PCB Effects on On-chip Capacitor Requirements and an Efficient Resonance-Prevention ASIC Methodology

Timothy Budell and Eric Tremble
IBM Systems and Technology Group
1000 River Street, Essex Junction, VT 05452
Phone: (802) 769-7335, Email: budell@us.ibm.com

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2010/1/000392/2252756/isom-2010-wa2-paper1.pdf by guest on 18 June 2024

Abstract

A method for determining adequate quantities and locations of on-chip capacitors to maintain supply voltages at all locations on a chip within pre-specified limits given the switching activity of on-chip circuits was presented in [3]. In this paper, we extend the method to include current flow from the package and PCB. The effects of on-chip capacitance and other system parasitics on the time it takes for additional supply current to flow into a chip are discussed. The relationship between switching current, capacitance, system parasitic inductances, and on-chip noise is presented. These concepts are then applied to the subject of power delivery network (PDN) resonance. A 1-dimensional model for simulating PDN resonance is presented. The model includes chip, package, and PCB components, along with explicit networks for each chip power supply and their interactions. The topology of the model and the contributions of each model component are described. A design methodology for avoiding PDN resonance, presently in use on all IBM ASIC modules, is presented.

Key Words: Power supply noise, decoupling capacitors, resonance prevention.

1. Introduction

Modern ASIC chips can draw many amps of transient current from on-chip power supply rails in just a few hundred picoseconds. To the extent that this current can not be supplied to the switching circuitry from local charge reservoirs, the current must be drawn through the impedance of the power supply and a transient on-chip voltage compression will occur [1]. Such power supply noise, when excessive, raises serious chip functionality concerns.

All electronic systems, including ASIC modules and the printed circuit boards (PCBs) to which they are attached, exhibit resonances at which the power supply impedance climbs substantially [2]. Switching events having significant frequency components near such resonant frequencies can dramatically increase overall power supply noise. Yet switching events of many of today's applications unavoidably occur near the natural resonant frequencies of chip-package-PCB power supply networks, which can range widely, but are typically in the vicinity of 50 MHz for core-logic supplies and 150 MHz to 300 MHz for I/O supplies.

Adequately suppressing power supply noise in today's microelectronic systems requires, at minimum, the availability of low-impedance on-chip decoupling capacitors as local charge reservoirs. A method for determining adequate quantities and locations of on-chip capacitors to maintain supply voltages at all locations on a chip within pre-specified limits given the switching activity of on-chip circuits was presented in [3]. The method accounts for charge moved by on-chip circuits, but assumed no additional charge is delivered to the chip during the period of switching activity. In this paper, we extend the method to include current flow from the package and PCB. The concepts discussed are then applied to the subject of power delivery network (PDN) resonance and how it can be avoided.

2. 1-Dimensional PDN Model

A 1-dimensional representation of a system level PDN is shown in Fig. 1. The model includes chip, package, and PCB components, along with explicit networks for each chip power supply and their interactions. Total on-chip capacitance for each supply is included on the left in Fig. 1, as well as on-chip resistances. An ideal current source represents on-chip switching activity of one supply for transient simulations; an AC source is used for frequency domain simulations.

A disadvantage of ideal current sources is that they continue to draw the same current even as the voltage across them decreases. We have found the error is positive (meaning it bounds the actual noise) and small (10% or less) provided charge reservoirs on chip are sufficient to limit supply voltage compression to less than 15%.

Impedance of a system at a given port can be readily simulated by injecting 1 Amp into the port. The resulting voltage across the port is thereby equivalent

to the port impedance. This technique is used to greatly simplify complicated package and PCB power network geometries to a few electrically equivalent components. Detailed PDN behavior can be extracted using commercially available tools such as Ansoft's SIwave or Sigrity's PowerSI. The extraction can capture 1-dimensional PDN behavior in a 2-port S-parameter block by commoning all like supply and return pins at each system interface into a single port. When 1 Amp AC is injected into the chip-package port of such an S-parameter block extracted from a flip-chip package, with return current through the package-PCB port, the resulting voltage response (and impedance) is found to vary linearly and with positive slope on a log-log plot versus frequency from about 1 MHz to nearly 1 GHz. Hence, the package is found to be almost purely inductive over this wide frequency range. This range completely includes the resonant responses that are the subject of Section 7. At lower frequencies, the impedance is found to be nearly constant and approximately equal to the DC resistance. Similar electrical behavior is found for PCB power delivery networks. Hence, a package and PCB can be accurately modeled with simple resistive and inductive elements in series for the purpose of investigating system resonances. These package and PCB elements are shown in the middle and right of Fig. 1.

The vertical R, L, and C elements in Fig. 1 optionally represent package capacitors. The dual set of package R and L elements to the right and left of the package capacitor elements are adjusted accordingly to model where in the package layer stack the package capacitor terminals are connected.

