عنوان: مدیریت توان مصرفی در سامانه های نهفته بی درنگ

چکیده:

با پیشرفت فناوری ساختCMOS ، تراشه های چندهسته ای به طور گسترده در سامانه های نهفته ی بی-درنگ به دلیل پتانسیل آن ها برای دست یابی به کارآیی بالا به همراه هزینه ی کم، استفاده می شوند. زمانی که اندازه ی ترانزیستورهای CMOS به صورت پیوسته کوچک می‌شود، اکثر هسته ها و دستگاه های الکترونیکی روی یک تراشه مجتمع می شوند. از آنجایی که تراکم توان برای یک تراشه¬ی چندهسته‌ایِ پیشرفته به سرعت در حال رشد است، باید مدیریت توان به عنوان یکی از مسائل بحرانی که بر روی کارآمدی خنک سازی تراشه و قابلیت اطمینانِ تراشه تأثیرات مخرب می¬گذارد، در نظر گرفته شود. برای این منظور، بدون نقض نیازمندی های زمانی، روش های مدیریت توان متعددی برای کمینه کردن انرژی مصرفی یا برای کمینه کردن دمای تراشه، به وسیله ی کنترل کردن متوسط توان مصرفی ارائه شده است. بنابراین مدیریتِ توان مصرفی یک تراشه¬ی چندهسته ای یک مسئله‌ی مهم در طراحی سامانه های نهفته ی بی درنگ است. لازم به ذکر است که اکثر روش های موجود از طریق کاهش انرژی مصرفی کل و یا کاهش دمای تراشه ، متوسط توان مصرفی را کاهش می¬دهند. حال آن¬که در این بین توجه اندکی به بدترین رفتار لحظه¬ایِ توان مصرفی در یک تراشه (اوج توان مصرفی در سطح تراشه) شده است .

بدترین رفتار لحظه ایِ توان مصرفی یک معیار مهمِ طراحی در تعیین هزینه، اندازه‌ی تراشه و منبع تغذیه است. روند افزایشی مجتمع سازی در تراشه های چندهسته ای، باعث افزایش منطق دیجیتال و به تبع آن افزایش توان مصرفی کل تراشه خواهد شد. این افزایشِ توان تحت عنوان سیلیکون تاریک بیان می گردد و بدین معناست که در سامانه های چندهسته ای، درصد مهمی از هسته های در دسترس، به دلیل محدودیت دمایی و توانی نمی توانند به طور همزمان فعال باشند. به طور معمول سازندگان تراشه‌های چندهسته‌ای، عددی تحت عنوان توان حرارتی طراحی برای هر تراشه ارائه می دهند. مقدارTDP ، حداکثر توانِ مصرفی قابل قبولی است که تراشه می‌تواند بدون استفاده از مکانیزم‌های تنزل کارآیی مصرف کند؛ از جمله‌ی این مکانیزم ها می توان به روش مدیریت حرارتی پویا اشاره نمود. در سامانه‌های نهفته به دلیل محدودیت منابع انرژی و اثر منفیِ وجود واحدِ خنک‌سازِ فعال بر قابلیت اطمینان، معمولاً در بسیاری از موارد خنک ساز فعال وجود ندارد و از خنک‌ساز غیرفعال استفاده می‌شود که موجب کاهش قابل توجه TDP می گردد .

دسته¬ی مهمی از سامانه‌های نهفته ی توزیع شده، سامانه هایی هستند که در کاربردهایی با محدودیت های بی درنگی سخت به کار گرفته می‌شوند. سامانه‌های نهفته‌ی بی‌درنگ سخت معمولاً در کاربردهایی بکار گرفته می¬شوند که در ارتباط مستقیم با حیات انسان‌ها، محیط زیست و یا اقتصاد و امنیت کشورها هستند. از جمله¬ی این کاربردها می‌توان به سامانه‌های تجهیزات پزشکی، هوایی و فضایی، ارتباطی و نظامی اشاره کرد. بنابراین بروز هرگونه اشکال در این سامانه‌ها می‌تواند منجر به وارد شدن خسارت سنگین و غیرقابل جبرانِ جانی، مالی و زیست‌محیطی گردد. بنابراین این سامانه‌ها باید دارای قابلیت اطمینان بسیار بالایی باشند که از طریق تحمل‌پذیری اشکال حاصل می¬شود. سامانه‌های نهفته‌ی بی‌درنگ سخت با قابلیت تحمل‌پذیری اشکال به گونه‌ای طراحی و ساخته می¬شوند که دارای سه ویژگی باشند: ۱) کار تخصیص داده شده به این سامانه‌ها، به درستی و با صحتِ کامل انجام شود، ۲) قیود زمانی را برآورده کرده، به طوری که موعد مربوط به بی‌درنگی سخت نقض نشود و ۳) محدودیتِ TDP مربوط به تراشه رعایت شود؛ زیرا با نقض هریک از موارد بالا ممکن است این سامانه‌ها با تنزل قابلیت اطمینان روبرو شوند. در این نوع از سامانه ها، تحمل پذیری اشکال از راه تکرار وظایف مبتنی بر افزونگی سخت‌افزاری و نرم‌افزاری بدست می‌آید. از سوی دیگر روش های مبتنی بر تکرار وظایف، باعث تحمیل سربارِ توان مصرفی بسیار قابل توجهی به سامانه می شوند .

