Microprocessor Assembly Interconnect Pathfinding Challenges

Raj Nair
raj.nair@intel.com
Assembly Technology Development
ATD – Pathfinding
Chandler, AZ

Abstract

It is well known that Moore's law [1, 2] and the progression of ultra-large-scale integration places increasingly challenging demands on microprocessor and high-density interconnect chip assembly technology. As the number of transistors on lead microprocessor products as well as the operating frequency doubles every two years, total power consumption continues to relentlessly increase. Reducing supply voltages require operating current increases that trend at a much faster rate. On another front, the data transfer rate required of chip-to-chip links is also increasing at a pace that in the foreseeable future is expected to keep up with the chip operating frequency increases.

These trends make the design of the "package" or the electro-thermo-mechanical interface to the chip all the more challenging. This paper highlights the challenges and details some of the solutions envisioned.

Introduction

Assembly technology has been said to have, in the past, the importance of the clothes that define a person. The present and the trends of the future on the other hand place such burden on these layers of clothing that assembly components are at risk of limiting microprocessor performance in much the same fashion as a heavy burden carried by a vehicle slows it's progress.

It is hence critical that the requirements placed upon assembly technology components be fully comprehended such that holistic solutions are developed. A clear understanding of the division of the burden, i.e., Technology Development Burden Partitioning (TDBP) between the CPU architecture, design and process as well as the assembly technology architecture and design is very necessary. It will be clear in the discussions that follow that CPU architectural and circuit solutions will play a large part in sharing the technology development burden for future microprocessor product solutions.

It will also be seen that cooperative efforts alone will not be sufficient in addressing the challenges of the future. Out-of-the-box, breakthrough architectures that provide holistic scalable solutions will be necessary alongside very aggressive advances to the technology solutions of the present.

Power Hungry Processors

Current predictions indicate that as lead processors follow Moore's law (Figure 1), the processor power consumption will increase to as much as 2X to 3X that of the previous generation [4, 5]. The challenges this trend imposes are two-fold. The heat generated because of such concentrated power consumption is increasingly harder to remove. More significant, perhaps, is the difficulty such a progression in power consumption places upon the systems that deliver power to the processor, and given that the performance of the processor is sustained by such power delivery, it is critical that this challenge be met effectively. A specific problem here is the very high current demand of future microprocessors.

Operating voltages reduce in each process generation dictated primarily by the reliability concerns with the scaling gate oxide. Yet the drive for performance is slowing this trend, and with increased capacitance per unit die area, increasing total die area as well as doubling frequencies, power consumption continues to increase at an alarming pace. The following calculations that investigate the impact of scaling according to Moore's law will make this situation clearer. As dimensions are scaled to a new silicon process, device dimensions (width W, length L and gate thickness T) scale by 0.7; the area capacitance therefore scales as

$$\frac{W \times L}{T} = 0.7$$

And the fringe capacitance scales as the width by 0.7. Hence the total capacitance for a device reduces by a factor of 0.7. Yet the area reduces by $(0.7)^2$ or a factor of 0.5 and hence the capacitance per unit area increases by (1/0.7) or about 43% [4].

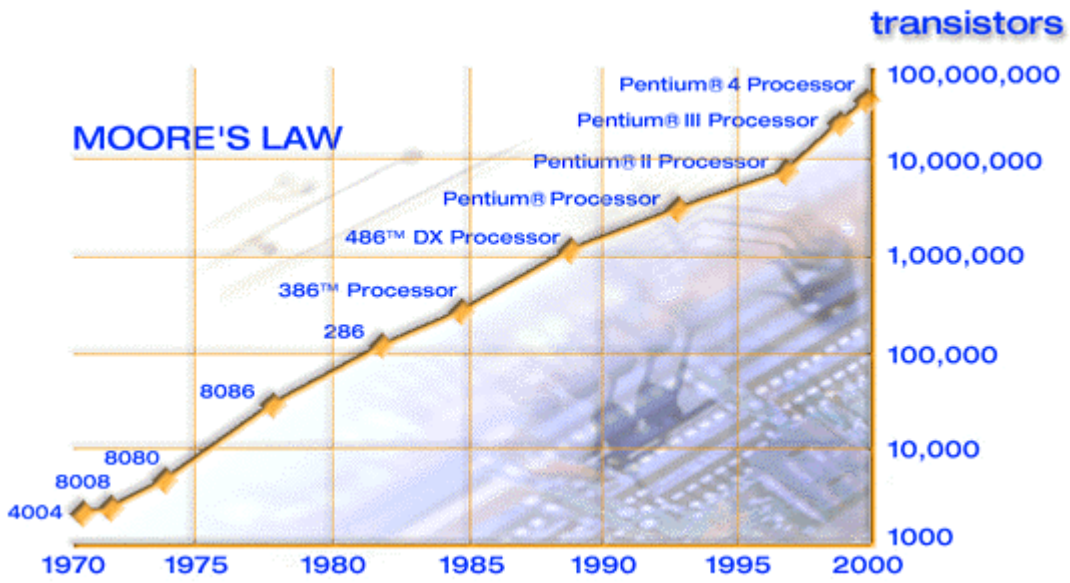

Figure 1: Microprocessors following Moore's law [2]

Current scaling scenario – 1: Quintuplet
Assume for the moment that a new microarchitecture requires 2X more transistors and 2X the frequency of the previous generation for maintaining the performance trend. Given the scaling in the process dimensions, this will require the same die area as the previous microprocessor. Next assume that voltages do not scale by the 0.7 factor, but scale only by ≈ 85% which is approximately $\sqrt{0.7}$. Since the power consumed is given as: $P \approx C_s V^2 f$, it can be seen that the power consumed will increase by 2X, or 100%, for the same die area. This is an increase in the current *I* of 2.4X, or 140%.

Current scaling scenario – 2: Triplet
Let us now consider a constant power architecture and design for the CPU. In the *Triplet* scenario, we require the microarchitects to squeeze the necessary performance out of the component with 100% more transistors implementing a new microarchitecture in the next generation process running at a frequency that is (1/0.7) or ≈1.43X higher. We simultaneously reduce the voltage by the full scaling factor of 0.7. Now the power P remains constant, and the current increases by a factor of (1/0.7) or ≈143% since the voltage scales by 0.7. The increase of 43% in the current would seem, superficially, to be a far more manageable situation.

An assumption made in determining the increase in component current draw in the simple analysis above is that all the spent power, based on standard switching power consumption is associated with performance. That will not be the case beyond deep-sub-micron device dimensions (L_{ch} – channel length, ≈100nm and T_{ox} – gate oxide thickness, ≈10Å) where temperature and applied voltage dependent leakage currents could increase to compose as much as 50% of the current demand [5].

Supply currents could therefore increase at the least by ≈ 43% or 140% depending on which of the above scenarios plays out in tune with market demand and process/micro-architecture/circuits optimization. A key consequence of the large supply current demanded combined with architectural controls that wake a microprocessor near-instantaneously from sleep, or a low power consumption state, is a series of what are commonly referred to as supply voltage droops. A challenge often shouldered unilaterally by assembly technology architecture, components, and design is the minimization of these transient reductions of the supply voltage differential within the microprocessor.

Simulations and droop theory will now be employed with cross-verification to analyze the consequences of the supply current trends envisioned in the preceding discussion. Lumped package electrical models (Figure 2) are often used in the verification of droops in the on-die power supply voltage. Load currents are modeled as anticipated in the architecture of the microprocessor using simple switched resistors and delay elements. An example of the disparate load currents anticipated (under ideal conditions) is shown in Figure 3. Both fast and slow ramping load currents as often used for Willamette simulation models are shown. Supply voltage droops caused by the combination of the fast load current ramp, the slower current ramp as well as the static leakage contribution is shown in Figure 4.

