Survey on Particle Swarm Optimization Techniques in Network-on-Chip

K. Sathis Kumar, K. Paramasivam

Abstract: ***Network-on-Chip (NoC) an interconnection framework is proposed by Numerous number of Intellectual Property cores in the nature of System-on-Chip(SoC). Communication challenges in a global nature with respect to nanoscale technology is provided by NoC. A configuration of NoC with its least average traffic in communication, consumption of power and area covered in chip is the needed in real time applications. Effective routing, mapping the cache hierarchy, memory and application mapping are the main parameter to increase the efficiency in aNoC. This can be done with optimization techniques. With optimization technique, NoC can be configured such that the latency, consumption of power, and chip area engaged in aNoC are made to be minimal. This paper provides a survey of Particle Swarm Optimization (PSO) algorithm techniques to optimize NoC routing, Mapping of memory and application mapping to provide performance improvement.***

Keywords: ***Network-on-chip (NoC); Cache Hierarchy; algorithms in routing; Particle Swarm Optimization (PSO); Quality of Service (QoS), NoC design, Chip Multiprocessor***

I. INTRODUCTION

Core elements and IP in a chip are integrated to meet the requirements with respect to performance in embedded system. The performance and power consumption is directly related to the interconnections in the chip. In comparison to bus architecture, NoC accommodates larger number of cores which is an important advantage in an inter-connection architecture. The IPs or cores are connected to on-chip routers[1]. NoC is an On-chip integrated network, provides the embedded core-based system chips with a set of standard for communication. The size and the density in a large scale IC, the transmission of dataincreases among processing elements in the environment. NoC Routing is the major design issue of interconnection networks[2]. The strategy of routing is determined by the path followed by each packet between the source node and the destination node. The communication performance is affected by type ofNoC routing. Finding out the best Routing in a NoC environment is an NP-Hard Problem, and these complexities makes it less possible to solve by a traditional common arithmetic. System power consumption, hardware resource consumption and latency are drastically affected by NoC design. NoC design is a most important step in NoCplatform [3]. Application-Specific Network-on-Chip (ASNoC) handles the problems in computational resources communicationincurrent real time applications. The design network in an on-chip in an ASNo Cis optimized to conform to the application requirements.

Revised Version Manuscript Received on 25 November, 2018.

K Sathis Kumar, Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam (Tamil Nadu), India.

Dr. K. Paramasivam, Department of Electrical and Electronics Engineering, Kumaraguru College of Technology, Coimbatore (Tamil Nadu), India.

Major characteristics of ASNoC based systems involvesless cost in communication, less overhead in area allocation,low consumption of energy, and high throughput. There are many NoC topologies proposed for ASNoC. With a survey on various topologies, the mesh topology has various advantages over other topology. It is easier to implement the mesh topology inside chips. Small and equally spaced lengths are maintained between the links. In a topology, all the routers are of same type other than the routers at the corner and edges. These fabrications are made on one layer of metal. Also, most of the techniques in application mapping techniques in and NoC isplanned based on mesh topology.

The flow of PSO Algorithm flow is given below [4]:

1. Initialize the position values and velocity. Binary values are assigned for the position value of particles. The velocity of the particle is assigned between known maximum and minimum.
2. Calculate the value of fitness for the entire particle in the pool.
3. Evaluatevalue of fitness for particles in swarm to best value of fitness, if the value of present fitness is smaller than the previous fitness value; update the best position with presentparticle position.
4. Compare the present position of each particle to best position of entire group considered, if the value is smaller than previous; update the position of group best with position of the presentparticle.
5. Check for satisfaction of convergence rule or on reaching the most figures of iteration, make theprocess to halt, otherwise loop to step 6.
6. Velocity of the particle and its position is updated based on the best values from previous steps.
7. Velocity of the particle and its position is updated in accordance with the operation of cross-variation.
8. Loop to step 2.

This survey manuscript is structured as given below. Section 2 presents optimization in routing the data in a NoC environment. Section 3 presents the mapping applications to the NoC. Section 4 presents optimization in cache. The summary of the paper is given in section 4.

II. NOCROUTING OPTIMIZATION WITH PSO

ANoC system consists of nodes with resource, communication and Resource NetworkInterface (RNI). Resource node is used in performing the computational tasks contains processors, memory devices, reconfigurable devices, input or output devices [5]. Across the network, the nodes of the resource are connected in line with the communication node.

This can be achieved with the help of RNI.The communication tasks are completed by routers which allocate all the nodes in the network.

The position values of eachparticle are initialized. Relative positions of source node and destination nodes are determined in accordance with the position value of the particles in the environment. Particles which move away from the direction of shortest path is removed [6]. The distance of hop between the nodes of source and destination decides the routing information. The binary arrays involved for the values of position is also determines by it. Particles in the swarm are considered for the validity by computation in binary arrays with 0 and 1. The link node of the entire network is determined by the particle fitness value.

