Microarquitectura Intel Core vs AMD Zen

Introducción

El procesador “Zen 2” está diseñado para satisfacer las necesidades de diversos mercados, como servidores, ordenadores de sobremesa, dispositivos móviles y estaciones de trabajo. El núcleo ofrece mejoras significativas de rendimiento y eficiencia energética respecto a Zen gracias a cambios en la microarquitectura, que incluyen un nuevo predictor de ramificación TAGE, una caché de operaciones de doble tamaño y una unidad de punto flotante de doble ancho. Basándose en el diseño del núcleo, un enfoque modular de chiplets proporciona flexibilidad y escalabilidad hasta 64 núcleos por zócalo con un total de 256 MB de caché L3[1].

Los modelos de rendimiento son fundamentales para optimizar el código sensible al rendimiento. Al modelar el uso de unidades funcionales de CPU x86-64 fuera de servicio, la disponibilidad de datos varía según el fabricante: existen asignaciones de instrucciones a puertos para los procesadores Intel, mientras que la información para los diseños de AMD es insuficiente. Esta disparidad se debe a que las técnicas estándar para inferir asignaciones de puertos exactas requieren contadores de rendimiento de hardware que AMD no proporciona[2].

La inferencia de asignaciones de puertos ha sido objeto de investigación reciente: uops.info proporciona asignaciones de puertos precisas para las microarquitecturas de Intel con microbenchmarks que analizan el uso del puerto de cada instrucción. Sin embargo, este enfoque no abarca otras microarquitecturas como las recientes arquitecturas Zen de AMD, ya que uops.info se basa en los contadores de rendimiento de hardware por puerto de Intel [2].

CPU Core

El núcleo de CPU Zen 2, que se muestra en la Figura 1, presenta dos mejoras principales con respecto a su predecesor, Zen. En primer lugar, la eficiencia energética se duplica gracias a una combinación de mejoras tecnológicas y de microarquitectura. En segundo lugar, el IPC aumenta aproximadamente un 15 % gracias a los cambios de microarquitectura en el front-end ordenado, la ejecución de enteros, la ejecución de coma flotante/vectorial, la carga/almacenamiento y la jerarquía de caché[1].

Floating-Point/Vector Execute

El motor de coma flotante/vectorial Zen 2 ha duplicado el ancho de la ruta de datos de 128 bits (Zen) a 256 bits. Ambos núcleos admiten instrucciones AVX-256, pero Zen realiza operaciones de doble bombeo utilizando sus rutas de datos de 128 bits, mientras que Zen 2 admite operaciones nativas con sus rutas de datos de 256 bits. El ancho de la PRF vectorial también se duplica a 256 bits. Los registros ahora pueden renombrarse con una granularidad de 256 bits en lugar de 128 bits. Por lo tanto, la capacidad efectiva de la PRF vectorial se duplica para el código AVX-256, aunque el número de entradas de la PRF vectorial se mantiene en 160[1].

Contribuciones

En resumen, ofrecemos las siguientes contribuciones:

• Un algoritmo de inferencia explicable para la asignación de puertos que no requiere los contadores de rendimiento por puerto de Intel y

• Una implementación que evaluamos en la arquitectura Zen+ de AMD, que anteriormente estaba fuera del alcance de los algoritmos de inferencia de asignación de puertos explicables.

• El resultado es, a nuestro leal saber y entender, la asignación de puertos más completa y precisa disponible para Zen+.

• Nuestro caso práctico documenta numerosos aspectos de Zen+ que no estaban documentados previamente o que estaban mal documentados[2].

Diseño multinúcleo eficiente

La línea Intel Core, derivada de la serie de procesadores Pentium M, se lanzó en 2006 y contó con la aprobación del 23% de los votantes (véase la Figura 4). La arquitectura Core sigue siendo la base de varios procesadores actuales dirigidos a los mercados de ordenadores de sobremesa y móviles. Mientras tanto, el AMD Opteron quedó en segundo lugar, con el 17% de los votos; sin embargo, cabe destacar que los procesadores Intel se llevaron tres de los cuatro primeros puestos[3].

Algunos admiradores del AMD Opteron escribieron:

• AMD64 cambió las reglas del juego y salvó el mundo (de x86).

• Opteron, en cierto modo, fue responsable de impulsar a toda la industria hacia los 64 bits.

Los comentarios sobre la serie Intel Core incluyen:

• Un excelente ejemplo de la compensación entre potencia y rendimiento.