The ideal voltage sources on the right in Fig. 1 represent the PCB supplies. In this example schematic, only three (RLC+voltage source) loops are shown. Note that a complete model would include a unique loop for each power supply on the chip.

3. Time of charge delivery

The topology of the circuit in Fig. 1 makes it clear that current cannot flow from the external ideal supply into the chip instantaneously in response to chip activity that moves charge from a supply rail to the common return rail. First, a voltage must be established across the package and PCB inductors, then current will start to flow. As described in detail in [3], such a voltage is established when switching circuits on the chip move charge between supply and return rails. Briefly, $Q=CV$, where Q is the charge stored on the chip when it is idle, C is the on-chip capacitance, and V is the voltage across the chip supply rails. If Q is reduced by a switching event that dumps charge to a return rail, then the voltage across the supply rails will be reduced linearly until that charge is resupplied from outside the chip. The above also implies that if C is larger, the switching event will create less on-chip voltage compression during the interval prior to resupply from the outside.

The circuit in Fig. 1 is a series LC resonator having a lossless natural ring frequency of

$$F_R = {}^{1}\!\!\diagup_{2\pi\sqrt{LC}} \qquad (1)$$

where L is the total system inductance including package and PCB. After an event that alters the initial voltage across the capacitor (such as removal of some on-chip charge), this voltage will vary sinusoidally until resistive losses dampen the response. An example of this sinusoidal response is shown in Fig. 3 (blue curve), resulting from the current step and plateau shown in Fig. 2 (blue curve). Consistent with the curves shown in Fig. 3, the resonant frequency of IBM ASIC flip-chip core logic supplies is typically in the 30 – 60 MHz range.

The resonant behavior described above is fundamental to electronic power delivery systems. The best one can do to suppress such responses is add huge amounts of on-chip capacitance to grossly over damp

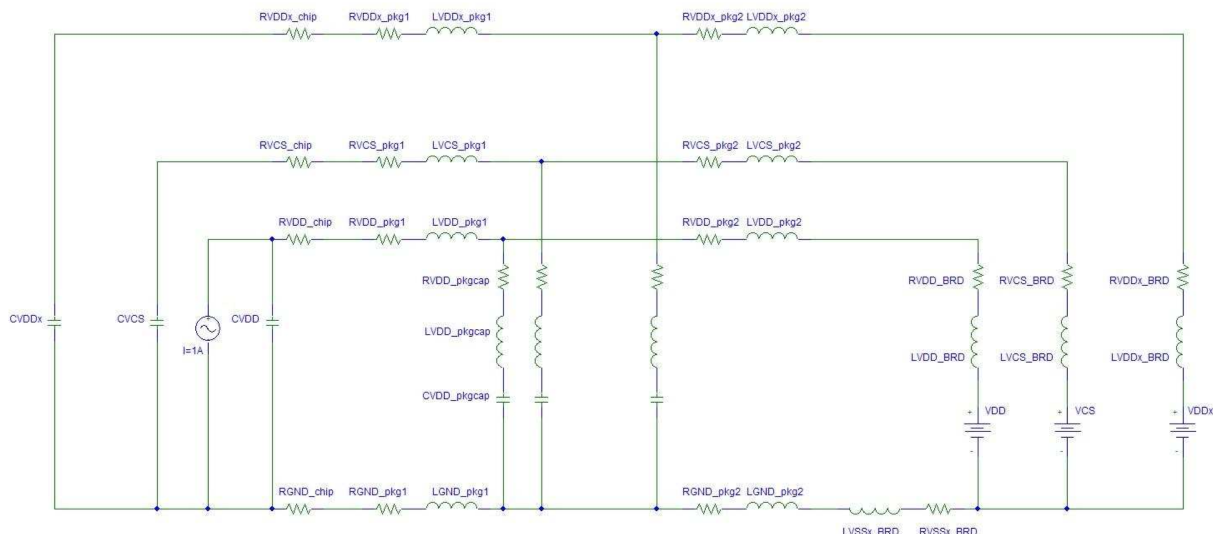

Fig. 1. 1-dimensional PDN model including chip, package, and PCB (3 of n supply loops shown).

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2010/1/000392/2252756/isom-2010-wa2-paper1.pdf by guest on 18 June 2024

them. But a more practical approach is to decide before hand what minimum supply voltage on-chip circuits can tolerate without incurring functionality issues, and then calculate the minimum amount of on-chip capacitance needed to keep on-chip voltages at or above that minimum.

Changes in on-chip capacitance, C, have two noteworthy effects that partially oppose each other: 1) an increase in C will reduce on-chip supply voltage compression resulting from a given switching event by immediately supplying charge (that does not have to first pass through the PDN impedance); 2) an increase in C will lengthen the time until charge flows across the system inductances to replenish charge removed by a switching event.