The proposed algorithm selects the best routing with bandwidth constraint, measurement of the algorithms performance. The standard of judgment is based on the figures of efficientsteeringsystemwith a set of permanent mapping [7]. There is a decrease in the effective mapping when number of cores and the traces of communication increases. When the application size is larger, PSOwill have best outcome in constrained bandwidth and routing.

III. NOC MAPPING WITH ENERGY- AND LATENCY-AWARE

An IP set includes all the IP cores. IP cores in an IP set can be of homogeneous resource or heterogeneous resource. An Application Task Graph (ATG) is represented by a directed graph. Tasks present in the application is represented by the vertex set C, communication relationship among the tasks is represented by arc set A [8]. Greedy searching is not applicable to NoC mapping since it falls under the category of NP-complete problem. Heuristic method of searching should be applied for the better output.

Mapping process involves selection of IP cores and assignment of the network tiles. In anNoC environment, the possibility of solutions can be represented as $\beta^{\alpha}\times\beta!$when mapping an application with α tasks to a – nodes β. Therefore, with the time complexity, directly obtaining the best solution is very less.[9].

The mapping process is consistsof two phases [10]:

- Estimation on common energy consumption and delay is considered for the selection of IP cores with an optimum output for an IPs in heterogeneous.
- Set the total particles, maximum iterations and randomly initialize the particles.
- Calculate fitness of every particle. Initialize the personal best solutions in response with the present solution; calculate the best solution among the entire particles – gBest.
- With discrete PSO algorithm, updated mapping is obtained from IPs to the NoC tiles in accordance with accurate delay and parameters of power consumption [11].
- Input to the next- second phase is the output of the previous - first phase. Perform injection mapping from IPs to tiles of the network. Only one IP is accommodated in a network tile in injection mapping.
- The solution fitness is calculated after the updating of each particle and the NoC-IP table is constructed with updated new values. With this way, the delay and consumption of power can be estimated with highest accuracy when the lengths between the tasks are known.

The population is generated in random at the initial state. 0.75 is the crossover probability and 0.1 is the mutation probability which is practiced in genetic algorithm of single-point-crossover. Convergence speed is fast with PSO algorithm in process of searching. The convergence rate is very fast with less iteration and fitness value also converges to a small value with view point from number of iterations in PSO.In comparison with PSO, mapping with GA makes convergence very slow and satisfactory performance [12]. By a permanent number of iterations, the performance of power and waiting is optimum in comparison with random mapping and genetic algorithm.

IV. OPTIMIZATION OF CACHE HIERARCHY

Leonid Yavits, Amir Morad, Ran Ginosar focused on major three cache configurations types: private single level, two-level cache (one private and shared) and three level (two private and one shared) cache. Combination of the number of levels with private, hybrid or shared can be extended via this framework [13].Delay due to data transfer, blocking and queuing constitutes the NoC delay. The shared rate of cache access, the capacity of network, the core number, etc. are the major parameters of congestion, i.e., delay of blocking and queuing.

The objective function is the average of delays in memory. These are obtained by the manipulation of the best of configurations numbered to three with a diversity of resource constraints [14]. Power consumption of the cache memory increases with the square root of its area. Traffic in a memory off-chip is optimized by reducing the data rate to the DRAM.

The minimal requirement for satisfying the constraint of off-chip memory bandwidth is two-level hierarchy, which in turn reduces the low access rate to off-chip DRAM. When the area size grows, hierarchy with three levels gets its optimal with area of ~40. A cache with single level provides a feasiblewithminimal solution at ~110. In accordance with the above, zero is allocated to the area. When area grows to level two then three, the allocation among each level is increased in accordance with the configurations and constraints [15].

The optimal solution is provided in single-level cache when the area is low and in a constrained NoC bandwidth. When the area grows, the point of optimality point moves from level one to the two-level hierarchy and to openly shiftingthe third level configuration. Two-level cache moves to suboptimal when the area grows further level [16].

V. MAPPING OF APPLICATION

Random creation of first population is made and the evaluation of value of fitness for each particle is made in the first generation. Initial particle value is take as the pbest i.e., each particle's local best value. The particle which gives the smallest amount communication cost with the fewest fitness function is considered as

the best among all the particles (gbest) in the first generation. Second generation is developed by random exchange positions of the coreinside the particle swarm [17]. The pbest and gbest are updated if the second generation gives better fitness value in comparison with the first generation. Swap operations is employed for the development of future generations by a series of operations. For each and every generation, the pbest and gbest of all the particles are updated when it gets a better fitness value than the previous generation. [18]

- Swap operator: Position index is the location of core element in a particle. The indexing value ranges from 0 to N-1. The positions are swapped for creation of a new particle.
- Swap sequence: combination of swap operators is used in swap sequence. A new particle is created by swap operators with the sequence of swaps applied in order on the particle P. The particle P is identified by swap sequence with its pbest and gbest.

