OUKA
The University of Osaka
Institutional Knowledge Archive

Title	スーパーコンピュータと半導体
Author(s)	小田中, 紳二
Citation	サイバーメディアHPCジャーナル. 2012, 2, p. 3-6
Version Type	VoR
URL	https://doi.org/10.18910/70527
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

スーパーコンピュータと半導体

小田中　紳二
大阪大学 サイバーメディアセンター コンピュータ実験科学研究部門

1. はじめに

地球環境、情報、生命、ナノテクノロジーなどの科学技術分野において、様々な数理モデルによるシミュレーションを通して、その理解を深め、新たな知見を得る試みが大きく進展している。スーパーコンピューティングによる大規模シミュレーションの実現は科学技術分野に大きな影響を与えている。

近年におけるスーパーコンピューティング技術の進展は大きな波を描いている。1970 年代後半のベクトル型スーパーコンピュータの登場が第一期の大きな波といえる。そのスーパーコンピュータの斬新なデザインと当時においては驚くべき性能は、1980 年代における科学技術計算分野に大きなインパクトを与えた。さらに、先端 LSI 技術を導入した日本製ベクトル型スーパーコンピュータは半導体産業、自動車産業、航空機産業などの産業界にも急速に導入され実用化が進むと、米国では、著名な科学者を集めて「科学技術における大規模計算に関するパネル討論会」(1982)がもたれ、有名な Lax 報告[1]としてまとめられ、米国におけるスーパーコンピュータ技術の必要性に対する警鐘となった。1980 年代後半から米国を中心にした超並列計算機と並列計算の研究は大きな盛り上がりを示し、第二期における大きな波となった。その当時、ボストン近郊には並列計算機に関連するベンチャー企業も数多く出現し、Hypercube マシーンや Connection マシーンなど様々な並列アーキテクチャーが提案された。しかしながら、1990 年代にはいってプロセッサー性能の飛躍的向上に伴い、高性能のパーソナルコンピュータやワークステーションが廉価に手にはいり、クラスタ型コンピュータの開発も進むと、その波は次第に静まっていくかに見えた。しかしながら、2000 年代にはいって再び並列計算とスーパーコンピュータへの期待度が増し、第三期のスーパーコンピューティングの技術的波が到来している。

スーパーコンピュータ技術は、図１が示すように様々な技術が階層を持って構築されており、各階層の技術が驚くべき進展を遂げている。その基礎を支えるのは半導体技術であり、スーパーコンピュータ技術の進展も半導体技術の進展と無縁ではない。半導体技術がスーパーコンピュータ技術の進展にいかに影響を与えてきたかを概観し、今後の技術を展望する。

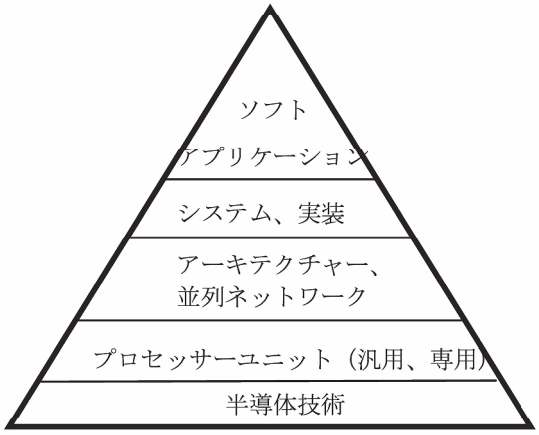

図１：スーパーコンピュータの技術

2. 半導体素子の性能向上

現在、エレクトロニクス機器に用いられている集積回路を構成している半導体素子は CMOS デバイスと呼ばれる電界効果型トランジスタである。この素子は電子が輸送する n-MOSFET と正孔が輸送する p-MOSFET から構成され、その性能の行方がエレクトロニクス機器の性能に大きく関わっている。トランジスタ性能は、その構造によって決まる電場内の電子輸送によって決まり、室温下のその輸送特性は Poisson-Boltzmann 方程式系によって記述される。電子輸送特性がわかれば、電気特性としての電流-電圧特性が得られる。

