(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2012년04월19일
(11) 등록번호 10-1137073
(24) 등록일자 2012년04월09일

(51) 국제특허분류(Int. Cl.)
G06F 1/32 (2006.01) *G06F 9/48* (2006.01)
G06F 9/38 (2006.01)
(21) 출원번호 10-2010-7003117
(22) 출원일자(국제) 2008년08월05일
심사청구일자 2010년03월12일
(85) 번역문제출일자 2010년02월11일
(65) 공개번호 10-2010-0041829
(43) 공개일자 2010년04월22일
(86) 국제출원번호 PCT/EP2008/060273
(87) 국제공개번호 WO 2009/024459
국제공개일자 2009년02월26일
(30) 우선권주장
11/840,711 2007년08월17일 미국(US)
(56) 선행기술조사문헌
EP1715405 A
US20060288241 A1
US5717319 A
JP2000155626 A

(73) 특허권자
인터내셔널 비지네스 머신즈 코포레이션
미국 10504 뉴욕주 아몬크 뉴오차드 로드
(72) 발명자
구딩 토마스
미국 미네소타주 55901 로체스터 22엔디 애비뉴 엔더블유 5715
잉글렛 토드
미국 미네소타주 55904 로체스터 하이랜드 릿지 래인 에스이 5465
(뒷면에 계속)
(74) 대리인
허정훈, 윤여원

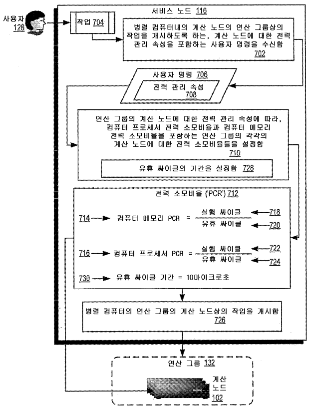
전체 청구항 수 : 총 10 항 심사관 : 박지은

(54) 발명의 명칭 **병렬 컴퓨터에서의 사전적 전력 관리**

(57) 요 약

병렬 컴퓨터에서의 사전적 전력 관리가 개시되며, 병렬 컴퓨터는 서비스 노드와 복수의 계산 노드를 포함하고, 서비스 노드는 대역외 서비스 네트워크를 통해 계산 노드에 연결되고, 각각의 계산 노드는 컴퓨터 프로세서와 이 컴퓨터 프로세서에 동작가능하게 결합된 컴퓨터 메모리를 포함한다. 실시예들은, 서비스 노드에 의해, 병렬 컴퓨터내의 계산 노드의 연산 그룹상의 작업을 개시하도록 하는 사용자 명령 - 이 사용자 명령은 계산 노드에 대한 전력 관리 속성을 포함함 - 을 수신하고; 연산 그룹의 계산 노드에 대한 전력 관리 속성에 따라 서비스 노드에 의해, 컴퓨터 프로세서 전력 소모비율과 컴퓨터 메모리 전력 소모비율을 포함하는 연산 그룹의 각각의 계산 노드에 대한 전력 소모비율을 설정하며; 서비스 노드에 의해, 병렬 컴퓨터의 연산 그룹의 계산 노드상의 작업을 개시하는 것을 포함한다.

대 표 도 - 도7



(72) 발명자
리에브취 토마스
미국 사우쓰다코타주 57212 알링톤 노쓰베이 드라이브 100
무스타 토마스
미국 미네소타주 55906 로체스터 스캔란 래인 엔이 3075
리드 돈
미국 미네소타주 55955 맨토빌 234티에이치 애비뉴 61329

특허청구의 범위

청구항 1

병렬 컴퓨터에서의 사전적 전력 관리 방법으로서, 상기 병렬 컴퓨터는 서비스 노드와 복수의 계산 노드들을 포함하고, 상기 서비스 노드는 대역외 서비스 네트워크를 통해 상기 계산 노드들에 연결되고, 상기 계산 노드들 각각은 컴퓨터 프로세서와 이 컴퓨터 프로세서에 동작가능하게 결합된 컴퓨터 메모리를 포함하는 것인, 상기 병렬 컴퓨터에서의 사전적 전력 관리 방법에 있어서,

상기 서비스 노드에 의해, 상기 병렬 컴퓨터내의 계산 노드들의 연산 그룹상의 작업을 개시하도록 하는 사용자 명령 - 상기 사용자 명령은 상기 계산 노드들에 대한 전력 관리 속성을 포함함 - 을 수신하고;

상기 연산 그룹의 계산 노드들에 대한 상기 전력 관리 속성에 따라 상기 서비스 노드에 의해, 컴퓨터 프로세서 전력 소모비율과 컴퓨터 메모리 전력 소모비율을 포함하는 상기 연산 그룹의 계산 노드들 각각에 대한 전력 소모비율들을 설정하며;

상기 서비스 노드에 의해, 상기 병렬 컴퓨터의 연산 그룹의 계산 노드들상의 작업을 개시하는 것

을 포함하는, 병렬 컴퓨터에서의 사전적 전력 관리 방법.

청구항 2

제1항에 있어서, 상기 컴퓨터 프로세서 전력 소모비율과 상기 컴퓨터 메모리 전력 소모비율은 실행 싸이클 대 유휴 싸이클의 비율을 더 포함하는 것인, 병렬 컴퓨터에서의 사전적 전력 관리 방법.

청구항 3

제2항에 있어서, 상기 전력 소모비율들을 설정하는 것은 상기 유휴 싸이클의 기간을 설정하는 것을 더 포함하는 것인, 병렬 컴퓨터에서의 사전적 전력 관리 방법.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 컴퓨터 프로세서 전력 소모비율은 상기 컴퓨터 메모리 전력 소모비율과 동일한 것인, 병렬 컴퓨터에서의 사전적 전력 관리 방법.

청구항 5

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 컴퓨터 프로세서 전력 소모비율은 상기 컴퓨터 메모리 전력 소모비율과 동일하지 않는 것인, 병렬 컴퓨터에서의 사전적 전력 관리 방법.

청구항 6

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 사용자에 의해, 적어도 하나의 이전 작업 실행 동안에 상기 병렬 컴퓨터의 연산 그룹의 계산 노드들의 전력 소모를 모니터링하는 것을 포함하는 상기 전력 관리 속성을 설정하는 것을 더 포함하는, 병렬 컴퓨터에서의 사전적 전력 관리 방법.

청구항 7

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 복수의 계산 노드들은 복수의 데이터 통신 네트워크들을 통해서 데이터 통신을 위해 연결되며, 상기 데이터 통신 네트워크들 중 적어도 하나의 데이터 통신 네트워크는 점대점 데이터 통신에 대해 최적화되어 있고, 상기 데이터 통신 네트워크들 중 적어도 하나의 데이터 통신 네트워크는 일괄적 연산에 대해 최적화되어 있는 것인, 병렬 컴퓨터에서의 사전적 전력 관리 방법.

청구항 8

병렬 컴퓨터에서의 사전적 전력 관리를 위한 서비스 노드에 있어서, 상기 병렬 컴퓨터는 서비스 노드와 복수의 계산 노드들을 포함하고, 상기 서비스 노드는 대역외 서비스 네트워크를 통해 상기 계산 노드들에 연결되고, 상기 계산 노드들 각각은 컴퓨터 프로세서와 이 컴퓨터 프로세서에 동작가능하게 결합된 컴퓨터 메모리를 포함하

며, 상기 서비스 노드는 컴퓨터 프로세서와 이 컴퓨터 프로세서에 동작가능하게 결합된 컴퓨터 메모리를 포함하며, 상기 서비스 노드가 포함하는 컴퓨터 메모리내에는,

상기 서비스 노드에 의해, 상기 병렬 컴퓨터내의 계산 노드들의 연산 그룹상의 작업을 개시하도록 하는 사용자 명령 - 상기 사용자 명령은 상기 계산 노드들에 대한 전력 관리 속성을 포함함 - 을 수신하고;

상기 연산 그룹의 계산 노드들에 대한 상기 전력 관리 속성에 따라 상기 서비스 노드에 의해, 컴퓨터 프로세서 전력 소모비율과 컴퓨터 메모리 전력 소모비율을 포함하는 상기 연산 그룹의 계산 노드들 각각에 대한 전력 소모비율들을 설정하며;

상기 서비스 노드에 의해, 상기 병렬 컴퓨터의 연산 그룹의 계산 노드들상의 작업을 개시하는 것

이 가능한 컴퓨터 프로그램 명령이 배치되는 것인, 서비스 노드.

청구항 9

제8항에 있어서, 상기 컴퓨터 프로세서 전력 소모비율과 상기 컴퓨터 메모리 전력 소모비율은 실행 싸이클 대 유휴 싸이클의 비율을 더 포함하는 것인, 서비스 노드.

청구항 10

병렬 컴퓨터에서의 사전적 전력 관리를 위한 컴퓨터 프로그램이 기록된 컴퓨터 판독가능 기록매체에 있어서, 상기 병렬 컴퓨터는 서비스 노드와 복수의 계산 노드들을 포함하고, 상기 서비스 노드는 대역외 서비스 네트워크를 통해 상기 계산 노드들에 연결되고, 상기 계산 노드들 각각은 컴퓨터 프로세서와 이 컴퓨터 프로세서에 동작가능하게 결합된 컴퓨터 메모리를 포함하며, 상기 컴퓨터 프로그램은,

상기 서비스 노드에 의해, 상기 병렬 컴퓨터내의 계산 노드들의 연산 그룹상의 작업을 개시하도록 하는 사용자 명령 - 상기 사용자 명령은 상기 계산 노드들에 대한 전력 관리 속성을 포함함 - 을 수신하고;

상기 연산 그룹의 계산 노드들에 대한 상기 전력 관리 속성에 따라 상기 서비스 노드에 의해, 컴퓨터 프로세서 전력 소모비율과 컴퓨터 메모리 전력 소모비율을 포함하는 상기 연산 그룹의 계산 노드들 각각에 대한 전력 소모비율들을 설정하며;

상기 서비스 노드에 의해, 상기 병렬 컴퓨터의 연산 그룹의 계산 노드들상의 작업을 개시하는 것

이 가능한 컴퓨터 프로그램 명령을 포함하는 것인, 컴퓨터 판독가능 기록매체.

명 세 서

기 술 분 야

[0001] 본 발명분야는 데이터 프로세싱이며, 또는 보다 구체적으로는, 병렬 컴퓨터에서의 사전적 전력 관리를 위한 방법, 장치, 및 제품이다.

배 경 기 술

[0002] 1948년도의 EDVAC(Electronic Discrete Variable Automatic Computer) 컴퓨터 시스템의 개발은 종종 컴퓨터 시대의 시초로서 언급되어지고 있다. 그 때 이후로, 컴퓨터 시스템은 매우 복잡한 장치로 진화되었다. 오늘날의 컴퓨터는 EDVAC과 같은 초창기때 시스템보다도 훨씬 정교해졌다. 컴퓨터 시스템은 일반적으로 하드웨어 및 소프트웨어 컴포넌트의 조합, 애플리케이션 프로그램, 오퍼레이팅 시스템, 프로세서, 버스, 메모리, 입력/출력 장치 등을 포함한다. 반도체 프로세싱 및 컴퓨터 구조의 발전이 컴퓨터의 성능을 점점 드높여갈수록, 컴퓨터 소프트웨어는 이러한 하드웨어의 보다 높아진 성능을 이용하도록 보다 정교해지게끔 진화되어가고, 이로써 오늘날의 컴퓨터 시스템은 불과 몇년 전보다도 훨씬 강력해졌다.

[0003] 병렬 컴퓨팅은 발전을 경험해온 컴퓨터 기술의 영역이다. 병렬 컴퓨팅이란 보다 신속하게 결과를 획득하기 위해 다수의 프로세서들을 통해 (분할되고 특별히 개조된) 동일한 업무를 동시적으로 실행하는 것을 말한다. 보통, 병렬 컴퓨팅은 과제를 해결하는 프로세스가 보다 작은 업무들로 분할될 수 있고, 이러한 보다 작은 업무들은 어느 정도 조화를 이루면서 동시적으로 수행될 수 있다는 사실에 기초한다.

[0004] 병렬 컴퓨터는 병렬 알고리즘을 실행한다. 병렬 알고리즘은 수 많은 상이한 프로세싱 장치들을 통해 한번에 한

조각씩 실행되도록 분할되고, 그런 다음 데이터 프로세싱 결과를 얻기 위해 최종적으로 다시 재결합될 수 있다. 몇몇 알고리즘들은 쉽게 조각으로 분할된다. 어느 숫자가 최상위인지를 알아보기 위해 일에서 십만까지의 모든 숫자들을 체크하는 업무를 분할하는 것은 예컨대 숫자들의 서브세트를 각각의 이용가능한 프로세서에 할당하고, 그런 다음 긍정적 결과 리스트를 재결합시킴으로써 행해질 수 있다. 본 명세서에서, 개별적인 병렬 프로그램 조각들을 실행하는 다수의 프로세싱 장치들을 '계산 노드(compute node)'라고 칭한다. 병렬 컴퓨터는 계산 노드뿐만이 아니라, 예컨대 입력/출력('I/O') 노드 및 서비스 노드를 포함하는 기타의 프로세싱 노드로 구성된다.

