システム設計から見た MCM と シングルチップパッケージの比較〔上〕

大塚 寛治*

Comparison between MCM and Single-Chip Package Standing on System Design

Kanji ÔTSUKA *

*明星大学情報学部電子情報学科 (〒198 東京都青梅市長淵2-590)

*Department of Electronics and Computer Science, College of Informatics, Meisei University (2-590 Nagabuchi, Ôme-shi, Tokyo 198)

1. はじめに他山の石を知る

小型で高密度な電子装置を、コストの上昇しない範囲で工夫して作る。これは何人も反対しない概念である。ところが、標準的な範囲を越えた小型化や高密度化が要求される。市場競争からの必然的な現象である。MCM (Multichip Module) もこの要求から生まれたものである。しかし、MCMで完成したサブシステムをオンチップにすることができれば、最も高密度実装が実現する。技術の発展途上では、この2つのサイクルは常に無限ループである。いままで、ハイブリッドICがこのサイクルの一部を担っていたように思われる。

MCMはハイブリッドICの延長ではないという議論が多い。MCMを使用する狙いは、

(1) ハイブリッドICより大きなシステムの完成をめざしたものである

(2) 配線を短くして、高速信号処理を有効に実現できる

(3) 汎用的な目的で作られているLSIは過剰な性能をもっているが、MCM用のLSIはMCMシステムとして最適設計ができる

(4) 現有LSIの組み合わせで、短納期で、高性能が実現する

(5) 1stレベルパッケージングを省略できるため、コストが安い

(6) LSI化しにくい回路群を、システムにまとめ

上げることができる (今までのハイブリッドICの概念)

(7) 小型高密度を先取りできる

(8) 高性能システムの先取りができる

などがある。本当にそうか、ここではコンピュータの高密度実装というテーマで考えることにする。

この世界はダウンサイジングの革命の中にある。この革命がなぜ起きたか解析することは、これから考えようとする比較の内容を明確にする。大型コンピュータはその時代のLSIの性能を越えて設計してきた (狙い(1), (4), (8))。多くのLSIを集合させてシステムを構築せざるを得なかった (狙い(3))。高速になると、LSI間の配線の長さが信号の遅れとして目立つようになった。高密度実装を考え、この問題に対応した (狙い(2), (7))。しかもLSIを無理に高速で働かせるため、一番やってはいけない電力を北京ダックのように供給した。したがって、高発熱密度となり、水冷が必要となった。高密度実装は冷却性能で制限されることになり、伝熱工学の問題となった。コンピュータは電子工学を極めるところであろう。話はどんどん横道にそれているようである。どんどんコスト高になったのはいうまでもない。MCMの主要な狙いがすべて入っていた大型コンピュータの設計思想は、競争に負けたのである。

一方、パーソナルコンピュータはどうだったであろうか。低コストが前提であったため、無理なLSIの使い方をせず、その時代のLSI技術で可能なシステ

ム構築をしてきた。すなわち、シングルチップで論理システムが可能な範囲で設計をしてきた。高性能な冷却装置もつけられないため、電力も抑えに抑えてきた。LSIはそれでも目をみはるほどの速度で集積度と性能を上げた。労せずして、シングルチップシステムの性能が上がったことになる。これはメモリLSIの考え方がしっかりしていたからである。このメモリの開発概念に少しふれよう。

メモリは、決まった大きさのボード上で集積度を大きくしなければならない。集積度を大きくしても、ボード全体の消費電力は一定である。ボード全体の価格も一定に抑えたいという顧客の要求も、満たさねばならない。また顧客が不特定多数のため、メーカーの事情を顧客に押しつけることもできない。このように八方ふさがりの中で開発が行われた。それでも技術的工夫で格段の進歩があった。八方ふさがりでない分野よりである。人間の英知は無限であることの良例である。図1¹⁾はそれを端的に表す進歩の過程である。元々メモリは多数並べて実装されていたため、パワーはかけられず、NMOSであった。MOSの電力消費は主としてゲートの容量であり、ゲート幅の縮小は常に努力されていた。これはまたスイッチング速度を速める効果や集積度を上げる効果もあり、まさに焦点の絞られた開発であった。

