

DESIGN FLOW OF ROBUST ROUTED POWER DISTRIBUTION FOR LOW POWER ASIC

Dae Woon Kang, Yong-Bin Kim

Department of Electrical and Computer Engineering
Northeastern University
dwkang@ece.neu.edu, ybk@ece.neu.edu

ABSTRACT

This paper deals with a detailed systematic methodology to design power and ground distribution networks in low power ASICs. A design flow for determining the width of each branch in the power and ground networks meeting the design requirements is presented, along with a design example, where the objective is to minimize the area of the power and ground networks for several given constraints such as IR drop and EM. This methodology was applied to a low power speech recognition ASIC design and was successfully verified.

1. INTRODUCTION

VLSI power distribution networks are designed to provide the required voltages and currents to the transistors that perform the logic functions of a chip reliably over the chip lifetime[1]. The advantage of routed power networks is that it requires a small portion of the total routing resources available on the chip compared to grid power networks[2]. This is important for technologies with a limited number of metal layers, where designs tend to be routing limited. The disadvantage is that a routed topology has very little redundancy, since only a few power trunks carry all the current to circuit blocks. If a power trunk is too narrow, current is not easily drawn through the power trunk. Thus, over-designing was usually an acceptable solution[3]. But as technology moves into deeper sub-micron regions, this is not a valid approach. Over-designing causes too much area penalty.

In this paper, a design methodology for robust routed power distribution network is presented along with the design approaches that reduce the severity of the power integrity problems such as voltage drops and electro-migration (EM). Typically the routed power distribution is designed hierarchically, first at the local block (section II), then at the global level (section III) to satisfy reliability and noise tolerance.

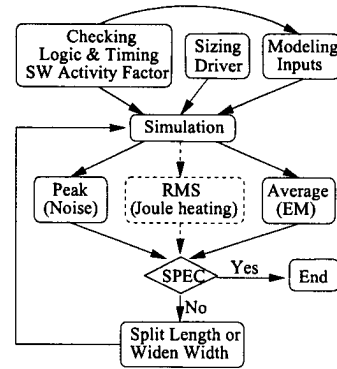


Fig. 1. Optimum local power bus line design methodology

2. LOCAL POWER DISTRIBUTION NETWORK DESIGN METHODOLOGY

Local power bussing is the power distribution within a Functional Unit Block (FUB) or module. At this level, the currents can be measured precisely by circuit simulation. The peak, average, and RMS currents are necessary to calculate the IR drop, EM, and Joule heating respectively. Power bus line widths are determined based on these measured currents. Fig.1 shows the proposed local power distribution network design flow developed for a low power speech recognition chip implementation.

The following design example shows how metal lines are designed for a data-path branch using the proposed methodology. It also explains how to avoid over-designing while keeping the optimum sizes of V_{dd} and V_{ss} rails to meet EM and noise requirements as shown in Fig.1. The power bussing branch shown in Fig.2 is shared by two sets of tri-state drivers that drive different signal busses, ABUS and BBUS. There are 64 drivers in each set, all of the same size. Assume the load of each driver to be 0.7pf.

Before a designer starts simulating, it is important to look at how the circuit is used and where it is situated. First, check the logic that enables the drivers. For example, the ABUS sources once every four cycles and the BBUS every other cycle on average. The data transitions are expected to

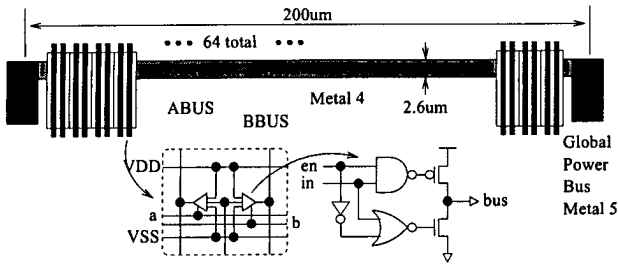


Fig. 2. Power bussing example of data-path bus drivers

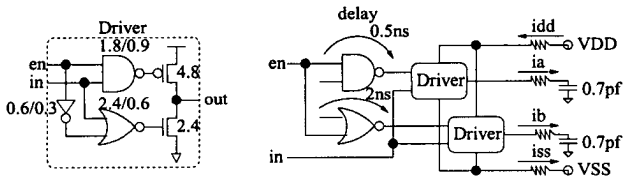


Fig. 3. Example of simulation schematic

be random, with 0-1, 1-0, 0-0, and 1-1 equally likely. Other useful facts are the enable and disable delays, and the transition times on the bus. In this case, enable delay should be added carefully in order to eliminate the possible instantaneous contention with other drivers. The NAND and NOR gates feeding the bus drivers should also be sized to keep sneak current flow between the P and N transistors in the drivers at minimal levels.

