

「実用化を目指した組込みシステム用ディペンダブル・オペレーティングシステム」
平成 18 年度採択研究代表者

佐藤 三久

(筑波大学計算科学研究センター 教授)

「省電力でディペンダブルな組込み並列システム向け計算プラットフォーム」

1. 研究実施の概要

本研究では、ユビキタス情報社会における高度な情報処理の要請に対し、これからの高性能組込みシステムはマルチコア・マルチチップになることを想定し、ディペンダブル OS の一部として、省電力高信頼組込み並列プラットフォームを研究開発する。並列システムでの耐故障性機能を実現し、実時間制約下で並列性制御と省電力化を行うための電力制御機構を研究開発する。さらに、並列処理を効率的に行うための低電力高性能高信頼通信機構として、複数のネットワークリンクを適宜用いることにより、電力制御・性能制御・耐故障性を包括的に実現するためのハードウェア及び通信機構の開発を行う。これらの技術を統合し、実証プラットフォームを研究開発する。

当該年度においては、研究領域全体のテーマである組込みディペンダブル OS に向けて、研究領域での本研究の位置づけを明らかにし、研究領域の他のチームとともに、システムの全体設計について検討した。それに基づいて、本研究の技術要素について見直し検討ならびに、基本設計を行った。

2. 研究実施内容

本研究の全体目標である実証プラットフォームならびに核となる低電力高速インターコネクトに関してはチーム全体で議論を行い、基本設計を行った。これからの並列組込みシステムに欠くことのできない低電力高速インターコネクトとして、次世代の IO 規格である **PCIExpress Gen2** を先取りして開発することとし、これをチップ間のインターコネクトとして用いるための、並列システム用インターコネクトブリッジを搭載した高速インターコネクトネットワークチップについて基本設計を行った。以下に、各研究項目の研究実施内容をまとめる。

1、並列組込み向け高信頼共有メモリ機構および省電力実時間並列実行制御機構（電力制御・高信頼並列グループ）

当該年度においては、電力制御機構と並列組込みシステム向け高信頼化機構について、

基本設計を行った。研究領域の他のチームとともに、システムの全体設計について再検討した結果、プロセッサ電力制御機構については、これからの並列組込みシステムのターゲットとしてサーバーアプリケーションなどの HA(high availability)システムにおいても耐故障機能と併せ、検討することとした。また、高信頼化技術については、個々のプロセッサにおいて実行されるプログラムにおいてもチェックポイントを行い、障害検出時に動的に fail-over する機構について開発し、広範囲の組込みシステムの高信頼化のための研究を行うこととした。

基本設計を進めるにあたって、現時点で入手可能な組込みプロセッサ (SH4) を利用した並列システムを構築し、Linux をインストールして組込み並列システムの開発プラットフォームを整備した。また、プラットフォームとして低電力プロセッサを用いたクラスタについても整備した。すでに準備研究がある分散共有メモリ機構については開発プラットフォーム上にて準備研究をベースに試作を行い、冗長性をもった共有ページ管理機構による高信頼技術について検討した。同時に、電力制御機構に関しては、組込み向けのプロセッサではマルチコアプロセッサが入手することがまだ難しいため、既存のマルチコアプロセッサにて実験環境を構築し、コアごとの電力制御についての検討を進めた。

また、基本設計と同時に組込みシステムのアプリケーションおよびベンチマーク、および組込みプロセッサのマルチコアの技術動向についての調査を行った。

2、並列システム内高信頼高性能通信機構 (通信システムアーキテクチャ・グループ)

耐故障・省電力ネットワークリンクをメイン CPU 側からアクセスするための基本的なアーキテクチャに関し、高速ネットワークグループと連携して新規開発予定の Communicator チップの機能・性能に関する基本仕様の検討を行った。Communicator チップ内では高速ネットワークリンクの制御、ルーティング、及びメイン CPU とのインタフェースのために、省電力組み込みプロセッサを内蔵し、DMA 制御装置を通じて各ブロック間の高速データ移動処理を行う(図1参照)。

また、メイン CPU と Communicator チップは PCI バスを介して接続することとし、この上でメッセージ通信及びページ単位でのリモート DMA 通信を効率的に実現するプロトコル API として、OFED により策定されている OpenIB インタフェースの一部を利用した仕様を用いることを検討した。これにより、既存のプラットフォーム上でのドライバ等のソフトウェア開発を円滑に行い、かつ全てのプロトコルを新規検討することなく、作業を進めることが可能になる。

