

Power Estimation and Power Optimization Policies for Processor-Based Systems

José L. Ayala Rodrigo
Dr. Ingeniero de Telecomunicación

Universidad Politécnica de Madrid
Escuela Técnica Superior de Ingenieros de Telecomunicación
Departamento de Ingeniería Electrónica

Madrid, 13 de marzo de 2006

1. Justificación

Los continuos avances en tecnología de semiconductores han conducido a dramáticos incrementos de prestaciones en los microprocesadores de propósito general y los sistemas empotrados. Dichas mejoras de prestaciones se pueden explicar tanto por las frecuencias de reloj alcanzadas como por las técnicas que se han desarrollado para explotar el paralelismo a nivel de instrucción y la localidad de memoria usando los transistores adicionales disponibles en cada nueva generación tecnológica. Sin embargo, como consecuencia negativa, este incremento de prestaciones también se traduce en un incremento en la potencia disipada, debido al hecho de que la potencia dinámica es directamente proporcional a la frecuencia de reloj y la capacidad de conmutación (la cual aumenta con el número de dispositivos integrados en el mismo chip). Por tanto, a pesar de los exhaustivos intentos de reducir las tensiones de alimentación en los dispositivos, y diseñar circuitos con orientación al bajo consumo, los niveles de disipación de potencia han venido incrementándose sin descanso con cada nueva generación de microprocesadores. Más aún, en la actualidad aparece un nuevo problema de diseño electrónico ya que el ahorro de consumo alcanzable con las técnicas de bajo nivel está alcanzando su máximo teórico.

La complejidad de las aplicaciones que se ejecutan en la actualidad, con un auge asombroso de las aplicaciones multimedia y de procesado en tiempo real, así como la necesidad de desarrollar sistemas flexibles y con amplias capacidades de expansión, han motivado que la práctica totalidad de los sistemas electrónicos que podemos encontrar estén basados en uno o varios procesadores integrados. Un claro ejemplo de ello son los sistemas empotrados que podemos encontrar en sistemas de alarma, de detección, en equipos de telefonía móvil, en dispositivos de computación portátil, etc. Todos ellos han visto incrementada su operatividad en gran medida, siendo capaces de ejecutar complejas aplicaciones e interactuar con diversos sensores e interfaces (reconocedores de voz, pantallas táctiles, teclados, infrarrojos...) gracias a que cuentan con procesadores empotrados de elevadas prestaciones. Así mismo, de esta forma se facilita la actualización frecuente de dichos sistemas (mediante actualizaciones de software, sistema operativo, o *firmware*) que aumenten su tiempo de vida y permitan la interacción con novedosos periféricos y accesorios.

Éste es el contexto en el que se enmarca esta Tesis Doctoral, abordando el diseño de sistemas complejos basados en procesador desde el punto de vista de la minimización de consumo.

Esta sección es una breve justificación de los aspectos más destacados del desarrollo de técnicas de estimación y reducción de consumo en sistemas basados en procesador. La idea consiste en cubrir el espectro completo de dichos sistemas basados en procesadores: desde los procesadores de propósito general, a los sistemas empotrados de alta especialización y los sistemas multi-procesador.

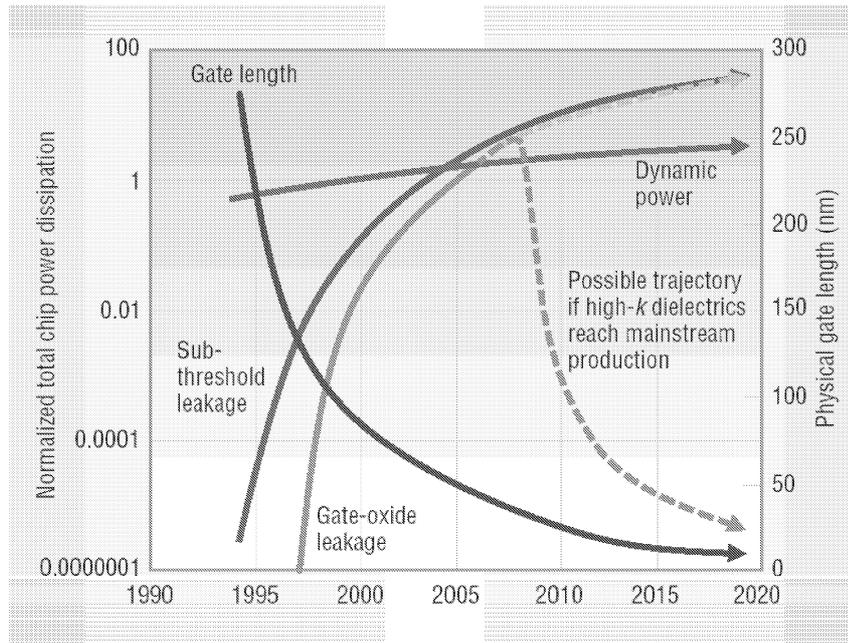


Figura 1: Predicciones de la ITRS

1.1. Introducción

Los avances y mejoras en tecnología de fabricación de circuitos integrados han permitido más que duplicar el número de transistores integrados en cada generación de procesadores. Al mismo tiempo, la corriente de fugas también se incrementa con cada generación de microprocesadores. De esta forma, el consumo de energía en el sistema en general y en los dispositivos de memoria en particular (unidad de memoria principal, cachés, bancos de registros, etc.) aumentará dramáticamente con los futuros procesos tecnológicos.

El consumo energético de los sistemas electrónicos se clasifica en potencia disipada estáticamente y potencia disipada dinámicamente. Históricamente, los dispositivos CMOS han disipado mucha menos potencia que tecnologías previas. De hecho, tradicionalmente se consideraba que cuando los transistores CMOS no conmutaban, su disipación energética (potencia estática) era despreciable. Sin embargo, esta potencia estática disipada ha incrementado sensiblemente con la velocidad del dispositivo y la densidad de integración.