The power feeds are determined by the global bussing scheme in the data-path. The branch for the drivers is M4, fed at both ends from the M5 global bussing. The goal of the example is to determine the width of this branch.

To simplify the simulation, it is only necessary to simulate one of the ABUS and BBUS drivers due to their regular feature. Since the timing of the signals is important in noise design, the paths generating the IN, EA, and EB signals should be modeled properly. The currents can be monitored by placing small resistors (0.01Ω) in series with the Vdd and Vss feeds to the cell and in the cell output. The simulation schematic is shown in the Fig.3, where the numbers on gate represent the widths of PMOS and NMOS with 0.18µ length respectively. And the simulation results in Fig.4 for one driver over two clock cycles are tabulated in Table I.

Table 1. Example of simulation results

Current	Peak (mA)	RMS (Ma)	Average (mA)
idd	1.457	0.387	-0.167
iss	1.663	0.448	0.165
iA(=iB)	1.633	0.580	0.243 µA

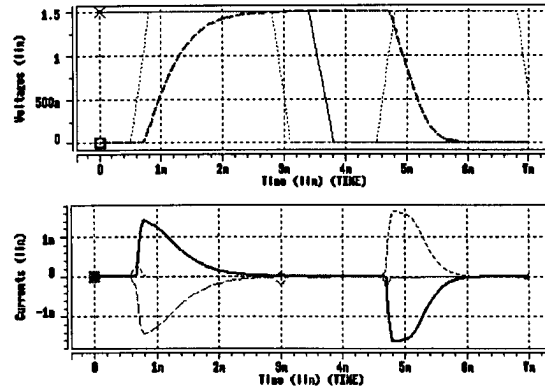


Fig. 4. Simulation waveforms

The bus noise should be within tolerable range. This requires a simulation or calculation of worst case instantaneous peak currents in the branch. In this example, the enable onto BBUS is 1.5nsec later than ABUS. This allows us to design for only one peak, instead of the sum of both. Here, the relative delay separates the signals, but it is useful to check any possible signal overlap in other cases.

Assume the enables for ABUS and BBUS were generated with logic that made them mutually exclusive. Each bus has a common enable signal, so all 64 drivers can be active at once. The worst case is when all 64 ABUS or BBUS drivers are driving high or low simultaneously.

Assume the width (W) calculated for EM gives less than 100mV drop in the branch, and metal width requirement for EM (See section III.) is 2.6µm for 200µm length of the branch and M4 sheet resistivity (0.71 Ω/square). Then, the total resistance of the branch can be calculated.

$$R(b) = \rho * L/W = 0.71 * 200 / 2.6 = 54.5 \Omega \quad (1)$$

The total peak current is 64 * 1.457 mA = 93.3mA in Vdd; Vss is about the same in this case. The voltage drop of 93.3mA through 54.5 Ω is about 5.1V drop. Again, this is a gross overestimate. Since the branch ties to the global bussing at each end, the current is zero in the center and the branch can be split for analysis. This reduces the current in each half branch to half the total, and the length is reduced to half the total length, and the total drop is reduced by a factor of four. In the case of the data-path, the drivers are evenly spread across the branch, so the average peak current in the half branch is half the total. This gives another 2X reduction in drop.

Consequently, these methods reduce the voltage drop by a factor of eight, giving a peak drop of 635mV. The M4 branches can be widened to reduce the resistance, as shown in Fig.5. This will result in the 4X improvement (159mV), since each side will contain half the cells (32) and run half as far (500µm).

