Chapter Three

Power estimation methodology

3.1 Introduction:

This chapter gives a description of the power dissipation on the NoC system and introduced the model used in the simulation.

3.2 Power Dissipation Limitation:

Power is an important architectural design constraint that is beginning to dominate MPSoC design, particularly for mobile applications such as cellular phones, MP3 players, and laptops. These portable devices run on batteries that have a limited energy budget between charges. MPSoC designs intended for use in such portable scenarios must have lower power consumption, to improve user experience. Reducing power consumption is also a priority for non-portable MPSoC applications, such as those used in server farms that tend to consume significant amounts of power (e.g., as much as 2 megawatts for a 25,000 square foot server farm with 8000 servers [23]). According to Semiconductor Industry Association SIA projections of future silicon technologies, the operating frequency and transistor density of MPSoCs will continue to increase, making power dissipation for these highly integrated and complex designs a major concern.

Thus, with scaling trends in emerging technologies, and the increasing proliferation of the Internet and mobile computing, the power problem has assumed a critical status that cannot be ignored by MPSoC designers. On-chip communication architectures have a considerable impact on MPSoC power consumption. There are several reasons why the interconnect fabric is receiving so much attention with respect to power consumption [24, 25]. First, unlike transistors, interconnects have not scaled exponentially in DSM technologies, as a result of which interconnect capacitance forms a larger portion of total chip capacitance [26]. Second, the problem of modeling DSM effects could be largely ignored in pre-DSM technologies, where transistors were the main focus due to their large sizes. However, in DSM technologies, effects such as coupling capacitance between adjacent wires become increasingly dominant [27, 28]. Third, interconnects in today's designs are proportionally longer, which implies that interconnect delay has increased. Fourth, the use of a large number of repeaters and vias to reduce wire delay almost doubles power consumption in interconnects [29]. Finally, state of the art communication architectures consist not only of bus wires, but also significant amounts of hardware logic (e.g., bridges, arbiters, decoders, buffers, etc.) that is comparable to the amount of logic in embedded processors of moderate complexity [30]. It has been predicted that communication architectures will consume a larger portion of on-

chip power in future technologies [31]. There is, therefore, a need to create models for estimating power consumption of on-chip communication architectures as early as possible in a design flow, to better design and optimize MPSoCs. Additionally, the thermal effects of power dissipation cannot be ignored, since they are beginning to have a significant impact on the power, performance, design, and reliability of on-chip buses [32].

3.3 Thesis Overview and Contribution:

This thesis focused mainly in estimating the energy consumed in on-chip communication system by making a system level analysis depending on the theory proposed by K.Lahiri and Raghunathan[29], the power estimation model is introduced first in R.kar[20] and Sudeep Pasricha and Nikil Dutt[34] there was some modification done to insert high frequency effect as an RLC effect, there was a focus in the wire bus model by inserting system activity response by proposing two bus model, high frequency model and high level power model with high frequency effect, the first model is presumed to be the normal activity model the second model is switched to in high activity situation, the system activity is checked by calculating the consumed power in the Master device by calculating E_{INP} which in high system activity will be influenced by the values of the control signals asserted when the register is loaded, and when the values are driven to the slave, taken that E_{INP} average value will be the transient value which will decides system activity(AVGEinp=13mj), as shown in the flowchart in Figure 3.1.

