Decoupling Solutions

John Prymak, Erik Reed, Michael Randall, Peter Blais, Bill Long

KEMET Electronics Corporation, 2835 KEMET Way, Simpsonville, SC 29681
Phone: +1-864-963-6573, FAX: +1-864-967-6867
e-mail: johnprymak@kemet.com

Abstract

Integrated circuit (IC) technology has advanced over time, requiring innovative decoupling solutions that address decoupling needs over the entire power path from the main power supply to the logic gates on the IC. As one follows the path from the power supply to logic gate, high-frequency content increases, required capacitance falls, and the importance of very low inductance (ESL) increases. Impedance (Z) must also remain low and stable over the frequency range of interest so that detrimental resonances are avoided.

Four tiers or levels of decoupling have been defined along the path from the power supply to the logic gates on the IC: (1) the voltage regulator level, (2) the board decoupling level, (3) the IC package decoupling level, and (4) the IC on-die decoupling level. Each level has impedance requirements that are best addressed by capacitor technologies which differ in capacitance density, capacitance stability, ESL, ESR and cost.

This paper addresses the relative performance of competing capacitor technologies for board level (2) and IC package level (3) decoupling. New decoupling paradigms are presented for both of these levels.

The Power Delivery Network (PDN)

Power delivery to high-speed microprocessors cannot be accomplished using any single device or component. Large current changes (30 to 100 A) can occur during time durations from seconds to nanoseconds. Proper microprocessor operation requires the voltage remain stable during those current transients (typically less than ±0.1 V variation around a 1-V supply) The power supply alone cannot maintain that voltage stability at those current levels and the shorter time scales due to limited bandwidth. Decoupling capacitors are required in the power delivery network (PDN) between the power supply and the microprocessor to maintain voltage stability.

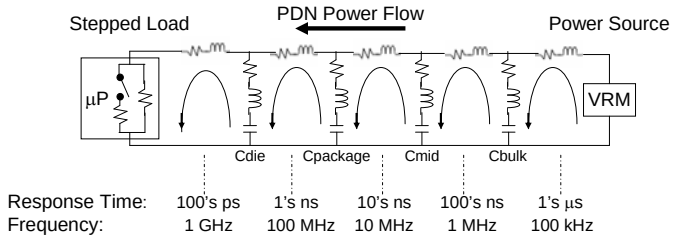

Figure 1. Progressively faster loops approaching the microprocessor (Load).

Figure 1 shows a typical decoupling scheme used for microprocessors today. Several types of capacitors are required. These capacitors are used in progressively faster loops as the PDN approaches the microprocessor. The extreme right of the diagram shows the power supply or voltage regulator module (VRM). The best VRMs with bandwidths of 100 KHz are able to respond to current changes on timescales of µs to 10's of µs. Moving to the left, the capacitors of Cbulk need to respond on time scale of 100's of ns or at a frequency of up to 1 MHz. The next loop is handled by the capacitors of Cmid (for mid-frequency) that respond in 10's of ns or up to 10 MHz. The capacitors for Cbulk and Cmid are located on the system motherboard. The final loop that can be chosen by the designer of the PDN is Cpackage (for the capacitors located on the microprocessor package). The response time needs to be in the nanosecond range or in the frequency domain up to 100 MHz. The final loop represents the capacitors on the microprocessor die (Cdie).

Through many years of trial, the capacitor types chosen for Cbulk are aluminum or tantalum electrolytic capacitors and have a combined capacitance in the 1000's of µF, an ESR of a few milliohms, and an inductance on the order of a nH. For Cmid, high-value MLCC capacitors with a combined capacitance of a few 100 µF, sub milliohm ESR, and inductance of 10's of pH are chosen. For Cpackage, low- inductance MLCC capacitors with a combined capacitance of 10's of µF and inductance on the order of a few pH are chosen. Within a loop, multiple capacitors are needed to satisfy requirements.

©2008 ECA (Electronics Components, Assemblies & Materials Association), Arlington, VA

Proceedings CARTS USA 2008, 28th Symposium for Passive Electronics, March, Newport Beach, CA

The Cbulk and Cmid capacitors are normally found on the motherboard in close proximity to the processor. Because the inductance of Cpackage is so critical, these capacitors are mounted on the same small interposer PCB as the processor is mounted, above the pins or solder ball attachment, and as close to the silicon die as possible.

Faster and more powerful microprocessors have resulted in increased current demands and faster response times. There is also a desire to decrease circuit size in order to decrease overall system size or to add more features to the same size system. This article discusses new decoupling schemes at the motherboard or package level that are responsive to these requirements.

