Automatic Package and Board Decoupling Capacitor Placement Using Genetic Algorithms and M-FDM

Krishna Bharath, Ege Engin and Madhavan Swaminathan
School of ECE
Georgia Institute of Technology
777 Atlantic Dr.
Atlanta, GA 30329
{kqhiharath, engin, madhavan}@ece.gatech.edu

ABSTRACT
In the design of complex power distribution networks (PDN) with multiple power islands, it is required that the PDN represents a low impedance as seen by the digital modules. This is to reduce the simultaneous switching noise (SSN), generated due to the switching activity of digital drivers. Typically, this reduction in impedance is accomplished by placing decoupling capacitors between the power and ground planes of a package or board. However, the performance of the decoupling solution is a function of capacitor selection and its placement. In this paper, an automatic capacitor placement optimization method has been proposed. This method relies on a genetic algorithm to provide a stochastic search of the design space, while employing an efficient core PDN simulator based on the multi-layer finite difference method (M-FDM). The technique has been employed to show optimized placements for split planes as well as for a realistic multi-layer server board.

Categories and Subject Descriptors
I.6 [Computing Methodologies]: Simulation and Modeling

General Terms
Algorithms, Design

Keywords
Signal/Power Integrity (SI/PI), Genetic Algorithm (GA), multi-layer finite difference method (M-FDM)

1. INTRODUCTION
In integrated digital systems, the design of the power distribution network (PDN) is a critical component to maintaining its functionality and performance. The switching activity of drivers on the digital module creates simultaneous switching noise (SSN) with energy concentrated at multiples of the clock frequency. This can lead to signal integrity problems and bit errors. This can also be the source of radiated emissions which can cause EMI/EMC concerns.

From a design perspective, the requirement on the PDN is that it should represent a low impedance at all frequencies at which current transients exist. This low impedance is denoted as the target impedance ($Z_{\text{tar}}$) and is calculated as

$$Z_{\text{tar}} = \frac{V_{\text{core}} \times 0.05}{I_{\text{core}} \times 0.5}$$

(1)

where $V_{\text{core}}$ is the core voltage of the active device and $I_{\text{core}}$ is the current drawn as described in [8]. This is based on the assumption that the tolerable noise voltage is 5% of the core voltage. Also, $I_{\text{core}}$ is the current over an entire clock period. Hence the denominator of (1) is the current in one clock edge. Traditionally, the solution to providing a low impedance power supply as seen by the device has been to place decoupling capacitors (decaps) across the power supply pins [7]. These capacitors act as a local store of charge and supply the transient current needs of the switching circuits. However, a realistic capacitor behaves as an inductor above its self resonance frequency (SRF), and hence is applicable only over a certain frequency bandwidth. Thus, to cover a broad frequency range, decoupling is provided using low frequency surface mounts discrete (SMD) capacitors on-board and high frequency capacitors on-chip, with an increasing tendency to cover the mid-band range using package-embedded discrete and planar capacitors [6].

The impedance profile of the PDN can be optimized by carefully choosing the values and placement locations of the decaps. This is critical since the anti-resonances (or parallel-resonances) that occur between decaps and between the decap and the PDN can lead to an impedance maximum. If this impedance maximum occurs at a frequency at which current transients exist, the severity of the SSN can be exacerbated. This choice and placement of decoupling capacitors can be accomplished [10] using an optimization engine based on a genetic algorithm (GA). However, the work in [10] relies on a cavity model for obtaining the PDN response. This is sufficient for simple rectangular two-layer geometries. However, for multilayer irregular geometries, the cavity model is inefficient.

An accurate impedance profile of the PDN can be obtained by simulating the stack as represented by the board and the package. Typically, this PDN is designed as alternating layers of power and ground planes with signal interconnects routed in between or on top of the planes. This...
arrangement serves two purposes - firstly, it allows a reduction in the package inductance, and secondly, signal lines on separate layers are isolated from each other due to the shielding effect of the planes.

