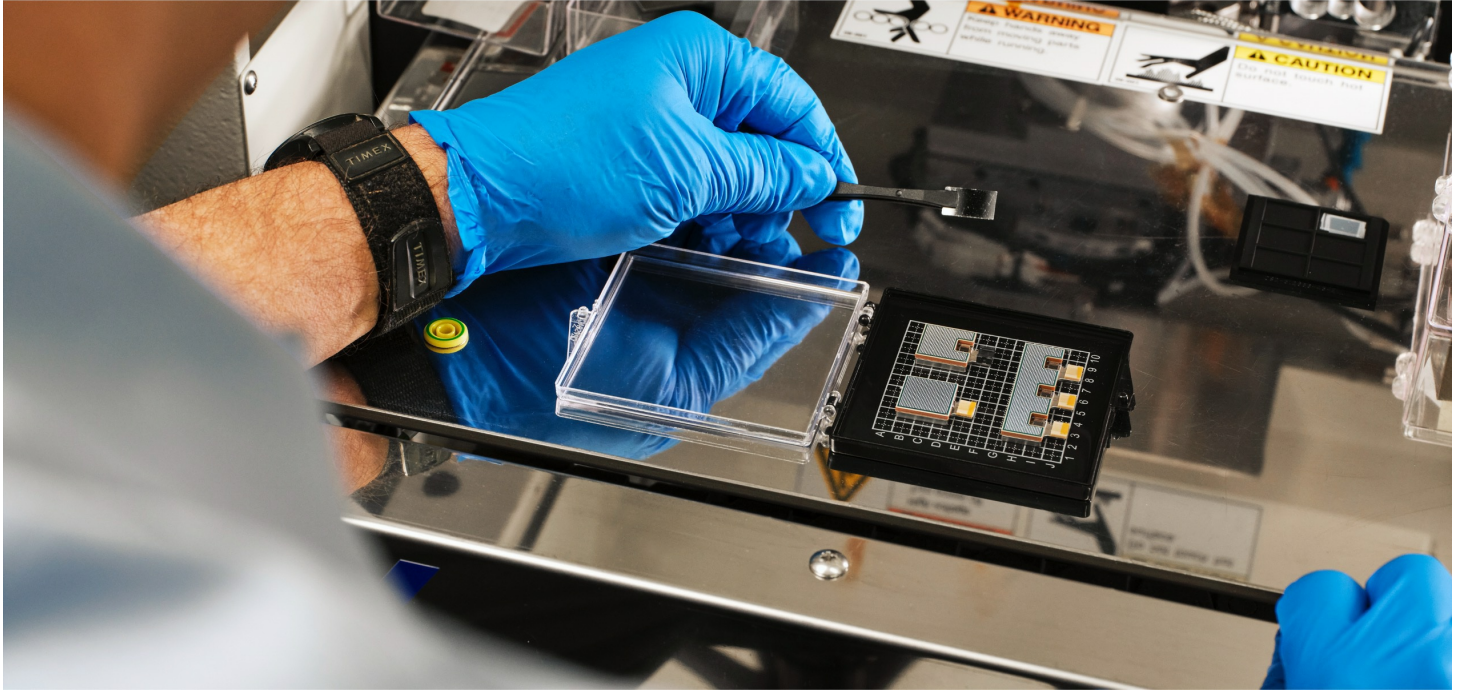


IBM lleva la velocidad de la luz a la era de la IA generativa con un avance en óptica

Una nueva innovación en óptica coempaquetada podría sustituir los interconectores eléctricos en los centros de datos, ofreciendo mejoras significativas en velocidad y eficiencia energética para la IA y otras aplicaciones informáticas.



YORKTOWN HEIGHTS, N.Y. – 9 de diciembre de 2024: IBM (NYSE: IBM) ha presentado un avance revolucionario en tecnología óptica que podría transformar drásticamente la forma en que los centros de datos entrenan y ejecutan modelos de IA generativa. Los investigadores han desarrollado un nuevo proceso para la tecnología de óptica coempaquetada (CPO, por sus siglas en inglés), la próxima generación de tecnología óptica, que permite la conectividad dentro de los centros de datos a la velocidad de la luz mediante óptica, complementando los cables eléctricos de corto alcance actuales. Al diseñar y ensamblar la primera exitosa guía de ondas óptica polimérica (PWG) anunciada públicamente para alimentar esta tecnología, los investigadores de IBM han demostrado cómo el CPO redefinirá la transmisión de datos de alta capacidad entre chips, placas de circuitos y servidores en la industria informática.

“ A medida que la IA generativa demanda más energía y potencia de procesamiento, los centros de datos deben evolucionar, y la óptica coempaquetada puede hacerlos a prueba de futuro ”

Actualmente, la tecnología de fibra óptica transporta datos a altas velocidades a largas distancias, gestionando la mayor parte del tráfico global de comercio y comunicaciones mediante luz en lugar de electricidad. Aunque los centros de datos utilizan fibra óptica para sus redes de comunicaciones externas, dentro de los racks los datos se comunican predominantemente a través de cables eléctricos de cobre. Estos cables conectan aceleradores GPU que pueden permanecer inactivos más de la mitad del tiempo, esperando datos de otros dispositivos en un proceso de entrenamiento distribuido, lo que genera costes y consumo de energía significativos.