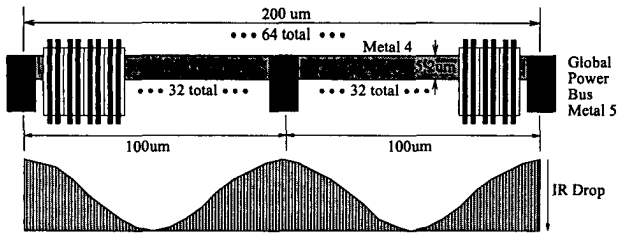


Fig. 5. Final bussing design

Activity determined by the signal timing and the control logic, can also be used to reduce the effective average current. This really depends on the type of logic. For example, one pin of a 32-decoder output may only change once in 32 clocks on average. With random inputs, a NOR gate will have a high output once every 2^n clocks on average, where n is the number of inputs. NAND gates with random inputs only go low at the same rate. This means that control logic gate activity factor using these gates can be assumed to be 0.25 on average.

3. GLOBAL POWER DISTRIBUTION NETWORK DESIGN METHODOLOGY

The primary function of on-chip power distribution is to provide a supply voltage with minimal IR drop and Ldi/dt. Equally important is the need to minimize the voltage difference between any two circuits on the chip because this condition influences the circuit noise margins directly.

The determination of the widths of the power networks branch is subject to a set of constraints such as maximum voltage drop allowed and EM requirement. Fig.6 shows the design flow for a global power distribution networks design. The following subsections explain the global power distribution network design issues in the proposed flow in more detail. And Fig.7 shows an example of a power distribution network in a low power speech recognition VLSI implementation. The network was designed and verified successfully using the proposed approach.

3.1. Maximum IR Drop

The current that flows through a branch is determined by the current requirements on the pins of the FUBs in the worst case. And if the branch current is constant, then the voltage drop across the branch can be expressed as:

$$V = \rho IL/W. \quad (2)$$

The FUBs can be modeled as constant current sinks (See section II). The voltage drop constraint specifies the total allowable voltage drop between the power pad and each power terminal of the FUBs. The optimum distribution with

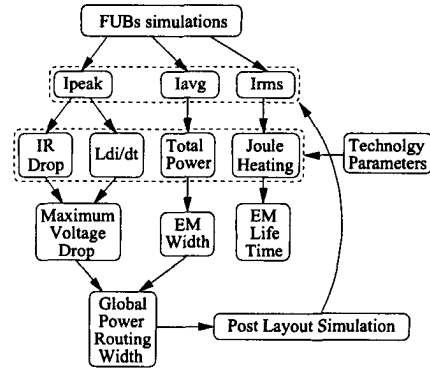


Fig. 6. Global power distribution networks design flow

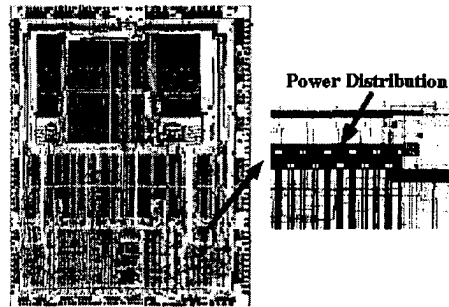


Fig. 7. A speech recognition chip and an example of its power distribution

the minimum area depends on the widths and lengths of branches. The global power distribution has to satisfy the following Equation.

$$\sum_{i \in U} \rho I_i \frac{L_i}{W_i} \leq v_{max} \quad (3)$$

where U is the total paths from power pad to FUB terminals, v_{max} is the allowable voltage drop corresponding to the path i , and ρ , W_i , and L_i are the sheet resistance, width, and length associated with path i , respectively.

3.2. Maximum Noise from Ldi/dt

Ldi/dt can cause voltage to drop severely. When a capacitive node is switching, the discharging current indicated in Fig.8(a) can often be approximated as having a triangular shape shown in Fig.8(b). Since the charge $Q (= C_{sw} V_{sw})$ in C_{sw} is $I_p \frac{t_f}{2}$, the internal Vss will experience a peak Ldi/dt noise of

$$V_{nmax} = L \frac{I_p}{t_a} = L \left(2 \frac{C_{sw} V_{sw}}{t_f} \right) \frac{1}{t_a}. \quad (4)$$

where L is the inductance in the Vss path, and I_p is the peak current through the driving buffer. If the total switching

