2J-4

コア温度情報による細粒度パワーゲーティング制御を行うOSスケジューラ

木村 一樹[†1]　近藤 正章[†2]　天野 英晴[†3]　宇佐美 公良[†4]
中村 宏[†5]　佐藤 未来子[†6]　並木 美太郎[†7]
東京農工大学大学院工学府情報工学専攻[†1]/ 電気通信大学情報システム基盤学専攻[†2]/ 慶應義塾大学大学院理工学研究科[†3]/
芝浦工業大学工学部情報工学科[†4]/ 東京大学大学院情報理工学系研究科システム情報学専攻[†5]/
東京農工大学大学院工学府[†6]/ 東京農工大学大学院工学研究院[†7]

1　はじめに

近年，システムLSIは高性能化の一方で消費電力の増大が顕著な問題となっており，各分野で省電力化の取り組みが行われている[1][2].

本プロジェクト[3]では，演算ユニットにランタイムパワーゲーティング(PG)技術を施したMIPS R3000ベースのCPU，Geyserの研究を行っている．Geyserではハードウェアが自律的にPGを行うことが可能だが，PG実施時のオーバヘッドと電力削減量の損益分岐点を考慮してソフトウェアでPGの実施方針を制御することで，省電力効果をさらに向上させることが期待できる.

本研究では，この損益分岐点となるサイクル数がコアの温度により変化することに着目し，ランタイムにコアの温度情報を取得してOSがPG実施ポリシーの決定を行う制御方式を提案する．本稿では，実際にFPGAボード上に評価システムを構築した上で，本提案方式をGeyserに実装し評価した結果を報告する.

2　Geyser概要

Geyserでは，ALU，SHIFT，MULT，DIVの各演算ユニット対して命令サイクルごとの細かい粒度で動的にパワーゲーティングを施すことができる.

2.1　PGStatusレジスタとスリープポリシー

パワーゲーティングの粒度をソフトウェア側から制御するために，特権レジスタであるPGStatusレジスタを有し，パワーゲーティング対象の各演算ユニットに対し，次の三つのスリープポリシーを定めることができる.

- 動的パワーゲーティング
- キャッシュミス時のみスリープ
- 常にアクティブ(スリープしない)

OSはPGStatusレジスタを活用し，細粒度パワーゲーティングのスリープポリシーを制御可能である.

2.2　細粒度PGの電力的特長

細粒度PGによりユニットがアクティブとスリープの間で状態遷移する際，電力のオーバヘッドが生じる．特に期間が短いスリープが頻発する場合スリープによる電力の削減量よりオーバヘッドが上回ることがある．このオーバヘッドに対しスリープによる省電力効果との損益分岐点となるスリープ期間（サイクル数）をBreak Even Point(BEP)と定義する．BEPより短いスリープを抑制することにより，省電力効果を高めることが期待できる．BEPの値はユニットにより異なり，また温度によって変化する．BEPの値に関しては，本稿の実験では先行研究[4]により求めた値を用いる.

3　コア温度情報を用いた制御方式の設計

3.1　スリープポリシーの制御

BEPは，図1に示すように，コア温度が高いほど短く，低いほど長くなるという特性を持つ．そこで本方式では，図2に示すようにユニットごとにある閾値温度θ_{TH}を定め，コア温度がθ_{TH}より高い場合は「動的パワーゲーティング」ポリシーを，低い場合は「キャッシュミス時のみスリープ」ポリシーを設定することで省電力効果の向上を図る.

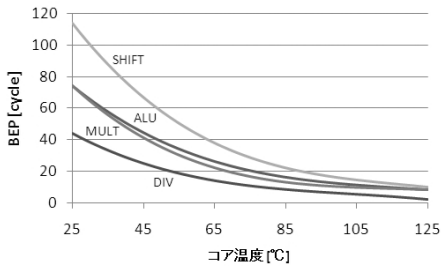

図 1: 各ユニットのBEPの温度変化

3.2　スリープ頻度特性による閾値温度の設定

それぞれのユニットのθ_{TH}の設定には，スリープ頻度情報を用いる．スリープ頻度情報とは，サイクル数nのスリープに対するその回数m_nの度数分布である．図3にスリープ頻度のイメージを図示する.

スリープ頻度情報$T_{sleep} = \{t_{n,m} | n, m \in \mathcal{N}\}$における全スリープサイクル数

$$t_{sleepAll} = \sum_{t \in T_{sleep}} t \tag{1}$$

に対し，

$$\sum_{n=1}^{n_{TH}} (m_n \times n) = \frac{1}{2} t_{sleepAll} \tag{2}$$

なるn_{TH}を求める．このn_{TH}に対して，ユニットごとに図4に示すようにθ_{TH}が得られる.

Fine Grain Power Gating Control with OS Scheduler Using Temperature Information of CPU Core

†1 Kazuki Kimura
Graduate school of Engineering, Tokyo University of Agriculture and Technology
†2 Masaaki Kondo
Graduate School of Information Systems, The University of Electro-Communications
†3 Hideharu Amano
Graduate School of Science and Technology, Keio University
†4 Kimiyoshi Usami
Department of Information Science and Engineering, Shibaura Institute of Technology
†5 Hiroshi Nakamura
Research Center for Advanced Science and Technology, The University of Tokyo
†6 Mikiko Sato
Graduate school of Engineering, Tokyo University of Agriculture and Technology
†7 Mitaro Namiki
Graduate school of Engineering, Tokyo University of Agriculture and Technology