Theoretical Analysis of Motherboard-Level Decoupling Performance

Simple analyses of PDNs have been accomplished by drawing circuits that contain multiple clusters of capacitance connected together by series resistance and inductance (see, for example, Figure 1). These clusters of capacitance are located between the microprocessor and the VRM. Such a circuit is useful for illustrating that each of these clusters of capacitance responds to a step change in processor current after its own, progressively longer time delay until finally the VRM takes over.

Inherent in this simplified model is the idea that charge delivery by these capacitor clusters may not smoothly overlap, so there could easily be peaks and droops in the voltage response at the processor during the critical response time interval. Figure 2 illustrates this schematically. The ideal response to the step load in current is shown by the dashed line in the voltage plot. The effect of mismatched impedances causes non-uniformity in the voltage response, as shown by the solid line. Unfortunately, this analysis approach does little to quantify the expected voltage response in an intuitive way that mimics actual decoupling behavior. Thus it is not a very useful tool beyond communicating the general nature of the PDN and explaining why lower-ESL capacitors are located closer to the processor and higher-ESL capacitors are located farther away.

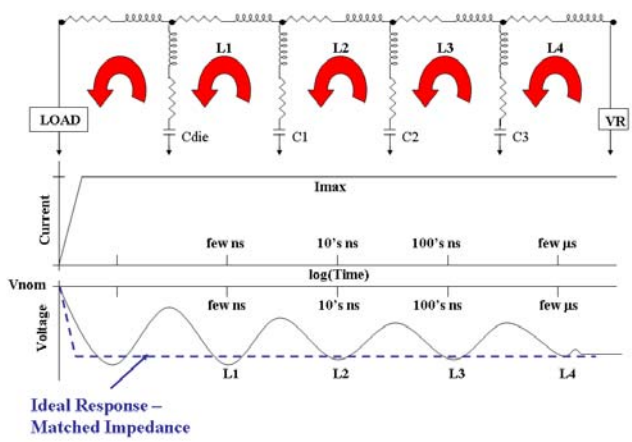

Figure 2. Ideal and Non-Ideal Responses of a Microprocessor PDN[1].

To fill this analysis gap, there is a powerful theoretical strategy for analyzing the decoupling performance of a PDN. The strategy is to characterize the impedance looking into the PDN from the microprocessor's perspective and then calculate the time-domain step response of the voltage across this impedance to a step change in load current[2].

Specifically, the frequency-domain complex impedance looking into the PDN from the processor's point of view is either predicted from models of the capacitors and circuit board or measured on actual motherboards with an impedance analyzer. This frequency-domain complex impedance is then transformed into the time domain via inverse fast Fourier transform (IFFT) to discover the impulse response of the network. The impulse response is then integrated to discover the step response, and the step response is multiplied by the anticipated current step amplitude.

The end product of this procedure is a time-domain plot of the theoretically-expected response of the PDN to a step change in microprocessor load current. If the impedance prediction or measurement is accurate, the resulting step response must also be accurate.

After applying this analysis over a variety of cases, it quickly becomes clear that the best step response is found in PDNs that present the processor with the most constant impedance spectrum. Of course, most real-life PDNs are not perfect, and their non-constant impedance gives rise to non-ideal step response. This will be demonstrated presently, but first an example is given of a nearly ideal board-level decoupling solution.

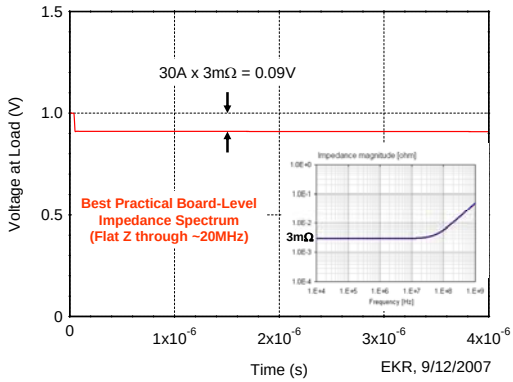

Figure 3. Time-Domain Step Response of an Idealized 3mΩ VRM, Bulk Decoupling, and Mid-Frequency Decoupling Solution Subjected to a 30A Current Step.

The solution presented in Figure 3 is not a real design, but rather is an idealized conception that presents a microprocessor in a notebook computer with a constant 3 mΩ impedance up to roughly 20 MHz before becoming inductive. This frequency range is assumed to cover the response of the VRM, the bulk decoupling capacitors, and the mid-frequency decoupling capacitors, all of which are located on the motherboard. Above 20 MHz it is further assumed that decoupling is smoothly handled by processor package and die capacitance.

