Earth Simulator: uma visão geral

Tiago J. Carvalho – RA:087343
Instituto de Computação
Universidade Estadual de Campinas
Campinas, São Paulo, Brasil
tiagojc@gmail.com

RESUMO

Esse trabalho traz uma visão do que é o "Earth Simulator" uma proposta de máquina com tecnologia vetorial que tem o objetivo de possibilitar a previsão de mudanças globais no ambiente. Dono de um sistema computacional com poder único, tem atraido pesquisadores do mundo inteiro. Dono de uma performance de execução de 35.86TFLOPS é considerado o computador mais rápido do mundo. Possui um hardware baseado em um sistema paralelo de memória distribuída e tem seu sistema operacional baseado no UNIX. Sempre utilizado simulações do meio ambiente, tem produzidos resultados únicos para simulações nas áreas de circulação oceânica global e reprodução de campos de ondas sísmicas, dentre outras.

Palavras chave

Earth Simulator, supercomputador, NEC, processamento vetorial

1. INTRODUÇÃO

O "Earth Simulator" é uma proposta de máquina especial feita pelo NEC com o mesmo tipo de tecnologia vetorial que a disponível no SX-6. O NEC é uma das maiores produtoras de computadores e eletrônicos no mundo. Ele é o segundo maior produtor de semicondutores (a Intel é a primeira) [1]. Ele também produz PCs e notebooks, controlando metade do marketing no japão. A decisão do NEC de se basear completamente em processadores vetoriais é uma nova tendência seguida no design de super computadores [2]. O "Earth Simulator" foi terminado em fevereiro de 2002 depois de cinco anos de desenvolvimento tendo a mais recente arquitetura na linha de computadores vetoriais, linha essa que começou com o Cray 1.

O objetivo do "Earth Simulator System (ESS)" é possibilitar a previsão de mudanças globais no ambiente devido ao seu enorme poder computacional, poder esse requerido em tais tipos de pesquisas [6]. O assunto de mudanças ambientais tem se tornado mais e mais importante com o passar das décadas, já que a poluição tem se tornado um sério problema que ameaça o clima do planeta em que vivemos. E o clima é exatamente onde o ESS vai dar suporte em pesquisas.

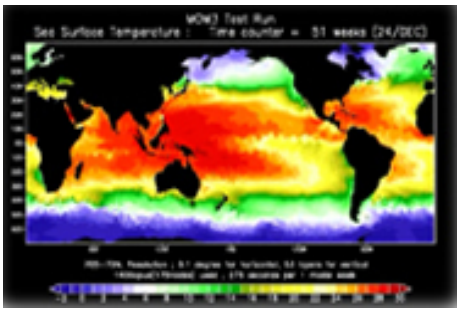

Figura 1: Tela do ESS

A possibilidade do uso de um sistema computacional com poder único tem atraído pesquisadores do mundo inteiro e estimulado o interesse na realização de projetos conjuntos na área das ciências e engenharia. Um número de pesquisas cooperativas entre o Japão e países europeus bem como os E.U.A. estão em andamento, resultando em grandes avanços na ciência e abrindo caminho para novos aspectos no ramo das simulações científicas.

2. DETALHES TÉCNICOS

Do ponto de vista da ciência da computação, os seguintes pontos tem que ser notados. Primeiro, o ESS demonstra a possibilidade e significancia de uma grande arquitetura paralela para computadores vetoriais. Segundo, o ESS tem empurrado o estado-da-arte nas linguagens de programação e facilita a questão do desenvolvimento de aplicações por suportar linguagens de alto-nível. Em terceiro lugar, a performance sustentável do "Earth Simulator" é 1000 vezes maior que o supercomputador mais usado em 1996 no campo de pesquisas climáticas. Sua performance de execução de 35.86TFLOPS foi aprovada pelo Linpack benchmark e o ESS foi registrado como o super computador mais rápido do mundo em 20 de junho de 2002 [5].

