

2024年7月19日

「データセンター冷却手法を革新する「水浸」冷却技術の確立と、 ITRI と ZYRQ による次世代機の共同開発実施について」

概要：株式会社 ZYRQ（以下、ZYRQ と言う）は、主たる冷媒として一般的な「水」のみを使用しながら既存の冷却技術に対する多くの優位性を有する、新しい「水浸」冷却技術を確立していました。この水浸冷却技術が有効に機能して非常に高い冷却性能を有することを、台湾最大の産業技術研究機構であり、台湾半導体業界のパイオニアでもある工業技術研究院（以下、ITRI と言う）が、3回に渡る ZYRQ 本社の実験サイトを訪問して実施した共同計測実験により、公式に確認を致しました。

この結果を受け、その非常に高い冷却性能を更に高めることを目的として、ITRI は、熱伝導効率を高めることが可能である開発済みのマイクロチャネル技術を最適化して ZYRQ に提供することとし、両者の共同開発によって、最新となる第 6 世代水浸冷却システムを開発することに合意し、年内の同システムの完成を目指すことと致しました。

背景：データセンターの世界市場では、全世界規模でのネット社会への移行によるデータセンター需要の増加に加えて、地球温暖化による影響も重なり、特に夏季のデータセンター稼働を維持するための冷却能力不足の問題が深刻となっております。

欧州では、ロシア・ウクライナ戦争の影響から電力コストが急騰し、十分な電力供給の確保にも困難が生じる中で、消費電力を低減する必要性も生じておりました。

OpenAI 社が GPT-4 をリリースした 2023 年 3 月を境に、生成 AI 処理を行うための AI データセンター需要が更に急増してきており、データセンターは「冷却能力」の不足、「電力」の不足、その双方を緩和するために必要となる「水」資源の不足という、三重苦に見舞われている状況です。

特に、AI データセンターで大量に使用される生成 AI 処理用の GPGPU（汎用グラフィックス処理ユニット）は、これまでのデータセンターで使用されてきた CPU が 100-200W 程度、GPU でも 300W 程度しか電力を消費して発熱しなかったものが、現時点の最高性能機種で 700W、既に発表済みで年内に納品が開始される次世代主力機種では 1,000W を超える電力消費と発熱となることが確定しており、この三重苦が更に深刻な問題となっていくことが明らかです。

市場レポートや最近の報道によれば、2030 年から 2035 年には、先進国の電力需要に占めるデータセンターの電力消費は 8-20% にも達するとされます。しかし、それだけの電力を供する発電所も、系統連系を担う電力網も、冷却のために必要な水資源の何れも、十分に確保される見通しが立っていない状況で、そのための本質的な解決策が求められているところです。

既存技術：データセンターの冷却方式としては、長い期間、空冷方式が採用されてきており、特に寒冷地

域においては、自然冷却により外部の冷たい空気を取り込んでデータセンター施設内の空気と熱交換することで、冷却コストを抑えながら必要な冷却能力を得ることが可能でありました。

しかしながら近年、データセンターが大型化し、高性能化する中で、一部のデータセンターでは外気による自然冷却や、室外冷凍機による強制冷却に加えて、冷却塔（クーリングタワー）を採用して湖沼水や河川水、地下水を活用して、その蒸発時の気化熱によって冷却を行う事例も増えています。特に米国南部のテキサス州などでは、大規模データセンターが地域の水資源を大量に冷却に使用することで、深刻な環境問題にまで発展している事態を報じるニュースも目にする様になりました。

生成 AI の勃興は、更に強力な冷却方式を採用する大規模データセンターを出現させており、今年に入って古い歴史を持つ「液冷」方式を採用したデータセンター用サーバー製品を投入するベンダーが増えてきています。この液冷方式は、既存の空冷サーバーの基本的な構造は維持したまま、CPU や GPGPU の直上にヒートシンクに代わってコールドプレートを設置し、そこに分岐配管（マニフォールド）により冷却水を供給することで、空気とヒートシンクの組み合わせよりも遥かに高い冷却能力を提供することが可能です。しかしながら、冷媒である水を供給する配管構造には直径が 10mm に満たないチューブが使用されており、そこに流せる水の流量には限界があります。また分岐配管となっていることから、末端の 1 個 1 個の GPGPU に提供される水流量は大きく限定されてしまっています。更に、分岐配管構造は、コールドプレートを有さない、空気冷却を行う必要がある GPGPU 以外の多数の電子部品に対しての風流を阻害する場合があります。GPGPU 以外の部分の冷却性能を低下させる問題点があります。

