CHAPTER 2

NOC Architecture: Literature Survey

Recent Systems on chips (SoCs) designs are integrated by IP (Intellectual Property) cores which are useful in designing reconfigurable architectures. A rapid evolution of interconnects is experienced in this century which have catered the compelling need for higher bandwidth at ultra-low power. In this Chapter, the background of NoC is discussed. Followed by the background section, the work done by several researchers are reviewed.

2.1. Background

Network on Chip uses layered communication system of OSI model like general data networks [73]. The OSI model consists of seven layers i.e. Physical layer, Data link layer, Network layer, Transport layer, Session layer, Presentation layer and Application layer. Each layer consists of some software and hardware components to perform certain assigned functionality. Each layer performs a task independently. The layers provide various services to their upper layer and acquire service from their bottom layer. Description of the functionalities of layers (in bottom to top order) of OSI model is given below. Physical layer provides hardware support and is responsible for sending and receiving data on a carrier. In Data link layer, encoding and decoding of data packet into bits is done. Routing and forwarding of data is the main function of this layer. Transport layer ensures data transfer from source to destination. Session layer sets up co-ordinates and terminates conversations, dialogues between the applications. Presentation layer
transforms data into a form that the application layer can accept i.e. encryption of data from application to network format and vice versa. Application layer supports applications and end user processes.

When a layer is performing a task, it is hidden from other layers. Each layer follows some certain set of rules which are called as protocols and the layers communicate with each other by some kind of bridges i.e. interfaces. There are many advantages of the layered communication system and it has some unavoidable overheads. The most important layers for a NoC are listed below. They are i) Physical Layer, ii) Data link layer and iii) Network layer. The important functionality of physical layer is to provide clock signals to every connection and to generate control to Data Link layer. The important functionalities of data link layer are flitization, de-flitization, error detection and correction. The functionalities of network layer are data packetization, routing, buffering, congestion detection and control. QoS (Quality of Service) is provided by this layer by improving latency, throughput and jitter of the network. In this thesis, a sole importance is given to network layer and its functionalities. Some network communication related terms are described below.

**Message**

Message is the information or data which are transmitted from source to destination resource. It is defined in application layer. A message can be of fixed length or can be of variable length according to the requirement. These messages that travels in the networks in various forms are described below.
Packet

In packetization process, message is divided into certain number of packets. Packets of the same message are independent of each other. Each packet has enough information to travel throughout the network. Generally, a packet has three parts namely i) Header ii) Payload or Body iii) Trailer. Header contains controlling and routing information such as source and destination address. Sometimes it contains the whole route for data transmission. Payload contain actual data or information. Trailer indicates the end of packet.

Flit

A packet can be divided into further smaller elements. Those small elements are called flits. They are flow control digits. Flits also have three parts i.e. Header flit, Body flit and Tail flit. The size of the flit is always fixed. Due to flits, the storage devices required in the switches or routers will be very small.

Phit

A flit can be divided into small units known as phits. They are physical transfer digits. Phits are travelled across a channel among the network
switches but as one unit. It can be considered as link width phit which can be 
considered as the measuring parameter of link width. It indicates the number 
of wires needed for data transfer between network routers. Size of a phit can 
be same as the size of flit or maybe not.

2.2. Case Study

Following sections describe some of the practical NoC structure.

(A) Argo

Argo is a globally asynchronous, locally synchronous architecture for a 
general-purpose multiprocessor platform [74]. The platform has one NoC 
which offers access to a shared memory and one NoC supporting message 
passing, each of them optimized for its purpose of use.

The Argo multiprocessor platform for 2D mesh topology is shown in 
Figure 2.2. It is made up of two different components namely Network 
Interfaces (NI) and routers. The NI converts the transaction based 
communication from the processor core to stream based communication 
towards other processor cores in the network. The router is routing the stream 
of packets that are injected from the NIs through the network according to the 
static TDM schedule. The routers are connected in a 2D bi-torus structure. The 
communication that goes through the network is controlled by the TDM 
schedule. A Direct Memory Access (DMA) placed in the NI controls the 
communication from that NI, enforcing the TDM schedule. A DMA controller is 
available for every communication channel.
The TDM schedule divides time into slots and during one time-slot one communication channel, i.e. one DMA, is active. The NI of Argo has two Open Core Protocol interfaces. The first is a 32-bit interface, connected to the processor core for configuring TDM schedule and DMA controllers. The second is a 64-bit interface, connected to the local memory of the processor, i.e. ScratchPad Memory (SPM) for accessing the data.

