ON CHIP DECOULING CAPACITOR

[1] ROHINI N. DEOKAR , [2] P. R. THORAT

[1, 2] Department of Electronics Engineering, Savitribai Phule Womens Engg. college, B.A.M.U. University, Aurangabad, Maharashtra, India.

rohinideokar30@gmail.com, Thorat.popat.r.@gmail.com

ABSTRACT : *Decoupling capacitors are widely used to reduce power supply noise. On-chip decoupling capacitors have traditionally been allocated into the white space available on the die based on an unsystematic or ad hoc approach. In this way, large decoupling capacitors are often placed at a significant distance from the current load, compromising the signal integrity of the system. This issue of power delivery cannot be alleviated by simply increasing the size of the on-chip decoupling capacitors. To be effective, the on-chip decoupling capacitors should be placed physically close to the current loads. The area occupied by the on chip decoupling capacitor, however, is directly proportional to the magnitude of the capacitor. The minimum impedance between the on-chip decoupling capacitor and the current load is therefore fundamentally affected by the magnitude of the capacitor. A distributed on-chip decoupling capacitor network is proposed in this paper. A system of distributed on-chip decoupling capacitors is shown to provide an efficient solution for providing the required on-chip decoupling capacitance under existing technology constraints. In a system of distributed on chip decoupling capacitors, each capacitor is sized based on the parasitic impedance of the power distribution grid. Various tradeoffs in a system of distributed on-chip decoupling capacitors are also discussed. Techniques presented in this paper are applicable not only for current technologies, but also provide an efficient placement of the on chip decoupling capacitors in future technology generations. Index Terms Power distribution systems, power distribution grids, decoupling capacitors, power noise.*

1. INTRODUCTION

On-chip decoupling capacitors (decaps) are widely used to reduce power supply noise. Capacitors play a critical role in the stable operation of digital electronics by protecting sensitive microchips from noise on the power signal which can cause anomalous behaviors. Capacitors used in this application are called decoupling capacitors and should be placed as close as possible to each microchip to be most effective, as all circuit traces act as antennas and will pick up noise from the surrounding environment. Decoupling and by-pass capacitors are also used in any area of a circuit to reduce the overall impact of electrical noise. A decoupling capacitor is a capacitor used to decouple one part of an electrical network (circuit) from another. Noise caused by other circuit elements is shunted through the capacitor, reducing the effect it has on the rest of the circuit. For example, if the voltage level for a device is fixed, changing power demands are manifested as changing current demand. The power supply must accommodate these variations in current draw with as little change as possible in the power supply voltage. When the current draw in a device changes, the power supply can't respond to that change instantaneously. Decoupling is the process of breaking coupling between portions of systems and circuits to ensure proper operation. Typically, designs use NMOS decaps between standard-cell blocks and NMOS+PMOS decaps within the blocks. Decoupling capacitors (decap) are often used to filter out noise in

the power distribution system (PDS). Decaps acts as a local source of energy for a short period. With the scaling of CMOS technologies the power supply voltage is lowered, clock frequency has gone up, and more functionality is integrated on-chip resulting in higher simultaneous switching noise (SSN). As a result signal integrity of on-chip power supply has become a major concern. Placement and sizing of decaps, effective utilization of on-chip whitespace, resonant free voltage responses for a wide range of frequencies are some of the key challenges faced with the shift towards deep submicron regime. Decoupling capacitors are used to reduce unwanted AC signals riding on DC supply circuits and also in places in a circuit where AC signals need to be eliminated. These are usually placed between the DC supply and the ground of the circuit or directly across another component. Decoupling capacitors known as "charge reservoirs" are placed between power and ground lines to maintain low target impedance and to reduce the noise in power distribution network. Low impedance and resonant free response are two requirements for maintaining power and signal integrity of the PDS. Power distribution system (PDS) spans different levels of hierarchy consisting of voltage regulator module, power distribution network on board, on package and on-chip the power distribution system (PDS). Decaps acts as a local source of energy for a short period. With the scaling of CMOS technologies the power supply voltage is lowered, clock frequency has gone up, and more