[0005] 현대의 프로세서들이 동작하는 방식으로 인해, 직렬(비병렬) 알고리즘을 통해서 수행하는 것보다 병렬 알고리즘을 통해 몇몇 종류의 방대한 컴퓨팅 업무들을 수행하는 것이 빠르기 때문에 병렬 알고리즘은 매우 유용하다. 동일한 처리량을 갖는 수 많은 저속 프로세서들을 갖는 컴퓨터보다는 하나의 고속 프로세서를 갖는 컴퓨터를 구축하는 것이 훨씬 더 어렵다. 또한 직렬 프로세서들의 잠재적 속도에 대해서는 일정한 이론적 한계가 존재한다. 한편, 모든 병렬 알고리즘들은 직렬 부분을 가지고 있으며, 그래서 병렬 알고리즘들은 포화점을 갖는다. 이 포화점 이후 프로세서들을 더 추가시키는 것은 더 이상의 처리량을 산출시키지는 않고 오버헤드와 비용을 증가시키기만 한다.

[0006] 병렬 알고리즘들은 또한 병렬 컴퓨터의 노드들간의 데이터 통신 요건인 하나 이상의 자원을 최적화하도록 설계된다. 병렬 프로세서들이 통신하는 방법에는 두 가지 방법들, 즉 메모리 공유 또는 메세지 전달 프로세싱이 있다. 메모리 공유 프로세싱은 데이터에 대한 추가적인 로킹(locking)을 필요로 하고, 추가적인 프로세서 및 버스 싸이클의 오버헤드를 가져다주며, 또한 알고리즘의 일부분을 직렬화시킨다.

[0007] 메세지 전달 프로세싱은 고속 데이터 통신 네트워크와 메세지 버퍼를 이용하지만, 이 통신은 데이터 통신 네트워크상의 전송 오버헤드를 추가시키는 것 뿐만이 아니라 노드들간의 데이터 통신에서의 메세지 버퍼 및 레이턴시를 위한 추가적인 메모리 필요성을 추가시킨다. 통신 오버헤드는 작아질 것이지만 트래픽량을 결정하는 것은 병렬 알고리즘이 되도록 병렬 컴퓨터의 설계는 특수하게 설계된 데이터 통신 링크를 이용한다.

[0008] 병렬 컴퓨터내의 노드들간의 메세지 전달을 위해 수 많은 데이터 통신 네트워크 구조가 이용된다. 계산 노드는 네트워크내에서 예컨대 '원환체(torus)' 또는 '메시(mesh)'로서 편성될 수 있다. 또한, 계산 노드는 네트워크내에서 트리(tree)로서 편성될 수 있다. 원환체 네트워크는 삼차원 메시내의 노드들을 랩 어라운드 링크(wrap around link)로 연결시켜준다. 모든 노드는 이러한 원환체 네트워크를 통해 자신의 여섯 개 이웃들에 연결되고, 각각의 노드들은 메시내에서 자신의 x,y,z 좌표에 의해 주소지정된다. 트리 네트워크에서, 일반적으로 노드들은 이진 트리로 연결되는데, 여기서는 각각의 노드가 하나의 부모 노드와, 두 개의 자식 노드를 갖는다(일부 노드들은 하드웨어 구성에 따라, 자식 노드를 갖지 않거나 또는 하나의 자식 노드만을 가질 수 있다). 원환체 및 트리 네트워크를 이용하는 컴퓨터에서, 두 개의 네트워크들은 일반적으로 개별적인 라우팅 회로, 개별적인 물리적 링크, 및 개별적인 메세지 버퍼를 가지면서 서로 독립적으로 구현된다.

[0009] 원환체 네트워크는 점대점 연산에 적합하지만, 트리 네트워크는 일반적으로 점대점 통신에서는 비능률적이다. 하지만, 트리 네트워크는 어떠한 일괄적 연산들, 즉 모든 계산 노드들이 동시적으로 참여하는 메세지 전달 연산들의 경우에서 높은 대역폭과 낮은 레이턴시를 제공해준다.

발명의 내용

해결하려는 과제

[0010] 병렬 컴퓨터는 작업동안에 동시적으로 동작하는 수 천개의 계산 노드들을 포함할 수 있기 때문에, 병렬 컴퓨터는 대량의 전력을 소모시킬 수 있다. 소비자가 특정 전력량, 즉 피크 전력량보다 많은 전력량을 소모시킨 후에는, 전기 공급자는 일반적으로 평소보다 높은 요금으로 소비자에게 비용을 청구한다. 작업동안에 동시적으로 동작하는 많은 갯수의 계산 노드들로 인하여, 병렬 컴퓨터는 종종 피크 전력량 이상을 소모한다. 이에 따라, 독자는 병렬 컴퓨터에서의 사전적 전력 관리에서 개선의 여지가 존재함을 이해할 것이다.

과제의 해결 수단

[0011] 병렬 컴퓨터에서의 사전적 전력 관리를 위한 방법, 서비스 노드, 및 제품이 개시되고, 병렬 컴퓨터는 서비스 노드와 복수의 계산 노드를 포함하고, 서비스 노드는 대역외 서비스 네트워크를 통해 계산 노드에 연결되고, 각각의 계산 노드는 컴퓨터 프로세서와 이 컴퓨터 프로세서에 동작가능하게 결합된 컴퓨터 메모리를 포함한다. 실시예들은, 서비스 노드에 의해, 병렬 컴퓨터내의 계산 노드의 연산 그룹상의 작업을 개시하도록 하는 사용자 명령을 수신하고(상기 사용자 명령은 계산 노드에 대한 전력 관리 속성을 포함함); 연산 그룹의 계산 노드에 대한

전력 관리 속성에 따라 서비스 노드에 의해, 컴퓨터 프로세서 전력 소모비율과 컴퓨터 메모리 전력 소모비율을 포함하는 연산 그룹의 각각의 계산 노드에 대한 전력 소모비율을 설정하며; 서비스 노드에 의해, 병렬 컴퓨터의 연산 그룹의 계산 노드상의 작업을 개시하는 것을 포함한다.

[0012] 본 발명의 전술된 특징과 장점, 및 기타의 특징과 장점은 첨부된 도면들에서 도시된 이하의 본 발명의 예시적인 실시예들의 보다 구체적인 설명을 통해 분명해질 것이며, 본 도면내에서 동일한 참조 번호는 일반적으로 본 발명의 예시적인 실시예들의 동일 부분들을 나타낸다.

발명의 효과

[0013] 병렬 컴퓨터에서의 개선된 사전적 전력 관리가 제공된다.

도면의 간단한 설명

[0014] 도 1은 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리를 위한 예시적인 시스템을 도시한다.

도 2는 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리에서 유용한 예시적인 계산 노드의 블럭도를 설명한다.

도 3은 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리에서 유용한 예시적인 서비스 노드를 포함하는 자동 컴퓨팅 머신의 블럭도를 설명한다.

도 4a는 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리가 가능할 수 있는 시스템에서 유용한 예시적인 점대점 아답터를 도시한다.

도 4b는 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리가 가능할 수 있는 시스템에서 유용한 예시적인 글로벌 결합 네트워크 아답터를 도시한다.

도 5는 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리가 가능할 수 있는 시스템에서 유용한 점대점 연산에 대해 최적화된 예시적인 데이터 통신 네트워크를 나타내는 선도를 설명한다.

도 6은 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리가 가능할 수 있는 시스템에서 유용한 일괄적 연산에 대해 최적화된 예시적인 데이터 통신 네트워크를 나타내는 선도를 설명한다.

도 7은 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리를 위한 예시적인 방법을 나타내는 흐름도를 설명한다.

도 8은 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리를 위한 추가적인 예시적 방법을 나타내는 흐름도를 설명한다.

발명을 실시하기 위한 구체적인 내용

[0015] 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리를 위한 예시적인 방법, 장치, 및 컴퓨터 프로그램 제품을 도 1을 시작으로, 첨부된 도면을 참조하여 설명한다. 도 1은 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리를 위한 예시적인 시스템을 도시한다. 도 1의 시스템은 병렬 컴퓨터(100), 데이터 저장 장치(118)의 형태의 컴퓨터용 비휘발성 메모리, 프린터(120)의 형태의 컴퓨터용 출력 장치, 및 컴퓨터 단말기(122)의 형태의 컴퓨터용 입력/출력 장치를 포함한다. 도 1의 예시에서의 병렬 컴퓨터(100)는 복수의 계산 노드(102)를 포함한다.

[0016] 계산 노드들(102)은 고속 이더넷 네트워크(174), JTAG(Joint Test Action Group) 네트워크(104), 일괄적 연산에 대해 최적화된 글로벌 결합 네트워크(106), 및 점대점 연산에 대해 최적화된 원환체 네트워크(108)를 포함하는 여러개의 독립적 데이터 통신 네트워크에 의한 데이터 통신을 위해 결합된다. 글로벌 결합 네트워크(106)는 계산 노드를 트리로서 편성하기 위해 계산 노드에 연결된 데이터 통신 링크를 포함하는 데이터 통신 네트워크이다. 각각의 데이터 통신 네트워크는 계산 노드들(102)간의 데이터 통신 링크로 구현된다. 데이터 통신 링크는 병렬 컴퓨터의 계산 노드들간의 병렬 연산을 위한 데이터 통신을 제공한다.

[0017] 또한, 병렬 컴퓨터의 계산 노드(102)는 적어도 하나의 계산 노드들의 연산 그룹(132)으로 편성된다. 계산 노드들의 연산 그룹은 작업 수행에 관여하는 병렬 컴퓨터내의 모든 계산 노드들의 서브세트이다. 연산 그룹은 일괄적 병렬 연산 또는 점대점 연산을 위해 구성될 수 있다.

[0018] 일괄적 연산은 연산 그룹의 계산 노드들간의 데이터 통신으로 이행된다. 일괄적 연산은 연산 그룹의 모든 계산 노드들이 참여하는 노드들의 기능이다. 일괄적 연산은 계산 노드들의 연산 그룹내의 모든 계산 노드들에 의해 동시적으로(즉, 거의 동시에) 실행되는 연산, 즉 메세지 전달 컴퓨터 프로그램 명령이다. 이와 같은 연산 그룹은 병렬 컴퓨터(100)내의 모든 계산 노드들 또는 모든 계산 노드들의 서브세트를 포함할 수 있다. 일괄적 연산은 종종 점대점 연산을 중심으로 구축된다. 일괄적 연산은 연산 그룹내의 모든 계산 노드들상의 모든 프로세스들이 일치하는 인수(argument)를 가지면서 동일한 일괄적 연산을 호출할 것을 필요로 한다. '브로드캐스트'는 연산 그룹의 계산 노드들간에 데이터를 이동시키기 위한 일괄적 연산의 예이다. '감축' 연산은 연산 그룹의 계산 노드들간에 분배되는 데이터에 대해 산술적 또는 논리적 기능을 실행하는 일괄적 연산의 예이다. 연산 그룹은 예컨대, MPI '통신자'로서 이행될 수 있다.

[0019] 'MPI'는 '메세지 전달 인터페이스(Message Passing Interface)를 말하며, 종래기술의 병렬 통신 라이브러리로서, 병렬 컴퓨터상의 데이터 통신을 위한 컴퓨터 프로그램 명령의 모듈이다. 본 발명의 실시예에 따른 시스템과 함께 이용하도록 개선될 수 있는 종래기술의 병렬 통신 라이브러리의 예에는 MPI 및 '병렬 가상 머신('Parallel Virtual Machine; PVM') 라이브러리가 포함된다. PVM은 테네시 대학, 오크 릿지 국립 연구소, 및 이모리 대학에 의해 개발되었다. MPI는 MPI 표준을 정의하고 유지하는 수 많은 조직들의 대표단으로 구성된 개방 그룹인 MPI 포럼에 의해 공표된다. 본 명세서를 쓰는 시점에서의 MPI는 분산된 메모리 병렬 컴퓨터상에서 병렬 프로그램을 구동시키는 계산 노드들간의 통신을 위한 사실상의 표준이다. 이 규정은 때때로 설명의 편의를 위해 MPI 기술을 이용하지만, 이러한 MPI의 이용은 본 발명의 요건 또는 한정사항은 아니다.

[0020] 몇몇의 일괄적 연산은 연산 그룹내의 특정한 계산 노드상에서 구동되는 단일한 발송 또는 수신 프로세스를 갖는다. 예를 들어, '브로드캐스트' 일괄적 연산에서, 나머지 다른 모든 계산 노드들에 데이터를 배분하는 계산 노드상의 프로세스는 발송 프로세스이다. '수집' 연산에서, 예컨대, 나머지 다른 계산 노드들로부터 모든 데이터를 수신하는 계산 노드상의 프로세스는 수신 프로세스이다. 이와 같은 발송 프로세스 또는 수신 프로세스가 구동되는 계산 노드를 논리적 루트라고 칭한다.