• Rendimiento y eficiencia energética que revolucionaron a Intel y dominaron la industria durante una década.

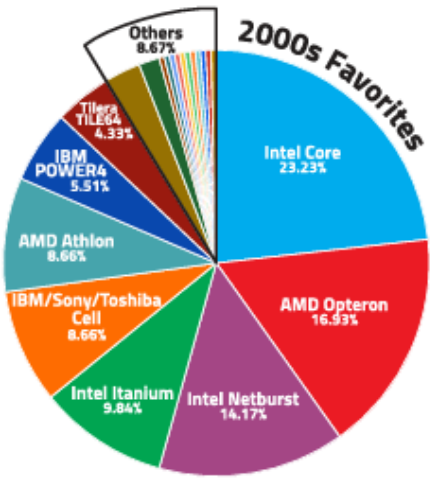

FIGURE 4. Favorite microprocessors from the 2000s.

Década de 2010: Arquitectura heterogénea

La racha de Intel finalmente llegó a su fin en la década de 2010, con AMD Zen (26%) y Apple M1 (24%) a la cabeza (véase la Figura 5). La arquitectura AMD Zen se introdujo en 2017 con la serie Ryzen y su diseño altamente flexible basado en chiplets ha permitido su implementación en portátiles de gama baja y servidores de gama alta[3].

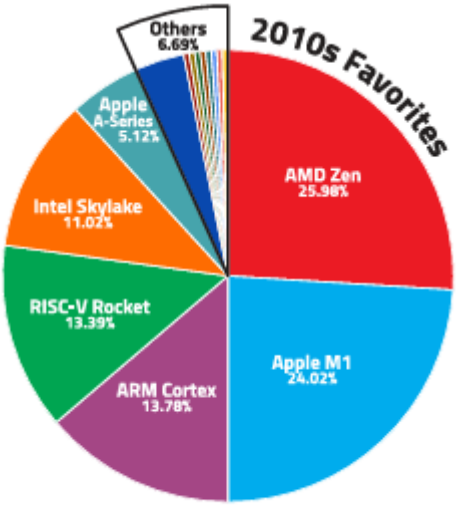

FIGURE 5. Favorite microprocessors from the 2010s.

La lista de favoritos de todos los tiempos sorprendió a todos, con DEC Alpha siendo elegido el número uno, superando por poco a AMD Zen (véase la Figura 6). Aunque no fue elegido el favorito en su propia década, el impacto duradero del procesador DEC Alpha en los procesadores modernos es evidente por su popularidad incluso hoy en día. AMD Zen quedó en segundo lugar, posiblemente debido a su popularidad desde la década de 2010 hasta la actualidad[3].

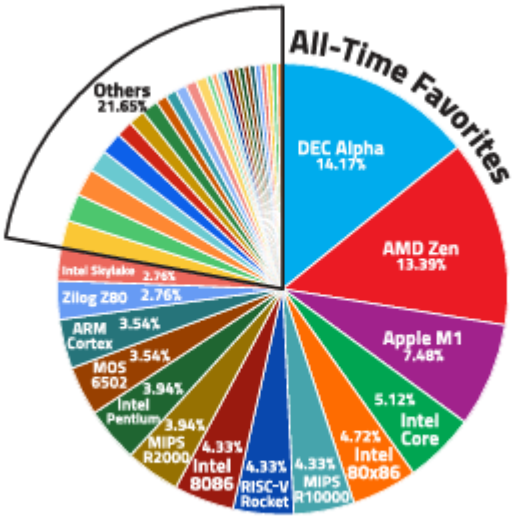

FIGURE 6. Favorite microprocessors of all time.

Arquitectura modular de Zen 2

Zen 2 utiliza un diseño modular en múltiples niveles [2, Sección 1.8.1]. La estructura del procesador se muestra en la Figura 2. Cuatro núcleos se agrupan en un Complejo de Núcleos (CCX, también Complejo de CPU). Un Matriz de Complejo de Núcleos (CCD) consta de dos CCX. Hasta ocho CCD se conectan a una matriz de E/S. (a) Matriz de Complejo de Núcleos (CCD) con Complejos de Núcleos (CCX) en procesadores de hasta 64 núcleos. Según el número de núcleos, dos o uno de los CCD se conectan al mismo conmutador dentro de la red de matrices de E/S. Cada conmutador de la matriz de E/S que conecta los CCD también conecta un controlador de memoria con dos canales de memoria, lo que puede generar cuatro nodos de acceso a memoria no uniforme (NUMA)[4].