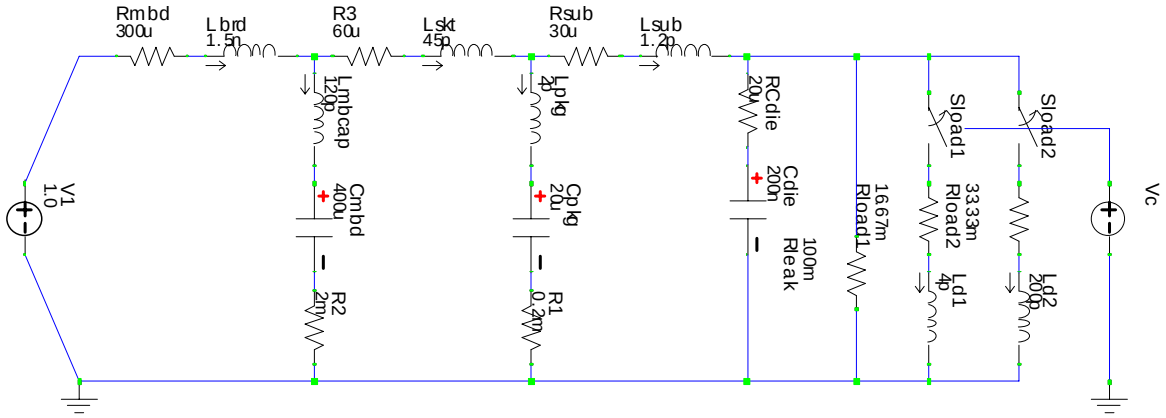

Figure 2: Lumped voltage droop simulation model

The simulation in this instance was carried out with a 1 Volt power supply (with no 'voltage positioning'), a total supply current of 100A corresponding to a processor consuming 100 W. It is assumed that 10 A is due to leakage and 90 A due to dynamic power consumption. The architecture is designed to ramp as much as 60% of the full load in a very short duration (few cycles of chip wake-up) and then to ramp the remnant 30% over a much longer period akin to the Willamette design.

Droop theory uses a very simple equation to predict the droops in the power delivery architecture. *At higher and higher frequencies of operation*, the fast turn-on of a microprocessor places a step-function current demand upon the power delivery architecture. Under the assumption that the large package capacitance behaves as a battery source (a model simplification), the transient response of the system can be solved as

$$V_f = V_i - \left(\Delta I \sqrt{\frac{L_p}{C_d}} \right) \sin(\omega_1 t)\, e^{-\alpha t} - I r_s \qquad (1)$$

Where α is given by

$\alpha = \frac{r_p}{2L_p}$, r_p is the damping resistance associated with the first resonant loop, L_p is the total effective inductance in the loop and C_d is the capacitance intrinsic to the microprocessor die across the power supplies, r_s is the resistance directly in series with the current path , ΔI is the step current demand, and I is the average current flowing to the load.

The resonant frequency of the loop ω_1 is given by

$$\omega_1 \approx \sqrt{\frac{1}{L_p C_d}} \qquad (2)$$

Equation (1) and (2) give simple forms for estimating the amplitudes and durations of the undershoot (droop) and overshoot caused by the step current demand. Assuming that the various resonant loops as seen in figure have little effect on each other, equation (1) can be expanded to include additional damped resonant terms that describe the transient nature of the voltage on the microprocessor die over much longer durations.

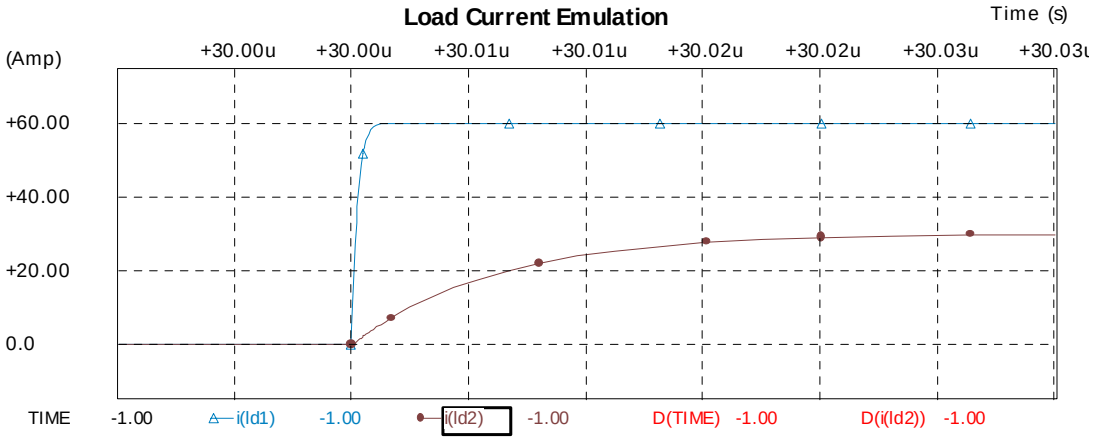

Figure 3: Die load emulation

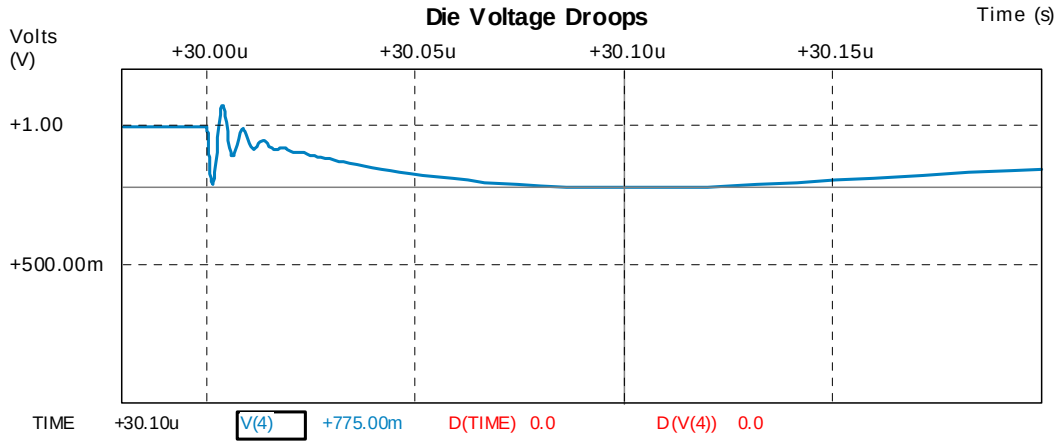

Figure 4: Die voltage droops caused by a step current demand

The purpose of this discussion is to validate through simple theory the droops seen in the simulations. The droop magnitude for the resonant loop that includes the die capacitance is given by

$$\Delta V = \Delta I \sqrt{\frac{L_p}{C_d}} \qquad (3)$$

Plugging in numbers from the circuit simulated, we get 60A×4mΩ = 240mV. This is close to the number seen in the simulation, which is ≈225mV; the simulation is expected to be more accurate through the inclusion of voltage dependency in the load and the finite current demand slope. The same formula with appropriate replacements for *L*, *C*, *ΔI* (a portion or all of the slower current ramp may be included) may be used to determine the reactive voltage droops for the 2nd and 3rd resonant loops of the architecture. These values could then be summed appropriately with drops across significant effective series resistances to determine the resonant droop amplitudes of the slower loops in the power delivery architecture.

Clearly, the knobs available to control the droops are limited from an assembly technology standpoint for the resonant loop of the highest frequency. Both *ΔI* and C_d are determined by the CPU architecture, circuit design and process scaling. Components of assembly technology, package caps and the substrate, contribute to the **effective loop inductance L_p** that determines the droop in this damped resonant circuit. As supply currents increase and explicitly added die capacitances are removed (for die cost, leakage and other die and circuit design reasons), reducing the power loop inductance to maintain droop ratios within control is a key challenge.