Figure 3 contains two graphs, a small frequency-domain impedance plot inset in a graph of the resulting time-domain step response. The impedance plot extends to 1 GHz. The time-domain plot extends to 4 µs which is long enough to capture any non-ideal response that might be caused by decoupling capacitors but not be corrected by the VRM.

Observe that the impedance of the idealized decoupling solution of Figure 3 is 3 mΩ and that the step current is 30 amperes. Thus it is expected that the voltage will shift downward by 0.090V which is what is seen in the time-domain graph. No unusual undershoots or oscillations are observed. This is the expected ideal voltage response of a constant-impedance decoupling solution.

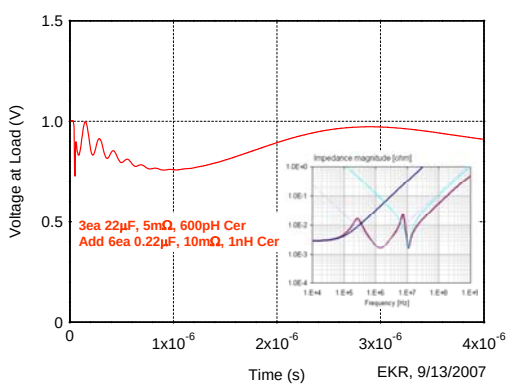

Figure 4. Time-Domain Step Response of a 100 kHz BW VRM plus Three 22 µF, 6 mΩ, 600 pH MLCCs and Six 0.22 µF, 10 mΩ, 1 nH MLCCs Subjected to a 30 A Current Step.

Figure 4 presents analysis of a much less ideal board-level decoupling solution. Even though the VRM has very impressive performance with control bandwidth of 100 kHz, it still does not provide fast enough response to deal with such a fast and large current step.

Three 22 µF, 6 mΩ, 600 pH MLCCs and six 0.22 µF, 10 mΩ, 1 nH MLCCs are connected in parallel with the VRM in an effort to extend and flatten the impedance curve up to a little over 1 MHz. The addition of these capacitors produces a better time-domain curve than the VRM would alone, but the impedance curve still has lumps

and bumps because of the sharp impedance response of the MLCCs. These deviations from constant impedance manifest themselves as oscillations in the time-domain response.

The slow oscillation is caused by the lower-frequency impedance peak while the faster oscillation is caused by the higher-frequency impedance peak. What is happening physically is resonance among the added capacitors. These resonances are excited by the current step and cause oscillations that are eventually damped by the ESR of the capacitors. Finally, the "too-soon" upturn of the impedance plot near 1 MHz causes the narrow downward spike in voltage at the very beginning of the time-domain response.

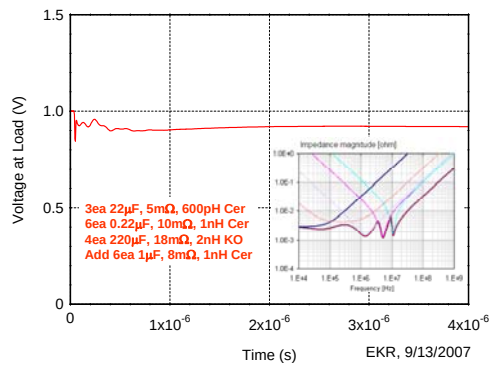

Figure 5. Time-Domain Step Response of a 100 kHz BW VRM plus Three 22 µF, 6 mΩ, 60 0pH MLCCs, Six 0.22 µF, 10 mΩ, 1 nH MLCCs, Four 220 µF, 18 mΩ, 2 nH Tantalum Polymer Capacitors, and Six 1 µF, 8 mΩ, 1 nH MLCCs Subjected to a 30 A Current Step.

In an effort to further improve the response of Figure 4, additional capacitors are added in parallel with the decoupling solution to further smooth and extend the impedance spectrum. Specifically, four 220 µF, 18 mΩ, 2 nH tantalum polymer capacitors and six 1 µF, 8 mΩ, 1 nH MLCCs are added. These specific capacitors are chosen because their impedance curves combine with those of the other capacitors to specifically blunt peaks at troublesome frequencies in the impedance spectrum.