Los investigadores de IBM han demostrado una forma de llevar la velocidad y capacidad de la óptica al interior de los centros de datos. En un artículo recientemente publicado en arXiv, IBM presenta un nuevo módulo prototipo de CPO que habilita la conectividad óptica de alta velocidad. Esta tecnología podría aumentar significativamente el ancho de banda de las comunicaciones en los centros de datos, minimizando el tiempo de inactividad de las GPUs y acelerando drásticamente el

procesamiento de IA. Según lo descrito, esta innovación permitiría:

- **Reducción de costes y consumo energético:** 5 veces menos consumo energético en comparación con interconectores eléctricos de rango medio, además de extender la longitud de los cables interconectores en centros de datos de uno a cientos de metros.
- **Entrenamiento más rápido de modelos de IA:** Los desarrolladores podrían entrenar un modelo de lenguaje grande (LLM) hasta cinco veces más rápido con CPO que con cableado eléctrico convencional, reduciendo el tiempo de entrenamiento de tres meses a tres semanas, con mayores ganancias en modelos más grandes y más GPUs.
- **Aumento significativo en la eficiencia energética:** Ahorros equivalentes al consumo anual de energía de 5.000 hogares estadounidenses por cada modelo de IA entrenado.

"A medida que la IA generativa demanda más energía y potencia de procesamiento, los centros de datos deben evolucionar, y la óptica coempaquetada puede hacerlos a prueba de futuro", afirmó Darío Gil, vicepresidente senior y director de investigación de IBM. "Con este avance, los chips del futuro se comunicarán de manera similar a cómo los cables de fibra óptica transportan datos hacia y desde los centros de datos, inaugurando una nueva era de comunicaciones más rápidas y sostenibles capaces de manejar las cargas de trabajo de IA del futuro".

Ochenta veces más ancho de banda que las comunicaciones entre chips actuales

En los últimos años, los avances en la tecnología de chips han incrementado significativamente la densidad de transistores; la tecnología de chips de 2 nanómetros de IBM puede contener más de 50.000 millones de transistores. La tecnología CPO busca escalar la densidad de interconexión entre aceleradores al permitir que los fabricantes de chips añadan rutas ópticas que conecten chips en un módulo electrónico, superando los límites de las rutas eléctricas actuales. Según el artículo de IBM, estas nuevas estructuras ópticas de alta densidad, combinadas con la transmisión de múltiples longitudes de onda por canal óptico, tienen el potencial de aumentar hasta 80 veces el ancho de banda entre chips en comparación con las conexiones eléctricas.

La innovación de IBM permitiría a los fabricantes de chips añadir seis veces más fibras ópticas en el borde de un chip de fotónica de silicio, conocido como "densidad de línea costera", en comparación con la tecnología CPO actual. Cada fibra, de aproximadamente tres veces el ancho de un cabello humano, podría abarcar desde centímetros hasta cientos de metros y transmitir terabits de datos por segundo. El equipo de IBM ensambló una guía de ondas óptica polimérica de alta densidad con canales ópticos de 50 micrómetros de separación, acoplados adiabáticamente a guías de onda de fotónica de silicio, utilizando procesos estándar de ensamblaje y empaquetado.

El artículo también señala que estos módulos CPO con PWG de 50 micrómetros de separación son los primeros en superar todas las pruebas de resistencia requeridas para fabricación. Los componentes fueron sometidos a entornos de alta humedad y temperaturas que varían entre -40 °C y 125 °C, así como pruebas de durabilidad mecánica para confirmar que los interconectores ópticos pueden doblarse sin romperse o perder datos. Además, los investigadores demostraron tecnología PWG con una separación de 18 micrómetros. Apilando cuatro PWG se podrían habilitar hasta 128 canales de conectividad con esa separación.

IBM continua su liderazgo en I+D de semiconductores

La tecnología CPO abre un nuevo camino para satisfacer las crecientes demandas de rendimiento de la IA, con el potencial de reemplazar las comunicaciones fuera del módulo de eléctricas a ópticas. Este avance se suma a la historia de liderazgo de IBM en innovación de semiconductores, que incluye la primera tecnología de chips de 2 nm, la primera implementación de procesos tecnológicos de 7 nm y 5 nm, transistores Nanosheet, transistores verticales (VTFET), memoria DRAM de celda única y fotoresistencias químicamente amplificadas.

Los investigadores completaron el diseño, modelado y simulación para el CPO en Albany, Nueva York, donde el Departamento de Comercio de los Estados Unidos seleccionó recientemente la ubicación para el primer Centro Nacional de Tecnología de Semiconductores (NSTC, por sus siglas en inglés). Los prototipos fueron ensamblados y probados en la planta de IBM en Bromont, Quebec, una de las mayores instalaciones de ensamblaje y prueba de chips en América del Norte. Parte del Corredor de Semiconductores del Noreste entre Estados Unidos y Canadá, la planta de Bromont ha liderado durante décadas la industria del empaquetado de chips.

Sobre IBM

IBM es un proveedor líder de nube híbrida global e inteligencia artificial, y experiencia en consultoría. Ayudamos a clientes en más de 175 países a aprovechar el conocimiento de sus datos, optimizar los procesos de negocios, reducir costes y obtener una ventaja competitiva en sus industrias. Más de 4.000 entidades gubernamentales y corporativas en áreas de infraestructura crítica como servicios financieros, telecomunicaciones y atención médica confían en la plataforma de nube híbrida de IBM y Red Hat OpenShift para lograr sus transformaciones digitales de manera rápida, eficiente y segura. Las innovaciones revolucionarias de IBM en inteligencia artificial, computación cuántica, soluciones de nube específicas de la industria y consultoría ofrecen opciones abiertas y flexibles a nuestros clientes. Todo esto está respaldado por el compromiso de larga data de IBM con la confianza, la transparencia, la responsabilidad, la inclusión y el servicio. Visite www.ibm.com para obtener más información.

For further information: Camila Cuetos, Dpto. Comunicación, camila.cuetos2@ibm.com
