



Patrocinado por



El Futuro del enfriamiento Suplemento



La preparación para un centro de datos de alta densidad

CONTENIDO

¿Qué hacer con el calor?

> Cada vatio de potencia del centro de datos debe usarse dos veces

Dentro de los microfluidos

> El enfriamiento dentro de un chip es mayor de lo que se cree

Aire acondicionado de precisión

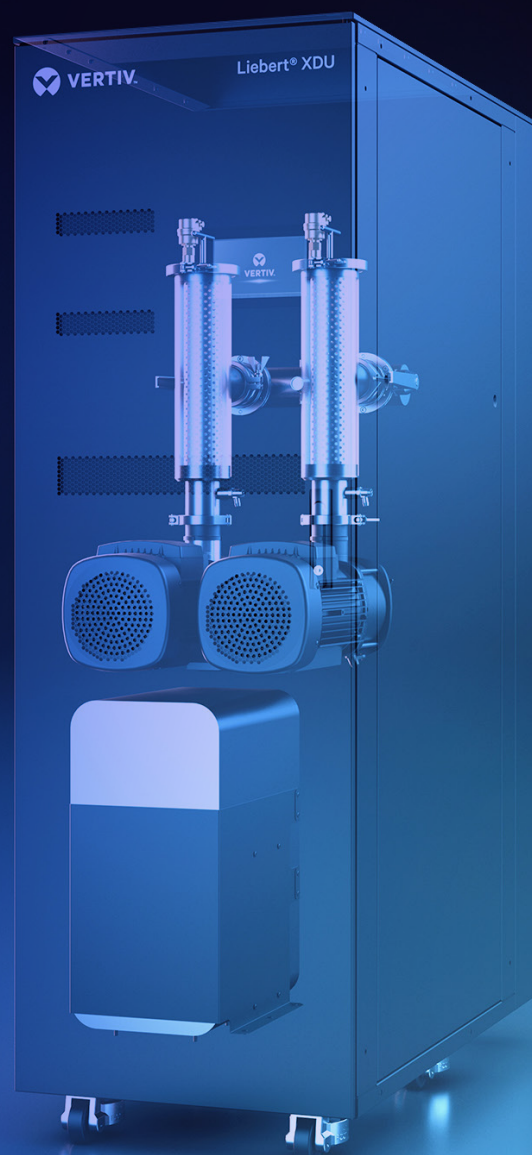
> Miles de diminutos chorros podrían llegar hasta un centro de datos cercano



Diseñado para
la densidad.

**Listo para la
aceleración.**

Gestione un mayor calor con la
tecnología de enfriamiento líquido.



Descubra cómo podemos ayudarlo en [Vertiv.com/GetCool](https://www.vertiv.com/GetCool)



Índice

4. La combinación de calor y computación

Cada vatio de potencia del centro de datos debe usarse dos veces

8. Publireportaje: Cuando menos es más

Los operadores de centros de datos necesitarán adoptar estrategias de enfriamiento híbrido para gestionar las crecientes densidades de TI. ¿Qué significa esto?

10. Los microfluidos: El enfriamiento dentro de los chips

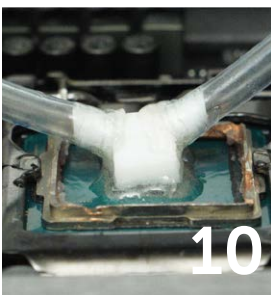
Si usted piensa que la inmersión es la fase final del enfriamiento líquido, piénselo de nuevo.

13. El enfriamiento de precisión al nivel del chip

De qué manera miles de diminutos chorros podrían llegar hasta un centro de datos cercano

15. ¿Por qué está interesado Intel en el enfriamiento?

¿Cómo enfriar procesadores de 2 kW?



La física del enfriamiento

La termodinámica nos encuentra tarde o temprano, y no se puede negociar con esa ley. En la actualidad, los centros de datos usan una enorme cantidad de energía y generan un nivel de calor que afecta su entorno.

Para hacer frente a esto, se debe encontrar la manera más creativa de trabajar dentro de las reglas decretadas por la física.

Esto tiene implicaciones en el macronivel, tanto para todo el edificio como el espacio circundante, y abre posibilidades en el micronivel, hasta los procesadores, e incluso dentro de ellos. Este suplemento muestra cómo los ingenieros pueden encontrar el camino hacia la próxima generación de tecnología para centros de datos.

Usar todo dos veces

Toda la energía utilizada en los sistemas de TI se transforma en calor. El calor es el único resultado inevitable, pero es un bien valioso.

Entonces, ¿qué sucede si cambiamos nuestra visión y vemos los centros de datos como una fuente de calor que también ofrece computación como subproducto?

Este pensamiento lleva a la reutilización térmica, lo cual plantea la posibilidad de sistemas de calefacción urbana, donde los centros de datos canalizan el aire y el agua calientes hasta sus vecinos.

Si fuésemos capaces de asegurarnos de que toda nuestra energía informática haga dos trabajos, podríamos contar con toda la energía necesaria, pero hay más de una forma de lograrlo.

En el interior de los chips

A medida que los chips se vuelven más complejos e incluyen más

transistores, enfriarlos resulta cada vez más difícil. Desde hace años, los rápidos y calientes chips han sido enfriados con disipadores térmicos especializados, los cuales suelen ser enfriados con líquidos, pero ¿qué pasaría si pudiéramos canalizar los fluidos más cerca de los transistores, dentro de los chips?

Pues bien, los procesos que graban e imprimen pequeños componentes electrónicos sobre un sustrato pueden adaptarse para crear canales microscópicos que llevan el refrigerante hasta el corazón del procesador.

Las estrategias hacia Coolerchips

La comunidad del silicio está buscando otras técnicas de enfriamiento menos invasivas, pero igualmente radicales, para crear chips con puntos de diseño térmico (TDP) cada vez más altos.

¿Qué tan serio es esto? El programa COOLERCHIPS del gobierno de EE. UU. incorpora múltiples enfoques para sistemas de enfriamiento de este nivel, tales como JetCool, el cual dispara chorros de refrigerante dentro de sistemas de enfriamiento directo al chip.

El programa DOE también ha reclutado a Intel. El gigante de silicio se ha dado cuenta de que sus planes para enviar silicio cada vez más potente chocarían contra una pared sin una manera efectiva de enfriar los nuevos procesadores.

Intel combina disipadores térmicos generados por computadora y nuevos materiales.

Todos están buscando técnicas radicales para aprovechar el potencial de crecimiento continuo.

La combinación de calor y computación



Peter Judge
Editor ejecutivo

Cada vatio de potencia del centro de datos debe usarse dos veces

La termodinámica afirma que todo tipo de energía eventualmente se convertirá en energía térmica. Si se utiliza 1 kWh de energía eléctrica para hacer funcionar una computadora, esta emitirá 1 kWh de calor.

Astrid Wynne de TechBuyer lo explicó de la siguiente manera en un artículo publicado en 2022: "La primera ley de la termodinámica, también conocida como ley

de conservación de la energía, establece que la energía no puede crearse ni destruirse, solo convertirse de una forma de energía a otra. La energía eléctrica consumida por los centros de datos a través de los servidores se convierte casi en su totalidad en calor. La generación de datos es un subproducto de esto, pero representa menos del 0,1 % de la conversión".

Para el ingeniero del centro de datos, esto supone un problema, ya que si el calor no se elimina, el equipo se sobrecalentará.

Por otro lado, para el planificador urbanístico, esto presenta una oportunidad; la comunidad quiere ese calor.

Usar la energía dos veces

De acuerdo con Wynne, "El término 'centro de datos' es inapropiado desde un punto de vista energético. Sería más preciso llamarlos 'centros de calor' que también pueden usarse para procesar y almacenar información. Un cambio en este pensamiento nos



permite ser más creativos con respecto a cómo planificamos y gestionamos las construcciones de centros de datos hoy y en el futuro”.

Las comunidades necesitan el calor, y la calefacción genera muchas emisiones en la actualidad. Por ejemplo, en el Reino Unido la calefacción representa cerca del 37 % de las emisiones totales. Parte de estas emisiones provienen del calor industrial de alta temperatura, pero más de la mitad procede de la calefacción y el agua caliente.

Justo ahora, la mayor parte de ese calor proviene directamente de los combustibles fósiles o de la electricidad, gran parte de la cual también proviene de combustibles fósiles.

Con el objetivo de alcanzar las cero emisiones netas, la calefacción debe descarbonizarse; por esta razón, los gobiernos buscan eliminar gradualmente los calentadores de gas y sustituirlos por fuentes de calor alternativas con bajas emisiones de carbono. Las bombas eléctricas de calor geotérmicas son una de estas fuentes, pero necesitan funcionar con energía renovable.