functionality is integrated on-chip resulting in higher simultaneous switching noise (SSN). As a result signal integrity of on-chip power supply has become a major concern. Placement and sizing of decaps, effective utilization of on-chip whitespace, resonant free voltage responses for a wide range of frequencies are some of the key challenges faced with the shift towards deep submicron regime. Decoupling capacitors are used to reduce unwanted AC signals riding on DC supply circuits and also in places in a circuit where AC signals need to be eliminated. These are usually placed between the DC supply and the ground of the circuit or directly across another component. Decoupling capacitors known as "charge reservoirs" are placed between power and ground lines to maintain low target impedance and to reduce the noise in power distribution network. Low impedance and resonant free response are two requirements for maintaining power and signal integrity of the PDS. Power distribution system (PDS) spans different levels of hierarchy consisting of voltage regulator module, power distribution network on board, on package and on-chip.

the power distribution system (PDS). Decaps acts as a local source of energy for a short period. With the scaling of CMOS technologies the power supply voltage is lowered, clock frequency has gone up, and more functionality is integrated on-chip resulting in higher simultaneous switching noise (SSN). As a result signal integrity of on-chip power supply has become a major concern. Placement and sizing of decaps, effective utilization of on-chip whitespace, resonant free voltage responses for a wide range of frequencies are some of the key challenges faced with the shift towards deep submicron regime. Decoupling capacitors are used to reduce unwanted AC signals riding on DC supply circuits and also in places in a circuit where AC signals need to be eliminated. These are usually placed between the DC supply and the ground of the circuit or directly across another component. Decoupling capacitors known as "charge reservoirs" are placed between power and ground lines to maintain low target impedance and to reduce the noise in power distribution network. Low impedance and resonant free response are two requirements for maintaining power and signal integrity of the PDS. Power distribution system (PDS) spans different levels of hierarchy consisting of voltage regulator module, power distribution network on board, on package and on-chip [1].

Histroy:

On-chip decoupling capacitors (decaps) are widely used to reduce power supply noise. Capacitors play a critical role in the stable operation of digital electronics by protecting sensitive microchips from noise on the power signal which can cause anomalous behaviors. Capacitors used in this application are called decoupling capacitors and should be placed as close as possible to each microchip to be most effective, as all circuit traces

act as antennas and will pick up noise from the surrounding environment. Decoupling and by-pass capacitors are also used in any area of a circuit to reduce the overall impact of electrical noise.

Typically, designs use NMOS decaps between standard-cell blocks and NMOS+PMOS decaps within the blocks. Starting at the 90nm CMOS technology node, the traditional decap designs may no longer be suitable due to increased concerns regarding thin-oxide gate leakage and electrostatic discharge (ESD) reliability. This thesis investigates new decap design approaches that address gate leakage and ESD. A cross-coupled design is described that has been recently introduced by cell library developers to handle ESD problems. Three modifications of the cross-coupled design are introduced here and the tradeoffs among transient response, gate leakage and ESD performance are analyzed. The modifications offer designers greater flexibility in decoupling capacitor design for 90nm and below. To improve the power-grid noise reduction capability in the areas between blocks, two versions of a switched-decap design are proposed. One provides excellent decap performance but consumes large power, whereas the other saves power but suffers from excessive delay. A novel low-power voltage regulator using switched decaps is proposed to better balance performance and power consumption. Decoupling capacitors (decap) are often used to filter out noise in the power distribution system (PDS). Decaps acts as a local source of energy for a short period. With the scaling of CMOS technologies the power supply voltage is lowered, clock frequency has gone up, and more functionality is integrated on-chip resulting in higher simultaneous switching noise (SSN). As a result signal integrity of on-chip power supply has become a major concern. Placement and sizing of decaps, effective utilization of on-chip whitespace, resonant free voltage responses for a wide range of frequencies are some of the key challenges faced with the shift towards deep submicron regime.

