FPGA에서 에너지 효율이 높은 데이터 경로 구성을 위한 계층적 설계 방법

장 주 욱[†] · 이 미 숙[††] · Sumit Mohanty[†††] · 최 선 일[†††] · Viktor K. Prasanna[††††]

요 약

본 논문은 FPGA 상에서 에너지 효율이 높은 데이터 경로 설계 방법론을 제안한다. 에너지, 처리시간, 그리고 면적간의 트레이드오프를 이해하기 위하여, 도메인 특성 모델링, coarse-grained 성능평가, 설계공간 조사, 그리고 로우-레벨 시뮬레이션 과정들을 통합한다. 도메인 특성 모델링 기술은 도메인의 특성에 따른 시스템 전체의 에너지 소모에 영향을 미치는 여러 가지 구성요소와 파라미터들을 식별함으로써 하이-레벨 모델을 명시한다. 도메인이란 주어진 어플리케이션 커널의 알고리즘에 대응하는 아키텍쳐 패밀리이다. 하이-레벨 모델 또한 에너지, 처리시간 그리고 면적을 예측하는 함수들로 구성되어 트레이드오프 분석을 용이하게 한다. 설계 공간 조사(DSE)는 도메인에 명시된 설계 공간을 분석하여 설계 셋을 선택하도록 한다. 로우-레벨 시뮬레이션은 설계 공간 조사(DSE)에 의해 선택된 설계와 최종 선택된 설계의 정확한 성능평가를 위하여 사용된다.

본 논문에서 제안한 설계 방법은 매트릭스 곱셈에 대응하는 알고리즘과 아키텍쳐 패밀리를 사용한다. 제안된 방법에 의해 검증된 설계는 에너지, 처리시간과 면적간의 트레이드오프를 보인다. 제안된 설계 방법의 효율성을 보이기 위하여 Xilinx 에서 제공되는 매트릭스 곱셈 커널과 비교하였다. 성능 비교 메트릭으로 평균 전력 밀도(E/AT)와 에너지 대 (면적 x 처리시간)비를 사용하였다. 다양한 문제의 크기에 대하여 Xilinx 설계들과 비교하였을 때 제안한 설계 방법이 전력밀도(E/AT)에서 평균 25% 우수하였다. 또한 본 논문에 제안한 설계의 방법을 MILAN 프레임워크를 이용하여 구현하였다.

키워드 : 에너지 최적화, 임베디드 시스템 설계, 재배열 컴퓨팅

A Model-based Methodology for Application Specific Energy Efficient Data path Design Using FPGAs

Jang, Ju-Wook[†] · Lee, Mi-Sook[††] · Sumit Mohanty[†††] · Seonil Choi[†††] · Viktor K. Prasanna[††††]

ABSTRACT

We present a methodology to design energy-efficient data paths using FPGAs. Our methodology integrates domain specific modeling, coarse-grained performance evaluation, design space exploration, and low-level simulation to understand the tradeoffs between energy, latency, and area. The domain specific modeling technique defines a high-level model by identifying various components and parameters specific to a domain that affect the system-wide energy dissipation. A domain is a family of architectures and corresponding algorithms for a given application kernel. The high-level model also consists of functions for estimating energy, latency, and area that facilitate tradeoff analysis. Design space exploration(DSE) analyzes the design space defined by the domain and selects a set of designs. Low-level simulations are used for accurate performance estimation for the designs selected by the DSE and also for final design selection.

We illustrate our methodology using a family of architectures and algorithms for matrix multiplication. The designs identified by our methodology demonstrate tradeoffs among energy, latency, and area. We compare our designs with a vendor specified matrix multiplication kernel to demonstrate the effectiveness of our methodology. To illustrate the effectiveness of our methodology, we used average power density(E/AT), energy/(area x latency), as themetric for comparison. For various problem sizes, designs obtained using our methodology are on average 25% superior with respect to the E/AT performance metric, compared with the state-of-the-art designs by Xilinx. We also discuss the implementation of our methodology using the MILAN framework.

Key Words : Energy Optimization, Embedded System Design, Reconfigurable Computing

1. 서 론

FPGA(Field Programmable Gate Arrays)는 신호 처리에 있어 DSP나 ASIC 보다 융통성이 있는 매력적인 대안이다. FPGA에서 사용할 수 있는 높은 처리 능력은 휴대용 제품에 사용되는 신호 처리 커널과 같은 복잡하고, 계산량이 많은 어플리케이션을 구현하기 위한 매력적인 수단이 된다 [11]. 휴대용 제품은 에너지 소모가 제한된 환경에서 동작되므로 처리시간, 면적과 더불어 에너지는 주요 성능 지표이다. 에너지 효율 데이터 경로 설계를 위한 전통적인 방법은

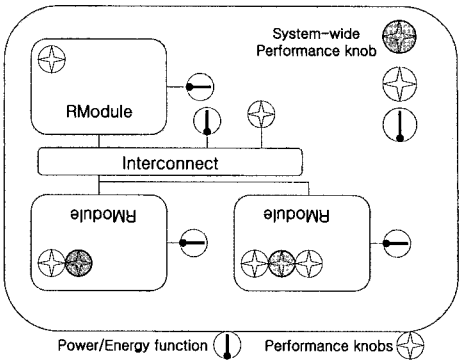
※ 본 논문은 정보통신부의 출연금으로 수행한 IT SoC 핵심설계인력양성 사업의 연구결과임.
† 정 회 원 : 서강대학교 전자공학과 교수
†† 준 회 원 : 서강대학교 전자공학과 박사과정
††† 비 회 원 : 미국 University of Southern California 전자공학 박사과정
†††† 비 회 원 : 미국 University of Southern California 컴퓨터학과 교수
논문접수 : 2005년 3월 4일, 심사완료 : 2005년 8월 5일

RTL이나 게이트 레벨에서 최적화를 이루기 위한 로우-레벨 설계 툴의 사용을 포함하는 것이다. 그 기술은 시간 소모적이고 에너지 효율성의 향상이 하이-레벨 최적화와 비교하여 훨씬 적다는 점에서 비 효과적이다. 본 연구는 알고리즘 레벨에서 에너지 손실의 최적화는 RTL이나 게이트 레벨의 최적화보다 시스템 총 에너지의 손실에 더 높은 영향을 미친다는 것을 보인다[17]. 에너지 최적화에 대한 영향의 비율은 알고리즘, 레지스터, 그리고 회로 레벨 기술에서 각각 20:25:1라 보고된 바 있다[16]. 더욱이 FPGA를 사용하는 설계는 에너지, 처리시간, 그리고 면적간의 성능의 균형을 이루어야 한다. 이 균형을 이루기 위하여 설계자는 에너지 대 처리시간, 에너지 대 면적, 에너지 대 I/O등 다양한 트레이드오프를 고려하여야 한다.

