Ethane: A Heterogeneous Parallel Search Algorithm for Heterogeneous Platforms

Sunday, the 5th of June, 2011

Julián J. Domínguez julian@lcc.uma.es
Enrique Alba eat@lcc.uma.es

Departamento de Lenguajes y Ciencias de la Computación
E. T. S. I. Informática

Index

Introduction

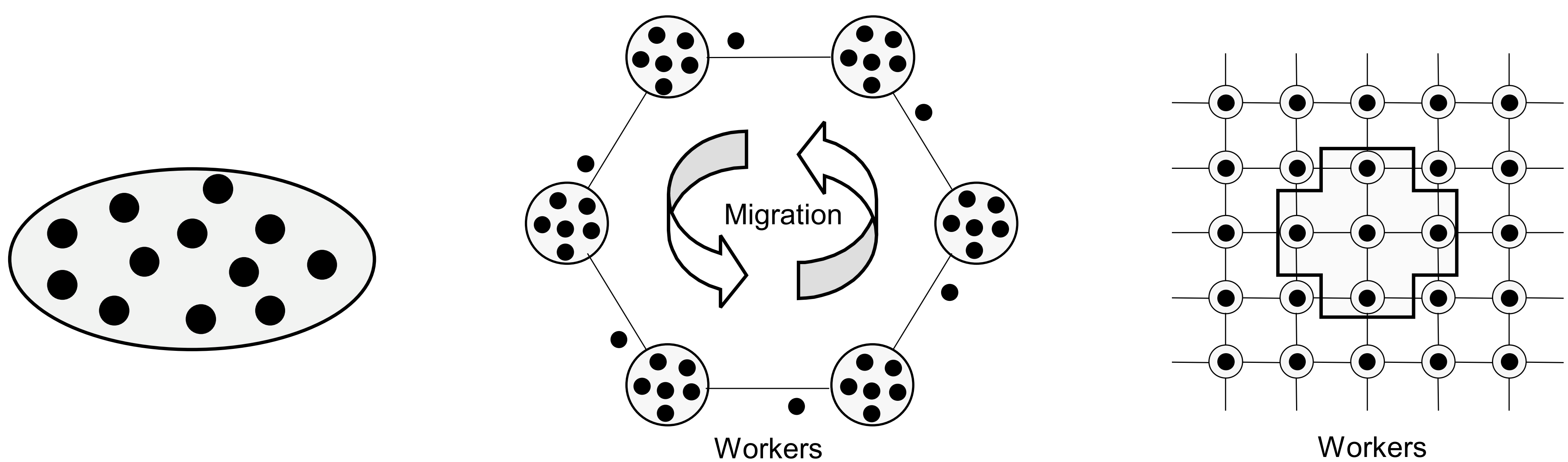
- Metaheuristics provide a fast and efficient way for solving problems, and in many cases parallelism is necessary. There exists many parallel models.
- In a common modern lab we can usually find many different hardware architectures.
- It has been proven that heterogeneous resources can be used efficiently with standard parallel models, but few works exists about the design of a specific parallel model for such a heterogeneous environment.
- We present Ethane, a heterogeneous parallel search algorithm designed for its execution on heterogeneous hardware environments.
- We also propose HydroCM (HydroCarbon inspired Metaheuristics), a general schema based in hydrocarbons for shaping a family of parallel heterogeneous metaheuristics.

Decentralized and Heterogeneous PMs

- There exist many parallel implementations of metaheuristics.
- We will focus in two of the more common families of metaheuristics: Evolutionary Algorithms (EAs) and Local Search Metaheuristics (LSMs):
  - Evolutionary Algorithms: EAs are population based methods; a random population is enhanced through a Nature-like evolution process. EAs provide a good exploration of the search space, they are also called exploration-oriented methods.
  - Local Search Metaheuristics: LSMs use only one candidate solution which is enhanced by moving through its neighborhood replacing the candidate solution. LSMs are able to find local optima so they are called exploitation-oriented methods.
- Many different parallel models have been proposed for each method, we present the more representative ones.