4. Accounting for Charge Resupply

In [3], on-chip power-supply voltage compression owing to on-chip switching events was formulated in terms of quiet and switched capacitances. The voltage compression can be equivalently formulated in terms of quiet capacitance and a current step. In either formulation, the change in supply voltage across on-chip circuits is a direct function of the change in charge stored on that chip supply. The initial charge is

$$Q_i = CV \qquad (2)$$

where C is the on-chip quiet capacitance and V is the supply voltage. The new charge on chip after the current step is

$$Q' = Q_i - Q_s + Q_{pkg} \qquad (3)$$

where Q_s is the charge moved by the current step and Q_{pkg} is the charge resupplied through the package and PCB. The new voltage resulting from the current step is found by rearranging the form of Eq. (2), namely

$$V' = Q'/C \qquad (4)$$

Combining Eq. (2) – (4), the compression in on-chip supply voltage, (V - V') as a result of the current step is

$$V_{compress} = (Q_s - Q_{pkg})/C \qquad (5)$$

For Eq. (5) to be useful, we need to define an interval of interest so that we can convert our assumption of a current step to a quantity of charge. Referring to the smooth blue dotted curve of Fig. 3, it can be seen that the minimum voltage and maximum compression occurs at one quarter of the natural ring period of the system, ($1/F_R$). In Fig. 4, current through the package for a lossy system is seen to follow a damped sinusoid response that reaches the same magnitude as the original current step at one-quarter of the natural ring period. These two observations suggest that the interval of interest is one quarter of the natural ring period. Let I be the current that turns on at time t=0 and T be the natural ring period. Then combining with Eq. (1), the switched charge is

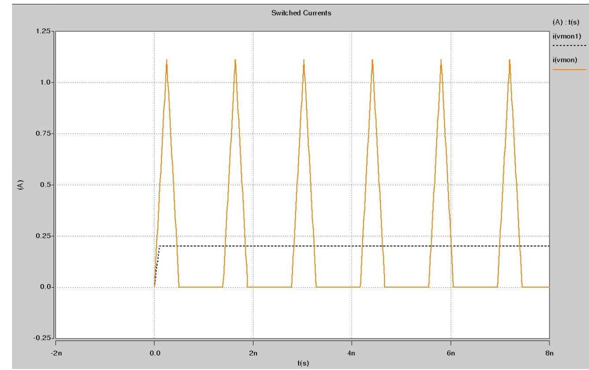

Fig. 2. Current of single power step event (blue dotted curve) and 720 MHz switching event (brown curve) of same average power.

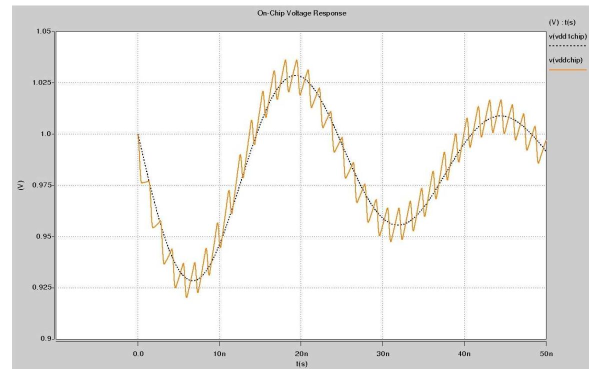

Fig. 3. On-chip voltage response owing to power steps of Fig. 2 - single power step (blue dotted curve) and 720 MHz switching (brown curve) with F_R=40 MHz.

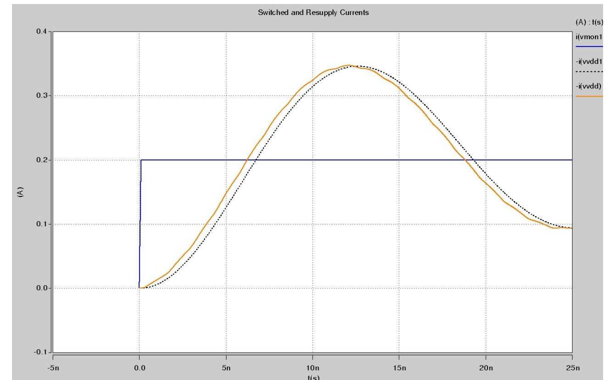

Fig. 4. Package current resulting from power steps of Fig 2 - single current step (blue dotted curve) and 720 MHz switching (brown curve).

$$Q_s = I(T/4) = I/4F_R = I\pi\sqrt{LC}/2 \qquad (6)$$

Referring again to Fig. 4, if we assume a lossless system rather than the damped sinusoid shown, current resupply through the package has a (1-cos(x)) form, and repeats every 2π radians with a period of T. Charge resupplied by the package in the T/4 interval can be found by integrating this current as shown below.