![Figure. 2.2 ARGO Mesh architecture](image)

The DMA and the SPM of each processor are mapped in the local address space and can be accessed by the processor through specific load and store instructions. The DMA is setup by the processor core to initiate a message transmission, creating a direct connection from the local SPM of one processor to the local SPM of a remote processor.

The Argo TDM router uses source routing, and in combination with the TDM scheme, which requires no flow control or buffering, the routers become very simple and efficient. The micro architecture of the Argo asynchronous router is shown in Figure 2.3.
Argo supports GALS organization with independently clocked processor cores and mesochronous NIs. The NI clock is the time base for the TDM scheduling. The asynchronous routers provide the timing elasticity necessary to cover for skew among the mesochronous NIs. The Argo TDM router is a three-stage pipeline. The pipeline stages use two-phase bundled-data handshake-latches [75] that are implemented using conventional enable latches and a Mousetrap controller [76]. The message-passing NoC in Argo platform uses statically scheduled time-division multiplexing (TDM) to implement end-to-end circuits and support time predictability. The primary reason for choosing TDM over alternative approaches for time predictability are: 1) simplicity of calculating communication latency and 2) simplicity of the hardware implementation of the routers—pipelined crossbars (Xbars) without any circuitry for arbitration or buffering. Intuitively this simplicity may result in disadvantages as well, for example: 1) increased complexity of
the NIs because they must now include schedule tables and data buffers and 2) increased communication latency.

The test case is set up to transmit four packets of random data from every node to every other node based on an all-to-all schedule. This setup results in average link utilization of 23% as measured during simulation. The experiment shows that the majority of the energy is consumed by the NIs, with a share of roughly 47%, and by the link driving buffers contributing approximately 40%. The routers contribute only approximately 10% of the energy, a fact attributed to the gating used in the routers.

(B) MANGO

The MANGO network (Message-passing Asynchronous Network-on-chip providing Guaranteed services over OCP interfaces), developed at the Technical University of Denmark, is a clock-less NoC, targeted for coarse-grained GALSType SoC [77].

MANGO NoC consists of Network Adapters (NA), routers and links. Each IP core is connected to the network through an NA, providing high level communication services, i.e. OCP transactions, on the basis of primitive services implemented by the network. Each NA, which also performs the synchronization between the clocked IP core and the clock less network, is connected to a router.

The routers are connected by links in a grid type structure, either homogeneous or heterogeneous. To keep speed up, long links can be implemented as pipelines. Figure 2.4 shows a conceptual picture of the MANGO
router. The router implements a number of unidirectional ports. Two of these are local ports which connect to the NA. The local ports consist of a number of physical interfaces.

![Figure 2.4 Block diagram of MANGO Architecture](image)

The remaining ports are network ports which, via point-to-point links, connect the router to neighboring routers. Each of these ports implements a number of independently buffered VCs.

Network ports and local ports implement the same type of interface. It is the function of the NA to translate communication to and from network packet format. Internally, the router consists of a Best Effort (BE) router, a Guaranteed Service (GS) router, output buffers, and link arbiters. The BE and the GS router are separately implemented. The BE router dynamically source-routes connection-less data packets, according to the routing path defined in the packet header. A subset of the VCs are allocated for BE routing. The GS
router uses the remaining VCs to route header-less data streams on statically programmable connections. In MANGO, a connection implements a logical point-to-point circuit between two different local ports in the network by reserving a sequence of independently buffered VCs. The GS router provides non-blocking switching between the input ports and the output buffers, therefore GS can be realized purely on the basis of link access arbitration.

The switching module makes it possible for all input ports to route flits to any combination of VC buffers. It is made entirely without any form of arbitration, making its performance highly predictable, simplifying its design, saving area and routing latency. Since a VC buffer is part of a connection, no congestion will occur as only one input will attempt to route to one VC buffer at any given time. The switching module, which constitutes a considerable part of the total router area, scales linearly with the number of VCs, and thus with the number of connections supported.

The MANGO NoC uses static paths for virtual circuits, constraining them when a throughput limit is reached, enforcing a throughput rate over a time interval. The downside of the design is the high hardware complexity caused by the virtual channel buffers and a larger Xbar. The large Xbar comes from the use of virtual channel buffers and from the arbitration and flow control needed in every output port. It is to be noted that a typical MANGO router is 10 times larger than an aelite router[77].