معمولاً برای تحمل پذیری اشکال در سامانه های نهفته ی بی درنگ سخت از افزونگی سخت افزاری و افزونگی زمانی استفاده می شود. افزونگی زمانی از زمان بیکاری برای بازگشت از اشکال استفاده می کند که نسبت به افزونگی سخت افزاری از سربار سخت افزاری کمتری برخوردار است. این افزونگی فقط زمانی کاربرد دارد که به قدر کافی زمان برای اجرای مجددِ وظایفی که دچار اشکال شده اند، وجود داشته باشد. در طرف مقابل با استفاده از افزونگی سخت‌افزاری، معمولاً چند نسخه از یک وظیفه به طور همزمان یا با همپوشانی در هنگام اجرای وظایف اجرا می شوند. بنابراین مدیریت توان مصرفی یکی از چالش های مهم در سامانه های نهفته تحمل پذیر اشکال است.

Title: Power Management in Real-Time Embedded Systems

With the advancement of CMOS fabrication technology, multi-core chips are widely used in real-time embedded systems due to their potential to achieve high performance with low cost. As the size of CMOS transistors is continuously reduced, most of the cores and electronic devices are integrated on a single chip. As the power density for an advanced multi-core chip is rapidly increasing, power management should be considered as one of the critical issues that adversely affects the chip cooling efficiency and chip reliability. To

this end, various power management methods have been proposed to minimize the power consumption or to minimize the chip temperature by controlling the average power consumption without violating the timing requirements. Therefore, power management of a multi-core chip is an important issue in the design of real-time embedded systems. It should be noted that most of the existing methods reduce the average power consumption by reducing the total energy consumption or reducing the temperature of the chip. However, little attention has been paid to the worst-case instantaneous behavior of the power consumption in a chip (peak power consumption at the chip level).

The worst-case instantaneous behavior of the power consumption is an important design criterion in determining the cost, chip size and power supply. The increasing integration trend in multi-core chips will increase the digital logic and consequently increase the power consumption of the entire chip. This increase in power is referred to as dark silicon and means that in multi-core systems, a significant percentage of the available cores cannot be active simultaneously due to temperature and power limitations. Typically, manufacturers of multi-core chips provide a number called the thermal design power for each chip. The TDP value is the maximum acceptable power consumption that the chip can consume without using efficiency degradation mechanisms; Among these mechanisms, we can mention the dynamic thermal management method. In embedded systems, due to the limited energy resources and the negative effect of the active cooling unit on reliability, there is usually no active cooling in many cases and a passive cooling is used, which significantly reduces TDP.

An important category of distributed embedded systems are systems that are used in applications with hard real-time constraints. Hard real-time embedded systems are usually used in applications that are directly related to human life, the environment, or the economy and security of countries. Among these applications, we can mention medical, aerospace, communication and military equipment systems. Therefore, any malfunction in these systems can lead to heavy and irreparable damage to life, money and the environment. Therefore, these systems must have very high reliability, which is achieved through fault tolerance. Hard real-time embedded systems with fault tolerance are designed and built to have three characteristics: 1) the work assigned to these systems is performed correctly and with complete accuracy, 2) the time constraints are met so that the hard real-time deadline is not violated, and 3) the TDP limit of the chip is respected; because violating any of the above may cause these systems to experience reliability degradation. In this type of system, fault tolerance is achieved through task repetition based on hardware and software redundancy. On the other hand, methods based on task repetition impose a very significant power consumption overhead on the system.

Hardware redundancy and time redundancy are usually used for fault tolerance in hard real-time embedded systems. Time redundancy uses idle time to recover from faults, which has less hardware overhead than hardware redundancy. This redundancy is only useful when there is enough time to rerun the failed tasks. On the other hand, with hardware redundancy, multiple instances of a task are usually executed simultaneously or overlapped during task execution. Therefore, power management is one of the important challenges in fault-tolerant embedded systems.