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2010/1/000392/2252756/isom-2010-wa2-paper1.pdf by guest on 18 June 2024

$$I_{pkg} = I(1-\cos(x)) = I\left(1-\cos\left(\frac{2\pi t}{T}\right)\right)$$

$$Q_{pkg} = \int_0^{T/4} I_{pkg}\,dt = I\int_0^{T/4} 1-\cos\left(\frac{2\pi t}{T}\right)dt$$

$$Q_{pkg} = I\left[t - \frac{T}{2\pi}\sin\left(\frac{2\pi t}{T}\right)\right]_{t=0}^{t=T/4}$$

$$Q_{pkg} = I\left(T/4\right)\left(1-\frac{2}{\pi}\right) = Q_s\left(1-\frac{2}{\pi}\right) \tag{7}$$

Substituting Eq. (7) into Eq. (5),

$$V_{compress} = 2Q_s/\pi C \tag{8}$$

Substituting in Eq. (6) into Eq. (8), it is found that

$$V_{compress} = I\sqrt{\frac{L}{C}} \tag{9}$$

The square root term is a standard frequency-independent expression of characteristic impedance, a concept we will return to in Section 8. Rearranging terms, the amount of on-chip capacitance required to stabilize on-chip supply voltage compression to less than a pre-specified value during a current step event is

$$C = I^2L/V_{compress}^2 \tag{10}$$

It should be noted that the above derivation assumes an infinitely fast current rise time followed by a constant current plateau, and as previously stated, minimum voltage at time $t=T/4$. The assumption of a constant current plateau is a good one when the current step is caused by thousands or millions of logic circuits becoming active, as might happen when an ASIC chip goes from sleep mode to functional mode. When current events are discrete and repetitive, an assumption of multiple switching events is more appropriate. This is discussed in Section 6.

A colleague recently demonstrated that closed form solutions exist for the minimum voltage and the time when it occurs assuming a finite slope current ramp of any duration followed by an infinite current plateau. Under these assumptions, the equations (verified by simulations) indicate that the voltage minimum occurs at one-quarter of the resonant period (i.e., $T/4$) plus one-half the switching transition time, for transition times faster than half the resonant period. The equations also indicate that the assumption of infinitely fast current transition in Eq. (9) introduces an error in predicted $V_{compress}$ of just 1% when the actual transition time is increased to as much as 30% of the $T/4$ interval.

The introduction of loss into the derivation of Eq. (9) increases predicted $V_{compress}$. The resulting RC time-constant slows charge resupply through the package and increases the $T/4$ interval, meaning that switching circuits draw charge off the supply rail for a longer interval without significant charge resupply. We have found empirically that losses in typical flip-chip applications (primarily on-chip resistance) increase $V_{compress}$ by 5% or less versus that predicted by Eq. (9).

In Section 6, it is noted that several practical aspects of chip floor planning and decoupling capacitor ("decap") insertion introduce more uncertainty than the errors discussed above.

5. Relationship between *I*, *C*, *L*, and Noise

Eq. (9) elucidates several fundamental aspects of the relationship between charge demand owing to on-chip switching events (I), on-chip capacitance (C), system parasitic inductances (L), and on-chip noise ($V_{compress}$). Holding two of these four variables constant, the relationships between the remaining two variables are as follows:

1. $V_{compress}$ is inversely proportional to the square root of C. For a given system inductance and current step, halving the desired supply voltage compression requires a quadrupling of on-chip capacitance.
2. ΔI is proportional to the square root of C. For a given system inductance and supply voltage compression, doubling the current step requires a quadrupling of on-chip capacitance. Note that for a given switching event, doubling the frequency of that event will double the charge moved per unit time, meaning the transient current is doubled. Hence, switching frequency F_{SW} is also proportional to square root C.
3. $V_{compress}$ is proportional to ΔI. For a given system inductance and quiet capacitance, increasing the current step linearly increases the amount of power-supply noise.
4. $V_{compress}$ is proportional to the square root of L. For a given current step and quiet system capacitance, halving the voltage compression requires a reduction in system inductance by a factor of four.

The first two relationships demonstrate that reducing transient currents and relaxing voltage compression targets can exponentially reduce the need for on-chip quiet capacitance and the area it occupies. The fourth relationship shows that reductions in system inductance result in relatively little supply noise benefit.

It is worth considering how much each of the variables above can be altered during the design of an ASIC module (chip+package). Total inductances of flip-chip packages vary widely depending on chip size, footprint configuration, and number of C4 connections (solder bumps) between chip and package for each supply, but partial L values tend to be fairly constant and in the 3 – 4.5 nH/C4 range. It is plausible that with careful design attention, inductance in localized regions of a package can be reduced by perhaps a factor of two, at a cost of increased design turn-around time.

Based on limited customer data, it appears that PCB loop inductances (supply+return) tend to be in the 3 nH range per package supply ball (BGA) on well

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2010/1/000392/2252756/isom-2010-wa2-paper1.pdf by guest on 18 June 2024

designed PCBs. When the number of C4s versus the number of BGAs is factored in, we have found that PCB loop inductance per supply is typically some 3.5 – 5 times higher than the corresponding package loop inductance and therefore dominates total system inductance. As an ASIC supplier, IBM has little control over the inductance of customer PCBs. Hence, when attempting to control PDN noise during the module design phase, total system inductance is essentially a fixed variable.