この MOSFET の性能には基本的な設計指針が見出されており、スケーリング則と呼ばれる[2]。それは、トランジスタ構造を縮小し、電源電圧を下げれば、トランジスタ性能が向上するというものである。では、どのように縮小し、どのように電圧を下げていけばよいのか、次の Poisson 方程式の相似変換から考察することができる[3]。

$$\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}+\frac{\partial^2\varphi}{\partial z^2}=\frac{q}{\varepsilon}(n-p-(N_D-N_A)) \quad (1)$$

ここで、φ はポテンシャル、n, p は電子・正孔密度、N_D, N_A は不純物密度、q は素電荷量、εは誘電率である。変数に以下のスケーリング

$$\begin{aligned}&\tilde{\varphi}=\varphi/\kappa\\&(\tilde{x},\tilde{y},\tilde{z})=(x,y,z)/\lambda\\&(\tilde{n},\tilde{p},\tilde{N}_D,\tilde{N}_A)=(n,p,N_D,N_A)\lambda^2/\kappa\end{aligned} \quad (2)$$

を施せば、(1)は

$$\frac{\partial^2\tilde{\varphi}}{\partial x^2}+\frac{\partial^2\tilde{\varphi}}{\partial y^2}+\frac{\partial^2\tilde{\varphi}}{\partial z^2}=\frac{q}{\varepsilon}(\tilde{n}-\tilde{p}-(\tilde{N}_D-\tilde{N}_A)) \quad (3)$$

となり、相似変換となることがわかる。すなわち、電位を 1/k 倍し、サイズを 1/λ 倍、密度をλ^2/k倍すれば、ポテンシャルは相似変換する。今、λ = kとし、MOSFET のゲート遅延、消費電力、駆動電流に関する解析式から、表 1 のスケーリング則を得る。

表 1 ：スケーリング則

Dimensions,	L, W, T_{ox},	$1/\kappa$
Potentials,	φ,	$1/\kappa$
Concentration,	N_D, N_A,	κ
Electric Field,	E,	1
Gate Capacitance,	C_g	$1/\kappa$
Current,	I,	$1/\kappa$
Gate Delay,	τ,	$1/\kappa$
ClockFrequency	f	κ
Power,	$\frac{1}{2}fCV^2$,	$1/\kappa^2$

ここで、1/k=0.7 とすれば、サイズを 0.7 倍に縮小すれば、ゲート遅延は 30%向上し、消費電力は 1/2 になり、さらに、ウエハー当たりに取れるチップ数は２倍になることがわかる。半導体技術は微細化することによって、性能向上と経済性、さらに、信頼性向上を達成しており、成長の原動力になっている。

半導体製造プロセスは、ゲート長を代表して、65nm 世代、45nm 世代、32nm 世代などと呼ばれ、世代毎に 0.7 倍に縮小された半導体製造プロセスを開発することによって新しい集積回路が開発されてきた。しかしながら、微細化が進むにつれ、プレナー構造 MOS 論理速度は物理的限界（おおよそ、0.25psec と予測される。）に近づいている。図２に、今まで世界で開発された MOSFET のゲート遅延をゲート長を関数にして示している。微細化だけでは性能向上のトレンドを保つことができず、90nm 世代以降、新構造と新材料の導入によって MOS 論理速度の性能向上が図られていることがわかる。