La figura 1 muestra la tendencia del consumo estático y dinámico basándose en estadísticas del año 2002 y normalizada según las directrices del *International Technology Roadmap for Semiconductors* (ITRS) del año 2001. ITRS prevé un descenso en el consumo dinámico por dispositivo. Sin embargo, si asumimos que la densidad de integración se dobla cada dos años, el consumo dinámico total del chip se verá incrementado.

La figura 1 también muestra el incremento exponencial que se prevé para los dos componentes principales del consumo estático: la corriente sub-umbral (debida a la corriente inversa a través del dispositivo); y la corriente de fugas de la puerta (corriente que aparece a través del óxido aislante de la puerta).

ITRS espera que la pendiente de estos incrementos se estabilice en el 2006 pero, aun así, permanecerá en valores significativos. Incluso hoy en día, la disipación total en el chip debida a corrientes de fugas se aproxima a la debida a consumo dinámico. Además, se prevé que el consumo estático supere al dinámico para tecnologías sub-micrónicas por debajo de 65 nm.

Por otro lado, los sistemas basados en procesador implementan aplicaciones cada vez más complejas. La complejidad de estas aplicaciones conduce a un incremento en las necesidades de recursos de computación. Sin duda, uno de los mayores problemas debido a este incremento es el consumo total de potencia en el sistema. De hecho, el consumo de potencia tiene diversas e importantes repercusiones en el sistema, ya que incrementa el tamaño de éste, su peso, y la complejidad y coste del sistema de ventilación y las

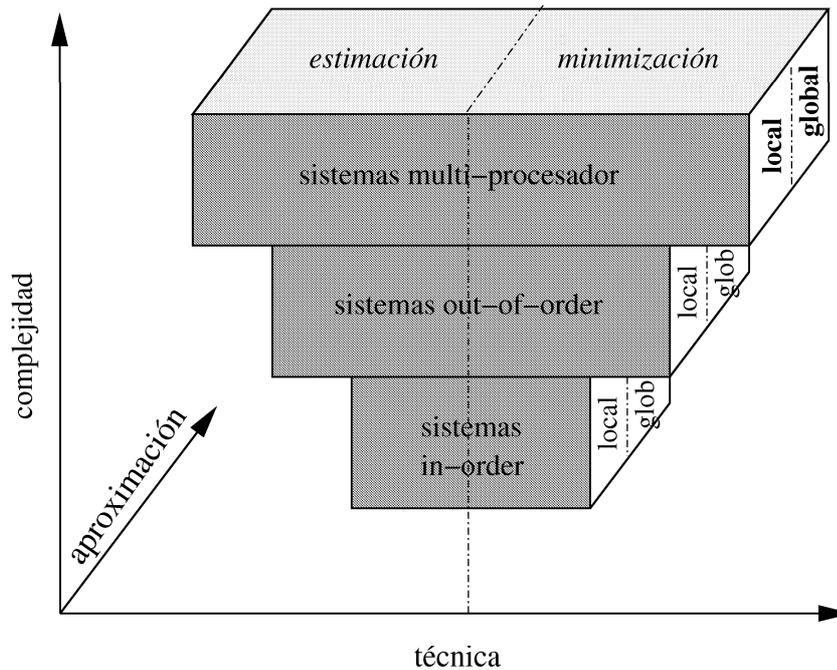


Figura 2: Visión esquemática de los objetivos de la Tesis

baterías.

En las próximas secciones se abordarán los diversos objetivos que se han tratado en esta Tesis Doctoral. En concreto, la Tesis aborda la estimación y reducción de consumo en sistemas basados en procesador desde diferentes puntos de vista, proponiendo herramientas de estimación y técnicas de compilación y modificaciones arquitecturales para la reducción del consumo energético. En primer lugar, se han abordado sistemas procesadores de reducida/media complejidad como son los sistemas empotrados con mono-procesador de ejecución “en orden” (conocidos como *in-order*, donde se han desarrollado técnicas de estimación de consumo para la jerarquía caché y donde se han propuesto mecanismos de reducción energética. Seguidamente, se han propuesto técnicas de reducción de consumo en sistemas procesadores de media complejidad, como son los sistemas empotrados con mono-procesador de ejecución “fuera de orden” (conocidos como *out-of-order*). Finalmente, se ha realizado un salto cuantitativo en materia de complejidad estableciendo como objetivo la reducción de consumo en sistemas procesadores de alta complejidad, como son los sistemas multi-procesador o redes-en-un-chip.

Otro de los puntos de vista que se ha empleado en el desarrollo de la Tesis es el método de abordar el problema que se plantea. Así, el problema del consumo energético se ha abordado inicialmente desde la estimación de dicho consumo, para continuar con la minimización de éste y, finalmente, desde la interacción consumo-prestaciones que aparece en los sistemas complejos cuando se emplean técnicas de reducción energética.

Finalmente, también se ha realizado todo el trabajo previamente descrito desde dos planteamientos diferentes: el establecimiento de técnicas de reducción local de consumo (en un módulo hardware determinado), y el establecimiento de técnicas que reduzcan el consumo global del sistema. La figura 2 muestra esquemáticamente los diversos puntos de vista que se han tratado en la Tesis y que acaban de ser comentados.

La complejidad de la materia, así como la amplitud de ésta, impiden su solución exhaustiva y completa. Esta Tesis ha cubierto algunos de ellos pero a su vez abre nuevas líneas de investigación futuras.

Las próximas secciones describen con más detalle los puntos principales en los que se trabajará.