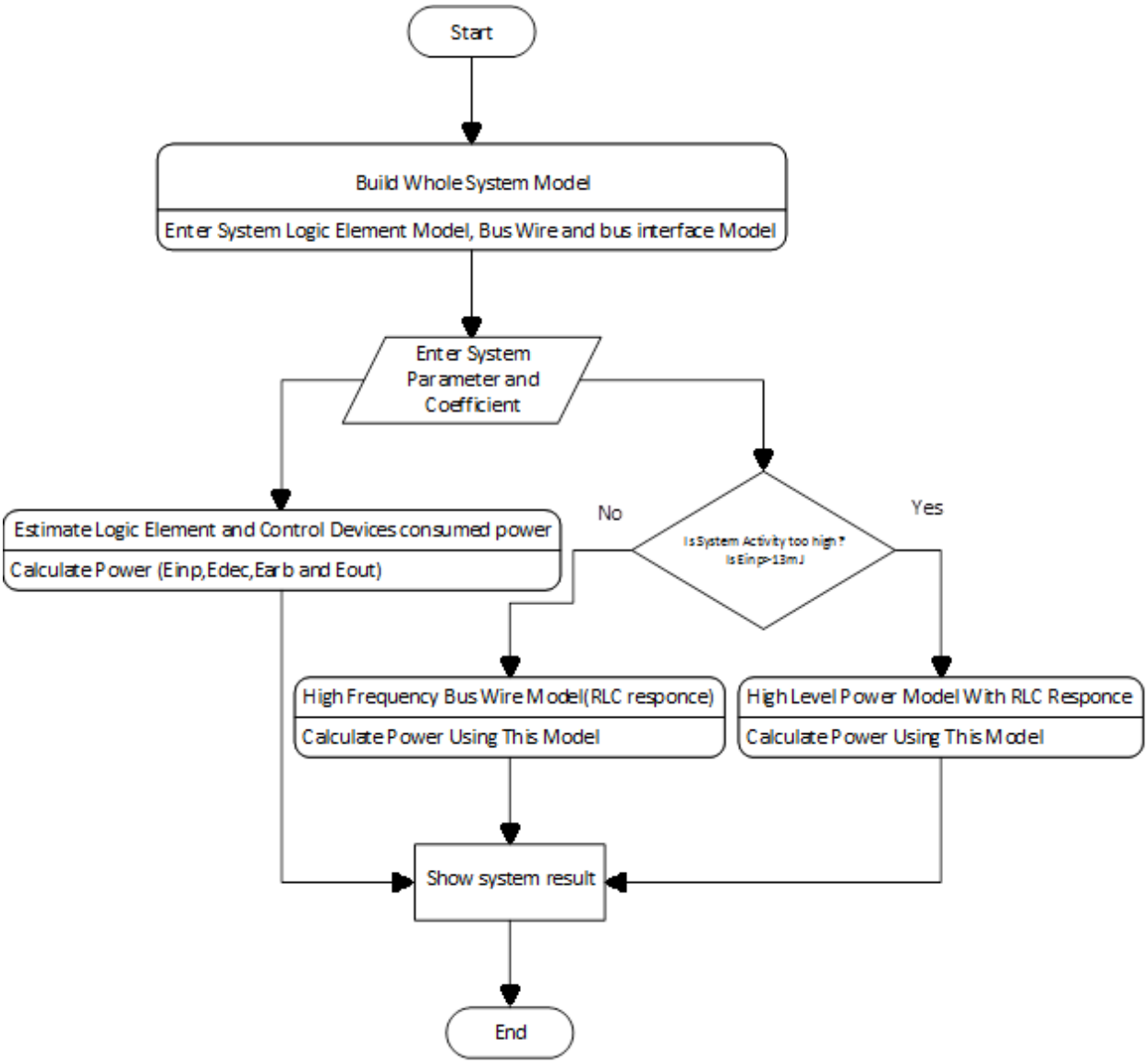

Figure 3.1: system flowchart

3.4 On-Chip Communication Model:

In this thesis a simplified implementation of the AMBA on chip bus is used, Figure 3.2 shows the proposed on chip system, its only consist of the Advanced High Performance Bus (AHB) which is a pipelined bus meaning that the address and the data for different transaction may overlap on time, All the slaves are memory mapped. For each transfer, the decoder generates slave select signals to contact the correct slave. Multiplexers properly route address, write data, and control parameters from the masters to the slaves, as well as slave responses and read data, from the slaves back to the masters. The arbiter regulates access to the shared bus using a configurable

arbitration scheme. Burst transactions enable a master to perform a sequence of transfers without requiring arbitration for each transfer.