また、本プロジェクトで新規開発する高速ネットワークリンク以外にも既存のコモディティネットワークリンクとして Gigabit Ethernet を用いた通信レイヤにおいても、耐故障ネットワークリンクを実現させ、最終的には同じユーザ API によってどちらのネットワークも利用可能にする予定であるため、Gigabit Ethernet に基づく耐故障通信ドライバの開発を先行して行った。一般的な PC サーバ Linux 上に本システムを実装し、2リンクの Gigabit Ethernet においてドライバレベルでの TCP/IP 通信が、平常時ではバンド幅として2倍の性能を実現し、リンク故障時には性能縮退しつつ通信を継続する機能が実装できたことを確認した。

3、低電力高速インターコネクトと省電力高密度並列ハードウェアプラットフォームの開発（高速ネットワークグループ）

ハードウェアプラットフォームの全体構成仕様及び各基本ユニットのモジュールレベルの検討を実施した(図1)。基本ユニットはプロセッサ、DRAM、フラッシュメモリ、及びPCI-express をベースとしたコミュニケーションチップ (Communicator) から構成する。Communicator チップは本ハードウェアプラットフォームの中核をなすもので基本ユニット間の高速ネットワークインターコネクト構成し、新規設計を実施する。図2に Communicator チップのブロック図を示す。

Communicator チップは、2008 年度のデモの時期に実用化されるであろう、65nm CMOS シリコン技術を用い、インターコネクト技術として、PCI-express の次世代規格である、PCI-express Generation2 を搭載する。

PCI-express Gen.2を構成するLink 層と物理層の各モジュールの基本検討を完了し、ディペンダビリティ機能として、次年度以降、以下の技術を開発する。

Generation 2(5Gbps)/Generation 1(2.5Gbps)の動的制御により、パワーマネジメントを可能とする。レーン構成としては、4 レーン対応のリンクモジュールを4個搭載する。未使用レーンでの電源ターンオフ等、各レイヤのモジュールも電力制御を行う。上位レイヤ設計にフレキシビリティを持たせるために、送受信に係わるパケット生成、及びヘッダ解析は図2にあるようにソフトウェアで制御する構成を採用した。物理層として、65nm CMOS シリコン技術による、アナログ要素回路技術の開発を実施した。

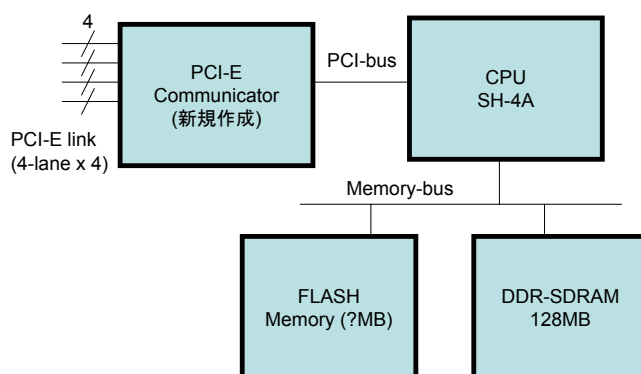


図1 基本構成

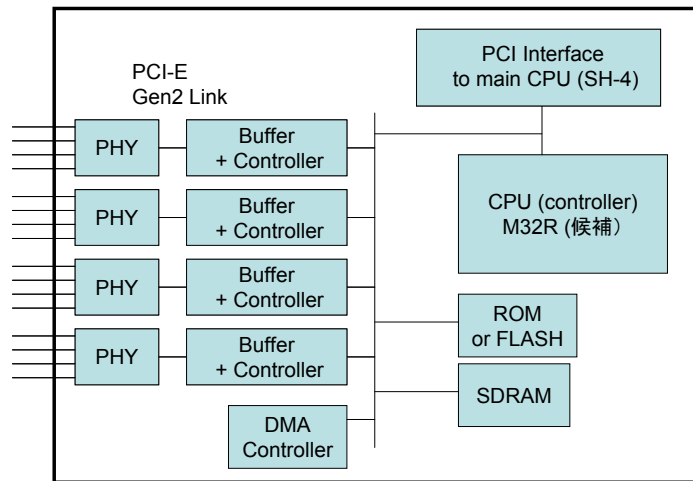


図2 PCIExpress Gen2 Communicator チップ構成

3. 研究実施体制

(1)「電力制御・高信頼並列システム」グループ

①研究分担グループ長:佐藤 三久(筑波大学 教授)

②研究項目

- ・ 並列組込み向け高信頼共有メモリ機構および省電力実時間並列実行制御機構

(2)「通信システムアーキテクチャ」グループ

①研究分担グループ長:朴 泰祐(筑波大学 教授)

②研究項目

- ・ 並列システム内高信頼高性能通信機構

(3)「高速ネットワーク」グループ

①研究分担グループ長:有本 和民(ルネサステクノロジ 副統括部長)

②研究項目

- ・ 低電力高速インターコネクトと省電力高密度並列ハードウェアプラットフォームの開発