설계자가 FPGA를 이용하여 에너지 효율 시스템을 설계하는 데는 직면하는 몇 가지 문제들이 있다. FPGA의 유연성은 설계 면적을 크게 한다는 것이다. Xilinx XPower등과 시간 소모적인 로우-레벨 시뮬레이션에 큰 면적을 사용하는 것은 비실용적이다[18]. 700MHz Pentium III Xeon 에서 돌아가는 시뮬레이터에서 단지 3x3 매트릭스 곱셈기의 에너지 손실을 예측하는데 평균 2~3시간이 필요하였다. 또한 FPGA는 RISC 프로세서와 같은 하이-레벨 구조를 표시하지 않는다. 만일 그런 하이-레벨 구조를 보일 수 있다면 알고리즘 레벨 설계, 최적화 그리고 분석 등을 하는 하이-레벨 모델을 명시하기 위하여 이용될 수 있다. FPGA를 위한 하이-레벨 모델은 실제로 매핑되는 아키텍쳐에 따라 좌우된다.

위의 이슈를 해결하기 위하여 본 논문에 제안한 설계는 아키텍쳐 패밀리를 모델로 도메인 특성 모델링 기술을 이용하고 어플리케이션을 구현하는 알고리즘에 대응하는 설계 방법을 제안한다. 하이-레벨 모델의 결과는 가상 파라미터 형태로 데이터 경로를 표현할 수 있다(그림 1). 예를 들어, 운영 주파수, 기억용량, 대역폭, 그리고 정밀도들은 설계에 필요한 파라미터들이다. 이들의 다양한 설정이 에너지, 처리시간 그리고 면적간의 트레이드오프를 제공하고 시스템 설계자로 하여금 성능 요구에 기반한 적절한 설정을 선택하도록 한다. 각 콤포넌트와 연관된 전력 함수는 콤포넌트의 전력 소모상의 다양한 성능의 영향을 반영한다. 본 논문에 제안한 설계의 접근방법은 도메인 특성 모델링, 도메인의 하이-레벨 파라미터를 추출한 후, 다양한 알고리즘 레벨 최적화에 적용하는 top-down 방식이다. 최적화의 기술들은 a) 파라미터들의 적절한 설정을 검증 b) 알고리즘 특성에 기반한 아키텍쳐 수정이다.

이 논문에서 어플리케이션이라 함은 매트릭스 곱셈, FFT 등과 같은 동작이다. 설계 공간 조사(DSE)는 아키텍쳐-알고리즘 도메인상에서 에너지, 처리시간 그리고 면적의 관점에서 설계의 질을 평가하는 것이다. 본 논문의 설계에서는 성능평가에 에너지, 처리시간 그리고 면적에 대한 측정 결과를 이용한다. 면적 메트릭은 FPGA의 특성에 따르는데, 예를 들면 Xilinx Virtex에서는 사용된 슬라이스의 수를 가리킨다. 본 논문에 제안한 설계 방법에서 설계자는 주어진 어



(그림 1) 데이터 경로

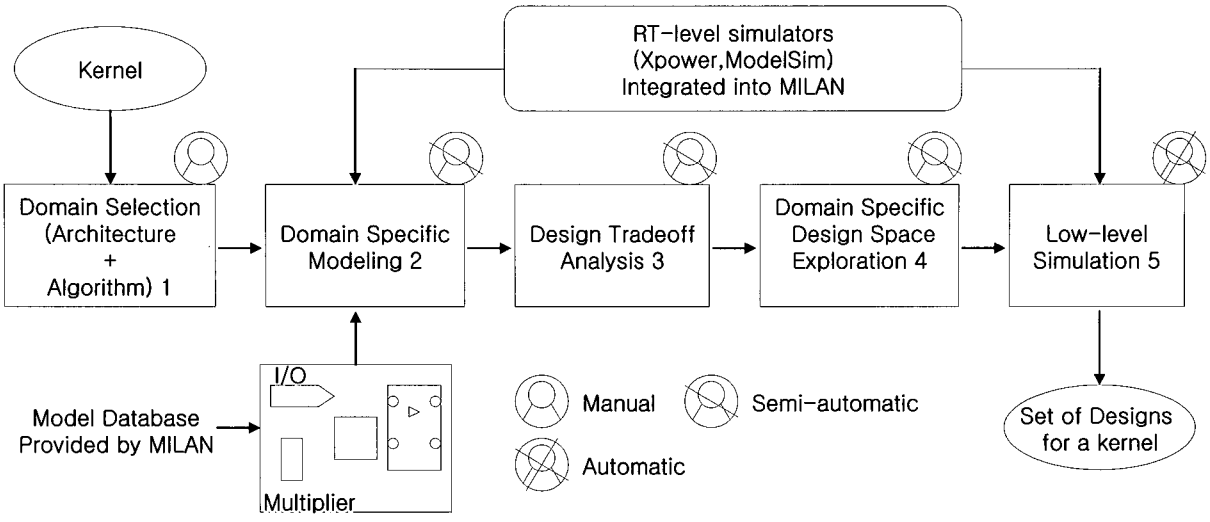
플리케이션을 위하여 처음에 아키텍쳐 패밀리와 알고리즘(도메인)을 선택하고 그 다음 하이-레벨 모델을 명시한다. 이 모델은 다양한 아키텍쳐 파라미터를 포함하고, 선택된 아키텍쳐-알고리즘 쌍에 대한 데이터 경로의 빠른 에너지 예측에 사용할 전력 함수를 제공한다. 설계자는 도메인과 하이-레벨 모델에 기반하여 DSE 기술을 검증한다. 이 모델은 알고리즘에서 최적화를 표현할 수 있을 정도의 추상화된 레벨에서 설계된다. 예를 들어 레지스터의 수, 곱셈기수, 운영 주파수, 곱셈기의 형태 등을 가리킨다. 로우-레벨 시뮬레이션은 전력함수의 정확도를 검증하기 위하여 사용된다. 이 논문은 어플리케이션에 적합한 설계 방법에 초점을 둔다. 모델 정의의 상세내용은 [3]에서 찾을 수 있다. 본 논문에 제안한 설계 방법을 구현하기 위하여 어떻게 MILAN 프레임워크[1]를 설정하였는지 보인다.

설계방법의 효율성을 보이기 위하여 제안된 방법을 이용한 설계와 Xilinx 매트릭스 곱셈 커널과 비교한다. 제안된 설계는 에너지, 처리시간, 그리고 면적간의 트레이드오프를 보인다. 성능 평가를 위하여 평균 전력 밀도(E/AT) 메트릭을 정의하였다. 이 메트릭에서는 문제의 크기와는 무관하게 설계가 면적과 시간에 대하여 최적화되었다고 가정하고, 단위 면적당 손실되는 평균 전력의 면에서 도메인을 결정한다. 이 가정을 기반으로 E/AT의 값이 적을수록 좋다. E/AT 메트릭에 기반하여 본 논문에 제안한 설계는 Xilinx의 설계보다 평균 25% 우수하였다. E/AT는 에너지/(면적x처리시간)으로 표현된다.

본 논문의 구성은 다음과 같다. 제2장은 관련연구를 소개하며, 제3장에서는 제안된 설계 방법을 논한다. 제4장에서는 MILAN을 이용하여 이 설계방법을 구현한다. 제안된 방법을 설명한 그림의 예와 기존의 설계를 이용한 방법과의 결과 비교가 제5장에서 제시된다. 제6장은 결론부분이다.

2. 관련연구

다양한 신호 처리 커널의 성능 최적화를 위한 FPGA 모델링에 대한 여러 연구가 이루어졌다[2, 4, 8]. 이 연구들은 효율적 데이터 경로 설계를 통한 처리시간의 최적화에 초점이 맞추어져 있다. Luk등은 동적 재설정 시스템의 모델링과



(그림 2) 설계 방법

시간 성능을 최적화하는 몇 가지 방법을 제안하였으나[11] 하이-레벨 모델링을 통한 FPGA기반 구현의 에너지 최적화에 대한 연구는 없다.