3. ARQUITETURA DO HARDWARE

O "Earth Simulator" é um sistema paralelo de memória distribuída que consiste de 640 nodos processadores (PNs) conectado por um barramento compartilhado de 640 x 640

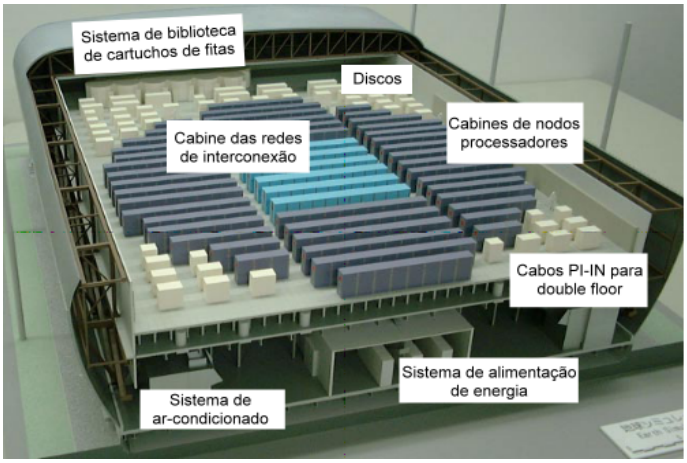

Figura 2: Um modelo do ESS. O local que abriga o ESS tem 50m x 65m x 17m e tem dois andares, além de incluir um sistema de isolação sísmica.

nodos internos [5]. Cada nodo é um sistema de memória compartilhado composto de de oito processadores aritméticos (APs), um sistema de memória compartilhada de 16GB, uma unidade de controle remoto (RCU), e um controlador de entrada e saída. O máximo de performance de cada AP é 8GFLOPS. O número total de processadores é 5120 e o máximo total de performance e capacidade total de memória são 40TFLOPS e 10TB, respectivamente. As demais especificações do hardware são mostradas na tabela 1.

Tabela 1: *A especificação de Hardware do Earth Simulator*

Número total de nodos processadores	640
Número de AP para cada nodo	8
Número Total de AP	5120
Máximo de performance para cada AP	8GFLOPS
Máximo de performance para cada PN	64GFLOPS
Máximo de performance do sistema total	40TFLOPS
Memória compartilhada de cada PN	16GB
Memória principal total	10TB

3.1 Processador

Cada AP contém uma unidade vetorial (VU), uma unidade de operações super-escalares (SU), e uma memória principal acessada pela unidade de controle, os quais são montados em um chip LSI operando numa freqüência de clock de 500MHz, parcialmente 1GHz.

A SU é um processador 4-way super escalar com uma cache de instruções de 64KB, uma cache de dados de 64KB, e 128 registradores escalares de propósito geral. A SU utiliza a previsão de "branch", "prefetching" de dados e execução de instruções "out-of-order".

A VU possui uma pipeline vetorial que pode manipular seis tipos de operações (add/shift, multiplicação, divisão, lógica, máscara e load/store), além de um haver um total de 8 registradores vetoriais e 64 registradores vetoriais de dados, cada um dos quais tem 256 elementos vetoriais.

Quando a adição e a multiplicação são realizadas concorrentemente, o máximo de performance será alcançado. Aqui o máximo de performance teórico pode ser computado por vários operadores padrões. Cada AP carrega 16 operações de ponto flutuante com um ciclo de máquina de 2 nseg pela utilização completa de um conjunto de pipelines - permite que ambos adição e multiplicação rode 8 operações. Isso indica 8 Gflops como máximo de performance, mas um vetor de comprimento suficientemente longo e uma alta velocidade na taxa transferência de dados entre o processador e a unidade de memória é essencial para garantir essa performance.

Dado que um elemento do vetor tem 8 bytes então, 8Gflops de performance requerem 64GB/s de taxa de transferência de dados, enquanto a taxa de transferência de dados com a unidade de memória é 32 GB/s para um processador. Isso implica que uma aplicação tem que ser computada intensivamente enquanto a freqüência do acesso à memória tem que ser mantida baixa para uma computação eficiente. A performance teórica máxima "F" pode ser derivada do número de instruções dentro de um laço "DO" com base na seguinte fórmula assumindo um vetor de comprimento suficientemente longo.

$$F = \frac{4\,(ADD + MUL))}{max\,(ADD, MUL, VLD + VST)} \qquad (1)$$

onde ADD, MUL, VLD, e VST são o número de adições, multiplicações e operações vetoriais de "load" e "store" respectivamente. A tabela 3 sumariza como a performance sustentável é computada para diferentes operações padrão. Aqui o índice do laço "DO" mais interno é denotado como "n" e o do laço externo como "j". Note que o acesso aos vetores C e D é tomado como "load" escalar em termos de "n", e não é incluído na contagem de VLD.

