International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:268
Vol.2, No.6, June 2013 DOI:10.15693/ijaist/2013.v2i6.78-84

Design and implementation of adaptive mapping and scheduling for noc

K.Muthulakshmi[1] S.Rameshkumar[2] Dr.R.Ganesan[3]
[1] PGScholar, Sethu Institute of technology, Madurai.
[2] Assistant Professor, Sethu Institute of technology , Madurai.
3 professor HOD, Sethu instute of technology, Madurai

***Abstract*—Task scheduling and core mapping have a significant impact on the overall performance of network on chip (NOC). In this paper, an adaptive task scheduling and core mapping algorithm is proposed for different NOC architectures including regular mesh, irregular mesh and custom networks. First, a unified model combining scheduling and mapping is introduced using mixed integer linear programming (MILP). To make the MILP-based system scalable, shared routers by multiple tiles is employed to speed up our method. Compared with two previous state-of-the-art works, experimental results show that more than 25% and 15% improvement on the execution time is achieved with similar energy consumption on average for regular mesh NOC. For irregular and custom NOC, the improvementis 30% and 14.5% on the execution time with 24.3% and 20% lower energy. Moreover, our method is scalable for large benchmark circuits.**

Keywords – core mapping,Network on chip(NOC),Networ topology,Task scheduling.

I.INTRODUCTION

AS VLSI technology advanced into deep submicrometer era, network on chip (NOC) which enhances on-die communicationby data packetization is considered as an alternativeof conventional bus-based interconnection in system-on-chip(SOC) design. As early stages of the NOC design flow, task scheduling and core mapping have a great impact on the overall performance of the entire system. Task scheduling is to assign each task in a task graph to different cores and decide the sequence of their executions (called execution table) if two tasks are scheduled on the same core. Then, a core graph, which defines the communication between cores instead of tasks, could be generated from the task graph. Core mapping is to assign each task in a task graph to different cores and decide the sequenceof their executions (called execution table) if two tasksare scheduled on the same core. Then, a core graph, which definesthe communication between cores instead of tasks, could be generated from the task graph. Core mapping is to assign cores in the core graph onto network tiles. Based on the floorplanwhich is generated after core mapping, network routing could beperformed statically in design time or dynamically in run time.

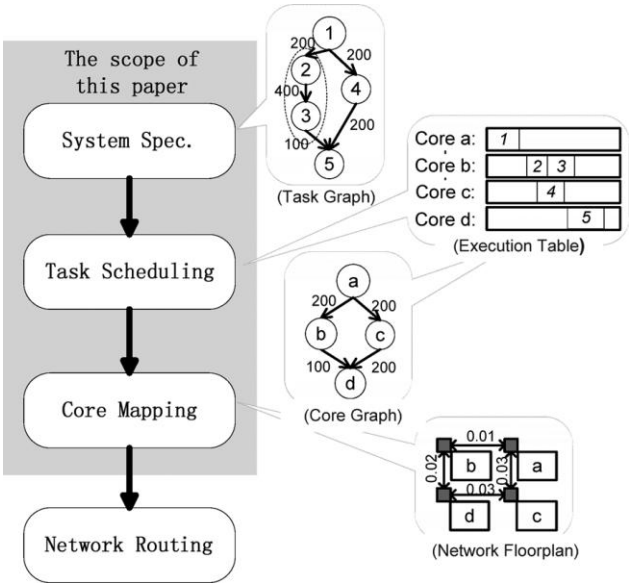

Fig 1.Traditional NOC design flow.

As a result,it is necessary to consider scheduling and mapping as an integral process for a globally-optimized communication, whichwill further lead to a better performance and lower energy of the whole system.Since scheduling and mapping are tightly coupled because of the communication, we will model them as a unified manner.Shin et al. [1] explored the design space of the unified methodology.Another two previous works were also carried out in aunified manner recently by Chi et al. [2] and Yu et al. [3]. However,only regular mesh was considered in these works. The benefitof this unified methodology has been shown by Ghosh et al.[4], but still for regular mesh Practically, there are different types of architectures for aNOC design, besides regular mesh. Fig. 2 shows three differentNOC architectures.