Loop Inductance Scaling and minimization

Loop inductance scaling can now be estimated by revisiting the supply current scenarios previously described. Using

$$\Delta V_1 = k_v V_1 = \Delta I_1 \sqrt{\frac{L_1}{C_1}} = k_i \frac{C_1 V_1^2 f_1}{V_1} \sqrt{\frac{L_1}{C_1}}$$

And

$$\Delta V_2 = k_v V_2 = \Delta I_2 \sqrt{\frac{L_2}{C_2}} = k_i \frac{C_2 V_2^2 f_1}{V_2} \sqrt{\frac{L_2}{C_2}}$$

Where X_i refers to a parameter *X* in the current and a scaled process[1], it can be shown that the scaling factor for the loop inductance is given by

$$S_l = \frac{1}{S_c S_f^2} \qquad (4)$$

Where S_x refers to the scaling factor for parameter *X*. It can now be seen that the ***Quintuplet "Q"*** scenario that increases the dynamic current drawn by 140% will require that the effective loop inductance scale by the **5th power of the process scaling factor** or to **≈1/6th** the previous value. The ***Triplet "T"*** scenario does not offer much relief either, requiring the loop inductance to scale by the **3rd power** of process scaling or to **≈1/3rd** the previous loop-L value.

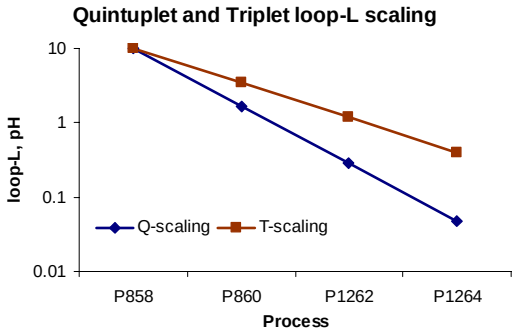

Figure 5: Loop-L scaling for Q & T power trends

[1] Though the capacitances contributing to power consumption are not the same as that serving the purpose of decoupling, they are represented by the same quantity C_i in both equations without loss of accuracy.

Starting from a hypothetical loop-L requirement of 10pH for a lead product in the P858 process, trends are plotted in Figure 5 indicating the loop-L requirement for both the **Q** and the **T** scaling trends. As seen, the **Q** trend leads to as little as 0.047pH by the P1264 generation and 0.008pH by the P1266 timeframe. The **T** trend leads to a loop-L requirement of 0.4pH in the P1264 timeframe and 0.067pH in the P1266 generation.

Both these scenarios are outside the scope of the POR packaging architectures and therefore it is clear that alternate solutions must be found for a manageable loop-L requirement-scaling trend.

Relaxation in the droop tolerance is an argument occasionally made (especially for the droop that has the least duration) in supporting less aggressive minimization of the loop-L requirement. Though this can be shown to be true, it is important to note that

a) This relaxation is a one-time benefit and the trends remain the same;
b) The removal of added gate oxide die decoupling capacitance, coincident with this relaxation, wipes out the benefit obtained;
c) Such relaxation is not a generic change in specifications across all architectures and
d) The droop duration is dependent upon L_p and C_d and does not necessarily scale with frequency.

To illustrate the points made above, let us review the recent experiments conducted on the Willamette and Coppermine processors that determined that in the presence of only a few of the low-inductance package capacitances, the die decoupling capacitance had little effect on performance [12]. Consequently, the droop specification has been relaxed tentatively from the traditional 10% to 20% while most of the added die decoupling was removed, with C_d reducing from the ≈900nF it used to be on the Willamette to ≈190nF. The end result is that the loop-L needs now to be **further reduced** by about 16%!

Let us now investigate **potential solutions** to this very intimidating challenge. Microprocessor scaling does not give any relief in terms of power or frequency reduction as foreseen in the near future. Therefore ΔI is expected to continue to increase, and at the very minimum, at a rate dictated by the **T** scenario. Since the voltage V does not seem to play a part in the loop-L scaling factor, we are left with the scaling factor of the capacitance as our only relief.

With some manipulation of the equations, it can be shown that a capacitance C_a added on-die transforms the loop-L scaling relationship to

$$S_l = \frac{1}{S_c S_f^2} + \frac{C_a}{C_1 S_c^2 S_f^2} \qquad (5)$$

If C_a is made equal to S_c times C_1 (thus effectively doubling the capacitance on the die), the loop-L scaling factor is increased by a factor of 2, or the inverse of the square of the process scaling factor 0.7. This also follows intuitively from the simple droop equation that determines that if the capacitance is doubled, the loop inductance could also be doubled while maintaining the droop voltage the same. The addition of such a capacitance limits the loop-L scaling factor requirement (the **"T"** scenario) to the process-scaling factor of 0.7.

Practical methods of accomplishing this include the addition of **MIM** (metal-insulator-metal) capacitors to the microprocessor die in the interconnect layers, where the insulator material is one with high dielectric permittivity such as Ta_2O_5. With **~10fF** per square micron, an area of 10 mm^2 (or a die space of ≈ 4mm by 4mm, counting area inefficiency) is sufficient to achieve as much as 100nF of capacitance that is close to the anticipated on-die intrinsic capacitance on products in the making.

Build-up Layer Capacitance (BLC) within a package substrate (Figure 6), enabled through metal (high temperature tolerant) core substrates and **high-K** material build up is a potential alternative solution that could take advantage of the much larger substrate area for greater capacitance values [15]. Though the close proximity to the die and the distributed nature of the capacitance will facilitate very low ESL values, the interconnect parasitics in the BLC are not expected to be as good as in MIM capacitance integration with the microprocessor.

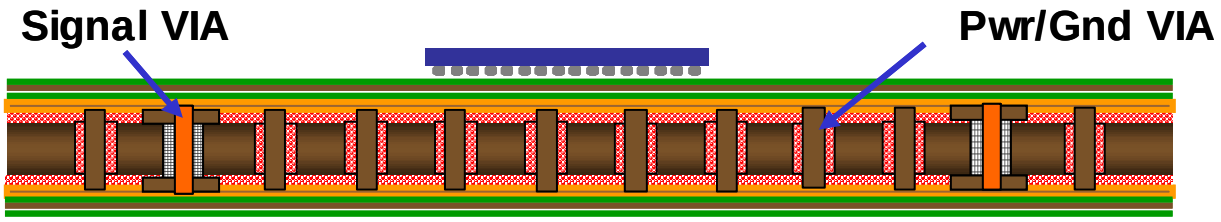

Figure 6: Substrate build-up layer capacitance (BLC, conceptual)

BBUL, a bump-less build-up layer technology internally developed by components research has demonstrated significant benefits in the lowering of the overall power path loop inductance [14]. BBUL eliminates the loop-L contribution of the substrate core (a significant portion in the loop-L budget) and the C4 bumps[2] of the current packaging architecture and is similar in power delivery benefits to 'core-less' architectures under investigation while simultaneously offering greater package routing scalability and component integration capability.

Power path series resistance scaling & reduction

Power path resistance losses are both a significant system efficiency reduction mechanism and a heat generation process within the boards, sockets and packages that support the microprocessor. The large current values anticipated in future microprocessors (as much as 150A) are forcing greater pin counts in sockets, larger copper thicknesses in boards, and are arresting the thickness reduction (for aspect ratio control in fine trace widths) for metal layers within substrates. This is perhaps a more difficult challenge as compared with loop-L reduction.

Consider a hypothetical processor consuming 40A of current at 1.75V with a power path series resistance of 3.12mΩ. Because of the 125mV voltage drop in the series resistance, the power delivery system (VRM) has to supply the 40A of current at 1.875V. More importantly, the power loss in the interconnect resistance is as much as **5W**, which incidentally is as much as 26.67% of the power spent in the VRM assuming 80% efficiency in the VRM functionality.