El calor producido por la computación es efectivamente libre de carbono, porque es un subproducto de la energía utilizada en otros lugares.

Además, si el calor se utiliza en otros lugares, puede decirse que cancela la carga de la computación sobre la infraestructura energética. Se obtiene 1 kWh de computación a partir de 1 kWh de energía que de cualquier manera se convertiría en calor.

En teoría, la demanda de calor residual casi no tiene límites.

Tal y como mencionó Mark Bjornsgaard, director ejecutivo de Deep Green Energy, en un [podcast](#) de DCD, la necesidad total de calefacción para los edificios en Europa es mucho mayor que la cantidad de energía utilizada en los centros de datos.

Por ejemplo, el Reino Unido utiliza unos 434 TWh de energía para calentar los edificios cada año, mientras que la energía total utilizada en los centros de datos es una pequeña fracción de dicha cifra, unos dos o tres TWh al año.

Esto equivale a 2 o 3 TWh de calor producido al año en los centros de datos. Si dicho calor pudiera usarse en la calefacción, en cierto nivel, estos centros de datos habrían alcanzado un consumo energético operativo cero neto. Además, como se siguen necesitando más de 400 TWh de calor, también hay margen para alimentar el aumento previsto de la IA.

Es una posibilidad tentadora, pero en la práctica no es tan sencilla, ¿cierto?

La dificultad de las redes térmicas

Los centros de datos han ofrecido su calor residual por muchos años, pero muy poco se ha utilizado con éxito.

Los centros de datos ofrecen una fuente de calor predecible, ya que mantienen sus servidores en operación continua, pero el calor es “de baja calidad”; es más cálido que caliente y se presenta en la forma de aire, el cual es difícil de transportar.

De manera que la mayoría de centros de datos descargan el calor en la atmósfera.

En algunas ocasiones, existen redes de calefacción urbana, las cuales proveen calor a los hogares y empresas locales a través de una red de tuberías. Si su centro de datos se encuentra cerca de una de estas redes, es cuestión de ampliarla para conectarla con el centro de datos y aumentar el grado de calor.

Sin embargo, es importante estar en el lugar correcto para conectarse a una. “En algunos países se han establecido o desarrollado redes de calefacción, pero la mayoría no cuenta con una red de calefacción en sí misma, por lo que se está haciendo de manera poco sistemática”, dijo a DCD Neal Kalita, director ejecutivo de potencia y energía para NTT.

De acuerdo con Rolf Brink, de la consultora de enfriamiento Promersion, es poco probable encontrar una red de estas en EE.UU.: “Estados Unidos es un ecosistema fundamentalmente diferente, pero Europa es mucho más densa en términos de población y hay una mayor demanda térmica”, afirmó.

Los países nórdicos tienen muchas redes de calefacción. Un buen ejemplo de ello son los parques de datos de Estocolmo; existe un campus de centros de datos en la zona urbana de Estocolmo, donde cada centro de datos tiene una conexión a la red de calefacción urbana y recibe una remuneración por su calor.

Además, Alemania tiene algunas redes de calefacción y planea ampliarlas, y el Reino Unido tiene algunas, pero la experiencia hasta ahora ha resultado difícil.

“Creo que el desafío de que los centros de datos conecten su calor a las redes de calefacción es cuándo estará en operación esa red de calefacción”, indica Kalita.

¿Qué tan bueno es su calor?

Otro de los principales desafíos se relaciona con la calidad del calor. “No es lo que llamamos calor de alta calidad o calor de alta temperatura”, afirma Kalita. “No se puede operar una red de calefacción completa a partir de la calefacción de un centro de datos. Para ello es necesario instalar una bomba de calor y poner más energía para aumentar la

temperatura, o la red de calefacción necesita tener algún tipo de fuente térmica”, agrega.

La red de calefacción debe estar muy cerca y el calor del aire caliente se debe complementar con otras fuentes de energía, como la quema de biomasa o el uso de una bomba de calor para aumentar la temperatura hasta que esta se pueda canalizar al sistema.

Por ejemplo, en Suecia, el experto en informática Cloud&Heat ha instalado un centro de datos, el cual consiste de dos contenedores de 6 metros (20 pies) enfriados por líquido y equipados con 1600 tarjetas gráficas de alto rendimiento en una instalación combinada de calor y energía (CHP, por sus siglas en inglés).

El sistema CHP quema biomasa para producir calor y electricidad. Los sistemas de Cloud&Heat usan parte de esa electricidad y el calor residual puede ir directamente al flujo térmico, de manera que nada se desperdicia.

Gracias a los nuevos sistemas de calefacción urbana de “cuarta generación”, el calor de baja calidad se ha convertido en un problema menor; estos sistemas se adaptan mejor para manejar el calor de baja calidad con medidas como tuberías de polietileno, las cuales pueden mantener el calor del agua caliente.

La incorporación de la agricultura

Naturalmente, los hogares en gran parte del mundo no necesitan tanto calor en el verano; así que la clave es planificar y encontrar otros usos para el calor. La producción alimenticia tiene muchos procesos que necesitan calor, desde invernaderos hasta piscifactorías, fermentación y producción de queso. De acuerdo con Wynne, el calor puede utilizarse incluso en países con veranos cálidos: “Por ejemplo, cuando el clima se vuelve más cálido, la energía se puede desviar de los hogares con calefacción a los sitios de fermentación”.

Green Mountain ha estado proporcionado el agua caliente de su centro de datos en Noruega a una piscifactoría; Lefdal Mine Datacenter hace lo mismo en otras partes de Noruega y el White Data Center en Japón, enfriado por nieve, aprovecha el calor para criar anguilas.

Dar valor a su calor

Otro problema es que, al considerar el calor como un producto residual, los centros de datos no lo valoran. Las redes de calefacción son un negocio y esperan comprarle el calor de forma comercial a un proveedor comprometido con un suministro predecible. Estas redes buscan acuerdos de compra de energía (PPA) para el calor.



“Al firmar un PPA de calor, se espera que ofrezca ese nivel de calidad y ese volumen de calor de manera constante”, indica Kalita. “Sin embargo, hay ocasiones en que los centros de datos sufren un fallo o se deben renovar las instalaciones y luego el calor desaparece.

Cuando uno se compromete a ofrecer calefacción durante 10-20 años, esto podría limitar su flexibilidad”, agrega.

Para algunos, esto no es insuperable. El consultor en reutilización térmica, David Gyulnazaryan, indicó a una audiencia en un centro de datos de Londres que el calor residual es “una fuente de calor altamente sostenible y confiable” y, en caso de ser necesario, puede contar con la “generación energética como respaldo en caso de fallos, lo cual permite un nivel de confiabilidad del 99,996 %”.

Desde el punto de vista financiero, Gyulnazaryan afirma que “el costo de convertir el calor de baja calidad del centro de datos es de unos 25 €/MWh, mientras que el precio del gas natural es de unos 200 €/MWh.

De manera que el potencial de ahorros es enorme para los organismos públicos, especialmente aquellos con un suministro de gas natural cada vez menor”, señala.

De acuerdo con Brink, es una cuestión de planificación: “Cuando se toma en cuenta de manera efectiva el calor durante la fase de construcción de una instalación de centro de datos, este puede significar un enorme activo

con considerables reducciones en su costo total de propiedad.

“A la hora de elegir una ubicación para construir un centro de datos, la reutilización térmica no se toma en cuenta en la mayoría de las ocasiones. “Si se toma en cuenta, podría ubicar un centro de datos junto a una piscina o cerca de algo que aprovechara el calor”, agrega Brink.

Tal y como afirma, “La discusión sobre la reutilización térmica está un poco desequilibrada, ya que mientras no se incluya en el estudio de evaluación del sitio, no tendrá sentido discutirlo más adelante, simplemente porque no se elige un sitio para dicho propósito”.

Las regulaciones promueven el intercambio de calor

El mayor problema de los sistemas de calefacción urbana podría ser que su implementación exige enormes inversiones y muchos años. Algunos gobiernos locales y nacionales están interviniendo para fomentar el uso de redes térmicas.

En la región de Ámsterdam de los Países Bajos, los centros de datos se deben diseñar para poder conectarse con las redes de calefacción urbana. Alemania cuenta con normas similares en todo el país, las cuales han sido consagradas en su recién aprobada Ley de Eficiencia Energética.

Esto no ha sido algo sencillo. Los anteproyectos exigían que las nuevas instalaciones construidas a partir de 2025 reutilizaran el 30 % de su calor. La Asociación Alemana de Centros de Datos se opuso al proyecto de ley y el acta final menciona que los centros de datos de más de 200 kW necesitan reutilizar únicamente el 20 % de su calor y no deben hacerlo hasta 2028.