[0021] 대부분의 일괄적 연산들은 브로드캐스트, 수집, 확산, 및 감축의 네 개의 기본적인 연산들의 변형 또는 조합이다. 이러한 일괄적 연산들에 대한 인터페이스는 MPI 포럼에 의해 공표된 MPI 표준에서 정의되어 있다. 하지만, 일괄적 연산을 실행하는 알고리즘은 MPI 표준에서 정의되어 있지 않다. 브로드캐스트 연산에서, 모든 프로세스들은 동일한 루트 프로세스를 지정하며, 이 루트 프로세스의 버퍼 콘텐츠가 송신될 것이다. 루트 프로세스 이외의 프로세스들은 수신 버퍼를 지정한다. 이러한 연산 이후, 모든 버퍼들은 루트 프로세스로부터의 메세지를 보유한다.

[0022] 확산 연산에서, 논리적 루트는 논리적 루트상의 데이터를 세그먼트들로 분할하고, 서로 다른 세그먼트들을 연산 그룹내의 각각의 계산 노드에게 분배한다. 확산 연산에서, 모든 프로세스들은 일반적으로 동일한 수신 카운트를 지정한다. 송신 인수는 오직 루트 프로세스에게만 중요하며, 루트 프로세스의 버퍼는 사살상 주어진 데이터 유형의 발송카운트 N개 엘리먼트를 보유하며, 여기서 N은 주어진 계산 노드들의 그룹에서의 프로세스들의 갯수이다. 발송 버퍼는 분할되어 (논리적 루트상의 프로세스를 포함한) 모든 프로세스들에게 확산된다. 각각의 계산 노드에는 '랭크'로 칭해지는 순차적 식별자가 할당된다. 이러한 연산 이후, 루트는 발송카운트 데이터 엘리먼트를 각각의 프로세스에게 랭크 오름순으로 보낸다. 랭크 0은 발송 버퍼로부터 제1 발송카운트 데이터 엘리먼트를 수신한다. 랭크 1은 발송 버퍼로부터 제2 발송카운트 데이터 엘리먼트를 수신한다.

[0023] 수집 연산은 확산 연산의 기술사항과는 완전히 반대인 다대점 일괄적 연산이다. 즉, 수집은 일정한 데이터유형의 엘리먼트가 랭크화된 계산 노드로부터 루트 노드내의 수신 버퍼로 수집되는 다대점 일괄적 연산이다.

[0024] 감축 연산도 두 개의 데이터 엘리먼트들상에서 수행되는 산술적 또는 논리적 기능을 포함하는 다대점 일괄적 연산이다. 모든 프로세스들은 동일한 '카운트' 및 동일한 산술적 또는 논리적 기능을 지정한다. 이러한 감축 연산 이후, 모든 프로세스들은 계산 노드 발송 버퍼로부터 카운트 데이터 엘리먼트를 루트 프로세스에게 보낸다. 감축 연산에서는, 대응하는 발송 버퍼 위치로부터의 데이터 엘리먼트가 산술적 또는 논리적 연산들에 의해 쌍형태로 결합되어 루트 프로세스의 수신 버퍼내에서 대응하는 단일 엘리먼트가 생성된다. 런타임시에 애플리케이션 특유적인 감축 연산이 정의될 수 있다. 병렬 통신 라이브러리는 미리정의된 연산을 지원할 수 있다. 예컨대, MPI는 다음과 같은 미리정의된 감축 연산을 제공한다:

[0025] MPI_MAX 최대

[0026] MPI_MIN 최소

[0027] MPI_SUM 덧셈

[0028] MPI_PROD 곱

[0029] MPI_LAND 논리 AND

[0030] MPI_BAND 비트단위 AND

[0031] MPI_LOR 논리 OR

[0032] MPI_BOR 비트단위 OR

[0033] MPI_LXOR 논리 배타적 OR

[0034] MPI_BXOR 비트단위 배타적 OR

[0035] 계산 노드에 더하여, 병렬 컴퓨터(100)는 하나의 데이터 네트워크(174)를 통해 계산 노드(102)에 결합된 입력/출력('I/O') 노드(110, 114)를 포함한다. I/O 노드(110, 114)는 계산 노드(102)와 I/O 장치들(118, 120, 122) 사이에 I/O 서비스를 제공한다. I/O 노드(110, 114)는 근거리 네트워크('LAN')(130)를 통해 데이터 통신 I/O 장치들(118, 120, 122)에 연결된다. 병렬 컴퓨터(100)는 또한 하나의 네트워크(104)를 통해 계산 노드들에 결합된 서비스 노드(116)를 포함한다. 서비스 노드(116)는 계산 노드에 프로그램을 로딩시키고, 계산 노드상에서 프로그램 실행을 시작하고, 계산 노드상의 프로그램 연산의 결과를 회수하는 것 등의 복수의 계산 노드들에 공통하는 서비스를 제공한다. 서비스 노드(116)는 서비스 애플리케이션(124)을 구동시키고, 컴퓨터 단말기(122)상에서 구동하는 서비스 애플리케이션 인터페이스(126)를 통해 사용자(128)와 통신한다.

[0036] 아래의 본 명세서에서 보다 자세하게 설명된 바와 같이, 도 1의 시스템은 일반적으로 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리를 위해 동작한다. 본 발명의 실시예에 따라 병렬 컴퓨터에서의 전력 관리는 병렬 컴퓨터에서의 미래의 문제점, 필요사항, 또는 변동의 예측으로 전력이 관리되기 때문에 사전적이라고 일반적으로 기술된다. 이와 같은 사전적 관리는 병렬 컴퓨터에서의 미래의 문제점, 필요사항, 또는 변동을 예측하지 않는 반응형 전력 관리와 대비된다. 도 1의 시스템은, 서비스 노드(116)에 의해, 병렬 컴퓨터(100)내의 계산 노드(102)의 연산 그룹(132)상의 작업을 개시하도록 하는 사용자(128) 명령을 수신하고; 연산 그룹(132)의 계산 노드(102)에 대한 전력 관리 속성에 따라 서비스 노드(116)에 의해, 컴퓨터 프로세서 전력 소모비율 및 컴퓨터 메모리 전력 소모비율을 포함하는 연산 그룹(132)의 각각의 계산 노드(102)에 대한 전력 소모비율을 설정하며; 서비스 노드(116)에 의해, 병렬 컴퓨터(100)의 연산 그룹(132)의 계산 노드(102)상의 작업을 개시할 수 있다.

[0037] 병렬 컴퓨터에 의해 수행될 작업은 애플리케이션의 실행 인스턴스이다. 이와 같은 애플리케이션은 연산 그룹내의 각각의 계산 노드에 대한 컴퓨터 프로그램 명령을 포함한다. 그러면, 작업을 수행하는 것은 애플리케이션의 컴퓨터 프로그램 명령을 실행하는 것을 포함한다.

[0038] 이와 같은 작업을 개시하도록 하는 사용자 명령은 계산 노드에 대한 전력 관리 속성을 포함한다. 전력 관리 속성은 서비스 노드가 연산 그룹의 하나 이상의 계산 노드들의 가변적 전력 소모 기능성을 조정하는데 이용되는 파라미터이다. 도 1의 예시에 따른 전형적인 계산 노드는 프로세서의 전력 소모가 작업마다 달라질 수 있도록 하고, 메모리의 전력 소모가 작업마다 달라질 수 있도록 해준다. 본 발명의 실시예에 따른 전력 관리 속성은 병렬 컴퓨터의 연산 그룹내의 계산 노드에 대한 전력 소모비율 및 유휴 싸이클 기간을 포함한다. 전력 소모비율은 컴퓨터 프로그램 명령의 실행 동안에 계산 노드의 전력 소모를 제어하는 파라미터이다. 전력 소모비율은 컴퓨터 프로세서 전력 소모비율 및 컴퓨터 메모리 전력 소모비율을 포함한다. 각각의 비율은 실행 싸이클 대 유휴 싸이클의 비율로서 표현된다. 실행 싸이클은 컴퓨터 메모리 또는 컴퓨터 프로세서가 명령을 실행하는 구간이다. 즉, 실행 싸이클은 컴퓨터 메모리 또는 컴퓨터 프로세서가 활성화상태에 있는 구간이다. 이와 대비되어, 유휴 싸이클은 컴퓨터 메모리 또는 컴퓨터 프로세서가 유휴상태에 있는 구간, 즉 어떠한 명령도 실행하지 않는 구간이다. 유휴상태로 있는 동안에는, 컴퓨터 메모리와 컴퓨터 프로세서 어떠한 것도 전력을 소모하지 않는다.

[0039] 상술한 바와 같이, 전력 관리 속성은 또한 유휴 싸이클의 기간을 포함한다. 유휴 싸이클의 기간이 클수록, 이와 같은 유휴 싸이클 동안에 계산 노드내의 컴퓨터 프로세서 또는 컴퓨터 메모리는 보다 오랫동안 유휴상태에 있다. 서비스 노드는 전력 소모비율을 설정할 때에 유휴 싸이클에 대한 기간을 설정할 수 있다.

[0040] 전력 관리 속성은 작업 특유적일 수 있다. 즉, 사용자는 하나의 작업에 대하여 특정한 전력 관리 속성 세트를 제공하고, 다른 작업에 대해서는 이와 완전히 다른 전력 관리 속성 세트를 제공할 수 있다. 대안적으로, 사용자

는 작업 그룹에 대하여 하나의 전력 관리 속성 세트를 제공할 수 있거나 또는 규칙 기반으로 적용되는 전력 관리 속성 세트를 제공할 수 있다. 예컨대, 사용자는 여름철동안에 수행되는 모든 작업들에 대해 적용되는 하나의 전력 관리 속성 세트와, 그 해 나머지동안에 적용되는 하나의 전력 관리 속성 세트를 제공할 수 있다.

[0041] 도 1의 서비스 노드(116)는 서비스 노드(116)에 의해, 병렬 컴퓨터(100)내의 계산 노드(102)의 연산 그룹(132)상의 작업을 개시하도록 하는 사용자(128) 명령을 수신할 수 있는 컴퓨터 프로그램 명령의 모듈인 서비스 애플리케이션(124)을 포함하며, 상기 사용자 명령은 계산 노드에 대한 전력 관리 속성을 포함한다. 병렬 컴퓨터(100)내의 계산 노드의 연산 그룹(132)상의 작업을 개시하도록 하는 사용자 명령을 수신하는 것은 단말기(122)로부터 서비스 애플리케이션 인터페이스(126)에 의해 제공된 그래픽 사용자 인터페이스('GUI')를 통해 사용자(128)에 의해 입력된 전력 관리 속성을 수신함으로써 수행될 수 있다. 이와 같은 GUI는 컴퓨터 프로세서 전력 소모비율 및 컴퓨터 메모리 전력 소모비율 모두에 적용되는 실행 싸이클 대 유휴 싸이클의 단일 비율에 더하여 유휴 싸이클 기간을 수용하도록 특별히 구성될 수 있다. 이와 달리, GUI는 컴퓨터 프로세서 전력 소모비율로서 설정되는 하나의 비율과 컴퓨터 메모리 전력 소모비율로서 설정되는 하나의 비율인, 두 개의 구별되는 실행 싸이클 대 유휴 싸이클의 비율들에 더하여 유휴 싸이클 기간을 수용하도록 특별히 구성될 수 있다.

[0042] 도 1의 서비스 애플리케이션(124)은 또한, 연산 그룹의 계산 노드(102)에 대한 전력 관리 속성에 따라 서비스 노드(116)에 의해, 컴퓨터 프로세서 전력 소모비율 및 컴퓨터 메모리 전력 소모비율을 포함하는 연산 그룹의 각각의 계산 노드에 대한 전력 소모비율을 설정할 수 있는 컴퓨터 프로그램 명령을 포함한다. 연산 그룹(132)의 각 계산 노드(102)에 대해 전력 소모비율을 설정하는 것은 연산 그룹(132)내의 각각의 계산 노드들(102)이 전력 소모비율을 갖도록 구성시킴으로써 수행될 수 있다. 즉, 각각의 계산 노드들내의 컴퓨터 메모리가 전력 소모비율을 갖도록 구성한다.

[0043] 도 1의 시스템에서, 예시적인 서비스 노드(116)는 대역외 서비스 네트워크, 즉 JTAG 네트워크(104)를 통해 전력 관리 속성을 설정한다. 비록 도 1의 서비스 네트워크는 JTAG 네트워크로서 도시되고 있지만, 본 발명분야의 독자라면 서비스 네트워크가 서비스 노드(116)와 계산 노드(102)사이의 대역외 통신의 인에이블이 가능한 임의의 통신 링크로서 구현될 수 있음을 이해할 것이다. 이와 같은 대역외 통신 링크는 예컨대, 본 발명분야의 당업자에게 떠오를 수 있는, 상호집적 회로('I^2C') 버스, 1-와이어 버스, 주변 컴포넌트 상호접속('PCI') 버스, 시스템 관리 버스('SMB'), 직렬 주변 인터페이스('SPI'), 지능형 플랫폼 관리 버스('IPMB') 등을 포함할 수 있다.