冷媒に、空気ではなく液体を使用する冷却方式としては、他に「液浸」方式が挙げられます。こちらは、電気を通さない合成油やシリコンオイル、フッ化炭素化合物 (PFAS) 液体に、電子基板全体を浸漬させ、その液体を循環させて冷却するか、低沸点で沸騰蒸発させて気化熱により局所の冷却性能を得るかに大別されます。しかし後者については、低沸点冷媒である PFAS 液体の使用が原則的に禁止されることと、サーバーラックに代わる液浸槽には高度な密閉構造が求められることから、冷却構造が複雑化、大型化してしまうために、今後、大規模データセンターで使用されるケースは減少していくものと思われます。前者の合成油やシリコンオイルを使用する方式では、コールドプレートや分岐配管構造を設ける必要がないために、原理的には大量の冷媒を流して冷却することができます。しかしながら、基板上の電子部品がショートして破損することを避けるために、最も冷却性能が高い冷媒である「水」を使用することは叶わず、冷却性能が水よりも大きく劣って、高い粘性で保守が非常に困難となり、部品の再利用にも支障を来す合成油やシリコンオイルを使用する必要があり、こちらも本格的な普及には困難が伴います。またそうした合成油やシリコンオイルは冷媒としては非常に高価でもあり、結果としては長い歴史を有する「液冷」方式が、現時点では AI データセンターの次の冷却方式の主流となるものと考えられてきました。

新技術の必要性：世界のデータセンター市場は、生成 AI の出現と発展により、そしてそれを支えるための非常に強力な GPGPU の普及によって、多くの問題に直面しています。冷却能力不足、電力不足、水不足の三重苦の他にも、液冷方式や液浸方式を導入するためのコストの高さや構造の複雑さでも、大きな支障を来しています。そして、現在のままでは先進国の電力供給能力の制約は、生成 AI の普及を阻害してしまい、世界の進歩と発展に必要となる計算能力をデータセンターから提供できないことが明らかです。その上に、次世代の GPGPU では、空冷方式でも、液冷方式でも、既存の液浸方式であっても、その冷却を十分には行えない可能性もあり、全く新しい冷却方式が必要な状況が生じています。

その新しい冷却技術は、1) これまでになく強力な冷却性能を有して、2) 十分低温に GPGPU を冷却することで消費電力を大きく低減し、3) 貴重な水資源を蒸発させて失うことがなく、4) 人体にも自

然環境にも完全に無害、無負荷であり、5) 初期設備投資費用と運用保守費用の双方が安価で、6) 短期間にその利用を開始することが可能であること、が強く求められます。

新技術の内容：上述の背景と新技術の必要性を鑑みて、2014 年以来、国内における液浸冷却技術の先駆者として世界的実績を多数有する ZYRQ 研究員の齊藤は、2023 年 2 月から単独での研究開発を敢行し、400 種類以上の試行による基礎実験を繰り返した結果として、これまでは不可能であると思われていた、超純水や純水といった特殊な水ではない水道水などの一般の「水」を主冷媒としながら、既存のどの冷却方式と比較しても非常に強力な冷却性能を、安価かつ最小限の構成で実現可能な全く新しい「水浸」冷却方式を確立するに至りました。これまでの液浸冷却方式の優位性の殆どを引き継ぎつつ、液浸冷却方式の問題点の全てを払拭した、その新技術の内容は以下の通りです。