Let's assume that the platform design requires that the loss in the power delivery interconnect remains constant. As currents scale by a factor determined by one of the two examples chosen previously, the resistance will need to reduce by the square function of the current scaling factor to satisfy the constant interconnect power requirement, or

$$S_{ppr} = \frac{1}{S_I^2} \qquad (6)$$

Where S_{ppr} is the power interconnect resistance scaling factor and S_I is the current scaling factor. It can be seen that the **"Q"** scenario for current scaling will require a series path reduction by a factor of **~0.175** and the **"T"** scenario will need an S_{ppr} of **~0.5**. Perhaps the **"T"** scenario is more manageable under the constraint of constant interconnect power dissipation as can be seen in Figure 7. It may well be argued that given the 2X increase in power consumption in the microprocessor in the **"Q"** scenario, the interconnect dissipation also ought to be allowed that scaling, in which case the resistance scaling S_{ppr} will change to **~0.35** which is still worse than the scaling necessary in the **"T"** scenario; the power dissipated in the interconnect resistance will eventually need to saturate given the difficulty in cooling segments of metal embedded in material of poor thermal conductivity.

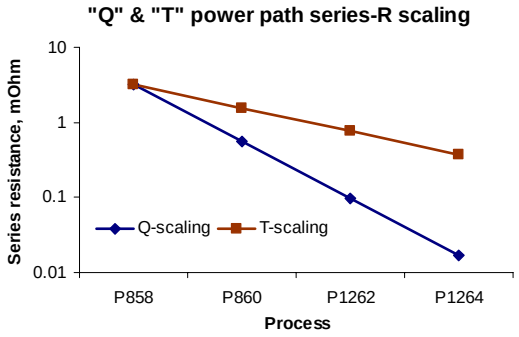

Figure 7: Supply path resistance scaling

The board, sockets, interposers and substrates in sum contribute to power path resistances. A 2X reduction in series path resistance to satisfy the least aggressive scenario in current scaling may lead to packaging solutions that are not cost-effective.

A potential solution to the high power delivery path resistance is the "**Separable–OPVR**" architecture[3] internally referred to as the ARARAT development [13]. This architecture (Figure 8) bypasses the motherboard and the socket for high current delivery by attaching a module integrating power conversion devices above the package and conveying power through a card edge connector between the module and the CPU substrate thus replacing the power pathway through the socket. Because power flows directly onto the substrate at close proximity to the CPU die, it is believed that the series resistance in the power path will be reduced in this architecture. The architecture also proposes that the cooling solution for the CPU and the power conversion devices be combined into a single heat sink because of the proximity of the components.

The best that can perhaps be done in minimizing the power path resistances and associated losses is the integration of the power conversion circuitry and components along with the microprocessor circuitry on the same semiconductor substrate. This is the focus of the **MVRM** work presently in progress at Virginia Tech [18]. This approach brings additional challenges in the integration of 'high' voltage (5 V as compared with the logic 1 V process) devices on the

[2] Controlled Collapse Chip Connection, a fine-dimension, array interconnection technique between microchips and substrates

[3] OPVR: On-Package Voltage Regulator

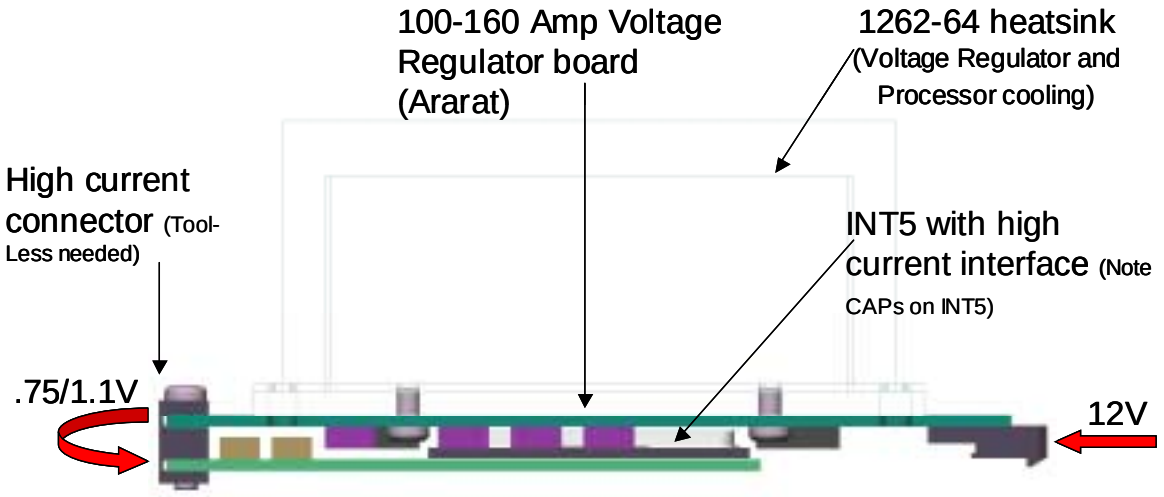

Figure 8: Ararat power delivery architecture (Courtesy Rick Nolte, PED)

same die as the high performance low-voltage logic devices and circuits, as well as the increased area and power consumption in the VLSI component due to the integration and operational efficiencies of the power conversion components. Additionally, the integration of large values of inductors and capacitances necessary for efficient power conversion continues to be a significant challenge as pointed out by Gordon Moore more than 3 decades ago [1].

Our approach, the **Silicon Sandwich**[4] (Figure 9) architecture currently in a phase–1 paper study, is to integrate the power conversion components into an active silicon layer called the "Silicon Transformer" that is bonded circuit–side to circuit–side with the microprocessor die. This architecture while *minimizing the power path series resistance* also facilitates *two–sided cooling* solutions. Instead of the logic process, the Silicon Sandwich allows the use of a fabrication process for the power conversion circuits that is different and potentially of lower cost. This architecture affords much *more area* for the *integration of micro-fabricated inductances* created as a layer above CMOS circuitry [20] and *high–K film based filter capacitances* such as MIM caps.

The Silicon Transformer[5] employs a *hybrid voltage regulation module (HVRM)* consisting of high frequency *switched buck converters (SBC)* designed for high efficiency in converting power from high to low voltages and *distributed charge valves* designed to manage high–Q transient events and thereby reduce the load filtering capacitance requirement [17, 19]. A key point to note is that the Silicon Sandwich architecture collapses the power delivery architecture into a highly scalable 2–resonant–loop design that could limit the voltage droops to 10% primarily due to the presence of the large, distributed filter capacitance (≈25µF) integrated into the Silicon transformer and the extremely low inductance in the CPU to Silicon Transformer interconnect that can scale along with the silicon processes.

The **thermal benefit** is another key advantage of the Silicon Sandwich architecture. Because silicon is a good conductor of heat, the use of copper to copper bonding between the Silicon Transformer and the CPU leads to very efficient two–sided cooling of the architecture. Simulations show that the maximum junction temperature for a Northwood power map reduces from 96.9°C to 80.1°C, despite a 25% power consumption increase due to the power conversion circuitry, which equates to a θjc reduction from 0.43 for the POR to 0.29 °C/W for this architecture [21].

Leakage in Future Generations

Fine device channel and gate oxide dimensions lead to leaky devices increasing static power consumption that could be 1/4 the chip power in the 95nm process technology and as much as *half of the total chip power* [4, 5] *around the 60nm node* (2008 per the ITRS, [3]).

[4] The Silicon Sandwich integrates all the components for power conversion into a multi–component active interposer that is bonded to the CPU and 'sandwiched' between two heat sinks. The name derives from the structure and the many technologies integrated.

[5] The *Silicon Transformer* serves both the power conversion function and the function of transforming fine CPU interconnect die to a typical *printed wiring board (PWB)* through, say, land grid array (LGA) contacts that allow separability of the assembly and PWB.