The performance in net-list simulations using worst-case timing parameters (1.08 V/125 C) is 515 MHz per port (795 MHz under typical timing
conditions). It internally uses a bundled-data circuit style and the links use a DI two-phase encoding. The router latency is estimated as 5.2ns. For the switch, VC Buffers/Control and VC merge, latency is estimated as 2.1ns, 1.2ns and 1.6ns respectively. MANGO provides hard latency/throughput guarantees unlike other VC prioritization based schemes.

(C) WiNoC

One of the possible ways to overcome the communication overhead incurred by the multi-hop channels in wired NoC is to adopt wireless communication fabric. Surface wave communication has been recently demonstrated as a feasible on-chip wireless solution [78-79]. Here, wireless communication layer of WiNoCs is replaced with a waveguide medium as the surface wave communication fabric which generates a NoC architecture as shown in Figure 2.5.

---

Figure 2.5 Block diagram of WiNoC Architecture
To cater wireless communication, the routers in WiNoCs must be equipped with a wireless transmission interface which serves as a bridge between the wireless and the wired communication layers. The wireless transmission interface works closely with the routing logic, virtual channel allocator, arbiter and crossbar switch for efficient wireless signal transmission. The wireless transmission interface is equipped with a retransmission buffer and a suitable error encoding and decoding scheme.

Nodes with both wireless transmission and reception capabilities possess a CMOS-based circulator which acts as a communication bridge between the transmitter, receiver and the 2-D waveguide medium [80]. The total power consumption of the transceiver is 36.7mW. A zigzag antenna is employed.

But the overhead of the erroneous packets as well as the retransmission process introduces contention on the wireless layers, and have drastic effects on the performance of the WiNoC. Moreover, buffer spaces contribute significantly to the total power consumption of the NoC [81] and should be used judiciously.

(D) Ref-2-D Mesh Architecture with cluster cores (Muhammad Ilyas & Saad Ahmed)

WNoC is a on-chip communication architecture based on Radio Frequency (RF) interconnection[82]. Recent advances in IC technology make it possible to integrate low-cost transceiver antenna onto a single chip which is termed as Radio-on-Chip technology. The WNoC consists of two basic
components viz., Transparent Network Interface (TNI) and Radio Frequency (RF) node. A RF node is a radio-frequency interface which is meant for bidirectional communication among IP cores. A number of RF nodes, equipped with low-cost, low-power transceivers and tiny antenna, are distributed on-chip to form a multi-hop wireless micro network. In a WNoC, the RF nodes are properly distributed owing to IP core placement, non-uniform core size, and different data transportation needs. Such flexible network architecture allows that several IP cores share one RF node to reduce power and area overhead. Therefore these IP cores are grouped into clusters. The data are transmitted in packets, each of which includes the destination address and the data payload.

The IP cores access the network via the TNI, and their packets are delivered to the destination through multiple hops across the network. This architecture will enable higher bandwidth, higher flexibility and reconfigurable interaction.

The IP cores are clustered on the chip in order to minimize the routing cost and to stabilize the communication workload consequently. These IP cores within the cluster are hardwired to RF node via TNI and are share it for data communication. All RF nodes are of limited range and they are connected to each other via adjacent RF nodes by wires. These RF nodes are connected through both wired and wireless medium. The wired medium is used for control logic and wireless medium is used for data transmission. The control logic through wires is for the purpose to access the channel. Each RF node has a set of ‘n’ control lines (Rx/Tx) connected to its n neighbors. Each pair of
control lines consists of a single bit input (Rx[i])/output (Tx[i]) line for
handshaking between a RF node and its i\textsuperscript{th} neighbor. Based on this
interconnection infrastructure, the multiple access controlling is performed on
these control wires and data are transmitted through network wirelessly.

(E) LHNoC

In conventional on-chip network, the on-chip routers are connected
through metal wires and the data are transferred in a multi-hop
communication manner. This network causes high latency and high power
consumption. To tackle the drawbacks in wired network, interconnect
innovation with optical, radio frequency combined with accelerated efforts in
design and packaging have been proposed [83-84]. With the advancements in
nanotechnology and wireless communication, on-chip wireless
communication has come into limelight and can achieve hundreds of GHz to
tens of THz of bandwidth. With these improvements, several hybrid on-chip
network architectures have been proposed. In hybrid networks, high
bandwidth and single hop communication is feasible. It offers high flexibility,
high bandwidth, low power, low energy and reduced interconnection delay.
The hybrid wired/wireless on-chip network is a revolutionary on-chip
communication infrastructure which can offer the benefits of both wired and
wireless connection in an optimized manner [85]. The key problem in hybrid
networks is the placement of the wireless links between a particular pair of
source and destination cores, which will affect the performance gain. Also, the
wireless link is a lossy channel compared to wired link. So, in order to
guarantee the reliable data transmission, a retransmission scheme is required in hybrid on-chip network. A communication model to optimize the placement of wireless links has been proposed to resolve the problems in link allocation.