Figure 1 shows a three layer package PDN supplying power to a mixed-signal IC. Multiple power supplies are typically required in modern packages and boards due to the various integrated components. Split planes are required to provide DC isolation to the different supply voltages. Also, holes are created in the solid power/ground planes in order to route signals or to provide via anti-pads. The switching activity of digital circuitry causes a time varying current to be drawn from its power supply terminals, Vdd1-Gnd1. Due to the associated inductance of the loop, SSN is generated. SSN can couple horizontally across a plane pair and across power islands. Also, SSN couples vertically through vias, and through apertures. This can be regarded as a coupling by means of a wrap-around current on the edges of the planes. Through these mechanisms, ground bounce can occur across the Vdd2-Gnd2 planes. Thus, it becomes critical to model split planes and apertures.

Of the methods in literature, the multi-layer finite difference method (M-FDM) provides good accuracy while maintaining high efficiency [1], as compared to full-wave EM solvers. This approach can model vertical coupling of energy through apertures or cut outs in the power planes, as well as horizontal coupling across slots separating split planes. Also, using an approach based on modal decomposition [1], it is possible to co-simulate signal interconnects along with the PDN. Due to this versatility, M-FDM has been proposed as the core simulator for the optimization engine.

Thus, the contributions of this paper are the following:

1. A robust GA based optimizer for the selection and placement of decoupling capacitors in multi-layer geometries
2. The use of M-FDM as the core simulator, which allows 1. Co-optimization of signal and power distribution networks, and 2. Reduction in the vertical coupling of noise.

The rest of the paper is organized as follows. A brief overview of the M-FDM technique is provided in Section 2. The details of the GA-based optimizer are discussed in Section 3. Test cases and results are shown in Section 4 and conclusions are presented in Section 5.

2. THE M-FDM FORMULATION

The underlying elliptic partial differential equation for the modeling of planes is a Helmholtz equation

$$\nabla^2 u + k^2 u = -j\omega \mu d J_z$$

where $\nabla^2$ is the transverse Laplace operator parallel to the planar structures, $u$ is the voltage, $d$ is the distance between the planes, $k$ is the wave number, and $J_z$ is the current density injected normally to the planes [9]. The problem definition is completed by assigning homogenous Neumann boundary conditions, which correspond to assuming a magnetic wall, or an open circuit, on the periphery of the planes. One method to solve the Helmholtz equation is by applying the finite-difference scheme. The 2-dimensional Laplace operator can be approximated as

$$\nabla^2 u_{i,j} = \frac{u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j}}{h^2}$$

where $h$ is the mesh length and $u_{i,j}$ is the voltage at node $(i,j)$ for the cell-centered discretization shown in Figure 2(a). This discretization results in a well-known bedspring unit cell model for a plane-pair consisting of inductors (L) between neighboring nodes, and capacitors (C) from each node to ground. Figure 2(b) shows the equivalent circuit obtained by discretizing a plane-pair into unit cells.
To obtain a model for the combined unit cell representing all the planes in the structure, consider the inductor elements in a unit cell as shown in Figure 3(a). $L_1$ is the per unit cell (p.u.c.) inductance between plane 1 and plane 2, whereas $L_2$ is the inductance between plane 2 and 3. Hence, reference planes are different in both models in Figure 3(b) and $L_2$ would be short-circuited if the same nodes on plane 2 were connected with each other. In order to avoid that, the p.u.c. inductances can be combined as shown in Figure 3(c) using a mutual inductance and assigning plane 3 as the reference plane. This model can be extended in a similar way to any number of planes. Physically, this model is based on the fact that there is a complete coupling of the magnetic field when the return current is on plane 3, as represented by the mutual inductance that is equal to $L_2$.

In terms of the admittance parameters, this model can be derived using the indefinite admittance matrix [2]. Following the formulation provided in [3], the total unit cell can be obtained as shown in Figure 4(b) for the example of three planes, where the bottom plane is chosen as the voltage reference plane. The equivalent circuit that would be obtained for a three layer geometry is shown in Figure 4(c). This allows for correct modeling of a multi-layer geometry with apertures.