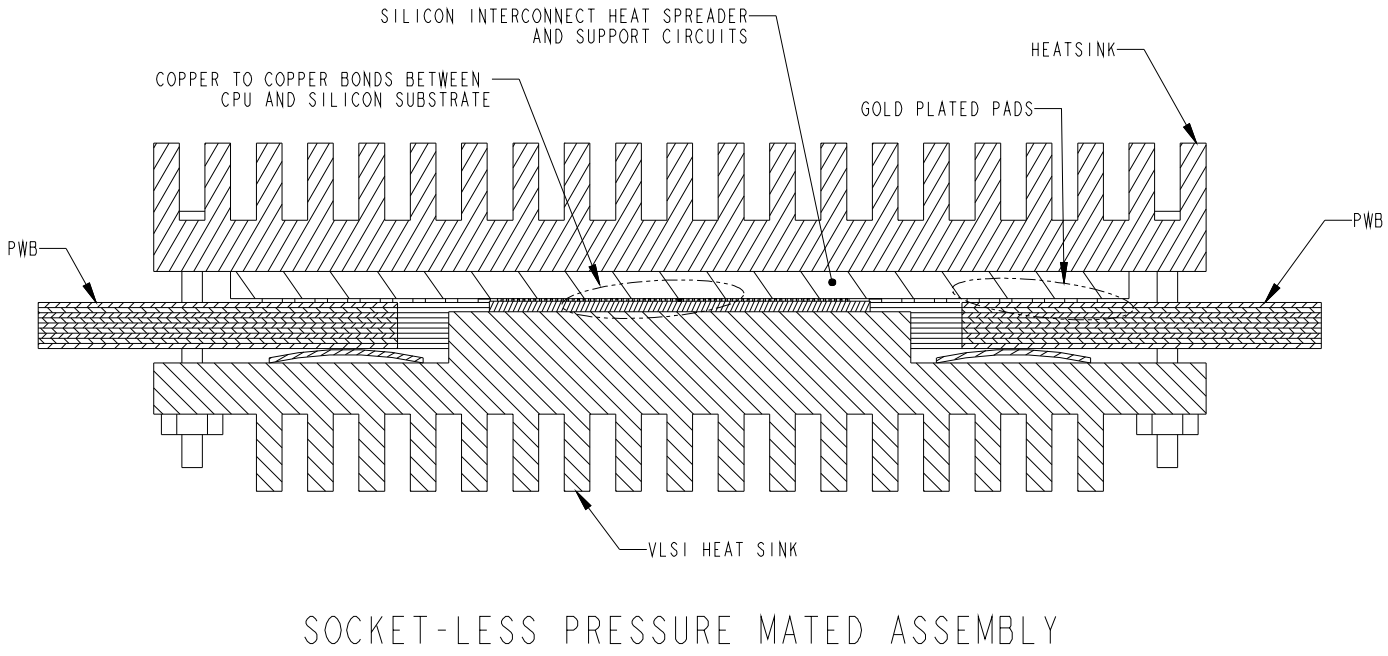

Figure 9: Conceptual drawing of the Silicon Sandwich assembly architecture

Leakage currents therefore will add an additional burden to assembly technology components that is not comprehended in either of the two current scaling scenarios discussed previously and could well be comparable in magnitude to the dynamic currents anticipated. Of interest is a look into some means of utilizing some of the otherwise wasted static power for computation.

Studies have shown [6, 7, 8] that **current-mode logic** (CML) circuits show better performance for power consumed as compared with typical CMOS logic circuits. CML circuits also utilize multiple devices stacked between VCC and VSS effectively increasing the impedance presented between PWR and GND and consequently reducing leakage related power wastage. Although not investigated specifically as a solution to leaky CMOS circuits, CML could perhaps be a solution to this problem. *Sub–threshold* drain-to-source leakage currents are the dominant component of leakage in DSM devices and are highly dependent on temperature. Stacked configuration circuits employed in CML circuits may potentially minimize wasted static sub–threshold currents.

Another very significant advantage of CML circuits is their *noise tolerance* and *very low-noise generation* characteristics that makes these circuits very suitable for mixed-signal integration. Analog circuits are seen to be increasingly necessary within high-speed microprocessors as for example in high-speed point-to-point links that employ analog signal processing circuitry. CML *circuit performance* is also *adaptively tunable* through the control of the bias currents that determine their speed; this characteristic allows both *process, voltage and temperature invariance* of performance as well as the flexibility of continuously modifying circuit and system performance for optimal power/performance management.

Although leakage currents as such do not burden the transient performance of microprocessor power delivery today, techniques for leakage control that switch these currents off may well be a cause for concern given that leakage is expected to match dynamic currents in the future. Minimizing these currents or utilizing them in a useful manner is therefore key to avoiding greatly increased assembly technology burden in future processes.

Bandwidth Hungry Processors

As device integration and operating frequencies continue increasing relentlessly, traditional processor communication bus structures are becoming the next bottleneck limiting platform performance. The front-side-bus (FSB) that connects multiple processors and a bridge component is seen to be running out of steam at **532 MT/s** on it's current data width, or at **4 GB/s** in maximum data transfer capacity. This limit is reached due primarily to the multi–drop or shared bus-wiring architecture that implements the FSB. Platforms are hence moving into point-to-point bus configurations for the processor communication buses that eliminate the multiple branches on the bus and hence allow for scaling to higher bandwidth.

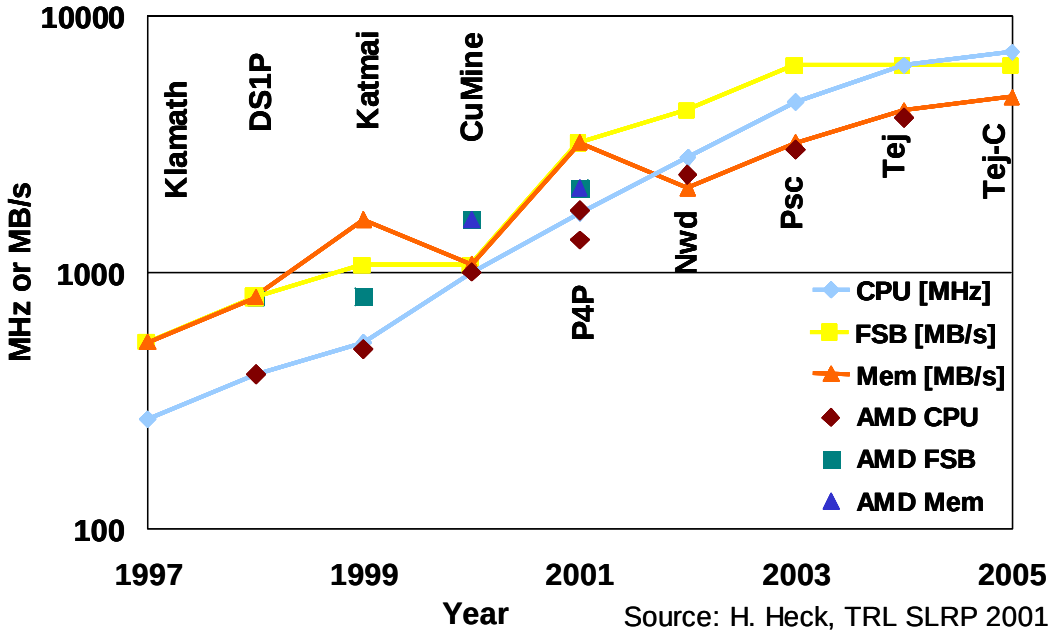

Figure 10: Bus data transfer rate progression

Examples of such point-to-point (P2P) buses include the transformed FSB and a processor backside bus that connects to a level–3 cache memory. The new FSB is expected to operate at multi-Gb/s rates following CPU core frequencies (Figure 10). A high data transfer rate per pin coupled with the requirement for many of these buses per chip due to the transition to point to point inter-chip interconnect architecture places significant additional challenges in the design of the package substrates, interposers, and sockets.