Parallel EA Models

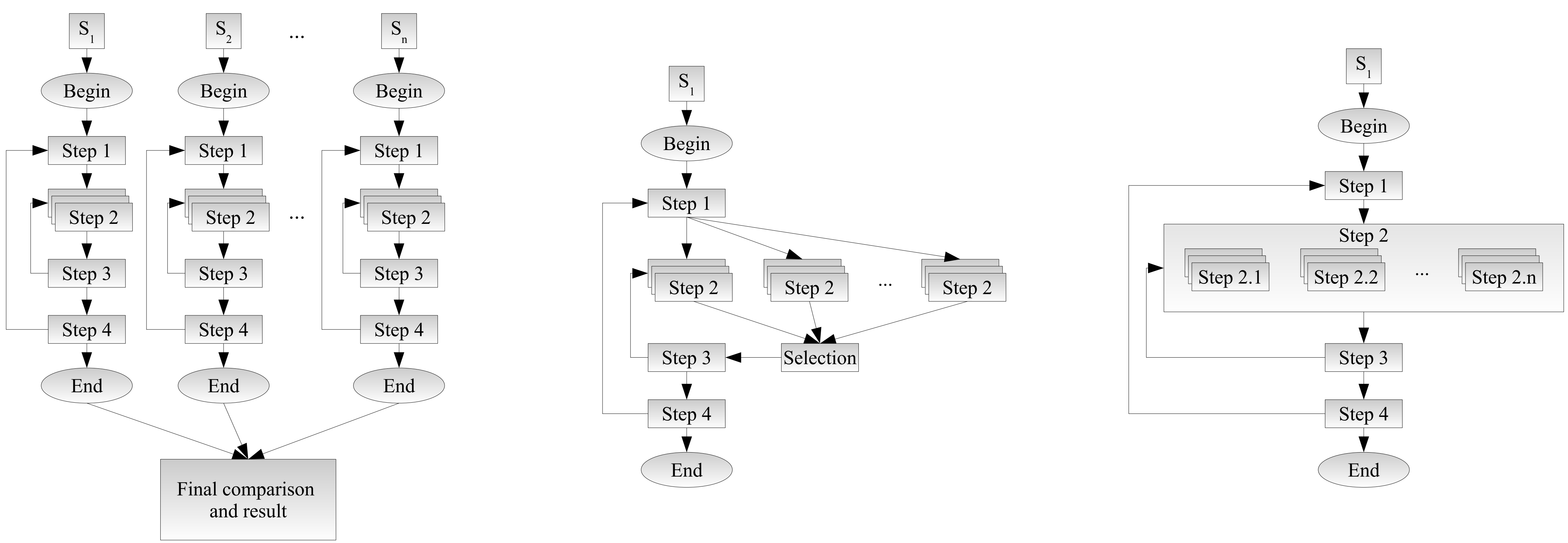
- There exist many parallelization schemas for EAs, but we are going to focus in structured populations, which leads to a distinction between:
  - Distributed EAs: the population is divided into a number of islands running an isolated instance of the EA. Some individual are exchanged following a migration scheme. Few sub-algorithms, loose coupling.
  - Cellular EAs: there exists only one population structured into neighborhoods. An individual can only interact with its neighbors. The structure of the neighborhood leads to different behaviors. Many sub-algorithms, tight coupling.



(a) (b) (c)

Parallel LSM Models

- Three parallel LSM models are specially extended in the literature:
  - Parallel multistart model: several independent instances are run simultaneously.
  - Parallel moves model: a kind of master-slave architecture where all the slaves perform a move and the master selects one of them.
  - Move acceleration model: The quality of each candidate is evaluated in a parallel centralized way.



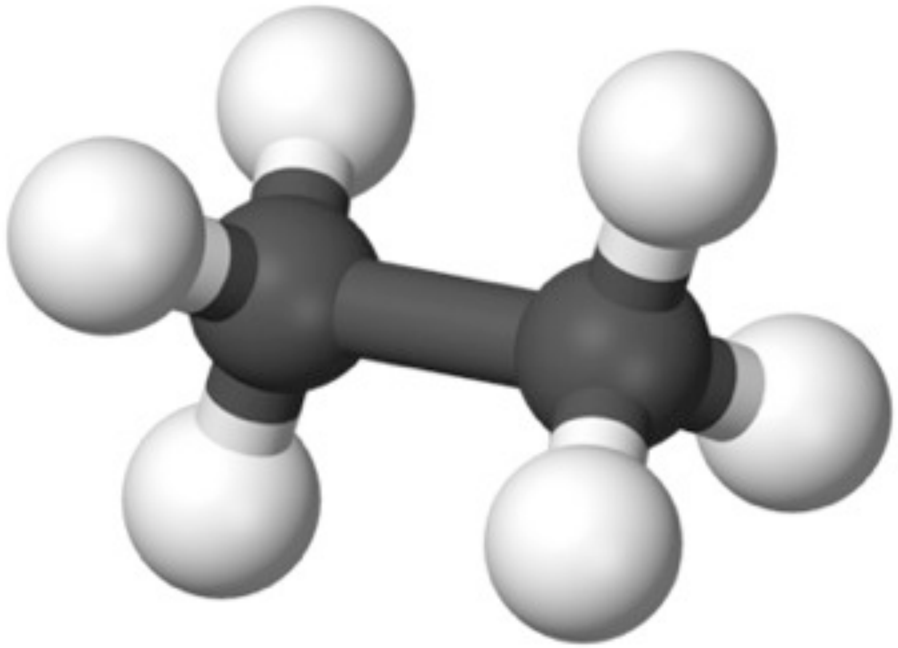
(a) (b) (c)