The addition of a decoupling capacitor to this equivalent circuit is simple. The 2-port admittance matrix of the capacitor is stamped on to the nodes that it is connected to in the discretized PDN model. The typical equivalent circuit for a decap is shown in Figure 5. Here, ESR is the effective series resistance and ESL is the effective series inductance. The self resonance frequency (SRF) of the capacitor is given by:

$$\frac{1}{2\pi \sqrt{C \times ESL}}$$

This combined equivalent circuit model can be solved using a standard circuit solver. However, direct solution of the M-FDM equation using a linear equation solver can improve the memory requirements and speed, since the resulting admittance matrix is a sparse banded matrix. With nested dissection reordering, the flop count can be improved to $O(N^{1.5})$ and memory to $O(N \log_2 \sqrt{N})$ [4].

3. THE GA - BASED OPTIMIZATION ENGINE

The first step in the GA-based engine is the representation of the data as a set of 'genes' that make up a 'chromosome'. In this case, it is sufficient if the chromosome represents a particular choice of decaps and its placement (x- and y-locations, as well as layer connectivity), with cost being an optional parameter. Hence, all the available decaps can be arranged and indexed in order of their resonance frequency. It is critical that the capacitors be arranged in order of their self resonance frequencies, as will be explained later in this section. Hence, a chromosome representing a particular decoupling solution is only dependant on the number...
of capacitors being placed.

The general flowchart of the optimization process is shown in Figure 6. The inputs to the algorithm are 1. The package or board layout, 2. The target impedance and the bandwidth, 3. The number of decaps to place, and 4. A library of SMD- and embedded-decaps. Other algorithm specific inputs include the size of the population ($N$), the probability that crossover ($p_c$) and mutations ($p_m$) occur and the maximum number of optimization iterations. Given these inputs, the engine generates a random initial population, with decoupling capacitors chosen from the library, and with random locations on the board or package.

Each of the potential solutions are evaluated using the M-FDM engine for their $Z$-parameters at multiple noise source locations. The fitness of the $i^{th}$ solution has been defined as

$$f_i = \sum_{j=1}^{N_{\text{port}}} \sum_{k=1}^{N_{\text{freq}}} \left( w_1 [Z_{\text{tar},j} - Z_{j,k}(k)] + w_2 [Z_{\text{tar},j} > Z_{j,k}(k)] \right)$$

(4)

where $w_1$ and $w_2$ are weights chosen based on application, $N_{\text{port}}$ is the total number of noise sources at which the target impedance criterion must be met and $N_{\text{freq}}$ is the total number of frequency points in the simulation. Typically $N_{\text{freq}}$ is chosen to adequately cover the frequency range. $Z_{\text{tar},j}$ is the target impedance spec at the $j^{th}$ port, and $Z_{m,n}(k)$ is the $(m,n)$ entry of the $Z$-matrix at the $k^{th}$ frequency point. Also, the logical operation in Equation 4 returns 1 if the condition is met, and 0 otherwise. An additional input to the fitness function may be cost of the capacitors used.

At this stage, the population is sorted in order of their fitness, and the best solution is checked to see if it satisfies the target impedance at all ports and at all frequencies, in which case the algorithm terminates successfully. Otherwise, the crossover and mutation steps are carried out.

A pair of individuals, parents, in the population are chosen. The key idea here is that the fitness of a solution is directly related to the probability that it will be selected as a parent. This selection pressure leads to cumulative improvement in the fitness of solutions over generations.

In the crossover phase, the chromosomes of the parent solutions are mixed together at a rate determined by the crossover rate, $p_c$. Essentially, a new solution is created by copying a portion of the solution from each parent, as shown in Figure 7. This is performed iteratively until a new population is generated.