![Hybrid Network on Chip](image)

Figure 2.6. Hybrid Network on Chip

In this on-chip network, there are two kinds of communication channels, conventional wired links and long-distance high-bandwidth wireless links. The neighboring IP cores are connected by the wired links and the wireless links are used to connect the distant cores.

![Microarchitecture of Hybrid on-chip router](image)

Figure 2.7 Microarchitecture of Hybrid on-chip router
These on-chip wireless links enable on-hop data transfer between distance cores, which in turn reduce the multi-hop long-distance wired communications. Figure 2.6 shows the mesh-based hybrid on-chip network. Basically, the cores are interconnected by a 2-D mesh wired on-chip network. And there are two long-distance wireless links: one connects Router 3 and 5, the other connects Router 8 and 14.

In this hybrid on-chip network, there are two routers namely basic router and hybrid router. The basic router has the same micro architecture as conventional on-chip router. Hybrid router combines basic router and a wireless communication unit. Figure 2.7 shows the micro architecture of a hybrid router in 2-D Mesh based hybrid on-chip network. It has five bidirectional ports, a crossbar switch, control logic (VC allocator and switch arbiter) and a wireless communication unit.

Figure 2.8. Block diagram of the wireless communication unit

The Wireless Communication Unit (WCU) shown in Figure 2.8 has a modulator, demodulator, checksum generation logic, checksum verification
logic, input buffer and retransmission buffer. For wireless channel allocation, Frequency Division Modelling technology is being used. The WCU provides a half-duplex wireless channel. Data loss has been reduced by the use of error deduction codes that are appended into the data flits. During the communication process, the WCU of sender accepts data flits from crossbar switch and after checksum generation and modulation, data are transmitted through the antenna. At the same time, the sender stores the transmitted data flits into retransmission buffer. And the WCU of receiver picks up the signal using the on-chip antenna. It is then demodulated and the checksum is verified. If the checksum is right, the flit enters the input buffer and then routes to the output port. Otherwise, an NACK signal will be sent back to the sender. When the sender receives an NACK signal, it will retransmit the corresponding flit from the retransmission buffer.

The key point in routing mechanism to be employed in hybrid architecture lies in making decision of using the wireless link. Comparing the transmission delay with and without wireless link, the decision is made. This method is applicable to any topology and any traditional routing. The proposed hybrid on chip network employs FDM like wireless technology at sub-THz or THz frequency with the area overhead of one wireless link around 200μm². The characteristics of the wireless link allocation are analyzed. The energy consumption and average flit delay for a 2-D mesh is found as 1.226 pJ/bit and 31.62 cycles. The simulation results show that the hybrid on-chip network improves the performance in terms of delay and energy consumption.
significantly. But the architecture proposed is computationally complex with increased area overhead.

(F) WrHNoC

The main drawback of the existing NoC Architecture is its high latency and power consumption due to multi-hop long-distance communication among Processing Elements (PE)[85]. This limitation is overcome by implementing a wireless on-chip communication interface between the PE as shown in Figure 2.9. Wireless links are inserted between subnets to form express communication links by replacing baseline wired routers with routers having wireless communication capabilities. Optimal locations for wireless routers (WRs) are calculated to minimize the average traversal distance.

![Figure 2.9 Hybrid NoC Architecture](image)

The architecture supports both wired and wireless connection for transferring the packets from source router to destination router. If the source and destination router is on same subnet, then the packets are routed using the
wired connection. If the source and destination are in different subnets, then source router routes the packets to the wireless router in subnet 1 using wired connection and wireless router in subnet 1 routes the packets to the wireless router in subnet 2 using wireless connection. Next, the wireless router in subnet 2 routes the packets to the destination router in subnet 2 using wired connection. The simulations are carried out in mesh topology with single-cycle channels. The hybrid architecture has exhibited a reduction in average power consumption by 11.1% in shuffle traffic mode, 12% in bit comp traffic mode, 10.7% in transpose traffic mode and 10.52% in bitrev traffic mode. It also reduces the latency over the conventional wired NoC router by 30% in shuffle traffic mode, 11.25% in bitcomp traffic mode, 12.85% in transpose traffic mode and 13.3% in bitrev traffic mode.