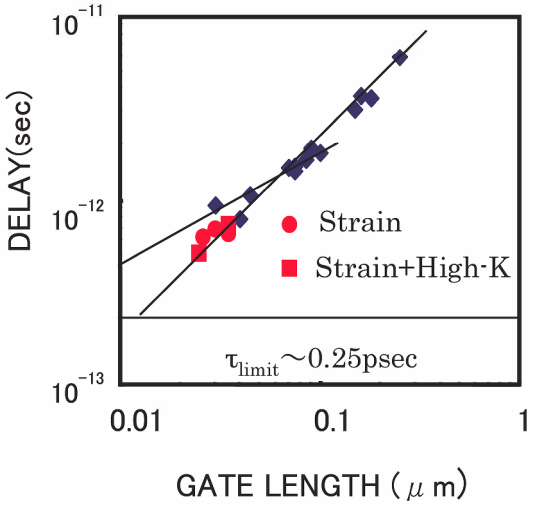

図 2 : ゲート遅延

図 3 に MOSFET 構造を示している。新構造とは、Si-結晶にストレスを加え結晶構造を変形させることによって、電子・正孔の有効質量を変え、輸送特性を向上させる技術である。n-MOSFET には Linear によって引っ張りのストレスを、p-MOSFET には、SiGe を埋め込んで圧縮ストレスを加える構造になっている。また、新材料には、ゲート絶縁膜材料に HfO_2 などの高誘電材料を、ゲート電極に金属材料を導入するものであり、32nm 世代で共に導入された。

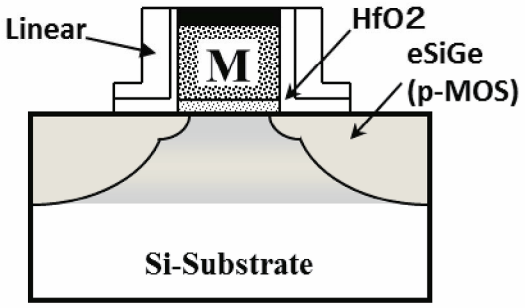

図 3：MOSFET 構造

3. プロセッサー技術と半導体

プレナー構造 MOS 論理速度の物理的限界を受けて、2005-2006 年頃から、プロセッサー技術に並列処理機能をもたせて、その性能向上をはかると共に、低消費電力化を進める技術となっている。　その一つは、マルチコア技術であり、もう一つは、NVIDIA 社などから提供されている SIMD アーキテクチャーを導入した GPU 技術である[4]。　この二つのプロセッサー技術は、第三期のスーパーコンピューティングの研究開発を促進し、スーパーコンピュータへの期待度を増している。

表２に近年開発されたスーパーコンピュータと、それに使われたプロセッサーと半導体製造プロセス世代をまとめている。Intel 社が提供する PC プロセッサーに最先端半導体プロセスが使われており、スーパーコンピュータ用プロセッサーには必ずしも最先端の半導体技術が使用されているわけではない。

表 2：スーパーコンピュータと半導体

スーパーコンピュータ	プロセッサー	半導体プロセス
京コンピュータ (2011)	SPARC チップ	45nm Bulk CMOS(Fujitsu)
Bluewater(IBM, 2011開発中止)	Power7	45nm SOI CMOS(IBM)
天河(China, 2010)	NVIDIA GPU Intel(Xeon)	40nm bulk CMOS(TSMC)
参考:PC	Intel (Westmere) (2010)	32nm bulk CMOS (第２世代High-K)

2012 年から Intel 社が投入している 22nm 世代は Tri-Gate トランジスタと呼ばれる立体構造を有した新しい MOSFET である。現在、日本企業にはこのタイプのトランジスタ技術を実現する半導体プロセスを量産化する計画はない。今後、日本のスーパーコンピュータ用プロセッサーに国産半導体素子技術が使われることはなく、海外ファウンドリーメーカーに製造委託されることになる。この場合、必ずしも先端技術が提供されるわけではない。NVIDIA 社の GPU の製造を請け負っているファンドリーメーカーの TSMC 社でさえ、このタイプの Multi-gate トランジスタの量産化は 2014 年と言われている。さらに、IBM 社は 45nmSOI/CMOS 技術(Bulk CMOS より高性能)によって開発されたプロセッサーPower7 によって進めていたスーパーコンピュータの開発を 2011 年に中止した。MOSFET 性能向上のために SOI 技術の薄膜化に注力した IBM 社及び開発アライアンス企業は、トランジスタ技術に遅れをとったのではないかと考えられる。今後、IBM がどのような半導体技術でプロセッサーの開発を進め、再びスーパーコンピュータ開発に参入するのかが注目される。Intel 社はすでに High-K 技術を有した Tri-Gate トランジスタを実現しており、今後、III-V 族 Tri-Gate トランジスタも視野に入れた開発が進められている[5]。