No.	DO loop	VLD	VST	ADD	MUL	Gflops	Taxa Máxima
1	X(n) = X(n) + A(n)	2	1	1	0	4/3=1.33	17 %
2	X(n) = X(n) + A(n)*B(n)	3	1	1	1	8/4=2.00	25 %
3	X(n) = X(n) + P(n, j)*C(n)	2	1	1	1	8/3=2.67	33 %
4	X(n) = X(n) + P(n, j)*C(j) + P(n, j+1)*C(j+1) + P(n, j+2)*C(j+2) + P(n, j+3)*C(j+3) Y(n) = Y(n) + P(n, j)*D(j) + P(n, j+1)*D(j+1) + P(n, j+2)*D(j+2) + P(n, j+3)*D(j+3)	6	2	8	8	64/8=8.00	100 %

Figura 3: Estimação de performance de vetores

3.2 O Sistema de Memória

O sistema de memória (MS) em cada PN é simetricamente compartilhado pelos 8 APs. Isso é configurado por 32 unidades de memória principal (MMU) como 2048 bancos. Cada AP em um PN pode acessar 32 portas de memória quando um vetor de instruções "load/store" é carregado [8]. Cada AP tem uma taxa de transferência de dados de 32 GB/s com os dispositivos de memória, o que resulta em um "throughput" agregado de 256 GB/s por PN. Assim sendo, uma série de acessos a endereços de memória contínuos resultará na performance máxima, enquanto isso será reduzido particularmente com o acesso a um número constante de intervalos de memória porque o número de portas de memória disponível para memória para acesso concorrente é diminuído. Logo, o acesso à memória em um dos 32 endereços com um grande espaçamento para outro endereço irá resultar em uma diminuição da eficiência de 1 GB/sec. O RCU em cada PN é diretamente conectado a um barramento compartilhado de dois modos, enviando e recebendo, e controlando a comunicação de dados entre os nodos. A capacidade total de memória é de 10 TB

3.3 Rede de Interconexão

A rede de interconexão (IN) consiste de duas partes: uma é a unidade de controle de barramento entre nodos (XCT) que coordena as operações de switch; o outro é um barramento entre nodos de "switch"(XSW) que funciona como um caminho real para dados. XSW é composto de 128 switches separados, cada um dos quais tem uma banda de 1Gbits/s operações independentemente. Qualquer par desses switches e nodos são conectados por cabos elétricos. A taxa de transferência de dados teóricos entre os pares de PNs é 12.3GB/s. Um modelo de programação para trocas de mensagens com uma biblioteca de Interface de Passagem de Mensagens (MPI) é suportado pelos PNs e dentro deles. A performance da função Put da MPI foi medida. O throughput máximo e a latência da função Put da MPI são 11.63GB/s e 6.63 μseg, respectivamente. Também deveria ser notado que o tempo para sincronização de barreiras é apenas 3.3 μseg por causa do sistema de hardware dedicado para sincronização de barreiras dentro dos PNs.

4. SISTEMA OPERACIONAL E LINGUAGENS

O sistema operacional que roda em um nodo processador é um sistema baseado no UNIX para suportar computações científicas de larga escala. Compiladores de Fortran90, Fortran de Alta Performance (HPF), C e C++ são disponibilizados com suporte automático para vetorização e paralelismo. Uma biblioteca de transmissão de mensagens baseada em MPI-1 e MPI-2 também está disponível.

Em especial, é focado o HPF/ES por ser desenvolvido especialmente para o ES. Ele é um compilador HPF desenvolvido para o Earth Simulator pelo melhoramento em muitos aspectos do HPF/SX V2 feito pelo NEC. Possui algumas extensões únicas bem como algumas características do HPF 2.0, suas extensões aprovadas e HPF/JA para suportar uma programação paralela eficiente e portátil. As extensões únicas incluem diretivas de vetorização, otimização de características de comunicação irregular, entrada e saída paralela, e assim por diante. HPF/ES detecta comunicações do tipo SHIFT automaticamente e gera chamadas para rotinas em tempo de execução altamente otimizadas do tipo SHIFT de comunicação. No mais, a geração de "schedule" de comunicação é re-usada quando o mesmo padrão de comunicação é iterado na execução de laços. O ESS tem ainda uma hierarquia de três níveis de paralelismo que serão apresentadas nas próximas sub-seções:

4.1 Paralelismo de Nível 1

O primeiro nível de processamento paralelo é o processamento vetorial em um AP individual. Esse é o mais fundamental nível de processamento no ES. Vetorização automática é aplicada por compiladores para programas escritos em linguagens convencionais como Fortran 90 e C [9].