Random logic macros (RLMs) and hard macros (often called "intellectual property" or IP) are typically pre-designed, often by 3rd parties. Customers want such macros to be as fast as possible, which typically means burning higher power. Meanwhile, the macro designers want supply voltage variation as low as possible to increase the design margins available to them. As full-chip and package designers, we can attempt to influence both of these variables, but in practice, I and $V_{compress}$ are also essentially fixed constraints.

On chip "background capacitance" for areas containing mostly dust logic in IBM's Cu-45 technology is about 1 nF/mm^2. Deep trench (DT) capacitors, which have been specifically designed as highly localized charge reservoirs in support of high activity circuits, have capacitance densities in the 100 – 200 nF/mm^2 range, depending on type. In contrast to the other three variables discussed above, we have found that intentionally placing DT capacitors near circuits that draw large transient currents is a highly efficient and viable design option with regard to supply noise, even though reducing such noise requires an exponential increase in capacitance. Our experience is that, owing to the very high charge-storage density of DT capacitors, such intentionally placed quiet capacitance can generally be fit on ASIC chips in existing "white space" between logic circuits, memories, and other functional blocks with minimal design effort and without the need for increasing the size of the die. Moreover, in contrast to oxide capacitors (which have much lower capacitance densities), the leakage of DT capacitors is essentially zero (10s of μAmps on chips that draw 10s of Amps of functional current).

6. Multiple Switching Events

In this section we consider the effects of discrete switching events rather than a current step to a constant current plateau. Let Q_{SE} be the charge moved by a single switching event (for example, a read or write access to an SRAM). When multiple such switching events occur in the T/4 interval, the on-chip supply voltage will collapse in a step wise fashion as charge is repeatedly removed from the on-chip capacitance prior to significant resupply from outside the chip. The brown curve in Fig. 3 shows an example of this behavior. As indicated by Eq. (8), supply voltage compression is a function of total charge moved by switching events in the T/4 interval.

Let F_{SW} be the frequency of the discrete switching events. Then Eq. (6) can be re-written as

$$Q_s = Q_{SE} F_{SW} \pi \sqrt{LC} \big/ 2 \qquad (11)$$

We have not completed a rigorous derivation, but have found empirically that the maximum supply compression of a multiple switching event scenario is predicted well and bounded by Eq. (8) when an additional term equal to one-half the charge moved by each switching event is included. That is,

$$V_{compress} = \left(1/C\right)\left(\left(2Q_S/\pi\right) + \left(Q_{SE}/2\right)\right) \qquad (12)$$

Substituting in Eq. (11), it is found that

$$V_{compress} = \frac{Q_{SE}}{C}\left(\left(F_{SW}\sqrt{LC}\right) + 1/2\right) \qquad (13)$$

While this equation is not readily solved for C, we have implemented it algorithmically to determine how many decaps (decoupling capacitors) are needed within RLMs or around hard-macro IP such as SRAMs, TCAMs, and register arrays to control supply compression within pre-specified limits. We should note that our decap guidelines and automatic decap algorithms use switched capacitance, C_{SE}, as the unit of measure. This is switched charge, Q_{SE}, divided by the supply voltage. This is a more convenient metric because we have found that C_{SE} is nearly constant across process, voltage, and temperature (PVT) for most IP.

After assessing decap needs around IP blocks and within RLMs and placing the decaps in the chip floor plan, IBM's ALSIM_TA full-chip simulator is used to verify that supply noise does not exceed the pre-specified $V_{compress}$ limit anywhere on the chip. The simulations include resistances of all on-chip power buses as well as internal resistance of the DT capacitors, which we have found can have a particularly strong effect on local supply compression. In addition to the effects of loss, such full-chip noise simulations have highlighted two practical aspects of decap estimation that tend to overwhelm the sources of error mentioned in Section 4 regarding Eq. (9) and (10), and by extension, Eq. (13). First, the resonant frequency of individual RLMs may differ significantly from the full-chip F_R. Second, proximity of RLMs and macros versus their activity and internal capacitance can significantly affect supply noise near their boundaries.

As mentioned in Section 5, effective L varies with RLM size. Likewise, depending on switching activity, Eq. (13) may call for capacitance density within an RLM that is markedly different from the chip-wide average. If the resulting F_R of the RLM as predicted by Eq. (1) is much higher than that of the full chip (which will dominate), supply noise within the RLM may be higher than predicted by Eq. (13) because charge resupply from the package will take longer than assumed by this equation.