[0044] 도 1의 서비스 애플리케이션(124)은 또한 서비스 노드(116)에 의해, 병렬 컴퓨터(100)의 연산 그룹(132)의 계산 노드(102)상의 작업을 개시할 수 있는 컴퓨터 프로그램 명령을 포함한다. 병렬 컴퓨터(100)의 연산 그룹(132)의 계산 노드(102)상의 작업을 개시하는 것은 연산 그룹(132)내의 각각의 계산 노드(102)를 노드 자신의 작업 특유적 컴퓨터 프로그램 명령을 갖추도록 구성하고, 컴퓨터 프로그램 명령을 실행하도록 하는 통지를 연산 그룹(132)내의 계산 노드(102)에게 보냄으로써 수행될 수 있다.

[0045] 도 1에서 도시된 예시적인 시스템을 구성하는 노드, 네트워크, 및 I/O 장치의 배열은 본 발명의 예시에 불과하며, 이것은 본 발명을 한정하려는 것은 아니다. 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리가 가능한 데이터 프로세싱 시스템은 도 1에서는 도시되지 않은 추가적인 노드, 네트워크, 장치 및 구조물을 포함할 수 있으며, 이러한 것들은 본 발명분야의 당업자에게 떠오를 수 있을 것이다. 도 1의 예시에서의 병렬 컴퓨터(100)는 16개의 계산 노드(102)를 포함하지만, 독자는 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리가 가능한 병렬 컴퓨터는 임의의 갯수의 계산 노드들을 포함할 수 있음을 유념할 것이다. 이더넷 및 JTAG에 더하여, 이와 같은 데이터 프로세싱 시스템에서의 네트워크는 예컨대, TCP(송신 제어 프로토콜), IP(인터넷 프로토콜), 및 본 발명분야의 당업자에게 떠오를 기타의 것들을 포함하는 수 많은 데이터 통신 프로토콜을 지원할 수 있다. 본 발명의 다양한 실시예들은 도 1에서 도시된 것에 더하여 다양한 하드웨어 플랫폼상에서 구현될 수 있다.

[0046] 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리는 복수의 계산 노드들을 포함하는 병렬 컴퓨터상에서 구현될 수 있다. 실제적으로, 이와 같은 컴퓨터는 이와 같은 수 천개의 계산 노드들을 포함할 수 있다. 이어서 각각의 계산 노드는 그 자체적으로 하나 이상의 컴퓨터 프로세서들, 각자 고유의 컴퓨터 메모리, 및 각자 고유의 입력/출력 아답터로 구성된 일정한 종류의 컴퓨터이다. 따라서, 추가적인 설명을 위해, 도 2는 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리에서 유용한 예시적인 계산 노드의 블럭도를 설명한다. 도 2의 계산 노드(152)는 랜덤 액세스 메모리('RAM')(156) 뿐만이 아니라 하나 이상의 컴퓨터 프로세서(164)를 포함한다. 프로세서(164)는 고속 메모리 버스(154)를 통해 RAM(156)에 연결되고, 버스 아답터(194)와 확장 버스(168)를 통해 계산 노드(152)의 다른 컴포넌트들에 연결된다. RAM(156)내에는 애플리케이션 프로그램(158), 즉

병렬 알고리즘을 이용하여 병렬적인 사용자 레벨 데이터 프로세싱을 수행하는 컴퓨터 프로그램 명령 모듈이 저장된다. 도 2의 애플리케이션 프로그램(158)은 송신용 메세지를 저장하기 위한 애플리케이션 버퍼를 다른 계산 노드에 할당한다.

[0047] RAM(156)에는 또한 메세징 모듈(160), 즉 점대점 연산뿐만이 아니라 일괄적 연산을 포함하여, 계산 노드들간의 병렬 통신을 수행하는 컴퓨터 프로그램 명령의 라이브러리가 저장된다. 애플리케이션 프로그램(158)은 메세징 모듈(160)내의 소프트웨어 루틴을 호출함으로써 일괄적 연산을 실행한다. 병렬 통신 루틴의 라이브러리는, C 프로그래밍 언어와 같은 전통적인 프로그래밍 언어를 이용하고, 두 개의 독립적인 데이터 통신 네트워크상의 노드들간에 데이터를 송신 및 수신하는 병렬 통신 루틴을 기록하기 위한 전통적인 프로그래밍 언어를 이용하여, 본 발명의 실시예에 따른 시스템내에서의 이용을 위한 스크래치로부터 개발될 수 있다. 이와 달리, 기존의 종래기술의 라이브러리가 본 발명의 실시예에 따라 연산되도록 개선될 수 있다. 종래기술의 병렬 통신 라이브러리의 예시에는 '메세지 전달 인터페이스'('MPI') 라이브러리 및 '병렬 가상 머신'('PVM') 라이브러리를 포함한다.

[0048] RAM(156)에는 또한 오퍼레이팅 시스템(162), 즉 계산 노드의 기타 자원들에 에 대한 애플리케이션 프로그램의 액세스를 위한 컴퓨터 프로그램 명령 및 루틴 모듈이 저장된다. 실행 쓰레드는 노드의 모든 자원에 대한 완전한 액세스 자격이 주어지기 때문에 병렬 컴퓨터의 계산 노드에서의 애플리케이션 프로그램 및 병렬 통신 라이브러리가 사용자 로그인 및 보안적 사항없이 단일한 실행 쓰레드를 구동시키는 것이 일반적이다. 따라서 병렬 컴퓨터내의 계산 노드상의 오퍼레이팅 시스템에 의해 수행될 업무의 양과 복잡성은 수 많은 쓰레드가 동시적으로 구동되기에 직렬 컴퓨터상의 오퍼레이팅 시스템의 업무의 양보다는 작고 덜 복잡하다. 또한, 도 2의 계산 노드(152)상에는 오퍼레이팅 시스템에 대한 수요를 감소시키는 다른 인자인, 비디오 I/O가 없다. 따라서, 본 오퍼레이팅 시스템은 원래 버전의 페어 다운된 버전(pared down version) 또는 특별한 병렬 컴퓨터상의 연산을 위해 특별히 개발된 오퍼레이팅 시스템으로서, 범용 컴퓨터의 오퍼레이팅 시스템에 비해 매우 경량화될 수 있다. 계산 노드에서의 이용을 위해 유용하게 개선되고 단순화될 수 있는 오퍼레이팅 시스템은 UNIX$_{TM}$, Linux$_{TM}$, Microsoft XP$_{TM}$, IBM® AIX®, IBM i5/OS® 및, 본 발명분야의 당업자에게 떠오를 기타의 것들을 포함한다.

[0049] 또한, RAM(156)에는 컴퓨터 프로세서 전력 소모비율(714)과 컴퓨터 메모리 전력 소모비율(716)을 포함하는 전력 소모비율(712)이 저장된다. 전력 소모비율들은 컴퓨터 프로그램 명령의 실행 동안에 계산 노드의 전력 소모를 제어하기 위한 파라미터들이다. 전력 소모비율(712)은 서비스 노드에 의해 설정된다. 애플리케이션 프로그램(158)의 실행 동안, 계산 노드(156)는 전력 소모비율(712)에 따라 동작한다.

[0050] 도 2의 예시적인 계산 노드(152)는 병렬 컴퓨터의 다른 노드들과의 데이터 통신을 이행하기 위한 여러 개의 통신 아답터들(172, 176, 180, 188)을 포함한다. 이와 같은 데이터 통신은 RS-232 접속을 통해, USB와 같은 외부 버스를 통해, IP 네트워크와 같은 데이터 통신 네트워크를 통해, 그리고 본 발명분야의 당업자에게 떠오르는 기타의 방법들을 통해 직렬로 수행될 수 있다. 통신 아답터는 하드웨어 레벨의 데이터 통신을 이행하며, 이를 통해 하나의 컴퓨터는 직접적으로 또는 네트워크를 통해 다른 컴퓨터에게 데이터 통신을 보낸다. 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리를 위한 시스템에서 유용한 통신 아답터의 예들에는 유선 통신을 위한 모뎀, 유선 네트워크 통신을 위한 이더넷(IEEE 802.3) 아답터, 및 무선 네트워크 통신을 위한 802.11b 아답터가 포함된다.

[0051] 도 2의 예시에서의 데이터 통신 아답터에는 데이터 통신을 위한 예시적인 계산 노드(152)를 기가비트 이더넷(174)에 결합시켜주는 기가비트 이더넷 아답터(172)가 포함된다. 기가비트 이더넷은 IEEE 802.3 표준에서 정의된, 1초당 10억 비트의 데이터율(1 기가비트)을 제공하는 네트워크 송신 표준이다. 기가비트 이더넷은 멀티모드 광섬유 광케이블, 단일모드 광섬유 광케이블, 또는 비차폐된 이가닥선을 통해 동작하는 이더넷의 변형이다.

[0052] 도 2의 예시에서의 데이터 통신 아답터에는 데이터 통신을 위한 예시적인 계산 노드(152)를 JTAG 마스터 회로(178)에 결합시켜주는 JTAG 슬레이브 회로(176)가 포함된다. JTAG란 경계 스캔을 이용하여 인쇄 회로 기판을 테스트하는데 이용되는 테스트 액세스 포트용 표준 테스트 액세스 포트 및 경계 스캔 구조로 칭해지는 IEEE 1149.1 표준에서 이용되는 통상적인 명칭이다. JTAG는 매우 폭넓게 적응되기에, 이 경우, 경계 스캔은 JTAG와 다소 동의적이다. JTAG는 인쇄 회로 기판뿐만이 아니라 집적 회로의 경계 스캔을 수행하는데 이용되고, 또한 임베디드 시스템을 디버깅하기 위한 메카니즘으로서 유용하며, 본 시스템내에 편리한 "백 도어"를 제공해준다. 도 2의 예시적인 계산 노드는 모두 이것들 중 세 개일 수 있다: 계산 노드는 일반적으로 인쇄 회로 기판상에 설치된 하나 이상의 집적 회로를 포함하며, 자신 고유의 프로세서, 자신 고유의 메모리, 및 자신 고유의 I/O 성능을 갖는 임베디드 시스템으로서 구현될 수 있다. JTAG 슬레이브(176)를 통한 JTAG 경계 스캔은 본 발명의 실시예에

따라 병렬 컴퓨터에서의 사전적 전력 관리에서의 이용을 위한 계산 노드(152)내의 프로세서 레지스터 및 메모리를 효율적으로 구성시킬 수 있다.

[0053] 도 2의 예시에서의 데이터 통신 아답터에는 데이터 통신을 위한 예시적 계산 노드(152)를 점대점 메세지 전달 연산에 대해 최적화된 네트워크(108), 예컨대 삼차원 원환체 또는 메시로서 구성된 네트워크에 결합시켜주는 점대점 아답터(180)가 포함된다. 점대점 아답터(180)는 세 개의 x, y, 및 z 통신축상의 여섯 개의 양방향 링크들, 즉 +x(181), -x(182), +y(183), -y(184), +z(185), 및 -z(186)을 통한 여섯 개의 방향으로의 데이터 통신을 제공한다.

[0054] 도 2의 예시에서의 데이터 통신 아답터에는 데이터 통신을 위한 예시적인 계산 노드(152)를 예컨대 이진 트리로서 구성된 글로벌 결합 네트워크상의 일괄적 메세지 전달 연산에 대해 최적화된 네트워크(106)에 결합시켜주는 글로벌 결합 네트워크 아답터(188)가 포함된다. 글로벌 결합 네트워크 아답터(188)는 세 개의 양방향 링크들(두 개는 자식 노드(190)를 향한 것이고, 하나는 부모 노드(192)를 향한 것임)을 통해 데이터 통신을 제공한다.

[0055] 예시적인 계산 노드(152)에는 두 개의 산술 논리 장치('ALU')가 포함된다. ALU(166)는 프로세서의 컴포넌트이며, 분리된 ALU(170)는 감축 연산의 산술적 및 논리적 기능을 수행하는데 이용하기 위한 글로벌 결합 네트워크 아답터(188)의 독점적 이용에 전용된다. 병렬 통신 라이브러리(160)내의 감축 루틴의 컴퓨터 프로그램 명령은 산술적 또는 논리적 기능을 위한 명령을 명령 레지스터(169)내에 래치시킬 수 있다. 감축 연산의 산술적 또는 논리적 기능이 예컨대 '합산' 또는 '논리적 OR'인 경우, 글로벌 결합 네트워크 아답터(188)는 프로세서(164)내의 ALU(166)의 이용에 의해 산술적 또는 논리적 연산을 실행시킬 수 있거나, 또는 전용 ALU(170)를 이용함으로써 일반적으로 훨씬 빠르게 실행시킬 수 있다.