4.2 Pararelismo de Nível 2

O segundo nível é o de processamento paralelo com memória compartilhada processado em um PN individual. Programação paralela com memória compartilhada é suportada por "microtasking" e "OpenMP". A capacidade do "microtasking" é similar em estilo àquelas fornecidas por um supercomputador Cray, e as mesmas funções são realizadas pelo ES. "Microtasking" é aplicado de duas formas: uma (AMT) é a paralelização automática pelos compiladores e a outra

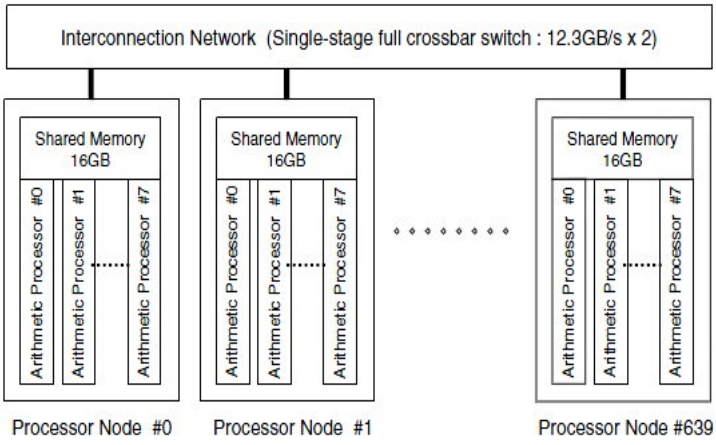

Figura 4: Esquema simplificado do ES

(MMT) é a inserção manual de diretrizes de "microtasking" antes do alvo dos loops.

4.3 Pararelismo de Nível 3

O terceiro nível é o processamento paralelo de memória distribuída que é compartilhado dentro das PNs. Um modelo de programação paralela com memória distribuída é suportado pela MPI. A performance desse sistema para funções de entrada (Put) da MPI foi medida com funções especificadas para o MPI-2.

5. APLICAÇÕES

Para aplicações práticas com performance média sustentável de 30% do máximo, o "Earth Simulator" tem produzido resultados únicos para simulações como "Kuroshio", "Gulf Stream" e "Agulhas Ring" na circulação oceânica global; furacões e início de raios na circulação atmosférica global; reprodução de campos de ondas sísmicas de toda terra na física de terra sólida; material de super-diamante de nano tubos de carbono em materiais científicos [7].

A figura 5 mostra os parâmetros de performance sustentável dos projetos rodando no Earth Simulator em 2002. A maior das performances foi condecorada com o prêmio "Gordon Bell" em 2002.

5.1 Simulação de terremotos

Para modelar a propagação de ondas sísmicas resultantes de grandes terremotos no Earth Simulator foram utilizados 1944 processadores segundo a implementação descrita em [3]. Esse tipo de simulação é baseado sobre o método de elementos espectrais, uma técnica com alto grau de elementos finitos com uma grande quantidade de matrizes diagonal exatas. É usada uma grande rede com 5.5 bilhões de pontos de rede (14.6 bilhões de graus de liberdade). Para isso é utilizada a complexidade total da Terra, isto é, uma velocidade de onda tridimensional e estrutura densa, um modelo de crosta 3D eliptico, bem como a topografia. Um total de 2.5 terabytes de memória é necessário. Tal implementação é puramente baseada sobre o MPI, com a vetorização de loops em cada processador. Obteve uma excelente taxa de vetorização de 99.3%, e foi alcançada uma performance de 5 teraflops (30% da performance máxima) em 38% da máquina. A resolução mais alta da malha permite a geração completa de calculos tridimensionais em períodos sísmicos com menos de 5 segundos.

