

논문 2004-41SD-10-13

Deep Submicron 공정의 멀티미디어 SoC를 위한 저전력 움직임 추정기 아키텍처

(Low-Power Motion Estimator Architecture for Deep Sub-Micron Multimedia SoC)

연 규 성*, 전 치 훈*, 황 태 진*, 이 성 수*, 위 재 경*

(Gyu Sung Yeon, Chi Hoon Jun, Tae Jin Hwang, Seongssoo Lee, and Jae-Kyung Wee)

요 약

본 논문에서는 0.13 μm 이하의 deep sub-micron 공정처럼 누설 전류가 심한 공정을 이용하여 멀티미디어 SoC를 설계할 때, 가장 전력 소모가 높은 움직임 추정 기법의 전력 소모를 줄이기 위한 저전력 움직임 추정기의 아키텍처를 제안하였다. 제안하는 아키텍처는 기존의 동적 전력 소모만을 고려한 구조와는 달리 정적 전력 소모까지 고려하여 누설 전류가 심한 공정에 적합한 구조로, 효율적인 전력 관리가 필수적인 동영상 전화기 등의 각종 휴대용 정보기기 단말기에 적합한 형태이다. 제안하는 아키텍처는 하드웨어 구현이 용이한 전역 탐색 기법 (full search)을 기본으로 하며 동적 전력 소모를 줄이기 위하여 조기 은퇴 (early break-off) 기법을 도입하였다. 또한 정적 전력 소모를 줄이기 위하여 전원선 잡음을 고려한 메가블록 전원 차단 기법을 사용하였다. 제안된 아키텍처를 멀티미디어 SoC에 적용하였을 때의 효율성을 검증하기 위해 시스템 수준의 제어 흐름과 저전력 제어 기법을 개발하였으며, 이를 바탕으로 시스템 수준에서의 소모 전력을 계산하였다. 모의실험 결과 0.13 μm 공정에서 전력 소모가 60% 정도로 감소함을 확인할 수 있었다. 선폭의 감소와 칩 내부 발열량의 증가로 인한 누설 전류의 증가를 고려할 때, 기존의 동적 전력 소모만을 고려한 구조는 전력 감소 효율이 점점 나빠짐에 반하여 제안하는 움직임 추정기 아키텍처는 안정적인 전력 감소 효율을 보여주었다.

Abstract

This paper propose a motion estimator architecture to reduce the power consumption of the most-power-consuming motion estimation method when designing multimedia SoC with deep submicron technologies below 0.13 μm . The proposed architecture considers both dynamic and static power consumption so that it is suitable for large leakage process technologies, while conventional architectures consider only dynamic power consumption. Consequently, it is suitable for mobile information terminals such as mobile videophone where efficient power management is essential. It exploits full search method for simple hardware implementation. It also exploits early break-off method to reduce dynamic power consumption. To reduce static power consumption, megablock shutdown method considering power line noise is also employed. To evaluate the proposed architecture when applied multimedia SoC, system-level control flow and low-power control algorithm are developed and the power consumption was calculated based on them. From the simulation results, power consumption was reduced to about 60%. Considering the line width reduction and increased leakage current due to heat dissipation in chip core, the proposed architecture shows steady power reduction while it goes worse in conventional architectures.

Keywords : deep sub-micron, leakage current, multimedia SoC, low-power, motion estimation

I. 서 론

* 정회원, 숭실대학교 공과대학 정보통신공학부
(School of Electronic Engineering, Soongsil University)

※ 이 논문은 2004년도 숭실대학교 교내연구비에 의하여 연구되었음.

접수일자: 2004년6월1일, 수정완료일 : 2004년9월22일

최근 들어 공정 기술의 발달에 따라 소자의 크기가 점점 작아지고 deep sub-micron 공정이 등장하였다. 이에 따라 예전에는 하나의 칩으로 구현하기 어려웠던 대

규모 시스템을 하나의 칩으로 집적하는 SoC (System-on-Chip) 기술이 제안 되었으며, 높은 연산량, 다양한 기능, 칩의 소형화가 절대적으로 요구되는 휴대용 멀티미디어 단말기용 SoC 설계 분야를 중심으로 활발한 연구가 진행 중이다. 이러한 멀티미디어 SoC는 그 특성상 전력의 효율적인 관리와 사용이 가장 큰 문제로 대두되고 있다. 저전력 SoC를 구현하기 위한 기존 연구는 대부분 RTL 수준, Gate 수준 및 회로 수준과 같은 하위 수준에 집중되어 있다. 최근들어 시스템 수준의 저전력 SoC 연구가 활발하게 연구되기 시작했지만, 이들 연구 대부분은 게이트의 스윙칭 횟수를 줄이거나 동적으로 전압 또는 주파수를 제어하여 동적 전력 (dynamic power)을 줄이는 기법이 주류를 이루고 있다^[1,8]. 그러나 deep sub-micron 공정이 점차 일반화되어감에 따라 누설 전류가 매우 커지기 때문에^[2] 이로 인한 정적 전력 (static power)은 저전력 SoC 설계에서도 더 이상 무시할 수 없는 요소가 되고 있다. 최근 들어 문턱 전압 (threshold voltage)을 동적으로 제어하여 누설 전류로 인한 정적 전력을 줄이는 연구^[2]가 제안되었지만, 하드웨어의 복잡도가 높기 때문에 실제 SoC 구현에서는 그다지 많이 채택되지 않고 있다.

멀티미디어 SoC는 일반적으로 비디오 부호화, 비디오 복호화, 오디오 부호화, 오디오 복호화 등 다양한 작업을 수행하며, 각각의 작업은 움직임 추정 (motion estimation), 이산 여현 변환 (discrete cosine transform), 양자화 (quantization) 등의 다양한 연산으로 구성된다. 이 중에서 움직임 추정은 멀티미디어 SoC에서 가장 큰 연산량을 차지하며, 멀티미디어 압축 및 복원의 국제 표준인 MPEG-4^[3]에서는 전체 연산량의 60% 이상이 움직임 추정에 사용되고 있다. 따라서 저전력 멀티미디어 SoC 구현을 위해서는 동적 전력뿐만 아니라 정적 전력도 효과적으로 줄이고, 특히 움직임 추정의 전력 소모를 효과적으로 줄이는 기법의 개발이 절실하게 요구되고 있다.