4倍、4倍の集積度の増大は、努力しなければそれに比例したパワーの増大につながり、ゲート幅縮小だけでは追いつかない状態であった。NMOSよりトランジスタの数が増えても、格段にパワーの小さいCMOSが256kビットより採用され、現在に至っている。以降、ゲート縮小の努力を経て、4Mビット

でボード上、通常冷却の限界である500mWとなったため、電圧を5Vから3.3Vに降圧する動きがでた。パワー $P=1/2 \cdot CV^2$ (Cは負荷容量, Vは電圧) の関係にあり、降圧の効果は絶大である。初めはシステムに影響のないLSI内部回路の降圧、システムが追従できるようになった現在、外部降圧でさらに低消費電力としている。

アクセスタイムはゲート縮小により、どんどんと短縮できたが、256kビット以前はシステムが追従できず、大きな努力は払われていなかった。最近ではシステムの高速化が進み、高速メモリへの要求が強くなり、開発の中心課題となっている。このように、パワーは一定で、集積度と速度の大幅な改善がなされてきた。

パーソナルコンピュータのCPUであるマイクロプロセッサチップも、このメモリで開発されたCMOS技術を使い、シングルチップでどんどん性能を上げてきた。大型コンピュータだけで支持されているECL回路に比べて、はるかに大きな市場をもつメモリの、すなわち、CMOS回路の発展速度は速い。マイクロプロセッサは全面的にこの利益を享受できたのである。そのCMOSの威力を図2²⁾、図3³⁾で示す。ECLに比べ、CMOSは集積度で5倍から1桁、ゲート速度は同じような負荷付きの回路で似たようなレベルまで追いついてきた。パワーも100ps/ゲートのデバイスと比較すると、ECL系の0.5mW/ゲートに対し、CMOSは0.009mWと1桁以上低いレベルにある³⁾。

図1. DRAMの消費電力とアクセスタイム

図2. CMOSとバイポーラのゲート遅延の改善動向

図3. CMOSとバイポーラの電力-遅延時間積の比較³⁾

大集積が可能なCMOSをうまく使うために、マイクロプロセッサの設計者はシングルチップ内の論理規模を拡大するだけでなく、周辺回路をどんどんこのチップ内に取り込んできた。このような大きなシステムを設計する能力は米国において優れ、残念ながら、日本はOEMでその一部を生産するにとどまっている。それはさておき、高速で処理しなければならないコンピュータの中味のほとんどを、小さなチップ内に閉じ込めることができたわけである。大型コンピュータで問題であった、配線の遅れというペナルティを負う必要がなく、演算速度は完全に大型コンピュータを凌駕することになったわけである。集積度の問題で処理能力は大型コンピュータに一步譲るが、普通の放熱構造でシングルチップ実装でよいわけであるから、コストは大型の1桁も2桁も安い。これを簡単に表すと図4のようになる。ダウンサイジング革命のゆえんである。

低コストで高密度なパッケージを開発するにあた

図4. コンピュータの価格-性能の概念

って、この事実は他山の石である。マイクロプロセッサは常に主流技術の中にいたわけである。技術に不可能はない。しかし、世界的に最もマンパワーをかけている技術が最も工夫され、コストも安くなり、勝者になるわけである。

パッケージ技術は幸い日本が強い。日本の中で最もマンパワーをかけている技術、それは何であろうか。言わずと知れたリードフレームをプラスチックモールドしたパッケージで、これをプリント配線板にはんだ付けで実装する形態である。世界的に見ても85%以上の生産量はこの形態であろう。マイクロプロセッサが行ったように、迷わずこのプラスチックモールド道を歩めばよいのであろうか。残念ながら、リードフレームの加工技術に遅れを生じていて、要求を満足できない状態となっている。パッケージ技術の前記したマイクロプロセッサと異なるところは、LSI技術やシステム技術の従属者であるところである。しかし、従属者という観点から見ると、メモリと同じである。すなわち、開発ペースが決められているメモリは、八方ふさがりからすばらしい開発を行ってきた。パッケージもそのペースに合わせて、メモリのようにすべきである。