Esta asociación se sigue oponiendo al proyecto, ya que este requisito limita los centros de datos a ubicaciones donde existen sistemas de calefacción.

Benjamin Brake, jefe del departamento de Políticas Digitales y de Datos del Ministerio Federal de Asuntos Digitales y Transportes (BMDV), reconoció en la conferencia de la GDA que “El uso del calor residual como único criterio para elegir una ubicación es problemático, especialmente cuando no existen requisitos ni obligaciones para los usuarios del calor residual”.

De acuerdo con la [Federación Nacional de Vivienda](#), en el Reino Unido existen unas 14.000 redes de calefacción, las cuales son en su mayoría pequeñas. El Fuel Poverty Action Group [critica](#) el estado de las redes de calefacción británicas y argumenta que los clientes nacionales tienen pocos derechos, ya que las redes no se encuentran reguladas. Además, son vulnerables a cortes y aumentos drásticos de los precios, pues la mayoría de las redes incluyen aportes térmicos del gas natural, cuyo precio se ha disparado.

El [gobierno](#) cuenta con una línea de nuevos proyectos de calefacción urbana, la cual asciende a casi £3.000 millones (\$3800 millones). Únicamente un proyecto en Park Royal, con una subvención de £36 millones, pretende aprovechar el calor residual del centro de datos y no se pondrá en marcha hasta 2040 (ver cuadro lateral).

Dado el enorme costo del proyecto y su largo plazo, muchos observadores han expresado a DCD sus dudas de que este algún día se complete.

"Londres odia absolutamente tener que romper las carreteras", afirma Kalita. "Imagínese romper una carretera para una nueva red de calefacción a lo largo de las distancias que recorrerá la red".

Otros se enfrentan a un problema mayor y se preguntan si debería hacerse.

"No me encuentro lo suficientemente cerca del proyecto para saberlo, pero cuando vi el costo tuve que mirarlo dos veces", indica Charlie Beharrell de Heata. "El costo de conexión por hogar, sin incluir cualquier TI o sobrecostos, es mayor que el costo total de instalación de uno de nuestros sistemas".

A estos sistemas se les llama calderas digitales y calderas de datos.

Beharrell, junto con Bjornsgaard y un grupo de otras empresas, apoyan un enfoque diferente, el cual podría ser más eficaz que las redes térmicas, pero replantea el centro de datos.

La caldera de datos.

La idea de una caldera de datos fue propuesta por primera vez por Microsoft en un [artículo](#) publicado en 2011 y escrito por un grupo que incluía a Sean James. Básicamente, la idea consiste en colocar racks de servidores en hogares y oficinas donde su calor sea útil.

El desarrollo de la nube se traduce en que los trabajos útiles puedan enviarse de forma cifrada a la caldera de datos. Los clientes pagan un precio por la computación, el cual cubre la electricidad utilizada y el anfitrión obtiene calor gratis.

"En los climas más fríos, unas 110 tarjetas madre podrían mantener un hogar tan cálido como lo haría un calentador convencional", afirmaron los autores del artículo. "El resto del año, los servidores seguirían en funcionamiento, pero el calor generado se expulsaría al exterior, tan inofensivo como el calor producido por una secadora de ropa", agregan.

Muchas compañías han desarrollado negocios alrededor de esta idea, entre ellas Qarnot en Francia y Cloud&Heat en Alemania.

La más reciente generación de compañías de informática y calefacción distribuida en el Reino Unido incluye a Deep Green y Heata.

Por un lado, Deep Green instala servidores con uso informático intensivo en tanques de inmersión enfriados por líquido y de esta manera su energía térmica puede canalizarse a los usuarios cercanos. Sus primeros clientes son las piscinas. Por otro lado, Heata ofrece electrodomésticos que generan calor para el hogar.

Ambos utilizan el mismo proveedor de nube, Civo, para enviar trabajos a sus unidades.

La idea tiene el potencial de superar a las redes de calefacción urbana, afirma Beharrell, ya que "es mucho más barato y más fácil trasladar los datos que el calor".

Por ejemplo, las calderas de datos pueden instalarse en cada vivienda de una urbanización, con necesidades de infraestructura mínimas. Los hogares ya cuentan con electricidad.

Es posible que algunos necesiten mejoras en la red, pero ese es un trabajo mucho menor a tener que tender tuberías desde un gran centro de datos, a través del sistema de calefacción, hasta cada hogar.

Beharrell no cree que los dos enfoques sean incompatibles y se siente feliz de ver el avance de proyectos como la red de calefacción de Park Royal: "Creo que es bueno ser optimista y, siendo sinceros, es bueno para todos si esto sigue adelante".

Sin embargo, le gustaría ver una mayor sensibilización sobre la posibilidad de distribuir la computación, en lugar de trasladar el calor: "Sería agradable ver una mayor inversión detrás de modelos más disruptivos en lugar de reforzar el status quo", afirma.

Decidir cuál es el mejor enfoque para una situación determinada exige un pensamiento genuinamente estratégico.

"Lograr reducir las emisiones de carbono necesarias para 2050 implica un enfoque de sistemas completo", indica Wynn. "Integrar a los centros de datos en esto contribuiría enormemente a alcanzar el éxito". ●



Foto de Luke Stackpoole

CALEFACCIÓN URBANA EN LONDRES

Este año, el gobierno del Reino Unido destinó £36 millones (\$44.5 millones) a un sistema de calefacción urbana en la zona occidente de Londres, el cual será construido por Aecom y compartirá el calor residual del centro de datos con hasta 10.000 nuevos hogares.

Dos centros de datos en construcción tienen el objetivo de ofrecerle 98,7 GWh de calor al sistema, el cual estaría en funcionamiento hasta 2040.

El Green Heat Network Fund (GHNF) del Departamento de Energía pagará el dinero a Old Oak and Park Royal Development Corporation (OPDC) para crear un sistema de calefacción que aproveche el calor residual a temperaturas entre 20 °C (68 °F) y 35 °C. (95 °F).

La ubicación es significativa, porque es precisamente allí donde una escasez de distribución eléctrica interrumpió los proyectos de construcción de viviendas en 2022, debido a que los centros de datos habían reservado toda la energía disponible de las subestaciones locales.

Además, el proyecto recibirá unas £400.000 (\$495.000) de financiamiento de experiencia técnica del programa Local Energy Accelerator (LEA) de parte del alcalde de Londres y David Lunts, director ejecutivo de OPDC, quien lo describe como "un ejemplo emocionante e innovador del apoyo de OPDC al ambicioso objetivo de cero emisiones netas del alcalde". ●

Cuando menos es más

*Los operadores de centros de datos necesitarán adoptar estrategias de enfriamiento híbrido para gestionar las crecientes densidades de TI.
¿Qué significa esto?*

"Menos es más" es uno de esos clichés que rara vez soportan un escrutinio real, especialmente cuando se trata de centros de datos. Año tras año, los clientes siempre quieren más poder, más espacio y más de todo, y los operadores de centros de datos deben satisfacer estas demandas sin titubeos en su marcha hacia las cero emisiones netas.

Estas demandas suponen todo un desafío en lo relacionado con el enfriamiento convencional de las salas de servidores, con racks abarrotados de CPU y GPU más potentes, las cuales generan más calor y ponen a los diseñadores en una encrucijada a la hora de maximizar el flujo de aire y el panorama térmico del centro de datos.

De hecho, se espera que la demanda de aplicaciones de IA y aprendizaje automático se intensifique; por lo tanto, los racks con múltiples GPU a bordo serán cada vez más comunes para manejar el aumento de la carga de trabajo. Sin embargo, ¿podrá el centro de datos soportar el calor generado por todo este silicio en operación?

La lucha por el aire

"Todas estas nuevas demandas aumentan en gran medida la complejidad del panorama térmico", afirma Nigel Gore, director ejecutivo de enfriamiento líquido en Vertiv.

Aunque el enfriamiento por aire ha evolucionado para afrontar los mayores desafíos de cada generación de tecnología, también se enfrenta al reto de adaptarse a la siguiente fase de la computación.

La industria solucionó el enfriamiento por aire de alta densidad al acercarlo a la carga de calor en los racks en forma de enfriamiento a nivel de filas y el enfriamiento basado en racks. El siguiente paso implicó gestionar la recirculación del aire de salida que entra al pasillo caliente a temperaturas

altas para evitar aumentar la temperatura ambiente de toda la sala de servidores.

Sin embargo, aunque la tecnología y las topologías de enfriamiento por aire se han adaptado ingeniosamente a lo largo de los años conforme crecieron las necesidades de TI, los desafíos actuales son tan grandes que se ha llegado a un punto de inflexión donde se deberán implementar tecnologías de enfriamiento líquido para manejar el aumento de potencia a nivel del procesador.