본 논문에서는 휴대용 멀티미디어 단말기용 SoC에서 동적 전력뿐만 아니라 정적 전력까지 효과적으로 줄인 움직임 추정 블록의 저전력 구조 및 시스템 수준의 제어 기법을 제안한다. 제안하는 움직임 추정 블록은 성능이 우수한 전역 탐색 기법 (full search)을 채택하였으며 $\pm 16 \times \pm 16$ 또는 $\pm 8 \times \pm 8$ 크기의 탐색 범위를 가지며 입력 영상의 크기는 CIF (352×288 화소, 30 프레임/초) 또는 QCIF (176×144 화소, 15 프레임/초)를 기본으로 한다. 동적 전력을 줄이기 위하여 조기 은퇴 개념을 적용

하였고 정적 전력을 줄이기 위하여 블록 단위 셋다운 알고리즘^[4,5]을 사용하였다. 기존의 블록 단위 셋다운은 셋다운 과정에서 큰 전력선 잡음 (power line noise)이 발생하기 때문에 잡음에 민감한 deep sub-micron 공정에는 적합하지 않기 때문에 파이프라인 방식 블록 단위 셋다운^[5] 구조를 적용 하였다.

본 논문의 구성은 다음과 같다. II 장에서는 deep sub-micron 공정의 누설 전류 특성을, III 장에서는 기존의 시스템 수준의 저전력 알고리즘에 대해서 알아보고 IV 장에서는 제안하는 저전력 움직임 추정 블록의 구조와 시스템 수준에서의 저전력화 방법에 대해서 논한다. V 장에서는 제안하는 저전력 움직임 추정 블록의 전력 소모를 모의실험하고 그 결과를 분석하며 VI 장에서는 결론을 맺는다.

II. Deep Sub-Micron 공정에서의 누설 전류

일반적으로 deep sub-micron 공정에서의 누설 전류의 종류 및 원인^[6]은 1) threshold voltage의 감소로 인한 subthreshold leakage (I_{sub})의 exponential한 증가, 2) gate oxide 두께의 감소에 의한 gate edge-direct-tunneling leakage (I_{EDT})와 3) gate-induced-drain-leakage (I_{GIDL})의 exponential한 증가, 4) lightly-doped-drain 또는 pocket-doping 농도에 의한 bulk band-to-band-tunneling leakage (I_{B-TBT})의 exponential한 증가를 들 수 있다. 이러한 누설 전류의 극심한 증가는 deep sub-micron 공정에서의 회로 설계에서 매우 중요한 고려 사항으로 떠오르고 있다.

표 1은 0.13 μm 공정에서 누설 전류 특성과 만 게이트 당의 누설 전류를 나타낸 것이다. 이때, LV (Low Power and High Performance), LP (Low Power), ULP (Ultra Low Power)는 각각 사용 목적에 따라 세분되어 개발되는 공정 종류를 가리킨다. 표 1에서 알 수 있듯이 0.13 μm LV 공정에서 누설 전류의 최대치는 만 게이트 당 일반적으로 260 μA , 최대로 780 μA 에 달하므로 최대 100만 게이트에 달하는 MPEG-4 멀티미디어 SoC에서는 무시할 수 없는 큰 값을 가지게 된다.

또한 0.13 μm 공정에서 전체 전력 소모에 대하여 누설 전류에 의한 전력 소모의 비는 약 10% 정도를 차지하며 향후 0.10 μm , 0.07 μm , 0.05 μm 공정으로 내려 갈수록 각각 20%, 40%, 50% 정도로 큰 비중을 차지하게 될 것이다. 더욱이 표 1의 누설 전류는 상온인 25°C를 기준으로 하고 있으나, 전력 소모가 높은 멀티미디어

표 1. 0.13μm 공정에서의 누설 전류
Table 1. Leakage current in 0.13μm technology.

	process		
	LV	LP	ULP
leakage-nom n/p(pA/μm)@1.1V _{DD}	1E4	3.3	0.12
만 게이트 당 누설 전류의 일반치	1E4	1.8	0.16
leakage-max n/p(pA/μm)@1.1V _{DD}	3E4	10	0.5
만 게이트 당 누설 전류의 최대치	3E4	10	0.5
	780μA	260nA	13nA

SoC에서는 중심부의 온도가 상온보다 훨씬 더 높아 지기 때문에 누설 전류의 의한 정적 전력도 더 커질 것으로 생각된다. 반면 deep sub-micron 공정으로 갈수록 공급 전압 (supply voltage)이 감소하기 때문에 스위칭에 의한 동적 전력은 공급 전압의 제곱에 비례하여 감소한다. 따라서 가까운 미래에는 동적 전력을 줄이는 저전력 기법보다 오히려 정적 전력을 줄이는 저전력 기법이 더욱 중요하게 될 것으로 생각된다.

III. 시스템 수준의 저전력 기법

SoC의 전력 소모는 크게 1) 회로가 어떠한 연산을 수행할 때 발생하는 동적 전력 소모와 2) 회로가 대기 상태에 있을 때에도 변함없이 발생하는 누설 전류에 의한 정적 전력 소모가 있다. 우선 동적 전력 소모를 줄이기 위한 대표적인 시스템 수준 저전력 기법에는 클록 게이팅^[7], 동적 전압 제어^[1,8], 동적 주파수 제어^[9] 등이 있다. 이러한 기법들은 동적 전력을 나타내는 수식 (1)에서 비롯되었다. 여기서 α 는 각 노드의 스위칭 수, C 는 로드 커패시턴스 값, f 는 동작 주파수, V_{DD} 는 공급 전압을 나타낸다.