1.1 How a decoupling capacitor works?

A decoupling capacitor serves as a storage device for small amounts of energy. In situations like the one mentioned above, in which a device's output must fluctuate current or voltage levels without putting strain on its power supply, a decoupling capacitor is fastened to the device in between its output and power supply in order to buffer any changes that occur. At high frequencies, a decoupling capacitor is able to maintain power levels to a device's output while the device's power supply has time to adjust t the changes.
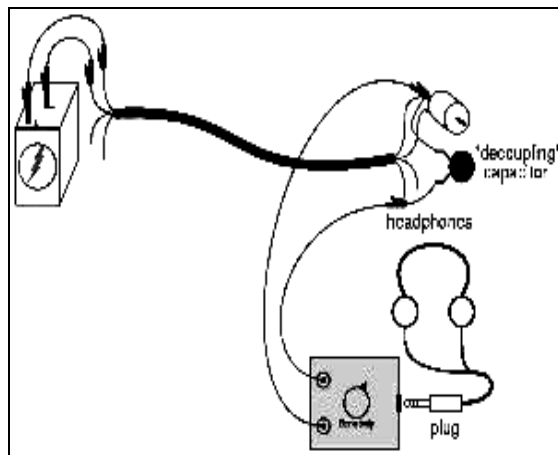
Figure.1: Decoupling capacitor

1.2 What is proper decoupling and why is it necessary

Most ICs suffer performance degradation of some type if there is ripple and noise on the power supply pins. A digital IC will incur a reduction in its noise margin and a possible increase in clock jitter. For high performance digital ICs, such as microprocessors and FPGAs, the specified tolerance on the supply ($\pm 5\%$, for example) includes the sum of the dc error, ripple, and noise. The digital device will meet specifications if this voltage remains within the tolerance. The traditional way to specify the sensitivity of an analog IC to power supply variations is the power supply rejection ratio (PSRR). For an amplifier, PSRR is the ratio of the change in output voltage to the change in power supply voltage, expressed as a ratio (PSRR) or in dB (PSR).

PSRR can be referred to the output (RTO) or referred to the input (RTI). The RTI value is equal to the RTO value divided by the gain of the amplifier. Graph 1 shows how the PSR of a typical high performance amplifier (AD8099) degrades with frequency at approximately 6 dB/octave (20 dB/decade). Curves are shown for both the positive and negative supply. Although 90 dB at dc, the PSR drops rapidly at higher frequencies where more and more unwanted energy on the power line will couple to the output directly. Therefore, it is necessary to keep this high frequency energy from entering the chip in the first place. This is generally done with a combination of electrolytic capacitors (for low frequency decoupling), ceramic capacitors (for high frequency decoupling), and possibly ferrite beads. Power supply rejection of data converters and other analog and mixed-signal circuits may or may not be specified on the data sheet. However, it is very common to show recommended power supply decoupling circuits in the applications section of the data sheet for practically all linear and mixed-signal ICs [3].