The response with these additions appears in Figure 5. It is clear that the additional capacitors did the job of flattening the impedance spectrum and the resulting time-domain response is substantially similar to the ideal response. But altogether it took a lot of capacitors to achieve this result and the response is still not really ideal. This demonstrates the challenge faced by power distribution engineers as they attempt to produce acceptable response to step changes in current while trying to reduce costs and conserve board space.

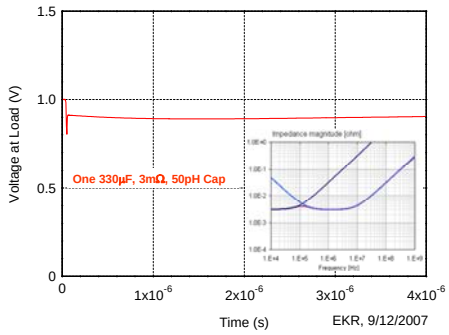

Figure 6. Time-Domain Step Response of a 100 kHz BW VRM plus One 330 µF, 3 mΩ, 50 pH Tantalum Polymer Capacitor Subjected to a 30 A Current Step.

A different decoupling solution appears in Figure 6. This solution consists of the same 100 kHz BW VRM and a single 330 µF, 3 mΩ, 50 pH tantalum polymer capacitor. Observe that the step response is nearly ideal except for the small undershoot at the leading edge. The nearly ideal response is produced by the low, constant impedance of the tantalum capacitor that extends up to almost 10 MHz before turning inductive.

Perhaps the leading edge undershoot could be adequately handled by package-level capacitance, or perhaps the inductance of the tantalum polymer capacitor needs to be further reduced for acceptable performance in real applications. But the important idea is that almost all of the bulk and mid-frequency decoupling task can be handled by a single capacitor rather than by more than a dozen capacitors, if that single capacitor has both low ESR **and** low ESL. If such devices can be produced with reasonable size, cost, and ease of application, they will likely populate the next generation of decoupling solutions.

Achieving Low ESR and ESL in Tantalum Polymer Capacitors

How might we achieve this low ESR and low ESL in a tantalum polymer capacitor? Much has been written on the subject of low ESR tantalum polymer capacitors[3]. The largest advance in ESR was moving from manganese dioxide as the primary cathode material to the intrinsically conductive polymer (most commonly based on polypyrrole or poly-3,4 ethylenedioxythiophene). However, to achieve even lower levels of ESR, significant modifications are also made to the anode, lead wire, lead frame, and carbon and silver layers. This has resulted in the availability of ESR values less than 5 mΩ in a 7343 package. Work continues to drive this ESR value even lower.

Up to recently, low ESL has not been an area of focus for tantalum polymer. However, as shown in the previous section, lowering the ESL to values of 50 pH can extend the useful frequency range for tantalum polymer for decoupling from less than 1 MHz to 10 MHz. At this point the tantalum polymer becomes useful not only for bulk decoupling, but also can replace some or all of the MLCCs used in mid-frequency decoupling.

The ESL is closely related to the loop area and width of the current path. Decreasing the loop area and making the current path wider will decrease the ESL.

Two-terminal tantalum polymer devices are normally packaged in plastic (injection mold process), and a leadframe is used to bring electrical connection from the capacitor's pellet structure out to a high-speed, pick-and-place plastic package for surface mount applications. That leadframe adds a lot of length to the current path for this device (Figure 7). At the anode connection it extends from the middle of the package where it contacts the riser wire, to the outside of the plastic, and is bent down and under the end of the body. At the bottom face of the device, this leadframe makes contact with the solder pad on the PCB. The cathode connection starts along the top of the pellet, is bent down to the middle of the plastic body, then out, down, and under to repeat the solder pad contact.

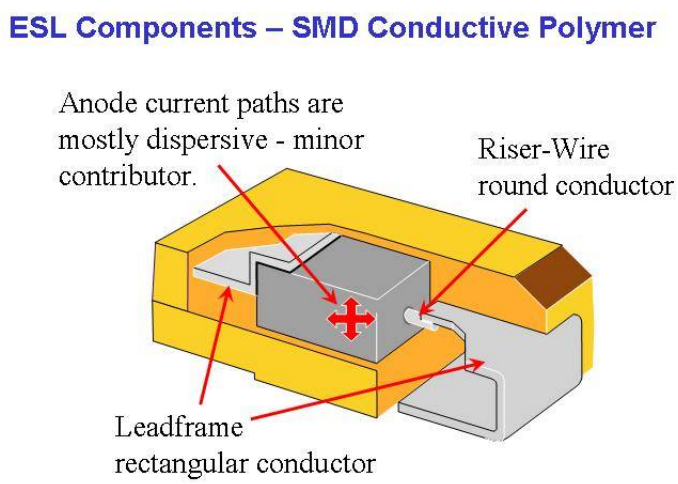

Figure 7. Standard, surface mount conductive polymer capacitor.