This new population is now subjected to mutation. By mutation, small random changes are made to each solution at the mutation rate, $p_m$. It is critical that these changes be small, as the probability of the solution improving reduces with the magnitude of the mutation. Hence, a mutation to a particular decap will result in choosing a new decap which resonates at a slightly higher or lower frequency. Hence, the library of decaps is indexed according to its SRF. At this step, a new generation has been created and the algorithm goes back to the selection phase.

A few of the best solutions are preserved intact (elitism,[5]) over all iterations to prevent the solutions from becoming worse over generations.

Termination occurs when the GA converges to a solution that satisfies the target impedance requirements. However, in some cases, there may be no improvement in fitness, whereupon the algorithm is terminated, and may have to be run again with different input parameters.

### 4. TEST CASES AND RESULTS

All simulations were performed on an Intel dual-processor, 3GHz Xeon Workstation with 3 GB of RAM. The crossover rate was chosen to be 0.5 (i.e., 50% of the solutions that form the next generation will be created by crossover). The mutation rate was 0.1. This implies that each character in the chromosome has a 10% chance of undergoing a mutation. The maximum number of iterations was 100.
4.1 Test Case 1

The first example is a simple three layer package showing a transmission line traversing a slot in the power plane. The slot introduces a return path discontinuity as well as an impedance discontinuity. The stack up and the geometry are shown in Figure 8. The insertion loss of the transmission line is shown in Figure 9(a). Also, assuming that transmission line is terminated in digital modules, there will be a need to reduce the impedance looking into the PDN. The impedance profile at the two ports between the power and ground plane (identical, due to symmetry) is shown in Figure 9(b). The optimizer was required to meet a target impedance of 500 mΩ, using fewer than 10 SMD capacitors, chosen from a library of 40 capacitors with an SRF ranging from 50 MHz up to 1 GHz. The parasitics of the SMD capacitors were included in the library. The optimizer converged in 16 iterations and with 9 decaps. The simulation required 8000 nodes, and the simulation time per frequency point in M-FDM was 140 ms (total time was 18 minutes). The top view of the placement is shown in Figure 10.

4.2 Test Case 2

The second test case is a realistic three layer server backplane. While the top and the bottom layers are continuous, the middle layer is not. The geometry of the middle layer and the cross section of the structure is shown in Figure 12. The location of the ports has also been shown. The board dimensions are 112 mm × 97 mm.

The impedance of the PDN looking into the planes is shown in Figure 13(a). Due to the discontinuous middle layer, although ports 1 and 2 are separated, they will still be coupled, as can be seen from the insertion loss, shown in Figure 13(b).

The optimizer was required to place 200 decaps to achieve a target impedance of 200 mΩ, over a frequency range of 100 MHz - 1 GHz. The optimizer converged in 90 iterations, and the optimized impedance profile is shown in Figure 14(a). Also, by providing the decoupling, the vertical coupling between ports has also been reduced, as shown in Figure 14(b).

The M-FDM model consisted of 89,000 nodes and required a simulation time of 2.5 sec/frequency point. The total time for convergence was 26 hours. This can be accelerated by seeding the initial population with reasonably good solutions rather than those generated at random.
5. CONCLUSIONS

With reduction in target impedance specifications, manual design of a board or package decoupling solution becomes a severe challenge. In this paper, a technique by which the selection and placement of decoupling capacitors can be automatized has been shown, thus removing the need for manual design iterations. The technique uses a GA-based optimizer, coupled with M-FDM as the core simulator. The optimizer was able to achieve the target impedance goals for a split plane structure as well as for a three layer irregular PDN. In the case of the split-plane structure, this optimized placement was able to improve signal integrity by providing a low-impedance return path. In the multi-layer structure, the placement also lead to isolation between the vertically separated ports. The technique can be further enhanced to allow cost-based optimization.

6. ACKNOWLEDGEMENTS

This work was supported by the Mixed Signal Design Tools Consortium (MSDT) at the Packaging Research Center, Georgia Tech under project number 2126Q0R.

7. REFERENCES