Numbers on the edges denote therelative As mentioned before, thisunified methodology has not been fully studied yet,especiallyfor irregular mesh and custom network, as in Fig. 2(b) and (c).Besides, the benefit on chip performance and energy by applyingthis unified flow on irregular and custom network will be greater than regular mesh.

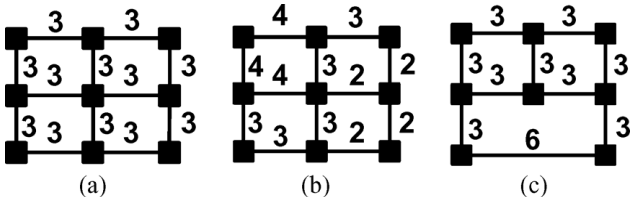

Fig2.(a)Regular mesh (b)Irregular mesh(c)Custom NOC

1) A thorough study of the unified taskscheduling and coremapping is presented using mixed integer linear programming
(MILP).
2) A novel communication model called Labeled Graph isadopted to extend our MILP model to general networks,especially to irregular and custom networks, where the
number of network hops is not accurate any more for thecommunication latency and energy.
3) A heuristic is developed to accelerate the algorithmand make it scalable for large benchmarks.

Energy efficiency is one of the most critical design issuesin embedded systems. For this reason, energy-aware taskallocation becomes an active research topic [4], [5], [6], [7].Existing studies that tackle this problem can be categorized into two main branches: software partitioning, which assigns each task to a particular processor statically, and dynamic scheduling, which relies on a global scheduler

To find an energy efficient application mapping, all four problems (i)-(iv) have to be solved. There are two option savailable - solve them sequentially or solve them in a unifiedway. Thesequentialapproach has manifold disadvantages.Firstly, decision taken at an early phase may turn out to be expensive later. Secondly, because of some earlier decisions may lead to violation of constraints at some later phase, and thus resulting in re-execution of all the steps multiple times involving an enormous amount ofcomputation.

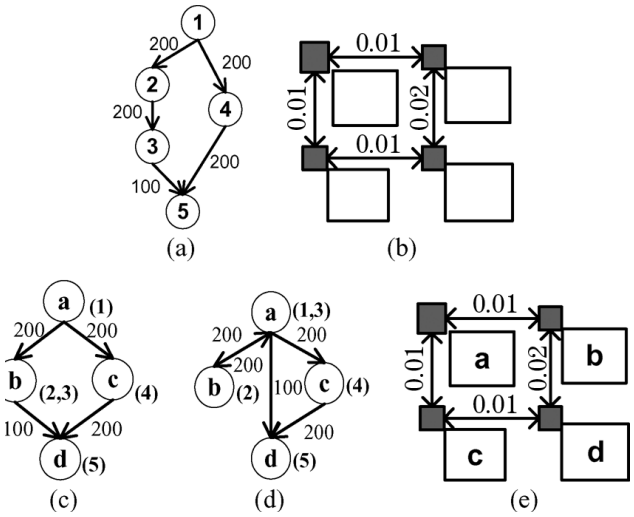

Fig. 3. Scheduling and mapping on irregular NOC. (a)Task graph. (b) Targetnetworks tiles. (c) Core graph with min cut. (d) Core graph with larger cut.(e) Mapping result of Fig. 2(b).