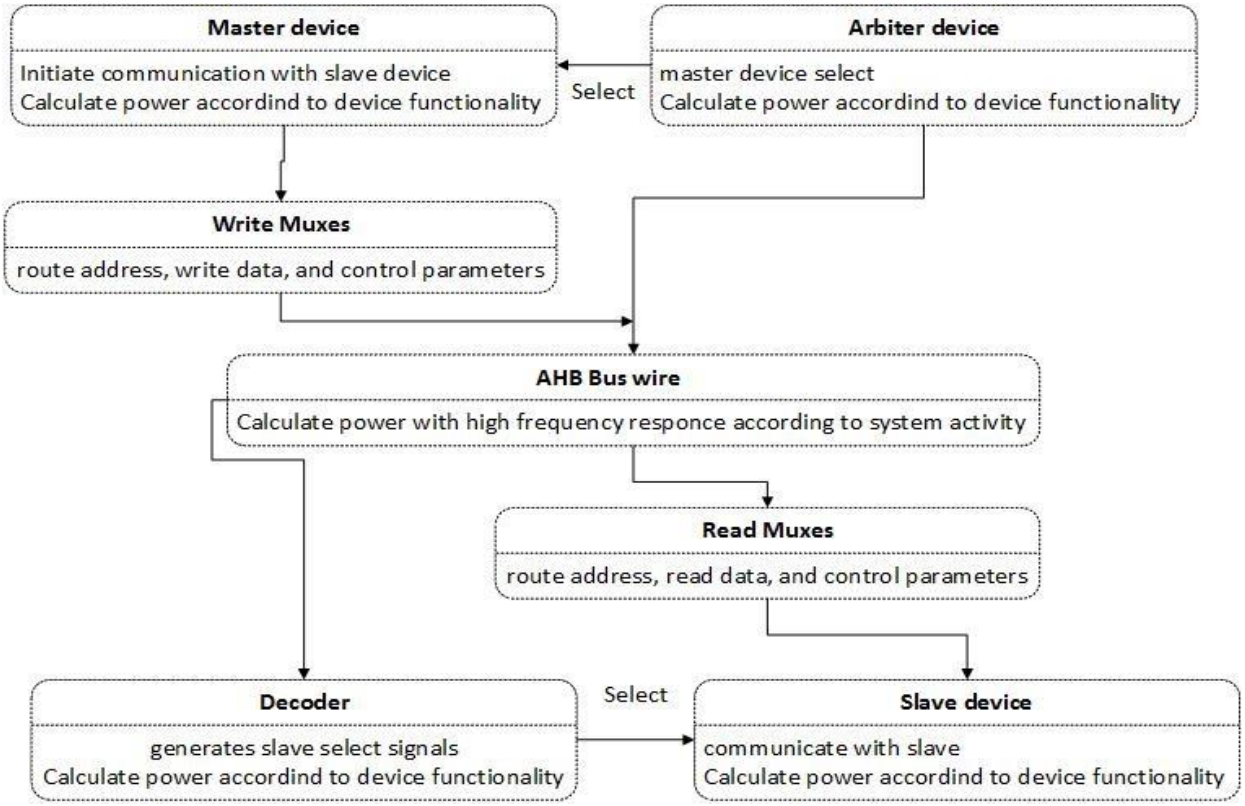

Figure 3.2: proposed model for power estimation

3.5 Power Dissipation in on Chip Communication Model:

To obtain the Energy consumption for the AMBA AHB bus matrix communication architecture shown in Figure 3.1 a macro model is used to give the total energy consumption of a bus matrix can be expressed as:

$$E_{MATRIX} = E_{INP} + E_{DEC} + E_{ARB} + E_{OUT} + E_{WIRE} \quad (3.1)$$

Where E_{INP} and E_{DEC} are the energy for the input and decoder components for all the masters connected to the matrix, E_{ARB} and E_{OUT} are the energy for arbiters and output stages connecting slaves to the matrix, and E_{WIRE} is the energy of all the bus wires that connect the masters

and slaves. Energy macro-models were created for the first four components, with E_{WIRE} being calculated separately.