2.3 Literature Review

For communication between the cores in System on Chip environment, myriad of methodologies have been proposed. Of which bus architecture is the simple and most common method in use. To overcome some of its limitations including bandwidth, wiring delay, speed and power, many advanced architectures like AMBA [86], Open Core [87], WISHBONE SoC interconnection [88] have been developed. Network on Chip have been adopted to mitigate wiring delay existing in previous techniques. NoC replaces SoC in terms of manufacturing costs, power, performance and speed [89-93]. SoC manufacturers such as ST Microelectronics, Samsung, Philips and also Universities such as Bologna University, M.I.T., Berkeley and more are all
proposing proprietary frameworks based on NoC interconnects. These frameworks help engineers in the switch of design methodology and speed up the development of new NoC based systems on chip. The research has been confined towards on chip communication to enhance all the performance metrics concerned with the Network on Chip architectures. A decade of continuing research has been going on in the performance enhancement of Network on Chip. The requirements of an efficient NoC include low power, reduced area and high speed.

The International Technology Roadmap for Semiconductors (ITRS) predicts that the future generations of high-end VLSI designs will operate in 10-20 GHz range with the communication between cores in Gbit/s. This requires designers to work within a tight power budget.

In microchip technology, the resources available in a single chip have doubled every second year which imposes an exponential trend in force. Physical limitations such as time-of-flight of electrical signals, power use in driving long wires/cables, etc. are associated with the scaling of microchip technologies. Also, single chip system has experienced communication issues and on-chip wires are expensive in terms of power and speed. In the year 1997 to 2002, there was a drastic decrease in the geometry which increased the design complexity by nearly 50 times. To ensure effective advantage of technology scaling, an effective use of network interconnects, routers and the routing algorithm are mandatory. The mainstay of advancement lies in maintaining a trade-off between flexibility, performance and hardware cost.
Chip design has four distinct aspects: computation, memory, communication and I/O. As processing power has increased and data intensive applications have emerged, the challenge of the communication aspect in single-chip systems, System on Chip (SoC), has attracted increasing attention. The major driving factors for the development of communication schemes are the ever-increasing density of on-chip resources and the drive to utilize these resources with a minimum of effort as well as the need to counteract the physical effects of deep sub-micron technologies. The trend is towards a subdivision of processing resources into manageable pieces. This helps reduce design cycle time since the entire chip design process can be divided into minimally interdependent sub problems. This also allows the use of modular verification methodologies, that is, verification at a low abstraction level of cores (and communication network) individually and at a high abstraction level of the system as a whole. Working at a high abstraction level allows a great degree of freedom from lower level issues. It also tends towards a differentiation of local and global communication. As inter core communication is becoming the performance bottleneck in many multicore applications, the shift in design focus is from a traditional processing-centric to a communication-centric one. One top-level aspect of this involves the possibility to save on global communication resources at the application level by introducing communication aware optimization algorithms in compilers [94]. System-level effects of technology scaling are further discussed by Catthoor [95]. A standardized global communication scheme, together with standard communication sockets for IP cores, would make Lego brick-like
plug-and-play design styles possible, allowing good use of the available resources and fast product design cycles.

From the advent of SoC in 90s, the communication structures have been under research which is characterised as custom designed ad hoc mixes of buses and point-to-point links [96]. The first model named as bus is simple and easy to model. But in a dense interconnected multi core system, it ends with communication bottleneck. As the number of units gets increased, the power utilized per communication event also upshots owing to the capacitive load caused by much more attached units. Followed by bus, cross bar overcomes some of the limitations of bus. But it has come up with intermediate solution and it is not ultimately scalable. Dedicated point-to-point links are optimal in terms of bandwidth availability, latency, and power usage. They are simple to design and verify and easy to model. But the number of links needed increases exponentially as the number of cores increases. Thus, an area and possibly a routing problem develop. For a system with lesser number of cores, a best way of communication is point to point dedicated link. With increasing number of cores, a much more scalable and flexible communication structure is required. The term NoC is used in research today in a very broad sense ranging from gate level physical implementation, across system layout aspects and applications, to design methodologies and tools. A major reason for the widespread adaptation of network terminology lies in the readily available and widely accepted abstraction models for networked communication. The OSI model of layered network communication can easily be adapted for NoC usage as done
in Benini and Micheli [97] and Arteris [98]. NoC facilitates communication-centric design as opposed to traditional computation centric design.