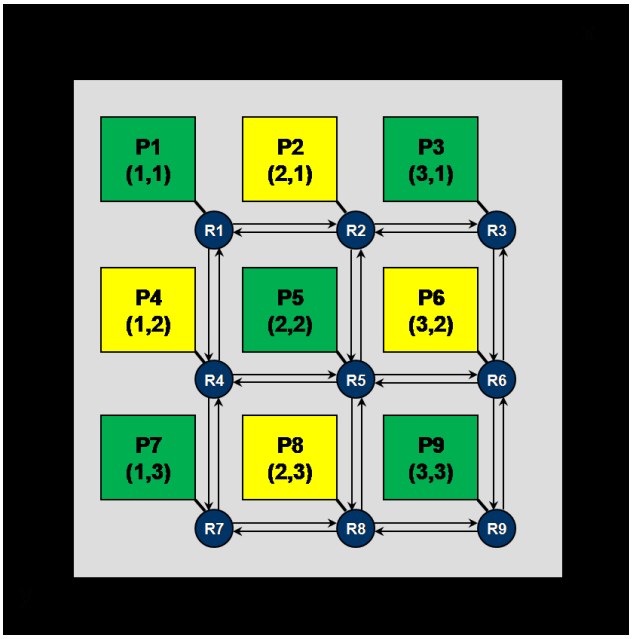

Figure 4. An example 2D-mesh NoC

For irregularmesh and custom NOC, core mapping is usually performed togetherwith floorplanning information, like [19], [20]. Theseworks take a core graph as an input, which describes the communications among different cores rather than tasks. As a result,a core graph needs to be extracted from the task graph .

II.PROPOSED WORK

A.MIXED INTEGER LINEAR PROGRAMMING

Mixed integer linear programming (MILP) isa powerful representation often used to formulate decision-making problems under uncertainty.However, it lacks a natural mechanism toreason about objects, classes of objects, and relations. Mixed integer linearprogramming has established itself as a successful formalism for decision-making under uncertainty in operation research and cooperative control, and has attracted attention more recently in AI and machine learning.Hence MILP helps in finding an optimal solution in our paper.

B. MOTIVATION

In this section, we illustrate the benefit of adaptive scheduling and mapping on irregular NOC with shared routers by multiple tiles. The challenge of supporting general network architectures will be discussed as well.
First, a task graph given in Fig. 3(a) is to be scheduled onCore a, b, c, and d. In some industrial benchmarks like E3S, the energy variance for the same task on different corescould be more than 10 times. Table I shows an example of thecomputing energy when tasks are executed

on the four cores.The numbers in Table I are assumed for illustration. Then, a coregraph could be generated as Fig. 3(c) or (d), where the nodesrepresent the cores and the edges represent the communication volume between two cores. At last, the core graph is mapped

TABLE I
TASK ENERGY ON EACH CORE

	Task 1	*Task 2*	*Task 3*	*Task 4*	*Task 5*
Core a	1	10	1	10	10
Core b	10	10	10	10	10
Core c	10	10	10	10	10
Core d	10	10	10	10	10

onto a given network in Fig. 3(b), where the weight on eachedge denotes communication energy for sending one bit on thislink.Fig. 3(c) and (e) show scheduling and mapping results generatedby a traditional "min-cut partitioning + mapping" flow,like [21]. In that case, the number of cut is 700 and the energyconsumption is

Ecore=1+10+10+10+10

Ecom=0.01*200*3+0.02*100

E=Ecore+Ecom=41+8=49

where and denote the core computing energy andcommunicationenergy,respectively.However, there is another result from a unified flow with alarger cut (i.e., 900) in Fig. 3(d), which has a smaller total energy.

E=Ecore+Ecom

=(1*2+10*3)+(0.02*800+0.02*100*2)

=32+10=42

C.ADAPTIVE UNISM MODELING AND ALGORITHM

.

Modeling the Unification of Scheduling and Mapping: Therefore, the unified modelingcombining these two stages becomes the only way to evaluatethese objectives. The first challenge is how to design the variables, which canbe used to calculate these objectives in a linear way.

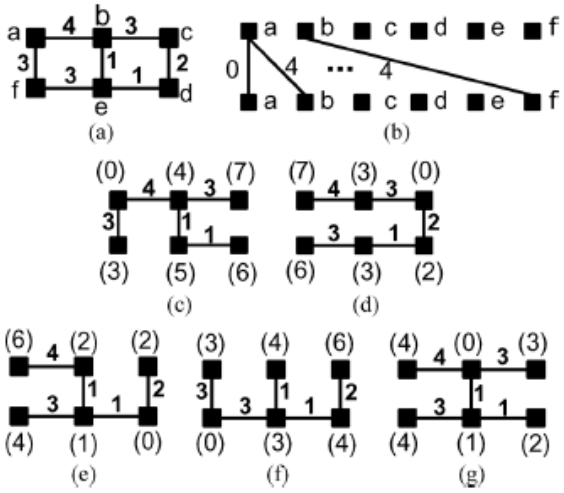

Fig5a)irregularmeshb)calculatingminimalenergy

a) The three groups are sorted in the order of Part 1, Part 2and App 2 as in Line 1 of Fig. 11.
b) Then, we budget the number of cores on each group. Part1 gets two cores while Part 2 and App 2 get one core.
c) A bipartite Graph B is then built up in Fig. 13 with thedefinition in Line 3 of Fig. 11. Note that Part 1 has adummy node, since it gets two cores in the last step.
d) The minimal weighted bipartite matching algorithm is executedon Graph B to allocate cores to each group, like
.e) From Line 5 to Line 12, we will allocate network tiles to
the three groups. First, Part 1, Part 2 and App 2 are sortedin order with total communication volume 200, 200, and
0. Then, Part 1 with Core b and c is picked up. Twotiles with link number 0.01 are assigned to Part 1, as in

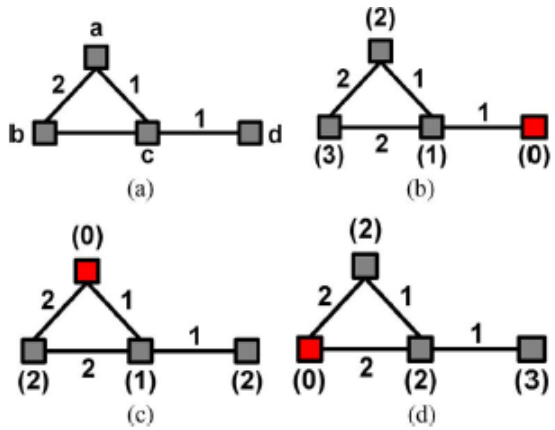

Fig6 Examples of labeled graph

Hence , a unified flow combining task scheduling andcore mapping named adaptive UNISM is proposed to support regular mesh, irregular mesh and custom NOC, using MILP. To enablethis MILP modeling on irregular and custom NOC, a novelgraph model called Labeled Graph is developed to calculate thecommunication latency and energy. Moreover, this model doesnot introduce any new variables,whichmakes our unified model as easy as a

single scheduling algorithm in terms of the numberof variables in MILP. Then, we accelerate our UNISM usingNOC localization. Compared with two latest previous works,experimental results show that 15% and 11.5% improvement on the execution time of the task graph is achieved on averagewith similar energy consumption for regular mesh NOC. For irregularmesh and custom NOC, the improvement is 27.3% and14.5% with 24.3% and 18.5% lower energy on average. Moreover,our method is scalable in terms of runtime.

First and foremost, NoC applications such as cache coherence protocolscannot tolerate dropping of packets unlike Internet protocols.To guarantee packet delivery the proposed DSB router uses creditsso that flits can only be timestamped and placed in a middle memorywhen its next-hoprouterhasbuffersavailableatthecorrespondinginput port.HenceNOC applications plays amajor role in real time applications . The need for small buffers due topower and area constraints andthe need to achieve ultra-low latency communication in NoCs .to support a wide range of delay-sensitive applications with diverse traffic characteristics necessitate a novel flow control

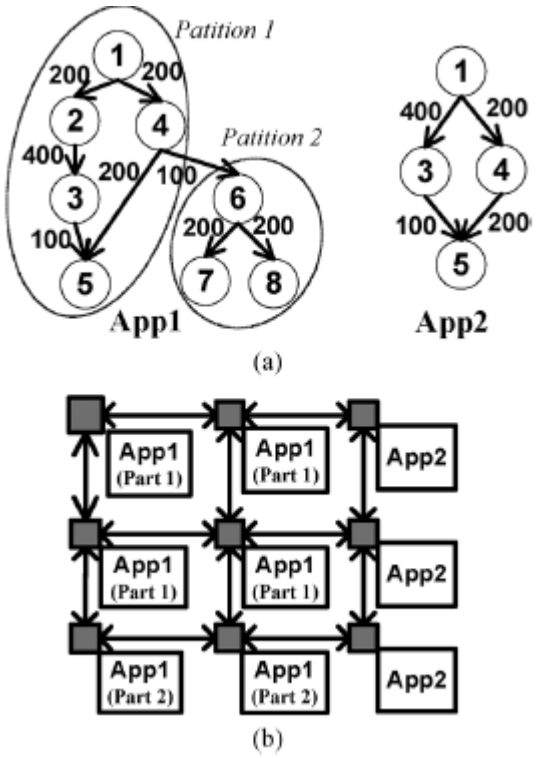

Fig. 7. Accelerating the optimization. (a) A task graph with two applications.
(b) A localized unification on 3 *3 mesh

III.SPEEDUP TECHNIQUES

In this situation, it is inefficient to merge all the sub graphsand perform an exhausting search on the solution space. A bettertradeoff between solution quality and runtime is to localize each subgraph on the network, All the cores and network tiles are partitioned into three groups and then eachgroup performs scheduling and mapping within a local domainof the network. As discussed in Section III, min-cut partitioning
may cause some loss on the solution quality. But different fromthe previous works which partition the task graph directly ontoeach core, we partition this graph onto different groups of cores,instead of one core. Within each group, we can still apply our UNISM. The granularity of our partitioning is much larger thanthe previous works which results in a better tradeoff between
CPUruntime and solution quality, as in SectionVII.Ourmethodfor accelerating UNISM is introduced as follows:

First, if the number of input applications (or called subgraphs) is greater than a threshold, they will be clustered tothis threshold. On the other hand, if the number of the applicationsis less than the threshold, they will be partitioned to thisthreshold. In our implementation, the threshold for groupingsub task graphs is empirically set to . One clustered or partitioned group of applications is called a sub
graph group.Then, cores and network tiles are assigned to each subgraph group.

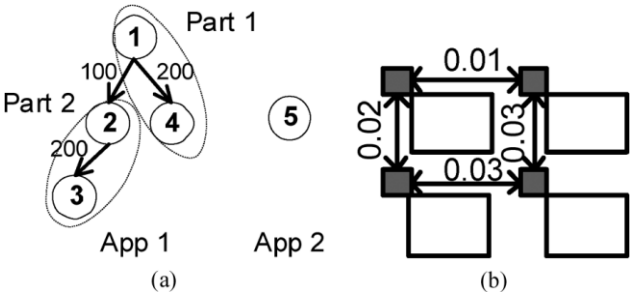

Fig. 8. Input task graph and network topology

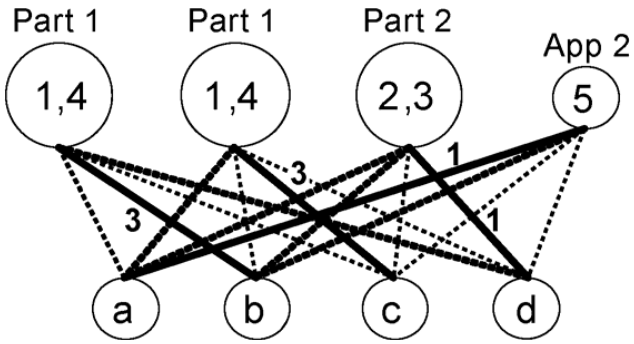

Fig. 9. Assigning cores to each sub graph group.
.

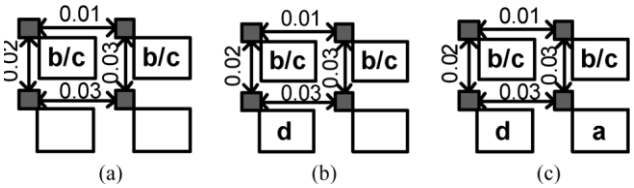

Fig. 10. Assigning tiles to each sub graph group

IV.SHARED ROUTERS BY MULTIPLE TILES:

In all the test cases ofthis paper, each tile owns one router. However, in custom NOCdesign, it is common to have shared routers by multiple tiles,as in Fig. 20(a). In this situation, we can generate an equivalenttopology by adding some dummy routers like Fig. 20(b). After that, Labeled Graph and UNISM still work.Due to increasing systems complexity and design productivity gap, the design of future multiprocessor systems-on-chip(MPSoCs) faces several challenges.From this perspective,Networks-on-Chip (NoCs)are a promising interconnect solution for complex chips consisting of many heterogeneous intellectual property (IP) cores. In NoC-based MPSoCs, the global wires are replaced by a network of shared links and multiple routers exchanging data packets simultaneously. Ideally, the traffic congestion in such a network should be avoided in order to maximize the system performance. In practice, however, it is difficult to completely eliminate the network contention which significantly degrades the system performance, particularly latency and throughput.Previous work attempts to minimize of the communication Previous work attempts to minimize the communicationenergy consumption.However the communication energy consumption is a good indicator of latency only if there is no congestion in the network. Indeed, in the absence congestion,packets are injected/transmitted through the network as soon as they are generated and then latency can be estimated by counting the number of hops from source to destination.

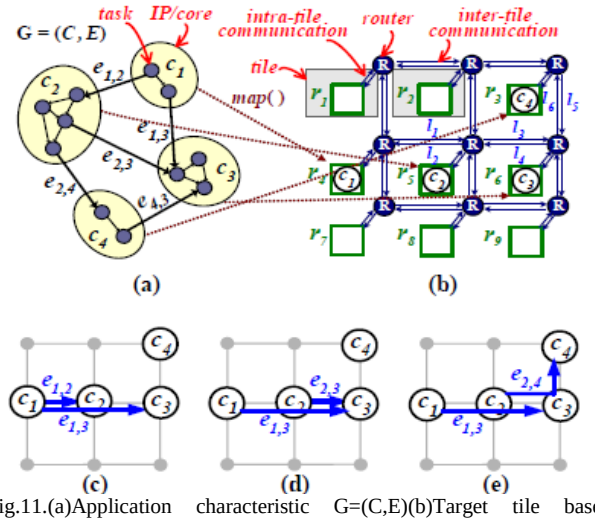

Fig.11.(a)Application characteristic G=(C,E)(b)Target tile based NoC(C)Source based contention(d)destination based contention(e)Path based contention

assigned and scheduled onto a set of available cores1 [1]. Theapplication graph is described by G = (C, E) as shown inFig.11(a). Each core ci C represents a cluster of tasks, whereeach edge ei,j E represents the communication between the
cores ci and cj. Note that the tasks belonging to the same coreare mapped onto the same tile of the NoC. Fig. 11(b) shows thegeneral design of a 2D tile-based NoC. Inside any tile, eachcore is attached to a router via the network interface whichallows for exchanging packets with other cores. Each router isconnected to five ports: four neighboring ports (east, south,west, and north), plus a local port connecting the core. Thecommunication between the routers is referred to as "inter-tile"communication (e.g., communication through links li wherei = 1 ~ 6 in Fig. 11(b)), while the local communication betweenthe core and the router is referred to as "intra-tile" communication.For reasons that will become clear later, our work in this We clarify the network contention due to the inter-tile datacommunication into three types: source-based, destinationbased,and path-based contention. The source-based contention(see Fig. 11(c)) occurs when two traffic flowsoriginating fromthe same source contend for the same links. The destinationbasedcontention (see Fig. 11(d)) occurs when two traffic flowswhich have the same destination contend for the same links.Finally, Fig. 11(e) shows the path-based contention when twodata flows which neither come from the same source, nor gotowards the same destination contend for the same links somewherein the network.In this paper, we first evaluate the impact of these three typesof contention on average packet latency; then present an integerlinear programming-based (ILP-based contention-aware mappingtechnique for minimizing the network contention andcommunication energy consumption. We show that by mitigatinga major source of contention, the end-to-end average packetlatency can be significantly decreased.

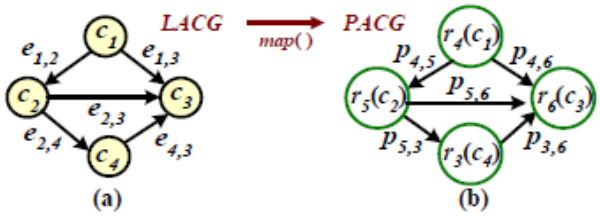

Fig12.Logical and physical application communication mapping

Compared to previous work, our focus in this paper is on the network contention problem; this highly affects the latency,throughput, and communication energy consumption. We show that, by mitigating the network contention, the packetlatency can be significantly reduced; this means that the networkcan support more traffic which directly translates into significant significant improvements.Furthermoe, our mapping solution can be used to improve the efficiency of congestion control techniques for best-effort communication .Lastbut not least, this idea can be applied to other NoC synthesisproblems, such as topology selection, or path selection and some mapping/scheduling heuristics on parallel systemsto achieve further packetlatency reduction and throughput improvements.

V. COMPARISON ON RANDOM BENCHMARKS

We tested five randombenchmarks for each group we do not use our method without acceleration, because the runtime of UNISM grows exponentially. Therefore, in this case,our acceleration technique becomes necessary when the original MILP model ends up with unacceptable CPU time.On these larger benchmarks of Table VI, we can still find that accelerated UNISM works better than NMAP and A3MAP-GA(22% and 14% less with 17%and 11%lower on theaverage of all the network architectures), even if some solutiondegradation is introduced by our accelerating heuristics.Also, improvements on execution time and energy are shown

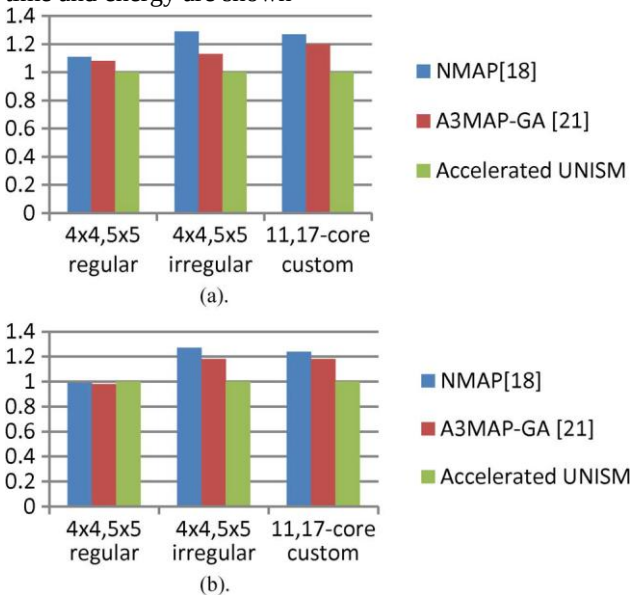

Fig 13.Comparision on industrial benchmarks

VI. CONCLUSION

In this paper, an adptive method combining task scheduling and core mapping is proposed to support regular mesh, irregular mesh and custom NOC, using MILP. To enable this MILP modeling on irregular and custom NOC, a novelgraph model called Labeled Graph is developed to calculate the communication latency and energy. Moreover, this model doesnot introduce any new variables,which makes our unified model as easy as a single scheduling algorithm in terms of the number of variables in MILP. Then, we accelerate our method using shared routers using NOC localization. Compared with two latest previous works,experimental results show that 25% and 14.5% improvement on the execution time of the task graph is achieved on averagewith similar energy consumption for regular mesh NOC. For irregularmesh and custom NOC, the improvement is 27.3% and14.5% with 24.3% and 18.5% lower energy on average. Moreover,our method is scalable in terms of runtime.

REFERENCES

[1] D. Shin and J. Kim,"Powerawarecommunication optimization fornetworks-on-chip with voltage scalable links," in Proc. Int. Conf.Hardw./Softw. Codesign Syst. Synth., 2004, pp. 170–175.

[2] H. C. Chi, C. M. Wu, and J. H. Lee, "Integrated mapping and schedulingfor circuit-switched network-on-chip architectures," inProc. 4thIEEE Int. Symp. Electron. Design, Test, Appl. (DELTA), pp. 415–420.

[3] H. Yu, Y. Ha, and B. Veeravalli, "Communication-aware application
mapping and scheduling for NoC-based MPSoCs," in Proc. IEEE Int.Symp. Circuits Syst. (ISCAS), pp. 3232–3235.

[4] P. Ghosh, A. Sen, and A. Hall, "Energy efficient application mappingto NoC processing elements operating at multiple voltage levels," inProc. ACM/IEEE Int. Symp. Netw. Chip, 2009, pp. 80–85.

[5] R. Marculecu, U. Y. Ogras, L. S. Peh, N. E. Jerger, and Y. Hoskote,"Outstanding research problems in NoC design: Circuit,microarchitecture,and system-level perspectives," IEEE Trans. Comput.-AidedDesign Integr. Circuits Syst., vol. 28, no. 1, pp. 3–21, Jan. 2009.

[6] J.-M. Chang, M. Pedram, "Codex-dp: co-design of communicatingsystems using dynamic programming," in IEEE Trans. on Computer-AidedDesign of Integrated Circuits and Systems, pp. 732-744, July 2000.

[7] A. Bender, "MILP based task mapping for heterogeneous multiprocessorsystems," Proc. EURO-DAC '96, European, pp. 190-197, 1996.

[8] S. Murali, G. De Micheli, "Bandwidth-constrained mapping of coresonto NoC architectures," Proc. DATE, pp. 896-901, Feb. 2004.

[9] G. Ascia, V. Catania, M. Palesi, "Multi-objective mapping for meshbasedNoC architectures," Proc. CODES+ISSS, pp. 182-187, Sept. 2004.

[10] J. Hu, R. Marculescu, "Energy- and performance-aware mapping forregular NoC architectures," in IEEE Trans. on Computer-Aided Design ofIntegrated Circuits and Systems, pp. 551-562, April 2005

[11]M. Alighanbari and J. P. How. Cooperative task assignment ofunmanned aerial vehicles in adversarial environments. In Proc.IEEE American Control Conference (ACC), 2005.V.

[12]Borkar, V. Chandru, and S. Mitter. Mathematical programmingembeddings of logic. Journal of Automatic Reasoning, 2002.

[13]R. Braz, E. Amir, and D. Roth. Lifted first-order probabilisticinference. In IJCAI-05, 2005.
Vijay Chandru and John Hooker. Optimization methods for logicalinference. Wiley, New York, 1999.B.C.Eavesand U.G.Rothblum.Formulation of linear problemsand solution by a universal machine. Mathematical Programming,65(1–3), 1994.

[14]R. Fourer, D. M. Gay, and B. W. Kernighan. AMPL: A ModelingLanguage for Mathematical Programming.DuxburyPress,2002.http://ampl.com.R. E. Gomory. Outline of an algorithm for integer solutions tolinear programs. Bull. Amer. Math. Soc., 64(5):275–278, 1958.G. Gordon, S.

[15]A. Hong, and M. Dud´ık. First-order mixed integerlinear programming. Technical Report CMU-ML-09-108,
2009.D. Heckerman, C. Meek, and D. Koller. Probabilistic entityrelationship

[16] Shang-Tse Chuang, A. Goel, N. McKeown, and B. Prabhakar. Matching

AUTHORS PROFILE

.

K.Muthulakshmi received my B.E degree in P.T.R college of engineering and Technology Madurai.Currently pursuing mu M.E degree in Sethu institute of Technology Kariapatti.My area of interest includes low power in vlsi design and vlsi technology.

S.Ramesh Kumar received his **M.Sc** degree in electronics from the St.Josephs College, Trichy, in 2004. He completed **M.Tech.** in Embedded Systems in SRM University, Chennai, India in the year 2006. Currently working as Assistant Professor in the department of ECE in Sethu Institute of Technology. His research interest includes VLSI Design and Embedded Systems.

Dr.R.Ganesan received his B.E. Instrumentation & Control Engineering from Arulmigu Kalasalingam College of Engineering and ME (Instrumentation) from Madras Institute of Technology in the year 1991 and 1999 respectively. He has completed his PhD from Anna University, Chennai, India in 2010. He is presently working as Professor and head in the department of M.E-VLSI Design at Sethu Institute of Technology, Madurai, India. He has published more than 25 research papers in the National & International Journals/ Conferences. His research interests are VLSI design, Image Processing, Neural Networks and Genetic algorithms.