3.5.1 Energy Consumed in the Input Stage:

Every master connected to a bus matrix has its own input stage that buffers address and control bits for a transaction, if a slave is busy. The input stage model can be expressed as:

$$\boldsymbol{E_{INP} = \alpha_{inp0} + \alpha_{inp1}.\psi_{load} + \alpha_{inp2}.\psi_{desel} + \alpha_{inp3}.\psi_{HDin} + \alpha_{inp4}.\psi_{drive}} \quad (3.2)$$

Where $\boldsymbol{\psi_{load}}$ load and $\boldsymbol{\psi_{drive}}$ are control signals asserted when the register is loaded, and when the values are driven to the slave, respectively; $\boldsymbol{\psi_{desel}}$ is the control signal from the master to deselect the input stage when no transactions are being issued; and $\boldsymbol{\psi_{HDin}}$ is the Hamming distance of the address and control inputs to the register.

3.5.2 Energy Consumed in the Decoder:

A decoder component is connected to every master, and consists of logic to generate the select signal for a slave after decoding the destination address of an issued transaction. It also handles multiplexing of read data and response signals from slaves. The decoder energy consumption model can be formulated as:

$$\boldsymbol{E_{DEC} = \alpha_{dec0} + \alpha_{dec1}.\psi_{slavesel} + \alpha_{dec2}.\psi_{respsel} + \alpha_{dec3}.\psi_{HDin} + \alpha_{dec4}.\psi_{sel}} \quad (3.3)$$

Where $\boldsymbol{\psi_{slavesel}}$ and $\boldsymbol{\psi_{respsel}}$ are control signals asserted in the cycle in which the slave select and the data/response MUX select signals are generated, respectively; $\boldsymbol{\psi_{HDin}}$ is the Hamming distance of the read data and response signals from the slave; and $\boldsymbol{\psi_{sel}}$ is a control signal where transition occurs when the decoder is selected or deselected.

3.5.3 Energy Consumed in the Output Stage:

Every slave is connected to the bus matrix through the output stage that handles multiplexing of address and control bits from the masters. It also calls the arbiter to determine

when to switch between accessing masters. The energy consumption for the output stage is given by:

$$\boldsymbol{E_{OUT} = \alpha_{out0} + \alpha_{out1}.\psi_{addrsel} + \alpha_{out2}.\psi_{datasel} + \alpha_{out3}.\psi_{HDin} + \alpha_{out4}.\psi_{noport}} \quad (3.4)$$

Where $\boldsymbol{\psi_{addrsel}}$ and $\boldsymbol{\psi_{datasel}}$ are control signals asserted when address and data values are selected after a call to the arbiter results in a change in the master accessing the slave; $\boldsymbol{\psi_{HDin}}$ is the Hamming distance of address and data inputs; and $\boldsymbol{\psi_{noport}}$ is a control signal from the arbiter, which goes high when no masters access the slave in a cycle.

3.5.4 Energy Consumed in the Arbiter:

The arbiter is invoked by the output stage, and uses an arbitration scheme to grant access to one of the potentially several masters requesting for access to the slave. The cycle energy model for the arbiter is calculated as:

$$E_{ARB} = \alpha_{arb0} + (\alpha_{arb1} + n.\alpha_{arb2}).\psi_{arb} + \alpha_{arb3}.\psi_{arb+1} + (\alpha_{arb4} + n.\alpha_{arb5}).\psi_{desel} + \alpha_{arb6}.\psi_{desel+1} \quad (3.5)$$

Where ψ_{arb} and ψ_{arb+1} are control signals representing the cycle when arbitration occurs, and the subsequent cycle when the master select signal is generated; ψ_{desel} and $\psi_{desel+1}$ are control signals representing the cycle when the arbiter is not selected by any master, and the subsequent cycle when it generates the no port signal for the output stage; and n represents the number of masters connected to the arbiter.