$$P_{dynamic} = \alpha \times C \times f \times V_{DD}^2 \quad (1)$$

클록 게이팅의 경우 수식 (1)의 α 값을 줄이기 위하여 대기 상태에 있거나 연산을 수행하지 않을 때는 AND 게이트와 같은 간단한 회로를 추가 하여 클록의 입력을 선택적으로 제어한다. 동적 전압 제어나 동적 주파수 제어에서는 각 블록이 최대 연산 수행 시보다 주어진 연산을 일찍 종료하였을 때 최대 연산 수행 시간 (worst-case execution time)과 실제 연산 수행 시간 (execution time)과의 차이를 연산 여분 시간 (slack tim)이라 하며, 이러한 시간이 생기는 경우 동작 주파수 f 나 V_{DD} (혹은 둘 다)를 낮추어 연산 속도를 느리게 만들어서 실제 연산 시간이 최대 연산 수행 시간에 최대

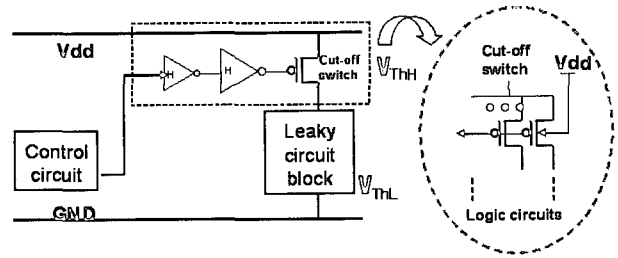


그림 1. 블록단위 섀다운 구조
Fig. 1. Architecture of block-wise shutdown.

한 가깝게 연산을 수행함으로써 전력 소모를 줄인다. 이들 기법을 비교해보면 이론적으로 클록 게이팅의 경우 약 1~2 배 정도 동적 전력 소모를 낮출 수 있고, 동적 주파수 제어는 동작 주파수 f 에 비례하여, 동적 전압 제어는 공급 전압 V_{DD} 의 제곱에 비례하여 동적 전력 소모를 낮출 수 있다. 그러나 실질적인 측면에서 살펴본다면 클록 게이팅의 경우 간단한 회로의 추가로 구현 할 수 있다는 용이성이 있으나 그 효율이 비교적 적고, 동적 주파수 제어 또는 동적 전압 제어는 연산의 모니터링과 스케줄링에 의한 제어 방식이 구현하기 복잡하고 주파수나 전압을 바꾸는 과정에서 잡음이 발생하여 회로에 악영향을 끼칠 가능성이 상존한다.

정적 전력 소모를 줄이기 위한 기본 개념은 수식 (2)를 바탕으로 누설 전류 $I_{leakage}$ 를 줄이는 것이다. 대표적인 알고리즘에는 그림 1과 같은 블록 단위 섀다운^[4,5]이 있다.

$$P_{static} = I_{leakage} \times V_{DD} \quad (2)$$

블록 단위 섀다운 알고리즘은 하나의 시스템에서 모든 블록이 동시에 사용되지 않는다는 전제 하에 대기 상태에 있는 블록의 누설 전류를 차단하기 위하여 전력 공급을 차단하는 방법이며 전력 차단을 위하여 Cut-off 스위치를 사용하며 스위치에 의한 누설을 차단하기 위하여 문턱 전압이 높은 스위치를 사용한다. 섀다운 시간이 길면 길수록 정적 전력 소모를 많이 줄일 수 있으며 너무 짧은 시간 동안에 섀다운 / wake-up을 하는 경우 제어 동작에 의한 전력 소모로 인하여 오히려 전력 소모를 증가시킬 수 있다. 그러나 실제 구현에 있어서는 전력선 잡음의 발생과 그로 인한 주변 블록의 오동작을 야기시킬 수 있다는 단점이 있다.

동적 전력 소모를 줄이기 위해서 가장 각광받고 있는 DVS^[1,8] (dynamic voltage scaling)와 DVTS^[2,10] (dynamic V_{Th} scaling), 정적 전력 소모를 줄이기 위한 블록 단위 섀다운^[4,5]을 하나의 SoC에 함께 적용했을 때의 개

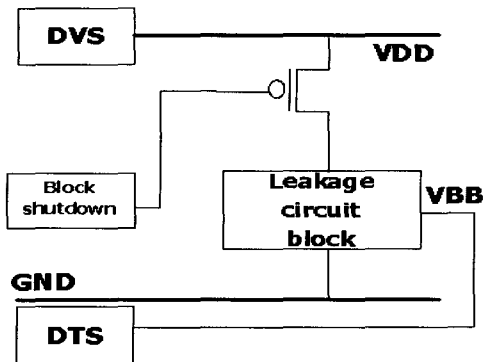


그림 2. DVS, DTS 및 블록 단위 섀다운
Fig. 2. DVS, DTS and block-wise shutdown.

념도는 그림 2와 같다. DVS는 V_{DD} 를 변화시키고 DTS는 V_{BB} (bulk bias voltage)를 변화시켜 전력 소모를 감소시키는 기법이다. 현재까지는 DVS와 DTS가 높은 전력 감소 효율 때문에 각광을 받고 있지만, 0.13 μm 이하의 deep sub-micron 공정에서는 많은 문제점이 발견되고 있다.

deep sub-micron 공정으로 갈수록 공급 전압은 낮아지고 누설 전류는 심해지므로, 이러한 환경에서 DVS를 구현하기 위하여 공급 전압을 낮추면 그만큼 문턱 전압도 낮추어야 한다. 예를 들어 공급 전압이 1.0 V이고 문턱 전압이 0.3 V 라면 DVS를 위하여 0.5 V로 공급 전압을 낮추면 문턱 전압은 0 V에 가깝게 낮아져야 한다. 그러나 이러한 문턱 전압의 변화는 누설 전류의 양을 기하급수적으로 증가 시킨다. 또한 DTS의 경우 고속연산을 위해 문턱 전압을 낮추고 저속 동작을 할 때는 누설 전류를 억제하기 위하여 문턱 전압을 높인다. 앞의 예에서 문턱 전압을 0.3 V를 0 V에 가깝게 낮추려면 V_{BB} 전압이 0 V에서 대략 0.3 V로 증가해야 할 것이다. 그러나 V_{BB} 를 제어하기 위해서는 일반적으로 charge pump를 사용하기 때문에 빠른 동작이 어려우며 또한 bulk capacitance가 크기 때문에 charge / recharge로 인한 전력소모가 커서 전력 효율이 저하된다는 문제점이 있다.

