



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2019년08월09일
(11) 등록번호 10-2009425
(24) 등록일자 2019년08월05일

(51) 국제특허분류(Int. Cl.)
G06F 9/50 (2018.01)
(52) CPC특허분류
G06F 9/5094 (2013.01)
G06F 2209/508 (2013.01)
(21) 출원번호 10-2018-0016213
(22) 출원일자 2018년02월09일
심사청구일자 2018년02월09일
(56) 선행기술조사문헌
KR1020120066189 A*
(뒷면에 계속)

(73) 특허권자
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
(72) 발명자
백웅기
울산광역시 울주군 언양읍 유니스트길 50 울산과
학기술대학교
박진수
울산광역시 울주군 언양읍 유니스트길 50 울산과
학기술대학교
박성범
울산광역시 울주군 언양읍 유니스트길 50 울산과
학기술대학교
(74) 대리인
제일특허법인(유)

전체 청구항 수 : 총 12 항

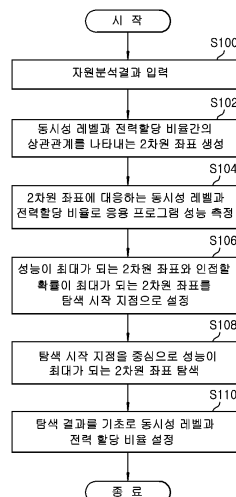
심사관 : 유진태

(54) 발명의 명칭 **컴퓨팅 시스템의 성능 관리 방법, 이를 위한 컴퓨터 프로그램 및 컴퓨터 판독 가능 기록 매체**

(57) 요약

본 발명은 컴퓨팅 시스템의 성능 관리 기술에 관한 것으로, 컴퓨팅 시스템의 자원 분석 결과를 기초로 상기 컴퓨팅 시스템의 동시성 레벨(concurrency level)과 전력 할당 비율(power allocation ratio)로 정의되는 시스템 상태를 나타내는 적어도 둘 이상의 2차원 좌표를 생성하는 단계; 상기 적어도 둘 이상의 2차원 좌표 각각에 대응하는 동시성 레벨과 전력 할당 비율로 상기 컴퓨팅 시스템에서 실행되는 응용 프로그램의 성능 및 LLC MPKI (Last-Level Cache Misses Per Kilo Instructions)를 측정하는 단계; 상기 측정되는 성능 및 LLC MPKI를 기초로 상기 응용 프로그램의 특성을 파악하여 상기 응용 프로그램의 성능이 최대가 되는 2차원 좌표와 인접할 확률이 최대가 되는 2차원 좌표를 탐색 시작 지점으로 설정하는 단계; 및 상기 탐색 시작 지점을 중심으로 상기 응용 프로그램의 성능이 최대가 되는 2차원 좌표를 탐색하는 단계를 포함할 수 있다.

대표도 - 도3



(56) 선행기술조사문헌

JP2016134120 A*

KR1020160063974 A*

KR101509938 B1

KR1020160085029 A

*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호 1711058252

부처명 과학기술정보통신부

연구관리전문기관 한국연구재단

연구사업명 SW전문인력역량강화

연구과제명 Peta Flops 스케일의 기계학습 프레임워크 개발

기 여 율 1/2

주관기관 울산과학기술원

연구기간 2017.08.01 ~ 2018.05.31

이 발명을 지원한 국가연구개발사업

과제고유번호 1711056733

부처명 과학기술정보통신부

연구관리전문기관 정보통신기술진흥센터

연구사업명 SW컴퓨팅산업원천기술개발

연구과제명 빅데이터 처리 고도화 핵심 기술개발 사업 총괄 및 고성능 컴퓨팅 기술을 활용한 성능 가속화 기술 개발

기 여 율 1/2

주관기관 울산과학기술원

연구기간 2017.06.01 ~ 2018.03.31

명세서

청구범위

청구항 1

컴퓨팅 시스템의 자원 분석 결과를 기초로 상기 컴퓨팅 시스템의 동시성 레벨(concurrency level)과 전력 할당 비율(power allocation ratio)로 정의되는 시스템 상태를 나타내는 적어도 둘 이상의 2차원 좌표를 생성하는 단계;

상기 적어도 둘 이상의 2차원 좌표 각각에 대응하는 동시성 레벨과 전력 할당 비율로 상기 컴퓨팅 시스템에서 실행되는 응용 프로그램의 성능 및 LLC MPKI(Last-Level Cache Misses Per Kilo Instructions)를 측정하는 단계;

상기 측정되는 성능 및 LLC MPKI를 기초로 상기 응용 프로그램의 특성을 파악하여 상기 응용 프로그램의 성능이 최대가 되는 2차원 좌표와 인접할 확률이 최대가 되는 좌표를 탐색 시작 지점으로 설정하는 단계; 및

상기 탐색 시작 지점을 중심으로 상기 응용 프로그램의 성능이 최대가 되는 2차원 좌표를 탐색하는 단계를 포함하고,

상기 시스템 상태는,

확장성(scalable)-코어 집중적(core-intensive), 확장성-메모리 집중적(memory intensive), 비확장성(non-scalable)-코어 집중적 및 비확장성-메모리 집중적 시스템 상태 특성을 갖는

컴퓨팅 시스템의 성능 관리 방법.

청구항 2

제 1 항에 있어서,

상기 생성하는 단계는,

상기 컴퓨팅 시스템의 코어 개수에 따라 상기 동시성 레벨을 구분한 좌표 성분을 2차원 좌표축의 일 축 상에 생성하는 단계;

상기 컴퓨팅 시스템의 CPU 및 메모리에 대한 CPU의 전력 할당 비율에 따라 상기 전력 할당 비율을 구분한 좌표 성분을 상기 2차원 좌표축의 타 축 상에 생성하는 단계;

상기 일 축 상의 좌표 성분과 상기 타 축 상의 좌표 성분을 조합하여 상기 시스템 상태를 나타내는 적어도 둘 이상의 2차원 좌표를 생성하는 단계를 포함하는

컴퓨팅 시스템의 성능 관리 방법.

청구항 3

삭제

청구항 4

제 2 항에 있어서,

상기 측정하는 단계는,

상기 코어 개수 별로 상기 응용 프로그램을 실행했을 때의 수행 시간 및 상기 LLC MPKI를 측정하는 단계를 포함하는

컴퓨팅 시스템의 성능 관리 방법.