A new configuration of the surface mount tantalum capacitor has been developed in the last several years. This configuration is called the "face down termination" version. It has several advantages: better volumetric efficiency, lower ESL, and higher ripple current rating. Rather than a wrap around leadframe, the anode and cathode are connected to terminals at the bottom of the chip. The reason for the lower ESL is the reduced loop area in the face down configuration as compared to the standard lead frame version (Figure 8). Devices that embody these performance characteristics are discussed in the next section.

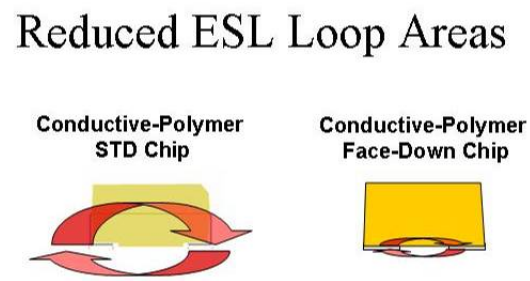

Figure 8. Illustration of reduced current loops created by facedown design.

New Solutions for Bulk and Mid-Frequency Decoupling

In the not-distant past, bulk decoupling for microprocessors was typically accomplished with wound-foil wet aluminum electrolytic capacitors while mid-frequency decoupling was accomplished with multilayer ceramic chip capacitors (MLCCs). As processor speed and power supply current continued to increase, the wet aluminum electrolytic capacitors were displaced by aluminum electrolytic capacitors with conducting polymer electrolyte. This change reduced capacitor ESR (equivalent series resistance) and capacitance roll-off with increasing frequency, both of which improved decoupling performance. Also, to meet the mid-frequency decoupling demands of the newer, higher-performance processors, more and higher-capacitance MLCCs were employed. The cost and required circuit board space occupied by the decoupling solution were growing.

Recently there appears to be a significant shift in thinking regarding bulk and mid-frequency microprocessor decoupling. New solutions are appearing which feature fewer, and often smaller, capacitors. The driver of this paradigm shift is the desire to control cost, maintain high performance, and minimize circuit volume.

Two facilitators of this paradigm shift are the evolution of voltage regulator module (VRM) technology and the development of very-low-ESL (equivalent series inductance) valve-metal capacitors (made from aluminum and tantalum). VRMs are now available with higher switching frequencies and wider control bandwidth, both of which reduce the required bulk capacitance. Valve-metal capacitors with very-low-ESL and very-low-ESR (equivalent series resistance) can now match or sometimes exceed the mid-frequency decoupling performance of MLCCs. With these new technologies, it is now possible to replace a large number of wound-foil aluminum electrolytic capacitors and MLCCs with just a few high-performance valve-metal capacitors while controlling costs and conserving board space.

This shift in thinking is still in its early stages and designers of microprocessor-driven systems have not been equally motivated to make the shift. Bulky low-cost desktop computer systems do not require much miniaturization, so there has not been much motivation to pursue a more space-efficient design unless substantial cost savings are possible. But in small, mobile systems (laptops, etc.), high-performance designs are appearing that employ substantially fewer capacitors. The few remaining valve-metal capacitors typically have much higher performance than their predecessors and significantly shorter device height. Because of the higher performance of the valve-metal capacitors, the number of MLCCs employed has fallen dramatically because they are no longer necessary. Overall, the bulk and mid-frequency decoupling solution is shrinking.

So just what are these new valve-metal capacitors and how do they work with advanced VRM technology to provide a high-performance decoupling solution in much less space while maintaining reasonable cost? Two broad categories of these capacitors are low-profile, very-low-ESL aluminum polymer capacitors and very-low-ESL tantalum polymer capacitors. In the first category are aluminum polymer capacitors with configurations similar to items A and B in Figure 9. In the second category are tantalum polymer capacitors with configurations similar to items C and D in Figure 9.