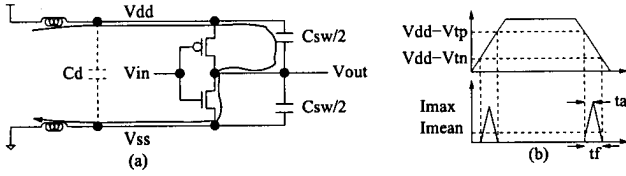


Fig. 8. (a) On-chip switched capacitance contributing to di/dt noise and (b) input switching waveform and model for short-circuit current

node capacitance is evenly distributed between power and ground, the actual switching capacitance is half of the total switching capacitance, as shown in Fig.8. However, if there is a decoupling capacitance, C_d , most of the initial current spike passes through C_d . Due to charge sharing, thus, the maximum noise is reduced to [4]

$$V_{nmax} = \frac{V_{dd}C_{sw}/2}{2C_d + C_{sw} + I_j^2/6L}. \quad (5)$$

3.3. EM Widths

Another constraint of power networks is EM. The metal width satisfying the EM requirement is calculated using

$$W = \frac{P_{tot}}{T J_{max} V_{DD}}, \quad (6)$$

where T is the thickness of metal, P_{tot} is total power consumption by the metal, and J_{max} is the maximum current density. Thus, if P_{tot} is known, power line width can be designed accordingly.

3.4. Joule Heating Effect

The EM lifetime reliability of metal interconnects is modeled by the Black's Equation given by

$$TTF = A j^{-n} \exp\left(\frac{Q}{k_B T_m}\right) \quad (7)$$

where TTF is the time-to-fail. A is a constant that is dependent on geometry and micro structure of the metal, and j is the average current density. The exponent n under normal conditions is typically 2, while Q is the activation energy for grain-boundary diffusion, k_B is the Boltzmann constant, and T_m is the metal temperature. The EM lifetime decreases exponentially with increasing temperature as shown in Equation (7), and the self heating temperature increases as metal resistance increases as shown in the Equation (8). Therefore, the power distribution network must comply with the requirements imposed by Eq. (7) and (8).

$$\Delta T_{selfheating} = I_{rms}^2 R R_\theta, \quad (8)$$

where R_θ is the effective thermal impedance of the metal line[5].

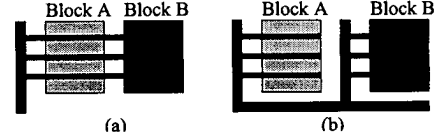


Fig. 9. Routings (a) through a block, and (b) around blocks

3.5. Area vs. IR Drop Balance

As shown in Fig.9(a), if power is routed through block A to block B, a larger IR drop will occur in block B. This is due to the fact that power is being consumed by block A before it reaches block B. To balance the IR drops, one could route around the block and feed power to each block separately, as shown in Fig.9(b), because the total IR drop is based on the resistance seen from the pin to the block. Although the main trunks are large enough to handle all the current flowing through separate branches, the T-junctions have a high current density and may be prone to EM problems. In addition, this configuration requires more chip area to implement.

4. CONCLUSIONS

Although routed power networks provide the best choice for ASIC applications with limited layers due to their small wiring area, it is difficult to design the routed power networks since it has very little redundancy. The hierarchical methodology for designing the routed power networks is proposed, and the methodology has been applied to a low power speech recognition ASIC design and was successfully verified. The developed power distribution network design methodology will be a good reference for any future low power ASIC design.

5. REFERENCES

- [1] S. Rochel et al, "Full-chip Reliability Analysis", *ACM Proc. IRPS* pp.356-361, 1998.
- [2] A. Chandrakasan et al, "Design of High-Performance Microprocessor Circuits", *IEEE Press* 2001.
- [3] S. Chowdhury et al, "The Construction of Minimal Area Power and Ground Nets for VLSI Circuits", *ACM Proc. DAC* pp.794-797, 1985.
- [4] P. Larsson, "di/dt Noise in CMOS Integrated Circuits", *Proc. Analog Integrated Circuits and Signal Processing*. no. 1/2, pp.113-130, 1997.
- [5] K. Banerjee et al, "Analysis and Optimization of Thermal Issues in High-Performance VLSI", *ACM Proc. ISPD* pp.230-237, 2001.