3.5.5 Energy Consumed in the Wire (High Frequency Model):

The wire will be modelled as an RLC network, signal propagation delay is a function of length were.

$$E_{bus} = r^2c^2 \times \frac{c(2l-cr^2)}{(r^2+\sqrt{c(2l-cr^2)})}.e^{-x\frac{r(\sqrt{c(2l-cr^2)}+\sqrt{cr}}{l-cr^2}} \quad (3.6)$$

Where r is resistance per length, l represents inductance per length and c is capacitance per length x is the bus length.

3.5.6 Total Power Consumed in Bus Wire (High Level Power Model):

The total power consumption of a bus in this model is given by equation below [33]:

$$P_{total} = P_{sw} + P_{vias} + P_{repeater} \quad (3.7)$$

Where P_{sw} is the power consumption due to switch interconnect capacitance and inter-wire coupling, P_{vias} is the power consumed by the vias due to the use of multiple metal layers, and $P_{repeaters}$ is the power consumed by repeaters inserted to minimize signal delay.

3.5.6.1 Switching Power (P_{sw}):

The model makes use of a table first presented by Taylor et al [33], where total switching power is determined by the types of transitions, instead of the number of transitions that can occur on the interconnect. Since coupling effects between wires decrease sharply the further apart they are, only interconnect and its adjacent wires need to be considered. Table 3.1 shows the set of various transitions that are possible on three-wire interconnects.

Table 3.1: A three wire lookup for 1mm, 0.18µm Cu interconnect.

Trans.	Energy(pJ)	Trans.	Energy(pJ)
S=no transition, t=transition, t'=opposing transition			
S S S	0.0	t t t	0.002
S t S	0.156	S t t	0.045
S S t	0.033	t t t'	0.153
t S t	0.131	S t t'	0.327
t S t'	0.0	t' t t'	0.560

3.5.6.2 Power Due to VIAS:

The purpose of vias is to (i) connect transistors residing on the substrate with the interconnect connecting these transistors, and (ii) connect interconnects running on multiple metal layers. The total power consumed by the vias is given by:

$$P_{vias} = V_N . P_{via} \quad (3.8)$$

Where V_N is the number of vias and P_{via} is the power consumption of a single via.

3.5.6.3 Power Due to Repeater:

With shrinking feature size, interconnect wires are getting proportionally longer and not scaling as well as transistors. A wire can be modeled as a simple RLC network, for which the signal propagation delay is a quadratic function of length, because both resistance, inductance and capacitance are functions of length. Inserting repeaters is a commonly used practice to reduce wire delay, since it reduces the quadratic dependence into a linear dependence on wire length, repeater power is given by:

$$P_{rep} = Z_{rep} . V_{dd}^2 . f . \sum_{i \in I} \rho_i N_{Ri} + P_{via} . \sum_{i \in I} V_{Ri} \quad (3.9)$$

Where Z_{rep} is given by equation (3.10), V_{dd} is the operating voltage, f is the clock frequency, ρ_i is the switching activity, N_R is the number of repeaters, P_{via} was discussed earlier and V_R the total number of repeater vias which are calculated to be twice the number of repeaters.

$$Z_{rep} = h.k.Z_o \quad (3.10)$$

$$h = \sqrt{\frac{R_0 . X_{int}}{X_0 . R_{int}}} \quad (3.11)$$

$$K = \sqrt{\frac{0.4(R_{int} . X_{int})}{0.7(R_0 . X_0)}} \quad (3.12)$$

$Z_0 = \sqrt{\frac{l}{C}}$ for lossless wire (3.13)

Where Z_0 is the impedance of the minimum sized inverter, h is the optimal size repeater for minimum wire delay, k number of repeaters for minimum wire delay, R_{int} is the resistance of the wire and R_0 is the resistance of the minimum sized repeater (or inverter), x_{int} is the acceptance of the wire, l is the inductance per wire length and is the capacitance per wire length.

3.6 Chapter Summary:

In this chapter the power significance in the NoC is discussed, then a system model used in this thesis is presented with explaining the system analysis made here to reach an accurate power estimation and describing the contribution made in the thesis.