1.2. Estimación de Consumo

Los diseñadores de sistemas digitales complejos demandan técnicas y herramientas para la estimación de disipación de potencia de forma que se puedan satisfacer las restricciones de potencia sin reducir significativamente las prestaciones globales, y todo ello desde las primeras fases del diseño. Así, se evitarán costosos rediseños. El objetivo es cumplir con los plazos temporales en el diseño, mientras que se explora el espacio de las posibles alternativas de diseño. La precisión y la eficiencia de las aproximaciones de alto nivel deberían ser los factores clave para alcanzar las restricciones de bajo consumo, evitando costosos procesos de rediseño. Es necesario indicar que en estimación de consumo son más interesantes los valores relativos que los absolutos de potencia disipada ya que, normalmente, el objetivo principal es comparar diversas alternativas en una exploración del espacio de diseño. Sin embargo, si el objetivo es el diseño de baterías o dispositivos de ventilación, la precisión de la estimación es una característica indispensable.

Cuando los diseñadores comenzaron a hacer mayor énfasis en el consumo como figura de mérito, se hizo claro que aunque existían herramientas que realizaban estimaciones de prestaciones y área, muy pocas tenían en cuenta el consumo energético. En los últimos años, esta situación ha mejorado drásticamente. Los desarrolladores más prestigiosos de herramientas de diseño ofrecen en la actualidad herramientas de estimación y análisis de consumo a nivel de circuito y puerta.

Cuando se trabaja con el procesador de un sistema empotrado, existen varios parámetros que tienen una relación directa sobre el balance prestaciones-consumo. Principalmente, el procesador empotrado y la jerarquía de memoria caché. En esta Tesis, se ha planteado el diseño y posterior uso de una herramienta de estimación precisa de consumo en la memoria caché. La herramienta se ha diseñado para ser fácilmente integrable en un proceso de diseño cruzado. Con este trabajo, las fuentes más importantes de disipación de potencia en el sistema procesador quedan completamente caracterizadas desde las primeras etapas de diseño y, de esta forma, el diseño subsecuente tiene en consideración la disipación de potencia como nueva restricción. La metodología propuesta puede extenderse a nuevas arquitecturas, proporcionando un entorno flexible y eficiente para el diseño de sistemas empotrados con restricciones de consumo.

1.3. Optimización de Consumo

Una vez que el problema de la estimación de consumo se ha solucionado parcialmente, el objetivo principal es la reducción de disipación energética en el sistema, actuando sobre los dispositivos y módulos con mayor consumo. El banco de registros consume una fracción nada despreciable de la potencia total disipada en los procesadores empotrados, y se convierte en una fuente dominante en la disipación energética después de aplicar otros mecanismos de reducción de consumo. Además, pese a que existe un trabajo previo muy completo en materia de reducción de consumo en la jerarquía caché, no existe tal desarrollo para el banco de registros. El consumo del banco de registros en sistemas empotrados depende en gran medida de la configuración del sistema, principalmente del número de registros integrados, el tamaño de la memoria caché y la existencia o no de una tabla de predicción de saltos (es decir, depende del tamaño relativo de este dispositivo en relación al resto de memorias). En la arquitectura del Motorola M.CORE, el consumo del banco de registros alcanza el 16 % del consumo total del procesador, y el 42 % del consumo en la línea de datos. En procesadores out-of-order con gran número de registros físicos, implementados muchas veces como parte del Re-Order Buffer (Pentium III, por ejemplo), el banco de registros disipa hasta el 27 % de la energía total de acuerdo a últimas estimaciones.

Las técnicas de reducción de consumo para procesadores in-order y out-of-order que se han desarrollado se caracterizan por la ausencia de penalización en prestaciones y reducido incremento en el tiempo de diseño. Se basan en la observación de que un registro sólo se usa cuando una instrucción lo lee o lo escribe, mientras que permanece en un estado “desocupado” el resto del tiempo. Si se mantienen dichos registros desocupados en un estado de bajo consumo (estado “de letargo”), se puede ahorrar una considerable porción energética. La mayoría de los registros del banco de registros permanecen desocupados en cada ciclo de reloj, ya que cada instrucción despachada accede a un máximo de tres registros.

Así, tal y como se ha descrito en las secciones previas, esta Tesis plantea el problema de la estimación y minimización de consumo en sistemas procesadores desde diferentes niveles de complejidad y análisis, desarrollándose un conjunto de técnicas, herramientas y modificaciones arquitecturales que permiten la estimación eficiente de la disipación energética, así como la minimización de ésta sin reducir las prestaciones del sistema.

1.4. Compromiso Consumo-Prestaciones en MPSoCs

El trabajo de investigación en este área se centra en el análisis a nivel de sistema de las interacciones entre diferentes procesadores de un sistema multi-procesador, cuando éstos se escalan en tensión y frecuencia para reducir el consumo energético del chip. En concreto, los factores principales a analizar son el impacto en las prestaciones globales del sistema, la disipación energética y el producto energía-retardo.

Cuando un único procesador (o varios) de un sistema MPSoC trabaja a una frecuencia relativa superior o inferior a la del resto de unidades procesadoras, aparece un impacto negativo en la utilización del bus de comunicaciones cuando se ejecutan tareas que requieren una elevada comunicación y sincronización inter-procesador. Dicho efecto se agrava cuando son varios los procesadores trabajando a diversas frecuencias de reloj, llegando a ocultar cualquier ahorro de consumo debido a la degradación del tiempo de ejecución.

El trabajo de investigación en este área facilita un conjunto de reglas de diseño que permiten la selección estática de la frecuencia de funcionamiento de los procesadores integrados para una tarea dada, de forma que se asegure el ahorro energético sin reducir las prestaciones del sistema. Del mismo modo, se ha analizado el impacto que tiene la existencia de un Sistema Operativo sobre la utilización del bus y los beneficios de incrementar la frecuencia de reloj en aquellos procesadores que ejecutan elevadas cargas asimétricas.

Finalmente, el conocimiento adquirido durante las fases de análisis y diseño ha permitido desarrollar técnicas de control de potencia que permiten el ahorro de consumo bajo restricciones de tiempo de ejecución mediante el escalado dinámico y selectivo de la frecuencia de los procesadores.