로우-레벨 전력 예측과 최적화에 대한 여러 연구가 빠르게 이루어졌다. Wolff등은 RTL과 IP(Intellectual Property) 콤포넌트간의 전력과 처리시간을 특징화하기 위하여 미리 계산된 테이블을 사용하는 방법을 제안하였다[17]. 이 방법은 주어진 처리시간의 요건을 맞추기 위한 최저 에너지 손실을 가지는 콤포넌트로 에너지 효율 설계를 이루었다. 그런데 이 방법은 주파수, 메모리의 크기등과 같이 다양한 변화로 향상된 에너지 성능을 이룰 수 있는 부분들은 개발하지 않았다. Nemani등은 조합회로의 함수설명과 평균적인 동작에 기반한 빠른 전력 예측 방법을 제안하였으나[14] 이 방법은 전력의 최적화에 대한 부분은 없다. 본 논문에서는 이 방법을 전력 예측의 툴로 사용하였다. Garcia등은 파이프라인 아키텍쳐를 통하여 에너지를 최적화 시키는 방법을 연구하였는데[6], 면적의 증가가 더 높은 에너지 손실의 필수적인 결과를 가져오진 않는다는 것을 보였다.

Raghunathan등은 임베디드 시스템상의 다른 콤포넌트들의 전력 손실을 측정하는 방법을 제안하였다[16]. 이 방법은 bottom-up 접근 방법을 이용하는데, 로우-레벨 측정 방법을 수정하여 전력 측정의 속도를 높인다. 그러나 본 논문에 제안한 설계의 방법은 반대로 예측과 최적화를 위하여 top-down 접근 방법을 사용한다. 제안된 설계의 방법은 로우-레벨 툴을 사용하는 설계를 구현하기 전에 효율적 설계임을 검증하기 위하여, 아키텍쳐-알고리즘(추상화) 레벨부터 알고리즘적 최적화를 수행한다.

Xilinx의 Virtex-II Pro FPGA를 가진 설계 툴은[18] 알고리즘과 아키텍쳐 레벨의 설계 공간을 조사하는 하이-레벨 추상화 대신 게이트-레벨 또는 RT레벨의 최적화에 집중한다. 본 논문에 제안한 설계의 방법은 상호보충적이며 Xilinx 툴을 사용하여 좀더 최적화 될 수 있다.

3. 설계방법

설계 방법의 목표는 어플리케이션에 따른 에너지 효율 데이터 경로 설계이다. 이 목표를 달성하기 위하여 본 논문에 제안한 설계의 방법은 에너지, 처리시간, 면적간의 트레이드오프를 제공하는 설계-셋을 제안한다. 설계자는 선택영역과 성능 메트릭을 기반으로 적절한 설계인지 조사하고 검증한다. 본 논문에 제안한 설계 방법이 (그림 2)에 설명되어 있다. 각 단계의 자동화 레벨은 제4장에서 논의될 것이다.

3.1 도메인 선택

각 커널을 위하여 몇 가지의 아키텍쳐 패밀리 후보들이 있을 수 있다. 예를 들어, 선형 어레이 프로세서, 2차원 어레이 프로세서 그리고 캐쉬 기반의 단일프로세서들은 광대하게 사용되는 아키텍쳐 패밀리이다. 각 아키텍쳐 패밀리와 더불어 이들을 구현하는 여러 알고리즘들이 존재한다.

과거의 여러 연구가 다양한 아키텍쳐 패밀리를 검증하였고[5, 10, 15], 각각은 I/O 복잡도, 메모리 요구사항, 면적 등의 관점에서 다른 특성을 가진다. 본 논문에 제안한 설계는 필수 성능, 용량 그리고 FPGA칩의 한계에 기반하여 적절한 아키텍쳐-알고리즘 패밀리를 검증한다. 적절한 도메인의 검증으로 요구되는 성능에 가장 적절하고 효율적인 설계와 에너지, 처리시간 그리고 면적간의 트레이드오프를 달성하기 위한 여러 아키텍쳐 파라미터를 찾아낸다. 이 단계는 설계자가 진행하는 과정으로써 도메인 검증을 위한 알고리즘과 아키텍쳐 선택에 설계자의 전문성이 필요하다.

3.2 도메인 특성 모델링

도메인 특성 모델링은 특정 도메인의 하이-레벨 모델을 개발한다. 설계결과의 에너지 손실 분석에 필요한 아키텍쳐 파라미터를 검증하기 위하여 도메인에 대한 상세한 지식이 요구된다. 하이-레벨 모델은 기본 콤포넌트로 재배치가능 모듈(RModule)과 내부연결(Interconnect)로 구성된다. 말하

자면, 운영 주파수(f), 정밀도(w), 메모리크기(s) 그리고 서로 연관된 각 콤포넌트의 전력 상태(ps)등의 아키텍쳐 파라미터들이다. 각 모델 파라미터 값의 범위를 검증하여 설계 공간을 줄인다. 예를 들어, 만일 최대 처리시간등과 같은 성능 제한이 있다면, 최대, 최소 문제크기는 몇 개의 파라미터들에게 영향을 미친다. 따라서 특정 도메인에서 변화시킬 수 없는 파라미터들이 있다면 그것은 모델링 과정에서 제외된다.

하이-레벨 모델의 또 다른 중요한 면은 각 콤포넌트와 연관된 전력 함수인데, 콤포넌트(재배치가능 모듈과 내부연결)의 전력 작용을 특성화한다. 함수는 전력 손실 콤포넌트와 관련된 파라미터들의 변화에 영향을 받는다. 로우-레벨 시뮬레이션에서 얻어진 곡선은 함수를 결정하는데 사용된다. 하이-레벨 모델은 또한 문제의 크기에 따라 면적과 처리시간을 평가하는 함수와 아키텍쳐 파라미터들을 반영한다. 결과적으로, 전력 함수와 콤포넌트 전력 상태(CPS) 매트릭스-셋이 시스템 전체적 에너지 함수를 추론해내는데 사용되며, 콤포넌트 전력 상태(CPS) 매트릭스들은 알고리즘 설명으로부터 유추된다. 이 매트릭스들은 각 사이클에서 모든 콤포넌트의 전력 상태를 표시한다. 도메인 특성 모델링의 상세한 부분과 전력 함수를 예측하는 방법 그리고 시스템 전체 에너지 함수는 [3]에서 찾을 수 있다.

3.3 설계 트레이드오프 분석

하이-레벨 모델은 각 콤포넌트와 연관된 전력 함수, 에너지, 처리시간 그리고 면적성능 함수등과 같은 몇 개의 함수를 표시한다. 이 함수들은 다른 성능 매트릭스들(에너지, 처리시간 그리고 면적)의 트레이드오프를 분석하는데 사용되며, 또한 다양한 아키텍쳐 파라미터들과 관련한 성능의 민감도을 표시한다. 예를 들어, 만일 파라미터를 변화시킬 수 있는 콤포넌트와 연관된 전력 함수는 이 파라미터와 관련한 전력 민감도를 분석할 수 있다.