- 1) 水浸槽内を内循環させる主冷媒としては、安全で安価に利用可能な一般的な水道水のみを利用する
- 2) 水浸槽毎に設置される熱交換器を介して内循環する水に熱伝導させるために、水浸槽外部で外循環させる二次冷媒としては、自然水である地下水、河川水、湖沼水、海水などを利用可能であるが、そうした自然水は、気化させることなく、僅か数度の温度上昇のみで自然環境に戻されることで水量が失われることがない。また、そうした自然水を二次冷媒として利用できない場合には、室外冷凍機により冷却される非フロン性冷媒などを使用することも可能である
- 3) 内循環させる主冷媒に GPGPU などを実装した電子基板を液浸させる際、高い導電性を有する水に浸漬させるため、この水は GPGPU の半導体表面には専用開発した熱伝導構造体を介して直接的に接触させながらも、その他の部位や電子部品については電気伝導性がなく、一定の熱伝導性と耐久性を有した柔軟なフィルムで被覆を行うことにより、電氣的短絡（ショート）を完全に回避している
- 4) 熱伝導構造体とフィルムの接合部は、長期使用によって接合部の水密が破れて、フィルム内部に水が浸入してしまうと、高額な GPGPU を含む電子基板全体が電氣的短絡により不可逆に破損してしまうことから、水密性能を長期間担保することが可能な新しい構造を考案して、実用化している
- 5) 単なる液浸冷却方式では、冷媒として最も強力な水を使用したとしても、1,000W もの発熱を生ずる GPGPU から絶えず十分な抜熱を行うことは不可能であり、GPGPU 温度が半導体として動作可能な上限値である 105°C を短時間で上回ってしまうことから、強力な抜熱を実現する手法として、水浸槽内には潤沢に存在している主冷媒の水を、水浸槽毎に奢られた個別の CDU（冷媒配分ユニット）で高速に内循環させた上で、更に GPGPU 毎に局所強制対流を強力に与えることで、既存技術では得ることが叶わなかった極めて高い抜熱能力が得られ、その結果、GPGPU 温度を 25°C にまで低下させることに成功している
- 6) GPGPU 温度を 25°C にまで低下させて安定稼働を行えることにより、GPGPU の種類によっては、その消費電力を半分以下に低減させることが可能となる一方で、水浸冷却方式を機能させるために必要な電力は、外循環用、内循環用のそれぞれのポンプ駆動のために使用される、相対的には非常に小さい電力だけであり、システム全体の消費電力を大きく削減可能である
- 7) 第 5 世代までの開発を行うことで、水浸冷却構造を突き詰めることが可能となり、水浸槽の内循環を水のみとすることに成功して大幅な構造の簡素化を実現したことで、水自体が本来持つ非常に強力な冷却能力を最小限の体積で引き出すことに繋がり、既存方式では到達不可能な非常に高い冷却熱容量密度を達成している
- 8) 上記 1) から 7) までは、極めて簡素で安価なシステム構成によって実現することに成功しており、必要に応じて、短期間で大量の大規模、及び小規模システムの立上げが可能であると考えられる

上記の一連の水浸冷却方式に関して、ZYRQ 研究員である齊藤によって発明された一連の新技术については、齊藤を発明者、ZYRQ を出願人とする 8 件の PCT 特許出願により約 200 項目の請求項目でカバーされています。

出願済みである 8 件の特許の内、6 件の PCT 特許出願については、既に日本の特許庁による評価書が受領しており、それら 6 件による合計 156 項目の請求項のうち、少なくとも 141 項目については新規性、進歩性、実用化可能性が有り、特許性を有するとの判断が得られています。これらは今後、順次、各国における特許国内移行手続きを進めて参ります。

新技术の優位性：既存技術として最も有力とされている液冷方式との比較において、水浸冷却方式は以下の優位性を有します。

- 1) 液冷方式を構成するために必要なサーバー内部での複雑な分岐配管（マニフォールド）構造が不要であり、配管の直径の細さに起因する冷媒流量の制約も実用上は存在しない
- 2) 液冷方式ではサーバー内部での分岐配管とは別に、専用サーバーラック内部でのサーバーへの分岐配管と、CDU から専用サーバーラックへの分岐配管も存在しているところ、液浸槽毎の個別 CDU を配した結果、CDU からの分岐配管が一切存在しない単純で効率的な流水構造を実現している
- 3) 液冷方式では不可欠であるサーバー内部での空冷部分を完全に排しており、GPGPU 以外の電子部品は水冷による熱伝導方式により冷却されている
- 4) 液冷方式では不可欠であるサーバー内部での空冷部分を完全に排したことで、液冷方式では必要であるサーバーラック前後の空冷用の空気の流路スペースを大きく確保しなくてもよい
- 5) 非常に小さい体積で、非常に大きな熱量を冷却することが可能であり、CDU 部を除いた場合の体積当たりの冷却熱容量密度の比較では、最新製品であってもサーバーラックの前後に空冷用の空気の流路スペースが必要である液冷方式に対して、現時点でも約 10 倍もの高い冷却能力を有している
- 6) 1,000W 級の GPGPU の冷却運用温度として、液冷方式では到達不可能な 25°C を実現可能である

ITRI が確認した客観的事実：ITRI は、2024 年 2 月 19 - 21 日、5 月 11 - 13 日、7 月 16 - 17 日の 3 回に渡って ZYRQ 本社実験サイトへの複数の専門技術員派遣を行って、現地での共同計測実験を施行し、3 編の公式な評価実験報告書を作成して以下の内容について、客観的な事実を確認致しました。

- 1) ZYRQ が開発した革新的な「水浸」冷却方式の、第 4 世代システム、第 5 世代第 1 システム、同第 2 システムが、ZYRQ が意図した形で安定して稼働して冷却が行えており、これまでの既存技術では得られなかった、非常に強力な冷却性能を有していること
- 2) 特に、CDU（冷媒分配ユニット）の構成部材を全て含めた上で、164,724cm³（立法センチメートル）または 0.165m³（立法メートル）の非常に小さい体積にも関わらず、33kW もの消費電力からの発熱を強力に冷却しきっており、これは単位体積当たりの冷却熱容量密度の数値としては 200kW/m³ または 0.2MW/m³（1 立方メートル当たり 200kW、または 0.2MW）という、通常空冷方式（2-5kW/m³）や最新の液冷方式（30-40kW/m³）（但し、空冷のための空気流路体積を含む）との比較においても、遥かに高い数値を達成していること
- 3) その上で、GPGPU 毎の熱抵抗は平均値で RTH=0.02 という低い値を実現していることで、半導体温度が 25°C という、既存技術では実際の AI データセンター内の環境において得ることが不可能と考えられる低温において安定して保持されること

4) GPGPU の半導体温度が 25°C という低温で稼働する状況を、実際の GPGPU を用いて 85°C というその他の既存技術で得られる通常の冷却温度での稼働時と比較した場合に、特定の GPGPU の消費電力が半減している事例を確認したこと

共同開発の実施：ITRI と ZYRQ は、これまでの 3 回に渡る共同計測実験を通して、相互の技術レベルについて理解を深め、特に水浸冷却技術についての深い認識を共有するに至りました。ZYRQ が開発して、今回、ITRI がその優位性を確認した新しい水浸冷却方式によるシステムに関しては、現在、鋭意開発を進めている第 6 世代以降の冷却システムで、ITRI がこれまでに開発して保有しているマイクロチャネル構造による高性能熱伝導技術を、本水浸冷却システム用に最適化して ZYRQ に提供して、GPGPU 毎の熱抵抗の平均値である $R_{th}=0.02$ を更に低い値として GPGPU 温度を更に低下させる目的で、ZYRQ がこれを水浸冷却システムに採用することに二者が合意を致しました。この合意に基づいて、既に 2024 年 6 月より、具体的な共同開発作業を開始しております。

今後の展望：ITRI と ZYRQ の共同開発により、2024 年中に水浸冷却技術による第 6 世代の「システム 1」、および「システム 2」の 2 種類を完成させる予定です。冷却熱容量密度は現時点の最新版である第 5 世代のシステム 2 で達成された $200\text{kW}/\text{m}^3$ に対して、それぞれ $300\text{kW}/\text{m}^3$ および $400\text{kW}/\text{m}^3$ という一層高い数値にまで達する計画としています。ITRI からの技術提供を受けて共同開発される第 6 世代の水浸冷却は、その優位性を更に明確なものとして、生成 AI 処理を、余裕をもって行うことができる大規模および小規模双方の AI データセンターの構築に不可欠なものとするべく、引き続き開発を弛まらずに推進して参ります。

量産化に向けては、特に長期安定運用に不可欠となる水密加工技術について、日本国内の専門メーカーからの協力も得ながら、最適な材料開発と、加速試験による長期信頼性の確保を得て参る予定です。

現在、世界の半導体市場では、生成 AI 処理用の GPGPU の需給が逼迫しており、市場が要求する必要な数量が十分に供給されずに、世界的な争奪戦が繰り広げられている状況にあります。この原因としては、GPGPU を構成するために必要である、GPGPU1 個当たり 4-8 個の HBM メモリの供給が不足していることと、GPGPU のロジック半導体と HBM メモリ半導体を一つのパッケージに封止して、 105°C の高温環境でも、その接続が全く破断せずに GPGPU が安定稼働するために不可欠な、液冷方式の最先端の 2.5 次元パッケージング工程が不足していることが挙げられます。今回開発された水浸冷却技術によって、 105°C の高熱環境において、シリコンと有機サブストレート基板材料間の線膨張率の違いから生ずる接続破断の問題が発生しなくなるため、この問題を解決するために必要となっているシリコンインタポザーを使用する最先端の 2.5 次元パッケージング工程ではなく、通常の 2.5 次元パッケージング工程で十分となる可能性があります。その上で、シリコンインタポザーを使用する最先端の 2.5 次元パッケージングによる面積制約から解放されることから、8 - 16 層の 3 次元積層構造である HBM メモリを、非常に小さい面積で実現する必要性から生ずる、HBM メモリの製造歩留まり問題や、最先端 DRAM プロセスを使用する HBM メモリ製造に拘る必要がなくなり、HBM メモリの供給量が大幅に増える可能性もあります。それらの結果としては、現在、世界中で不足している GPGPU の製造上の制約がなくなって、潤沢に、そして安価に GPGPU を提供することも可能になるものと考えられます。この仮説を検証するために、本水浸冷却方式と組み合わせて使用することが可能である、最先端 DRAM プロセス以外で製造する、現在のものよりは大きなサイズの HBM メモリと、標準的な 2.5 次元パッケージング工程を使用して製造される GPGPU を準備して、その安定動作を確認して参る計画です。

また、1,000W 級の GPGPU が 25°C の低温で安定稼働が可能になると、GPGPU の半導体設計が革新される可能性があります。論理演算を行う半導体は 53 年超の歴史を有しますが、その間、半導体設計手法は常に、25°C 条件と 105°C（または 100°C）条件の 2 種類の条件において設計開発と検証作業が行われてきました。特に条件の厳しい高温側の 105°C においても半導体を問題なく動作させるために、膨大な設計工数が割かれ、また半導体面積が大きくなる原因にもなっていました。今回、初めて実現された水浸冷却方式が普及して、AI データセンターにおいては GPGPU を 25°C 条件で稼働させることが可能となった暁には、53 年間の半導体設計が一新されて、25°C 条件のみでの設計開発が確立される可能性があります。その場合、半導体設計期間の圧縮、半導体面積の縮小、半導体消費電力の削減、半導体駆動速度の向上などの副次的な、しかし大きなメリットを享受できることが期待されるため、今後、半導体設計開発のフロントエンドベンダー、およびバックエンドベンダーと共同して、この重要な仮説の検証についても進めて参る所存です。

問い合わせ先：

台湾工業技術研究院（ITRI）

日本事務所：

〒108-0073 東京都港区三田 1 - 2 - 1 8 TTD ビル 3 階

Tel.: 03-5419-3836 E-mail: itritokyo@itri.org.tw

株式会社 ZYRQ：

〒940-0061 新潟県長岡市城内町 3 - 4 - 1 4

Tel.: 0258-84-7516 E-mail: info@zyrq-dc.com

報告書

本報告書は、株式会社 ZYRQ（以下、「当社」といいます。）の求めに応じて、弊所での特許申請状況について概要を報告するものです。

当社から弊所に対し、2023年2月に新規に国際特許出願（以下「PCT 出願」といいます。）の依頼があり、弊所はこれを受任して、2023年4月4日以降現在までに8件のPCT出願の手続きを代理しています。

なお、弊所の簡易な調査によると、コンピュータ、サーバー等の電子機器又は半導体装置の液浸冷却に関する特許数は、日本では1990年1月以降に公告または登録されたものが127件確認されました。そのうち弊所が出願代理をして登録された特許の割合は約20%（25/127）に当たります（2024年7月調査時）。

現在、当社の依頼に基づく8件のPCT出願を日本国特許庁に提出済みですが、そのうちの6件の出願済みPCT出願に関し受領した国際調査機関の見解書では、5件の出願の請求項のすべてと、残る1件の出願の請求項の一部について、新規性・進歩性・産業上利用可能性有の見解が示されています。

よって、出願済みの6件について、将来、日本国その他の各国国内段階に移行後の、各国特許庁における審査において特許性が認められ、特許される可能性が十分にあると考えます。

残り2件については、国際調査機関の見解書が未着の段階です。

弊所の過去のコンピュータ機器の液浸冷却関連の特許事例を確認したところ、国際調査機関の見解書において日本国特許庁の審査官が出願発明に対して特許性（新規性、進歩性、産業上の利用可能性）有りとして認めた出願については、その後日本において100%（一旦拒絶理由通知を受けて特許請求の範囲を補正した場合を含みます。）が特許査定されています。

本書は、弊職が出願代理人の立場から、出願特許についての予備的評価を述べていますが、日本国特許庁または他の国の特許庁が同一の見解をとることを保証するものではありませんので、この点特にご留意ください。また、本書は、投資または事業判断の根拠とされることは意図しておらず、そのようなものと解釈されてはなりません。

本書では、本書に記載された事実以外の事実や事情は考慮されていません。

また、本書の目的外使用、無断転載、複製、開示は固く禁じます。

以上

弁護士法人黒田法律事務所
弁理士 松本 孝