2. 高密度実装を幾何学的に見る

なぜ高密度実装が必要か、配線長さを短くするためである。なぜ短くしなければならないかの真の理由は、電磁気学を知らなければならない。本章では、これはさておき、高密度実装に対するパッケージ設

計の考え方を幾何学的に見ることにする。

チップを隙間を開けずに実装するためには、チップ上の配線と同じレベルの微細配線を、取り付け基板側に必要とする。フリップチップでSi on Si実装しなければ解決しない。配線板の技術レベルに応じて、チップ間の隙間、すなわち、チップは位置ピッチを決めることになる。この概念には落とし穴がある。サブシステムとして回路が閉じたチップを、少数個集合させる場合を考えると、チップ間の相互結線数は少なく、入出力バスへの配線のみでよい。このとき、チップを隙間なく並べても、配線板の配線ピッチは粗くてもよい。チップ間の結線が、論理レベルで必要なとき、レント則に従う。これを最高結線数とすると、入出力バスへの結線は1/10から1/100になるかもしれない。

この入出力バスへの結線の場合の、幾何学的な配列を見ることにする。必要結線数がレント則、その1/10、1/100の3つに分類して、基板配線技術レベルと対比する。これを図5に示す。1 μ mプロセスがMCM-Dであるとすると、(a)(A)の組み合わせでもチップの隙間を必要としない。前述のSi on Siである。10 μ mプロセスをMCM-Cとすると、チップサイズ4mm(この場合、配線専有面積=チップサイズとする)の10倍のスペース、40mmピッチが必要となる。これが(b)(A)の組み合わせである。これが大型コンピュー

タの苦勞した点である。40mmピッチを多層にすることで近づけたわけである。2層が一般的なハイブリッドIC技術では、40mmのままであり、パッケージングしたチップが十分搭載できる。(b)(B)の組み合わせでは、問題なく隙間0が可能である。MCM-Lでは、(c)(A)の組み合わせは使えない。(c)(B)の組み合わせでは40mmのピッチを取る必要がある。このような広いスペースをあけて裸チップを載せる意味があるのだろうか。これもパッケージングしたチップを十分に載せることができる。すなわちMCM-Lの世界ではなく、SMTの世界である。(c)(C)で初めてMCM-Lが使えるレベルとなる。前述のように、プラスチックパッケージのSMTが最低のコストでできる実装方法である。無駄なところでMCMを使うべきでないことが、この分類でわかる。

さて、サブシステムLSIがどのくらいのピン数を必要とするのか、これを明確に知らないかぎり、高密度実装の最適方法を決めることはできないのである。一方、そもそも高密度実装はなぜ必要かも知っておく必要があろう。前者(ピン数)を知るには、LSI設計技術、コンピュータ設計技術、信号伝送技術などパッケージ技術者の知らない戦略的な内部知識をマスターする必要がある。後者(高密度実装)は上記技術に加え、モバイル情報処理、資源有効利用等の社会的ニーズについても考える必要がある。範囲が広がり過ぎるため、ここでは主として、電気信号をいかに高速に、大量に伝えるかの観点で述べてみたい。そして、MPUのピン数とその内容について、市場の例をあげ、多ピンの必然性を確認する。

3. 電磁気学の観点から

3.1 バスによるコミュニケーション

今、MCMの騒がれている背景は、シングルチップパッケージの電気特性に限界があるからである、としている。どこが限界かの、明確な説明なしである。筆者は現状パッケージング技術に電気特性上の限界は感じていない。この混乱をまず正すことにしたい。

LSI論理回路は1つのトリガパルスにより、動作する。これを定期的に与えるのが、クロック信号源である。クロックの周波数を高くすると、LSIの動作はそれにつれて速くなる。さらに周波数を上げると、