[0056] 추가적인 설명을 위해, 이에 따라, 도 3은 본 발명의 실시예에 따라 병렬 컴퓨터에서의 사전적 전력 관리에서 유용한 예시적인 서비스 노드(252)를 포함하는 자동화된 컴퓨팅 머신의 블럭도를 설명한다. 도 3의 서비스 노드(252)는 적어도 하나의 컴퓨터 프로세서(256) 또는 'CPU'를 포함할 뿐만이 아니라, 고속 메모리 버스(266) 및 버스 아답터(258)를 통해 프로세서(256) 및 서비스 노드의 다른 컴포넌트들에 연결된 랜덤 액세스 메모리(268)('RAM')를 포함한다.

[0057] RAM(268)에는 서비스 애플리케이션(124), 즉 본 발명의 실시예에 따라 병렬 컴퓨터에서의 전력을 사전적으로 관리할 수 있는 컴퓨터 프로그램 명령의 모듈이 저장된다. 도 3의 서비스 애플리케이션(124)은, 서비스 노드(252)에 의해, 병렬 컴퓨터내의 계산 노드들의 연산 그룹상의 작업을 개시하도록 하는 사용자 명령(706)(계산 노드에 대한 전력 관리 속성(708)을 포함함)을 수신하고; 연산 그룹의 계산 노드에 대한 전력 관리 속성에 따라 서비스 노드에 의해, 컴퓨터 프로세서 전력 소모비율과 컴퓨터 메모리 전력 소모비율을 포함하는 연산 그룹의 각각의 계산 노드에 대한 전력 소모비율을 설정하며; 서비스 노드에 의해, 병렬 컴퓨터의 연산 그룹의 계산 노드들상의 작업을 개시할 수 있는 컴퓨터 프로그램 명령을 포함한다.

[0058] RAM(268)에는 또한 오퍼레이팅 시스템(254)이 포함된다. 본 발명의 실시예에 따라 서비스 노드에서 유용한 오퍼레이팅 시스템은 $UNIX_{TM}$, $Linux_{TM}$, Microsoft $Vista_{TM}$, Microsoft XP_{TM}, IBM® AIX®, IBM i5/OS® 및, 본 발명분야의 당업자에게 떠오를 기타의 것들을 포함한다. 도 3의 예시에서는 RAM(268)내에 오퍼레이팅 시스템(254) 및 미디어 서버 애플리케이션 프로그램(202)이 있는 것이 도시되지만, 이와 같은 소프트웨어의 수 많은 컴포넌트들은 일반적으로 예컨대 디스크 드라이브(270)상의 비휘발성 메모리내에 저장된다. 도 3의 서비스 노드(252)에는 버스 아답터(258), 고속 버스를 위한 드라이브 전자장치를 포함하는 컴퓨터 하드웨어 컴포넌트, 전면측 버스(262), 비디오 버스(264), 및 메모리 버스(266) 뿐만이 아니라, 저속 확장 버스(260)를 위한 드라이브 장치가 포함된다. 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리에 유용한 버스 아답터의 예에는 인텔 노쓰브릿지, 인텔 메모리 제어기 허브, 인텔 사우쓰브릿지, 및 인텔 I/O 제어기 허브가 포함된다. 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리에 유용한 확장 버스의 예에는 산업 표준 구조('ISA') 버스 및 주변 컴포넌트 상호접속('PCI') 버스가 포함된다.

[0059] 도 3의 서비스 노드(252)에는 확장 버스(260)와 버스 아답터(258)를 거쳐서 프로세서(256)와 서비스 노드(252)의 기타 컴포넌트들에 결합된 디스크 드라이브 아답터(272)가 포함된다. 디스크 드라이브 아답터(272)는 디스크 드라이브(270) 형태의 비휘발성 데이터 저장소를 서비스 노드(252)에 연결시킨다. 서비스 노드에서 유용한 디스크 드라이브 아답터에는 집적 드라이브 전자장치('IDE') 아답터, 소규모 컴퓨터 시스템 인터페이스('SCSI') 아답터, 및 본 발명분야의 당업자에게 떠오르는 기타의 것들이 포함된다. 또한, 비휘발성 컴퓨터 메모리는 광 디스크 드라이브, 전기적으로 제거가능한 프로그램가능 판독 전용 메모리(소위, 'EEPROM' 또는 '플래쉬' 메모리로

칭해짐), RAM 드라이브, 및 본 발명분야의 당업자에게 떠오르는 기타의 것들로서 서비스 노드를 위해 구현될 수 있다.

[0060] 도 3의 예시적인 서비스 노드(252)에는 하나 이상의 입력/출력('I/O') 아답터(278)가 포함된다. 서비스 노드내의 I/O 아답터는 예컨대, 컴퓨터 디스플레이 스크린과 같은 디스플레이 장치로의 출력뿐만이 아니라, 키보드와 마우스와 같은 사용자 입력 장치(281)로부터의 사용자 입력을 제어하기 위한 소프트웨어 드라이버 및 컴퓨터 하드웨어를 통한 사용자 지향 입력/출력을 구현시킨다. 도 3의 예시적인 서비스 노드(252)에는 비디오 아답터(209)가 포함되며, 이 비디오 아답터(209)는 디스플레이 스크린 또는 컴퓨터 모니터와 같은 디스플레이 장치(280)로의 그래픽 출력을 위해 특별히 설계된 I/O 아답터의 예이다. 비디오 아답터(209)는 고속 비디오 버스(264), 버스 아답터(258), 및 전면측 버스(262)(이것 또한 고속 버스임)를 거쳐 프로세서(256)에 연결된다.

[0061] 도 3의 예시적인 서비스 노드(252)는 다른 컴퓨터(282)와의 데이터 통신 및 데이터 통신 네트워크(200)와의 데이터 통신을 위한 통신 아답터(267)를 포함한다. 이와 같은 데이터 통신은 RS-232 접속을 통해, 범용 직렬 버스('USB')와 같은 외부 버스를 통해, IP 데이터 통신 네트워크와 같은 데이터 통신 네트워크를 통해, 그리고 본 발명분야의 당업자에게 떠오르는 기타의 방법들을 통해 직렬로 수행될 수 있다. 통신 아답터는 하드웨어 레벨의 데이터 통신을 구현시키며, 이를 통해 하나의 컴퓨터는 직접적으로 또는 데이터 통신 네트워크를 통해 다른 컴퓨터에게 데이터 통신을 보낸다. 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리에 유용한 통신 아답터의 예들에는 유선 다이얼업(dial up) 통신을 위한 모뎀, 유선 데이터 통신 네트워크 통신을 위한 이더넷(IEEE 802.3) 아답터, 및 무선 데이터 통신 네트워크 통신을 위한 802.11 아답터가 포함된다.

[0062] 추가적인 설명을 위해, 도 4a는 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리가 가능한 시스템에서 유용한 예시적인 점대점 아답터(180)를 도시한다. 점대점 아답터(180)는 점대점 연산에 대해 최적화된 데이터 통신 네트워크에서의 이용을 위해 설계되며, 이 네트워크는 계산 노드들을 삼차원 원환체 또는 메시로 편성한다. 도 4a의 예시에서의 점대점 아답터(180)는 네 개의 단방향 데이터 통신 링크를 통해 x축을 따라 -x 방향(182)으로의 다음번째 노드로의 데이터 통신 및 이로부터의 데이터 통신과, +x 방향(181)으로의 다음번째 노드로의 데이터 통신 및 이로부터의 데이터 통신을 제공한다. 점대점 아답터(180)는 또한 네 개의 단방향 데이터 통신 링크를 통해 y축을 따라 -y 방향(184)으로의 다음번째 노드로의 데이터 통신 및 이로부터의 데이터 통신과, +y 방향(183)으로의 다음번째 노드로의 데이터 통신 및 이로부터의 데이터 통신을 제공한다. 도 4a에서의 점대점 아답터(180)는 또한 네 개의 단방향 데이터 통신 링크를 통해 z축을 따라 -z 방향(186)으로의 다음번째 노드로의 데이터 통신 및 이로부터의 데이터 통신 및, +z 방향(185)으로의 다음번째 노드로의 데이터 통신 및 이로부터의 데이터 통신을 제공한다.

[0063] 추가적인 설명을 위해, 도 4b는 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리가 가능한 시스템에서 유용한 예시적인 글로벌 결합 네트워크 아답터(188)를 도시한다. 글로벌 결합 네트워크 아답터(188)는 일괄적 연산에 대해 최적화된 네트워크에서의 이용을 위해 설계되며, 이 네트워크는 병렬 컴퓨터의 계산 노드들을 이진 트리로 편성한다. 도 4b의 예시에서의 글로벌 결합 네트워크 아답터(188)는 네 개의 단방향 데이터 통신 링크(190)를 통해 두 개의 자식 노드들로의 데이터 통신 및 이 노드들로부터의 데이터 통신을 제공해준다. 글로벌 결합 네트워크 아답터(188)는 또한 두 개의 단방향 데이터 통신 링크(192)를 통해 부모 노드로의 데이터 통신 및 부모 노드로부터의 데이터 통신을 제공한다.

[0064] 추가적인 설명을 위해, 도 5는 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리가 가능한 시스템에서 유용한 점대점 연산에 대해 최적화된 예시적인 데이터 통신 네트워크(108)를 도시하는 선도를 나타낸다. 도 5의 예시에서, 점들은 병렬 컴퓨터의 계산 노드(102)를 나타내고, 점들사이의 점선들은 계산 노드들사이의 데이터 통신 링크(103)를 나타낸다. 데이터 통신 링크(103)는 도 4a에서의 예시에 대해서 설명된 것과 유사한 점대점 데이터 통신 아답터로 구현되며, 이 데이터 통신 링크는 세 개의 x, y, 및 z 축상에서 여섯 개의 +x(181), -x(182), +y(183), -y(184), +z(185), 및 -z(186) 방향들로 앞뒤로 놓여 있다. 링크와 계산 노드는 점대점 연산에 대해 최적화된 이러한 데이터 통신 네트워크에 의해 삼차원 메시(105)로 편성된다. 메시(105)는 메시(105)의 반대측상의 메시(105)내의 최외각 계산 노드들을 연결시켜주는 각각의 축상의 랩 어라운드 링크(wrap-around link)를 갖는다. 이러한 랩 어라운트 링크는 원환체(107)의 일부를 형성한다. 원환체내의 각각의 계산 노드는 x, y, z 좌표 세트에 의해 고유하게 지정된 원환체에서의 위치를 갖는다. y 및 z 방향으로의 랩 어라운드 링크는 명확함을 위해 생략되었지만, 이것은 x 방향으로 도시된 랩 어라운드 링크와 유사한 방식으로 구성됨을 독자는 유념할 것이다. 설명의 명확함을 위해, 도 5의 데이터 통신 네트워크는 단지 27개의 계산 노드들만을 갖는 것으로 도시되지만, 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리에서의 이용을 위한 점대점 연산에 대해 최적화된 데이터 통신 네트워크는 단지 소수개만의 계산 노드들을 포함할 수 있거나 또

는 수천 개의 계산 노드들을 포함할 수 있음을 독자는 이해할 것이다.

[0065] 추가적인 설명을 위해, 도 6은 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리가 가능한 시스템에서 유용한 일괄적 연산에 대해 최적화된 예시적인 데이터 통신 네트워크(106)를 나타내는 선도를 설명한다. 도 6의 예시적인 데이터 통신 네트워크는 계산 노드들을 트리로서 편성하기 위해 계산 노드들에 연결된 데이터 통신 링크를 포함한다. 도 6의 예시에서, 점들은 병렬 컴퓨터의 계산 노드(102)를 나타내고, 점들사이의 점선들(103)은 계산 노드들사이의 데이터 통신 링크를 나타낸다. 데이터 통신 링크는 도 4b에서의 예시에 대해서 설명된 것과 유사한 글로벌 결합 네트워크 아답터로 구현되며, 각각의 노드는 어느정도 예외를 갖지만, 일반적으로는 두 개의 자식 노드들로의 데이터 통신 및 이로부터의 데이터 통신과, 부모 노드로의 데이터 통신 및 이로부터의 데이터 통신을 제공한다. 이진 트리(106)내의 노드들은 물리적 루트 노드(202), 브랜치 노드(204), 및 리프 노드(206)로서 특징지워질 수 있다. 루트 노드(202)는 두 개의 자식 노드를 갖지만 부모 노드는 갖지 않는다. 리프 노드(206) 각각은 부모 노드를 갖지만, 자식 노드를 갖지 않는다. 브랜치 노드(204) 각각은 부모 노드와 두 개의 자식 노드들 모두를 갖는다. 이에 따라, 링크와 계산 노드는 일괄적 연산에 대해 최적화된 이러한 데이터 통신 네트워크에 의해 이진 트리(106)로 편성된다. 설명의 명확함을 위해, 도 6의 데이터 통신 네트워크는 단지 31개의 계산 노드들만을 갖는 것으로 도시되지만, 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리를 위한 시스템에서의 이용을 위한 일괄적 연산에 대해 최적화된 데이터 통신 네트워크는 단지 소수개만의 계산 노드들을 포함할 수 있거나 또는 수천 개의 계산 노드들을 포함할 수 있음을 독자는 이해할 것이다.