Impedance discontinuities

As signals are routed from the output of the CPU die driver circuits through the substrate, an interposer if necessary, the socket, and through board routing layers and connectors, a number of minute segments of interconnect that do not display the ideal 50Ω impedance desired are encountered. Specific to assembly components, these are, for example, non-ideal trace and return path impedances through the substrate, the inductance and capacitance of the core plated-through-holes (PTH) in the current substrate architecture, the socket interconnect, the interface between the socket and the board etc. collectively referred to as **reactive discontinuities**. For example, with data rates of 5Gb/s and a 500mV signal swing, the edge voltage rate of change is 10V/ns for a 50ps rise/fall time. A lumped capacitive discontinuity of 1pF at the output of the driver will now require a current $i(t) = C\frac{dv}{dt}$ = 10mA during the bit transition simply to charge this capacitance at the desired voltage ramp rate while the transmission line, in ideal terminated form, requires a current of 10mA (500mV/50Ω). Hence the presence of such a discontinuity along the interconnect channel is certain to degrade the edge-rate for the above example significantly, thus closing the eye both in time and potentially, amplitude. P2P links running at 2.5Gb/s and beyond also utilize true-differential signal pairs, which amplifies the impact of these discontinuities both in attenuation of signals and in delay variations between the signals in a differential pair. *Minimizing the value or impact* of these reactive discontinuities within assembly components is hence critical to enabling very high speed P2P links.

Link energy loss is typically dominated by dielectric and skin effect losses called **resistive losses**. Since skin effect losses increase proportional to $\sqrt{f}$, whereas dielectric ($\tan\delta$) losses tend to increase much more with frequency for typical materials used in PCB layers, the dielectric losses dominate beyond a few GHz. Hence though package trace 'DC' and 'AC' or skin resistances continue to increase as substrates scale signal trace dimensions, losses in a link are dominated by trace losses in the board. With resistive losses in the link compensated for through amplification and other circuit-based techniques, no immediate concern about losses within package substrate traces is apparent.

Resonant effects are a cause for concern for assembly components as frequencies increase and bit-cell durations shrink. The free-space time delay for electromagnetic waves is ~84.75 ps/inch; with an ε_r of 3.2, it takes about 152ps to traverse an inch of

trace length on a substrate. Counting in the delays due to discontinuities within the substrate, this could be ~200ps to 250ps within the substrate. With this delay corresponding closely to bit-cell widths for 4 to 5Gb/s signals, trace lengths within the substrate between discontinuities could potentially be excited into resonance resulting in much higher ISI. Greater loss within substrate traces could actually help squelch such effects since resonant reinforcement incurs twice the penalty of the length and loss. Transmitter and receiver **equalization** techniques implemented in circuits help compensate for the loss-induced signal eye distortion and inter-symbol interference (ISI). Inverse filter functions or passive equalization circuits can compensate for the filtering effects of channel reactive discontinuities. Ground surrounded *pin configuration design* is expected to help minimize the impact of socket discontinuities and crosstalk. *Embedded passive filters* through the integration of passive microwave components into substrates is an approach that could perhaps assist in compensating for substrate discontinuities.

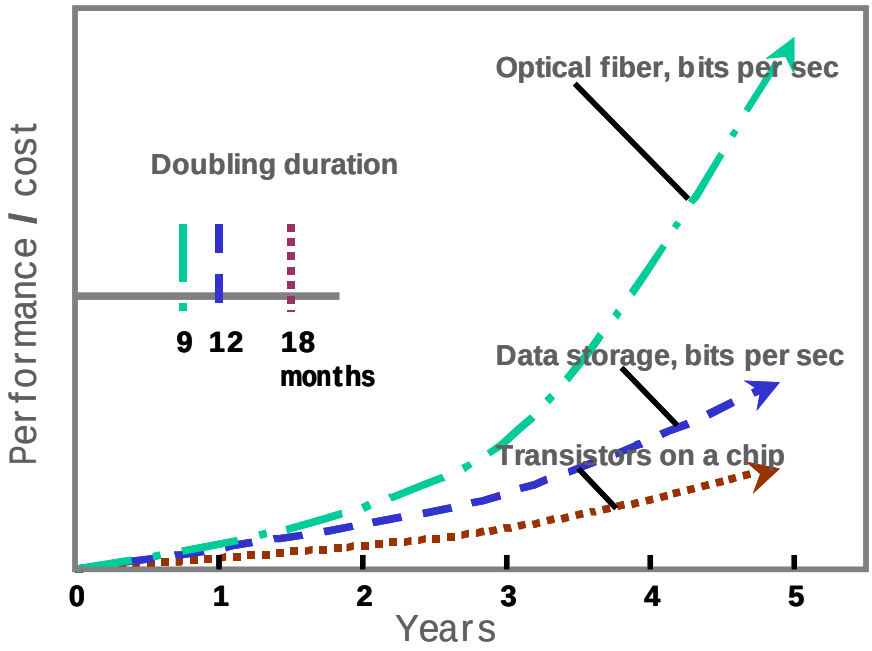

Figure 11: Progression of optical fiber interconnect [9]

Optical On-Board Interconnect

Figure 11 gives a view of the progression of optical interconnect in the data communications industry; it is seen that optical interconnect performance-to-cost ratio doubles in half the time for silicon [9]. With the expectation that costs will come down dramatically, a question often asked is of when and how optical interconnect is expected to come 'into the box', or be integrated into commonplace computing systems. The performance of a computing platform is determined largely by the raw processing capability of the microprocessor(s) within and the rate at which data can be continuously fed to the processor(s) to keep them busy. It is well known that optical data transmission medium provides bandwidth and data transfer capability supporting the requirements for the next few decades and perhaps more. Recognizing the limits of electrical interconnect due to link losses degrading the signal-to-noise ratio, it is logical to think that optical interconnect architectures could enhance computing platform performance.

This is not necessarily true in the context of small computing platforms, or 'boxes'. It has been argued in recent internal discussions (P1264 I/O ISTR) that latencies are less in optical interconnections for links over a few inches in length with some technology dependency [22]. This argument is based upon *time-of-flight* calculations showing that for refractive indices η of ~1.5, delays in optical links are about 125 ps/inch while electrical links exhibit delays of about 170 ps/inch, thus giving optical waveguides a 45ps advantage per inch. This calculation is based on the assumption that material of dielectric permittivity ε_r of 4 is used in the layers that form the electrical interconnect. The use of a material with ε_r of 2.25, when made practical, will nullify the benefit optical waveguides have over electrical ones at least as far as the flight time goes. Loss in electrical or copper-based interconnect, as discussed briefly previously, is dominated by the dielectric loss in the transmission medium, and replacement of the cheap FR4 material with the necessary low-loss dielectric layers for high-

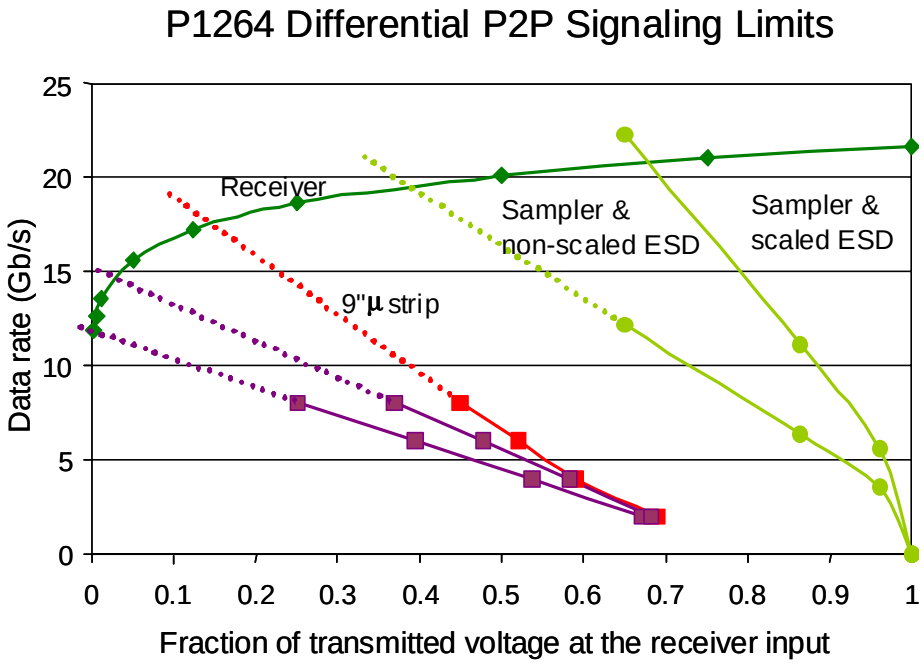

Figure 12: Signaling limits in the 65nm generation (R. Mooney [22])

speed interconnect will minimize that concern for a price. In short point-to-point interconnection architectures, therefore, it is not necessarily true that optical interconnections will be any better in performance. Additionally, latencies within the devices that communicate through P2P links (for data retrieval, packetization, transmission and reception) form a large portion the latency budget, particularly with shrinking platform form factors, rendering advantages gained in the interconnect physical link not very significant.