2. Objetivos de la Tesis

La importancia de desarrollar sistemas eficientes energéticamente ya ha sido brevemente descrita en la introducción del presente documento. Sin duda, el mercado actual demanda sistemas portátiles operados con baterías que realizan cada día aplicaciones más complejas y que demandan gran capacidad computacional (PDAs, telefonía móvil multimedia, GPS, transmisión de datos inalámbrica, etc). Al mismo tiempo, estos sistemas vienen reduciendo drásticamente su tamaño, y el peso de las baterías se traduce en una fracción determinante del sistema final. Los usuarios demandan baterías de larga duración y esto únicamente puede obtenerse incrementando la capacidad de la batería o incrementando la eficiencia de la lógica. La velocidad de innovación en tecnología de almacenamiento energético es lenta (no es capaz de seguir la velocidad de innovación en sistemas de computación) y, por tanto, es deber del diseñador digital mejorar las prestaciones del diseño electrónico.

Podemos apuntar otras razones por las que el consumo de potencia se ha convertido en un objetivo prioritario en el diseño de sistemas electrónicos basados en procesador. Al aumentar la disipación térmica en los circuitos integrados (debido al aumento de consumo energético que se traduce directamente en disipación calorífica), se hace más complicado y costoso el proporcionar suficientes mecanismos de ventilación a través de encapsulados específicos de alta conductividad térmica y resistividad calorífica, ventiladores y disipadores térmicos. Más aún, las elevadas temperaturas que pueden alcanzarse en los circuitos integrados aumentan la carga que deben soportar éstos y reducen su fiabilidad.

También es importante tener en cuenta consideraciones eléctricas. Proporcionar una fuente de energía de la suficiente capacidad requiere un gran número de interconexiones entre el chip y el encapsulado, y una elevada proporción del espacio destinado a interconexión de señales es ocupado por la distribución de la fuente de energía. Las altas densidades de corriente pueden desencadenar electromigraciones en las metalizaciones. A nivel de sistema, los dispositivos con grandes necesidades de consumo requieren complejas y costosas baterías o fuentes de energía.

Finalmente, no hay que olvidar que gran parte del consumo total energético se debe a los millones de sistemas de computación que permanecen continuamente activos y demandando energía eléctrica de las fuentes del planeta.

El problema que se ha resuelto en esta Tesis consiste en la estimación de consumo y posterior minimización de éste en sistemas basados en procesador, mediante la definición de técnicas de compilación y la inclusión de modificaciones arquitecturales. Entenderemos por sistemas basados en microprocesador todos aquellos sistemas electrónicos que cuentan con una o varias unidades de procesamiento programables independientes o comunicadas mediante un bus, ya sean de diseño y aplicación específica (sistemas

empotrados) o sistemas de propósito general (procesadores de altas prestaciones).

Así, se ha abordado el problema del consumo energético en sistemas procesadores desde diversos puntos de vista (la estimación, y la minimización local o global de éste) y desde diversos niveles de complejidad (sistemas mono-procesadores de baja y media complejidad, y sistemas multi-procesadores de alta complejidad). Además, las técnicas y modificaciones propuestas analizan los diversos efectos de segundo orden que aparecen cuando se establecen técnicas de reducción de consumo, y evitan cualquier penalización en las prestaciones globales del sistema.

Dado el impacto que dichas modificaciones y técnicas pueden tener en la funcionalidad, prestaciones y flujo de diseño de los sistemas convencionales, se han establecido los siguientes criterios de diseño:

- Ausencia o bajo impacto sobre las prestaciones del sistema final. Así, se han establecido las métricas que permiten monitorizar el impacto en prestaciones después de aplicar las políticas de reducción de consumo: producto $energía - retardo$, producto $energía^2 - retardo$, IPC .
- Simplicidad de la lógica adicional necesaria para establecer las políticas de minimización energética. Dicha lógica adicional ha sido perfectamente caracterizada en términos de tiempo de operación y consumo energético. De esta forma, los valores de simulación obtenidos para el sistema completo incluyen el impacto de los módulos hardware adicionales.
- Integración perfecta en el flujo de diseño habitual de sistemas electrónicos. La justificación de este punto es facilitar al diseñador de sistemas procesadores la inclusión de la disipación energética como restricción adicional a las tradicionales de prestaciones y coste.
- Adecuada flexibilidad en las técnicas y modificaciones arquitecturales propuestas que permiten su aplicación al conjunto de los sistemas procesadores existentes en la actualidad y su futura incorporación en arquitecturas de próximo desarrollo. De esta forma, se evitan los diseños de lógica adicional restringidos a la arquitectura de procesadores específicos, o el empleo de características de procesadores concretos en el desarrollo de técnicas de reducción de consumo. Igualmente, se ha trabajado en la medida de lo posible con modelos de sistemas basados en procesadores comerciales modernos, evitando el empleo de modelos simplificados de arquitecturas virtuales.

Cabe destacar que el objetivo de esta Tesis se ha planteado desde diversos niveles de complejidad y abstracción en el conjunto de los sistemas electrónicos basados en procesadores. Así, se han planteado soluciones para sistemas monoprocesador de propósito específico (sistemas empotrados) de altas prestaciones y especificidad, sistemas monoprocesador de propósito general de altas prestaciones, y sistemas multiprocesador en un chip (sistemas multiprocesadores con topología en red e integrados sobre el mismo sustrato).

2.1. Estimación de Consumo

Tal y como se ha comentado con anterioridad, el desarrollo de herramientas precisas y eficaces para la estimación de consumo en sistemas procesadores es una de las demandas de los diseñadores actuales de sistemas integrados.

Son varias las fuentes principales de consumo en un sistema de las características que nos ocupa, como son el núcleo procesador y la jerarquía de memoria caché. La presente Tesis abordará la memoria caché para el desarrollo de la herramienta de estimación de consumo.