3.4 설계 공간 조사(DSE)

설계 공간 조사 중 도메인 특성 설계 공간은 설계자가 선택영역 기준에 기반하여 검증한다. 본 논문에 제안한 설계의 방법은 최소 에너지 손실을 가진 설계와 최소 면적x처리시간을 가진 설계 중 선택한다. 도메인 특성 모델링 방법은 유효한 설계에만 설계 공간을 허용하므로 설계 공간이 크지 않다. 게다가, 서로 다른 성능 메트릭스들과 관련된 함수들을 사용하므로, 선정 기준에 맞는 설계-셋을 엄격한 방법으로 평가할 수 있다. 또한 이것은 설계자로 하여금 다양한 성능의 본질과 설계 공간 조사(DSE)를 위한 효율적인 방법을 구현하기 위한 전력 함수들을 개발할 수 있게 한다.

모델 파라미터들과(하이-레벨 모델로 정의된)범위 그리고 다양한 성능 메트릭스를 평가하는 함수들은 DSE단계에서 입력되며, 선택영역 기준은 더 추가될 수 있다. DSE의 출력은 단일 설계 또는 선택영역 기준을 만족시키는 설계-셋이다.

3.5 로우-레벨 시뮬레이션

로우-레벨 시뮬레이션은 DSE단계에 의해 선택되는 설계에 적용된다. DSE단계에서는 설계를 평가하기 위하여 다양한 함수들을 사용한다. 예측이 합리적이고 정확하다면(제5장에 보인다), 두 개의 다른 목적으로 로우-레벨 시뮬레이션을 사용한다. 본 논문에 제안한 설계의 하이-레벨 예측에 의한 에러는 대개 ±10% 범위임을 보인다. 그러므로 두 개의 후보 설계가 어떤 성능 메트릭에 대해서도 서로 10%이내에 있을 때, 로우-레벨 시뮬레이션을 이용한다. 또 다른 용도는 하이-레벨 모델에서 제공된 함수를 사용하여 평가된 성능 예측을 검증하는데 사용한다.

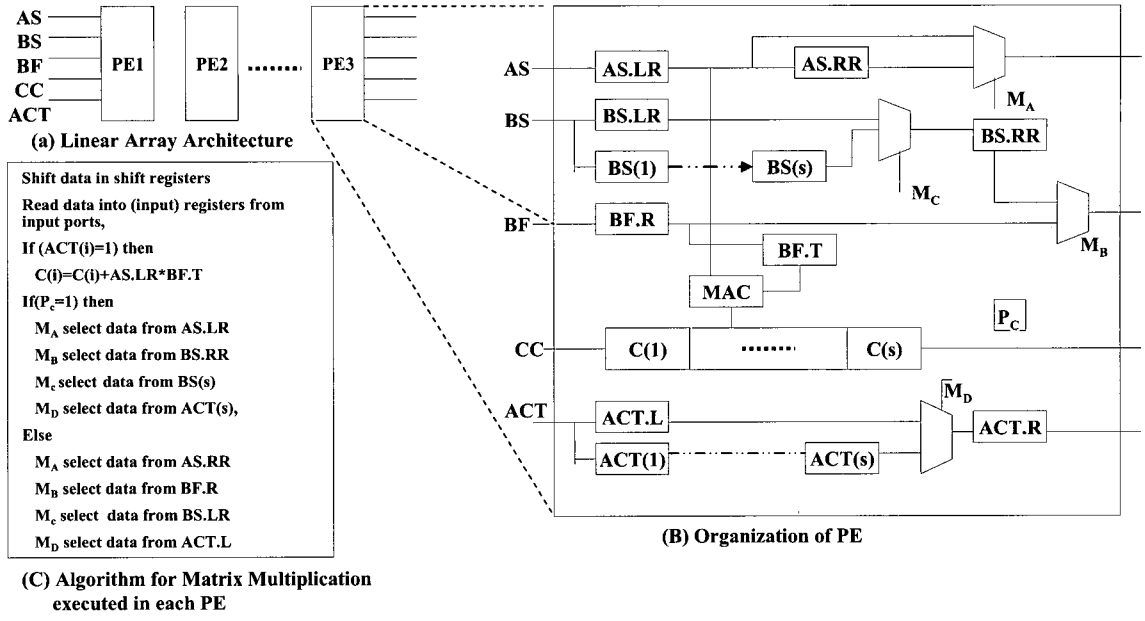
또한 로우-레벨 시뮬레이션을 이용하여 하이-레벨 모델과 연관된 전력 함수들을 예측할 수 있다. 전력 함수 예측을 위한 로우-레벨 시뮬레이션 사용의 상세설명은 [3]에서 찾을 수 있다.

4. MILAN을 이용한 방법 구현

모델 기반 통합 시뮬레이션(MILAN)은 모델 기반의 확장 프레임워크이다. MILAN은 특정 도메인에 적합한 모델링 언어를 제공하도록 설정될 수 있고, 다양한 시뮬레이터들과 툴들을 통합할 수 있으므로, 본 논문에 제안한 설계 방법에 적절하다. 시뮬레이터들과 툴들(이후로 둘 다 툴이라 한다)은 모델 번역기(MI)라 알려진 소프트웨어 콤포넌트들을 통하여 MILAN으로 통합된다. 각 툴은 자신의 모델 번역기(MI)들과 연관된다. 모델 번역기(MI)들은 모델에 저장된 정보를 툴에서 요구되는 형식으로 번역한다. 예를 들어, SimpleScalar(사이클이 정확한 MIPS 프로세서를 위한 시뮬레이터)의 통합은 설정파일과 어플리케이션을 위한 C코드를 생성하며, 시뮬레이션결과로 얻은 피드백을 제공하는 모델 번역기(MI)들의 셋을 포함한다. MILAN 프레임워크에 대한 상세한 설명은 [1]에서 찾을 수 있다.

본 논문에 제안한 설계는 제3장에서 설명된 설계 방법을 구현하기 위하여 어떻게 MILAN을 설정하였는지 설명한다. 첫 단계는 MILAN을 포함하지 않고, 대상 어플리케이션에 적절한 도메인 검증을 알고리즘과 아키텍쳐에서 설계자의 전문성을 요구한다. 도메인이 검증되면 MILAN이 도메인 특성 모델링에 적절한 언어를 제공하기 위하여 설정되고, 그래픽 모델링 툴인 GME2000을 통하여 사용자 인터페이스(UI)를 제공한다. UI는 2단계에서 후보 도메인을 모델링 하는데 사용되는 레지스터, 곱셈기, 덧셈기, SRAM등과 같은 기본 모듈들에 대한 대표적인 그래픽 블록들로 구성된다. 각 아키텍쳐 파라미터들은 도메인 분석을 통하여 얻은 적절한 범위의 값이다. MILAN은 로우-레벨 시뮬레이션과 전력 함수를 자동적으로 예측하기 위하여 하이-레벨 모델에 제공된 정보를 사용한다[3].