청구항 5

제 4 항에 있어서,

상기 설정하는 단계는,

상기 측정되는 수행 시간에 따라 상기 동시성 레벨의 확장성 또는 비확장성 여부를 결정하는 단계;

상기 확장성 또는 비확장성 여부가 결정되면 상기 측정되는 LLC MPKI에 따라 상기 전력 할당 비율의 코어 집중적 또는 메모리 집중적 여부를 결정하는 단계를 포함하는

컴퓨팅 시스템의 성능 관리 방법.

청구항 6

제 2 항에 있어서,

상기 탐색하는 단계는,

상기 탐색 시작 지점을 중심으로 상기 일 측 상의 좌표 성분 또는 상기 타 측 상의 좌표 성분을 기 설정된 단위로 증감시키면서 상기 응용 프로그램의 성능이 최대가 되는 좌표 성분을 결정하는 단계를 포함하는

컴퓨팅 시스템의 성능 관리 방법.

청구항 7

제 6 항에 있어서,

상기 탐색하는 단계는,

상기 일 측 또는 상기 타 측 상에서 기 설정된 단위로 좌표 성분을 증가시키는 단계;

상기 증가된 좌표 성분에서의 상기 응용 프로그램의 성능을 측정하는 단계;

상기 증가된 좌표 성분에서의 성능이 이전 좌표 성분에서의 성능보다 높으면 상기 기 설정된 단위로 상기 좌표 성분을 증가시키는 단계;

상기 증가된 좌표 성분에서의 성능이 이전 좌표 성분에서의 성능보다 낮으면 상기 이전 좌표 성분에서 상기 기 설정된 단위로 상기 좌표 성분을 감소시키는 단계;

상기 감소된 좌표 성분에서의 성능이 상기 이전 좌표 성분에서의 성능보다 높으면 상기 기 설정된 단위로 상기 좌표 성분을 감소시키는 단계; 및

상기 감소된 좌표 성분에서의 성능이 상기 이전 좌표 성분에서의 성능보다 낮고, 상기 일 측 및 상기 타 측에 대한 성능 측정이 완료되었으면 상기 이전 좌표 성분을 갖는 2차원 좌표를 탐색 결과로 결정하는 단계를 포함하는

컴퓨팅 시스템의 성능 관리 방법.

청구항 8

제 7 항에 있어서,

상기 성능은 상기 응용 프로그램의 단위 작업 당 수행 시간의 역수인

컴퓨팅 시스템의 성능 관리 방법.

청구항 9

제 7 항에 있어서,

상기 탐색 결과를 기초로 상기 컴퓨팅 시스템의 코어 개수와, 상기 컴퓨팅 시스템의 CPU 및 메모리에 대한 CPU의 전력 할당 비율을 설정하는 단계를 더 포함하는

컴퓨팅 시스템의 성능 관리 방법.

청구항 10

제 2 항에 있어서,

상기 탐색하는 단계는,

볼록 함수(convex function) 또는 오목 함수(concave function)의 특성을 이용한 탐색 알고리즘으로 상기 2차원 좌표를 탐색하는 단계를 포함하는

컴퓨팅 시스템의 성능 관리 방법.

청구항 11

제 1 항에 있어서,

상기 컴퓨팅 시스템은 전력 제한 컴퓨팅 시스템(Power-constrained computing system)을 포함하는

컴퓨팅 시스템의 성능 관리 방법.

청구항 12

컴퓨팅 시스템의 자원 분석 결과를 기초로 상기 컴퓨팅 시스템의 동시성 레벨과 전력 할당 비율로 정의되는 시스템 상태를 나타내는 적어도 둘 이상의 2차원 좌표를 생성하는 단계 - 상기 시스템 상태는, 확장성(scalable)-코어 집중적(core-intensive), 확장성-메모리 집중적(memory intensive), 비확장성(non-scalable)-코어 집중적 및 비확장성-메모리 집중적 시스템 상태 특성을 가짐 -;

상기 적어도 둘 이상의 2차원 좌표 각각에 대응하는 동시성 레벨과 전력 할당 비율로 상기 컴퓨팅 시스템에서 실행되는 응용 프로그램의 성능 및 LLC MPKI를 측정하는 단계;

상기 측정되는 성능 및 LLC MPKI를 기초로 상기 응용 프로그램의 특성을 파악하여 상기 응용 프로그램의 성능이 최대가 되는 2차원 좌표와 인접할 확률이 최대가 되는 좌표를 탐색 시작 지점으로 설정하는 단계; 및

상기 탐색 시작 지점을 중심으로 상기 응용 프로그램의 성능이 최대가 되는 2차원 좌표를 탐색하는 단계를 수행하는 명령어를 포함하는 프로그램이 기록된

컴퓨터 판독 가능 기록 매체.

청구항 13

컴퓨팅 시스템의 자원 분석 결과를 기초로 상기 컴퓨팅 시스템의 동시성 레벨과 전력 할당 비율로 정의되는 시스템 상태를 나타내는 적어도 둘 이상의 2차원 좌표를 생성하는 단계 - 상기 시스템 상태는, 확장성(scalable)-코어 집중적(core-intensive), 확장성-메모리 집중적(memory intensive), 비확장성(non-scalable)-코어 집중적 및 비확장성-메모리 집중적 시스템 상태 특성을 가짐 -;

상기 적어도 둘 이상의 2차원 좌표 각각에 대응하는 동시성 레벨과 전력 할당 비율로 상기 컴퓨팅 시스템에서 실행되는 응용 프로그램의 성능 및 LLC MPKI를 측정하는 단계;

상기 측정되는 성능 및 LLC MPKI를 기초로 상기 응용 프로그램의 특성을 파악하여 상기 응용 프로그램의 성능이 최대가 되는 2차원 좌표와 인접할 확률이 최대가 되는 좌표를 탐색 시작 지점으로 설정하는 단계; 및

상기 탐색 시작 지점을 중심으로 상기 응용 프로그램의 성능이 최대가 되는 2차원 좌표를 탐색하는 단계를 수행하는

컴퓨터 판독 가능 기록 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 컴퓨팅 시스템의 성능 관리 기술에 관한 것으로, 특히 전력 제한 컴퓨팅(Power-constrained computing) 환경에서의 성능 관리 기술에 관한 것이다.