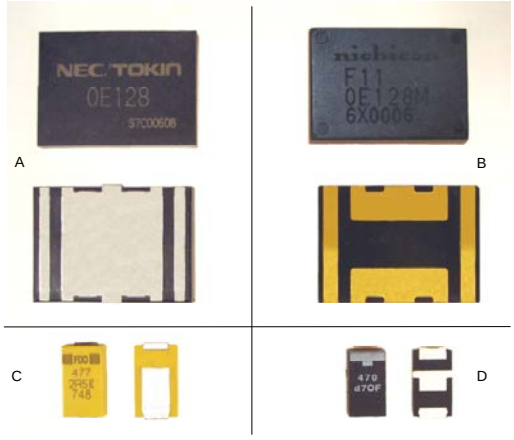

Figure 9. Photos of Low ESL Aluminum Polymer Devices (A and B) and Low ESL Tantalum Polymer Devices (C and D).

More details on ESL measurement and a comparison of performance of these devices is given in a companion paper[4]. Either of these devices could fulfill the requirements of Figure 6.

Each of these capacitors competes for the designer's attention with various desirable performance characteristics: very-low-ESL, very-low-ESR, moderately high capacitance, small footprint, low-profile, and modest cost. These all contribute to an extension of the low impedance capability of tantalum or aluminum polymer capacitors, up into the regions of the high capacitance MLCC capacitors. Ultimately, the designer will choose a particular solution based on a tradeoff between size, performance, and cost.

On-Package Decoupling

Decoupling capacitors are also required in close proximity to the microprocessor to handle decoupling on short time scales or at high frequencies (Cpackage). Multilayer ceramic capacitors are the choice for on-package decoupling as MLCCs offer the best combination of capacitance density, low ESR, very low ESL and cost for this application. Depending upon the design scenario, control of ESR may be important so as to provide impedance similar to the upstream power distribution network. Control of ESR may also be required to dampen resonance between microprocessor capacitance (Cdie) and on-package decoupling capacitance. Cost is also an important factor depending upon the application.

Package designers may opt to use two or more different types of on-package decoupling capacitors depending upon cost and value added to the design. These two types of MLCCs are normal or bulk decoupling MLCCs and high- speed decoupling MLCCs. Examples of each are illustrated in Figure 10. Bulk decoupling MLCCs are of traditional capacitor design while high-speed decoupling MLCCs are available in numerous formats that typically involve either reverse terminations or more than two terminals. The main objective in the development of high-speed decoupling capacitors has been to reduce ESL.

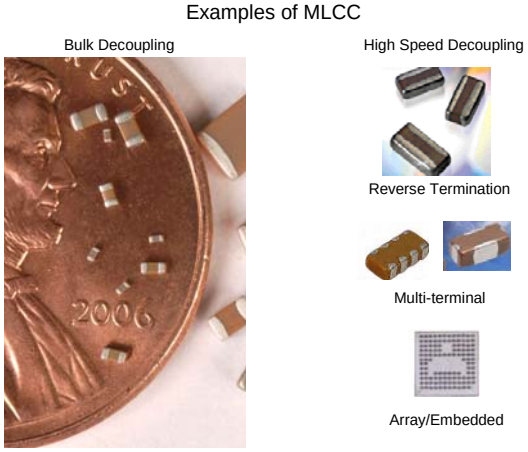

Figure 10. Examples of bulk decoupling MLCCs and high-speed decoupling MLCCs.

On-package decoupling capacitors may be placed either on the top of the microprocessor package or on the bottom of the microprocessor package as illustrated in Figure 11, or they may be embedded within the microprocessor package. Each of these locations has its associated advantages and disadvantages. Embedded capacitance for high-speed decoupling applications is not yet commercially successful.

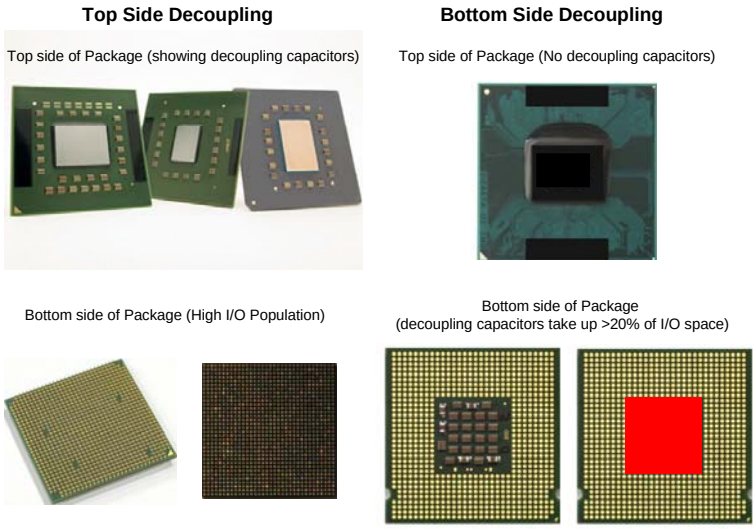

Figure 11. Top-side and bottom-side decoupling.