Los operadores de centros de datos deberán tomar la siguiente decisión: ¿Cómo incorporarán el enfriamiento líquido en sus salas de servidores para soportar la TI de alto rendimiento y, en consecuencia, atraer y retener a los clientes adinerados que lo necesitan?

Es fundamental que la instalación de tecnologías de enfriamiento directo al chip, de empresas como Vertiv, en las salas de servidores sea sencilla.

Los racks de servidores ocupan el mismo espacio que los racks convencionales y pueden instalarse junto con TI enfriados de forma convencional en la misma sala.

Además, el mayor "riesgo" identificado que presenta el enfriamiento líquido (los periodos de inactividad como resultado de una fuga) se ha reducido

considerablemente por medio de un diseño de racks con distribuidores de líquido integrados que incorporan tubería de alta calidad. Esta tubería incluye múltiples acoplamientos, válvulas y sistemas de seguridad, para que los ingenieros puedan realizar actualizaciones y mantenimiento de manera rápida y segura en los racks o incluso en los servidores individuales montados en racks.

De hecho, si su impresión del enfriamiento líquido proviene exclusivamente del mundo de las computadoras, donde los entusiastas compiten por construir el equipo sobrealimentado y más potente que su dinero pueda comprar (ya sea con un enfriador de agua o incluso nitrógeno líquido en algunos casos extremos), la realidad de un moderno rack de servidores enfriado por líquido le dejará sin aliento.

La potencia y el agua

De acuerdo con Gore, en una instalación típica en rack, lo primero es asegurarse de soportar la densidad de potencia. "Muchos centros de datos convencionales han sido diseñados alrededor de 5 kW por rack, pero ahora vemos 25 kW, 50 kW e incluso 70 kW por rack", señala Gore. De hecho, añade que es posible que las densidades de potencia lleguen a alcanzar los 100-150 kW por rack en poco tiempo, ya que se espera que la cantidad de transistores aumente exponencialmente en el silicio del futuro.

Debajo de la barra de alimentación, las tuberías para el refrigerante suelen conectarse de manera similar a cada rack. Para ello, Vertiv usa tuberías de acero inoxidable de alta calidad, las cuales permiten garantizar que la corrosión galvánica no contaminará el refrigerante ni afectará la transferencia y el flujo de calor. Además, la compañía está analizando la evolución de los plásticos resistentes a altas temperaturas como una posible alternativa.

Esta red de tuberías de enfriamiento líquido, a la cual se le llama en términos técnicos una red de fluido secundaria (SFN, por sus siglas en inglés), se encuentra diseñada con válvulas de cierre para el mantenimiento y los acoplamientos desde los racks hasta la tubería principal cuentan con válvulas en línea para aislar aún más el flujo, en caso de ser necesario. "Son acoplamientos manuales que pueden desconectarse, un diseño con una válvula de bola que evita automáticamente cualquier goteo al momento de la desconexión", indica Gore.

Como alternativa, existen válvulas electrónicas que facilitan la manipulación en el control del flujo de refrigerante en combinación con la unidad de distribución de refrigerante.

Todas estas tuberías y accesorios se conectan con los racks por medio de un



La barra de alimentación y los distribuidores suspendidos del rack son componentes vitales para soportar las implementaciones de enfriamiento líquido.

suministro (refrigerante frío) y un retorno (refrigerante caliente), y los distribuidores en las filas permiten aislar los racks individuales, en caso de ser necesario. Por lo tanto, el trabajo puede llevarse a cabo en un rack —por ejemplo, para instalar actualizaciones mientras los demás se encuentran en funcionamiento—.

En el caso de las salas con piso elevado en el centro de datos, esta red de tuberías puede instalarse debajo del piso. “Sin embargo, personalmente creo que es más fácil de operar cuando las tuberías se encuentran arriba, por ejemplo, cuando es necesario apagarlas para realizar tareas de mantenimiento o instalar actualizaciones en el rack”, indica Gore.

Sin importar dónde coloque las tuberías, es indispensable mantener la presión y la temperatura adecuadas. Para ello, las unidades de distribución de refrigerante —ubicadas en la fila, en el estante o en el perímetro— monitorean el flujo de forma inteligente para controlar el caudal, además de ofrecer filtrado, puertos de llenado de agua de fácil acceso y drenaje.

“Además, somos testigos de implementaciones en las que los equipos de las instalaciones quieren controlar automáticamente el líquido de cada rack para tener control grupal y cerrar el flujo hacia los racks individuales, y no solo cerrar el flujo, sino también controlar el volumen del flujo, como lo hacen los equipos de las instalaciones”, agrega Gore. En otras palabras, el enfriamiento de los racks puede ajustarse de formas que no son posibles con el enfriamiento por aire.

En la parte trasera de los racks, cada servidor montado en rack cuenta con acoplamientos para el refrigerante líquido (una línea de agua de suministro y de retorno, codificada por colores para mayor conveniencia), mientras que los servidores tienen un diseño estandarizado para las CPU o GPU, al fijar de forma segura las placas de refrigeración directamente sobre los chips.

Dentro del servidor, los contactos de cobre ofrecen el mejor material conductor de calor para transferir el calor de la CPU/GPU al líquido y, mientras el agua vuelve a la unidad de distribución de refrigerante y al intercambiador de calor, los ventiladores

del servidor expulsan el calor residual al aire por parte trasera del rack hasta el pasillo caliente de la sala de servidores, lejos de la TI.

Totalmente sumergido en enfriamiento

Naturalmente, el enfriamiento líquido a nivel del chip no es la única opción líquida. El enfriamiento por inmersión sigue siendo una opción, pero supone una curva de aprendizaje para muchos operadores, con servidores sumergidos en tanques horizontales, en lugar de verticales, en un líquido no conductor que cubre todos los componentes en el rack, no solo las CPU y/o GPU. Hemos visto los avances en las especificaciones del líquido dieléctrico del Open Compute Project adoptados por Intel.

“Significa poner todo el servidor en un líquido dieléctrico con una composición química específica. Debido a que todos los componentes estarán sumergidos (los transistores, los condensadores, los módulos de memoria, los cables, los conectores, etc.), debe asegurarse de que no existan riesgos de contaminación para garantizar que sus servidores sean compatibles con el sistema”, indica Gore.

Por ejemplo, el cableado y los plásticos revestidos de PVC podrían disolverse en los líquidos o aceites dieléctricos, así que deberán cambiarse por plásticos o cables Plenum. Los proveedores o el Open Compute Project han puesto a disposición material de referencia, el cual muestra los requisitos de inmersión y las guías de compatibilidad de los materiales.

Además, habrá pérdidas de calor radiado desde el tanque, las cuales la instalación deberá soportar. La guía operativa también señalará como riesgo de contaminación cada vez que un ingeniero abra el tanque para dar mantenimiento a un servidor o instalar actualizaciones.

De acuerdo con Gore, al comparar los líquidos, ninguno de ellos puede acercarse a la eficacia del agua en lo relacionado con la física de la transferencia de calor.

“Cuando se examinan las propiedades de un fluido dieléctrico en comparación con el agua en peso o volumen, la diferencia es significativa en términos de las capacidades y propiedades de transferencia de calor. En el caso de un fluido dieléctrico, cualquier cambio en la viscosidad tiene un efecto significativo en la capacidad de transporte térmico. Al igual que con el enfriamiento por aire, la complicación del “flujo de derivación”, donde los obstáculos dentro del servidor impiden un flujo óptimo a través de los componentes más calientes, con lo cual el enfriamiento puede ser menos efectivo en comparación con el enfriamiento directo al chip.

“Entonces se debe diseñar en función de canalizar el líquido para asegurarse de que siga siendo un refrigerante eficaz”, señala Gore. Por lo tanto, la captura óptima del calor generado a medida que el líquido es bombeado alrededor del tanque puede suponer todo un desafío en servidores y tanques densamente poblados. De acuerdo con Gore, esto es algo que se comprende bien y los equipos de desarrollo han hecho grandes avances en disipadores de calor optimizados y controles de distribución de flujo interno.

Sin embargo, quizás donde el enfriamiento directo al chip obtiene la puntuación más alta en comparación con el enfriamiento por inmersión es en la utilización de agua en lugar de un aceite mineral no conductor. “El líquido por excelencia antes de cualquier otro fluido para enfriar es básicamente agua; esta tiene la mayor capacidad de transporte térmico de los líquidos. Con el enfriamiento líquido, estamos llevando el líquido de mayor conducción de calor directamente a los componentes más calientes del rack lo más rápido posible. El enfriamiento líquido forma parte de nuestro futuro como una tecnología emergente que aprovecha el cambio de fase para acelerar la recolección de calor”, afirma Gore.