Figure 12 is an indication of the capability of P2P copper based interconnects in the 0.065-micron process generation [22]. Considering the impact of both circuit and interconnect limitations, it can be seen that a data rate of ~12Gb/s per wire has been determined to be feasible under the assumption that additional equalization built into the transmitters and the receivers will enable robust extraction of data despite full eye closure. Current architectural studies indicate that 5Gb/s per wire is sufficient for platforms in the P1264 generation. Hence copper-based interconnects are not viewed as a bottleneck to platform performance in the near future (5 years out).

Despite the promise of available bandwidth in P2P copper-based interconnects, scalability continues to be a challenge in P2P platform architectures, and P2P links introduce additional complexities in managing cache coherency and in physical design. A 4-CPU symmetric multi-processing architecture requires that each CPU be linked to the others, and this needs 4 links per CPU, and more than 4 links for the bridge component that interfaces with shared memory, graphics and other peripheral functions. Because high-speed P2P links are implemented through differential signaling techniques, the number of physical pins required to implement four 2 byte-wide bi-directional P2P links could total 500. The challenge in scaling the platform to 8-CPU and higher configurations is far greater in terms of assembly component requirements and board design; scaling leads to significantly increased latencies in alternate architectures that connect the bridge chips of two 4-CPU clusters to form an 8-way system.

An ***optical shared bus*** [23], on the other hand, could conceivably be implemented in a single layer on a board. Optical back planes are currently in development in the industry; back planes implement multi-drop parallel optical links. A wide, shared bus removes the processing delays associated with data packetization[6] and hence minimizes latency. Low-loss waveguide-to-package coupling techniques and waveguide material could additionally enable the capacity for a large number of drops on the bus. The loss in optical waveguides is independent of the rate of data transmission and hence the interconnect architecture could conceivably scale indefinitely in bandwidth. Additionally, optical wave-guides are well isolated from each other and can be routed at fine pitches with little interference between adjacent lines.

[6] Squeezing large data vectors through narrow (2 or 3 byte-wide) P2P buses requires serialization and de-serialization.

Choosing the correct package architecture (Figure 13 is an embodiment) to enable optical interconnects is another challenge. Given the widely varied nature of the devices (VCSEL arrays, high-speed drivers, photo-detector arrays, trans-impedance amplifiers etc. made on non-CMOS semiconductor substrates) that enable opto-electronic links, the current logical choice appears to be a system-on-package solution. Assemblies that are BBUL-like and the Silicon Sandwich architecture that integrates multiple chips into a package are candidates for the development of opto-electronic assemblies for computing platforms; silicon integration promises lower costs in the future.

The transition to P2P links in copper-based buses in computing platforms has occurred due to the fact that the multi-drop, wide-bus copper-based interconnects fail beyond a certain bus transfer rate. It is here, in multi-drop applications with wide buses that copper breaks; it is here that optical links could find a niche application, perhaps in high-end multi-processing computing systems, moving in the future to desktop systems as the technology matures.

Figure 14 illustrates how optical wave guide conduits and branches may be configured to couple to arrays of photo-emitters and photo-detectors integrated into packages. A key challenge in developing optical interconnect replacements for electrical interconnect is maintaining the capacity for feeding into and deriving energy from a single conduit, or in essence making the junctions bi-directional while ensuring low loss and the ability to tap fractions of the energy flowing within the conduit. The racetrack architecture achieves this by constraining the flow of photons to be unidirectional along the concentric wave guides and feeding into and tapping out of the conduit through carefully designed ‘entry’ and ‘exit’ ramps that are in the same plane as the conduits and terminate in 45° facets that redirect the photons normal to the plane of the guides and into the package. This architecture lends itself towards scaling to wide buses and a large number of ‘drops’. Key limitations of this architecture include the path length matching and ISI caused by unspent energy pulses returning to the transmit junction through the closed loop conduits.

While optical interconnect offers outstanding data transfer rate scalability, bus width scalability, multi-drop capability and potentially lesser power and EMI concerns, platform performance necessarily requires large amounts of low-latency shared memory accessible by the processors in a multi-processor architecture. It is critical therefore that memory device or memory array module technology is advanced to match the performance of CPU’s and the high-speed buses that link them together. Perhaps the future will see a large memory module with low-latency opto-electronic interface capability sitting directly on a shared optical front-side bus that connects the processors

Conclusions

Recent disclosures have indicated that 30nm and 20nm devices are well within Intel process technology capabilities. Process scaling is therefore expected to continue for at least another decade.

From an assembly technology standpoint, the burden that microprocessors place on assembly components will increase at an enormous rate if chip power continues to increase while operating voltages scale down. Containing power consumption while taking advantage of scaling for higher frequencies and better performance will hence be key to cost-effective assembly technology solutions. In short, responsible design practices will need to be followed to obtain performance through exceptional optimization and innovation in architecture and circuit design rather than through power consumption increases and die-size growth. It is also critical that the looming issue of leakage in CMOS circuits be addressed proactively before that turns into a possibly unmanageable power delivery hurdle for assembly technology & platform design.

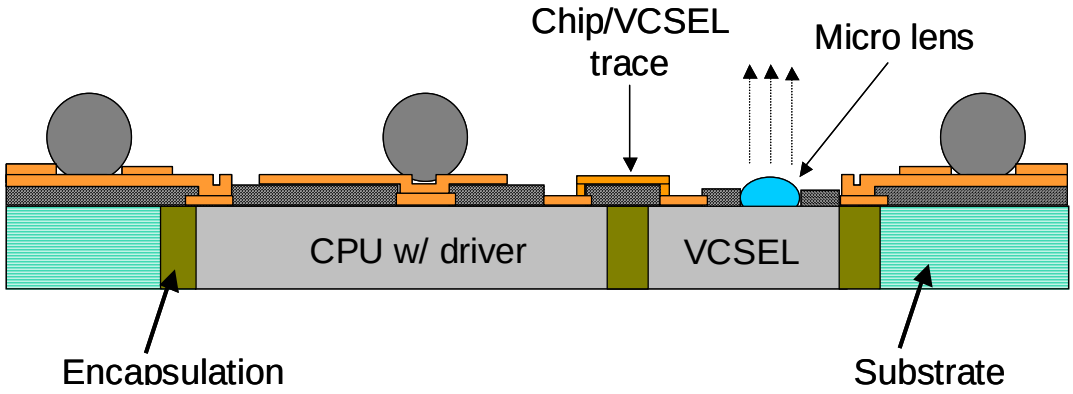

Figure 13: Optical component assembly integration (J. Swan, ATD)

Though recent investigations show that high-speed signaling at rates well beyond platform requirements is feasible in the 1264 timeframe and beyond, it is believed that minimizing total interconnect length will be key to reducing latencies and increasing overall performance as chips speed up. Package and socket interconnect will in this situation play greater roles in determining the bandwidth of these links. Hence advancing the assembly interconnect architecture minimizing the number of discontinuities along high-speed signal pathways and simultaneously improving the interconnect technology for these components will be critical in future generations.

There are clear indications of the implausibility of the upward trend of platform performance, particularly for the commonly used desktop platforms and the current system architectures, due to power delivery, thermal as well as interconnect performance limits and associated cost constraints. Arrays of low-power specialized processors at close proximity interconnected through very low latency and highly scalable low-power buses could be considered a more feasible architectural direction for platform performance scaling. Wide, shared optical buses could satisfy the interconnect bandwidth and scalability requirements; hence reducing component and assembly costs for such links enabling scalable low-latency buses could, in the architecture described above, be key to moving computing to a new plateau of performance and scalability. Developing innovative assembly solutions enabling such platforms both at the component level and the system level will hence be critical in continuing to advance computing performance while keeping costs low.