Cachés en Zen 2

Cada núcleo tiene un front-end común que obtiene instrucciones para dos hilos de hardware independientes [17]. La ventana de obtención tiene un ancho de 32B y alimenta a un decodificador de 4 vías. El back-end se divide en dos partes: una parte comprende cuatro Unidades Lógicas Aritméticas y tres Unidades de Generación de Direcciones (AGU), la otra contiene dos unidades de Multiplicación-Adición de Punto Flotante (FMA) de 256 bits de ancho y dos unidades de adición de Punto Flotante de 256 bits de ancho. Las AGU se pueden usar para dos cargas y un almacenamiento por ciclo, donde cada una de estas puede transferir hasta 32B de datos. Cada núcleo del procesador contiene una caché de operaciones para 4096 operaciones, cachés L1I y L1D de 32 KiB, y una caché L2 de 512 KiB, que se utiliza para instrucciones y datos. Además, cada CCX contiene 16 MiB de caché L3, distribuidos en cuatro porciones con 4 MiB cada una [4].

Uso de micro-op cache como optimización

Los procesadores modernos Intel, AMD y ARM traducen instrucciones complejas en microoperaciones internas más sencillas que se almacenan en caché en una estructura dedicada en chip denominada caché de microoperaciones. Este trabajo presenta un estudio exhaustivo de la caracterización de la caché de microoperaciones, aplicando ingeniería inversa a muchas características no documentadas y describe, además, ataques que la explotan como canal de temporización para transmitir información secreta [5].

Comparación de decodificadores Intel Skylake vs AMD Zen

La microarquitectura Skylake incluye: (a) múltiples decodificadores 1:1 que pueden traducir macrooperaciones simples que solo se descomponen en una microoperación; (b) un decodificador 1:4 que puede traducir macrooperaciones complejas que pueden descomponerse en entre una y cuatro microoperaciones; y (c) una ROM de microsecuenciación (MSROM) que traduce instrucciones microcodificadas más complejas. Una sola macrooperación puede traducirse en más de cuatro microoperaciones, lo que potencialmente implica múltiples ramificaciones y bucles, ocupando varios ciclos de decodificación. El pipeline de decodificación puede proporcionar un ancho de banda máximo de 5 microoperaciones por ciclo [33]. Por el contrario, AMD Zen cuenta con cuatro decodificadores 1:2 y relega a una ROM de microcódigo cuando encuentra una instrucción compleja que se traduce en más de dos microoperaciones[5].

Política de compartición en SMT: Intel vs AMD

En las arquitecturas Intel, la cola de operaciones macro, IDQ, y el caché de operaciones micro permanecen particionados entre diferentes subprocesos SMT que se ejecutan en el mismo núcleo físico, mientras que AMD permite que el caché de operaciones micro se comparta de manera competitiva entre los subprocesos SMT ubicados en el mismo lugar, al igual que el resto de las estructuras en la tubería de decodificación, incluidos los decodificadores y la MSROM[5].

Bibliografía

[1] D. Suggs, M. Subramony, and D. Bouvier, "The AMD 'Zen 2' Processor," *IEEE Micro*, vol. 40, no. 2, pp. 45–52, Mar. 2020, doi: 10.1109/MM.2020.2974217.

[2] F. Ritter and S. Hack, "Explainable Port Mapping Inference with Sparse Performance Counters for AMD's Zen Architectures," in *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3*, New York, NY, USA: ACM, Apr. 2024, pp. 317–330. doi: 10.1145/3620666.3651363.

[3] B. Hanindhito, K. Swaminathan, V. Narayanan, and L. K. John, "Intel Wins in Four Decades, but AMD Catches Up," *IEEE Micro*, vol. 41, no. 6, pp. 168–171, Nov. 2021, doi: 10.1109/MM.2021.3119825.

[4] R. Schone, T. Ilsche, M. Bielert, M. Velten, M. Schmidl, and D. Hackenberg, "Energy Efficiency Aspects of the AMD Zen 2 Architecture," in *2021 IEEE International Conference on Cluster Computing (CLUSTER)*, IEEE, Sep. 2021, pp. 562–571. doi: 10.1109/Cluster48925.2021.00087.

[5] X. Ren, L. Moody, M. Taram, M. Jordan, D. M. Tullsen, and A. Venkat, "I See Dead µops: Leaking Secrets via Intel/AMD Micro-Op Caches," in *2021 ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA)*, IEEE, Jun. 2021, pp. 361–374. doi: 10.1109/ISCA52012.2021.00036.