Low frequency noise requires larger electrolytic capacitors which act as charge reservoirs to transient currents. High frequency power supply noise is best reduced with low inductance surface mount ceramic capacitors connected directly to the power supply pins of the IC. All decoupling capacitors must connect directly to a low impedance ground plane in order to be effective. Short traces or via are required for this connection to minimize additional series inductance. Ferrite beads (nonconductive ceramics manufactured from the oxides of nickel, zinc, manganese, or other compounds) are also useful for decoupling in power supply filters.
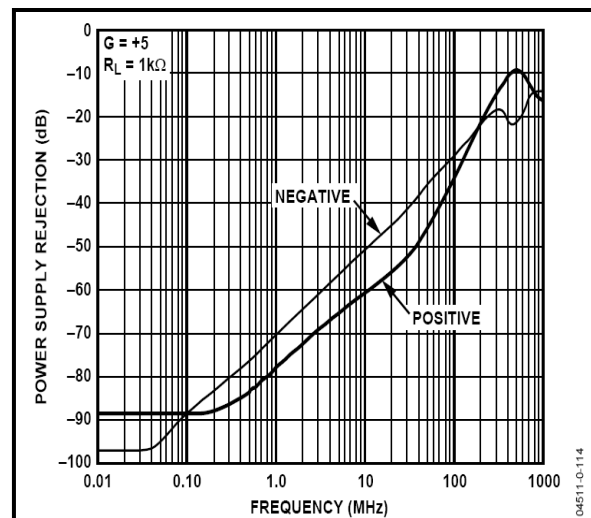
Graph 1: Power Supply Rejection vs. Frequency For the AD8099 High Performance Op Amp

At low frequencies (<100 kHz), ferrites are inductive; thus they are useful in low-pass LC filters. Above 100 kHz, ferrites becomes resistive (high Q). Ferrite impedance is a function of material, operating frequency range, dc bias current, number of turns, size, shape, and temperature. The ferrite beads may not always be necessary, but they will add extra high frequency noise isolation and decoupling, which is often desirable. Possible caveats here would be to verify that the beads never saturate, especially when op amps are driving high output currents. When a ferrite saturates it becomes nonlinear and loses its filtering properties. Note that some ferrites, even before full saturation occurs, can be nonlinear. Therefore, if a power stage is required to operate with a low distortion output, the ferrite should be checked in a prototype if it is operating near this saturation region.

2. ON CHIP DECOUPLING CAPACITOR

As integrated circuit (IC) technology scales, more and more transistors are being placed within a single chip, while the clock frequency continues to increase into the gigahertz range. The result is that large transient currents are drawn from the power supply rails in just a few hundred picoseconds in modern custom and application-specific integrated circuit

(ASIC) chips. Meanwhile, the supply voltage is scaled with technology to reduce overall power consumption, and as a consequence, the circuitry becomes more prone to power-supply noise. The management and regulation of the quality of the on-chip power supply is a major challenge. The power grid, which provides VDD and VSS (or ground) signals throughout the chip, experiences fluctuations in value due to a variety of noise sources. If the supply voltage noise or variation is excessively large, it may lead to problems such as delay variation, timing unpredictability, or even improper functionality. A commonly used metric, noise budget, is defined as the maximum allowable noise amplitude. Typically, it is required to keep the power supply noise within a certain percentage (e.g., 10%) of the nominal supply voltage VDD. Namely, 10% of VDD – VSS is typically the noise budget, a rule-of-thumb used in the industry. Circuit designers must ensure that the chip operates correctly if the maximum voltage difference between VDD and VSS is 10% or smaller than the nominal value [2].

In today's advanced deep-submicron (DSM) technology, the power grid noise is due to two main issues:

(1) As the power lines made of metal wires become thinner, the wire resistance R increases. When logic gates switch and a current I flow through the power lines to deliver charge to the gates, the voltage drop ΔV at the gates is $\Delta V1 = IR\cdot$. This type of power-supply noise is known as IR drop.