[0066] 도 6의 예시에서, 트리내의 각각의 노드들은 '랭크'(250)로서 칭해지는 유닛 식별자가 할당된다. 노드의 랭크는 트리 네트워크내의 점대점 및 일괄적 연산들 모두에서의 이용을 위한 트리 네트워크내의 노드의 위치를 고유하게 식별해준다. 이 예시에서의 랭크는 루트 노드(202)에 0이 할당되는 것을 시작으로, 트리의 제2층에서의 제1 노드에 1이 할당되고, 트리의 제2층에서의 제2 노드에 2가 할당되고, 트리의 제3층에서의 제1 노드에 3이 할당되고, 트리의 제3층에서의 제2 노드에 4가 할당되는 식으로 정수들로서 할당되어진다. 설명의 용이함을 위해, 여기서는 트리의 첫번째 세 개의 층들의 랭크들만을 도시하고 있지만, 트리 네트워크내의 모든 계산 노드들은 고유한 랭크가 할당된다.

[0067] 추가적인 설명을 위해, 도 7은 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리를 위한 예시적인 방법을 도시하는 흐름도를 설명한다. 병렬 컴퓨터는 연산 그룹(132)으로서 편성된 복수의 계산 노드들(102)을 포함한다. 병렬 컴퓨터는 또한 JTAG 네트워크(도 1상의 도면번호 104)와 같은, 대역외 서비스 네트워크를 통해 계산 노드(102)에 연결된 서비스 노드(116)를 포함한다. 각각의 계산 노드(102)는 컴퓨터 프로세서 및 이 컴퓨터 프로세서에 동작가능하게 결합된 컴퓨터 메모리를 포함한다. 몇몇 실시예들에서, 복수의 계산 노드들(102)은 데이터 통신을 위해 복수의 데이터 통신 네트워크를 통해 연결된다. 복수의 데이터 통신 네트워크는 점대점 데이터 통신에 대해 최적화된 데이터 통신 네트워크(도 1상의 도면번호 104)를 포함할 수 있다. 복수의 데이터 통신 네트워크는 또한 일괄적 연산에 대해 최적화된 데이터 통신 네트워크(도 1상의 도면번호 106)를 포함할 수 있다.

[0068] 도 7의 방법은, 서비스 노드(116)에 의해, 병렬 컴퓨터내의 계산 노드(102)의 연산 그룹(132)상의 작업(704)을 개시하도록 하는 사용자 명령(706)을 수신하는 것(702)을 포함하며, 이 사용자 명령(706)은 계산 노드(102)에 대한 전력 관리 속성(708)을 포함한다. 서비스 노드(116)에 의해, 병렬 컴퓨터내의 계산 노드(102)의 연산 그룹(132)상의 작업(704)을 개시하도록 하는, 계산 노드(102)에 대한 전력 관리 속성(708)을 포함하는 사용자 명령(706)을 수신하는 것(702)은, 예컨대, 컴퓨터 프로세서 전력 소모비율과 컴퓨터 메모리 전력 소모비율에 대한 실행 싸이클과 유휴 싸이클을 포함하는 전력 소모비율뿐만이 아니라, 유휴 싸이클에 대한 기간을 수신하는 것을 포함한다.

[0069] 병렬 컴퓨터(100)내의 계산 노드의 연산 그룹상의 작업(704)을 개시하도록 하는 사용자 명령(706)을 수신하는 것(702)은 서비스 애플리케이션 인터페이스(126)에 의해 제공된 그래픽 사용자 인터페이스('GUI')내로 사용자(128)에 의해 입력되는 전력 관리 속성을 단말기(122)로부터 수신함으로써 수행될 수 있다. 이와 같은 GUI는 컴퓨터 프로세서 전력 소모비율과 컴퓨터 메모리 전력 소모비율 모두에 적용되는 실행 싸이클 대 유휴 싸이클의 단일 비율에 더하여 유휴 싸이클 기간을 수용하도록 특별히 구성될 수 있다. 이와 달리, GUI는 두 개의 구별되는 실행 싸이클 대 유휴 싸이클의 비율들, 즉 컴퓨터 프로세서 전력 소모비율로서 설정되는 하나의 비율과 컴퓨터 메모리 전력 소모비율로서 설정되는 하나의 비율에 더하여 유휴 싸이클 기간을 수용하도록 특별히 구성될 수 있다.

[0070] 도 7의 방법은 또한, 연산 그룹(132)의 계산 노드(102)에 대한 전력 관리 속성(708)에 따라 서비스 노드(116)에 의해, 컴퓨터 프로세서 전력 소모비율(716)과 컴퓨터 메모리 전력 소모비율(714)을 포함하는 연산 그룹(132)의 각각의 계산 노드(102)에 대한 전력 소모비율들(712)을 설정하는 것(710)을 포함한다. 연산 그룹(132)의 계산 노드(102)에 대한 전력 관리 속성(708)에 따라 서비스 노드(116)에 의해, 컴퓨터 프로세서 전력 소모비율(716)과 컴퓨터 메모리 전력 소모비율(714)을 포함하는 연산 그룹(132)의 각각의 계산 노드(102)에 대한 전력 소모비율들(712)을 설정하는 것(710)은 연산 그룹(132)내의 각각의 계산 노드(102)가 전력 소모비율을 갖도록 구성시킴으로써 수행될 수 있다. 즉, 각각의 계산 노드들내의 컴퓨터 메모리가 전력 소모비율을 갖도록 구성시킨다.

[0071] 도 7의 방법에서, 전력 소모비율(712)은 실행 싸이클(718, 722) 대 유휴 싸이클(720, 724)의 비율로서 표현된다. 실행 싸이클은 컴퓨터 메모리 또는 컴퓨터 프로세서가 명령을 실행하는 구간이다. 즉, 실행 싸이클은 컴퓨터 메모리 또는 컴퓨터 프로세서가 활성화상태에 있는 구간이다. 이와 대비되어, 유휴 싸이클은 컴퓨터 메모리 또는 컴퓨터 프로세서가 유휴상태에 있는 구간, 즉 어떠한 명령도 실행하지 않는 구간이다. 유휴상태로 있는 동안에는, 컴퓨터 메모리와 컴퓨터 프로세서 어떠한 것도 전력을 소모하지 않는다. 사용자(128)가 전력 관리 속성(708)으로서 다음의 전력 소모비율을 제공하는 예를 고려해보자:

[0072] · 컴퓨터 프로세서 전력 소모비율 = 1/10

[0073] · 컴퓨터 메모리 전력 소모비율 = 2/5

[0074] 서비스 노드가 계산 노드에서의 전력 소모비율을 설정하고 계산 노드상의 작업을 개시한 후, 컴퓨터 프로세서는 하나의 싸이클 동안 명령을 실행하고 그 후 열 개의 싸이클 동안 유휴상태에 놓여있을 것이다. 컴퓨터 메모리는 두 개의 싸이클 동안 명령을 실행하고 그 후 다섯 개의 싸이클 동안 유휴상태에 놓여있을 것이다.

[0075] 도 7의 방법에서, 전력 소모비율(712)을 설정하는 것(710)은 유휴 싸이클 기간(730)을 설정하는 것을 포함한다. 유휴 싸이클 기간(730)을 설정하는 것은 각각의 계산 노드들내의 컴퓨터 메모리가 유휴 싸이클 기간(730)을 갖도록 구성시킴으로써 수행될 수 있다. 유휴 싸이클의 기간이 클수록, 계산 노드내의 컴퓨터 프로세서 또는 컴퓨터 메모리는 보다 오랫동안 이와 같은 유휴 싸이클 동안에 유휴상태에 있게된다. 도 7의 예시에서, 유휴 싸이클 기간(730)은 10마이크로초로 설정된다. 즉, 컴퓨터 프로세서 또는 컴퓨터 메모리가 유휴상태로 있는 각각의 싸이클은 10마이크로초동안 지속된다.

[0076] 도 7의 예시에서, 컴퓨터 프로세서 전력 소모비율(714)은 컴퓨터 메모리 전력 소모비율(716)과 동일할 수 있다. 즉, 사용자는 전력 관리 속성(708)의 일부로서 두 개의 전력 소모비율들(714, 716)에 대한 단일 비율을 제공할 수 있다. 이와 달리, 사용자는 각각의 전력 소모비율마다 서로 다른 비율을 선택할 수 있다. 이와 같은 경우, 컴퓨터 프로세서 전력 소모비율(716)은 컴퓨터 메모리 전력 소모비율(714)과 동일하지 않다. 사용자가 전력 소모비율들(714, 716)을 구별시켜 설정할 수 있도록 하는 것은 병렬 컴퓨터에서의 전력 소모의 정확한 제어를 가능하게 해준다.

[0077] 도 7의 방법은 또한, 서비스 노드(116)에 의해, 병렬 컴퓨터의 연산 그룹(132)의 계산 노드(102)상의 작업(704)을 개시하는 것(726)을 포함한다. 계산 노드(102)상의 작업(704)을 개시하는 것(726)은 연산 그룹(132)내의 각각의 계산 노드(102)가 자신의 작업 특유적 컴퓨터 프로그램 명령을 갖도록 구성하고 컴퓨터 프로그램 명령을 실행하도록 하는 통지를 연산 그룹(132)내의 계산 노드(102)에게 보냄으로써 수행될 수 있다.

[0078] 추가적인 설명을 위해, 도 8은 본 발명의 실시예에 따른 병렬 컴퓨터에서의 사전적 전력 관리를 위한 추가적인 예시적 방법을 도시하는 흐름도를 나타낸다. 도 8의 방법은, 서비스 노드(116)에 의해, 병렬 컴퓨터내의 계산 노드(102)의 연산 그룹(132)상의 작업(704)을 개시하도록 하는 사용자 명령(706)(이것은 계산 노드(102)에 대한 전력 관리 속성(708)을 포함한다)을 수신하고(702); 연산 그룹(132)의 계산 노드(102)에 대한 전력 관리 속성(708)에 따라 서비스 노드(116)에 의해, 컴퓨터 프로세서 전력 소모비율(716)과 컴퓨터 메모리 전력 소모비율(714)을 포함하는 연산 그룹(132)의 각각의 계산 노드(102)에 대한 전력 소모비율(712)을 설정하며(710); 서비스 노드(116)에 의해, 병렬 컴퓨터의 연산 그룹(132)의 계산 노드(102)상의 작업(704)을 개시하는 것(726)을 포함한다는 점에서, 도 8의 방법은 도 7의 방법과 유사하다.

[0079] 도 8은 또한, 사용자(128)에 의해, 전력 관리 속성(708)을 설정하는 것(802)을 포함한다는 점에서, 도 8의 방법은 도 7의 방법과 다르다. 도 8의 방법에서, 전력 관리 속성(708)을 설정하는 것(802)은 적어도 하나의 이전 작업 실행(804) 동안에 병렬 컴퓨터의 연산 그룹(132)의 계산 노드(102)의 전력 소모를 모니터링함(802)으로써 수행된다. 전력 관리 속성은 작업의 실행 동안에 병렬 컴퓨터에 의해 소모되는 전력량을 미리결정된 문턱값 미만의 양으로 감소시키도록 설정된다. 병렬 컴퓨터의 전력 소모를 사전적으로 관리하는 시스템에서의 이와 같은 미

리결정된 문턱값은 일반적으로 피크 전력, 즉 전기 공급자가 소비자에게 높은 요금을 청구하는 전력량이다. 비록 본 명세서에서는 단일 단계로서 전력 소모비율을 설정하는 것을 설명하였지만, 전력 관리 속성(708)을 설정하기 전의 수 많은 작업 실행 동안 사용자가 계산 노드의 전력 소모를 모니터링할 수 있음을 본 발명분야의 독자라면 즉시 이해할 것이다.

[0080] 병렬 컴퓨터에서의 사전적 전력 관리의 완전한 기능적 컴퓨터 시스템의 배경으로 본 발명의 예시적인 실시예가 개략적으로 설명되었다. 하지만, 본 발명은 또한 임의의 적합한 데이터 프로세싱 시스템과의 이용을 위한 신호 수록 매체상에 배치된 컴퓨터 프로그램 제품으로 구현될 수 있음을 본 발명분야의 독자라면 이해할 것이다. 이와 같은 신호 수록 매체는 자기 매체, 광학 매체, 또는 기타의 적절한 매체를 포함하여, 머신 판독가능 정보를 위한 전송 매체 또는 기록가능 매체일 수 있다. 기록가능 매체의 예에는 하드 드라이브 또는 디스켓내의 자기 디스크, 광학 드라이브를 위한 컴팩트 디스크, 자기 테이프, 및 본 발명분야의 당업자에게 떠오르는 기타의 것들이 포함된다. 전송 매체의 예에는 음성 통신을 위한 전화 네트워크 및 예컨대 이더넷$_{TM}$과 같은 디지털 데이터 통신 네트워크 및 인터넷 프로토콜 및 월드 와이드 웹과 통신하는 네트워크뿐만이 아니라 예컨대 IEEE 802.11계 규격에 따라 구현된 네트워크를 포함한다. 적합한 프로그래밍 수단을 갖춘 임의의 컴퓨터 시스템은 프로그램 제품으로 구현된 본 발명의 방법의 단계들을 실행할 수 있을 것임을 본 발명분야의 당업자는 즉시 이해할 것이다. 비록 본 명세서에서 설명된 몇몇 예시적인 실시예들은 컴퓨터 하드웨어상에서 소프트웨어 설치되고 실행되는 것에 맞춰지고 있지만, 펌웨어 또는 하드웨어로서 구현된 대안적인 실시예들이 본 발명의 범위내에 충분히 속해있음을 본 발명분야의 당업자는 즉시 이해할 것이다.