3단계로써 MILAN은 설계자의 시각적 검증에 의해 다양한 함수들을 만들어내는데 사용된다. MILAN에서 도출되는 하이-레벨 모델의 상세한 결과는 4번째 단계인 설계 공간 조사(DSE)에서 사용된다. 설계 공간 조사(DSE) 툴들은 MILAN에 통합되어 모델 번역기(MI)로부터 툴로 전달되는



(그림 3) 매트릭스 곱셈 아키텍쳐, PE 구조, 상응하는 알고리즘

정보와 연관된다. MILAN을 사용한 설계 공간 조사(DSE)의 그림을 [13]에서 찾을 수 있다. 2번째, 3번째 그리고 4번째 단계는 반-자동 단계로써, 설계자가 하이-레벨 모델의 상세한 내용, 설계 공간 조사(DSE) 툴의 선택, 설계 공간 조사(DSE)에 대한 선택영역 기준 등을 입력하여야 한다.

DSE단계(5단계)는 후보 설계-셋을 검증하여 MILAN 설계 데이터베이스에 저장한다. 설계자는 적절한 모델 번역기를 불러내어, 후보 설계들을 위한 로우-레벨 시뮬레이터를 설정한다. 본 논문에 제안한 설계는 전력과 처리시간 각각의 로우-레벨 시뮬레이션 수행을 위하여 MILAN에 XPower와 ModelSim을 통합하였다.

5. 사례연구: 매트릭스 곱셈을 위한 데이터 경로 설계

본 논문에 제안된 설계의 방법에 사용된 매트릭스 곱셈(MM), 신호와 이미지 처리의 커널에 사용된 주파수를 (그림 3)에 설명한다.

5.1 도메인 명시

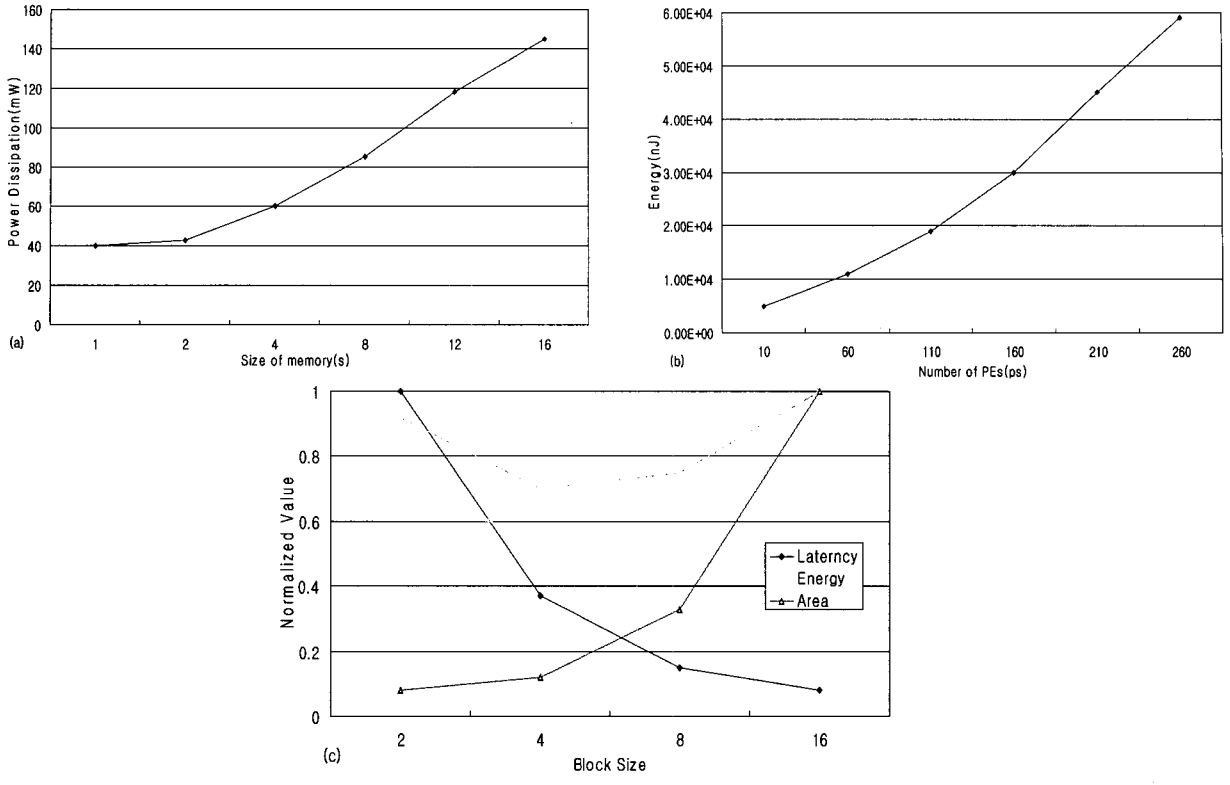
본 논문에 제안된 설계는 (그림 3)과 같이 Processing Elements(PE)의 선형 어레이를 후보 아키텍쳐로 선택한다. 이 아키텍쳐는 다른 아키텍쳐와 비교하여 낮은 I/O 대역폭을 요구하며, 문제 크기에 따라 확장이 가능하다. 이 아키텍쳐를 위한 최적의 알고리즘은 [15]를 참조한다. 선형 어레이의 각 PE는 고정 크기 $s, 1 \le s \le n$ 의 저장소를 가진다. 아키텍쳐는 $n\lceil n/s \rceil$ PE를 사용하여 $O(n^2)$ 시간 동안에 $n x n$ 매트릭스 곱셈을 수행한다. PE들의 수는 n부터 n^2까지 다양하다. 그러나 본 논문에 제안한 설계의 대상 제품은 제한된 양의 로직과 메모리를 합성하는 FPGA이므로, 두 개의 관련된 아기텍쳐-알고리즘을 고려한다.

- n이 작을때, $n x n$ 매트릭스 곱셈에 총 n^2저장소를 가지는 선형 어레이 아키텍쳐
- N이 n의 배수일 때 $n x n$ 크기의 서브-매트릭스에 선형 어레이 구현을 이용한 $N x N$블록 매트릭스 곱셈 알고리즘

5.2 선형 어레이 프로세서의 매트릭스 곱셈을 위한 하이-레벨 모델

(그림 3)(a)는 선형 어레이의 구조인데, 두 개의 콤포넌트 즉, PE와 인접하는 PE들을 연결하는 버스들로 구성된다. 본 논문에 제안한 설계는 하이-레벨 모델링의 목적으로 RModule과 두 개의 인접하는 PE 사이의 버스인 내부연결로 PE를 검증한다. PE는 (그림 3)(b) 정밀도 w와 메모리 크기 s의 MAC을 가진다. PE는 두 개의 전력 상태 *on*과 *off*를 가지는데, PE가 *on* 상태에서는 곱셈기가 *on*이므로 곱셈기가 *off*인 *off* 상태보다 더 많은 에너지를 손실한다. 곱셈기의 전력 상태는 클럭 게이팅에 의해 조절된다. PE는 또한 6개의 레지스터와 w 비트를 가지는 3개의 멀티플렉서를 포함한다. 에너지에 영향을 미치는 파라미터는 정밀도 (w), PE의 개수 (pe), PE 내의 메모리양(s) 그리고 전력 상태(ps)이다.