반면에 누설 전류를 줄이는 블록 단위 섀다운 기법은 문턱 전압이 높은 cut-off switch와 shutdown / wake-up을 위한 간단한 제어 블록만이 있으면 되므로 비교적 문제점이 적으며, 실제적인 SoC 구현이 용이하고 누설 전류를 거의 완벽하게 차단할 수 있다는 장점이 있다. 특히 deep sub-micron 공정으로 갈수록 누설 전류의 비중이 급격하게 커지고, 가까운 미래에 상용화 될 것으로 여겨지는 수십 nm 급 공정에서는 누설 전류의 억제 없이는 저전력화를 기대하기 어려울 정도이기 때

```
sadmin = MAXINT; mvx = 0; mvy = 0;
for (u = -sx; u<=sx; u++)
  for (v=-sy; v<=sy; v++)
  {
    sad = 0;
    for (i=0; i<16; i++)
    {
      for (j=0; j<16; j++)
        sad += |R(i,j) - S(i+u,j+v)|;
      if (sad >= sadmin)
        break; ← 조기 은퇴 기법
    }
    if (sad < sadmin)
    {
      sadmin = sad; mvx = u; mvy = v;
    }
  }
}
```

그림 3. 조기 은퇴 기법
Fig. 3. Early Break-off.

문에 블록 단위 섀다운 기법의 중요성이 매우 커질 것이다.

IV. 제안하는 저전력 움직임 추정기 아키텍처

1. 제안하는 움직임 추정기의 구조

전역 탐색 기법^[11] (full Search: FS)은 블록 매칭 방식의 움직임 추정 알고리즘 중에 가장 우수한 성능을 보여주고 있으며, 규칙적인 데이터의 흐름으로 인하여 SoC 설계가 간단하다는 장점을 가지고 있다. 그러나 연산량이 너무 많아서 하드웨어 구현시 면적이 너무 커지기 때문에 멀티미디어 SoC에서는 대부분 전역 탐색 기법 대신에 고속 탐색 기법 (fast algorithm)이 사용되어 왔다. 그러나 공정 기술의 발달에 따라 면적에 대한 제약 조건이 완화되고 휴대용 멀티미디어 단말기와 같이 가시 화면의 크기가 작아짐에 따라 FS 또는 FS에 준하는 알고리즘의 SoC 구현이 가능하게 되었다.

전역 탐색 기법을 비롯한 대부분의 움직임 추정 기법은 기준 블록 (reference block)과 가장 닮은 블록을 찾는 것을 목표로 하며, '가장 닮았다'라는 판단을 하기 위한 기준 척도 (matching criterion)로 SAD (sum of absolute difference)를 계산하여 SAD가 가장 작은 블록을 '가장 닮은' 블록으로 간주한다. 움직임 추정기가 탐색 범위 내에서 SAD 연산을 수행 할 때 현재 맞춰보고 있는 블록에서 일부분만 맞춰본 SAD 중간값이 이전에 탐색을 마친 탐색 범위의 SAD 최소값보다 클 경우에는 현재 맞춰보고 있는 블록이 '가장 닮은' 블록일 가능성이 없으므로 그림 3과 같이 더 이상의 SAD 계산을 하지 않고 다음 탐색 범위로 넘어가면 된다. 이러한 방법을 조기 은퇴 기법이라 한다. 이때, 예정된 계산의 일부분을 수행하지 않으므로 움직임 추정기가 동작하지

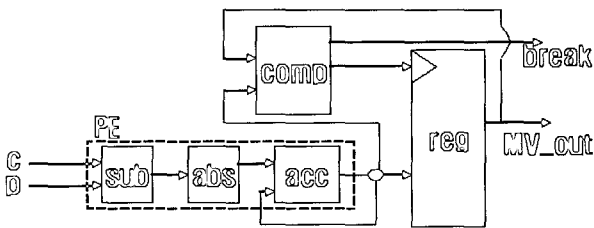


그림 4. SAD 연산과 조기 은퇴를 위한 연산기
Fig. 4. Processing element for SAD and early break-off.

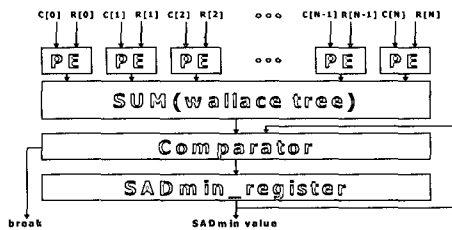


그림 5. 병렬 연산을 위한 아키텍처
Fig. 5. Parallel processing architecture.

않고 쉬는 구간이 발생하며, 이때 움직임 추정기를 셧다운시킴으로서 전력 소모를 줄일 수 있다. 본 논문에서는 이러한 조기 은퇴 기법을 채택하였으며, 이에 따른 단위 연산기 (processing element: PE) 구조는 그림 4와 같다.

입력 영상의 크기와 탐색 범위에 따라 연산량이 변하므로 주어진 입력 영상의 크기 및 프레임 속도 (frame rate)에 따라 필요한 연산 능력을 확보하기 위해서는 적당한 수만큼 PE를 추가하여 병렬처리를 해야 한다. 병렬 처리 연산을 위한 구조는 그림 5와 같다. PE가 16개인 경우 MB(macroblock)의 1행씩을 병렬 연산 할 수 있는 구조로 만들 수 있다. CIF 크기의 입력 영상과 $\pm 16 \times \pm 16$ 의 탐색 범위를 가질 경우 32 개의 PE를 사용하여 최대 15 frame/s에 적합한 형태를 만들 수 있으며 $\pm 8 \times \pm 8$ 의 탐색 범위를 가질 경우 16 개의 PE를 사용하여 최대 30 frame/s에 적합한 형태를 만들 수 있다. QCIF 크기의 입력영상 및 $\pm 16 \times \pm 16$ 탐색 범위일 경우 16 개의 PE를 사용하여 최대 30 frame/s 에 적합한 형태를 만들 수 있으며 $\pm 8 \times \pm 8$ 탐색 범위를 가질 경우 4 개의 PE 만으로 최대 30 frame/s에 적합한 형태를 만들 수 있다.