図5. チップの幾何学的な位置とプロセス/
I/Oピンの関係

1クロックサイクルの間で処理できないパスができる。このパスは転送処理でも論理処理でも、次のクロックの処理とぶつかり、混乱する。すなわち機能しない。その周波数において遅延により誤動作する恐れのあるパスを、クリティカルパスという。処理時間がかかる理由は、トランジスタの動作遅れと配線遅れがあるからである。高速処理のために、この両者をできるだけ小さくする努力が必要となる。以前はトランジスタの動作速度が遅く、配線の遅れは目立たなかったが、その改善がめざましく、最近では配線遅れが支配的になってきた⁴⁾。しかし、このクリティカルパスが2つのLSI間を渡るとき、初めてパッケージの責任分が生まれる。

考えてみると、LSI内のパスとLSI間のパスを混在して処理するのは、よい設計とはいえない。大型コンピュータはチップ内で論理が閉じず、このような混在を余儀なくされたため、高密度実装の先鞭をつけたわけである。チップ内で論理が閉じるようにすればよい。これがパーソナルコンピュータやワークステーションの基本となっている。チップの外に出るときは、他のサブシステムLSIに電話線のように信号を伝えればよい。国際電話で話をするときは、会話間の間が伸びるが、必要事項は国内と同じように十分伝えることができる。LSI間をつなぐこのような電話線をバスと呼んでいる。電話線は2本しかなく、1クロックで4通りのデジタル信号しか伝えることができないが、n本のバスでは1クロックで、 2^n 通りの信号を同時に伝えることができる。多くの信号のやり取りを必要とするとき、バス幅nを大きくとればよい。バスがどれだけ長くて、受け側のLSIが自分の好みのタイミングで、この信号をバスの幅分、同時に取り込めばよい。これを示すと図6

のようになる。パッケージ技術者にとって、4ビットから8ビットの制御コンピュータ、16～64ビットのパーソナルコンピュータやワークステーションのクリティカルパスは考えなくてもよい。これで話をだいたい単純化できたことになる。

もう少し高度な話をすると、CPUのマシンサイクルで同期するメモリのアクセス（例えば、2次キャッシュメモリとの間など）に配線遅れがあると、これはクリティカルパスとなる。ハイエンドのMPUではここに焦点をあてて論議していることが多い。

3.2 パルス波形の変形による伝達の遅れと伝達エネルギーの損失

しかし、バスの信号の流れだけをとっても、パルス信号周波数が大きくなるといろいろな問題がでてくる。第1の問題は、デジタル信号の反射や負荷による波形の変形である。問題の理解の前にこの現象を見よう。これを図7に示す（説明は後述）。この問

図7. 開放終端における信号線への電圧の伝わり方

図6. 4ビットバスの入出力結合形態

題はデジタル信号設計の問題でなく、電磁気学に関する問題である。回路周辺に複雑にからんだ電磁界のエネルギーが交錯しているとき、エネルギー保存の法則から、損失エネルギーを最小にするように作用する。自然界はよくできているもので、常にエネルギー的には最適解が求まっているのである。しかし高周波数になるほどプラス・マイナスの往復運動回数が多いわけであるから、その損失が大きくなり、小さくする工夫をしなくてはならない。いろいろな具体的な問題を理解するために、まずこの電磁損失エネルギーについて説明する。これにより、第1の問題は理解できるようになる。

電磁界のエネルギー変化に対して、その損失エネルギーを最小にするパッケージ構造、これはどんなものであろうか。信号エネルギーは必ず電源から供給され、グラウンドに放電する。しかし、ある瞬間を考えると、電気エネルギーの伝播回路は、2本往復線路があればよい。パッケージの信号線に信号パルス(エネルギー)が伝播するときは、電源線との往復回路、信号パルスエネルギーを0にするとときは、グラウンド線との往復回路となる。この2線を十分に結合(カップリング)して、その状態を連続してやればエネルギー損失を少なくすることができる。なんのことはない、同軸ケーブルである。これで10GHzは十分伝送できるのである。しかし、瞬間瞬間で組み合わせの違う回路となるパッケージ内の信号系には、この同軸線は使えない。