Philips developed a NoC named, AETHERAL provided a satisfactory throughput [99-101]. The researchers of KTH Royal Institute of Technology, Stockholm identified the rising complexity associated with the dense VLSI technology and proposed the NOSTRUM NoC [93,102-104]. The Technical University of Denmark developed the MANGO network (Message-passing Asynchronous Network-on-chip providing Guaranteed services over OCP interfaces) which is targeted for coarse-grained GALS type SoC.

University of Manchester implemented CHAIN network which has credited from low power capability in asynchronous circuits [105]. Technical University of Denmark who developed MANGO network (Message-passing Asynchronous Network-on-chip providing Guaranteed services over OCP interfaces) [106] is a kind of clock less NoC. University of Bologna and Stanford University developed a light weight implementation named XPIPE NoC which possessed average hop delay, optimised link latency, area and power [107-108].

Hybrid NoC structures are proposed by a number of researchers in wherein, the traditional metal interconnect-based NoCs are supplemented with high speed interconnect channels between distant clusters of communication blocks. The high-speed interconnect channels can be any of the optical, CNT based and RF based interconnects or wired high-speed metal interconnects. It is shown in that the hybrid NoCs with the high-speed
interconnect channels implemented using wireless RF interconnects outperform the implementation using wired high-speed metal interconnects in terms of throughput, latency and power consumption. In comparison, the packet energy per bandwidth for a wireless RF interconnect based hybrid NoC is only 10.89 nJ/TBps—a substantial power saving. Despite being specific to the characterization of the systems (antenna size, frequency, substrate modeling etc.), the comparison in demonstrates the potential savings for the targeted wireless RF interconnects on a NoC system. This prompts for further research into the four design challenges. Quality of service aware hybrid wireless NoC protocols has been presented for a collision free communication which would be needed for multiple transmitting antennas with the same carrier frequency. A rapid evolution of interconnects is experienced in this century which have catered the compelling need for higher bandwidth at ultra-low power.

Power and speed considerations

M. Modarressi and H. Sarbazi-Azad et al. presented a reconfigurable NoC architecture which was formed by embedding programmable switches between routers of a mesh-based NoC [109]. The evaluation results exhibited a reduction of power consumption of up to 32% compared to a conventional mesh network. R. Parikh et al. have designed a power aware NoC which exhibited a reduction in total network power consumption by 14.5% on average, with only a 1.8% degradation in performance, when all processor nodes are active [110]. At times when 15-25% of the processor cores are
communication-idle, it has enabled a leakage power savings of 36.9% on average.

K. Lee et al. designed low power NoC technique for high performance SoC system [111]. The chip designed consumed 160 mW and the on-chip network dissipated less than 51 mW which ensured a reduction in power dissipation by 38%. F. Pakdaman et al. designed a NoC and compared it with EVC and baseline NoC [112]. The proposed NoC outperformed EVCs by 11% and the very efficient baseline NoC by 27%, on average. The energy consumption offered by the proposed method also outperformed the baseline by 21%and EVCs by 7%, on average.

M. V. Theertha et al. have proposed architecture to support multiple applications and have enhanced the performance and power consumption of NoC by 29% and 7% respectively [113]. An optimized tradeoff has been achieved between area and flexibility through the reconfiguration algorithm. A. Karkar et al. developed a novel approach to tackle various challenges experienced during on chip implementation [114]. This approach which encompassed mixed wire and Surface wave communication fabrics have achieved an improvement in power reduction and communication speed up to 63% and 12X, respectively.

P. Lotfi-kamran, M. Modarressi, and H. Sarbazi-azad, introduced CIMA – a hybrid circuit-switched and packet-switched mesh-based interconnection network and has documented an improvement in performance by 21% compared to other networks [115]. M. O. Agyeman et al. combined wired and
wireless channels [116]. The communication fabric proposed by them has improved maximum sustainable load of NoC by an average of 20.9% and 133.3% compared to existing WiNoCs and wireline NoCs, respectively.

At the meantime, changes in software without replacing the hardware also emerged. In the past, any change of the network protocol usually required replacing the hardware, since the protocol definition was closely coupled with the hardware in the form of specific ASICs designed for each protocol [117]. The high hardware replacement cost is unacceptable, and not able to catch up the frequent technology updating. Moreover, to satisfy different requirements of applications, application-specific network-on-chip design introduces additional task of designing network with different configurations and interconnection. These steps require significant design time and need to verify network components and their communications for every design [118]. Thus, Liu Cong et al. designed software defined on chip network which separate on-chip network into the control plane and data forwarding plane, so that control logic will be decoupled from the underlying chip hardware, and applications will be able to configure the network according to their requirements [119]. The simulation results have evaluated an improvement in network performance and power consumption with the aid of programmable control logic and application-specific configuration.