Circuits near the edge of a die will tend to have higher noise for the same switching activity be-

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2010/1/000392/2252756/isom-2010-wa2-paper1.pdf by guest on 18 June 2024

cause at the die edge, half the area that might otherwise act as a charge reservoir is absent. Similarly, an RLM adjacent to another RLM having very high (or low) capacitance density may experience relatively lower (or higher) supply noise owing to the presence of this adjacent charge reservoir.

These uncertainties make Eq. (13) a good engineering starting point for decap insertion, but full-chip supply noise simulation is required to verify supply stability within the pre-specified $V_{compress}$ limit at all locations on the chip.

7. Resonance in the System

As discussed in Section 2, the impedance (Z) vs. frequency of the entire PDN for each supply in Fig. 1 can be determined by replacing the switching current source with a 1 Amp AC source and sweeping the frequency in simulation. Trans-impedances can be determined in the same simulation by observing the resultant voltage across the capacitor of each of the other supplies. The current source can then be moved in successive simulations to each of the other supplies to find all self and trans-impedances for all supplies.

Recall that impedance versus frequency for an inductor is $Z = j\omega L = j(2\pi f)L$, and for a capacitor is $Z = 1/(2\pi f)C$. Since the on-chip capacitance loop in Fig. 1 is in parallel with the system inductances loop and both connect to the AC current source, impedance is low at low frequencies (where the inductor has low impedance) and low at high frequencies (where the capacitor has low impedance). At intermediate frequencies, both elements have appreciable impedance and the total impedance of the system peaks. This behavior is shown in Fig. 5. Such peaks in impedance will occur at the system resonant frequency, F_R, described by Eq. (1).

Eq. (9) and Eq. (13) were derived in a frequency independent manner, but it can be seen in Fig. 5 that the system PDN impedance in the range of 50 MHz is more than an order of magnitude higher than at other frequencies. If significant transient current is drawn by the chip in this mid-frequency range, substantial voltage drop through the PDN will occur, resulting in much larger supply voltage compression at the on-chip circuits than predicted by Eq. (13). Such unanticipated supply noise can cause increased jitter, reduced timing margins on critical paths, and intermittent failure of data busses and other functional circuitry. An example of such excessive noise is shown in Fig. 6 (red dashed curve).

It is not necessary for core logic or an on-chip bus to run at the system resonant frequency for resonance to be excited. Data patterns can cause switching to occur at many frequencies other than the fundamental clock frequency. For example, a DDR bus running at 250 MHz (500 Mb/s) will transmit a repeating 0101 data pattern at a frequency of 250 MHz. But a repeating 011 data pattern switches at 166.7 MHz; a repeating 0011 data pattern switches at 125 MHz; a repeating pattern of 00001111 switches at 62.5 MHz; and so on. Since in Fig. 6 it can be seen that it takes 3 – 4 repetitions of a switching event to fully excite resonant noise, it may seem that occurrences of long repetitive data patterns are not likely and therefore the risk of pumped resonance must be very low. But our knowledgeable customers assure us that such exact data patterns do occur in real data traffic. Indeed, a jitter related, data dependant, inter-chip bus failure was observed during system test by one of our customers two years ago. Root cause was determined to be I/O supply PDN resonant interaction of chip capacitance and package and PCB inductive parasitics. Since ways in which resonance can be excited on core logic or I/O supplies can be envisioned or have been directly observed, it is prudent to ensure that a system is immune to such activity by design.

8. Resonance-Prevention ASIC methodology

The greatest uncertainty for an ASIC resonance-prevention methodology is the effective inductance of the PCB. Our experience is that during the design phase of an ASIC module, the PCB has rarely been designed yet. Moreover, some of our customers do not know how to assess PCB inductance as seen by the ASIC module. Such inductance is dominated by the PCB through-via length, the number of PCB capacitors closest to the ASIC module, the effective series inductance (ESL) of those capacitors, and their distance from the module. We have found that effective PCB inductance tends to be in the vicinity of 3 nH/BGA, but does vary considerably. When exact data is not available from a customer, we assume 3 nH/BGA plus and minus a factor of two. This results in three impedance curves per supply as shown in Fig. 5. We bound impedance for subsequent calculations by taking the envelope of these curves, as shown in Fig. 7.

We assess resonance and the possible need for PDN design changes by comparing supply noise calculations done in three ways – by pure charge sharing as described in [3]; by multi-event charge sharing as described by Eq. (13); and by $V=IZ$ using impedance at the switching frequency, F_{SW}. The largest of these calculated noises is then compared with the $V_{compress}$ limit, which is pre-specified jointly with the customer and is typically 8% of the supply voltage. If the limit is exceeded, then PDN design changes are indicated. Which method gave the largest noise informs what design changes should be made.