[0081] 본 발명의 본질적인 사상으로부터 벗어나지 않고 본 발명의 다양한 실시예들에서 변형과 변경이 행해질 수 있음을 전술한 설명으로부터 이해할 것이다. 본 명세서에서의 설명은 단지 설명을 위한 것이며 한정적인 의미로 해석되어서는 안된다. 본 발명의 범위는 이하의 청구항들의 어구에 의해서만 한정된다.

[0082] 의혹을 피하기 위해, 본 명세서에서의 상세한 설명과 청구항들에서 사용된 용어 "... 포함한다"는 "... 만으로 구성된다"를 의미하는 것으로서 해석되어서는 안된다.

부호의 설명

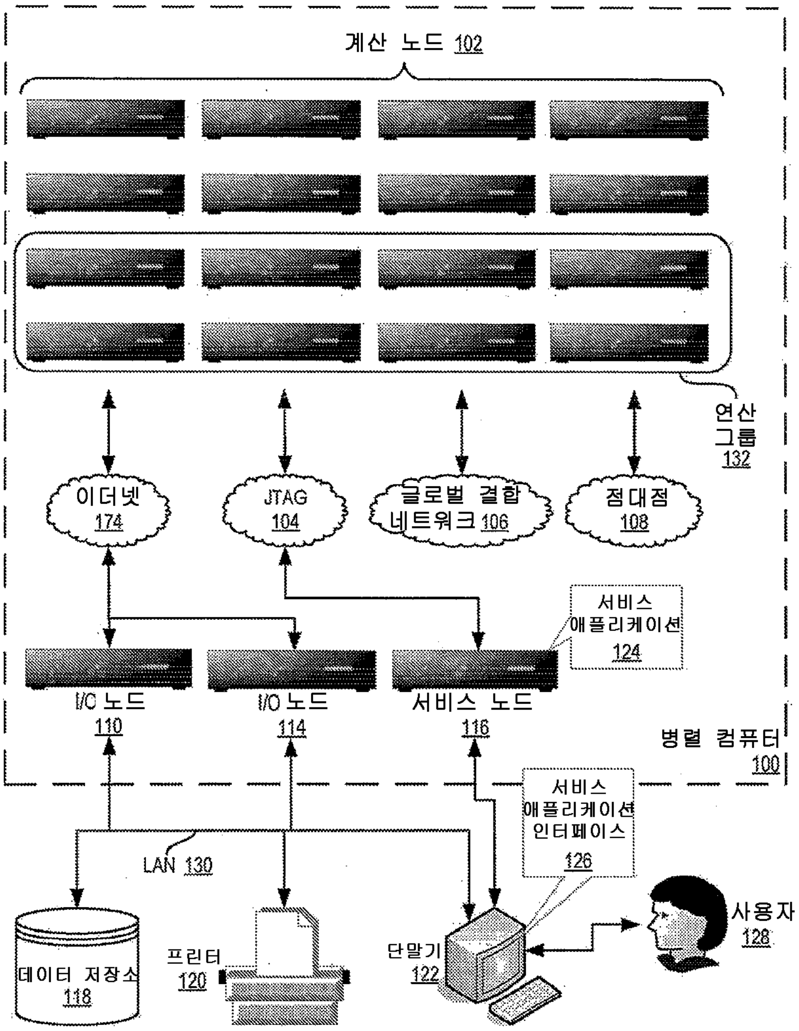
[0083]

102: 계산 노드,	132: 연산 그룹
174: 이더넷,	106: 글로벌 결합 네트워크
108: 점대점,	124: 서비스 애플리케이션
110, 114: I/O 노드,	116: 서비스 노드
100: 병렬 컴퓨터,	126: 서비스 애플리케이션 인터페이스
128: 사용자,	118: 데이터 저장소
120: 프린터,	122: 단말기
152: 계산 노드,	164: 프로세서
154: 메모리 버스,	194: 버스 아답터
168: 확장 버스,	158: 애플리케이션
160: 메세징 모듈,	712: 전력 소모비율('PCR')
714: 컴퓨터 프로세서 PCR,	716: 컴퓨터 메모리 PCR
162: 오퍼레이팅 시스템,	172: 이더넷 아답터
176: JTAG 슬레이브,	180: 점대점 아답터
188: 글로벌 결합 네트워크 아답터,	174: 기가비트 이더넷
178: JTAG 마스터,	108: 점대점 네트워크
190: 자식 노드,	192: 부모 노드
106: 일괄적 연산 네트워크,	252: 서비스 노드

256: 프로세서,
262: 전면측 버스
264: 비디오 버스,
280: 디스플레이 장치
209: 비디오 아답터,
258: 버스 아답터
266: 메모리 버스,
124: 서비스 애플리케이션
706: 사용자 명령,
708: 전력 관리 속성
254: 오퍼레이팅 시스템,
260: 확장 버스
267: 통신 아답터,
278: I/O 아답터
272: 디스크 드라이브 아답터,
178: JTAG 마스터
200: 네트워크,
282: 기타 컴퓨터
281: 사용자 입력 장치,
270: 데이터 저장소
176: JTAG 슬레이브,
180: 점대점 아답터
152: 계산 노드,
190: 자식 노드
105: 메시,
107: 원환체
103: 링크,
103: 링크
202: 물리적 루트,
250: 랭크
204: 브랜치 노드,
206: 리프 노드

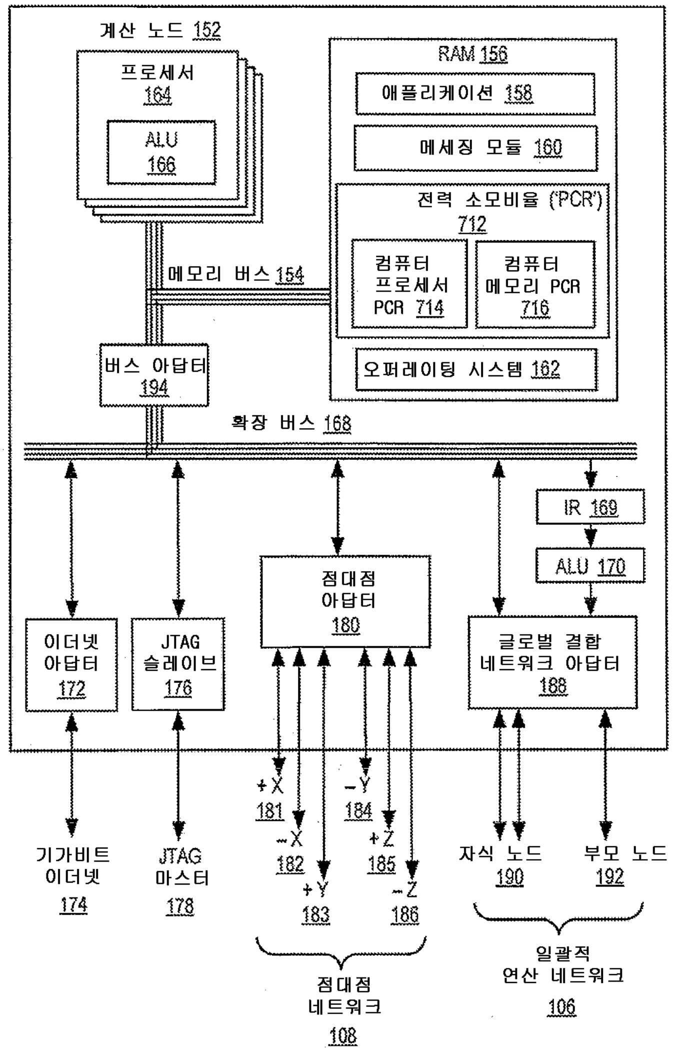
도면

도면1



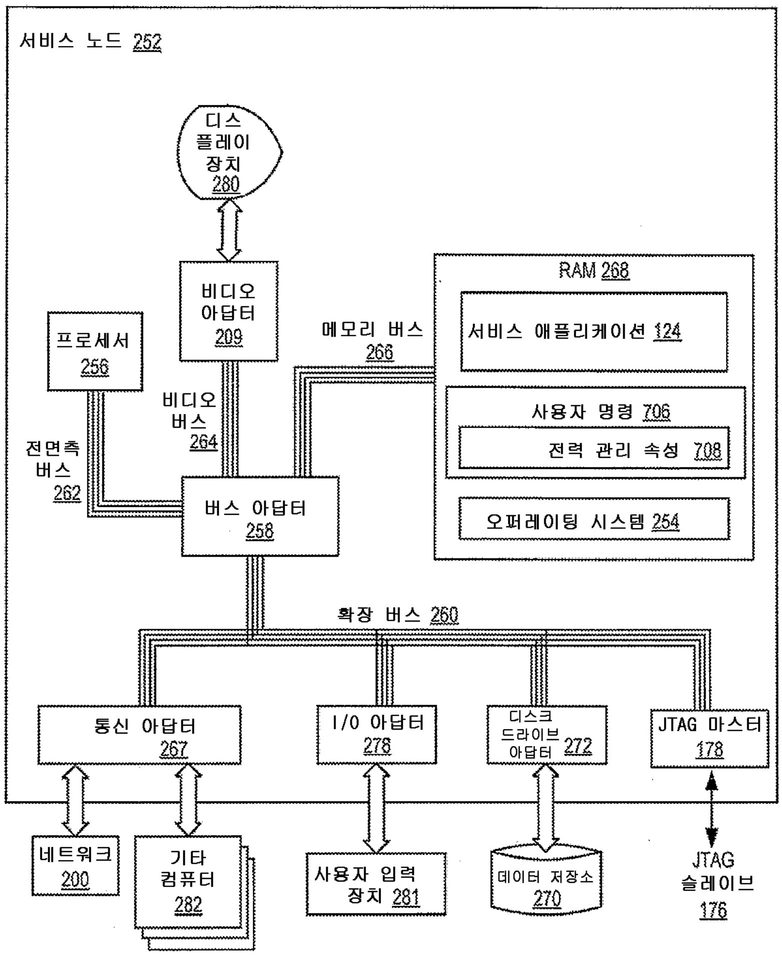
계산 노드 102
연산 그룹 132
이더넷 174
JTAG 104
글로벌 결합 네트워크 106
점대점 108
서비스 애플리케이션 124
I/O 노드 110
I/O 노드 114
서비스 노드 116
병렬 컴퓨터 100
서비스 애플리케이션 인터페이스 126
LAN 130
사용자 128
프린터 120
단말기 122
데이터 저장소 118

도면2

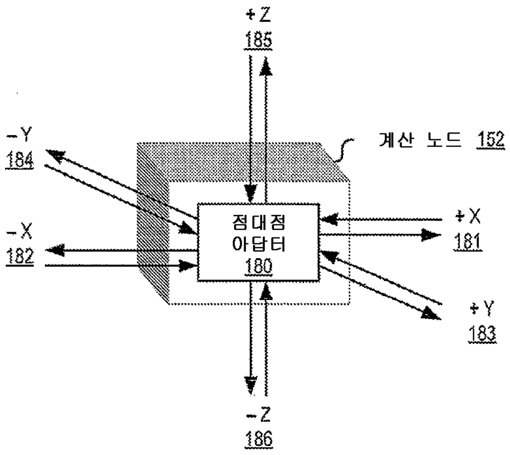


계산 노드 152
프로세서 164
ALU 166
RAM 156
애플리케이션 158
메세징 모듈 160
전력 소모비율 ('PCR') 712
컴퓨터 프로세서 PCR 714
컴퓨터 메모리 PCR 716
메모리 버스 154
버스 아답터 194
오퍼레이팅 시스템 162
확장 버스 168
IR 169
ALU 170
점대점 아답터 180
이더넷 아답터 172
JTAG 슬레이브 176
글로벌 결합 네트워크 아답터 188
+X 181
-X 182
+Y 183
-Y 184
+Z 185
-Z 186
기가비트 이더넷 174
JTAG 마스터 178
자식 노드 190
부모 노드 192
일괄적 연산 네트워크 106
점대점 네트워크 108