선형 어레이 아키텍쳐를 위한 매트릭스 곱셈 알고리즘은 [15]에 제안되어 있다. 콤포넌트 특성의 파라미터와 그 범위를 검증하기 위하여 개발되는 알고리즘에 몇 가지의 제한이 있다. 또한 최소의 처리시간을 달성하기 위하여, $n x n$ 매트릭스 곱셈에 필요한 최소의 PE의 수는 n이다[15]. 그러므로 s의 범위는 $1 \le s \le n$로 주어진다. 최소의 I/O 복잡도 $O(n^2)$을 달성하기 위하여 모든 PE를 가로 지르는 총 메모리 양은 n^2이 되어야 한다. 그러므로 총 PE의 수(pe)는 $n\lceil n/s \rceil$이다. $n\lceil n/s \rceil$개의 PE와 PE당 s의 메모리를 사용하는 이 설계의 처리시간(T)는 [15]이다:



(그림 4) 아키텍쳐의 두 패밀리와 연관된 다양한 트레이드오프

$$T=\frac{1}{f}(n^2+2n\lceil n/s\rceil-\lceil n/s\rceil+1) \quad (1)$$

본 논문에 제안한 설계는 $1\le n\le 16$ 범위의 문제를 고려한 것이다. w를 8로 고정한모델 파라미터들은 <표 1>과 같다.

본 논문에 제안한 설계는 f=166MHz에서 동작하는 Virtex-II FPGA를 사용하는 PE를 구현하였고, PE와 버스를 위한 전

〈표 1〉 모델 파라미터

파라미터	범위값
s	$1\le s\le n$
pe	$1\le pe\le n\lceil n/s\rceil$
w	8
ps	on, off

력 함수를 획득하기 위한 시뮬레이션을 수행하였다. PE를 위한 전력함수는:

$$PE.p.ps=\begin{cases}7.01s+31.04mW, & (ps=on)\\ 7.01s+14.04mW, & (ps=off)\end{cases} \quad (2)$$

버스는 39.74mW의 전력 손실의 상수 값을 가진다. 설계 면적은 A_{mult}가 곱셈기의 면적이고 A_{reg8}는 8비트 레지스터의 면적일 때, $A=A_{mult}\,x\,pe+A_{reg8}\,x(3s+6)\,x\,pe+50\,x\,n$로 표현될 수 있다. 이 방정식은 PE(그림 3)의 설계에 기반하여 도출되고, 숫자 50은 각 PE의 멀티플렉서등과 같은 콤포넌트들의 계산에 더해진다.

5.3 트레이드오프 분석

트레이드오프 분석은 하이-레벨 모델과 연관된 다양한 함수들을 만들어 낸다. 예를 들어, (그림 4)는 PE의 전력 손실상의 메모리 모듈 개수의 편차의 영향(s)과 특정 문제의 크기 (16x16)에 대한 시스템 전체 에너지상의 PE의 메모리 모듈 개수의 편차(pe)의 영향을 보인다. 이 곡선에 기반하여 최소 에너지를 소모하는 설계는 문제의 크기가 n일 때 pe=s=n인 설계이다.

큰 매트릭스 곱셈을 위하여 본 논문에 제안한 설계는 잘 알려진 블록 매트릭스 곱셈 기법을 사용한다. 16x16 매트릭스 곱셈에 대하여 에너지, 처리시간, 그리고 면적 관점에서의 성능이 (그림 4)(c)에 보인다. 분석을 위하여 각 범주의 최대값을 고려하여 이 그래프에서 에너지, 처리 시간 그리고 면적을 표준화하였다. 따라서 성능 요구에 따라서 최적의 블록 크기가 선택될 수 있다. 예를 들어 에너지의 관점에서 가장 효율적인 블록의 크기는 4이다. 그런데 만일 본 논문에 제안한 설계가 처리시간 또는 면적에 대하여 최적화한다면, 블록크기 16과 면적 2가 각각 도메인에 의해 명시된 설계 공간내의 최적의 설계이다.

5.4 설계 공간 조사

(그림 4)에서 양 도메인에 대한 최소 에너지 손실을 선택

〈표 2〉 성능 비교

크 기	Xilinx library에 기반한 설계				본 논문에서 제안한 방법에 기반한 설계				성능향상	
n x n	T cycles	A slices	E nJ	E/AT	T cycles	T slices	E nJ	E/AT	E %	T times
6 x 6	480	207	414.8	0.007	49	107.4	158.9	0.0050	62	10
9 x 9	1620	207	1400.0	0.007	100	193.5	590.6	0.0050	58	16
15 x 15	7500	207	6481.5	0.007	256	430.5	3459.9	0.0052	47	29
33 x 33	79860	207	69015.0	0.007	21296	429	28485.1	0.0052	59	4
48 x 48	245760	207	212385.8	0.007	25088	1074	81331.5	0.0050	62	10

영역 기준으로 고려한다. <표 2>는 매트릭스 곱셈을 수행하기 위하여 소모되는 총 에너지를 기반으로 하여 다른 문제의 크기들에 대한 검증된 에너지 효율적 설계들을 보인다. 이 설계를 Xilinx에서 제공되는 매트릭스 곱셈 설계와 비교한다[18].

본 논문에 제안한 설계의 전력 함수들은 최소값에서 잘 동작하는 함수들임에 따라 트레이드오프 곡선상의 가장 에너지 효율적 설계를 시각검증 하였다. 3x3 매트릭스 곱셈에 Xilinx 에서 제공되는 설계와 본 논문에 제안한 설계의 성능을 비교하였는데[18], 모든 설계는 166MHx의 동일 클럭 주파수에서 수행되었다. 더 큰 크기의 매트릭스 곱셈을 위하여 3x3 기본 설계를 사용하는 블록 매트릭스 곱셈을 사용하였다. 각 문제의 크기에 대하여 Xilinx 에서 제공되는 설계와 본 논문에 제안한 방법의 가장 효율적 설계를 비교하였다. <표 2>는 다양한 문제의 크기에 대하여 이 설계들의 에너지, 처리시간 그리고 면적 값을 보인다. 평균적으로 본 논문에 제안한 설계가 시스템 전체의 에너지 손실을 고려하면 Xilinx 설계보다 57% 나은 성능을 보였다. 처리시간 향상은 4x에서 30x까지 다양하였다. 또한 평균적으로 본 논문에 제안한 설계는 평균 전력 밀도(E/AT)면에서 25% 우수하였다. E/AT의 동일하거나 유사한 값은 유사한 설계가 다른 문제 크기에 사용된 것을 가리킨다.

이 장에서 논의된 다양한 설계의 에너지 손실은 도메인에 대한 시스템 전체 에너지 함수를 사용하는 하이-레벨 예측에 기반한 것이다. 본 논문에 제안한 설계의 에너지 예측 방법을 유효화 하기 위하여 다음의 실험을 수행하였다. 특정 설계를 위한 총 에너지 손실을 예측하기 위하여 시스템 전체 에너지 함수를 사용하였다. 이 결과를 Xilinx툴을 이용하여 본 논문에 제안한 설계의 완전한 VHDL 시뮬레이션 결과를 비교하였다. 샘플 시뮬레이션에서 전력 손실을 예측하는 콤포넌트에 들어가는 입력 데이터는 랜덤하게 생성되었고 스위칭 동작(sa)은 25%임이 발견되었다. 본 논문에 제안한 설계는 <표 3>과 같이 다른 문제의 크기들에 대해 다양한 설계들로 수행하였다. 본 논문에 제안한 설계의 에너지 예측은(평균적으로) 로우-레벨 시뮬레이션 툴을 사용하는 예측의 6.4% 이내였다. 최악의 경우 에러는 7.4%였다.