FS의 경우, 조기 은퇴가 전혀 일어나지 않는 경우, 즉 모든 MB에서 모든 연산을 빼놓지 않고 수행하는 경우가 최대 연산 수행 시간이 되고, 실제로 당작할 때 조기 은퇴가 일어나서 수행 시간이 단축되었을 때가 실제 연산 수행 시간이 되며, 이 연산 시간의 차가 컸다운을

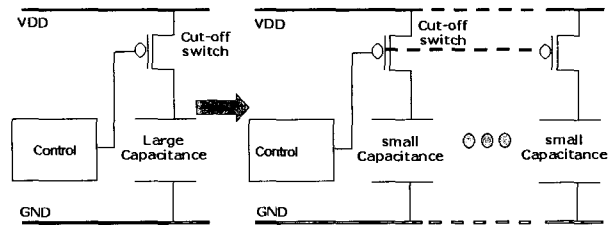


그림 6. 파이프라인 블록 단위 셧다운 아키텍처
Fig. 6. Pipelined block-wise shutdown architecture.

위한 여분의 시간이 된다. 조기 은퇴를 이용한 FS 알고리즘의 경우 연산량이 60~70%정도 줄어들고 그만큼 여분의 시간이 생기게 된다는 것을 알 수 있다. 연산이 끝나고 남은 시간 동안 대기 상태의 정적 전력 소모를 차단하기 위하여 전력선 잡음을 고려한 블록 단위 셧다운 기법^[5]을 적용하였고 제안하는 구조는 그림 6와 같다.

기존의 블록 단위 셧다운 방식은 단순히 대기 상태에 있거나 사용하지 않는 블록을 끄고 다시 사용하고자 할 때 블록을 켜는 형태이다. 그러나 블록의 크기를 고려하지 않은 이러한 기법은 블록의 크기가 클 경우 블록을 급속하게 끄거나 켜는 동작을 할 때에 주변의 블록들에게 전력선 잡음을 발생 시켜 오동작을 야기시킬 수 있다^[5]. 꺼져있는 블록의 크기가 클 경우, 이 블록은 하나의 큰 캐패시터와 같은 작용을 하며, 빠른 시간 내에 다시 동작을 시키기 위하여 wake-up 동작을 할 때 큰 캐패시턴스를 채우기 위하여 순간적으로 많은 양의 전류를 필요로 한다. 이때 순간적으로 흐르는 큰 전류는 주변 블록의 전력 공급에 잡음 효과를 일으킬 수 있다. 본 논문에서는 이러한 문제점을 제거하기 위하여 해당 블록을 여러 단계의 파이프라인 구조로 나누어 차례차례 wake-up 동작이 이루어지는 파이프라인 블록 단위 셧다운 기법^[5]을 적용하였다. 이때, wake-up 되는 블로가 여러 개로 나뉘어지기 때문에 캐패시턴스도 이에 상응하여 작아져서 전력선 잡음이 오동작을 일으키지 않도록 할 수 있었다. HDL로 움직임 추정기를 설계해 본 결과, 16개의 PE를 사용한 구조의 경우 약 2만 제이트 정도의 면적을 차지하였다.

2. 시스템 수준의 제어 방법

기존에 MPEG-4 멀티미디어 SoC를 설계할 때에 각 단위 연산의 흐름은 그림 7(a)와 같이 움직임 추정 (motion estimation: ME)가 끝나고 이산 역변환 (discrete cosine transform: DCT), 재배열 (reordering), 양자화 (quantization: Q), 가변길이 부호화 (variable

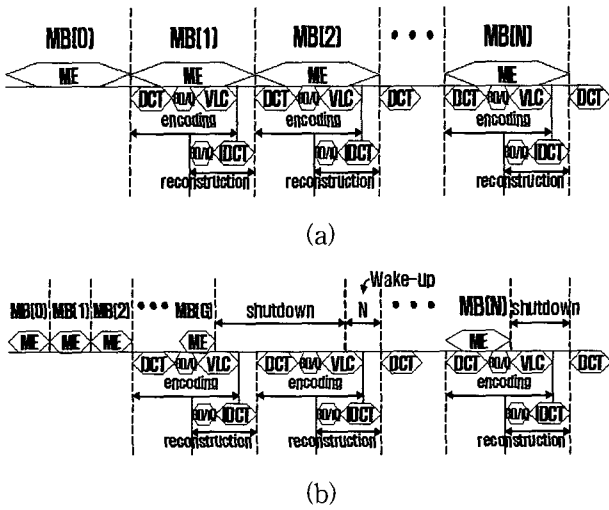


그림 7. 단위 연산 사이의 흐름도. (a) 기존 멀티미디어 SoC, (b) 제안하는 멀티미디어 SoC
 Fig. 7. Flow chart of unit processes.(a) Conventional multimedia SoC. and (b) Proposed multimedia SoC.

length coding: VLC)의 순으로 부호화를 수행하며 다음 프레임의 ME를 위하여 reconstruction frame을 만든다. 본 논문에서 제안하는 아키텍처의 연산 흐름은 그림 7(b)와 같이 ME가 조기에 연산을 끝마쳐 여러 개의 MB를 미리 연산하고 남는 시간 동안 섯다운을 하게 되며 전력선 잡음을 고려하여 파이프라인 방식의 wake-up 동작을 하기 위하여 최대 연산 시간으로부터 N cycle 이전에 wake-up신호를 발생한다. 이때, N은 PE의 개수를 뜻한다.