(2) Due to package pin inductance and thin-interconnect inductance, the power lines experience inductance effect when the current flow changes with respect to time. This second source of voltage drop is given by $\Delta V2 = IR + L\, dI/dt$. The two power supply noise components are illustrated in Figure 2, which depicts two inverters connected to an off-chip voltage supply through the on-chip power grid. Considering the two components together, the overall voltage drop ΔV at any point in the power grid.
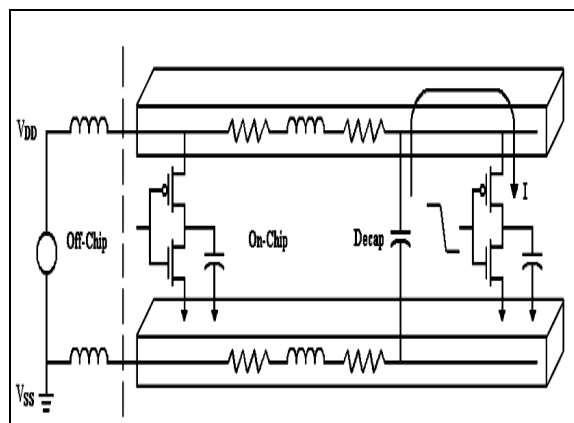
$$\Delta v = IR + L\, dI/d$$

Figure 2 : Use of decoupling capacitor for power grid.

To illustrate the IR drop, all the nodes in the power grid are initially charged to VDD with no activity in the circuit. As the second inverter starts to switch, the wire resistance along the VDD line creates voltage drops as current flows from the external voltage source towards the second inverter. Similarly, the ground grid is subject to the same type of problem when the outputs of the buffers switch low, except that the voltage level of the ground line will increase. This is sometimes referred to as ground bounce. In practice, IR drop can be caused by simultaneous switching of clock buffers, bus drivers, memory decoder drivers, and so on, when there is high activity in the circuit. These simultaneous switching activities can happen anywhere on the chip. Thus, all regions in the chip are susceptible to IR drop. In a wire-bond (e.g., dual-inline) package, the supply voltage level remains relatively high at the periphery of chip where the voltage supply I/O pads are located, and drops noticeably at the centre of the chip. In contrast, in a flip-chip (or ball-grid array) package, the centre of the die has rather high voltage level, whereas the periphery of the die experiences larger IR drops.

A common technique for reducing power supply noise and keeping the noise within the noise budget is through the use of on-chip decoupling capacitors (decaps). Decaps are essentially capacitors that hold a reservoir of charge and are placed close to the power pads and near any large drivers. When large drivers switch, the decaps provide instantaneous current to the drivers to reduce IR drop and Ldi/dt effects, and hence keep the supply voltage relatively constant. As shown in Figure 2, the on-chip decap delivers current to charge up the load capacitance of the second inverter when it switches. The supply voltage level is relatively constant at the inverter tap point since the decap is nearby, so ΔV is minimal.[4]

2.1 On chip decoupling capacitor using NMOS & PMOS

The concept of the proposed on-chip decoupling capacitors with series resistors and ideal switches is shown in Fig.3.-(a). In order to change the ESR of on-chip decoupling capacitors, resistors and switches are connected to the decoupling capacitors additionally. Fig.3-(b) shows also the proposed scheme of on-chip decoupling capacitors in a CMOS process. The proposed scheme has NMOS capacitors and PMOS switches that can be controlled by digital bit of the ESR controller. In order to control the total resistance of ESR linearly, the resistance of their parallel combination should be an arithmetical series. As the result, sizes of PMOS switch can be decide [2].
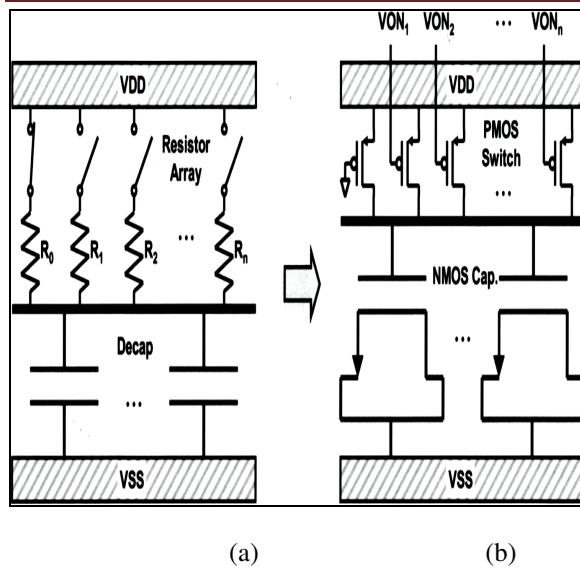
(a)   (b)

Figure 3 (a) Conceptual scheme of on-chip decoupling   capacitors with series resistors and ideal switches.

        (b) The proposed scheme of on-chip decoupling capacitors in   a CMOS process

2.2 Decoupling capacitor for NMOS

A standard decap is usually made from NMOS transistors in a CMOS process. As shown in Figure 4, the gate of the NMOS transistor is connected to VDD, whereas source, drain and substrate of the transistor are tied to VSS. This approach is considered effective because the thin-oxide capacitance of the transistor gate provides a higher capacitance than any other oxide capacitance available in a standard CMOS fabrication process. For this MOS decap, the first-order calculation of the capacitance is WLCOX, where W is the transistor width, L is the transistor length, and COX is the oxide capacitance per unit area. Accurate capacitance model needs to include the parasitic fringing and overlap capacitance of the transistor. At the 90nm technology node, the oxide thickness of a transistor is reduced to roughly 2.0nm or less. The thin oxide causes two new problems possible electrostatic discharge (ESD) induced oxide breakdown and gate tunneling leakage . Potential ESD oxide breakdown increases the likelihood that an IC will be permanently damaged during an ESD event, and hence raises a reliability concern. Higher gate tunneling leakage increases the total static power consumption of the chip. As technology scales further down, with a thinner oxide, the result is an even higher ESD risk and more gate leakage. The standard decap design using NMOS transistors experiences these two problems and therefore becomes rather inappropriate for 90nm and below [4].
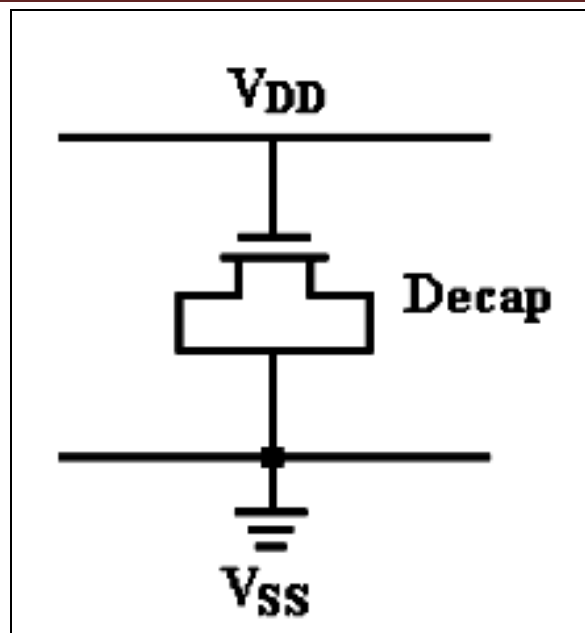

Figure 4 : Decoupling capacitor implemented using an NMOS device.

While satisfying ESD reliability and gate leakage limitations, decap designs must also meet the transient performance requirements. Since a 90nm process (or below) usually provides the capability of running at gigahertz frequencies, the decap must respond in the order of a hundred picoseconds.

2.3 Cross-Coupled Decap

Knowing that the standard N+P decap design for standard cells may no longer be suitable for 90nm technology due to increased ESD risk, a new cross-coupled decap design has been proposed to address this issue. In the new cross coupled design (Figure 5), the drain of the PMOS connects to the gate of the NMOS, whereas the drain of the NMOS is tied to the gate of the PMOS.
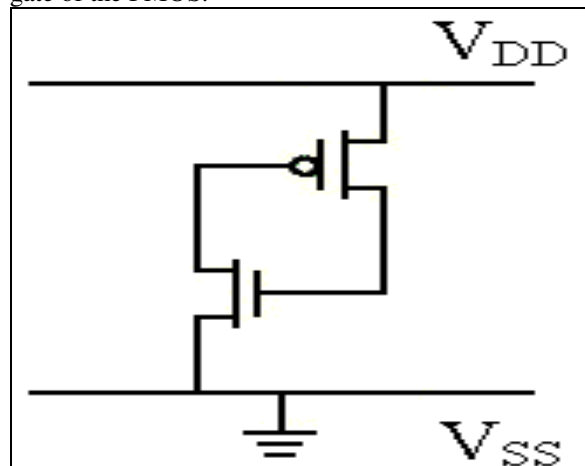

Figure 5 : Cross-coupled decap schematic

From the layout perspective, this cross-coupled circuit can be seen simply as a terminal-swapped version of the standard decap. In other words, the decap transistor areas need not to be modified, while only the metal wire connections are modified. Thus,

this new design does not require additional area in layout, compared to the standard design. Both transistors in this design are still in the linear region. In the standard decap design, the gates of the transistors are directly connected to either VDD or VSS, depending on the transistor type. In this case, the gate of the NMOS device is connected to VDD through the channel resistance of the PMOS device. Similarly, the gate of the PMOS device is tied the channel resistance of the NMOS device and then connected to VSS. The added channel resistance to the gate provides the input resistance Rin for ESD protection. The input resistance can help to limit the maximum current flow to the decap so that the voltage seen from the gate of the decap is also limited [1].

2.4 Gate Decouping

A new decap structure that saves gate leakage. The structure is called gated decap, as shown in Figure 6. A control transistor is inserted between the standard NMOS-only decap and the VSS line. The source and drain of the decap are connected to the source of the control transistor, making the node a virtual ground (V_GND). The drain of the control transistor is tied to the real VSS. As shown in Figure the substrate of the decap is still attached to VSS. There are two major components in gate leakage: leakage current from gate to channel (Igc), and from gate to substrate (Igb). The current Igc can be partitioned into two: leakage current from gate to source (Igcs) and from gate to drain (Igcd) . The amount of gate leakage from gate to substrate Igb is roughly 10x smaller than the leakage from gate to channel Igc . Thus, the substrate of the decap does not need to be tied through the control transistor, and the leakage current Igb is neglected.
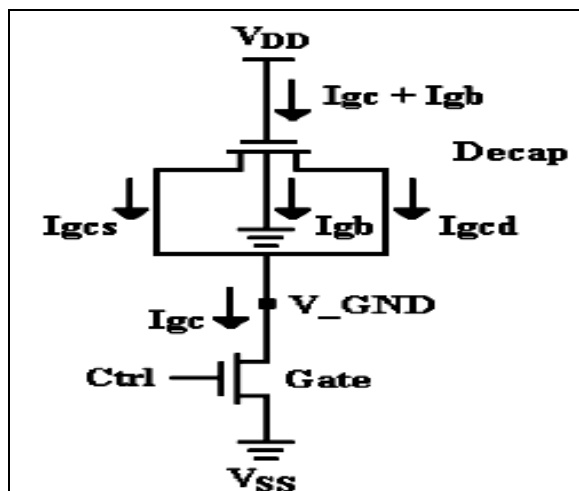

Figure.6: Basic gate schematic.

There are two modes of operation of the circuit: active mode and power saving mode. When in the active mode, the Ctrl signal of the control transistor is turned high. The gated decap operates almost like the standard decap, except that there is a small channel resistance of the control transistor. The size of the

control transistor needs to be large to have the channel resistance small since a large resistance will reduce the transient response of the decap. When in power saving mode, the Ctrl signal is turned low so that the control transistor operates in the sub threshold regime. The node V_GND can be considered a virtual ground (floating), where the voltage at V_GND can be determined by the series resistance of Reff of the decap and the channel resistance of the control transistor. In this configuration, the gate leakage saving is projected to be 99% in a 70nm process [4].

The basic idea of the gated decap is from multi-threshold CMOS (MTCMOS). The control transistor comes from the concept of the sleep transistor in MTCMOS. As expected, the control transistor should have a high VT to keep the sub threshold leakage small. The largest challenge of this gated decap would be the proper selection of the Ctrl signal. At the top level, the Ctrl signal can be driven by the hardware/software interface. When there is no activity in the system, the operating system (software) will set up the signal to force the chip into power saving or standby mode. From the hardware architectural level, the Ctrl signal can be managed by some self-predictive architecture. At the circuit level, it is desired that the gated decap is self-maintained, and no external circuitry is required to control it on or off. In that case, it may need to have a special clock, as shown in Figure 6 Before the regular clock rises, the Ctrl signal can be set high to allow some setup time for the decap to fully setup. When the regular clock falls, the Ctrl signal can also fall simultaneously to save power. The time period when the Ctrl signal is low can be considered as the power saving period.

3. CONCLUSIONS AND FUTURE SCOPE

With this system capacitor used to decouple one part of an electric network from another. As technology scales further into the deep submicron regime, with increasing clock frequency and decreasing supply voltage, maintaining the quality of power supply becomes a critical issue. On-chip power supply noise, due to IR drop and Ldi/dt effects, has a great impact on delay variation, and may even cause improper functionality. Power supply noise can be reduced by placing decoupling capacitors close to power pads and large drivers throughout the power distribution system. Decaps provide instantaneous current to the switching drivers and keep the power supply within certain noise budgets.

A number of issues regarding decoupling capacitors will have to be addressed in the near future. First, knowing that the thin-oxide decaps leak a significant amount of current in 90nm and below, it is important to place only the necessary amount of decaps in a certain design to avoid overdesign. The use of thick-oxide decaps may not solve the issue completely because the effective capacitance is much less for

thick-oxide devices and the total free area for decaps is limited. Also, the active decaps provide better noise reduction performance but at a cost of increased standby power requirement, compared to the passive decaps. Therefore, to determine the optimal number of thick-oxide, thin-oxide and active decaps to be placed into a design remains a challenge.

4. APPLICATIONS

Decoupling capacitors are often used in devices that have several settings, such as power tools, hair dryers, and conventional ovens. They allow a device to switch back and forth between varying settings, such as speed or temperature, before the device's power supply is able to maintain that setting. Devices could still have varying settings without a decoupling capacitor, but the effect would not be instantaneous and the user would have to wait for the power supply to adjust itself to the new setting.

5. ADVANTAGES

 Decoupling capacitors are advantageous because they are small and hardly noticeable, do not require any additional energy, and allow devices to change between various settings instantaneously. Decoupling capacitors also prevent damage to the power supply from occurring by buffering any fluctuations that occur in the device's output. This is especially true in devices that contain a loop-back mechanism that induces unused energy from the device's output back into the device's input.

6. REFERENCES

[1] N. Na; T. Budell, C. Chiu, E. Tremble, and I Wemple, "The Effects of On-Chip and Package Decoupling Capacitors and an Efficient ASIC Decoupling Methodology," Proceedings of Electronic Components and Technology (ECTC '04), Volume 1, pp. 556-567, June 2004.

[2] H. Su, S. S. Sapatnekar, and S. R. Nassif, "Optimal Decoupling Capacitor Sizing and Placement for Standard-Cell Layout Designs," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Volume 22, Issue 4, pp. 428-436, April 2003.

[3] D. A. Hodges, H. G. Jackson and R. A. Saleh, Analysis and Design of Digital Integrated Circuits in Deep Submicron Technology, 3rd Ed, McGraw-Hill, 2004.

[4] J. Chia, "Design, Layout and Placement of On-Chip Decoupling Capacitors in IP Blocks", M.A.Sc Thesis, University of British Columbia, 2004.