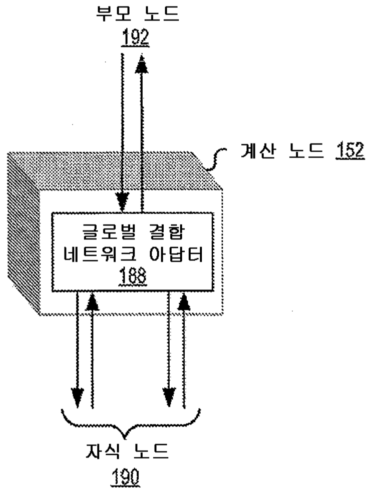
도면3


서비스 노드 252
디스 플레이 장치 280
비디오 아답터 209
프로세서 256
비디오 버스 264
전면측 버스 262
메모리 버스 266
RAM 268
서비스 애플리케이션 124
사용자 명령 706
전력 관리 속성 708
오퍼레이팅 시스템 254
버스 아답터 258
확장 버스 260
통신 아답터 267
I/O 아답터 278
디스크 드라이브 아답터 272
JTAG 마스터 178
네트워크 200
기타 컴퓨터 282
사용자 입력 장치 281
데이터 저장소 270
JTAG 슬레이브 176

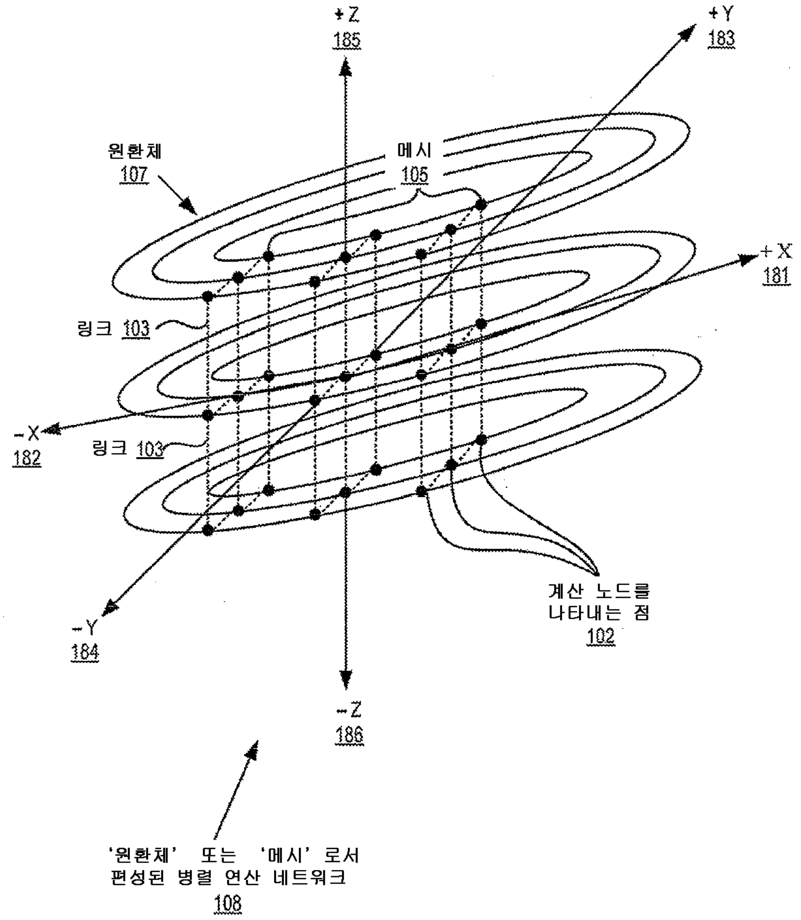
도면4a


+Z 185
-Y 184
계산 노드 152
-X 182
점대점 아답터 180
+X 181
+Y 183
-Z 186

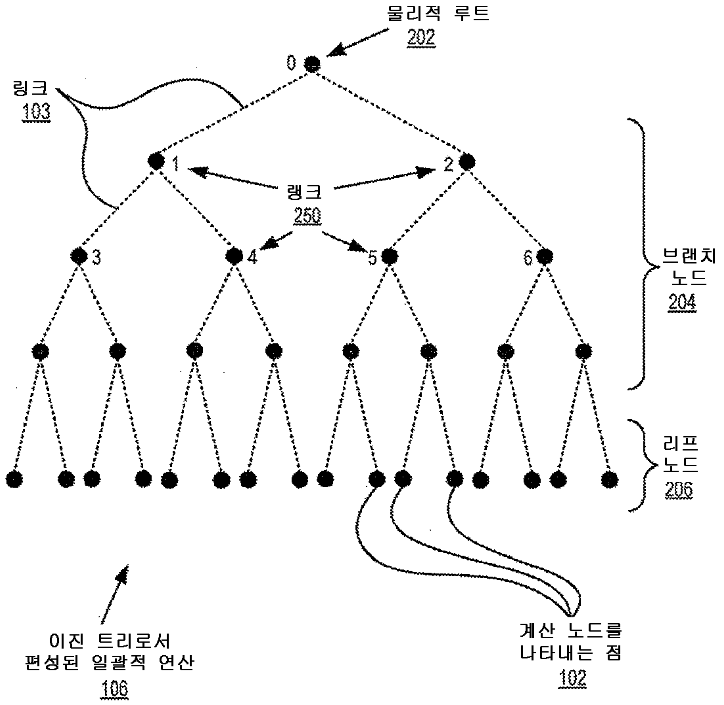
도면4b


부모 노드
192
계산 노드 152
글로벌 결합
네트워크 아답터
188
자식 노드
190

도면5

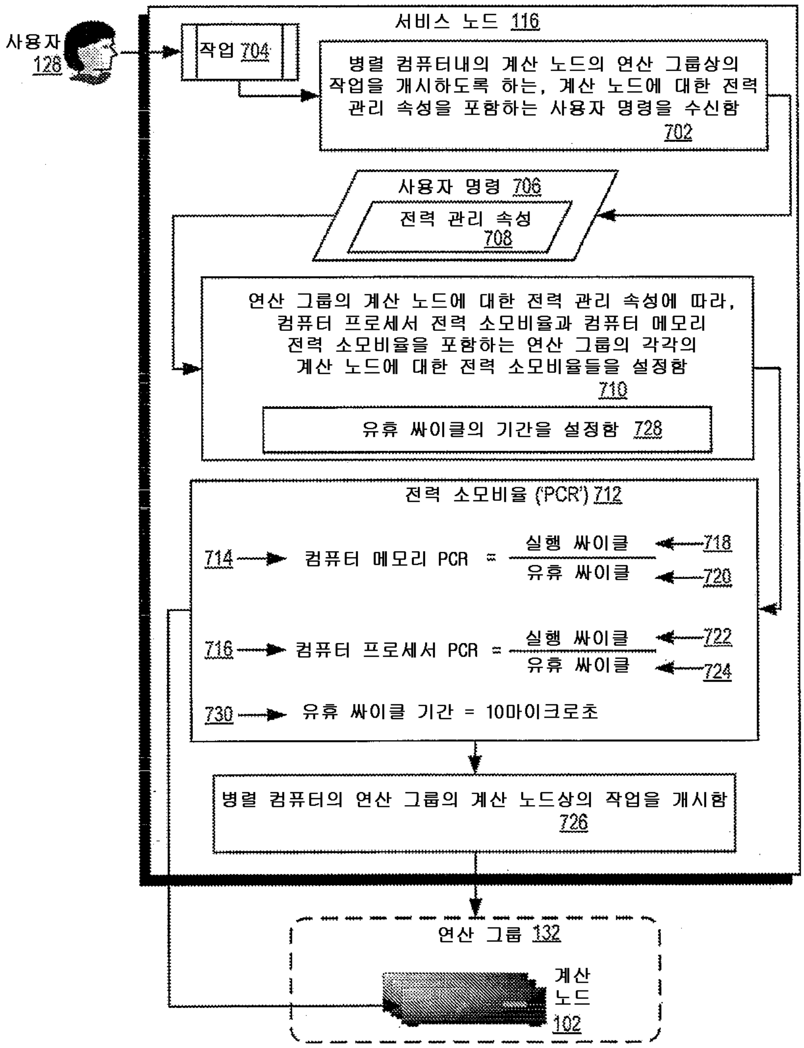

+Z
185
+Y
183
원환체
107
메시
105
+X
181
링크 103
-X
182
링크 103
계산 노드를
나타내는 점
102
-Y
184
-Z
186
'원환체' 또는 '메시'로서
편성된 병렬 연산 네트워크
108

도면6

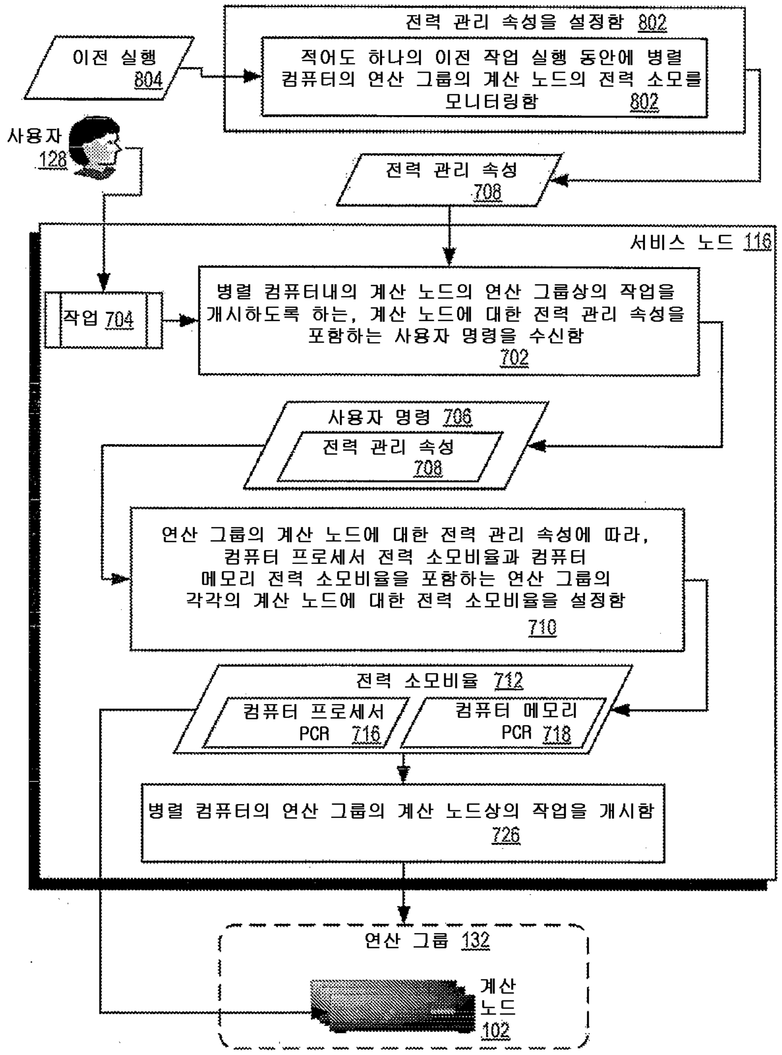


물리적 루트
202
0
링크
103
1
2
랭크
250
3
4
5
6
브랜치
노드
204
리프
노드
206
이진 트리로서
편성된 일괄적 연산
106
계산 노드를
나타내는 점
102

도면7


사용자 128
서비스 노드 116
작업 704
병렬 컴퓨터내의 계산 노드의 연산 그룹상의 작업을 개시하도록 하는, 계산 노드에 대한 전력 관리 속성을 포함하는 사용자 명령을 수신함 702
사용자 명령 706
전력 관리 속성 708
연산 그룹의 계산 노드에 대한 전력 관리 속성에 따라, 컴퓨터 프로세서 전력 소모비율과 컴퓨터 메모리 전력 소모비율을 포함하는 연산 그룹의 각각의 계산 노드에 대한 전력 소모비율들을 설정함 710
유휴 싸이클의 기간을 설정함 728
전력 소모비율 ('PCR') 712
714
컴퓨터 메모리 PCR = 실행 싸이클 / 유휴 싸이클
718
720
716
컴퓨터 프로세서 PCR = 실행 싸이클 / 유휴 싸이클
722
724
730
유휴 싸이클 기간 = 10마이크로초
병렬 컴퓨터의 연산 그룹의 계산 노드상의 작업을 개시함 726
연산 그룹 132
계산 노드 102

도면8



전력 관리 속성을 설정함 802
이전 실행 804
적어도 하나의 이전 작업 실행 동안에 병렬 컴퓨터의 연산 그룹의 계산 노드의 전력 소모를 모니터링함 802
사용자 128
전력 관리 속성 708
서비스 노드 116
작업 704
병렬 컴퓨터내의 계산 노드의 연산 그룹상의 작업을 개시하도록 하는, 계산 노드에 대한 전력 관리 속성을 포함하는 사용자 명령을 수신함 702
사용자 명령 706
전력 관리 속성 708
연산 그룹의 계산 노드에 대한 전력 관리 속성에 따라, 컴퓨터 프로세서 전력 소모비율과 컴퓨터 메모리 전력 소모비율을 포함하는 연산 그룹의 각각의 계산 노드에 대한 전력 소모비율을 설정함 710
전력 소모비율 712
컴퓨터 프로세서 PCR 716
컴퓨터 메모리 PCR 718
병렬 컴퓨터의 연산 그룹의 계산 노드상의 작업을 개시함 726
연산 그룹 132
계산 노드 102