Con el aumento de racks compatibles con el enfriamiento líquido, la incorporación de este al centro de datos ya se encuentra en marcha. Vertiv está desarrollando guías y soporte para que las instalaciones puedan agregar racks enfriados por líquido, los cuales pueden coexistir en salas de servidores con racks enfriados por aire y expandirse a medida que crezca la demanda de hardware de servidores enfriados por líquido.

A medida que la próxima generación de tecnologías de chips llega al centro de datos, el problema del enfriamiento se intensificará. Para poner en perspectiva las características de generación de calor de las CPU y GPU que llegarán inminentemente, 750 W son suficientes para alimentar un mini calentador de aceite que podría mantener una habitación modesta caliente durante todo el invierno.

Con ocho GPU de 750 W por servidor montado en rack y entre 14 y 42 servidores por grupo, eso equivale a casi 340 calentadores por grupo. Multiplicados en un centro de datos típico, los que no cuentan con una solución de enfriamiento eficaz corren el riesgo de colapsar.

>>> Para conocer más sobre la tecnología de enfriamiento líquido de Vertiv, eche un vistazo al [sitio web de opciones de enfriamiento exclusivas de Vertiv](#)



Líneas de refrigerante de suministro y retorno directas al chip

Los microfluidos: El enfriamiento dentro del chip



Peter Judge
Editor ejecutivo

Si usted piensa que la inmersión es la fase final del enfriamiento líquido, piénselo de nuevo. DCD escucha a los ingenieros que quieren que el refrigerante fluya dentro de sus chips

Todos sabemos que el enfriamiento líquido es el futuro de los centros de datos. El aire simplemente no puede soportar las densidades de potencia que llegan a las salas de datos, por lo cual los fluidos densos con una alta capacidad térmica poco a poco tomarán el control.

A medida que aumenta la densidad de calor de los equipos de TI, los líquidos se acercan cada vez más a ella, pero ¿qué tan cerca pueden llegar los líquidos?

Poner en funcionamiento un sistema de circulación de agua a través de las puertas traseras de los gabinetes del centro de datos ha tenido bastante aceptación. Después, los sistemas han estado haciendo circular agua hasta las placas frías en los componentes particularmente calientes, como las GPU o las CPU.

Más allá de eso, los sistemas de inmersión han sumergido racks enteros en tanques de fluido dieléctrico, así el líquido refrigerante puede entrar en contacto con cada parte del sistema. En la actualidad, los principales proveedores ofrecen servidores optimizados para la inmersión.

Sin embargo, hay un paso adicional. ¿Qué pasaría si el fluido pudiera acercarse a la fuente de ese calor: los transistores dentro de los propios chips de silicio? ¿Qué pasaría si los refrigerantes fluyeran dentro de los procesadores?

Husam Alissa, director de tecnología de sistemas de Microsoft, considera esto como una opción futura interesante: "En los microfluidos, algunas veces llamados enfriamiento integrado, 3D heterogéneo

o enfriamiento integrado, llevamos el enfriamiento al interior del silicio, muy cerca de los núcleos activos que realizan el trabajo".

Esto es algo más que un mejor sistema de enfriamiento. "Cuando uno trabaja con microfluidos, ya no está resolviendo solo un problema térmico" afirma. Los chips con su propio sistema de enfriamiento podrían solucionar el problema en su origen, en el propio hardware.

El nacimiento de los microfluidos

En 1981, los investigadores David Tuckerman y R. F. Pease de Stanford sugirieron que el calor podría eliminarse de manera más efectiva con "microcanales" grabados en un disipador térmico con técnicas similares a las utilizadas en las fundiciones de silicio.

Los pequeños canales tienen una mayor superficie y eliminan el calor de forma más eficaz.

Según sugirieron, el disipador térmico podría convertirse en parte integral de los chips VLSI y su demostración probó que un disipador térmico de microcanales podría soportar un impresionante flujo térmico de 800 W por m².

Desde entonces, la idea ha persistido en las universidades, pero solo afectó de manera tangencial al silicio de la vida real en los centros de datos.

En 2002, los profesores de Stanford Ken Goodson, Tom Kenny y Juan Santiago fundaron Cooligy, una empresa reciente con un impresionante diseño de "microcanales activos" en un disipador

térmico integrado directamente en el chip, junto con una inteligente bomba electrocinética silenciosa de estado sólido para hacer circular el agua.

Las ideas de Cooligy han sido absorbidas por parte de las tendencias principales. La empresa fue comprada por Emerson Network Power en 2005 y su tecnología y parte de su personal todavía trabajan en la nueva encarnación de Emerson, Vertiv.

La idea de integrar el enfriamiento y el procesamiento se volvió más práctica a medida que la fabricación de silicio se desarrolló y pasó a ser tridimensional. A partir de la década de 1980, los fabricantes realizaron experimentos y construyeron múltiples componentes uno encima del otro sobre una pastilla de silicio.

Crear canales en los pisos superiores de un chip de silicio multicapa es potencialmente una ganancia rápida para el enfriamiento, ya que se puede comenzar con la implementación de pequeñas ranuras similares a las aletas vistas en los disipadores de calor.

Sin embargo, la idea no tuvo mucho éxito, ya que los proveedores de silicio querían utilizar técnicas 3D para apilar los componentes activos. Actualmente, dicho enfoque se acepta para la memoria de alta densidad y las patentes sugieren que Nvidia podría tener la intención de apilar las GPU.

En la industria de los microprocesadores, el enfriamiento y el procesamiento se consideraban disciplinas separadas. Los chips se diseñaron para disipar su calor,

pero esto se hizo por medios relativamente sencillos, que utilizan materiales térmicos para desviar el calor al gran disipador térmico de cobre en la superficie.

El disipador térmico podía mejorarse por medio de grabar canales más pequeños, pero era un elemento separado y el calor tenía que cruzar una barrera de adhesivo para llegar allí.

Sin embargo, algunos investigadores pudieron ver las posibilidades. En 2020, Tiwei Wei, del Centro Interuniversitario de Microelectrónica y la KU Leuven en Bélgica integraron la electrónica y el enfriamiento en un solo chip.

Wei, cuyo trabajo fue publicado en Nature en 2020, no creía que la idea se volvería popular en los microprocesadores y afirmó que los microcanales de enfriamiento serían más útiles en la electrónica de potencia, donde los chips de gran tamaño fabricados a partir de semiconductores como el nitrato de galio (GaN) gestionan y convierten la electricidad dentro de los circuitos.

Posiblemente esto explica la razón por la cual Emerson/Vertiv quería adquirir Cooligy, pero Wei no veía que la tecnología fuese más allá: "Este tipo de solución de enfriamiento integrado no está prevista para los procesadores y chips modernos como la CPU", indicó a la revista IEEE Spectrum.

Profundizar en los chips

En ese tiempo, los investigadores ya llevaban varios años trabajando en el grabado de canales de microfluidos en la superficie de los chips de silicio. Un equipo de Georgia que trabajó con Intel en 2015 podría haber sido el primero en fabricar chips FPGA con una capa de enfriamiento de microfluidos integrada, sobre el silicio, "a unos cientos de micrones [micrómetros] de donde operan los transistores".

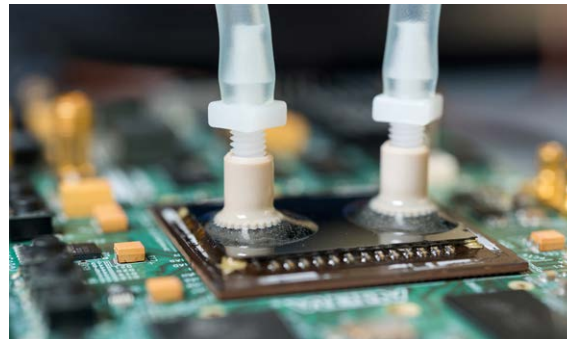
"Hemos eliminado el disipador térmico sobre la pastilla de silicio por medio de alejar el enfriamiento líquido unos cientos de micrones del transistor", indicó Muhannad Bakir, líder del equipo y profesor del Instituto de Tecnología de Georgia, en un comunicado de prensa de este Instituto. "Creemos que la integración confiable de enfriamiento de microfluidos directamente sobre el silicio se convertirá en una tecnología disruptiva para una nueva generación de electrónica", agregó.

En 2020, algunos investigadores de la Escuela Politécnica Federal de Lausana, en Suiza, fueron un paso más allá y, de hecho, hicieron pasar fluido por túneles debajo de los transistores generadores de calor.

El profesor Elisa Matioli vio la posibilidad de acercar aún más las cosas: "Diseñamos la electrónica y el enfriamiento en conjunto desde el principio", indicó en 2020, cuando se publicó el artículo de su equipo en Nature.