Pure charge sharing assumes charge is moved at time t=0 with no resupply of charge from the package and PCB. It applies when a switching event occurs after a long quiet period and does not repeat until long after the T/4 interval, when the PDN voltage response has settled. When a pure charge sharing calculation gives the largest noise and that noise exceeds the $V_{compress}$ limit, it indicates there is not enough on-chip quiet

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2010/1/000392/2252756/isom-2010-wa2-paper1.pdf by guest on 18 June 2024

capacitance to adequately support the switching event. The appropriate PDN design change is to add on-chip capacitance until the $V_{compress}$ limit is satisfied.

Multi-event charge sharing applies when discrete switching events occur repetitively and continuously within the $T/4$ interval after a long quiet period. It assumes an average current based on the charge moved by a switching event and the time until the next event. It also assumes charge resupply from the package and PCB as per the derivation of Eq. (13). When a multi-event charge sharing calculation gives the largest noise and that noise exceeds the $V_{compress}$ limit, the appropriate PDN design change is to add on-chip quiet capacitance until Eq. (13) satisfies the limit. Another option is to add package capacitors to shorten the $T/4$ charge-resupply interval (a concept discussed in greater detail below).

Using the impedance envelope discussed above, the $V=IZ$ method first determines the impedance of the PDN at the switching event frequency, F_{SW}. An example is shown in Fig. 7 (vertical dashed line where Z = 0.18 ohms). This impedance is multiplied by the average current as determined by the multi-event charge sharing method above, and the result is compared with the $V_{compress}$ limit. Note that this is a bounding calculation. Strictly speaking, the voltage response at a given frequency is a function of current and impedance at that frequency – that is, $V(f) = I(f)Z(f)$ – and the total voltage response is the integral of $V(f)$ over all frequencies. But $IZ(F_{SW})$ overstates the actual noise by assuming all of the current switches at frequency F_{SW} (as if it were a full-current pure sine wave). If this overstated noise is still less than the $V_{compress}$ limit, then there can not be an issue with pumped resonance on the PDN. Even if switching occurs right at F_R, the resonant frequency of the system, supply noise will remain within acceptable limits and will cause no functionality issues.

If $IZ(F_{SW})$ is larger than $V_{compress}$, then there is risk of pumped PDN resonance. A straight forward design change to eliminate the possibility of resonance is to add package capacitors. The addition of these very large charge reservoirs near the chip makes the PDN as seen by the on-chip circuitry look more ideal (the capacitors effectively bypass the PCB inductance and part of that of the package, thereby lowering total system L by some 80% or more). The system resonant frequency is thereby increased significantly (typically by 80 MHz or more), moving it above the frequency of switching. Peak impedance is also substantially decreased.

However, package capacitors add cost and can create wiring blockages in the package, so we consider them a last resort. First, we ask the customer if they can provide a better estimate of PCB inductance, which could narrow the bounding calculations. We also typically run transient simulations to see if the $V=IZ$ calculation is overly pessimistic. We assess whether adding

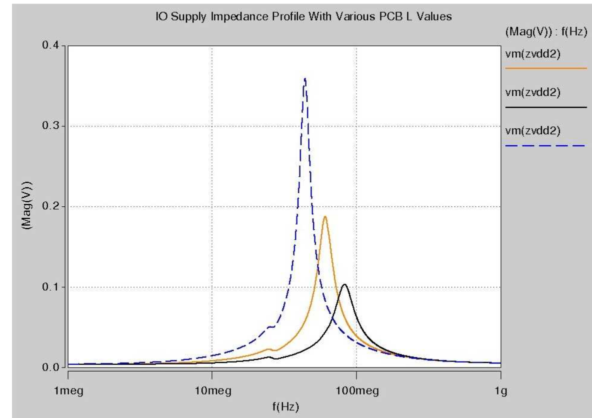

Fig. 5. System PDN impedance of an I/O supply vs. frequency as seen from the chip with PCB L = 6 nH/BGA (blue dashed curve); 3 nH/BGA (brown curve); and 1.5 nH/BGA (black curve).

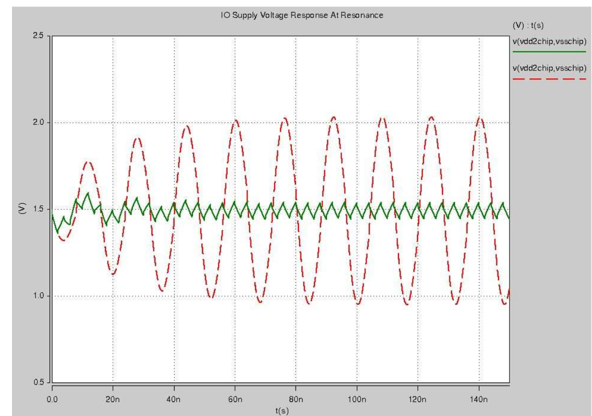

Fig. 6. I/O supply noise resulting from I/O switching at 62.5 MHz (red dashed curve) and 250 MHz (green curve) given the brown-curve impedance profile of Fig. 5.