今、図8の1段目の図に注目する。信号線と電源線があり、左のスイッチは開いているものが閉じたとする。左の電池からプラスチャージが信号線に供

図8．線路への電気エネルギーの流れ方

給され、電源線にはマイナスチャージが供給される。オンした瞬間が1段目の状態である。右にある負荷はこのときは何も感じていない。しかし、左にある電池は、配線の小さな抵抗の制限を受けるだけでフル回転している。オン後の時間経過と共に、2段目のように、プラスチャージは信号線に供給され続け、電子は電源線に注入される。この供給のされ方は電線の中の小さな抵抗に応じた量となる。この抵抗がよく耳にする特性インピーダンス Z_0 である。

時間経過と共に3段目の図のようになる。このとき初めて負荷にエネルギーが流れ、プラスチャージとマイナスチャージが衝突消滅することでエネルギーが消費される。その消費の速度は負荷の抵抗の大きさによる。すなわち抵抗が高ければ、パイプの太さが細いため、消費の速度は遅い。ところが、電線の端までは、 Z_0 という小さな抵抗、すなわち、太いパイプでエネルギーが流れてきている。この境界でエネルギーが余ることになる。その結果、余ったエネルギーは反射して戻ることになる。これを4段目の図に示した。ここでは2個に1個が反射しているモデルを示している。すなわち、 $2Z_0=R_L$ であり、 $Z_0<R_L$ の条件のときにあたる。図8は、出口の抵抗の不整合についてモデル化したものである。図7は入口の抵抗の不整合について述べられている。反射という概念は、接続部の抵抗の不整合で起こるもので、その部分がどこにあると同じである。

線路のインピーダンスを一定に保つことの重要性が判明する。一瞬に起こる現象は、実はこのように数ステップの作用の合成でできているのである。この説明で図7の現象が理解できると思われる。開放端は、電気の行き場所がないから100%反射する。すなわち、反射係数 $m_r=1$ である。電源からのパルスは、伝送線路に入る前に、抵抗 R_s というパイプと線路のパイプ Z_0 との継ぎ目で、パイプの太さの違う量に応じてパルスが反射する。その反射係数 m_s は図7のような関係となる。

(b)は細いパイプに太いパイプがつながっているため、太いパイプに電気をためるのに時間がかかる。すなわちなかなか立ち上がらない。

(c)はちょうど同じパイプがつながっているため、素直に電圧が上昇する。②の伝送路の入り口です

に抵抗による電圧低下 iR_s がある。 $v-iR_s$ が③の開放端に達すると、100%反射するため、 $2(v-iR_s)$ が③の端子の電圧となる。今、 $R_s=Z_0$ のため、この値は v そのものとなる。①と③の立ち上がり形状はまったく同じで、線路をパルスが進む時間 t_{pd1} だけ遅れが生ずる。

(d)は太いパイプから細いパイプに入るため、電源へ反射する電圧を細いパイプの線路が感じる。(c)の②より反射係数だけ高い電圧となり、③の状態は、電源電圧5Vより高くオーバーシュートし、ハンチングが起こる。このようにパイプのつなぎ目で大きな問題が生じることが判明する。

③の開放端に負荷である Z_L をつなぐと、もっと複雑な波形が負荷に伝わることになる。昔は立ち上がり時間 t_r が t_{pd1} より数倍以上大きかったため、この波形変形は立ち上がり変化の中に隠れて見えなかったわけである。今のパッケージ配線は、 $t_{pd1} > t_r$ とな

り、パイプの太さを気にしなくてはならない。LSIより相手が必要と思われる正しいエネルギーを出しても、パイプの太い細いの不連続部分で反射し、反射波と進行波の合成によるエネルギーロスで、100%エネルギーが伝わらない。これだけではなく、変形した波形のどの部分を信号として取ればよいかわからなくなってしまうのである。

以下次号につづく (1994. 3 .24-受理)

文 献

- 1) 伊藤清男, 他: “サブハーフミクロンDRAM技術”, 電子通信学会, 1993年春季大会予稿集, pp.10-23
- 2) 日経マイクロデバイス, No.9, p.37, 1990
- 3) 日経マイクロデバイス, No.8, p.41, 1991
- 4) H. Wessely et al.: “Electronic Packaging in the 90s”, Proceedings of Electronic Component & Technology Conference, pp.16-33, 1990