배경 기술

[0002] 전력 제한 컴퓨팅 시스템은 주어진 전력 설계(power budget) 하에서 컴퓨팅 성능을 최적화하는 것을 목표로 하

고 있으며, 데이터 센터(Data Center)에서 임베디드 시스템(Embedded System)까지 다양한 분야에 걸쳐 그 중요성이 커지고 있다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) 한국등록특허 제10-1620103호, 2016.05.03 등록

발명의 내용

해결하려는 과제

- [0004] 본 발명의 실시예에서는 전력 제한 컴퓨팅(power constrained computing) 환경에서 동시성 레벨(concurrency level)과 전력 할당 비율(power allocation ratio)을 동시에 적응적으로 조절하여 응용 프로그램 별 시스템 성능을 최적화할 수 있는 기술을 제안하고자 한다.
- [0005] 구체적으로, 본 발명의 실시예에서는 전력 제한 컴퓨팅 환경에서 수행되는 응용 프로그램 별로 CPU(Central Processing Unit)의 코어(core) 개수, CPU 및 메모리에 대한 CPU의 전력 할당 비율 등을 적응적으로 설정하여 컴퓨팅 수행 시간을 최소화할 수 있는 전력 제한 컴퓨팅 환경에서의 시스템 성능 관리 기술을 제안하고자 한다.
- [0006] 본 발명이 해결하고자 하는 과제는 상기에서 언급한 것으로 제한되지 않으며, 언급되지 않은 또 다른 해결하고자 하는 과제는 아래의 기재들로부터 본 발명이 속하는 통상의 지식을 가진 자에 의해 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0007] 본 발명의 실시예에 따르면, 컴퓨팅 시스템의 자원 분석 결과를 기초로 상기 컴퓨팅 시스템의 동시성 레벨(concurrency level)과 전력 할당 비율(power allocation ratio)로 정의되는 상기 컴퓨팅 시스템의 상태를 나타내는 적어도 둘 이상의 2차원 좌표를 생성하는 단계; 상기 적어도 둘 이상의 2차원 좌표 각각에 대응하는 동시성 레벨과 전력 할당 비율로 상기 컴퓨팅 시스템에서 실행되는 응용 프로그램의 성능 및 LLC MPKI(Last-Level Cache Misses Per Kilo Instructions)를 측정하는 단계; 상기 측정되는 성능 및 LLC MPKI를 기초로 상기 응용 프로그램의 특성을 파악하여 상기 응용 프로그램의 성능이 최대가 되는 2차원 좌표와 인접할 확률이 최대가 되는 2차원 좌표를 탐색 시작 지점으로 설정하는 단계; 및 상기 탐색 시작 지점을 중심으로 상기 응용 프로그램의 성능이 최대가 되는 2차원 좌표를 탐색하는 단계를 포함하는 컴퓨팅 시스템의 성능 관리 방법을 제공할 수 있다.
- [0008] 여기서, 상기 생성하는 단계는, 상기 컴퓨팅 시스템의 코어 개수에 따라 상기 동시성 레벨을 구분한 좌표 성분을 2차원 좌표축의 일 축 상에 생성하는 단계; 상기 컴퓨팅 시스템의 CPU 및 메모리에 대한 CPU의 전력 할당 비율에 따라 상기 전력 할당 비율을 구분한 좌표 성분을 상기 2차원 좌표축의 타 축 상에 생성하는 단계; 상기 일 축 상의 좌표 성분과 상기 타 축 상의 좌표 성분을 조합하여 상기 상관 관계를 나타내는 적어도 둘 이상의 2차원 좌표를 생성하는 단계를 포함할 수 있다.
- [0009] 또한, 상기 상관 관계는, 확장성(scalable)-코어 집중적(core-intensive), 확장성-메모리 집중적(memory intensive), 비확장성(non-scalable)-코어 집중적 및 비확장성-메모리 집중적 상관 관계 특성을 가질 수 있다.
- [0010] 또한, 상기 측정하는 단계는, 상기 코어 개수 별로 상기 응용 프로그램을 실행했을 때의 수행 시간 및 상기 LLC MPKI를 측정하는 단계를 포함할 수 있다.
- [0011] 또한, 상기 설정하는 단계는, 상기 측정되는 수행 시간에 따라 상기 동시성 레벨의 확장성 또는 비확장성 여부를 결정하는 단계; 상기 확장성 또는 비확장성 여부가 결정되면 상기 측정되는 LLC MPKI에 따라 상기 전력 할당 비율의 코어 집중적 또는 메모리 집중적 여부를 결정하는 단계를 포함할 수 있다.
- [0012] 또한, 상기 탐색하는 단계는, 상기 탐색 시작 지점을 중심으로 상기 일 축 상의 좌표 성분 또는 상기 타 축 상의 좌표 성분을 기 설정된 단위로 증감시키면서 상기 응용 프로그램의 성능이 최대가 되는 좌표 성분을 결정하는 단계를 포함할 수 있다.

- [0013] 또한, 상기 탐색하는 단계는, 상기 일 축 또는 상기 타 축 상에서 기 설정된 단위로 좌표 성분을 증가시키는 단계; 상기 증가된 좌표 성분에서의 상기 응용 프로그램의 성능을 측정하는 단계; 상기 증가된 좌표 성분에서의 성능이 이전 좌표 성분에서의 성능보다 높으면 상기 기 설정된 단위로 상기 좌표 성분을 증가시키는 단계; 상기 증가된 좌표 성분에서의 성능이 이전 좌표 성분에서의 성능보다 낮으면 상기 이전 좌표 성분에서 상기 기 설정된 단위로 상기 좌표 성분을 감소시키는 단계; 상기 감소된 좌표 성분에서의 성능이 상기 이전 좌표 성분에서의 성능보다 높으면 상기 기 설정된 단위로 상기 좌표 성분을 감소시키는 단계; 및 상기 감소된 좌표 성분에서의 성능이 상기 이전 좌표 성분에서의 성능보다 낮고, 상기 일 축 및 상기 타 축에 대한 성능 측정이 완료되었으면 상기 이전 좌표 성분을 갖는 2차원 좌표를 탐색 결과로 결정하는 단계를 포함할 수 있다.
- [0014] 또한, 상기 성능은 상기 응용 프로그램의 단위 작업 당 수행 시간의 역수일 수 있다.
- [0015] 또한, 상기 탐색 결과를 기초로 상기 컴퓨팅 시스템의 코어 개수와 상기 컴퓨팅 시스템의 CPU 및 메모리에 대한 CPU의 전력 할당 비율을 설정하는 단계를 더 포함할 수 있다.
- [0016] 또한, 상기 탐색하는 단계는, 볼록 함수(convex function) 또는 오목 함수(concave function)의 특성을 이용한 탐색 알고리즘으로 상기 2차원 좌표를 탐색하는 단계를 포함할 수 있다.
- [0017] 또한, 상기 컴퓨팅 시스템은 전력 제한 컴퓨팅 시스템을 포함할 수 있다.