5.2 Simulação de fluidos tridimensional para fusão científica

Dentre vários experimentos testados no "Earth Simulator", um código para simulação de plasma (IMPACT-3D) paralelizado com HPF, rodando em 512 nodos do "Earth Simulator" alcançou uma performance de 14.9 TFLOPS [5]. A performance máxima teórica dos 512 nodos foi de 32 TFLOPS, o qual representa 45% do máximo de performance que foi obtido com HPF. IMPACT-3D é um código de análise de implosões utilizando o esquema TVD, o qual apresenta uma compressão tridimensional e uma computação de fluido "Euleriano" com o esquema de cópia explícito de 5 pontos para a diferenciação espacial e o passo de tempo fracional para a integração de tempo.O tamanho da malha é 2048x2048x4096, e a terceira dimensão foi distribuída para a paralelização. O sistema HPF utilizado na avaliação é HPF/ES, desenvolvido para o "Earth Simulator" através do melhoramento do NEC HPF/SX V2 principalmente na escalabilidade de comunicação. A comunicação foi manualmente ajustada para dar a melhor performance utilizando extensões HPF/JA, os quais foram projetados para dar aos usuários um maior controle sobre paralelismo sofisticado e comunicações otimizadas.

5.3 Simulação de Turbulência

Outro tipo de simulação executada no "Earth Simulator" foram as simulações numéricas diretas de alta resolução (DNSs) de turbulência sem compressão, com número de pontos de "grid" superior a 4096^3 [9].As DNSs são baseadas nos métodos do espectro de Fourier, tal que as equações para conservação de massa são precisamente resolvidas. Nas DNSs baseadas no método espectral, a maior parte do tempo computacional é consumido calculando a transformada rápida de Fourier (FFT) em três dimensões (3D), a qual requer uma escala gigantesca de transferência de dados e tem sido o maior impecilho para uma computação de alta performance. Implementando novos métodos para possibilitar a 3D-FFT no ESS, a DNS alcançou 16.4 Tflops em 2048^3 pontos de "grid".

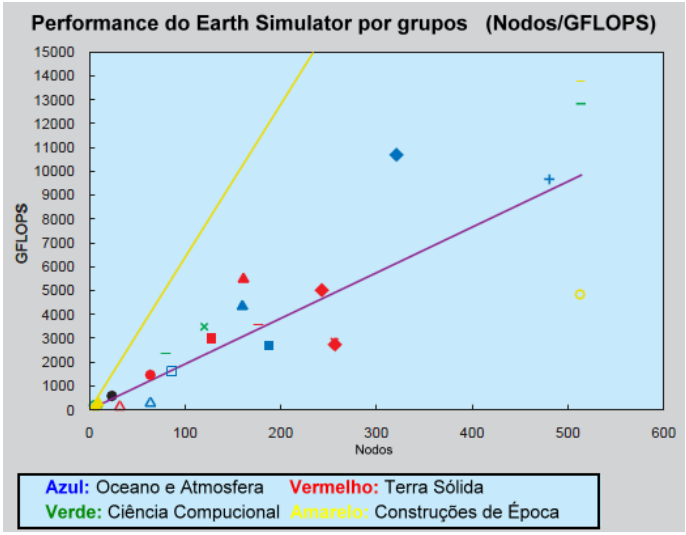

Figura 5: Perfomance teórica e sustentável do Earth Simulator. A linha amarela indica o máximo teórico e a linha rosa indica 30% da performance sustentável. As três maiores foram condecoradas com o prêmio Gordon Bell em 2002.

A DNS ainda produz um espectro de energia exibido em um completo subdomínio inerte, em contraste com as DNSs anteriores com baixa resolução e então fornece dados valiosos para o estudo de características universais de turbulência nos maiores números de Reynold.

5.4 Simulação da Atmosfera Global

Um modelo geral de circulação espectral atmosférico chamado AFES (AGCM para o "Earth Simulator") foi desenvolvido e otimizado para a arquitetura do Earth Simulator (ES) [8]. A performance sustentável de 26.58 Tflops foi conseguida para uma simulação de alta resolução (T1279L96) com AFES utilizando uma configuração com todos os 640 nodos do ES. A eficiência de computação resultante é de 64.9% da performance máxima, bem maior que das aplicações convencionais de estado atmosférico (clima) que têm em torno de 25-50% de eficiência para computadores vetoriais paralelos. Essa excelente performance prova que a efetividade do ES é um significantemente viável para aplicações práticas.