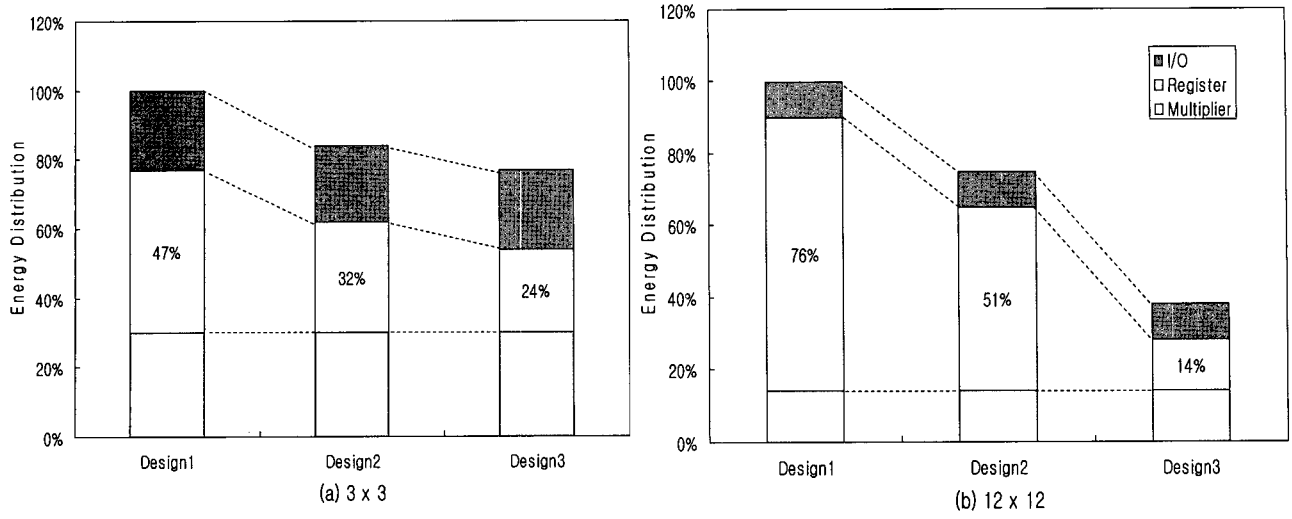
〈표 3〉 본 논문에 제안된 설계의 하이-레벨 예측 방법의 정확성

문제의 크기	n	3	6	8	9	12	16
에너지 (nJ)	Estimated	21.4	159.9	399.6	590.6	1,576.9	4,357.8
	Measured	23.1	169.1	430.2	633.0	1,671.5	4,646.4
	Error	7.4%	5.4%	7.1%	6.7%	5.7%	6.2%

5.5 에너지 최적화에 본 논문에 제안한 설계 방법 적용

본 논문에 제안한 설계 방법은 하이-레벨 모델과 트레이드오프 분석으로 알고리즘을 최적화하며 이 과정은 설계자의 프로세스로 수행된다. 설계자는 완성된 설계의 에너지 효율성을 향상시키기 위하여, 여러 콤포넌트간의 에너지 분배를 이용하는 시나리오를 도해한다. 본 논문에 제안한 설계는 이전 예에서 설명된 설계1을 후보 설계로 한다. 이미 보인 것처럼, 설계1에서 $s=n$은 도메인 내에서 에너지 효율적 설계의 결과이다. 본 논문에 제안한 설계는 어떻게 하이-레벨 모델에서 얻어진 상세 정보에 기반하여 에너지 효율성을 더 향상될 수 있는지를 도해한다. (그림 5)는 설계1에서 $s=n$과 $n=3,12$에 대하여 에너지 손실의 분배를 보인다. 설계1 (그림 3)에 대하여 3x3과 12x12 크기의 문제를 위한 레지스터에서 각각 47%와 76%의 에너지가 손실되었다. 벌크 에너지는 레지스터에서 손실된다. 또한 곱셈기와 I/O포트에서 처리시간을 증가시키지 않고 에너지 손실을 줄이는 것은 어렵다(그러므로 에너지 손실). 본 논문에 제안한 설계는 알고리즘과 아키텍처 레벨에서 매트릭스 곱셈에 사용되는 레지스터의 수를 줄이는 방법을 개발하였다.

본 논문에 제안한 설계 방법은 레지스터의 수를 $2n^2+6n$에서 n^2+4n으로 줄인다[7]. 예를 들어 설계1 (그림 3)(b)의 두 개의 레지스터 (*AS.LR and AS.RR*)는 매트릭스B의 레지스터를 채우고, n 사이클 후 매트릭스A를 채우는 방법으로 한 개의 레지스터로 대체된다. 데이터 이동에 단지 두 개의 레지스터 (*B.T1 and B.T2*)이면 B의 성분들을 저장하는데 충분하다는 것을 조심스럽게 분석할 수 있다. 이 새로운 설계를 (그림 5)의 설계2로 참조한다. (그림 5)에 보이는 바와 같이,



(그림 5) 최적화중 모듈간의 에너지 분배의 변화

설계2에서 전체적인 에너지 손실은 문제의 크기 3x3과 12x12 에서 각각 16%와 25%가 줄어든다. 레지스터에서 에너지 손실량은 47%와 76%에서 32%와 52%로 각각 줄어든다.

아키텍처와 알고리즘 레벨에서 최적화를 통한 에너지의 감소 이외에도 구현 레벨에서 추가적인 감소가 가능하다.

〈표 4〉 설계3과 Xilinx 설계간의 성능비교

메트릭	Xilinx	설계3	비율(Ratio)
에너지(nJ)	25	17	68%
지연시간(cycles)	45	16	35%
면적(slices)	180	415	2.3x
E/AT	0.67	0.43	64%

예를 들어, 중간 결과를 저장하는 레지스터 (그림 3)의 $C[j]$)는 단위 저장소당 전력을 줄이기 위하여 CLB 기반의 SRAM에 의해 대체될 수 있다. 그런데 대상 FPGA에서 SRAM당 최소 워드 수[17]는 16이어야만 한다. 본 논문에 제안한 설계는 이 설계를 설계3이라 한다.

설계3에서 원본 설계(설계1)와 비교하였을 때, 에너지 손실은 23%와 62%가 문제의 크기 3x3과 12x12 에서 각각 줄어들었다. 이 설계는 총 에너지의 24%와 14%만이 레지스터에서 손실되었다.

Xilinx 설계와 에너지 효율을 비교하기 위하여 3x3 크기에서 설계3을 로우-레벨 합성에 선택하였다. 로우-레벨 합성은 Synopsys FPGA Express와 Xilinx ISE 4.1i 설계 환경의 Xilinx XST를 사용하여 수행되었다. Virtex-II XC2V1500의 place-and-route 파일(.ncd 파일)과 Mentor ModelSim5.5e가 시뮬레이션과 결과생성에 사용되었다(.vcd 파일). 에너지 손실을 예측하기 위하여 Xilinx XPower 툴에 이 두 파일이 제공되었다. 그 결과는 <표 4>의 최적화된 Xilinx 참조 설계[18]에 대응하여 비교되었다. 설계3은 Xilinx 설계와 비교하여 32%, 64% 그리고 35%가 에너지, 처리시간 그리고 평균 전력 밀도의 면에서 각각 우수하였다. 그런데 면적 면에서 본 논문에 제안한 설계는 Xilinx 설계의 2.3x 이다. 설계의 상세, 에너지 함수 그리고 추가적인 결과는 [7]에서 볼 수 있다.