Así, se ha presentado el diseño y posterior uso de una herramienta precisa y multi-plataforma para la estimación de consumo en la jerarquía caché de procesadores. La metodología propuesta es extensible a futuras arquitecturas de procesadores, proporcionando un entorno eficiente para el diseño de sistemas empotrados con restricciones de consumo. El objetivo de nuestro trabajo de investigación en esta área ha sido el de proporcionar una herramienta precisa de estimación de consumo fácilmente integrable en el flujo de diseño de sistemas empotrados. Dicho entorno de diseño integra el simulador funcional del núcleo procesador del sistema, junto con el simulador de la memoria caché, y junto con la herramienta que proporciona las estimaciones de consumo para diversas arquitecturas destino. Más aún, el impacto de esta herramienta de estimación de consumo en el tiempo de diseño o en las fases requeridas para completar la especificación y validación del sistema electrónico es reducido, pero a la vez extiende la

información de salida del procesador funcional (información de prestaciones, accesos a memoria, etc) con la información de consumo.

La metodología propuesta puede ser aplicada a futuras arquitecturas, proporcionando así un mecanismo rápido y eficiente para analizar el consumo energético de dichos diseños.

De esta forma, la metodología de simulación proporciona un mecanismo sencillo para que los diseñadores puedan evaluar diversas alternativas de diseño con respecto a las restricciones de prestaciones y consumo existentes. Las principales ventajas de esta aproximación son: la generación automática de las herramientas cruzadas de diseño, la obtención en paralelo de información funcional y estimación de consumo en la memoria caché para cada una de las simulaciones realizadas, tiempo despreciable de aprendizaje para el usuario y flexibilidad en la elección del procesador destino.

Son diversos los elementos arquitecturales y estadísticos (correlaciones de datos) que impactan en el consumo de la memoria caché, y el diseñador de sistemas debe conocerlos para orientar el diseño de forma que se minimice la disipación energética. Este entorno será de gran utilidad para el diseñador, el cual puede realizar la exploración del espacio de diseño en diferentes escenarios, tal y como son las modificaciones del código fuente a través de optimizaciones del compilador, la variación de la estructura de la jerarquía caché, la evaluación de diversos conjuntos de datos de entrada, etc. De forma paralela a la simulación funcional, obtiene el consumo asociado a diferentes versiones de código, arquitectura de la jerarquía caché o conjunto de datos de entrada.

2.2. Optimización de Consumo en Procesadores In-Order

Los sistemas empotrados, o sistemas de aplicación específica basados con núcleo microprocesador, cuentan con procesadores de arquitectura in-order. En estos sistemas, el banco de registros consume una porción nada despreciable de la potencia total, más aún si tenemos en cuenta la disipación asociada al árbol de distribución del reloj que alimenta dicho dispositivo. Además, el banco de registros se convierte en la fuente dominante de consumo una vez que se han aplicado las técnicas tradicionales de reducción energética.

Finalmente, la tendencia de fabricación de los actuales procesadores claramente muestra cómo el tamaño del banco de registros se viene incrementando drásticamente con cada nueva generación. A esto colabora la tecnología de compiladores que permite aprovechar eficientemente la totalidad de registros existentes en la arquitectura, y las optimizaciones de compilación existentes, que incrementan la presión sobre los registros. Además, los bancos de registros de elevadas dimensiones presentan diversas ventajas: reducen el consumo de energía en la jerarquía caché y memoria principal disminuyendo el número de accesos a éstas y aumentan las prestaciones del sistema reduciendo el tráfico de memoria.

Así, nuestro objetivo en esta área ha sido el desarrollo de técnicas de compilación y definición de modificaciones arquitecturales que permiten la reducción local de consumo en procesadores in-order. En concreto, minimizando el consumo del banco de registros.

La idea fundamental en la que se basan las técnicas y modificaciones arquitecturales propuestas es el fenómeno de sobredimensionamiento que aparece en este dispositivo y la existencia de registros en estado "desocupado" que pueden aprovecharse para reducir el consumo. Cuando una sección de código, o una micro-instrucción del procesador, no emplea todos los registros contenidos en el banco de registros, debería ser posible inhabilitar aquéllos que se encuentran desocupados y mantener activos sólo los necesarios para el acceso. El inhabilitar los registros desocupados proporcionará importantes ahorros de energía debido al decremento de potencia estática en el dispositivo y debido a la posibilidad de inhabilitar también el árbol de distribución del reloj que alimenta estos registros. Cuando otra porción de código, u otra micro-instrucción, requiera alguno de los registros inhabilitados, éste deberá habilitarse con tiempo previo suficiente para que no se produzca ningún impacto en las prestaciones.

Sin embargo, es necesario facilitar algún mecanismo para que los registros inhabilitados no pierdan la información contenida en ellos, ya que en caso contrario sería necesario recuperarla de la caché o la memoria principal, requiriendo ciclos extras de acceso a memoria e incrementando el consumo en estos dispositivos de memoria. Así, los registros que no se empleen en el acceso deben mantenerse en un estado de bajo consumo que mantenga la información contenida pero reduzca el consumo estático a un mínimo. Para ello, se han analizado y aplicado, con las modificaciones necesarias, las investigaciones y técnicas de Escalado Dinámico de Tensión, que demuestran la reducción cuadrática de consumo que se produce

al reducir la tensión de alimentación de los dispositivos. En concreto, se ha partido de la aplicación del estado de bajo consumo “estado de letargo” propuesto primeramente por Flautner et al. para la reducción de consumo en memorias caché. Dicho esquema se ha modificado y ampliado para nuestros propósitos.

El problema fundamental que debe resolverse para evitar cualquier penalización en prestaciones que pueda ocurrir a la hora de habilitar o inhabilitar los registros es la detección temprana de su inactividad o la necesidad de éstos por parte de las micro-instrucciones ejecutadas. Así, se han desarrollado técnicas de predicción y/o detección hardware que permiten la (in)habilitación temprana de los registros. Dichas técnicas de predicción consideran características generales existentes en la codificación de micro-instrucciones para máquinas MIPS, y en la estadística de accesos al banco de registros. La lógica diseñada satisface las restricciones de retardo de ejecución necesarias para que el pipeline del sistema no se alargue, y el consumo energético de los módulos hardware adicionales ha sido estimado y es lo suficientemente despreciable para que el ahorro resultante satisfaga las expectativas del diseño.