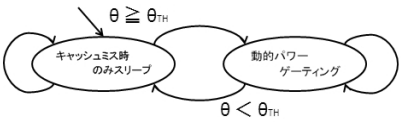

図 2: 温度閾値による PG ポリシー変更

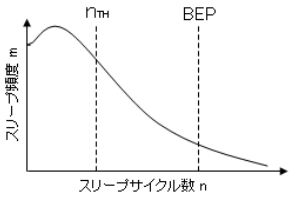

図 3: スリープ頻度

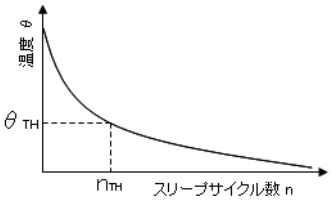

図 4: BEP 対温度

スリープ頻度特性から n_{TH} を求める方法には，

方法 1 アドミッションテスト（テストラン）により事前に得た情報を用いる方法

方法 2 実行時に動的情報を用いる方法

が考えられる．「方法 1」は実行時にスリープ頻度の測定のためのモジュールを必要としないものの，スリープ頻度の特性がタスクの性質によって異なるため，効率のよい制御がしにくいというデメリットがある．一方「方法 2」では実行時にスリープ頻度の測定のためのモジュールが必要だが，タスクの性質に応じた制御を行うことで省電力効果を効率よく向上させることが期待できる．そこで本稿では「方法 2」を用いる．

3.3 スリープ率による制御適用の判定

スリープ頻度の観測期間 t_{All} に対し

$$\frac{t_{sleepAll}}{t_{All}} \tag{3}$$

を「スリープ率」と定義する．使用頻度が極めて低いユニットは，温度によるスリープポリシー制御を適用するより「常に PG」とした方が消費電力を抑えられる場合がある．そこでスリープ率が高いケースでは「常に PG」ポリシー固定とする．

4 実装と評価

提案方式の PG ポリシー制御機構を，東京農工大学並木研究室で開発している組込み向け OS である Geyser OS[5] に実装した．また評価環境として FPGA ボード上に構築した Geyser による計算機環境，Geyser on FPGA を用いた．Geyser on FPGA はスリープ頻度の計測モジュールである PG パフォーマンスカウンタを含む各種入出力と主記憶を持つ．コア温度については，25[℃] から 125[℃] の範囲で変化する温度をエミュレーションした．また温度によるスリープポリシー制御を適用するスリープ率を 0.9 未満とした．

以下に，Geyser OS 上で MATRIX，QSORT，Dhrystone の各種ベンチマークタスクを実行したケースの評価結果を示す．パワーゲーティングを行わなかった場合，提案方式による制御を行わず常に「動的 PG」とした場合，提案方式の制御を行った場合の平均リーク電力の比較を図 5 に示す．ALU や MATRIX 実行時の MULT では提案方式による制御で PG 実施時のオーバヘッドを抑えた一方，MATRIX 以外の MULT や DIV

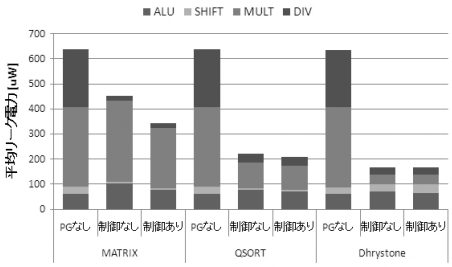

図 5: 平均リーク電力

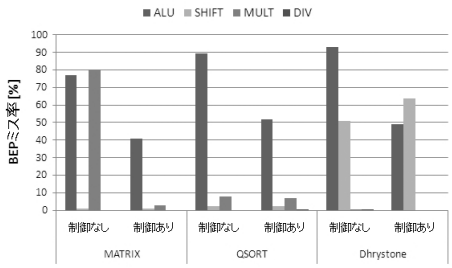

図 6: BEP ミス率

では動的 PG の効果が得られていることがわかる．全ケースの平均では約 9[%] のリーク電力削減を実現した．

BEP ミス率は，図 6 に示すように Dhrystone 実行時の SHIFT を除きすべてのケースで削減し，全体の平均で約 44[%] の改善を達成した．

提案方式によるパフォーマンスオーバヘッドは，およそ 5.6[%] であった．

5 おわりに

本稿では，コア温度情報を用いて OS が PG 実施方針を決定する制御方式を提案した．また提案方式について評価を行い，平均リーク電力の削減を確認した．

方法 1 との比較やスリープ頻度特性の判定方法を改良し更なる効果の向上を図ることは今後の課題である．

謝辞 本研究は, 科学技術振興機構「JST」の戦略的創造研究推進事業「CRSET」における研究領域「情報システムの超低電力化を目指した技術革新と統合化技術」の研究課題「革新的電源制御による次世代超低電力高性能システム LSI の研究」によるものである．

参考文献

[1] James Donald, Margaret Martonosi, "Power Efficiency for Variation-Tolerant Multicore Processors", International Symposium on Low Power Electronics and Design (Proc. ISLPED'06) SESSION10, No.3, pp.304-309, (Oct 2006).

[2] Pratap Ramamurthy et al., "Performance-directed Energy Management using BOS", ACM SIGOPS Operating Systems Review, Vol.41, pp.66-77, (2007).

[3] 中村宏, 天野英晴, 宇佐美公良, 並木美太郎, 今井雅, 近藤正章, 「革新的電源制御による超低消費電力高性能システム LSI の構想」, 情報処理学会研究報告 ARC-173, pp.79-84 (Jun 2007).

[4] 関直臣 他「MIPS R3000 プロセッサにおける細粒度動的スリープ制御の実装と評価」, 情報処理学会研究報告 2008-ARC-176, pp.71-76, (Jan 2008).

[5] 砂田徹也 他「省電力 MIPS プロセッサにおける OS の試作とシミュレーションによる電力評価」, 情報処理学会「システムソフトウェアとオペレーティング・システム」第 108 回研究報告, Vol.2008-OS-108, pp.163-170 (Apr 2008)