Acknowledgments

Thanks to *Vasu Atluri* and the iATTJ committee for inviting a paper from ATD-Pathfinding this year.

Much thanks also to *Roy Pulvino, Kris Frutschy* and *Ram Viswanath* for helping out at short notice with various architectural illustrations, to *Henning Braunisch* for deriving analytical solutions to portions of the power delivery architecture as well as for a thorough proofreading, *Larry Mosley* and *Ed Stanford* for many power delivery discussions and conversations, *Kim Merley* and *Lesley Polka* for information and for thought-provoking debates, *Gil Vandentop* and *Chuan Hu* for discussions and modeling on the Silicon Sandwich architecture, *Valluri (Bob) Rao* for many productive discussions on optical interconnect architectures, *Johanna Swan* and *Sam Calvin* for team discussions and activities relating to optical packaging and interconnect at ATD, *Siva Narendra* and other *CRL folks* for discussions on the sandwich implementation and sharing the CRL investigations into high-bandwidth switching and linear buck converters, and many more of my past and present colleagues for contributing to the thoughts and solutions discussed in this paper.

Finally, thanks to *Bala Natarajan* for his input and continual encouragement.

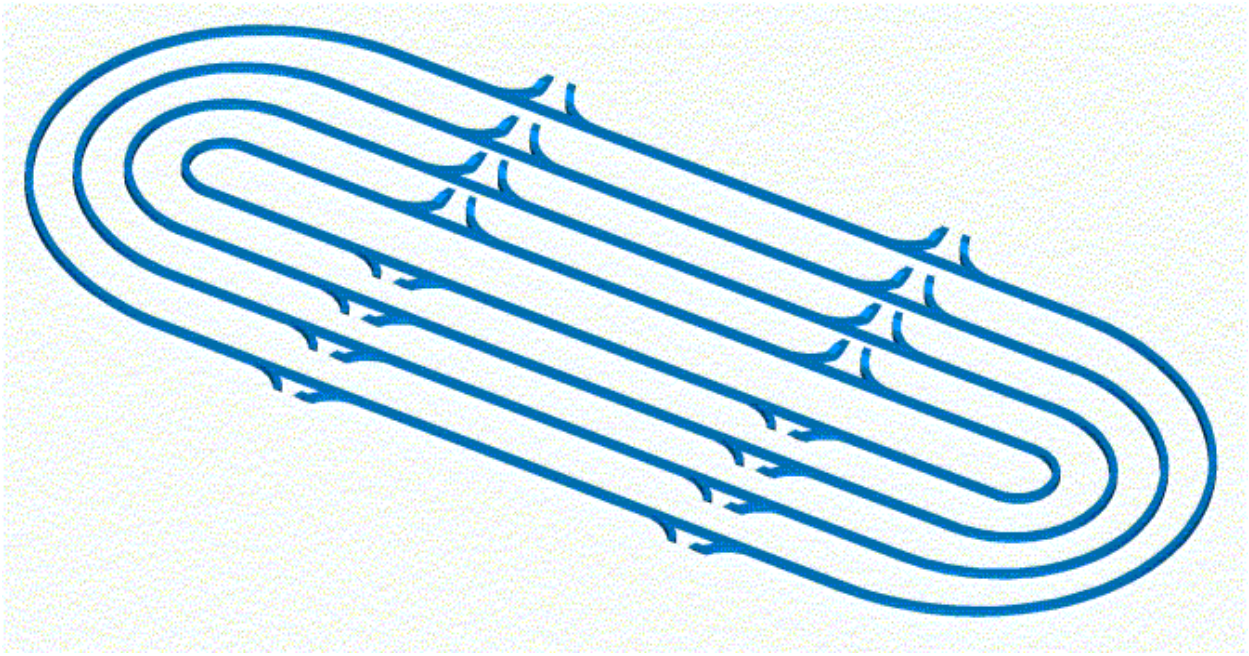

Figure 14: The 'Racetrack', a conceptual optical multi-drop bus interconnect

References

1. "*Cramming more components into integrated circuits*", Gordon E. Moore, Electronics, Volume 38, No. 8, April 19, 1965
2. "*Moore's Law*", Intel research document online, http://www.intel.com/research/silicon/mooreslaw.htm
3. "*Assembly & Packaging*", International Technology Roadmap for Semiconductors, 2000 update, http://public.itrs.net/Files/2000UpdateFinal/AssemblyPkg2000final.pdf
4. "*New Microarchitecture Challenges in the Coming Generations of CMOS Process Technologies*", Fred Pollack, Micro32 Keynote presentation, 1999
5. "*Moore's law and Microprocessors*" Shekhar Borkar, invited paper, iATTJ 2000 vol. 3
6. "*Dynamic Current Mode Logic (DyCML), A New Low-Power High-Performance Logic Family*", Allam, M. W. and Elmasry, M. I., IEEE 2000 Custom Integrated Circuits Conference proceedings, pp. 421-424
7. "*MOS Current Mode Logic for Low Power, Low Noise Cordic Computation in Mixed Signal Environments*", Musicer, J. M. and Rabaey, Jan, ISLPED 2000, pp 102-107
8. "*A 10Gb/s Demultiplexer IC in 0.18µm CMOS using Current Mode Logic with Tolerance to the Threshold Voltage Fluctuation*", **T**anabe, A., et. al., ISSCC 2000, pp 62-63
9. "*The Triumph of The Light*", Gary Stix, Scientific American January 2001 issue http://www.scientificamerican.com/2001/0101issue/0101stix.html
10. "*Managing the Impact of Increasing Microprocessor Power Consumption*", Stephen H. Gunther et. al., Intel Technology Journal Q1 2001, http://developer.intel.com/technology/itj/q12001/articles/art_4.htm
11. "*The Evolution of Microprocessor Packaging*", Ravi Mahajan et. al., Intel Technology Journal, 3rd Quarter 2000, http://developer.intel.com/technology/itj/q32000/articles/art_1.htm
12. "*Willamette and Coppermine on-die DECAP removal experiment reports*", Arabi T. R. et. al., Intel internal documents, 2000
13. "*ARARAT Separable OPVR Development*", Nolte, Rick, PED Power Delivery Architecture development, Intel Internal documents, http://pedforepg.patch.intel.com/pages/project.htm
14. "*Bumpless Build-Up Layer Packaging*", Towle, Steven N., Braunisch, Henning et. al., ASME 2000 pre-print, Intel Internal document
15. "*Substrate High-K Build-up Layer Capacitance*", Nair, R., 2001 ATD-Pathfinding proposal, Intel Internal document
16. "*Silicon Sandwich: A New Assembly Architecture for Power, Thermal and I/O*", Nair, Raj & Vandentop, Gilroy, Aug. 2001 Intel Internal project proposal document
17. "*Low Cost On-Die CMOS Distributed Voltage Regulation for Microcontrollers*", Nair, Raj & Tang, Bao, ESSCIRC 1998 Proceedings
18. "*Monolithic Voltage Regulation Module development report*", Huang, Alex & Lee, Fred C., August 2001, SRC funded project report
19. "*Distributed Charge Valves*", Nair, Raj; 0.1µm ISTR Power Delivery WG investigations, November 1999
20. "*High Frequency (GHz) and Low Resistance Integrated Inductors Using Magnetic Materials*"; Gardner, D. and Crawford, A. M., IEEE IITC Proceedings, June 2001
21. "*Study of the Thermal Performance of the Silicon Sandwich Package*"; Hu, Chuan, Intel Internal project task report
22. "*P1264 I/O ISTR Investigations Summary*"; Mooney, Randy; Intel Internal WG report, 2001
23. "*Optical Shared Bus*", Nair, Raj & Calvin, Sam, TRL SLRP 2001 contributions.