While advanced high-speed decoupling MLCC designs have enabled high-speed on-package decoupling to keep up with the advancements of microprocessor capabilities, the microprocessor package design community continues to search for decoupling solutions that are simpler and more cost effective. Cost-effective MLCC solutions that result in simplified microprocessor package design will gain favor in the design community. Simplified package design may be achieved by moving the MLCCs to the bottom side of the microprocessor package. This action enables to the designer to reduce or eliminate the number of power and ground fan-out layers in the multilayer microprocessor package, while leaving space on the top side of the package for lid-attach or heat-sinking structures. The reduced number of package layers will also enable the package designer to reduce or eliminate expensive and complicated package interlayer connections. Moving the MLCC from the top of the package to the bottom of the package also may enable the use of thicker MLCCs, with greater spatial capacitance density, since the height of the mounted microprocessor die no longer is a factor in MLCC selection. Finally, removing the MLCC from the top of the package is effective in reducing the overall ESL of the decoupling capacitor path as the current loop of top-side decoupling schemes is typically considerably larger in area than the current loops associated with bottom-side MLCC mounting as illustrated in Figure 12.[5]

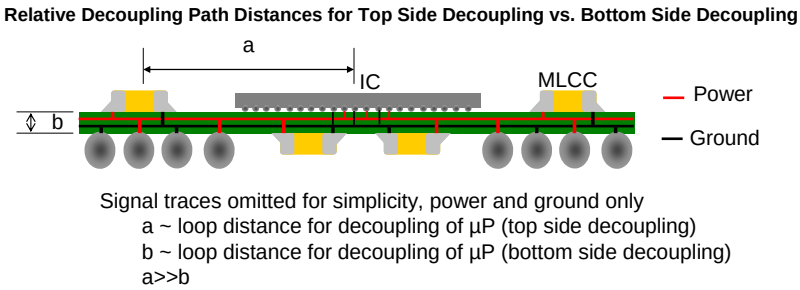

Figure 12. Limitations of top-side decoupling.

Moving the decoupling MLCC to the bottom side of the microprocessor package also has drawbacks. Bottom-side mounting of decoupling MLCCs takes up precious package space normally allocated to signal and power I/Os, resulting in a larger package design footprint requirement for the microprocessor package compared to a fully populated bottom I/O scheme (for example, see red square in Figure 11).

Significant design and development work has been focused on trying to combine the advantages of top-side and bottom-side decoupling without the associated disadvantages of either. One way to achieve that is to add functionality to the bottom-side on-package decoupling capacitors such that they may also act as power and ground

conductors, thus eliminating the need for additional space on the bottom side of the package for power and ground interconnects. This may also reduce cost as fewer I/O pins are needed. In this situation, the decoupling capacitor becomes a power feed-through capacitor as illustrated in Figure 13. These discrete devices may then be combined into an array at a predetermined pitch and sandwiched between the microprocessor package and the supporting circuit board so as to achieve the benefits of both bottom-side decoupling and top-side decoupling as illustrated in Figure 14.[6] These devices are called D-Packs. D-Pack interconnects attach to the power and ground I/Os and may be custom configured to different package I/O layouts where power and ground are concentrated. Full interposer D-Packs (FID-Packs) add further to assembly efficiency by adding additional signal I/O for the microprocessor package, thus eliminating all other interconnects between circuit board and microprocessor package as well as reducing the number of pick-and-place steps for decoupling capacitor attach to one step.

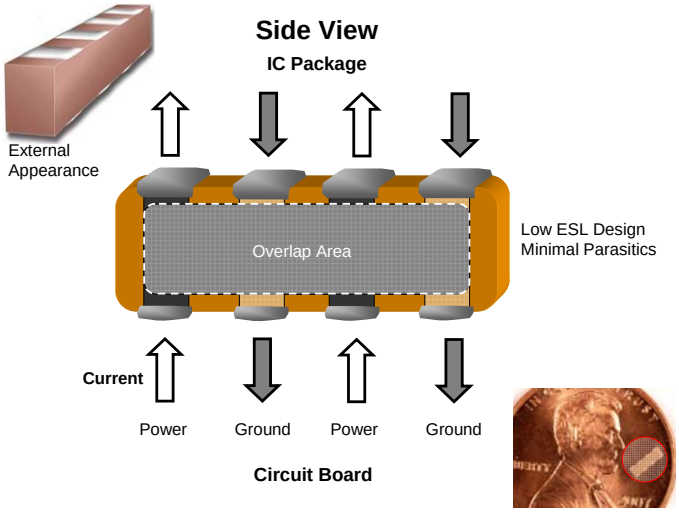

Figure 13. Low-ESL power feed-through capacitor.