발명의 효과

- [0019] 본 발명의 실시예에 의하면, 전력 제한 컴퓨팅 환경에서 동시성 레벨과 전력 할당 비율을 동시에 적응적으로 조절하여 응용 프로그램 별 시스템 성능을 최적화할 수 있다. 구체적으로, 전력 제한 컴퓨팅 환경에서 수행되는 응용 프로그램 별로 CPU 코어 개수와, CPU 및 메모리에 대한 CPU의 전력 할당 비율을 동시에 적응적으로 설정하여 컴퓨팅 수행 시간을 최소화시키면서 동적인 성능 관리가 가능하다. 또한, 본 발명의 실시예에 의하면, 컴퓨팅 시스템에 대한 온라인 프로파일링(Online Profiling)이 가능하여 특정 응용 프로그램에 대한 특성을 런타임(run-time)에 파악할 수 있다.

도면의 간단한 설명

- [0020] 도 1은 본 발명의 실시예에 따른 전력 제한 컴퓨팅 시스템의 성능 관리 장치에 대한 블록도이다.
- 도 2는 본 발명의 실시예에 따른 전력 제한 컴퓨팅 시스템의 성능 관리를 위한 동시성 레벨과 전력 할당 비율 간의 상관 관계 그래프이다.
- 도 3은 본 발명의 실시예에 따른 전력 제한 컴퓨팅 시스템의 성능 관리 방법을 예시적으로 설명하는 흐름도이다.
- 도 4는 도 3의 탐색 시작 지점을 중심으로 응용 프로그램의 성능이 최대가 되는 2차원 좌표를 탐색하는 과정을 보다 상세히 나타낸 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0021] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명의 범주는 청구항에 의해 정의될 뿐이다.
- [0022] 본 발명의 실시예들을 설명함에 있어서 공지 기능 또는 구성에 대한 구체적인 설명은 본 발명의 실시예들을 설명함에 있어 실제로 필요한 경우 외에는 생략될 것이다. 그리고 후술되는 용어들은 본 발명의 실시예에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [0023] 본 발명의 실시예는 전력 제한 컴퓨팅(Power-constrained computing) 환경에서 동시성 레벨과 전력 할당 비율을 동시에 적응적으로 조절하여 응용 프로그램 별 시스템 성능을 최적화하도록 하며, 구체적으로, 전력 제한 컴퓨팅 환경에서 수행되는 응용 프로그램 별로 CPU(Central Processing Unit)의 코어(core) 개수와, CPU 및 메모리에 대한 CPU의 전력 할당 비율을 동시에 적응적으로 설정하여 컴퓨팅 수행 시간을 최소화시키면서 동적으로 성

능을 관리할 수 있는 컴퓨팅 시스템의 성능 관리 기술을 제안하고자 한다.