El equipo de Matioli había logrado diseñar una red 3D de canales de enfriamiento de microfluidos dentro del propio chip, justo debajo de la parte activa de cada transistor, a unos cuantos micrómetros de donde el calor es generado. De acuerdo con Matioli, este enfoque podría mejorar el rendimiento del enfriamiento en un factor de 50.

Además, Matioli grabó ranuras de un micrómetro de ancho



en una capa de nitrato de galio sobre un sustrato de silicio y luego amplió las ranuras en el sustrato de silicio para formar canales lo suficientemente grandes como para bombear un líquido refrigerante.

Luego, las pequeñas aberturas en la capa de nitrato de galio se sellaron con cobre y se creó un dispositivo de silicio normal en la parte superior. "Únicamente tenemos microcanales en la pequeña zona de la oblea en contacto con cada transistor", indicó en ese momento. "Eso permite que la técnica sea eficiente", agregó.

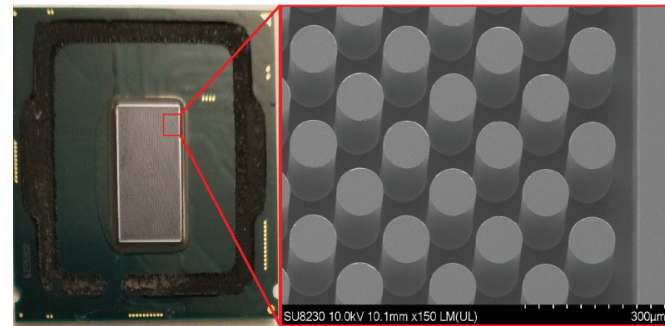
Matioli fue capaz de fabricar dispositivos con un alto consumo energético, como un circuito rectificador de CA a CD de 12 kV, el cual no necesitaba un disipador térmico externo. Los microcanales llevaron el fluido directamente hasta los puntos críticos y manejaron increíbles densidades de potencia de 1,7 kW por cm², o, dicho en otras palabras, 17 MW por m², varias veces el flujo de calor generado actualmente por las GPU.

Manos a la obra con el silicio estándar

De forma paralela, se sigue trabajando para agregar microfluidos al silicio estándar, por medio de crear estructuras de microfluidos en la parte trasera de los microprocesadores existentes.

En 2021, un equipo liderado por Microsoft, junto con Husam Alissa, utilizó aletas de "micropines" grabadas directamente en la parte trasera de una CPU Intel Core i7-8700K estándar disponible en el mercado.

"En realidad, tomamos un procesador de escritorio listo para la venta y le quitamos la carcasa", afirmó Alissa. Sin la cubierta del disipador térmico y el material de interfaz



térmica (TIM, por sus siglas en inglés), la pastilla de silicio del chip quedó expuesta.

"Cuando esa pastilla quedó expuesta, aplicamos métodos de grabado para crear los canales que queríamos ver", agregó. La parte trasera de la pastilla se grabó de forma selectiva, a una profundidad de 200 micrones, lo cual dejó un patrón de rastreo con barras de 100 micrones de espesor: los "micropines" que forman la base del sistema integral de enfriamiento directo al chip.

De acuerdo con Alissa, es una tarea delicada. "Uno debe tomar en cuenta la profundidad del grabado, de manera que las áreas activas del silicio no se vean afectadas", indicó.

Finalmente, la parte trasera de la CPU se selló en un distribuidor impreso en 3D, el cual permitía el flujo de refrigerante entre los micropines. Luego, el chip se sobrecalentó para disipar 215 W de potencia, más del doble de su potencia de diseño térmico (TDP, por sus siglas en inglés), la energía que puede manejar de forma segura sin sobrecalentarse.

Por sorprendente que parezca, el chip funcionó a dicho nivel con únicamente agua a temperatura ambiente proporcionada a través del distribuidor. El experimento mostró una reducción del 44 % en la resistencia térmica de la unión a la entrada y usó una trigésima parte de la cantidad de refrigerante por vatio que habría sido necesaria con una placa fría convencional. El rendimiento fue evaluado con programas de comparación estándar.

Esta fue la ocasión en que se crearon canales de microfluidos directamente en una CPU de consumo estándar y se alcanzó la mayor densidad de potencia con enfriamiento de microfluidos en un dispositivo CMOS activo. Tal y como informó el grupo en IEEE Xplore, los resultados demuestran el potencial para operar centros de datos de forma más eficiente sin necesidad de sistemas de enfriamiento con un alto consumo energético.

Todo lo que se necesitaría es que el fabricante de chips produjera procesadores en masa con micropines grabados y los vendiera empaquetados con un distribuidor

acoplado en lugar de la tapa habitual del disipador térmico. De acuerdo con Alissa, si las fundiciones como TSMC pudieran ofrecerles a sus chips enfriamiento líquido incorporado, esto cambiaría la dinámica de la adopción y le permitiría a la tecnología superar las barreras.

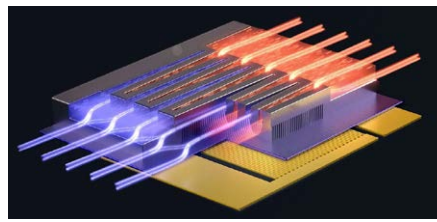
“Con las placas frías, es posible que obtenga agua a 40 °C (104 °F), pero con microfluidos probablemente podría tener una temperatura de 80 °C (176 °F) o mayor procedente de estos chips porque el refrigerante está muy cerca de los núcleos activos”, afirma. “Naturalmente, esto mejora la eficiencia y los beneficios de recuperación térmica, junto con menores requisitos de caudal”, agrega.

El futuro de los microfluidos

“Existen dos tipos de microfluidos principales”, señala Alissa. La opción más ligera, que según él podría implementarse “en un par de años”, es el enfoque que mostró su equipo, grabar canales en chips comerciales: “Se compran los chips, se hace el grabado y listo”.

Una versión más desarrollada de este enfoque sería que las fundiciones hicieran el grabado antes de que el chip llegara al consumidor, ya que no todos están dispuestos a quitarle la parte trasera a un procesador y atacarlo con ácido.

Además de eso, existe lo que Alissa llama el enfoque de “contacto más fuerte”, en el cual “se intercepta temprano la



fundición y se comienza a construir estructuras en 3D”.

De esta manera, se refiere a los chips porosos que apilan los componentes uno encima del otro con canales de refrigerante en las capas intermedias.

Se trata de una evolución basada en el enfoque utilizado por Matioli en Lausana. Según Alissa, “Eso resulta más prometedor, pero, naturalmente, implica más trabajo”.

Además, Alissa tiene un objetivo. “La Estrella Polar a la que queremos llegar es donde podamos optimizar conjuntamente este chip para el enfriamiento y eléctricamente al mismo tiempo, por medio de apilar varios tintes uno encima del otro, con grabado [de microcanales] en el medio”, explicó.

Si el enfriamiento permite apilar muchos componentes en una pastilla de silicio profunda, conectados por “vías a través de chips” (TCV, por sus siglas en inglés), conexiones de cobre a través de la pastilla de silicio. Estos chips de torre podrían necesitar menos energía y

funcionar mucho más rápido, ya que los componentes se encuentran más cerca: “En general, hay ganancias en rendimiento, se mejora el enfriamiento y también la latencia debido a la proximidad”, indicó.

Además, hay otro beneficio. Si los microfluidos permiten que los chips alcancen un punto de diseño térmico (TDP, por sus siglas en inglés) más alto, esto podría eliminar uno de los obstáculos que enfrentan actualmente los diseñadores de silicio.

La dificultad de eliminar el calor significa que los chips más grandes de la actualidad no pueden utilizar todos sus transistores a la vez o de lo contrario se sobrecalentarán. Los chips tienen áreas de “silicio oscuro” (vea el cuadro) y la aplicación de microfluidos les permitiría a los diseñadores iluminarlas, lo cual mejora el rendimiento de los chips.

Sin embargo, no espere a que los microfluidos lo resuelvan todo. En 2012, el profesor Nikos Hardavellas predijo un problema. “Aun si se emplearan tecnologías de enfriamiento exóticas, como el enfriamiento líquido junto con microfluidos, la entrega de energía al chip probablemente impondría nuevas limitaciones”, indicó.

Una vez que descubramos cómo extraer más calor del chip, tendremos que desarrollar formas de entregar una gran cantidad de energía, la cual pueda proporcionar integridad de la señal a los bajos voltajes requeridos por los transistores.

¿Estamos preparados para ello? ●

EL SILICIO OSCURO

Las generaciones de chips actuales y futuras se enfrentan a un problema importante. El rendimiento siempre ha aumentado a medida que se ponen más transistores en un solo procesador, pero en la actualidad hay tantos que no se pueden utilizar todos a la vez sin que el chip se sobrecaliente.