Como posibilidad adicional al empleo de modificaciones hardware, se han desarrollado técnicas de compilación que de forma independiente, o en colaboración con las modificaciones hardware, permiten la (in)habilitación de determinados registros. Dichas técnicas de compilación se establecen como optimizaciones adicionales a las tradicionales de prestaciones, y están fundamentadas por los estudios de perfilado y análisis de código efectuado sobre el conjunto de bancos de prueba de interés en el ámbito de los sistemas empotrados.

2.3. Optimización de Consumo en Procesadores Out-Of-Order

Una vez abordada la optimización de consumo en el banco de registros de sistemas empotrados, el siguiente objetivo que se ha cubierto es la traslación de las técnicas y módulos arquitecturales desarrollados hacia los sistemas de altas prestaciones y propósito general, incrementando de esta forma la complejidad del sistema final.

En estos sistemas out-of-order aparece un factor adicional de indeterminismo generado por el renombrado de registros, etapa adicional del pipeline que permite resolver las dependencias entre instrucciones cuando éstas se ejecutan de forma “desordenada” en procesadores superscalares, incrementando las prestaciones del sistema en gran medida.

Así, si queremos aplicar de forma directa las aproximaciones desarrolladas para procesadores in-order, es necesario modificar el renombrado hardware de registros que aparece en estas arquitecturas. Es decir, es necesario reducir el indeterminismo del renombrado de registros para que, en todo momento, podamos conocer los registros específicos que pueden ser inhabilitados por nuestra política de reducción de consumo.

De esta forma, se han desarrollado las necesarias modificaciones arquitecturales que permiten reducir el indeterminismo del renombrado y, al igual que en el caso anterior, la lógica adicional ha sido caracterizada en consumo y retardo para asegurar el ahorro de consumo y la ausencia de penalización en las prestaciones del sistema final. Del mismo modo, se han desarrollado las técnicas de compilación necesarias para, de forma independiente o adicionalmente a las modificaciones hardware, incrementar el ahorro de consumo que es posible obtener.

2.4. Optimización de Consumo en Sistemas Multi-Procesador en Un Chip

Una vez desarrolladas las técnicas de reducción de consumo para sistemas monoprocesador (tanto de aplicación específica como de propósito general) se ha abordado el problema de la optimización de consumo en arquitecturas cuantitativamente más complejas, los sistemas multiprocesador en un chip.

Los sistemas multiprocesador en un chip (MPSoCs) son objeto de intensa investigación en los últimos años debido a los elevados requerimientos en la ejecución de tareas multimedia con elevadas restricciones de prestaciones, y debido a los fenómenos complejos que aparecen en la comunicación de los procesadores mediante el bus integrado en el sistema, y que impiden aún hoy establecer modelos analíticos de comportamiento. En estos sistemas, el consumo energético no sólo vendrá determinado por las necesidades de consumo de cada uno de los procesadores integrados, sino también por el acceso a recursos comunes como la memoria compartida o el bus de comunicaciones. De especial interés resulta este último debido a que las colisiones que se producen en el recurso de interconexión en función del tipo de bus, la política de asignación de recursos, o la aplicación en ejecución, repercuten en gran medida en el consumo total del sistema.

Las técnicas y modificaciones arquitecturales propuestas han sido introducidas en un entorno de simulación precisa de este tipo de sistemas. El objetivo ha sido la definición de técnicas y modificaciones arquitecturales que reducen el consumo energético del sistema manteniendo las prestaciones del mismo. Estas técnicas se basan en el empleo del escalado selectivo de tensión y frecuencia en los diversos procesadores y se han analizado los efectos de segundo orden que aparecen cuando los elementos procesadores de los MPSoCs no operan sincronizadamente.

Del mismo modo, el análisis de dichos efectos de segundo orden ha permitido la definición de políticas de bajo consumo que aprovechan el conocimiento generado para el establecimiento de reglas o líneas de diseño cuando se procede al escalado en frecuencia y tensión de los procesadores. Así, el comportamiento observado y posteriormente analizado permite seleccionar de forma estática la configuración del sistema que minimiza el consumo de potencia para unas prestaciones dadas, y también permite la implementación de mecanismos dinámicos para la selección de configuraciones óptimas en función del comportamiento dinámico de la aplicación. Dichas modificaciones dinámicas pueden ser reguladas tanto por módulos hardware dedicados a tal efecto, o por la existencia de un sistema operativo eficiente en consumo que opere como base de las aplicaciones ejecutadas por el sistema.

Al igual que en casos anteriores, las modificaciones arquitecturales (módulos de control de tensión y frecuencia, interfaces, adaptadores, etc.) introducidas en el sistema base, satisfacen las restricciones temporales (evitando cualquier penalización de prestaciones) y están perfectamente caracterizadas en consumo para asegurar que los ahorros de consumo obtenidos cumplen las expectativas de diseño.

2.5. Conclusiones

Teniendo en cuenta las limitaciones de las aproximaciones actuales, se han presentado los objetivos para el desarrollo de nuevas técnicas de minimización de consumo en sistemas procesadores que permiten afrontar los retos futuros en el diseño de sistemas empotrados, sistemas de altas prestaciones, y sistemas multi-procesador en un chip, así como para la definición de herramientas eficientes de estimación de consumo en jerarquías caché.

Las principales líneas de acción, que definen los objetivos concretos de esta Tesis, podemos resumirlas en:

- Ausencia o bajo impacto sobre las prestaciones del sistema final.
- Simplicidad de la lógica adicional.
- Integración perfecta en el flujo de diseño habitual de sistemas electrónicos.
- Adecuada flexibilidad en las técnicas y modificaciones arquitecturales propuestas.
- Integración de la estimación de potencia con la simulación funcional en el flujo de diseño de sistemas procesadores.