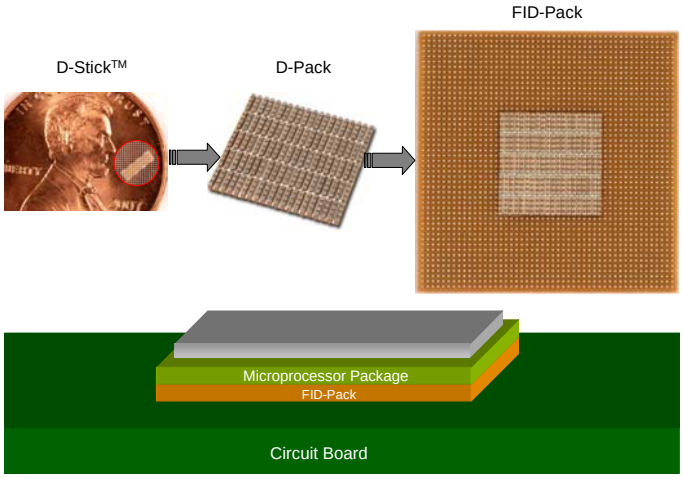

Figure 14. 3-D full interposer (FI) high-speed decoupling array (FID-Pack).

Design with D-Pack is relatively simple. The I/O pitch of the microprocessor package is set as well as the maximum thickness requirements. Discrete D-Sticks™ are then designed to meet the pitch and space requirements. The number of D-Sticks™ within the associated D-Pack design is determined by the customer's power, capacitance, ESL and ESR needs.

The FID-Pack approach to high-speed decoupling is excellent for high-performance applications where the package designer can work closely with the FID-Pack supplier. An example of this is a 100 D-Stick FID-Pack for a high-performance microprocessor application. In this case, D-Stick™ pitch and dimensions are set to accommodate

a 1 mm package pitch (X and Y) and the Z-dimension is set at <=1.2 mm. In this example, the microprocessor is a 200W design working at 1Vdc. The high-speed microprocessor requirements are very low ESL (1.5 pH or less) with high capacitance density under use conditions (>300 µF required at 105C and 1Vdc). The D-Stick™ is designed for >5 µF per unit (6.2 µF in this case) and 100 D-Sticks are used in the D-Pack design resulting in 200 power and 200 ground I/Os. A frame to accommodate the signal I/O pins is added around the D-Pack and the FID-Pack is complete. The FID-Pack is then mounted on the I/O pads of the microprocessor package or is mounted in a socket on the support circuit board. The above requirements are achieved with less than 1% power dissipation in the FID-Pack. Of course, design for other D-Stick™ counts is possible as well.

Conclusions

Recent advances in capacitor design and construction give new options in design of microprocessor decoupling circuits. Today, a combination of several aluminum and tantalum capacitors along with high-capacitance MLCCs are used on the motherboard close to the processor. Advances in low-ESL aluminum and tantalum capacitors allow this complex network to be replaced with possibly just one or two of these new-design capacitors. This can reduce circuit space while actually increasing performance.

On-package, D-Pack, a new interposer decoupling solution, is a very effective decoupling solution, having very high capacitance (ca. 1,000 µF/in2) in very close proximity to the microprocessor package and with extremely low inductance (<1 pH depending upon design). This is achieved via bottom-side mounting, but without the typical bottom-side requirement of increasing microprocessor footprint.

References

[1] L.E. Mosley, "Capacitor Impedance Needs for Future Microprocessors," Passive Component Industry, May/June 2006, p. 12, Figure 1.

[2] The authors would like to acknowledge Dr. Istavan Novak of Sun Microsystems for suggesting this technique.

[3] R. Hahn and J. Paulsen, "Low Impedance Ta Capacitors To Serve the Needs of the Capacitor Industry," CARTS USA 2007 Proceedings, pp. 349-360.

[4] E. Reed, "Low ESL Tantalum Capacitors," CARTS USA 2008 Proceedings.

[5] M. Randall, et al., "Decoupling Solutions," CARTS USA 2007 Proceedings, pp. 217-32.

[6] US Patent 7,068,490 and 7,291,935 others pending.