In future deep submicron designs, the interconnect effect will definitely dominate performance and Networks-on-Chip has become a promising solution to present communication infrastructure limitations. Usually, NoC
area dominates the target SoC devices and thus greatly influences power consumption and increases the clock distribution/skew problem.

Designing a NoC topology that is highly optimized for low power consumption and latency have been well addressed by several researchers. All the works are confined with pure wired or wireless architecture. But the present research concentrates on utilizing the benefits of both wired and wireless architecture.

Architectural power estimation is extremely important in order to

(1) verify that power budgets are approximately met by the different parts of the design and the entire design, and

(2) evaluate the effect of various high-level optimizations, which have been shown to have much more significant impact on power than low-level optimizations.

To tackle this problem, ORION, a set of architectural power models for network routers, was proposed in 2002, and have been widely used for early-stage NoC power estimation in literature and industry.

Energy Considerations

M. Modarressi et al. designed an arbitrary application-specific topology in a reconfigurable NoC which outperformed the conventional NoC with 22% energy reduction for the 1-hot flow and 9% reduction for the 3-hot flow traffic loads[120]. P. T. Wolkotte et al. designed a energy-efficient reconfigurable circuit-switched Network-on-Chip by physically separating the
concurrent data streams[121]. The proposed architecture consumed 3.5 times less energy compared to its packet switched equivalent.

Dominic DiTomaso[122] states that adaptive channel buffers (on-link storage) can considerably reduce power consumption and area overhead by reducing or replacing the power-hungry router buffers. However, channel buffer design can lead to head-of-line (HoL) blocking, which eventually reduces the throughput of the network. Here, they designed channel buffers and router crossbars to improve the performance (latency, throughput) while reducing the power consumption.

In addition, they implemented the proposed channel buffers and crossbar organizations in a Concentrated Torus (CTorus) topology which is a dual network without the additional area overhead. When the buffers are sized to 1x, bi-directional routing consumes 15% less energy at an increased cost of 63% in area compared to uni-directional routing.

Many researches have been done in NoC interconnection field. The following Table 2.1 summarizes the state of art of NoC proposals. It represents plethora of proposals developed by industries and academicians. Different prototypes along with their adopted topologies and switching methods are tabulated. It is observed the table, that mesh topology has been commonly used. It can be noticed that wormhole flow control are used in many architectures. Their evaluation metrics in terms of power, energy, speed, throughput and latency are also tabulated.
Table 2.1. Comparison of NoC Architectures