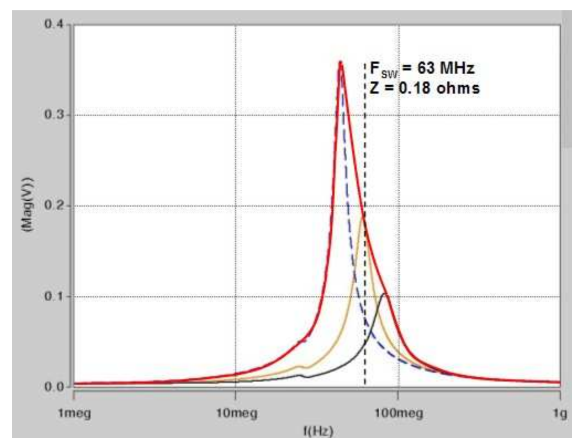

Fig. 7. Envelope of impedance of Fig. 5 (red curve) and an example switching frequency F_{SW} (black vertical dashed line).

additional on-chip capacitance is an adequate solution. When resonance is a potential issue, however, this latter approach often does not help much. Any reasonable

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2010/1/000392/2252756/isom-2010-wa2-paper1.pdf by guest on 18 June 2024

amount of on-chip capacitance often does not move the resonant frequency far enough (F_R is lowered only as the square root of on-chip C), and may simply move it down onto another data dependant switching harmonic. Our experience is that when a fundamental or data dependent F_{SW} sits near the frequency of peak PDN impedance and creates noise exceeding the $V_{compress}$ limit, the addition of package capacitors is typically the only viable design option.

Lastly, we check for trans-impedance problems. Noise on a victim supply is the product of the current step of an aggressor supply and the trans-impedance from the aggressor to the victim at the switching frequency of the aggressor supply. Treating each PDN as a victim in succession, we sum the noise contributions from all other PDNs and verify that total noise owing to trans-impedances is less than $V_{compress}$. If not, the most straight forward design change is to add on-chip capacitance to the supply having less capacitance to lower the trans-impedance between aggressor and victim supplies. We have found that trans-impedance is rarely an issue, but occasionally a large current-step supply is found to be a significant trans-impedance aggressor, particularly when the victim supply has relatively low capacitance.

Until now, for simplicity, we have discussed self and trans-impedance calculations separately. Note that in actual practice, we do these calculations concurrently and sum their effects to ensure that the total supply noise does not exceed the $V_{compress}$ limit.

A practical complication of all of the above calculations is that some supplies support circuitry at several different switching frequencies. The calculations can be done at each switching frequency, using only the portion of power associated with the circuits of a particular F_{SW}, and the noise results summed. More commonly, switching at one F_{SW} strongly dominates the total power of a supply and the calculations can be done at that F_{SW} alone without missing a significant portion of calculated supply noise.

A prototype of this resonance-prevention methodology has been in use on all IBM production ASICs since 4Q2009. Our experience is that about 10% of all supplies require some design modification to avoid the possibility of resonance. Typically, but not always, these are I/O supplies having wide synchronous busses.

9. Conclusions

A 1-dimensional PDN model was introduced with the goals of determining on-module capacitance requirements and assessing system resonances. The model includes chip, package, and PCB components, along with explicit networks for each chip power supply and their interactions. The magnitude of on-chip supply voltage compression was derived as a function of transient current draw, on-chip capacitance, and system parasitic inductances. The relationship between these four variables was described. The mechanisms of system resonance were described. Efficient ASIC methodologies for stabilizing on-chip supply compression within a pre-defined limit and avoiding supply resonance were introduced. These methodologies are a part of IBM's production ASIC design flows.

Acknowledgments

The authors would like to thank Charlie Chiu for his derivation, mentioned toward the end of Section 4, of $V_{compress}$ under the assumption of finite current step transition time; and Umberto Garofano, Ivan Wemple, Ed Sayre, and Kurt Carlsen, for their thoughtful contributions to the work discussed in this paper.

References

1. L. D. Smith and J. Lee, "Power Distribution System for JEDEC DDR2 Memory DIMM," *Proc. IEEE 12th Topical Meeting on Electrical Performance of Electronic Packaging*, Princeton, NJ, October 2003, pp. 121-124.
2. L. D. Smith, R. Anderson, and T. Roy, "Power Plane SPICE Models and Simulated Performance for Materials and Geometries," *IEEE Transactions on Advanced Packaging*, Vol. 24, No. 3, August 2001, pp. 277-287.
3. Nanju Na, Timothy Budell, Charles Chiu, Eric Tremble, and Ivan Wemple, "The Effects of On-chip and Package Decoupling Capacitors and an Efficient ASIC Decoupling Methodology", *Proc 54th Electronic Components and Technology Conf*, May 2004, pp. 556 - 567.

Downloaded from http://meridian.allenpress.com/ism/article-pdf/2010/1/000392/2252756/isom-2010-wa2-paper1.pdf by guest on 18 June 2024