이렇게 설계된 움직임 추정기를 전체 멀티미디어 SoC에 적용하여 저전력을 구현하는 방법은 다음과 같다. main controller와 power management unit(PMU)가 있고 제안하는 움직임 추정 블록이 on-chip-bus 형태로 그림 8과 같이 연결되어 있다. 우선 제안하는 구조의 움직임 추정 블록은 FS를 기본으로 하고 있기 때문에 FS를 이용하여 한 MB를 처리하는데 걸리는 시간을 계산하고 그 값을 최대 연산 수행 시간으로 정한다. 그 다음 매 MB 마다 조기 은퇴 기법을 적용하여 연산이 조기 종료 되어 최대 연산 수행 기간에 이르기까지 남는 시간 동안 섯다운을 하는 것보다 여러 개의 MB를 grouping을 하여 오랫동안 섯다운을 하는 것이 효율이 더 좋기 때문에 grouping 될 MB의 개수를 정한다. grouping할 MB의 개수와 최대 연산 수행 시간이 결정 되면 main controller가 PMU의 timer에 grouping 할 MB의 개수와 최대 연산 수행 시간을 동작 주파수 f의 한 주기 시간으로 나누어 cycle로 표현한 수를 timer에

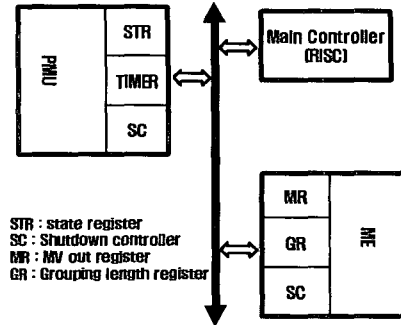


그림 8. 시스템 수준 아키텍처
 Fig. 8. System-level architecture.

셋팅하고 제안된 움직임 추정 블록의 GR (grouping register) 레지스터에 grouping 될 MB의 개수를 셋팅한다. GR은 일종의 포인터이며 블록 섯다운을 하더라도 다른 블록들을 위하여 움직임 벡터 (motion vector: MV)를 저장하기 위한 FIFO 구조의 register인 MR (motion vector register) 레지스터가 grouping 개수만큼 채워졌는지를 감지하는 역할을 한다. 시스템이 동작하면서 제안하는 구조의 움직임 추정 블록이 GR이 가지고 있는 MB 개수만큼 MV가 MR에 저장이 되면 PMU에게 블록 섯다운을 요청한다. PMU는 timer를 체크하여 움직임 추정 블록을 섯다운 시키고 상태 레지스터인 STR(status register)에 현재 상태(섯다운)를 기록하고 timer가 다운 카운팅을 하여 PE의 개수 N과 같아지는 순간 움직임 추정 블록을 wake-up하면서 STR을 다시 wake-up 상태로 기록 한다. 전력선 잡음을 고려하여 PE를 각각 1 cycle에 한 개씩 분할해서 파이프라인 방식으로 wake-up 동작을 수행하여 주변 블록의 오동작 발생을 방지하는 것이 가능하다.

이러한 제안하는 시스템 적용 구조를 HDL로 구현할 때 GR의 최대값을 16이라고 가정하면 MV를 저장하기 위한 MR과 GR에 저장된 값과 MR에 저장된 MV의 개수가 일치하는지 감시하는 포인터 블록이 약 3천 게이트 정도 추가 되며 제어 신호 요청 및 승인과 현재 상태를 나타내기 위한 제어 스테이트머신이 약 1천 게이트 정도 추가 된다. 제안하는 움직임 추정기 구조에 약 20%정도의 추가적인 면적으로 시스템 수준에서 동작할 수 있다.

V. 모의실험 결과

1. 전력 계산 방법

본 논문에서는 디지털 회로의 스위칭으로 인해 발생

하는 동적 전력 소모와 누설 전류로 인해 발생하는 정적 전력 소모를 각각 따로 따로 계산하였다. 일반적으로 ME의 경우 동일한 하드웨어에서 하나의 동영상 처리에 소모되는 단위 cycle당 전력은 평균적으로 같다. 따라서 전체 전력 소모는 수행되는 사이클에 직접 비례하게 된다. 그러므로 감소된 동적 전력은 단위 cycle당 소모하는 평균 전력이 동일하다는 가정 하에 FS를 이용하여 단위 MB의 MV를 연산하는데 걸리는 cycle과 초기 은퇴 개념을 이용하여 단위 MB의 MV를 연산하는데 걸리는 cycle의 비로 나타 낼 수 있다. 감소된 정적 전력은 FS를 이용하여 단위 MB의 MV를 연산하는데 걸리는 cycle과 FS를 이용한 cycle 에서 초기 은퇴 개념을 이용하여 단위 MB의 MV를 연산하는데 걸린 cycle의 차이 cycle (셋다운과 wake-up 을 하기 위한 control cycle은 제외)과의 비를 나타낸다.

전체 감소된 전력은 각 공정 별로 정적 전력 소모가 전체 전력에서 차지하는 비율(static power ratio: SPR)을 감소된 전력 소모 값에 곱하고 동적 전력 소모가 전체 전력에서 차지하는 비율(dynamic power ratio: DPR)을 감소된 동적 전력 소모 값에 곱하여 두 값을 더한 값이 된다. 예를 들어 0.13 μm 공정일 경우 정적 전력 소모가 전체의 10% 정도를 차지하고 동적 전력 소모가 90% 정도를 차지하므로 감소된 정적 전력 소모에 0.1을 곱하고 감소된 동적 전력에 0.9를 곱하여 더한 값이 된다. 이를 정리하면 수식 (3)-(5)와 같다. 이때, $\Delta P_{N,D}$, $\Delta P_{N,S}$, $\Delta P_{N,T}$ 는 각각 정규화된 동적 전력 소모 감소, 정적 전력 감소, 전체 전력 소모 감소이며, $N_{NoBreakOff}$, $T_{NoBreakOff}$ 는 초기 은퇴 없이 FS를 수행했을 때의 수행 사이클 수 및 수행 시간, $N_{BreakOff}$, $T_{BreakOff}$ 는 초기 은퇴를 적용하였을 때의 수행 사이클 수 및 수행 시간이다. $\Delta P_{N,D}$ 를 계산할 때에는 수행 사이클 수를, $\Delta P_{N,S}$ 를 계산할 때에는 수행 시간을 사용한 이유는 동적 전력 소모의 경우 전력 소모가 회로의 스위칭에 의해 발생하고 정적 전력 소모는 회로가 동작하는 시간 동안 흐르는 누설 전류에 의해 발생하기 때문이다.

$$\Delta P_{N,D} = \frac{N_{NoBreakOff} - N_{BreakOff}}{N_{NoBreakOff}} \quad (3)$$

$$\Delta P_{N,S} = \frac{T_{NoBreakOff} - T_{BreakOff}}{T_{NoBreakOff}} \quad (4)$$

$$\Delta P_{N,T} = \Delta P_{N,D} \times DPR + \Delta P_{N,S} \times SPR \quad (5)$$

표 2는 차세대 공정에 대한 동적 전력과 정적 전력의 비를 모의실험을 통해 측정된 결과이다. 표 2에서 알

표 2. 공정 세대에 따른 동적 전력과 정적 전력의 비율

Table 2. Ratio of dynamic power and static power with technology generation.

