

Resumen

Conforme se incrementa el número de núcleos en las nuevas generaciones de multiprocesadores en chip, los CMPs deben de escalar en prestaciones, área y consumo energético para cumplir con las demandas de un número núcleos mayor. Los protocolos basados en directorio constituyen la alternativa más escalable. Un directorio convencional, no obstante, sufre de una utilización ineficiente de almacenamiento y energía. En primer lugar, los grandes y poco escalables vectores de compartidores consumen una cantidad de energía de fuga y de área innecesaria, especialmente si se tiene en consideración que la mayoría de los bloques en un directorio solo se encuentran en la cache de un único núcleo. En segundo lugar, aunque incrementar el tamaño y la asociatividad del directorio aumentaría las prestaciones del sistema, esto supondría un incremento notable en el consumo energético.

Esta tesis estudia las diferencias significativas entre el comportamiento de bloques privados y compartidos en el directorio, lo que nos lleva hacia una gestión separada para cada uno de los tipos de bloque. Proponemos el PS-Directory, una cache de directorio de dos niveles que mantiene el reducido número de las entradas compartidas, que son los que se acceden con más frecuencia, en una estructura pequeña de primer nivel (concretamente, la *Shared Directory Cache*) y que utiliza una estructura más grande y lenta en el segundo nivel (*Private Directory Cache*) para poder mantener la información de los bloques privados. Las entradas en la Private Directory Cache no implementan el vector de compartidores, lo que conlleva importantes ahorros de energía y área. Debido a temas de área y latencia, se nos sugiere la utilización de tecnología eDRAM, mucho más densa pero más lenta que la tecnología SRAM, para la Private Directory Cache, consiguiendo así ahorros de energía mayores. Los resultados experimentales muestran que, comparado con un directorio convencional, el PS-Directory consigue mejorar las prestaciones a la vez que reduce el área de silicio y el consumo energético.

Ya que el ratio compartido/privado de las entradas en el directorio varía entre aplicaciones y entre las diferentes fases de ejecución dentro de las aplicaciones, proponemos el Dynamic Way Partitioning (DWP) Directory. El DWP-Directory reduce el número de

vías que almacenan entradas compartidas y permite que éstas se enciendan o apaguen en tiempo de ejecución según los requisitos dinámicos de las aplicaciones según un algoritmo de reparticionado. Los resultados muestran unas prestaciones similares a un directorio tradicional de alta asociatividad y un área similar a otros esquemas recientes del estado del arte. Adicionalmente, el DWP-Directory obtiene importantes reducciones de consumo estático y dinámico.

Esta disertación también se enfrenta a los problemas de escalabilidad que se pueden encontrar en las memorias cache. Las caches on-chip consumen una parte significativa del consumo total del sistema. Estas caches implementan un alto nivel de asociatividad (las caches L1 ya se implementan con ocho vías para potenciar las prestaciones del sistema). En un acceso a la cache, se accede a cada vía del conjunto en paralelo, siendo así una acción costosa en energía. Esta tesis presenta la arquitectura PS-Cache, un diseño energéticamente eficiente que reduce el número de vías accedidas sin perjudicar las prestaciones. La PS-Cache utiliza la información del estado privado-compartido del bloque referenciado para reducir la energía, ya que tan solo accedemos a un subconjunto de las vías que mantienen los bloques del tipo solicitado. Los resultados muestran unos importantes ahorros de energía dinámica.

Finalmente, proponemos otro diseño de arquitectura energéticamente eficiente que se puede aplicar a cualquier tipo de memoria cache asociativa por conjuntos y no solo a caches de procesador. La propuesta, la Tag Filter (TF) Architecture, filtra las vías accedidas en el conjunto de la cache, de manera que solo se mira un número reducido de vías tanto en el array de etiquetas como en el de datos. Esto permite que nuestra propuesta reduzca el consumo de energía dinámico de las caches sin perjudicar su tiempo de acceso. Para esto, la arquitectura sugerida mantiene los X bits menos significativos de cada etiqueta en una estructura auxiliar. Estos bits se utilizan para filtrar aquellas vías en las que estos bits menos significativos de la etiqueta no se correspondan. Los resultados experimentales muestran que este mecanismo de filtrado es capaz de obtener un consumo energético en caches asociativas por conjunto similar de las caches de mapeado directo.

Nos gustaría señalar que los esquemas propuestos han sido evaluados y comparados contra otras propuestas del estado del arte en prestaciones, energía y área. Los resultados

experimentales muestran que las propuestas presentadas en esta tesis consiguen un buen compromiso entre estos tres importantes pilares de diseño.