- [0024] 이하, 첨부된 도면을 참조하여 본 발명의 실시예에 대해 상세히 설명하기로 한다.
- [0025] 도 1은 본 발명의 실시예에 따른 컴퓨팅 시스템, 예를 들어 전력 제한 컴퓨팅 시스템(1)의 성능 관리 장치(100)를 나타낸 블록도이다.
- [0026] 도 1의 성능 관리 장치(100)는 시스템 자원 분석부(102), 시스템 성능 관리부(104), 저장부(106) 및 응용 프로그램 실행부(108)를 포함하며, 이러한 성능 관리 장치(100)는 사용자 인터페이스부(10) 및 전원부(20)와 연결될 수 있다.
- [0027] 도 1에 도시한 바와 같이, 시스템 자원 분석부(102)는 전력 제한 컴퓨팅 시스템(1)의 하드웨어 사양(hardware specification), 예를 들어 CPU 코어, 스레드(thread), 메모리 등과 관련된 시스템 자원과, 전력 설계(power budget) 또는 전력 정책과 관련된 시스템 자원을 분석할 수 있다.
- [0028] 시스템 성능 관리부(104)는 시스템 자원 분석부(102)에서 분석된 시스템 자원 분석 결과를 기초로 전력 제한 컴퓨팅 시스템(1)의 성능 관리를 위한 2차원 좌표를 생성할 수 있다. 이러한 2차원 좌표는 전력 제한 컴퓨팅 시스템(1)의 동시성 레벨(concurrency level)과 전력 할당 비율(power allocation ratio)로 정의되는 시스템 상태를 나타내는 특성 정보로서, 예를 들어 도 2에 예시한 바와 같이 총 4개의 2차원 좌표로 표시될 수 있다.
- [0029] 도 2에 도시한 바와 같이, 본 발명의 실시예에 따른 전력 제한 컴퓨팅 시스템(1)의 성능 관리를 위한 시스템 상태 특성 정보는 x축과 y축에 전력 할당 비율(CPU 및 메모리에 대한 CPU의 전력 할당 비율) 및 동시성 레벨(코어 개수)을 각각 설정하고, 이러한 x축 및 y축에 대한 2차원 좌표를 확장성(scalable)-코어 집중적(core-intensive) 시스템 상태(A), 확장성-메모리 집중적(memory intensive) 시스템 상태(B), 비확장성(non-scalable)-코어 집중적 시스템 상태(C) 및 비확장성-메모리 집중적 시스템 상태(D)로 각각 분류하였다.
- [0030] 확장성-코어 집중적 시스템 상태(A)는, 예를 들어 x축의 전력 할당 비율이 0.8%이고 y축의 동시성 레벨이 16코어인 경우의 2차원 좌표일 수 있다. 또한, 확장성-메모리 집중적 시스템 상태(B)는, 예를 들어 x축의 전력 할당 비율이 0.67%이고 y축의 동시성 레벨이 16코어인 경우의 2차원 좌표일 수 있다. 또한, 비확장성-코어 집중적 시스템 상태(C)는, 예를 들어 x축의 전력 할당 비율이 0.8%이고 y축의 동시성 레벨이 8코어인 경우의 2차원 좌표일 수 있다. 또한, 비확장성-메모리 집중적 시스템 상태(D)는, 예를 들어 x축의 전력 할당 비율이 0.67%이고 y축의 동시성 레벨이 8코어인 경우의 2차원 좌표일 수 있다.
- [0031] 시스템 성능 관리부(104)는 이러한 동시성 레벨과 전력 할당 비율로 정의되는 시스템 상태를 나타내는 2차원 좌표, 즉 시스템 상태 특성 정보를 기초로, 컴퓨팅 시스템에서 실행되는 응용 프로그램의 성능이 최대가 되는 2차원 좌표와 인접할 확률이 최대가 되는 좌표를 탐색 시작 지점으로 설정하고, 설정된 탐색 시작 지점을 중심으로 응용 프로그램의 성능이 최대가 되는 2차원 좌표(동시성 레벨과 전력 할당 비율로 정의되는 시스템 상태)를 찾기 위해 2차원 좌표 공간(시스템 상태 공간)을 탐색할 수 있다. 이러한 시스템 성능 관리부(104)가 시스템 상태 공간의 탐색 시작 지점을 설정하는 과정과 시스템 상태를 구체적으로 탐색하는 과정은 하기에 기술하는 도 3 및 도 4의 흐름도에서 보다 상세히 다루기로 한다.
- [0032] 저장부(106)는 시스템 성능 관리부(104)를 통해 생성되는 전력 제한 컴퓨팅 시스템(1)의 동시성 레벨과 전력 할당 비율로 정의되는 시스템 상태를 저장할 수 있다. 저장부(106)에 저장되는 시스템 상태 특성 정보는 시스템 성능 관리부(104)에 의해 취사 선택될 수 있다. 이러한 저장부(106)는, 예를 들어 캐시 메모리(Cache memory)를 포함할 수 있다. 다만, 도 1에는 저장부(106)가 성능 관리 장치(100) 내에 포함되는 경우를 예시한 것이며, 필요에 따라 저장부(106)를 성능 관리 장치(100)와는 별도의 기능 블록으로 표현할 수도 있다. 이 경우, 저장부(106)는, 예를 들어 SSD(Solid State Drive)와 같은 플래시 메모리(Flash memory) 기반의 기록 매체, 하드 디스크 드라이브(Hard Disc Drive)와 같은 자기 디스크 기반의 기록 매체 등을 포함할 수 있다.
- [0033] 응용 프로그램 실행부(108)는 후술하는 사용자 인터페이스부(10)의 실행 명령, 또는 성능 관리 장치(100)의 실행 명령에 따라 임의의 응용 프로그램을 실행시킬 수 있다. 본 발명의 실시예에 따라 실행되는 응용 프로그램은, 예를 들어 멀티-스레드(multi-threaded) 응용 프로그램이 적용될 수 있다.
- [0034] 상술한 바와 같은 시스템 자원 분석부(102), 시스템 성능 관리부(104), 응용 프로그램 실행부(108)(또는 저장부(106))는 각각 별개의 기능 블록으로 구분하였으나 이는 실시예일뿐이며, 성능 관리 장치(100) 내의 제어 수단, 예를 들어 마이크로프로세서(microprocessor)에 의해 이들 시스템 자원 분석, 시스템 성능 관리, 응용 프로그램 실행(또는 상관 관계 특성 정보 저장 및 독출) 등의 기능이 수행되도록 구현할 수도 있다.

- [0035] 사용자 인터페이스부(10)는 컴퓨팅 시스템의 성능 관리를 위한 응용 프로그램 실행 명령과 같은 입력 신호를 성능 관리 장치(100)로 제공하거나, 도 2에 도시한 바와 같은 동시성 레벨과 전력 할당 비율 간의 상관 관계를 나타내는 2차원 좌표를 외부로 표시할 수 있다. 이때, 이러한 2차원 좌표는 컴퓨팅 시스템(1)의 성능 테스트 결과를 사용자가 확인하기 위한 시뮬레이션 정보로서 제공될 수 있으며, 본 발명의 실시예에 따라 응용 프로그램의 성능을 측정하거나 2차원 좌표의 탐색 시작 지점을 설정하거나 2차원 좌표를 상세히 탐색하는 과정에서는 표시되지 않을 수 있다. 이러한 사용자 인터페이스부(10)는, 예를 들어 터치 입력이나 키 입력이 가능한 터치 모니터 또는 터치 패드 등을 포함할 수 있다.
- [0036] 전원부(20)는 컴퓨팅 시스템(1)으로 전력을 공급할 수 있다. 이때의 전력은 전력 설계 또는 전력 정책에 의해 제한된 전력일 수 있다.
- [0037] 이하, 상술한 구성들과 함께, 본 발명의 실시예에 따른 컴퓨팅 시스템의 성능 관리 방법을 첨부하는 도 3 및 도 4의 흐름도를 참조하여 보다 구체적으로 설명하기로 한다.
- [0038] 먼저, 도 3은 본 발명의 실시예에 따른 컴퓨팅 시스템의 성능 관리 방법을 나타낸 흐름도이다.
- [0039] 도 3에 도시한 바와 같이, 시스템 성능 관리부(104)는 시스템 자원 분석부(102)를 통해 분석된 컴퓨팅 시스템(1)의 시스템 자원 분석 결과를 입력받을 수 있으며(S100), 입력된 시스템 자원 분석 결과를 기초로 컴퓨팅 시스템(1)의 동시성 레벨과 전력 할당 비율로 정의되는 시스템 상태를 나타내는 적어도 둘 이상의 2차원 좌표를 생성할 수 있다(S102).
- [0040] 여기서, 적어도 둘 이상의 2차원 좌표는, 예를 들어 도 2에 예시한 바와 같은 확장성-코어 집중적 특성을 갖는 시스템 상태(A), 확장성-메모리 집중적 특성을 갖는 시스템 상태(B), 비확장성-코어 집중적 특성을 갖는 시스템 상태(C) 및 비확장성-메모리 집중적 특성을 갖는 시스템 상태(D)로 각각 분류될 수 있다. 구체적으로, 컴퓨팅 시스템(1)의 CPU 및 메모리에 대한 CPU의 전력 할당 비율에 따라 전력 할당 비율을 구분한 좌표 성분을 2차원 좌표축의 x축 상에 생성하고, 컴퓨팅 시스템(1)의 코어 개수에 따라 동시성 레벨을 구분한 좌표 성분을 2차원 좌표축의 y축 상에 생성하며, x축 상의 좌표 성분과 y축 상의 좌표 성분을 조합하여 시스템 상태를 나타내는 적어도 둘 이상의 2차원 좌표를 생성할 수 있다.
- [0041] 이후, 시스템 성능 관리부(104)는 응용 프로그램 실행부(108)를 통해 임의의 응용 프로그램이 실행될 때 해당 응용 프로그램에 대해서 상기 적어도 둘 이상의 2차원 좌표 각각에 대응하는 동시성 레벨과 전력 할당 비율에서의 성능 및 LLC MPKI를 측정할 수 있다(S104). 예를 들어, 시스템 성능 관리부(104)는 확장성-코어 집중적 특성을 갖는 시스템 상태(A)에 대응하는 2차원 좌표에서의 전력 할당 비율(0.8%)과 동시성 레벨(16코어)로 해당 응용 프로그램의 실행 성능을 측정하고, 확장성-메모리 집중적 특성을 갖는 시스템 상태(B)에 대응하는 2차원 좌표에서의 전력 할당 비율(0.67%)과 동시성 레벨(16코어)로 해당 응용 프로그램의 실행 성능을 측정하며, 비확장성-코어 집중적 특성을 갖는 시스템 상태(C)에 대응하는 2차원 좌표에서의 전력 할당 비율(0.8%)과 동시성 레벨(8코어)로 해당 응용 프로그램의 실행 성능을 측정하고, 비확장성-메모리 집중적 특성을 갖는 시스템 상태(D)에 대응하는 2차원 좌표에서의 전력 할당 비율(0.67%)과 동시성 레벨(8코어)로 해당 응용 프로그램의 실행 성능을 각각 측정할 수 있다.
- [0042] 이후, 시스템 성능 관리부(104)는 이러한 각각의 실행 성능 측정 결과를 기초로 성능이 최대가 되는 2차원 좌표와 인접할 확률이 최대가 되는 좌표를 탐색 시작 지점으로 설정할 수 있다(S106). 예를 들어, 확장성-코어 집중적 특성을 갖는 시스템 상태(A)에 대응하는 2차원 좌표에서 해당 응용 프로그램의 동시성 레벨과 전력 할당 비율에 따른 성능이 최대로 판정된 경우, 확장성-코어 집중적 특성을 갖는 시스템 상태(A)에 대응하는 2차원 좌표를 탐색 시작 지점으로 설정할 수 있다. 구체적으로, 시스템 성능 관리부(104)는 응용 프로그램의 성능에 따라 동시성 레벨의 확장성 또는 비확장성 여부를 결정하고, 확장성 또는 비확장성 여부가 결정되면 측정된 LLC MPKI에 따라 전력 할당 비율의 코어 집중적 또는 메모리 집중적 여부를 결정하는 과정을 포함할 수 있다. LLC MPKI는 대략 1,000개의 명령(instruction)을 수행하면서 Last-level cache(LLC)에 발생한 미스(miss) 횟수를 의미한다. LLC는 프로세서에 부착된 메모리의 일종으로 main memory(예를 들어, DDR4, RAM 등) 보다 빠르지만 용량이 작다. 메모리 집중적인 응용 프로그램의 경우 메모리 사용량이 많고 메모리에 자주 접근해야 하기 때문에 LLC MPKI 수치가 높게 나오고, 코어 집중적인 응용 프로그램의 경우에는 그렇지 않기 때문에 LLC MPKI가 낮게 나온다. 본 발명의 실시예에서는 여러 벤치마크들을 사용한 오프라인 실험을 통해 메모리 집중적인 경우와 코어 집중적인 경우를 구분하는 LLC MPKI 기준을, 예를 들어 10으로 가정하였다. 즉, 1,000개의 명령어당 LLC 미스가 10회 이상 발생하면 메모리 집중적인 것으로, 그렇지 않으면 코어-집중적인 것으로 판단할 수 있다.

- [0043] 이때, 본 발명의 실시예에 따라 측정되는 응용 프로그램의 성능이라 함은, 예를 들어 해당 응용 프로그램이 임의의 전력 할당 비율과 동시성 레벨로 정의되는 시스템 상태에서 실행될 때의 단위 작업 당 수행 시간의 역수일 수 있다. 예컨대, 확장성-코어 집중적 특성을 갖는 시스템 상태(A)에서의 응용 프로그램의 단위 작업 당 수행 시간이 2초라고 가정하면, 응용 프로그램의 성능은 $1/2$ 인 0.5로 표현될 수 있다.
- [0044] 이후, 시스템 성능 관리부(104)는 단계(S106)에서 설정된 탐색 시작 지점을 중심으로 응용 프로그램의 실행 성능이 최대가 되는 2차원 좌표(동시성 레벨과 전력 할당 비율로 정의되는 구체적인 시스템 상태)를 탐색할 수 있다(S108). 예를 들어, 탐색 시작 지점이 시스템 상태(A)로 설정되었다면, 시스템 성능 관리부(104)는 시스템 상태(A)를 나타내는 좌표를 중심으로 도 2의 x축 상의 좌표 성분 또는 y축 상의 좌표 성분을 기 설정된 단위로 증감시키면서 응용 프로그램의 성능이 최대가 되는 좌표 성분을 결정할 수 있다.
- [0045] 이러한 2차원 좌표 탐색 과정은 도 4에 보다 구체적으로 도시되어 있다.
- [0046] 도 4에 도시한 바와 같이, 예를 들어 시스템 상태(A)를 나타내는 좌표를 중심으로 x축(전력 할당 비율) 상에서 기 설정된 단위로 좌표 성분을 양(+)의 방향으로 증가시킬 수 있다. 이때의 기 설정된 단위는, 예를 들어 전체 전력 설계량의 10% 단위일 수 있으며, 필요에 따라 가변될 수 있다.
- [0047] 이후, 시스템 성능 관리부(104)는 기 설정된 단위로 증가된 좌표 성분에서의 응용 프로그램의 성능을 측정할 수 있다(S202). 응용 프로그램의 성능은 앞서 설명한 바와 같이 응용 프로그램의 단위 작업 당 수행 시간의 역수로 나타낼 수 있다.
- [0048] 기 설정된 단위로 증가된 좌표 성분에서의 응용 프로그램 성능이 측정되면, 시스템 성능 관리부(104)는 증가된 좌표 성분에서의 응용 프로그램 성능과 이전 좌표 성분, 즉 탐색 시작 지점인 시스템 상태(A)에서의 좌표 성분에서의 응용 프로그램 성능을 비교하고(S204), 증가된 좌표 성분에서의 응용 프로그램 성능이 이전 좌표 성분에서의 응용 프로그램 성능보다 좋으면, 즉, 증가된 좌표 성분에서의 응용 프로그램의 수행 시간이 이전 좌표 성분에서의 응용 프로그램의 수행 시간보다 짧으면, 단계(S200)로 피드백하여 상기의 과정들(S202, S204)을 반복 수행할 수 있다.
- [0049] 반면, 증가된 좌표 성분에서의 응용 프로그램의 성능이 이전 좌표 성분에서의 응용 프로그램 성능보다 떨어진다면, 즉, 증가된 좌표 성분에서의 응용 프로그램의 수행 시간이 이전 좌표 성분에서의 응용 프로그램의 수행 시간보다 길다면, 시스템 성능 관리부(104)는 이전 좌표 성분으로 탐색 지점을 변경한 후(S206), 기 설정된 단위로 좌표 성분을 감소시킬 수 있다(S208). 예를 들어, 이전 좌표 성분으로부터 x축(전력 할당 비율) 상에서 기 설정된 단위로 좌표 성분을 음(-)의 방향으로 감소시킬 수 있다.
- [0050] 좌표 성분이 음의 방향으로 감소되면, 시스템 성능 관리부(104)는 해당 좌표 성분에서의 응용 프로그램 성능을 측정할 수 있다(S210).
- [0051] 기 설정된 단위로 감소된 좌표 성분에서의 응용 프로그램 성능이 측정되면, 시스템 성능 관리부(104)는 감소된 좌표 성분에서의 응용 프로그램 성능과 이전 좌표 성분에서의 응용 프로그램 성능을 비교하고(S212), 증가된 좌표 성분에서의 응용 프로그램 성능이 이전 좌표 성분에서의 응용 프로그램 성능보다 좋으면, 즉, 증가된 좌표 성분에서의 응용 프로그램의 수행 시간이 이전 좌표 성분에서의 응용 프로그램의 수행 시간보다 짧으면, 단계(S208)로 피드백하여 상기의 과정들(S210, S212)을 반복 수행할 수 있다.
- [0052] 반면, 감소된 좌표 성분에서의 응용 프로그램의 성능이 이전 좌표 성분에서의 응용 프로그램 성능보다 떨어진다면, 즉, 증가된 좌표 성분에서의 응용 프로그램의 수행 시간이 이전 좌표 성분에서의 응용 프로그램의 수행 시간보다 길다면, 시스템 성능 관리부(104)는 도 2의 x축 및 y축에 대한 성능 측정이 완료되었는지를 판단할 수 있다(S214).
- [0053] x축 및 y축에 대한 성능 측정이 완료되지 않은 경우, 시스템 성능 관리부(104)는 단계(S200)로 피드백할 수 있다. 이때의 단계(S200) 수행은, 예를 들어 상관 관계(A)를 나타내는 좌표를 중심으로 y축(동시성 레벨) 상에서 기 설정된 단위로 좌표 성분을 양(+)의 방향으로 증가시킬 수 있다. 이때의 기 설정된 단위는, 예를 들어 전체 코어 개수의 25% 단위일 수 있으며, 필요에 따라 가변될 수 있다.
- [0054] 이후의 과정은 단계(S200) 내지 단계(S214)까지의 과정과 동일하므로 중복되는 설명은 생략하기로 한다.
- [0055] 한편, x축 및 y축에 대한 성능 측정이 완료되었다면, 시스템 성능 관리부(104)는 이전 좌표 성분을 갖는 2차원 좌표(x축 좌표 성분 및 y축 좌표 성분)를 탐색 결과로 결정할 수 있다(S216). 예를 들어, 시스템 상태(A)의 2차원 좌표가 이전 좌표 성분을 갖는 2차원 좌표라면, 시스템 상태(A)의 2차원 좌표를 탐색 결과로 결정할 수 있

으며, 시스템 상태(A)의 2차원 좌표에서 기 설정 단위로 증가된 좌표 성분을 갖는 2차원 좌표가 이전 좌표 성분을 갖는 2차원 좌표라면, 증가된 좌표 성분을 갖는 2차원 좌표를 탐색 결과로 결정할 수 있다.

[0056] 본 발명의 실시예에서, 상기 탐색하는 단계(S108)는, 예를 들어 볼록 함수(convex function) 또는 오목 함수(concave function)의 특성을 이용한 탐색 알고리즘, 예를 들어 *hill-climbing* 알고리즘으로 2차원 좌표를 탐색하는 단계를 포함할 수 있으며, 이와 같은 탐색 알고리즘은 본 발명의 기술분야에서 통상의 지식을 가진 자라면 용이하게 알 수 있는 바, 구체적인 설명은 생략하기로 한다.

[0057] 단계(S216)에서와 같이 탐색 결과가 결정되면, 도 3의 단계(S110)와 같이 탐색 결과를 기초로 동시성 레벨과 전력 할당 비율을 설정할 수 있다. 예를 들어, 탐색 결과가 상관 관계(A)의 2차원 좌표에 해당한다면, 0.8%의 전력 할당 비율과 16코어의 동시성 레벨로 시스템 상태를 설정할 수 있으며, 탐색 결과가 시스템 상태(A)의 2차원 좌표에서 기 설정된 단위로 감소된 2차원 좌표에 해당한다면, 그에 해당하는 전력 할당 비율과 코어 개수의 동시성 레벨로 시스템 상태를 설정할 수 있다.

[0058] 이상 설명한 바와 같이 본 발명의 실시예에 의하면, 전력 제한 컴퓨팅 환경에서 동시성 레벨과 전력 할당 비율을 동시에 적응적으로 조절하여 응용 프로그램 별 시스템 성능을 최적화하도록 하며, 구체적으로, 전력 제한 컴퓨팅 환경에서 수행되는 응용 프로그램 별로 CPU의 코어 개수와, CPU 및 메모리에 대한 CPU의 전력 할당 비율을 동시에 적응적으로 설정하여 컴퓨팅 수행 시간을 최소화시키면서 동적으로 성능을 관리할 수 있는 컴퓨팅 시스템의 성능 관리 기술을 구현하고자 한다.