Los fabricantes de procesadores especifican una potencia de diseño térmico (TDP, por sus siglas en inglés) para cada chip; esta representa la cantidad de energía que puede manejar y disipar de forma segura, y asumirán que hay un buen disipador térmico en el chip.

Los TDP han crecido mucho. Por ejemplo, la GPU Nvidia H100 SXM5 tiene un TDP de 700 W, enorme en comparación con las CPU estándar como Intel Xeons, que consumen unos 130 W.

Pero ¿cuánto se puede hacer con esa potencia? Actualmente, los transistores fabricados a 4 nm consumen apenas 10 attojulios (10-18 julios) cada uno, por lo que, si uno de ellos conmutara a 1,8 GHz, consumiría 18 microvatios (18 x 10-9 W).

Es una cantidad pequeña, pero los procesadores actuales cuentan con una enorme cantidad de transistores.

Jon ummers, de RISE, el instituto sueco de investigación, calcula que la GPU Nvidia H100, la cual tiene 80.000 millones de transistores, generaría 1440 W, más del doble del TDP que Nvidia establece para ella.

“Con un TDP de 700 W, debe significar que el 51 % del chip es silicio oscuro”, indicó Summers ante una audiencia en el DCD Connect London en noviembre de 2024.

La miniaturización continua no solucionará la situación. Los transistores más pequeños tienen una energía de conmutación menor, por lo cual se pueden encender más dentro de la propuesta del TDP, pero el número de transistores también está aumentando.

Summers indica que Intel planea tener un billón de transistores en un chip para 2030, cada uno de los cuales usará alrededor de 1aJ por conmutador. Si la velocidad del reloj ha subido a 4 GHz, el chip tiene 1000 mm cuadrados y un flujo térmico; eso significa que el 40 % de los transistores deben permanecer oscuros.

En la actualidad, los TDP se basan en el flujo térmico máximo que se puede eliminar de un chip. La Nvidia H100 tiene un área de

814 mm², por lo que el flujo térmico es de 860kW por m². Esto puede compararse con los niveles que se ven en las demostraciones de fusión nuclear y Summers espera que Intel alcance los 2,4 MW por m².

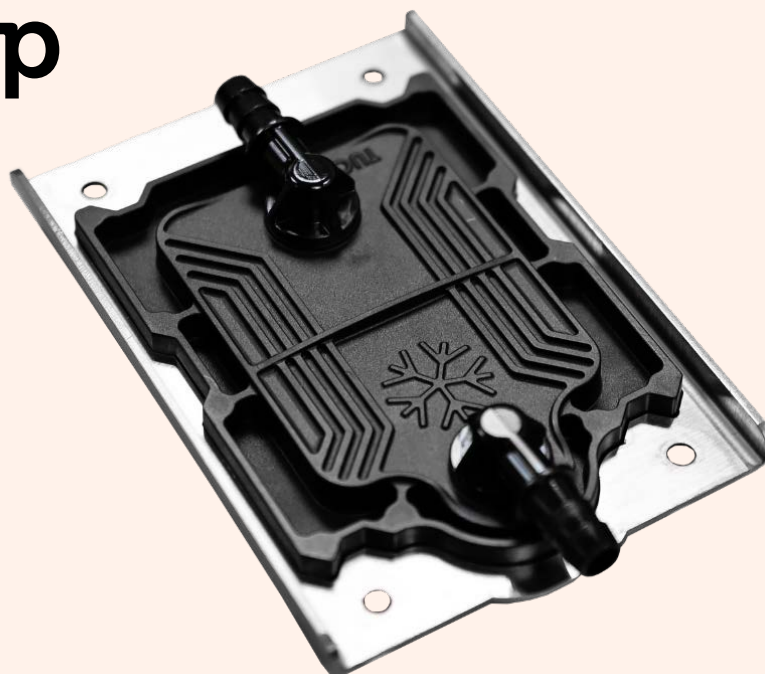
El problema del silicio oscuro se conoce desde hace mucho tiempo. En 2012, el profesor Nikos Hardavellas de la Universidad Northwestern, mencionó lo siguiente en Usenix, la revista de la Asociación de Computación Avanzada: “A falta de un milagro tecnológico, nos dirigimos hacia una era del “silicio oscuro”, capaz de construir dispositivos densos que no podemos darnos el lujo de alimentar. Sin poder utilizar más transistores ni operarlos más rápidamente, es probable que las mejoras de rendimiento se estanquen a menos que cambiemos de rumbo”.

Han existido muchas soluciones al problema, como aumentar el uso de núcleos especializados dentro de los chips, los cuales solo se usan en caso de ser necesarios, pero una posible forma de reducir el silicio oscuro sería si los fluidos pudieran fluir dentro del propio chip, donde podrían eliminar más energía y permitir un mayor flujo térmico. ●

El enfriamiento de precisión a nivel del chip



Sebastian Moss
Jefe de redacción



El director ejecutivo de JetCool explica cómo miles de pequeños chorros podrían llegar hasta un centro de datos cercano

A medida que aumentan las temperaturas de los chips y la densidad de los racks, muchas compañías se han manifestado para presentar su visión del futuro.

Las necesidades de enfriamiento de la inteligencia artificial y otras cargas de trabajo de alta densidad han superado las capacidades de los sistemas de aire y exigen algún tipo de enfriamiento líquido.

“Al pensar en el panorama del enfriamiento líquido, observamos tres categorías técnicas diferentes”, explicó Bernie Malouin, director ejecutivo de JetCool.

“Se trata de una inmersión monofásica, en el aceite. Y esto resulta interesante, pero existen algunas limitaciones en la potencia del chip. Por mucho tiempo han estado estancados en 400W. Hay quienes están tratando de mejorarla un poco, pero no tanto como se necesita”. La segunda categoría son los dieléctricos bifásicos: “Vemos a quienes manejan los procesadores [de punto de diseño térmico (TDP)] más altos, por lo cual

pueden alcanzar los 900-1000 W. Son tecnológicamente aptos para el futuro de la computación, pero se ven limitados por los productos químicos”, indicó Malouin.

Muchas soluciones bifásicas utilizan sustancias perfluoroalquilas (PFAS), también conocidas como sustancias químicas permanentes, las cuales se asocian a riesgos para la salud humana y tienen restricciones en Estados Unidos y Europa. Algunas compañías como ZutaCore se han comprometido a cambiar a otras soluciones para 2026, pero el cambio ha sido lento.

“Es una preocupación para muchos de nuestros clientes, los cuales acuden a nosotros debido a que les preocupa la seguridad de esos fluidos”, señaló Malouin. “Están preocupados por la disponibilidad continua de esos fluidos.

Además, existe una tercera categoría, las placas frías de enfriamiento líquido directo (DLC, por sus siglas en inglés). “Somos uno de ellos”, agregó Malouin, pero “existen otras”.

Las placas frías de DLC son una de las formas más antiguas de enfriamiento líquido



de TI; transportan líquido frío a las placas metálicas montadas directamente en los componentes más calientes: Estas han sido utilizadas durante mucho tiempo por la comunidad de computación de alto rendimiento, pero JetCool cree que el concepto necesita actualizarse.

En lugar de hacer pasar fluido sobre una superficie, sus chorros de enfriamiento dirigen el fluido directamente hasta la superficie de un chip. "Tenemos estos grupos de mil chorros de fluido pequeños y trabajamos directamente con los principales fabricantes de chips (Intel, AMD, Nvidias). Además, diseñamos estos chorros de manera inteligente para alinearlos con las fuentes de calor en un procesador determinado".

En lugar de tratar todo el chip como un todo con necesidades de enfriamiento únicas, el enfoque de enfriamiento microconvectivo "intenta equilibrar las cargas de calor dispares, los requisitos térmicos dispares de ciertas partes de esa pila de chips", indicó Malouin.

"Cuando se piensa en paquetes realmente integrados, los núcleos mismos podrían funcionar a una temperatura un poco más alta, pero podrían existir secciones de memoria de alto ancho de banda (HBM, por sus siglas en inglés) que no consumen tanta energía, pero tienen un límite de temperatura más bajo", agregó.

En lugar de intentar diseñar para los núcleos de alta potencia y el HBM sensible a la temperatura, cada sección puede enfriarse a un ritmo ligeramente diferente. "Esto le permite desacoplar estos elementos y tener un enfriamiento de precisión donde lo necesita", indicó Malouin.

Aunque Malouin cree que el enfriamiento líquido a nivel de instalaciones es el futuro de

los centros de datos, la compañía también cuenta con un sistema independiente para aquellos que buscan sumergirse en aguas más frías, con una asociación de Dell centrada en implementaciones de doble entrada.