4. メニコア時代のスーパーコンピューティング

科学技術分野における数理モデルの多くは偏微分方程式で記述されており、偏微分方程式の数値解法は並列計算の中心的課題である。その一つは大規模行列解法

$$Ax = b \qquad (4)$$

であり、今、一つは大規模固有値問題

$$Ax = \lambda x \qquad (5)$$

の解法である。近年、京コンピュータ上で開発された大規模固有値問題の並列計算手法[6]が密度汎関数法に適用され、大規模材料設計への道を開いているのは、近年におけるよい例である[7]。

並列計算アルゴリズムは、一般には次の基本的要素からなる。すなわち、並列計算のためには、なんらかのブロック化が必要であり、データを各プロセッサーエレメントに配信し、そこでロードバランスを保って高速演算を実行し、再び得られた結果を収集することが必要である。このため、そのアルゴリズム性能はコンピュータのアーキテクチャーとも深くかかわることになる。

並列アルゴリズム

1. Parallel: 何らかの Block 化が必要。
2. Distribute: 並列ネットワーク、共有メモリーや分散メモリー構造に関連。
3. Compute: シリアル計算、ロードバランス。
4. Gather: 並列ネットワーク、共有メモリーや分散メモリー構造に関連。

今後、マルチコア CPU は先端半導体素子技術を取り込みながら、ゆるやかなメニコア化に向かって進むと考えられる。このとき、同一チップ内のコア間の通信速度はチップ間通信速度よりも格段に高速になり、並列計算アルゴリズムにおけるデータ配信・収集時間に影響する。このため、先端半導体素子技術を用いたメニコアプロセッサーによるスーパーコンピュータの実現は、並列計算アルゴリズム開発にも直接的に影響すると考えられる。また、CPU と GPU との半導体技術に差が出てくる可能性があり、並列計算を行うプロセッサーとしての CPU vs GPU の議論を、そこに使われている半導体技術も含めて進めていく必要がある。

5. おわりに

スーパーコンピュータ技術の進展も半導体技術の進展と無縁ではない。プレナー構造 MOS 論理速度の物理的限界を受けて、プロセッサー技術の並列処理機能が進み、スーパーコンピューティングの第三期の技術的波が到来している。先端半導体素子技術を取り込みながら、メニコアプロセッサーによるスーパーコンピュータが今後実現されることが期待され、並列計算アルゴリズム開発にも直接的に影響すると考えられる。どのような半導体技術を使ってスーパーコンピュータを実現していくかは、スーパーコンピューティングのピーク性能だけでなく消費電力にも影響し、今後の実用化と普及にとって重要な要素である。

参考文献

(1) P.D.Lax(Chairman), Report of the pannel on large-scale computing in science and engineering, 1982.

(2) R.H.Dennard et al., Design of ion-implanted MOSFET's with very small physical dimensions, IEEE Journal of Solid-State Circuits, vol.9, pp.256-268, 1974.

(3) G.Baccarani et al., Generalized scaling theory and its application to a 1/4 micrometer MOSFET design, IEEE Trans. on Electron Devices, vol.31, pp.452-462, 1984.

(4) J.Y.Chen, GPU technology trends and future requirements, Proc. of IEDM, pp.1-6, 2009.

(5) M.Bohr, The evolution of scaling from the homogeneous era to the heterogeneous era, Proc. of IEDM, pp.1-6, 2011.

(6) T.Imamura, S.Yamada, and M.Machida, Development of a high-performance eigensolver on a peta-scale next-generation supercomputer system, Progress in Nuclear Science and Technology, vol.2, pp.643-650, 2011.

(7) J.Iwata, D.Takahashi, A.Oshiyama, T.Boku, K.Shiraishi, S.Okada, K.Yabana, A massively-parallel electronic-structure calculations based on real-space density functional theory, J. Comp., Phys., vol.229, pp.2339-2363, 2010.