3. Método y Fases de Trabajo

En este capítulo se describen los pasos que se han seguido en el desarrollo de esta Tesis para la consecución de los objetivos que se planteaban en el capítulo anterior.

3.1. Estimación de Consumo

La metodología que se ha seguido para alcanzar los objetivos fijados en el área de estimación de consumo es la siguiente:

- Selección del modelo analítico de consumo. Análisis de los parámetros involucrados y detección de debilidades del modelo que puedan ser mejoradas gracias a la información de simulación.
- Implementación del modelo e integración del mismo con el simulador de la memoria caché.
- Integración del simulador en el entorno de diseño de sistemas empotrados.

- Verificación de la funcionalidad obtenida con la herramienta.
- Definición del banco de pruebas.
- Empleo de la herramienta para el análisis de diversos factores que impactan en el consumo de la caché (datos de entrada, configuración de la jerarquía, etc) y análisis de resultados para establecer un conjunto de reglas de diseño. De esta forma, se ha realizado una completa exploración del espacio de diseño atendiendo a los principales factores que intervienen en el consumo de la jerarquía de memoria caché.

3.2. Optimización de Consumo en Arquitecturas In-Order y Out-Of-Order

La metodología que se ha seguido para alcanzar los objetivos fijados en el área de optimización de consumo en arquitecturas in-order y out-of-order es la siguiente:

- Selección del mecanismo de reducción de consumo en los registros inhabilitados (definición del estado de bajo consumo) y aplicación de éste al banco de registros.
- Análisis y detección de parámetros arquitecturales que permitan la habilitación e inhabilitación de los registros seleccionados mediante el empleo de hardware adicional.
- Diseño de la lógica hardware adicional, implementación de ésta, y caracterización de la misma en retardo y consumo energético.
- Definición del banco de pruebas.
- Verificación mediante simulación del ahorro energético alcanzado.
- Análisis y detección de parámetros de compilación y código fuente que permitan la (in)habilitación de los registros seleccionados mediante optimizaciones del compilador.
- Diseño de las optimizaciones de compilación e implementación de éstas en el seno de un compilador de altas prestaciones.
- Verificación mediante simulación del ahorro energético alcanzado.

3.3. Optimización de Consumo en Sistemas Multi-Procesador en Un Chip

La metodología que se ha seguido para alcanzar los objetivos fijados en el área de optimización de consumo en sistemas multi-procesador en un chip es la siguiente:

- Definición de la arquitectura multi-procesador base (procesadores, dispositivos compartidos, bus de comunicaciones, etc.).
- Diseño e implementación en el seno del simulador del sistema de los módulos hardware necesarios para el escalado selectivo de frecuencia y alimentación de los procesadores. Caracterización de la lógica adicional.
- Definición del banco de pruebas.
- Verificación mediante simulación del ahorro energético alcanzado.
- Análisis de los resultados y empleo de los mismos para la definición de un conjunto de reglas de operación.
- Diseño e implementación en el seno del simulador de un módulo controlador de consumo basándose en las conclusiones extraídas. Caracterización de la lógica adicional.
- Verificación mediante simulación del ahorro energético alcanzado.

4. Conclusiones y Resultados

Este documento presenta el resumen de una Tesis Doctoral encaminada al diseño de técnicas y modificaciones arquitecturales para la reducción de consumo en sistemas basados en procesador.

El objetivo final es la reducción significativa del consumo energético en sistemas de creciente complejidad: sistemas empujados de aplicación específica, sistemas de altas prestaciones y propósito general, y sistemas multiprocesador en un chip.

Se ha realizado una exhaustiva revisión de los trabajos publicados en el tema central de investigación, destacando en todo momento las publicaciones más recientes y de mayor relevancia. El análisis de dichos trabajos ha permitido el establecimiento de los objetivos y líneas de investigación a seguir.

Así, los problemas que se han abordado de forma ordenada en esta Tesis son:

1. La estimación de consumo en la jerarquía caché de sistemas procesadores;
2. La minimización del consumo en un módulo hardware concreto de los sistemas empujados con mono-procesador *in-order*;
3. La extensión de las técnicas previas a arquitecturas *out-of-order*;
4. Y el análisis y minimización del consumo global en sistemas complejos multi-procesadores;

Para ello, ha sido necesario combatir algunas de las limitaciones de las técnicas, lo que ha pasado por centrar los criterios de diseño en los siguientes:

- Ausencia o bajo impacto sobre las prestaciones del sistema final.
- Simplicidad de la lógica adicional.
- Integración perfecta en el flujo de diseño habitual de sistemas electrónicos.
- Adecuada flexibilidad en las técnicas y modificaciones arquitecturales propuestas.

Se ha planteado un plan de trabajo para llegar a la consecución de los objetivos propuestos, que incluye la verificación mediante simulación de las modificaciones propuestas con ejemplos reales.

De esta forma, los resultados más sobresalientes que se han obtenido en los diversos campos que cubre esta tesis han sido los siguientes.

En el campo de estimación de consumo:

- Desarrollo de un modelo analítico de consumo para la jerarquía caché de sistemas empujados. Este modelo analítico ha tenido en cuenta por primera vez hasta la fecha el impacto del factor de actividad en buses y unidades funcionales, evitando la aproximación de la distribución estadística homogénea realizada en los trabajos previos.
- Diseño de una herramienta de análisis de consumo basada en el modelo analítico.
- Integración en un entorno de diseño y generación automática de herramientas, siendo ésta la primera herramienta de la que se tiene constancia que realiza una verificación funcional y de consumo de la jerarquía caché de sistemas empujados de forma simultánea. Así mismo, la generación automática de la herramienta de estimación, de forma totalmente transparente al diseñador, es una característica muy apreciada y que es completamente novedosa en su dominio de aplicación.
- Finalmente, la aplicación de la herramienta al análisis de diversas aplicaciones clave, donde se ha detectado y analizado la influencia de la actividad en los buses, la correlación en los datos, y la jerarquía caché en el consumo de potencia. Hasta la publicación de estos resultados en diversas publicaciones internacionales, no se tenía constancia de la influencia de estas variables de alto nivel en el consumo energético, abriendo nuevas vías de optimización energética.