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Author Name</th>
<th>NoC Name</th>
<th>Ref.</th>
<th>Topology</th>
<th>Routing</th>
<th>Switching Technique</th>
<th>Flow Control</th>
<th>Area [mm²]</th>
<th>Power [mW]</th>
<th>Latency [ns]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Soteriou</td>
<td>Orion/</td>
<td>92</td>
<td>Mesh</td>
<td>Deterministic/Adaptive</td>
<td>Packet Switching</td>
<td>Wormhole</td>
<td>0.11</td>
<td>-</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Luna</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Bertozzi</td>
<td>Xpipes</td>
<td>10</td>
<td>Crossbar</td>
<td>Deterministic</td>
<td>Packet Switching</td>
<td>-</td>
<td>0.59</td>
<td>4.1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Kasapaki</td>
<td>Argo</td>
<td>47</td>
<td>Mesh</td>
<td>Deterministic</td>
<td>Packet Switching</td>
<td>Wormhole</td>
<td>0.75</td>
<td>3.56</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Andriah</td>
<td>SPIN</td>
<td>3</td>
<td>Ring</td>
<td>Adaptive</td>
<td>Packet Switching</td>
<td>Wormhole</td>
<td>0.29</td>
<td>7.6</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Bjerreg.</td>
<td>Mango</td>
<td>13</td>
<td>Mesh</td>
<td>Deterministic</td>
<td>Packet Switching</td>
<td>Wormhole</td>
<td>0.19</td>
<td>9.2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Kariniemi</td>
<td>XGFT</td>
<td>47</td>
<td>Fat-tree</td>
<td>Adaptive</td>
<td>Packet Switching</td>
<td>Wormhole</td>
<td>0.16</td>
<td>3.5</td>
<td>15.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Rijpkema</td>
<td>Aethereal</td>
<td>83</td>
<td>Mesh</td>
<td>Adaptive</td>
<td>Packet / Circuit</td>
<td>Wormhole</td>
<td>0.26</td>
<td>2.8</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Chang</td>
<td>Ref. pkt</td>
<td>18</td>
<td>Mesh</td>
<td>Deterministic</td>
<td>Packet Switching</td>
<td>Store and Forward</td>
<td>0.65</td>
<td>9.4</td>
<td>29.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mesh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Cidon</td>
<td>XHiNoC</td>
<td>21</td>
<td>Mesh</td>
<td>Adaptive</td>
<td>Packet Switching</td>
<td>Wormhole</td>
<td>0.10</td>
<td>12.5</td>
<td>26.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Moraes</td>
<td>Hermes</td>
<td>64</td>
<td>Mesh</td>
<td>Deterministic/Adaptive</td>
<td>Packet Switching</td>
<td>Wormhole</td>
<td>0.6</td>
<td>7.5</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Chang</td>
<td>Crossroad</td>
<td>18</td>
<td>Mesh</td>
<td>Deterministic</td>
<td>Circuit Switching</td>
<td>-</td>
<td>0.06</td>
<td>-</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Mullins</td>
<td>Lochside</td>
<td>66</td>
<td>Mesh</td>
<td>Deterministic</td>
<td>Packet Switching</td>
<td>Wormhole</td>
<td>0.5</td>
<td>5.3</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Salmine n</td>
<td>Ref. 2-D</td>
<td>86</td>
<td>Mesh</td>
<td>Adaptive</td>
<td>Packet Switching</td>
<td>Wormhole</td>
<td>0.08</td>
<td>4.9</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mesh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Lee</td>
<td>Slimspider</td>
<td>56</td>
<td>Crossbar</td>
<td>Deterministic</td>
<td>Packet / Circuit</td>
<td>-</td>
<td>-</td>
<td>7.2</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Beigne</td>
<td>Anoc</td>
<td>7</td>
<td>Mesh</td>
<td>Adaptive</td>
<td>Packet Switching</td>
<td>Wormhole</td>
<td>0.25</td>
<td>1.7</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Rostislav</td>
<td>Qnoc-</td>
<td>84</td>
<td>Mesh</td>
<td>Deterministic</td>
<td>Packet Switching</td>
<td>Wormhole</td>
<td>0.96</td>
<td>4.5</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sync</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.4 Issues and Challenges in NoC Design

The interconnection system on Network on Chip requires the development of protocols and algorithms to solve the following issues.

- **Routing Algorithm**: The routing algorithm implemented for a NoC must be simple, fast and the buffer requirements must be
minimal. An appropriate forwarding algorithm has to be introduced towards these constraints. The algorithm must optimally utilize the power in flexible traffic patterns.

- **Deadlock**: An important issue that has to be solved by smart way is deadlock. A complex solution based on Virtual Channels must to be adopted to address the issue

- **Mapping and Resource Allocation**: A prime challenge in predicting the performance and scalability of NoC architecture relies on the present and future traffic patterns. An efficient resource allocation can reduce the power dissipation by 60%.

- **Architecture**: To mitigate the increasing propagation delay that limits the scalability of the system, several research groups have advocated the use of a communication-centric approach to integrate IPs in complex SoCs. The architecture of the system has to address the latency and power constraints.

### 2.5 Problem Statement

The problem of power efficient router design with better operating performance is a challenging area for investigation. Designing wired and wireless NoC router is a wide scope for researchers. However, the concept off NoC is not very popular among the researchers.

This thesis addresses the problems of designing a new router paradigm under the following criteria.

1. Efficient routing algorithm formulation
2. Efficient power characteristics

3. Efficient use of wired/wireless or hybrid links for better performance.

The above research criteria are addressed in details in the various section of this thesis.

2.6 Proposed Methodology

To address the problems faced in the existing techniques, a novel architecture will be proposed. With the growing demand, NoC is the best option for interconnection networks. With a detailed literature survey, the mesh topology and worm hole switching technique best suited for wide range of traffic will be adopted. The power is mainly consumed by the longest path taken by the packets in wired connection. The system complexity is increased with the packet exchange in shortest wireless path. This will be mitigated by the design of efficient routing algorithm. The power starvation algorithms in the existing architectures will be replaced with the aid of reconfigurable routing scheme which manages the shift between wired and wireless connection. This tackles the power and complexity issue faced by the pure wired and wireless connection. In hybrid architectures, the routing scheme must be designed to be capable of optimizing the latency. To evaluate the performance of proposed architecture, the key issue challenges namely power, latency, I/O utilization, Junction Temperature and throughput will be calculated.