6. TRABALHOS FUTUROS

No dia 12 de maio de 2008 a corporação NEC anunciou que está sendo contratada para construir o "Novo Earth Simulator", um sistema de computador ultra veloz para a Agência Japonesa para Ciência e Tecnologia de Terra e Mar (JAMSTEC) [4]. O novo ESS será uma atualização do Earth Simulator já existente e irá introduzir novas características para possibilitar uma precisão melhor e análise de alta velocidade, além de projeções em escala global dos fenômenos ambientais. O sistema também será usado para produzir simulações numéricas para campos de pesquisa avançados que vão além do escopo de outros sistemas computacionais.

O novo sistema principalmente consiste de um sistema de supercomputador principal, unidades de sub-sistema e um sistema de gerenciamento de operações, e é projetado para alcançar uma performance máxima de 131TFLOPS (TFLOPS: um trilhão de operações de ponto flutuante por segundo). A performance de aplicações eficazes deverá ser aumentada para duas vezes aquela conseguida com o Earth Simulator existente. O novo sistema está previsto para ser instalado no Instituto para Ciências da Terra (JAMSTEC) de Yokohama.

7. CONCLUSÕES

O Earth Simulator tem contribuído de maneira extremamente importante para pesquisas relacionadas a fenômenos ambientais tais como aquecimento global, poluição atmosférica e marinha, El Niño, chuvas torrenciais dentre outros, uma vez que ele é capaz de executar tais simulações com alta velocidade desde que os mesmos sejam especialmente preparados para rodar no ES. Tais fatos são de grande importância para o desenvolvimento de atividades econômicas e sociais, para ajudar a resolver problemas ambientais além de melhorar o entendimento (de estudiosos da área) de fenômenos terrestres tais como placas tectônicas e terremotos. Ele possui um hardware dedicado e muito eficiente, que conta principalmente com computação vetorial para melhorar ainda mais o tempo e confiabilidade das simulações nele executadas. Assim sendo, é um dos supercomputadores de maior importância (se não o mais) nos dias atuais.

8. REFERÊNCIAS

[1] Nec. Internet: http://www.webopedia.com/TERM/N/NEC.html (acessado em 21-05-2009), 2001.

[2] M. Y. chief designer of NEC. Earth simulator system. Internet: http://www.thocp.net/hardware/nec_ess.htm (acessado em 19-05-2009), 2002.

[3] D. Komatitsch, S. Tsuboi, C. Ji, and J. Tromp. A 14.6 billion degrees of freedom, 5 teraflops, 2.5 terabyte earthquake simulation on the earth simulator. In *SC '03: Proceedings of the 2003 ACM/IEEE conference on Supercomputing*, page 4, Washington, DC, USA, 2003. IEEE Computer Society.

[4] NEC. Nec awarded contract for new earth simulator system. Internet:

http://www.nec.co.jp/press/en/0805/1601.html (acessado em 16-06-2009), 2008.

[5] H. Sakagami, H. Murai, Y. Seo, and M. Yokokawa. 14.9 tflops three-dimensional fluid simulation for fusion science with hpf on the earth simulator. In *Supercomputing '02: Proceedings of the 2002 ACM/IEEE conference on Supercomputing*, pages 1–14, Los Alamitos, CA, USA, 2002. IEEE Computer Society Press.

[6] T. Sato. The earth simulator: roles and impacts. *Parallel Comput.*, 30(12):1279–1286, 2004.

[7] T. Satu. The current status of the earth simulator. Internet: http://www.jamstec.go.jp/esc/publication/journal/jes_vol.1/pdf/JES1-2.pdf (acessado em 16-06-2009).

[8] S. Shingu, H. Takahara, H. Fuchigami, M. Yamada, Y. Tsuda, W. Ohfuchi, Y. Sasaki, K. Kobayashi, T. Hagiwara, S.-i. Habata, M. Yokokawa, H. Itoh, and K. Otsuka. A 26.58 tflops global atmospheric simulation with the spectral transform method on the earth simulator. In *Supercomputing '02: Proceedings of the 2002 ACM/IEEE conference on Supercomputing*, pages 1–19, Los Alamitos, CA, USA, 2002. IEEE Computer Society Press.

[9] M. Yokokawa, K. Itakura, A. Uno, T. Ishihara, and Y. Kaneda. 16.4-tflops direct numerical simulation of turbulence by a fourier spectral method on the earth simulator. In *Supercomputing '02: Proceedings of the 2002 ACM/IEEE conference on Supercomputing*, pages 1–17, Los Alamitos, CA, USA, 2002. IEEE Computer Society Press.