[0059] 한편, 첨부된 블록도의 각 블록과 흐름도의 각 단계의 조합들은 컴퓨터 프로그램 인스트럭션들에 의해 수행될 수도 있다. 이들 컴퓨터 프로그램 인스트럭션들은 범용 컴퓨터, 특수용 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서에 탑재될 수 있으므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서를 통해 수행되는 그 인스트럭션들이 블록도의 각 블록에서 설명된 기능들을 수행하는 수단을 생성하게 된다.

[0060] 이들 컴퓨터 프로그램 인스트럭션들은 특정 방식으로 기능을 구현하기 위해 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 지향할 수 있는 컴퓨터 이용 가능 또는 컴퓨터 판독 가능 기록매체(또는 메모리) 등에 저장되는 것도 가능하므로, 그 컴퓨터 이용 가능 또는 컴퓨터 판독 가능 기록매체(또는 메모리)에 저장된 인스트럭션들은 블록도의 각 블록에서 설명된 기능을 수행하는 인스트럭션 수단을 내포하는 제조 품목을 생산하는 것도 가능하다.

[0061] 그리고, 컴퓨터 프로그램 인스트럭션들은 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에 탑재되는 것도 가능하므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에서 일련의 동작 단계들이 수행되어 컴퓨터로 실행되는 프로세스를 생성해서 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 수행하는 인스트럭션들은 블록도의 각 블록에서 설명된 기능들을 실행하기 위한 단계들을 제공하는 것도 가능하다.

[0062] 또한, 각 블록은 특정된 논리적 기능(들)을 실행하기 위한 적어도 하나 이상의 실행 가능한 인스트럭션들을 포함하는 모듈, 세그먼트 또는 코드의 일부를 나타낼 수 있다. 또, 몇 가지 대체 실시 예들에서는 블록들에서 언급된 기능들이 순서를 벗어나서 발생하는 것도 가능함을 주목해야 한다. 예컨대, 잇달아 도시되어 있는 두 개의 블록들은 사실 실질적으로 동시에 수행되는 것도 가능하고 또는 그 블록들이 때때로 해당하는 기능에 따라 역순으로 수행되는 것도 가능하다.

부호의 설명

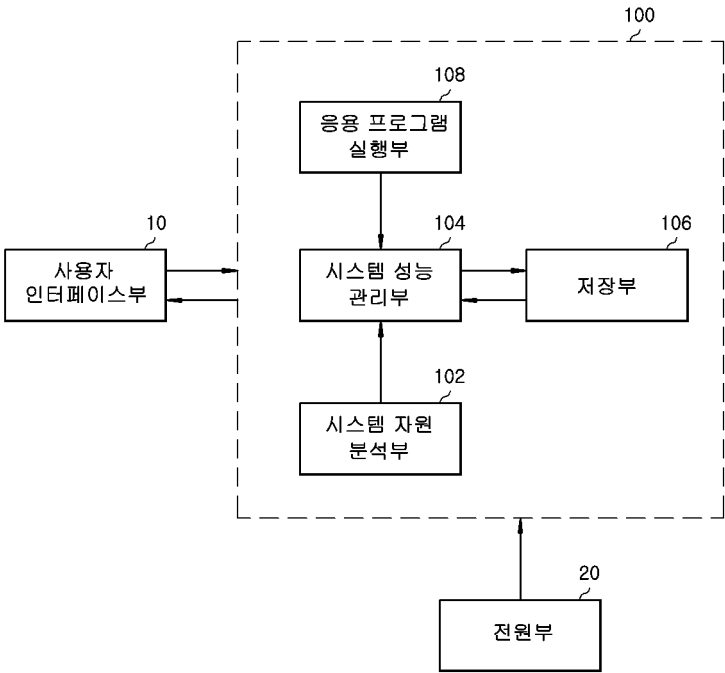
- [0063]
- 1: 전력 제한 컴퓨팅 시스템
 - 10: 사용자 인터페이스부
 - 20: 전원부
 - 100: 시스템 성능 관리 장치
 - 102: 시스템 자원 분석부
 - 104: 시스템 성능 관리부
 - 106: 저장부

108: 응용 프로그램 실행부

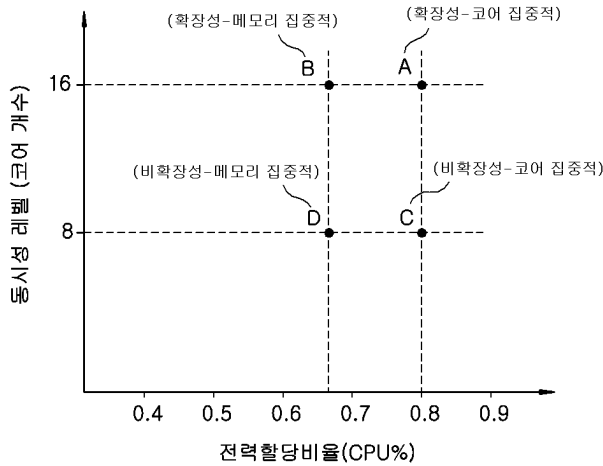
도면

도면1

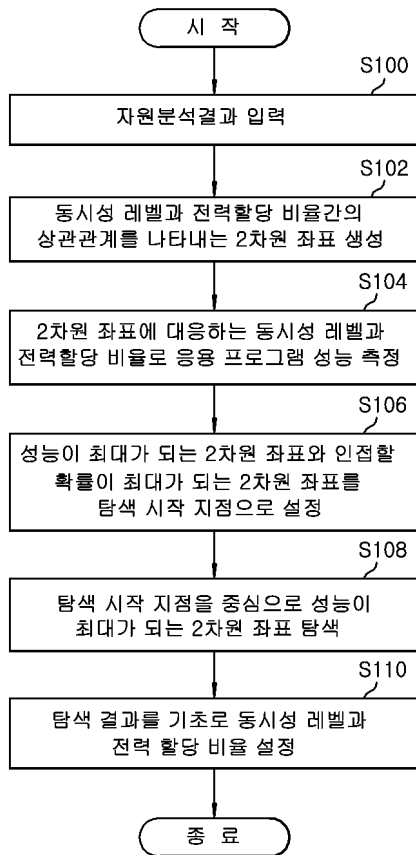
1



도면2



도면3



도면4

