conditions; Case Studies in Thermal Engineering 27 (2021), 101276

[3] A. Shehabi u.a.: United State Data Center Energy Usage Report, 2016

[4] Ralph Hintermann: Marktentwicklungen und Energieeinsparpotentiale bei Rechenzentren; DKV-Seminar Kühlung von Rechenzentren, 08. – 09. September 2022, Karlsruhe

[5] https://www.pcmasters.de/news/1337424-intel-core-2-quad-qx6700-und-intel-d975xbx2.html; Zugriff am 21.10.2022

[6] Julius Neudorfer: High Density IT Cooling Breaking the Thermal Boundaries; Data Center Frontier Special Report, 2022

[7] L. Baroso, U. Hozle: Datacenter as a Computer: An introduction to the design of warehouse-scale machines, in Synthesis Lectures on Computer Architecture, ed. M. Hill. The Vol. 6, (Morgan & Claypool, U.S., 2009)

[8] Wahlroos, M et al: Utilizing data center waste heat in district heating – Impacts on energy efficiency and prospects for low-temperature district heating networks, Energy, Elsevier BV, 2017, Volume 140, Seiten 1228-1238

[9] M. Saini, R. L. Webb: Heat rejection limits of air-cooled plane fin heat sinks for computer cooling; IEEE Transaction on Components and Packaging Technologies 26 (1) (2003)71-79

Nomenklatur

Formelzeichen

A	Wärmeübertragungsfläche [m²]
k	Wärmedurchgangskoeffizient [W/m²K]
$\dot{Q}$	Wärmeleistung [W]
t	Temperatur [°C]
Δ	Differenz

Indizes

A	Austritt
E	Eintritt
H	Heizer
m	mittlere
KüW	Kühlwasser