	Process Generation (μm)			
	0.13	0.10	0.07	0.05
DPR	0.9	0.8	0.6	0.5
SPR	0.1	0.2	0.4	0.5

표 3. 모의실험 결과

Table 3. Simulation results.

사용 영상	PSNR 성능 (dB)	$\Delta P_{N,D}$	$\Delta P_{N,S}$	$1 - \Delta P_{N,T}$
Trevor	36.52	0.39	0.43	0.61
Carphone	32.34	0.32	0.36	0.66
Miss America	32.70	0.34	0.38	0.66
Flower Garden	32.10	0.31	0.34	0.70

수 있듯이 공정 기술이 발전할수록 전체에서 정적 전력 소모가 차지하는 비율이 높아지는 것을 알 수 있다.

2. 전력 감소 결과

앞에서 설명한 바와 같이 초기 은퇴 기법을 이용하여 연산량을 평균적으로 60~70% 정도 줄임으로서 동적 전력 소모를 1/2 정도 줄일 수 있으며 연산이 초기 종료되기 때문에 여분의 시간 동안 셋다운을 하여 정적 전력 소모도 줄일 수 있다. 표 3은 일반적인 테스트 영상을 사용한 모의실험 결과이다. 이때, 파이프라인 방식의 wake-up을 위해서 추가로 소모되는 N개의 cycle 때문에 $\Delta P_{N,D}$ 와 $\Delta P_{N,S}$ 는 약간의 차이를 보이게 된다. 여기서 $\Delta PSNR$ 은 FS를 수행한 PSNR 값과 제안하는 움직임 추정기의 PSNR 값의 차분을 나타낸다. 표 3에서 보듯 정규화된 전력 감소 이득인 $1 - \Delta P_{N,T}$ 는 약 0.6에 달하며, 제안된 움직임 추정기는 기존에 비해 약 60~70% 정도의 전력 소모를 절감할 수 있다는 것을 알 수 있다.

그림 9는 동적 전력만을 고려하였을 때(P_{act})와 정적 전력과 함께 고려하였을 때($P_{act+leak}$) 전력 감소 이득 $1 - \Delta P_{N,T}$ 의 온도에 따른 변화를 나타낸 것이다. 온도가 증가함에 따라 누설 전류가 커지기 때문에 정적 전력 소모가 차지하는 비율도 높아진다. 멀티미디어 SoC의 집적도가 높아질수록 칩 중심부의 전력 소모가 높아져서 내부 온도가 증가하기 때문에, 기존의 기법처럼 동적 전력만을 고려하는 경우 효율이 떨어지게 된다. 반면에 제안하는 기법의 경우 온도 변화에 따른 누설 전류의 증가에 의한 효율의 저하가 매우 적기 때문에 칩

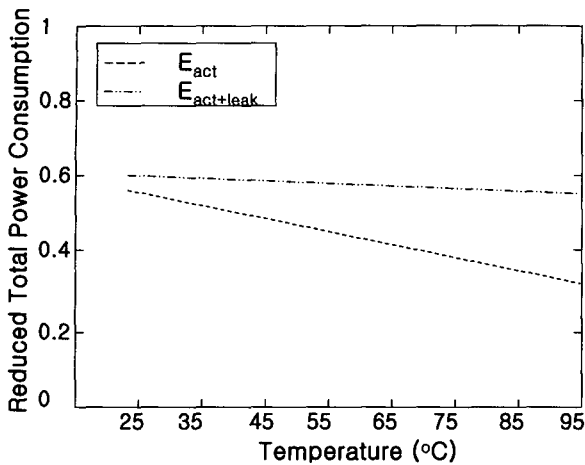


그림 9. 온도 변화에 따른 전력 감소 이득
Fig. 9. Power reduction gain vs. temperature.

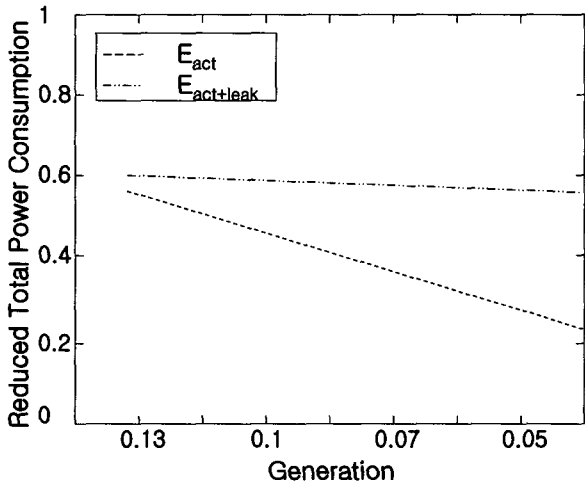


그림 10. 공정 기술의 변화에 따른 전력 감소 이득
Fig. 10. Power reduction gain vs. process technology.

내부의 온도가 높아져도 큰 영향을 미치지 않음을 알 수 있다.

그림 10은 공정 기술이 변화할 때의 전력 감소 이득을 나타낸다. 공정 기술이 발전하면서 선폭이 감소할 때 기존의 기법에서는 동적 전력만을 고려하기 때문에 누설 전류의 극심한 증가로 인하여 전력 감소 이득이 심하게 나빠지지만 제안하는 기법에서는 큰 영향을 미치지 않음을 알 수 있다.

VI. 결 론

본 논문에서는 공정 기술이 deep sub-micron 공정으로 발전해나감에 따라 동적 전력 소모뿐만 아니라 정적 전력 소모도 함께 고려하여 전력 소모를 줄이는 새로운 움직임 추정기 아키텍처와 이를 위한 시스템 수준 기법을 제안하였다. 제안하는 저전력 움직임 추정기 아키텍

처는 조기 은퇴 기법을 사용하여 연산량을 줄이며 움직임 추정기가 주어진 연산을 끝마치면 움직임 추정기 전체를 섯다운 시킴으로서 동적 전력 소모뿐만 아니라 정적 전력 소모까지 함께 줄일 수 있다. deep sub-micron 공정에서 블록 단위 섯다운이 전력선 잡음을 일으켜서 오동작을 초래하는 문제점을 방지하기 위해서 파이프라인 블록 단위 섯다운 기법을 사용하였다. 제안하는 저전력 움직임 추정기 아키텍처는 하드웨어 구현이 용이하며 기존의 멀티미디어 SoC에 seamless한 제어 신호로 연결이 가능하였다.

제안하는 움직임 추정기 아키텍처를 HDL로 구현하고 모의실험을 수행한 결과, 여러 개의 테스트 영상에 걸쳐서 고르게 전력 소모가 감소하였으며, 전체 전력 소모의 60~70% 정도를 감소시킬 수 있음을 확인하였다. 특히 칩 내부의 온도 변화와 공정 기술 발전에 따른 선폭의 감소에 관련한 모의실험 결과, 기존의 기법이 누설 전류의 증가로 인해 전력 감소 효율이 감소하는 반면에 제안하는 기법은 큰 문제 없이 효율적으로 전력을 감소시킬 수 있었다.

참 고 문 헌

- [1] A. Chandrakasan and R. Brodersen, *Low Power Digital CMOS Design*, Kluwer Academic Publishers, 1995.
- [2] K. Nose, M. Hirabayashi, H. Kawaguchi, S. Lee, and T. Sakurai, "V_{TH}-Hopping Scheme to Reduce Subthreshold Leakage for Low-Power Processors," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 3, pp. 413-419, Mar. 2002.
- [3] ISO/IEC JTC1/SC29/WG11 14496-2, "Coding of Audiovisual Object: Visual," 1998.
- [4] T. Shimizu, F. Arakawa, and T. Kawahara, "Autonomous Decentralized Low-power System LSI Using Self-Instructing Predictive Shutdown Method," *Proceedings of Symposium on VLSI Circuits*, pp. 55 - 56, 2001.
- [5] J. Choi, Y. Kim, J. Wee and S. Lee, "Pipelined Wake-Up Scheme to Reduce Power line Noise for Block-Wise Shutdown of Low-Power VLSI Systems," *IEICE Transactions on Electronics*, vol. E87-C, no. 4, pp. 629 - 633, Apr. 2004.
- [6] Y. Lin, C. Wu, C. Chang, R. Yang, W. Chen, J. Liaw, and C. Diaz, "Leakage Scaling in Deep Submicron CMOS for SoC", *IEEE Transactions on Electron Devices*, vol. 49, no. 6, pp. 1034 - 1041, Jun. 2002.

- [7] Q. Wang and S. Roy, "Power Minimization by Clock Root Gating," Proceedings of Asia South Pacific Design Automation Conference, pp. 249 - 254, 2003.
- [8] T. Simunic, L. Benini, A. Acquaviva, P. Glynn, and G. de Micheli, "Dynamic Voltage Scaling and Power Management for Portable Systems," Proceedings on Design Automation Conference, pp. 524- 529, 2001.- 529
- [9] G. Magklis, G. Semeraro, D. Albonesi, S. Dropsho, S. Dwarkadas, and M. Scott, "Dynamic Frequency and Voltage Scaling for a Multiple-clock-domain Microprocessor," *IEEE Micro*, vol. 23, no. 6, pp. 62 - 68, Nov. 2003.
- [10] C. Kim and K. Roy, "Dynamic V_{TH} Scaling Scheme for Active Leakage Power Reduction," Proceedings of Design, Automation and Test in Europe, pp. 163 - 167, 2002.
- [11] J. Jain and A. Jain, "Displacement Measurement and Its Application in Interframe Image Coding," *IEEE Trans. on Communications*, vol. 29, no. 12, pp. 1799-1808, Dec. 1981.

저 자 소 개



연 규 성(정회원)
 2003년 한림대학교 전자공학과
 학사 졸업.
 2004년 현재 숭실대학교
 정보통신공학과 석사과정.

<주관심분야: Mobile 시스템 설계, block-wise shutdown 기법, 저전력 멀티미디어 시스템, 저전력 MPEG 프로세서, motion estimation, SIP/SOC를 위한 IP 집적화 및 on-chip 데이터 통신>



전 치 훈(정회원)
 2003년 한림대학교 전자공학과
 학사 졸업.
 2004년 현재 숭실대학교
 정보통신공학과 석사과정.

<주관심분야: On-chip 데이터 통신 알고리즘 및 구조, 저전력 MPEG, SIP/SOC를 위한 IP 집적화 및 on-chip 데이터 통신>



황 태 진(정회원)
 2003년 한림대학교 전자공학과
 학사 졸업.
 2004년 현재 숭실대학교
 정보통신공학과 석사과정.

<주관심분야: 저전력 CMOS 회로 설계, 고속 및 고성능 DLL/PLL 설계, 고속 clocking 시스템, 아날로그 설계 기법>



이 성 수(정회원)
 1991년 서울대학교
 전자공학과 학사
 1993년 서울대학교
 전자공학과 석사
 1998년 서울대학교
 전기공학부 박사

1998년~2000년 University of Tokyo Research Associate

2000년~2002년 이화여자대학교 정보통신학과 연구교수

2002년~현재 숭실대학교 정보통신전자공학부 조교수

<주관심분야: 저전력 SoC 설계, 저전력 멀티미디어 신호 처리, MPEG4/H.264 SoC 설계>



위 재 경(정회원)
 1988년 연세대학교 물리학과
 학사 졸업.
 1990년 서울대학교 물리학과
 석사 졸업
 1998년 서울대학교 전자공학과
 박사 졸업.

1990년~2002년 하이닉스전자 메모리연구소 근무

2002년~2004년 한림대학교 정보통신공학부 조교수

2004년~현재 숭실대학교 정보통신전자공학부 조교수

<주관심분야: System-on-Package 설계, high speed I/O interface, DLL/PLL, Mixed Mode 설계, 저전력 고속 SoC>