VI. CONCLUSION

Optimization in routing method helps to achieve efficient, deterministic solution, an environment with no deadlocks, with least routing paths in the allocation program. The link load of system is also balanced. Processing units can be used to its potential by re-organizing all the tasks and conveying those tasks to dissimilar units of processing. This shows the better performance can be achieved and can also resolve NoC mapping with respect to energy and latency. The optimizationis improved further bythe inclusion ofthe impact on cache miss rate due to data sharing.

REFERENCES

1. XuChuan-pei, Yan xiao-feng and Chen Yu-qian, “A Technique for NoC Routing Based on Hybrid Particle Swarm Optimization algorithm”, Third International Conference on Genetic and Evolutionary Computing, 2009.
2. Wang Lei, Ling Xiang, “Energy- and Latency-Aware NoC Mapping Based on Chaos Discrete ParticleSwarm Optimization”, International Conference on Communications and Mobile Computing, 2010.
3. Leonid Yavits, Amir Morad, Ran Ginosar, “Cache Hierarchy Optimization”, IEEE Computer Architecture Letters, Vol. 13, No.2, 2014.
4. Pradip Kumar Sahu, PuttaVenkatesh, SunilrajuGollapalli, “Application Mapping onto Mesh Structured Network-on-Chip using Particle Swarm Optimization”, IEEE Computer Society Annual Symposium on VLSI, 2011
5. Hu J, Marculescu R. “Energy-aware mapping for tile-basedNoC architectures under performance constraints”, Proceedings of the 2003 Conference on Asia South PacificDesignAutomation , Kitakyushu , 2003, 233 – 239.
6. Lei T, Kumar S, “A two-step genetic algorithm for mappingtask graphs to a network on chip architecture”, Proceedings of the Euro micro Symposium on DigitalSystem Design, Belek-Antalya, 2003, 180 – 187.
7. Ascia G, Catania V, Palesi M, “An evolutionary approach to network-on-chip mapping problem”, Proceedings of the2005 IEEE Congress on Evolutionary Computation,Edinburgh, 2005, 112 – 119.
8. Murali S, De Micheli G, “Bandwidth-constrained mapping ofcores onto NoC architectures”, Proceedings of the Design,Automation and Test in Europe Conference and Exhibition, Paris, 2004, 896 – 901.
9. L. Benini, G, .De Micheli, “Networks on chips: a newSoCparadigm,”IEEE Computer, vol. 35, 2002. pp. 70-78.
10. William J. Dally, Brian Towles, “Route packets, notwires: on-chip inteconnectionnetworks”,Proceedings of the38th annual Design Automation Conference, ACM, NewYork, NY, USA, 2001, pp. 684-689.
11. Tang Lei, Shashi Kumar, “A two-step genetic algorithmfor mapping task graphs to a network on chip architecture”, Proceedings of the Euromicro Symposium on Digital SystemsDesign, IEEE Computer Society, Washington, DC, USA,2003, pp. 180-187.
12. Alameldeen, “Using compression to improve chip multiprocessorperformance”, PhD thesis, University of Wisconsin, Madison, WI, 2006.
13. Cassidy and A. Andreou, “Beyond Amdahl Law -An objectivefunction that links performance gains to delay and energy”, IEEETransactions on Computers, vol. 61, no. 8, pp. 1110-1126, Aug 2012.
14. Krishna, A. Samih, and Y. Solihin. "Data sharing in multi-threadedapplications and its impact on chip design", ISPASS, 2012.
15. Morad, T. Morad, L. Yavits, R. Ginosar, U. C. Weiser. "GeneralizedMultiAmdahl: Optimization of Heterogeneous Multi-Accelerator SoC,"IEEE Computer Architecture Letters, 2012.
16. L.Benini, “Application Specific NoC Design,” Proceedings ofIEEE Design, Automation and Test in Europe Conference, 2006 vol. 1, pp. 1–5.
17. P. Pande, C. Grecu, M. Jones, A. Ivanov and R. Saleh,“Performance Evaluation and Design Trade-offs for MP-SOC Interconnect Architectures,” IEEE Transactions on Computers, Vol.54, No. 8, pp.1025–1040, 2005
18. N. Koziris et al.,”An Efficient Algorithm for the Physical Mappingof Clustered TaskGraphsontoMultiprocessorArchitectures,”Proceedingsof 8thEuroPDP, pp. 406-413, 2000.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication