

# Parallel BMDA with an Aggregation of Probability Models

Jiri Jaros, and Josef Schwarz, *Member, IEEE*

**Abstract**—The paper is focused on the problem of aggregation of probability distribution applicable for parallel Bivariate Marginal Distribution Algorithm (pBMDA). A new approach based on quantitative combination of probabilistic models is presented. Using this concept, the traditional migration of individuals is replaced with a newly proposed technique of probability parameter migration. In the proposed strategy, the adaptive learning of the resident probability model is used. The short theoretical study is completed by an experimental works for the implemented parallel BMDA algorithm (pBMDA). The performance of pBMDA algorithm is evaluated for various problem size (scalability) and interconnection topology. In addition, the comparison with the previously published aBMDA [24] is presented.

## I. INTRODUCTION

The capability and performance of traditional parallel genetic algorithm (PGA) is well known. It flows from the idea that the large problem can be successfully solved using decomposition of the original one into smaller tasks. Consequently, the tasks can be solved concurrently using multiple processors.

Mostly, the original population is divided into a few subpopulations or demes, and each of these demes evolves separately on different processor. Exchange of information among subpopulations is realized via a migration operator. In this context, the term island model is commonly used. Island populations are free to converge toward different optima. The migration operator is supposed to mix good features that emerge locally in the different demes.

Many topologies are used for connecting the demes and migration can occur between neighboring demes. In [4], the theory is published providing rational decisions for the proper setting of control parameters. An interesting survey of PGA is published in [2]. An effective technique for the massive parallelization of compact GA was published in [12], [33]. An extremely prestigious PGA which is capable to solve billion-variable optimization problems was recently published in [7].

Manuscript received November 1, 2008. This work was partially supported by the Grant Agency of the Czech Republic under No.~102/07/0850 Design and hardware implementation of a patent-invention machine and the Research Plan No. MSM 0021630528 - Security-Oriented Research in Information Technology.

Jiri Jaros is with the Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic (e-mail: jarosjir@fit.vutbr.cz).

Josef Schwarz is with the Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic (e-mail: schwarz@fit.vutbr.cz).

This paper concerns the application of three differently interconnected island-based models for one of EDA algorithms known as Bivariate Marginal Distribution Algorithm (BMDA). The new approach, using the idea of probability model migration, is conceptually different from the traditional parallel evolutionary algorithms with migration of individuals/solutions. It differs also from the EDAs using parallel building of pseudo-sequential probabilistic models.

The paper is organized as follows: Section II introduces the basic concept of EDA algorithm and current approaches used in the parallelization of EDA algorithms. In Section III, a short discussion is dedicated to the theory of probability distribution aggregation. The description of the sequential BMDA is presented in Section IV. Section V includes the specification of island based topology and the principles of quantitative combination of probabilistic models. The parallel implementation of pBMDA algorithm is presented in Section VI. Experimental works and results are included in Section VII, Section VIII concludes the paper.

## II. TRADITIONAL EDAS

EDAs belong to the advanced evolutionary algorithms based on the estimation and sampling of graphical probabilistic models [8], [10], [19], [20], [22]. The canonical sequential discrete EDA is described in Fig. 1.

```
Set  $t \leftarrow 0$ ;  
Generate initial population  $D(0)$ ;  
While termination criteria is false do  
begin  
  Select a set of promising solution  
   $D^s(t)$ ;  
  Construct a new probability model  $M$   
  from  $D^s(t)$  using chosen metric;  
  Sample offspring  $O(t)$  from  $M$ ;  
  Evaluate  $O(t)$ ;  
  Create  $D(t+1)$  as a subset of  $O(t) \cup$   
   $D(t)$  with cardinality  $N$ ;  
   $t \leftarrow t + 1$ ;  
end
```

Figure 1. The pseudocode of canonical EDA

EDAs often surpass classical EAs in the number of required fitness function evaluations. However, the execution time is still a factor which determines the size of practically tractable problems. Referring to Fig. 1, the most time consuming task is the estimation of probability model. Most papers on EDAs concentrate on parallel construction and

sampling of probabilistic models. The well-known algorithm employing parallel construction of Bayesian network is EBNA algorithm targeted for MIMD architecture and designed both for MPI and POSIX threads, published in [14], [17], [18]. In [21], the theory of population sizing and timing to convergence is published.

A new idea of the multideme parallel estimation of distribution algorithm (PEDAs) based on PBIL algorithm was published in [1]. In [13], mixtures of distribution with Bayesian inference are discussed. Parallel learning of belief networks in large domains is investigated in [23]. Using the concept of PBIL algorithm [3], [9], [16], the classical phenomenon of migration in island based EAs was carried over into probability distribution of EDAs. A new approach of probability vector crossover was implemented with very good performance.

### III. AGGREGATION OF PROBABILITY MODELS

Aggregation of probability models is a technique frequently used in statistics, expert knowledge combination, and in Artificial Intelligence (AI). Interesting surveys can be found in [25]. From the area of AI, they can be mentioned [26], [27] which are close to the topic of structured distribution aggregation. A new approach to learned probabilistic belief is published in [28]. Recently a very efficient tool for probability graphical model combination is published in [29] - the problem of aggregation of probabilities is mostly reduced to the linear opinion pool (LinOP), so named by Stone (1961):

$$\text{LinOP}(\beta_1, p_1, \dots, \beta_k, p_k)(X) = \sum_{m=1}^k \beta_m p_m(X), \quad (1)$$

where  $p_1, \dots, p_k$  are given probability distributions from  $k$  data sources and weights  $\beta_m \geq 0$ , such that  $\sum \beta_m = 1$

Another typical combination approach uses multiplicative averaging and is often called a logarithmic opinion tool:

$$\text{LogOP}(\beta_1, p_1, \dots, \beta_k, p_k)(x) = r \prod_{m=1}^k p_m^{\beta_m}(x), \quad (2)$$

where  $r$  is a normalizing constant and  $x$  is atomic state

In general, the graphical probabilistic model defined one data source can be formalized as  $M = (G, \Theta)$ , where  $G$  is dependency graph and  $\Theta = (\theta_0, \theta_1, \dots, \theta_{n-1})$  is a set of parameters represented by local conditional or marginal probability for each node of the dependency graph. The previous references mentioned in Section II deal with the simple univariate probability model only. The aggregation/combination of probability models could be thus reduced to a quantitative combination of models only, resulting just in a combination of conditional probability tables (CPT), determining the set  $\Theta$ . If the dependency graph is not trivial, the aggregation of particular models consists of qualitative and quantitative phases. Qualitative aggregation deals with the estimation of consensus models structure, the

quantitative aggregation deals with the estimation of the consensus model parameters [21].

It is evident that the order in which the both type of aggregation are performed determines the proper approach to solve the task of model aggregation. If a consensus structure is firstly obtained and then the model's parameters are estimated, we are referring to topological fusion. If we use the opposite order of these steps, we are referring to graphical representation of consensus.

Let us assume, we have  $m$ -th particular graphical probability models  $\mathcal{M}_m$  each one stated by  $m$ -th engine (e.g. by EA subpopulation) or  $m$ -th observer in case of expert system. The aim is to construct an aggregated model which is capable to cover the whole world state. Generally it holds

$$\mathcal{M} = \sum_{m=1}^k \beta_m \mathcal{M}_m, \quad (3)$$

where  $\beta_m$  denotes the weight/importance of the particular model  $\mathcal{M}_m$ .

Each dependency graph  $G_m(V_m, E_m)$ ,  $m = 1, 2, \dots, k$  is represented by a set nodes  $V_m$  (representing variables/parameters) and set of oriented edges  $E_m$  (representing mutual dependency of variables). The dependency graph  $G_m$  can be represented by Bayesian network, tree, polytree or chain. The systematic approach how to aggregate Bayesian networks (BN) was published in [29]. In the first step, qualitative combination has to include a technique how to avoid cycles after combining two candidate models. In the second step, the problem of the possible increase of dependency level must be solved, after addition an edge. It can cancel some independent relationships among the variables. In the third step, the quantitative computation must be performed. Generally, Conditional Probability Tables (CPTs) may be different in combined models, namely their sizes.

In this paper we will deal with a simpler EDA model which includes only the bivariate dependency between nodes. The graphical model is represented by tree (forest) and associated algorithm is called Bivariate Marginal Distribution Algorithm (BMDA). As known, each node of the bivariate dependency graph (BDG) has only one parent.

Next, we make some assumption about possibility to utilize the principles specified in [29].

In Fig. 2, two candidates of BDGs to be aggregated are shown. Unlike [29], the both dependency graphs have got the same number of nodes. Adding any arc to the BDG1 breaks the limitation on bivariate dependency, e.g. adding the edge  $BE$  from BDG2 to BDG1 is that case. On the other hand adding the edge  $AC$  from BDG1 to BGD2 leads to cycle between  $A$  and  $C$  nodes. The solution of the problem lies in an adjusting of variable ordering to maintain directed acyclic graph. The Order( $v$ ) of a node  $v \in V$  in BDG is defined as the longest path from a rooted node to node  $v$ .

The variable ordering  $\gamma$  is specified as the sequence of

nodes sorted by  $\text{Order}(v)$  in ascending order for all nodes of BDG.

Cycles can be avoided when each edge in the candidate BDG originates in a node with lower order value and terminates in a node with higher order value. Two examples of candidate BDGs are shown in Fig. 2, and the final order values are explicated in Table 1.

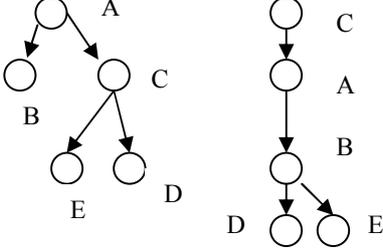


Figure 2. Example of two candidates of BDGs (BDG1 left, BDG2 right)

Node	BDG1	BDG2	Final ordering
Ordervalue (A)	0	1	1
Ordervalue (B)	1	2	2
Ordervalue (C)	1	0	3
Ordervalue (D)	2	3	4
Ordervalue (E)	2	3	5

Using the final ordering the orientation of edges in BDG1 remains unchanged, and in the BDG2 the arc  $CA$  is replaced by  $AC$  arc, see Fig.3. If we use the operation of intersection of BDG1 and BDG2 we achieve a common subtree involving the nodes  $A, B, C$  and it is possible to start the quantitative phase including the point combination of relevant CPT tables. The resulting tree can be then obtained as an alternation of the process of qualitative and quantitative aggregation with some limitations and optimization criterion.

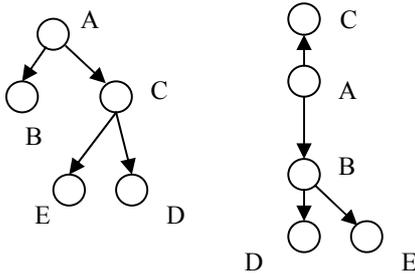


Figure 3. Resulting edge reversion in BDG2 (right) according to Table 1

Generally, the fusion of BDG is not without serious problems. Our attempt with the topological fusion of particular BDGs resulted in not feasible efficiency. Therefore we have proposed a second strategy mentioned in the previous analysis as a graphical representation of consensus, see next sections.

#### IV. SEQUENTIAL BMDA

Before we will specify a new parallel version of the BMDA a short review about the principles of sequential BMDA algorithm is presented.

The well known representative of bivariate EDAs is the Bivariate Marginal Distribution Algorithm (BMDA) proposed by Pelikan and Mühlenbein [1], [20]. This algorithm uses a factorization of the joint probability distribution that exhibits second-order dependencies.

EDAs are also population based algorithm but unlike GAs the new population is generated by sampling the recognized probability model.

Let us denote:

$D = (X^0, X^1, \dots, X^{N-1})$  with  $X \in D$ , is the population of strings /solutions/individuals,

$\mathbf{X} = (X_0, X_1, \dots, X_{n-1})$  is a string/solution of length  $n$  with  $X_i$  as a variable,

$x = (x_0, x_1, \dots, x_{n-1})$  is a string/solution with  $x_i$  as a possible instantiation of variable  $X_i$ ,  $x_i \in \{0,1\}$ ,

$p(\mathbf{X}) = p(X_0, X_1, \dots, X_{n-1})$  denotes the  $n$  dimensional probability distribution,

$p(x_0, x_1, \dots, x_{n-1}) = p(X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1})$  denotes a probability of a concrete  $n$  dimensional vector.

The probabilistic model used in BMDA can be formalized by  $M = (G, \Theta)$ , where  $G$  is dependency graph and  $\Theta = (\theta_0, \theta_1, \dots, \theta_{n-1})$  is a set of parameters which are estimated by local conditional or marginal probability for each node/variable of the dependency graph.

A greedy algorithm for building dependency graphs is used. At the beginning, the root node is randomly selected and subsequently the nodes with maximum dependency value are searched among the remaining nodes and joined.

These pairwise dependencies in BMDA are discovered by Pearson's chi-square statistics:

$$\chi_{i,j}^2 = N \left( \sum_{\forall x_i \in \text{Dom}(X_i)} \sum_{\forall x_j \in \text{Dom}(X_j)} \frac{m^2(x_i, x_j)}{m(x_i)m(x_j)} - 1 \right) \quad (4)$$

where  $N$  is the size of parent population and  $m(x_i, x_j)$ ,  $m(x_i)$  resp.  $m(x_j)$  denote the number of individuals in the parent population with concrete values of  $x_i$  and/or  $x_j$ ; if the denominator acquires the zero value, the subsume will be set to zero. These values are stored in the contingency tables. From the theoretical point of view this metric can be seen as statistical testing of hypothesis – for example binary variables  $X_i$  and  $X_j$  are considered to be independent at 95 percent confidence level if  $\chi_{i,j}^2 < 3.84$ . Algorithm BMDA

uses a variant of minimum spanning tree technique to learn a model. However, during the tree construction, if none of the remaining variables can be “rooted” to existing tree, BMDA starts to form additional tree from remaining variables. The final probability distribution is thus a forest distribution (a set of mutually independent dependency trees):

$$p(\mathbf{X}) = \prod_{X_r \in R} p(X_r) \prod_{X_i \in V \setminus R} p(X_i | X_{j(i)}) \quad (5)$$



square is symmetric, and dependencies between the same variables have no sense, the number of transferred tables is reduced to  $2(n^2 - n)$ . Thus the spatial complexity of the transferred data is  $O(n^2)$ .

In comparison with the probabilistic model migration, the migration of individuals used in [24] works with the spatial complexity  $nkN$ , where  $kN$  is the number of migrating individuals. Since the communication overhead in modern interconnection networks depends more strongly on the start-up latency of communication than on a transported message size, we can consider that the communication overhead will be nearly the same for both approaches. Using asynchronous or non-blocking type of migration [11], the communication overhead could be simply overlapped.

The information exchange between the resident and the immigrant island/islands, common to both model combination types, starts with initiation of receive requests. During the receive procedure, the resident island can compute its own contingency tables and the mean fitness value  $F_R$  of the resident population. Next, all computed data are packed into a simple send buffer and sent using non-blocking communication to the neighbor islands.

```

Procedure MakeExchangeIslandInformation();
for ( $i < \text{Neighbour\_count}$ ) do begin
  MPI_IRecv(Receive_buffer $_i$ );
end;

Calculate the mean fitness value  $F_R$  of the resident island;
Calculate resident contingency tables;
Pack resident contingency tables and  $F_R$  into a send buffer;

for ( $i < \text{Neighbour\_count}$ ) do begin
  MPI_ISend(Send_buffer $_i$ );
end;

MPI_WaitAll(Waiting for receiving finish);

Model_Combination();

MPI_WaitAll(Waiting for sending finish);
end

```

Figure 6. The scheme of MPI communication between the resident and the immigrant islands

Now, the resident island waits until all data from all immigrant islands will be delivered. Then, the model combination can start. Finally, the MPI\_WaitAll function is called to synchronize the islands after migration phase.

Besides the described type of communication, the MPI\_Gather [15] operation was used after each generation. During this operation, all necessary information from all processing nodes is collected to compute global statistics including the best global solution, the global mean fitness value, etc.

#### A. Two-phase Model Combination (aBMDA)

Previously proposed method [24] of adaptive learning of probabilistic model uses the two-phase combination of contingency tables, see Fig. 7. For simplicity, let consider only one immigrant island.

The resident chisqr-table is computed from the resident contingency tables, and the immigrant chisqr-table is computed according to unpacked data from receive buffer. Consequently, the probabilistic model composition is started. In the first phase (1) the resident and the immigrant chisqr-tables are combined together using the beta parameter to produce a new chisqr-table. A new dependency graph is created based on the information stored in the learned chisqr-table. The second phase (2) determines the parameters values. A set of parameters  $\Theta'_R$  is calculated using new dependency graph, and the original resident and the immigrant contingency tables. As a result, the new probabilistic model  $M'_R = (G'_R, \Theta'_R)$ , is determined.

```

Calculate Chisqr_Table_Resident $[i,j] = \chi^2_{i_R,j_R}$ 
Unpack immigrant contingency tables and  $F_I$  from receive buffers;
Calculate  $\beta$ ;
Calculate the Chisqr_Table_Immigrant $[i,j] = \chi^2_{i_I,j_I}$ ;
(1) Calculate items of the composed Chisqr_Table $[i,j]$ :
 $\chi^2_{i_R,j_R} \leftarrow \beta \chi^2_{i_R,j_R} + (1-\beta) \chi^2_{i_I,j_I}$ 
Build the new dependency graphs  $G'_R$  according to new Chisqr_Table;
(2) Calculate set of the parameters:  $\Theta_R(G'_R)$ ,  $\Theta_I(G'_R)$  using contingency tables ;
Learning of the parameters:  $\Theta'_R = \beta \Theta_R + (1-\beta) \Theta_I$ 
Compose new resident model:  $M'_R = (G'_R, \Theta'_R)$ 

```

Figure 7. Two-phase model combination (aBMDA)

#### B. One-phase Model Combination (pBMDA)

In the newly proposed parallel pBMDA algorithm the contingency tables of resident and immigrant islands are joined together prior to graph  $G'_R$  building and parameters learning. The model building and parameters setting procedure is then same as in sequential version.

```

Unpack immigrant contingency tables and  $F_I$  from receive buffers;
Calculate  $\beta$  ;
Calculate the joined contingency tables
 $C\_Table_R[i,j] \leftarrow \beta * C\_Table_R[i,j] + (1-\beta) * C\_Table_I[i,j]$ 
Calculate Chisqr_Table $[i,j]$  using new  $C\_Table_R[i,j]$ 
Build the new dependency graphs  $G'_R$  according to Chisqr_Table;
Learning of the parameters:  $\Theta'_R$  according to Chisqr_Table $[i,j]$ 
Compose new resident model:  $M'_R = (G'_R, \Theta'_R)$ 

```

Figure 8. One-phase model combination (pBMDA)

## VII. EXPERIMENTAL RESULTS

We pursued two aims. The first one is a mutual comparison of pBMDA and aBMDA algorithms. The second aim was directed to the investigation of the scalability and the influence of underlying topology of the proposed pBMDA that was not covered by [24], [32].

We investigated two concepts of subpopulation size. The first one described in subsections *B*, *C* works with the fixed size of 256 individuals per island derived partially from our experience and from the experimental results published in [21] for the Quadratic problem. This fixed subpopulation size of each island has been used for the whole range of problem and topology size. The second concept presented in subsection *D* used the phenomenon of limited resources. The whole population of island system was set to 2048 individuals so each subpopulation size was determined by the number of islands (e.g. 256 individuals for each of 8 islands). In pBMDA, truncation-based selection strategy was used, i.e. all individuals were ordered by their fitness value and the better half was used for model building. The truncation-based replacement strategy was also used for the replacement operator, i.e. the new generated solutions (offspring) replace the worse half of the subpopulation. The probabilistic model is built in each generation. The frequency of the model migration is permanent - once per five generations. First stop condition was met after 500 generations; the second condition was activated if there was no improvement in the interval of 50 generations. In all measurements 30 independent runs were carried out.

### A. Specification of Benchmarks

For our experimental study, the well known Quadratics and 3-Deceptive benchmarks with various complexities/sizes and known global optima were used:

$$\text{Quadratic: } f_{\text{Quadratic}}(x) = \sum_{i=0}^{\frac{n}{2}-1} f_2(x_{\pi(2i)}, x_{\pi(2i+1)}) \quad (12)$$

$$\text{where } f_2(u, v) = 0.9 - 0.9(u + v) + 1.9uv$$

$$\text{3-Deceptive: } f_{\text{3-Deceptive}}(x) = \sum_{i=0}^{\frac{n}{3}-1} f_3(x_{\pi(3i)} + x_{\pi(3i+1)} + x_{\pi(3i+2)}) \quad (13)$$

$$\text{where } f_3(u) = \begin{cases} 0.9 & \text{if } u = 0 \\ 0.8 & \text{if } u = 1 \\ 0 & \text{if } u = 2 \\ 1 & \text{otherwise} \end{cases}$$

In our previous paper [31], we have applied the complete set of benchmarks including OneMax, TwoMax, Quadratic and 3-Deceptive problem. The Quadratic problem represents the adequate benchmark that should be solvable just by any BMDA algorithm. The 3-Deceptive task belongs to the hard deceptive benchmark for BMDA and is often used for the testing of BOA algorithms.

### B. Comparison of pBMDA and aBMDA Algorithms

We have compared performance of the newly proposed one-phase aggregation model (pBMDA) with the recently published two-phase one (aBMDA) on eight island configurations, see Fig. 9.

Three basic interconnection topologies are mutually compared. Firstly, we installed the simplest possible topology - the unidirectional ring. Secondly, the islands were connected by Octagon topology [31], and finally, the fully connected topology has been studied.

The success rate curves for Quadratic problem are quite similar but better results were obtained by the one-phase model, see Fig. 9. The one-phase algorithm pBMDA provides higher variable range with 100% success rate on the octagon topology. Whereas aBMDA variant loses its searching capability after 280 variables, the pBMDA variant preserves its searching capability even up to 400 variables. On the contrary, the one-phase model loses its power in case of unidirectional ring topology. The success rate under fully connected topology is comparable for the both variants.

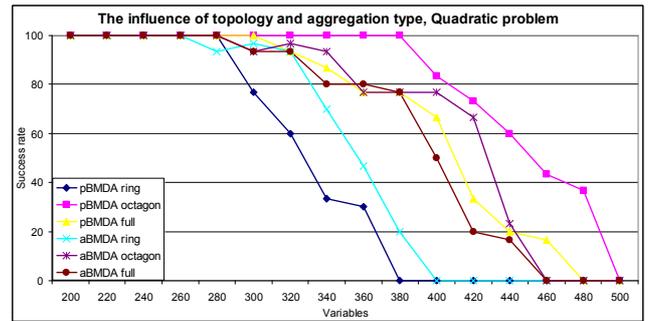


Figure 9. The dependency of success rate on the aggregation type and the interconnection topology.

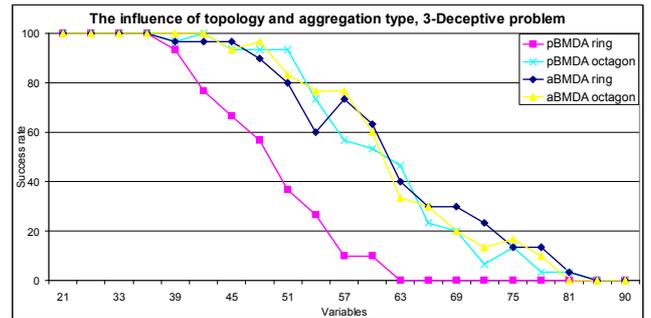


Figure 10. The dependency of success rate on the aggregation type and the interconnection topology.

Similar results were obtained for 3-Deceptive benchmark, see Fig. 10 but the success rate curves are rather compressed. The success rates of both variants of model combination on octagon topology and fully connected (not shown for lucidity of the Fig. 10) are comparable, such that it is not possible to choose simply the best one. The aBMDA provides better results with the simple ring topology.

### C. Scalability of One-phase Model

In the following set of experiments, the scalability of the proposed method for fixed subpopulation size of 256 individuals has been investigated. The influence of the number of islands on success rate is demonstrated in Fig. 11. It can be recognized that the success rate grows with the increasing number of connected island. The addition of the other island to 1-island algorithm expands the area of successfully solved benchmark instances approximately twice. The addition of another two islands also scales the benchmark size very well. Naturally, with higher number of islands, the improvement of success rate tends to decrease.

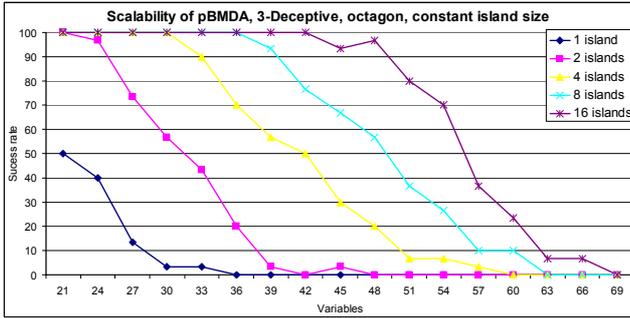


Figure 11. The dependency of success rate on the number of islands with fixed size of subpopulation (256 individuals per island).

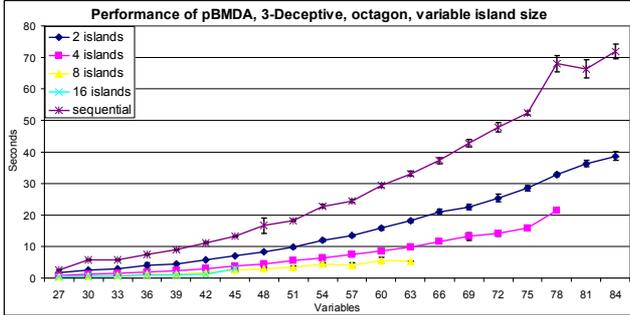


Figure 12. The execution time of pBMDA algorithm for varying number of islands with fixed subpopulation size.

The real execution times were also measured for increasing problem size. Fig. 12 presents the mean values and standard 0.95 confidence interval of execution time from runs that achieved global optima. The curves of execution times are very similar, it is caused by a very small overhead of model aggregation and data exchange between islands that grows very gently with the number of islands. The evolution will consume almost the same time for both one or 16 islands, but naturally it will produce much better results. Consequently, we can utilize all available computational resources to achieve the highest possible success rate.

### D. Scalability of One-phase Model with limited resources

This set of experiments examines the scalability using limited total resources, such that the total population of 2048 individuals is uniformly distributed over varying number of islands. Fig. 13 illustrates the success rate according to the

number of islands. The best results were produced by one-island (sequential) BMDA with the population of 2048 individuals. Similar results were obtained with two and four islands (with adequate subpopulation size of 1024 and 512 individuals). But it is evident that the 16-islands topology is not able to offer sufficient genetic material to produce good residential model. Its subpopulation size is too small.

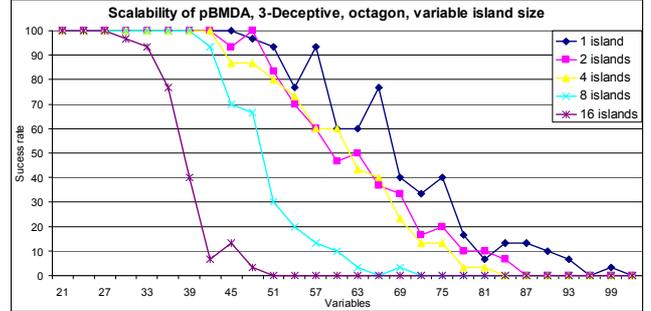


Figure 13. The dependency of success rate on the number of islands with limited total resources.

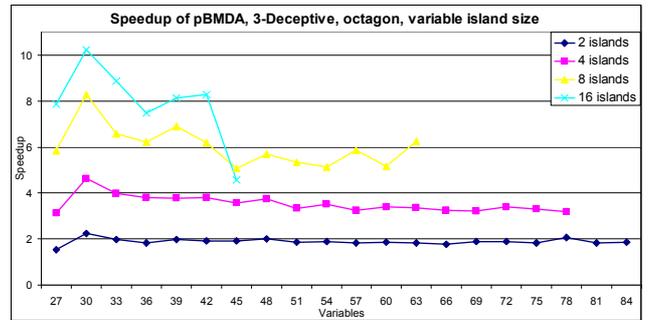


Figure 14. The speedup of pBMDA with the limited resources.

Fig. 14 demonstrates the speedup towards the sequential variant under limited resources calculated from runs that achieve global optima. The speedup of two and four islands attacks the maximal possible values (2 and 4 respectively). Actually, a super-linear speedup can be observed in some benchmark instances. The speedups for 8 and 16 islands do not reach the theoretical values due to low population size.

## VIII. CONCLUSIONS

In the paper, the problem of probabilistic model aggregation used for efficient parallelization of BMDA algorithm was investigated. In Section III, a short survey of references dealing with the theory and practice in this area is presented. The published results of research are mostly oriented on Bayesian network combination and very often are focused on various expert systems from medicine area. Much of this work is rather theoretical without a possibility to utilize its result for concrete applications.

The reference [29] is an exception to the rule, but its topic is focused on the aggregation of particular probabilistic models with relatively small intersection of variables. A perspective theoretical tool is published in [30]. The concept

of extraction of independency model from each Bayesian graphical model and their consequential union or intersection appears to be very promising.

The first experimental works have been arranged for island topologies with fixed subpopulation size. From Fig. 9 it flows that the one-phase algorithm pBMDA outperforms aBMDA for Quadratic problem on the octagon topology. In case of 3-Deceptive problem, the results are comparable due to difficulty of the solved benchmark (see Fig.10).

From Fig. 11 and Fig. 12 it is evident that using the concept of fixed subpopulation size, it is beneficial to use as much as possible number of islands to achieve higher value of success rate. But it is clear that execution time does not evidently increase since the process of model aggregation and data exchange is appropriately overlapped.

The second experimental works have used the concept of limited resources. From Fig. 13 it can be derived that the optimal number of island lies between 2 and 4. The speedups of these configurations are close to theoretical limits, moreover in some cases, the super-linear speedup occurs. The 8-island configuration provides still comparable success rate on the 100% level. The speedup does not meet the theoretical value, but it provides very good values about 6.

It can be concluded that the aBMDA is more suitable for unidirectional ring, generally for sparser interconnection topologies. On the other hand, the more complex topologies were advantageous for the newly proposed pBMDA.

In the future, we have an intension to test the approach published in [26], [30] for the aggregation of probabilistic model utilizing Bayesian networks. The great challenge is to propose an aggregation technique which is based on a partial learning of the residential probabilistic model by promising parts of immigrant models only.

#### REFERENCES

- [1] C. W. Ahn, D. E. Goldberg, and R. S. Ramakrishna, Multiple-Deme Parallel Estimation of Distribution Algorithms: Basic Framework and Application. Illigal Report No. 2003016, May, 2003.
- [2] E. Alba, and J. M. Troya, A Survey of Parallel Distributed Genetic Algorithms. *Complexity*, 4: 303-346, 1999.
- [3] S. Baluja, Population-Based Incremental Learning: A method for Integrating Genetic Search Based Function Optimization and Competitive Learning. Technical report CMU-CS-94-163, Carnegie Mellon University, 1994.
- [4] E. Cantu-Paz, Efficient and Accurate Parallel genetic algorithm. Kluwer Academic Publishers, 2000.
- [5] L. delaOssa, J. A. Gámez, and J. M. Puerta, "Migration of probability models instead of individuals: an alternative when applying the island model to edas," in *Parallel Problem Solving from Nature - PPSN VIII*, volume 3242 of LNCS, pp. 242-252. Springer, 2004.
- [6] L. delaOssa, J. A. Gámez, and J. M. Puerta, "Improving model combination through local search in parallel univariate EDAs," *Evolutionary Computation*, Vol. 2, pp. 1426-1433.
- [7] D. E. Goldberg, K. Sastry 1, and Xavier Llorà, Toward routine billion-variable optimization using genetic algorithms. *Complexity* 12: 2007.
- [8] D. Heckerman, D. Geiger, and M. Chickering, Learning Bayesian networks: The combination of knowledge and statistical data. Technical Report MSR-TR-94-09, Microsoft, Redmond, WA, 1994.
- [9] M. Höhfeld, and G. Rudolph, "Towards a theory of population-based incremental learning", in *Proceedings of the 4th International Conference on Evolutionary Computation*, pp. 1-5, IEEE Press, 1997.
- [10] P. Larrañaga, and J. A. Lozano, *Estimation of Distribution Algorithms*. Kluwer Academic Publishers, London 2002, ISBN 0-7923-7466-5.
- [11] S. C. Lin, W. F. Punch, and E. D. Goodman, "Coarse-grain parallel genetic algorithms: categorization and a new approach," in *Sixth IEEE Conference on Parallel and Distributed Processing*, pp 28-37, IEEE Press, Piscataway, NJ, 1994.
- [12] F. G. Lobo, C. F. Lima, and H. Martires, "Massive parallelization of the compact genetic algorithm," in R. Ribeiro et al., editors, *Proceedings of the International Conference on Adaptive and Natural Computing Algorithms*, pp 530-533. Springer, 2005.
- [13] J. M. Marin, K. Mengersen, and C. P. Robert, "Bayesian Modelling and Inference on Mixtures of Distribution", in *Handbook of Statistics*, 25, 16, 459-507.
- [14] A. Mendiburu-Alberro, "Parallel implementation of estimation of Distribution Algorithms based on probabilistic graphical models. Application to chemical calibration models," PhD thesis, the University of Basque Country, Donostia-San Sebastian, January 2006.
- [15] MPI-2 Extension to message-Passing Interface, Message Passing Interface Forum, 2003, Document available at url: <http://www.mpi-forum.org/docs/mpi2-report.pdf>
- [16] H. Mühlenthein, "The equation for response to selection and its use for prediction," in *Evolutionary Computation* 5(3), pp. 303-346, 1997.
- [17] J. Ocenasek, "Parallel Estimation of Distribution Algorithms," PhD. Thesis, Faculty of Information Technology, Brno University of Technology, Brno, Czech Rep., pp. 1-154, 2002.
- [18] J. Ocenasek, J. Schwarz, and M. Pelikan, "Design of Multithreaded Estimation of Distribution Algorithms," in: *Genetic and Evolutionary Computation Conference*. Springer, Berlin, pp. 1247-1258, 2003.
- [19] M. Pelikan, D. E. Goldberg, E. Cantú-Paz, Linkage problem, distribution estimation, and Bayesian networks. (IlliGAL Report No. 98013, University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, Urbana, Illinois), .1998.
- [20] M. Pelikan, D. E. Goldberg, F. Lobo, A survey of optimization by building and using probabilistic models. (IlliGAL Report No. 99018, University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, Urbana, Illinois) (1999)
- [21] M. Pelikan, K. Sastry, E. Cantú-Paz, Bayesian Optimization Algorithm, Population Sizing and Time to Convergence. (IlliGAL Report No. 2000001, Illinois Genetic Algorithm Laboratory, University of Illinois at Urbana-Champaign, p 13), 2000.
- [22] Pelikan M, Sastry K, Goldberg DE (2001) *Evolutionary Algorithms + Graphical Models = Scalable Black-Box Optimization*. IlliGAL Report No. 2001029, University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory, Urbana, Illinois
- [23] Y. Xiang, and T. Chu, "Parallel Learning of Belief Networks in Large and Difficult Domains," in *Data Mining and Knowledge Discovery*, vol.3, pp. 315-339, 1999.
- [24] J. Jaros, J. Schwarz, Parallel BMDA with Probability Model Migration, In: *Proceeding of 2007 IEEE Congress on Evolutionary Computation*, Singapore, SG, IEEE CS, 2007, pp. 1059-1066, 2007.
- [25] R. T. Clemen, R. L. Winkler, Combining probability distributions from experts in risk analysis. In *Risk Analysis*, 19(2):187-203, 1999.
- [26] I. Matzkevitch, B. Abramson, The topological fusion of Bayes set. In *Proc. UAI'92*, pp. 191-19 8, 1992.
- [27] D. M. Pennock, M. P. Wellman, Graphical representations of consensus belief. In *Proc. UAI'99*, pp.531-540,1999.
- [28] M. Pedrito, Reid, U. Chajewska. Aggregating Learned Probabilistic Beliefs. In *Proc. of UAI'2001*, pp. 354-361.
- [29] Ch. Jiang, T-Y. Leong, K.-L. Poh, PGMC: A Framework for Probabilistic Graphical Model Combination, *Proc. of AMIA 2005*, pp.370-374
- [30] J. Sagrado, S. Moral. Qualitative combination of Bayesian networks. *Int. J. Intell. Syst.* 18(2): 237-249 (2003).
- [31] F. Karim, A. Nguyen, An Interconnect Architecture for Networking Systems on Chips. *IEEE Micro*, Sept. – Oct. 2002, pp.36-45.
- [32] J. Schwarz, J. Jaros, Parallel Bivariate Marginal Distribution Algorithm with Probability Model Migration, In *Linkage in Evolutionary Computation*, Berlin / Heidelberg, DE, 2008, p. 3-23, ISBN 978-3-540-85067-0
- [33] M. Tomassini, *Spatially Structured Evolutionary Algorithms*, Springer-Verlag, Berlin-Heidelberg, 2005, ISBN 10-540-24193-0