Dos pequeños módulos de bombeo proporcionan la circulación del fluido y un intercambiador de calor de aire expulsa calor al otro extremo del sistema Smart Plate.

"Al agregar estas bombas, se agrega algo de consumo eléctrico, pero no es necesario que los ventiladores funcionen con tanta intensidad.

Así que hace que sea 15-20 decibelios más silencioso, y en red, extraemos alrededor de 100 W por servidor después de haber quitado la penalización a las bombas", indicó Malouin.

Además, señala que cuando se llega a 10 racks o más, pasar al nivel de instalación tiene más sentido. Cuando se le preguntó sobre la temperatura de entrada preferida, Malouin dijo que el sistema era flexible, pero agregó que "en realidad nos gustan mucho los fluidos calientes".

Según mencionó, "Hoy contamos con instalaciones que nos proveen temperaturas de enfriamiento de entrada de 60 °C (140 °F) o más, y seguimos enfriando esos dispositivos bajo carga completa". Eso no es algo común, pero Malouin cree que las aguas más cálidas ganarán popularidad en lugares como Europa debido al potencial de reutilización del calor.

En lo que respecta a EE. UU., la compañía forma parte del proyecto COOLERCHIPS del Departamento de Energía, cuyo objetivo es avanzar en gran medida en los sistemas de enfriamiento de los centros de datos.

El objetivo del premio de más de 1 millón de dólares de JetCool no es solo el potencial de enfriamiento, sino también un tentador beneficio secundario: "Tenemos casos en los que hemos hecho que el silicio sea intrínsecamente entre 8 y 10 % más eficiente desde el punto de vista eléctrico", indicó Malouin.

"Eso no tiene ninguna relación con el consumo energético del sistema de enfriamiento, sino con las fugas", agregó.

Malouin no se refiere a una fuga del sistema de enfriamiento, sino más bien al fenómeno cuántico de las corrientes de fuga de semiconductores que pueden afectar considerablemente el rendimiento de un chip.

La historia reciente del enfriamiento de centros de datos ha tendido a asumir que permitir que las temperaturas aumenten más ahorrará energía porque se utiliza menos energía en el enfriamiento. Los resultados, incluida la investigación realizada por Jon Summers en el RISE, muestran que las corrientes de fuga en el silicio limitan los beneficios de funcionar a mayor temperatura.

"Una gran parte de nuestra iniciativa de COOLERCHIPS es justificarla con evidencia científica más rigurosa y extrapolarla a diferentes entornos para ver qué es lógico y qué no", indicó Malouin.

Mirando aún más hacia el futuro, Malouin ve una oportunidad de profundizar en el silicio. "En algunos casos, podría integrarse como una capa incrustada dentro del silicio y luego acoplarla a un sistema externo de reutilización térmica. Al pensar en esto de manera integral, creemos que existe una oportunidad real para un cambio radical en la eficiencia del centro de datos".

Por el momento, la compañía afirma ser capaz de soportar los 900 W de carga de las GPU de Nvidia más grandes y actualmente está enfriando chips "personalizados" que consumen 1500 W.

"Finalmente, se deberá tomar en consideración el enfriamiento líquido si se quiere hacer funcionar no solo el futuro de la IA, sino también la IA generativa en la actualidad" ●

¿Por qué Intel está desarrollando tecnología de enfriamiento para centros de datos?



Sebastian Moss
Jefe de redacción

Los procesadores de 2 kW y el futuro de los centros de datos

Solía ser muy sencillo. Intel no es una compañía de enfriamiento y no quiere serlo, pero la compañía se ha visto obligada a ir en esa dirección debido al rápido aumento de los puntos de diseño térmico de sus procesadores.

La compañía de chips se encuentra cada vez más preocupada de que los chips del futuro sobrecarguen los centros de datos del futuro y sus investigaciones buscan determinar cómo puede ayudar.

“Queremos estar preparados. A medida que la potencia del procesador siga aumentando, los desafíos de enfriamiento se incrementan exponencialmente”, explicó Tejas Shah, ingeniero principal del grupo de plataformas de supercomputación de Intel. A manera de respuesta, Intel intenta cambiar radicalmente el enfriamiento del centro de datos.

Las crecientes densidades de energía han llevado a los desarrolladores a utilizar el enfriamiento por inmersión y algunos están tomando en consideración la inmersión en dos fases, en la cual los fluidos en ebullición eliminan aún más calor. Shah lidera un esfuerzo en estos sistemas con disipadores de calor en forma de coral de resistencia térmica ultrabaja e integrados con una cavidad de cámara de vapor en 3D.

Este fue uno de los 15 proyectos elegidos por COOLERCHIPS, un programa administrado por la ARPA-E (Agencia de Proyectos de Investigación Avanzados-Energía) de EE. UU, con el objetivo de cambiar radicalmente la forma de enfriar los centros de datos, pero antes de entrar en tecnología, hablemos del elefante en la sala; los fluidos de inmersión de dos fases han resultado controvertidos. “No estamos desarrollando fluidos, usaremos los fluidos estándar de la industria”, indicó Shah.

“Somos conscientes del problema

con todas las PFAS y sus regulaciones”, reconoció Shah, en referencia a las tóxicas sustancias químicas permanentes utilizadas por la mayoría de las soluciones bifásicas. “Básicamente, 3M está saliendo del negocio de la inmersión en dos fases; estamos trabajando con algunos de los proveedores de fluidos”.

El otro problema se relaciona con el potencial de calentamiento global (GWP) de muchos de los fluidos, el cual podría compensar cualquier ahorro de eficiencia energética de la inmersión. “Sacar las PFAS es el mayor desafío, pero tenemos la seguridad de cumplir con los requisitos de GWP”, afirmó Shah.

Hablemos de nuevo de la tecnología. La compañía está desarrollando disipadores térmicos con forma de coral por medio de la fabricación aditiva o impresión 3D.

“Básicamente se busca optimizar el diseño para que siga una función matemática específica”, indicó Shah. “De manera que se trata de la optimización de la topología, un procedimiento matemático iterativo para optimizar una determinada función de costo, independientemente de cuál sea. En este caso, se busca reducir la resistencia térmica y asegurarse de llegar a los puntos críticos correctos”, agregó.

Los disipadores térmicos están acoplados a cavidades de cámaras de vapor 3D: bolsas metálicas planas y selladas llenas de líquido para distribuir la capacidad de ebullición del sistema de dos fases.

“La forma del coral aumenta la superficie, pero también queremos reducir la resistencia a la expansión, por esta razón hay una cámara de vapor integrada dentro de la estructura 3D”, indicó Shah.

Cuando la corriente fluye a través de un material semiconductor, encuentra una “resistencia a la expansión” donde los electrones se extienden a través de

un área más amplia que la nucleación. “Esto contribuye con la ebullición que ocurre en la superficie”, señaló Shah.

Con los tres enfoques combinados, la empresa puede enfriar los procesadores hasta un punto de diseño térmico de 2 kW. Esto va más allá de lo que el mercado actual ofrece o que probablemente ofrecerá en los próximos años, aunque el Superchip Grace Hopper de Nvidia tiene una TDP de 450 W a 1 kW al combinar una CPU y una GPU.

De acuerdo con Shah, el sistema de inmersión puede soportar “al menos 8 procesadores” con un consumo de 2 kW cada uno y aumentar el número en un rack se convierte más en una limitación energética que térmica.

Aunque las pruebas son prometedoras, superar todos los problemas no es una conclusión predecible, especialmente si el sector de inmersión no puede superar el problema de las PFAS —el proveedor ZutaCore afirma que eliminará las PFAS en 2026, pero aún no ha proporcionado detalles específicos, mientras que su rival, LiquidCool, está diversificando hacia inmersión monofásica.

“No es algo fácil; supone todo un desafío”, mencionó Shah sobre el proyecto y otros esfuerzos de Intel. “Pero el objetivo es reducir la resistencia térmica y apuntar a la próxima generación de procesadores con la mejor solución térmica existente”, agregó.

Sin embargo, si funciona, Intel podría posicionarse como un proveedor de enfriamiento. “Lo que ‘podríamos’ terminar haciendo —y es una gran posibilidad—, pero podríamos incluso vender la solución térmica con el procesador inicialmente”, indicó Shah.

“Esto construirá el ecosistema y la próxima generación podría ponerse al día con nuestros procesadores”. ●



La IA está que arde.

Asegúrese de que
su centro de datos
esté frío.

Adopte estrategias de enfriamiento híbrido para hacer
frente a las crecientes densidades de TI hoy mismo.



Descubra cómo podemos ayudarlo en [Vertiv.com/GetCool](https://www.vertiv.com/GetCool)