6. 결 론

본 논문은 FPGA상에서 특정 어플리케이션의 에너지 효율 데이터 경로를 위한 모델 기반 설계 방법을 제안한다. 이 방법은 도메인 특성 하이-레벨 모델링, 설계 공간 조사, 하이-레벨 에너지 예측 그리고 로우-레벨 시뮬레이션을 통합한다. 에너지 효율 데이터 경로 설계를 그림으로 설명하기 위하여 매트릭스 곱셈이 후보 어플리케이션으로 선택되었다. 본 논문에 제안한 설계는 매트릭스 곱셈을 구현하는 두 개의 도메인을 고려하였다. 또한 매트릭스 곱셈을 구현하는 설계의 에너지 효율성을 향상시키기 위한 알고리즘 레벨 최적화 방법을 제안하였다.

양 도메인에서, 본 논문에 제안한 설계는 다양한 크기의 매트릭스 곱셈을 위한 에너지, 처리시간 그리고 면적간의 트레이드오프를 보였다. 본 논문에 제안한 설계의 방법에 기반한 설계는 Xilinx에서 제공되는 매트릭스 곱셈 설계와 비교하였을 때, 평균에너지 손실에서 57%, 처리시간에서 14x 향상을 보였다. 아키텍처 패밀리에 기반한 하이-레벨 모델링과 대응하는 알고리즘(도메인)은 시스템 전체 에너지 손실의 정확도가 높은 예측의 결과를 가져왔다. 본 논문에 제안한 설계의 에너지 예측은 로우-레벨 시뮬레이션을 사용하여 7.4% 이내였다.

참 고 문 헌

[1] A. Agrawal, A. Bakshi, J. Davis, B. Eames, A. Ledeczi, S. Mohanty, V. Mathur, S. Neema, G. Nordstrom, V. Prasanna, C. Raghavendra, M. Singh, "MILAN: A Model Based

Integrated Simulation for Design of Embedded Systems," Language Compilers and Tools for Embedded Systems, 2001.

[2] K. Bondalpati and V. K. Prasanna, "Loop Pipelining and Optimization for Reconfigurable Architectures," Reconfigurable Architectures Workshop(RAW), May, 2000.

[3] S. Choi, S. Mohanty, J. Jang, and V. K. Prasanna, "Domain-Specific Modeling for Rapid System-Level Energy Estimation of Reconfigurable Architectures," Intl. Conference on Engineering of Reconfigurable Systems and Algorithms, 2002.

[4] A. Dandalis, and V. K. Prasanna, "Signal Processing using Reconfigurable System-on-Chip Platforms," International Conference on Engineering of Reconfigurable Systems and Algorithms, June, 2001.

[5] J. A. B. Fortes, K. S. Fu, and B. Wah, ≪Systematic Approaches to the Design of Algorithmically Specified Systolic Arrays≫, International Conference on Acoustics, Signal and Speech Processing, 1985.

[6] A. Garcia, W. Burleson, and J. L. Danger, "Power Modeling in FPGAs", 9th International Conference on Field Programmable Logic and Applications, 1999.

*[7] J. Jang, S. Choi, and V. K. Prasanna, "Energy-Efficient Matrix Multiplication on FPGAs," FPL 2002, Lecture Notes in Computer Science, pp.534-544.

[8] D. Kumar and K. Parhi, "Performance Trade-off of DCT architectures in Xilinx FPGAs," The 33rd Asilomar Conference on Signals, and Computers, 1999.

[9] V. Kumar, A. Grama, A. Gupta, and G. Karypis, "Introduction to Parallel Computing: Design and Analysis of Algorithms," Benjamin Cummings, November, 1993.

[10] S. Lei and K. Yao, "Efficient Systolic Array Implementations of Digital Filtering," IEEE International Symposium on Circuits and Systems, 1989.

[11] W. Luk, N. Shirazi, and P.Y.K. Cheung, "Modeling and Optimizing Run-time Reconfigurable Systems," IEEE Symposium on FPGAs for Custom Computing Machines, 1996.

[12] Model-based Integrated Simulation, http://milan.usc.edu.

[13] S. Mohanty, V.K. Prasanna, S. Neema, and J. Davis, "Rapid Design Space Exploration of Heterogeneous Embedded Systems using Symbolic Serach and Multi-Granular Simulation," Language Compilers and Tools for Embedded Systems, 2002.

[14] M. Nemani and F. N. Najm, "High-level Area and Power Estimation for VLSI Circuits," IEEE/ACM International Conference on Computer-Aided Design, 1997.

[15] V. K. Prasanna Kumar and Y. Tsai, "On Synthesizing Optimal Family of Linear Systolic Arrays for Matrix Multiplication," IEEE Transactions on Computers, Vol.40, No.6, 1991.

[16] A. Raghunathan, N.K. Jha, and S. Dey, "High-level Power Analysis and Optimization," Kluwer Academic Publishers, 1998.

[17] F. G. Wolff, M. J. Knieser, D. F. "Low Power FPGA Design Methodology," National Aerospace and Electronics Conference, 2000.

[18] Xilinx Application Note: Vertex-II/Vertex-II Pro Series and Xilinx ISE 4.1 Design Environment, http://www.xilinx.com.

[19] Sumit Mohanty and Viktor K. Prasanna, "Energy Efficient Application Design using FPGAs," FPGA and Programmable Logic Journal, October, 2004.

[20] Jingzhao Ou and Viktor K. Prasanna, "A Methodology for Energy Efficient Application Synthesis Using Platform FPGAs," International Conference on Engineering of Reconfigurable Systems and Algorithms(ERSA), June, 2004.

장 주 욱

e-mail : jjang@sogang.ac.kr
1983년 서울대학교 전자공학(학사)
1985년 한국과학기술원(석사)
1993년 미국 University of Southern California 컴퓨터공학(박사)
1985년~1994년 삼성전자 컴퓨터개발실
1995년~현재 서강대학교 전자공학과 교수
관심분야 : 병렬처리, IT SoC, 인터넷 프로토콜

이 미 숙

e-mail : rosalee@eeca1.sogang.ac.kr
1984년 한양대학교 가정학과 졸업(학사)
1986년~1990년 SGS-Thompson Korea Ltd.
1995년~2003년 Ericsson Korea Ltd.(IT Manager)
1999년~2002년 서강대학교 정보통신대학원 졸업(공학석사)
2004년~현재 서강대학교 전자공학과 박사과정
관심분야 : IT SoC, 인터넷 프로토콜

Sumit Mohanty

e-mail : smohanty@usc.edu

현 재 미국 University of Southern California 전자공학 박사과정

최 선 일

e-mail : seonilch@usc.edu

현 재 미국 University of Southern California 전자공학 박사과정

Viktor K. Prasanna

e-mail : prasanna@usc.edu

인도 Bangalore University 전자공학(학사)

인도 Indian Institute of Science 전자공학(석사)

미국 Pennsylvania State University 컴퓨터학(박사)

현 재 USC(Univ. of Southern California) 컴퓨터학과 교수

관심분야 : 연산처리, 병렬 분산 시스템, 네트워크 컴퓨팅, 임베디드 시스템