En el campo de la optimización de consumo:

- Diseño e implementación de unos mecanismos hardware y software (compilación) de reducción de consumo en el fichero de registros de sistemas monoprocesador. Ambos mecanismos de optimización obtienen elevadas reducciones de consumo en el dispositivo de interés y, al contrario que las técnicas presentadas anteriormente en la literatura, no suponen ningún impacto negativo en las prestaciones del sistema.
- Obtención de resultados sobresalientes en ahorro energético: 84 % (mecanismo hardware en arquitecturas *in-order*), 65 % (mecanismo software en arquitecturas *in-order*), 62 % (mecanismo combinado hardware-software en arquitecturas *out-of-order*)
- Análisis de los complejos mecanismos que relacionan las prestaciones y el consumo energético en los sistemas multi-procesadores. Este análisis no se había tenido en cuenta previamente debido a lo novedosas que son las arquitecturas (no existiendo un simulador funcional a tal efecto) y a la falta de un modelo preciso de consumo.
- Diseño e implementación de un mecanismo de control dinámico de consumo en el sistema multi-procesador basado en una arquitectura productor-consumidor, alcanzando un ahorro del 20 % de consumo energético en el sistema completo.

5. Publicaciones de la Tesis

Capitulos de Libros

State-of-the-Art SoC Communication Architectures.” José L. Ayala, Marisa López-Vallejo, Davide Bertozzi, Luca Benini. To be published in ”The Embedded Systems Handbook”, (CRC Press, Florida), October 2004.

Revistas Internacionales

“Energy-Aware Compilation and Hardware Design for VLIW Embedded Systems” José L. Ayala, David Atienza, Praveen Raghavan, Marisa López-Vallejo, Francky Catthoor, Diederik Verkest. To be published in International Journal of Embedded Systems , Fall 2006.

“Reducing the Energy Consumption of the Register File for In-Order Architectures” José L. Ayala, Marisa López-Vallejo, Alex Veidenbaum, Carlos A. López. International Journal of Embedded Systems , February 2005.

“Integrating Functional and Power Simulation in Embedded Systems Design” José L. Ayala, Marisa López-Vallejo. Journal of Embedded Computing Vol. 1, No. 3, September 2004

“Power-Aware Compilation for Register File Energy Reduction” José L. Ayala, Alexander Veidenbaum, Marisa López-Vallejo. International Journal of Parallel Programming (Kluwer Academic Publishers), Vol. 31, No. 6, pp. 449-465, December 2003.

“A Unified Framework for Power-Aware Design of Embedded Systems” José L. Ayala, Marisa López-Vallejo. Lecture Notes on Computer Science (Springer Verlag) , vol. 2799, 2003

Comunicaciones en Congresos Internacionales

“Energy-Aware Compilation and Delay Impact Minimization for VLIW Embedded Systems” José L. Ayala, David Atienza, Praveen Raghavan, Marisa López-Vallejo, Francky Catthoor. IEEE Workshop on Innovative Architectures, Hawaii, January 2006.

“Optimal Loop-Unrolling Mechanisms and Architectural Extensions for an Energy-Efficient Design of Shared Register Files in MPSoCs” José L. Ayala, David Atienza, Marisa López-Vallejo, José M. Mendías,

Román Hermida, Carlos A. López. IEEE Workshop on Innovative Architectures, Hawaii, January 2005.

“Improving Register File Banking with a Power-Aware Unroller” José L. Ayala, Marisa López-Vallejo. Workshop on Power-Aware Real-Time Computing (in conjunction with ACM International Conference on Embedded Software), Pisa (Italy), September 2004.

“A Compiler-assisted Banked Register File Architecture” José L. Ayala, Marisa López-Vallejo, Alex Veidenbaum. IEEE Workshop on Application Specific Processors (in conjunction with IEEE International Conference on Hardware/Software Codesign and System Synthesis), Stockholm (Sweden), September 2004.

“Power Estimation and Power Optimization Policies for Processor-Based Systems” José L. Ayala, Marisa López-Vallejo. SIGDA Ph.D. Forum at Design Automation Conference, San Diego (California, USA), June 2004. (also the poster)

“Power-aware Register Renaming in High-Performance Processors using Compiler Support” José L. Ayala, Marisa López-Vallejo, Alex Veidenbaum. IEEE Workshop on Innovative Architectures, Hawaii, January 2004.

“Energy-Efficient Register Renaming in High-Performance Processors” José L. Ayala, Marisa López-Vallejo, Alex Veidenbaum. IEEE Workshop on Application Specific Processors (in conjunction with IEEE International Symposium on Microarchitecture), San Diego (California, USA), December 2003.

“A Case Study on Power Dissipation in the Memory Hierarchy of Embedded Systems” José L. Ayala, Marisa López-Vallejo and Carlos A. López Barrio. Design of Circuits and Integrated Systems Conference, Ciudad Real (Spain), November 2003.

“A Unified Framework for Power-Aware Design of Embedded Systems” José L. Ayala, Marisa López-Vallejo. IEEE International Workshop on Power and Timing Modeling, Optimization and Simulation, Torino (Italy), September 2003.

“Energy Aware Register File Implementation through Instruction Predecode” José L. Ayala, Marisa López-Vallejo, Alex Veidenbaum, Carlos A. López-Barrio. IEEE International Conference on Application-Specific Systems, Architectures and Processors, The Hague (Netherlands), June 2003.

“Reducing Register File Energy Consumption using Compiler Support” José L. Ayala, Alex Veidenbaum. IEEE Workshop on Application Specific Processors (in conjunction with IEEE International Symposium on Microarchitecture), Istanbul (Turkey), November 2002. Best Student Paper Award