Achieving the Heterogeneity

- We can modify the behavior of an algorithm by changing the search features between sub-algorithms, obtaining what we call a parallel heterogeneous metaheuristic.
- We can classify parallel heterogeneous metaheuristics in four levels depending on the source of heterogeneity:
  - Parameter level: the same algorithm is used in each node, but different parameters are used in one or more of them.
  - Operator level: heterogeneity is achieved using different mechanisms for exploring the search space, such as different operators.
  - Solution level: a different encoding is used for the solutions in each node.
  - Algorithm level: each component can run a different algorithm.
- The algorithm proposed in our work is an algorithm level parallel heterogeneous metaheuristic, which is based in ssGA and SA.

Description of the Model

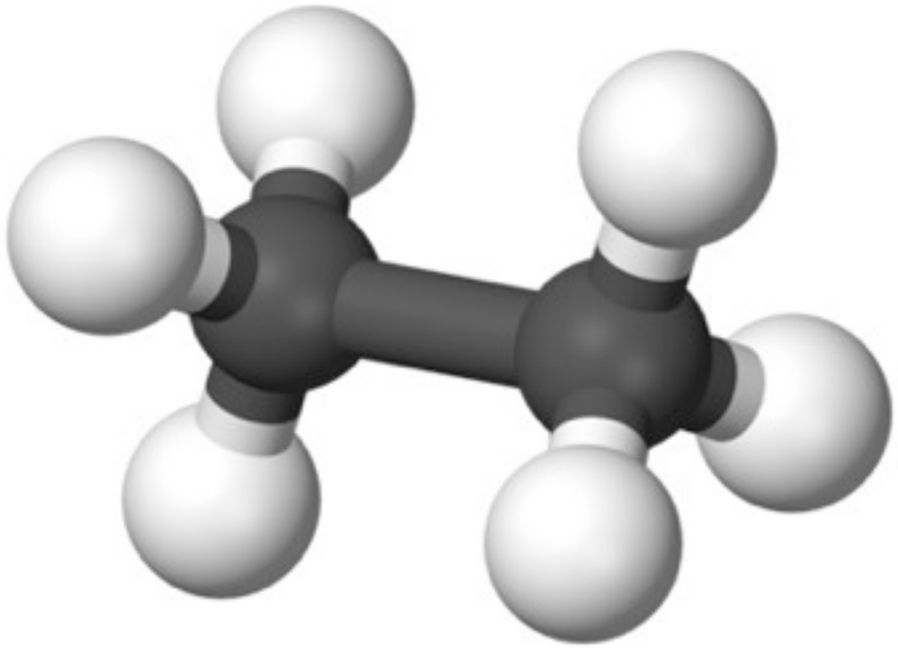
- Ethane is a Nature inspired heterogeneous parallel search algorithm designed for its execution in a heterogeneous hardware platform.
- We propose a communication schema where most of the communication load lays over the fastest nodes, and the slowest ones are placed as *slaves*.



- Ethane is inspired by the chemical compound of the same name, which consists of 6 hydrogen atoms and 2 carbon atoms, where carbon atoms are bond together and 3 hydrogen atoms are bond to each of them.
- We propose the same schema using two basic algorithms resembling different atoms, and communication channels resembling chemical bonds.

Description of the Model

- We have implemented two instances of Ethane:
  - Ethane G: ssGA sub-algorithms are assigned to central nodes and SA islands are assigned to the *slave* nodes.
  - Ethane S: SA islands are placed in the central nodes while ssGA components are placed as *slaves*.

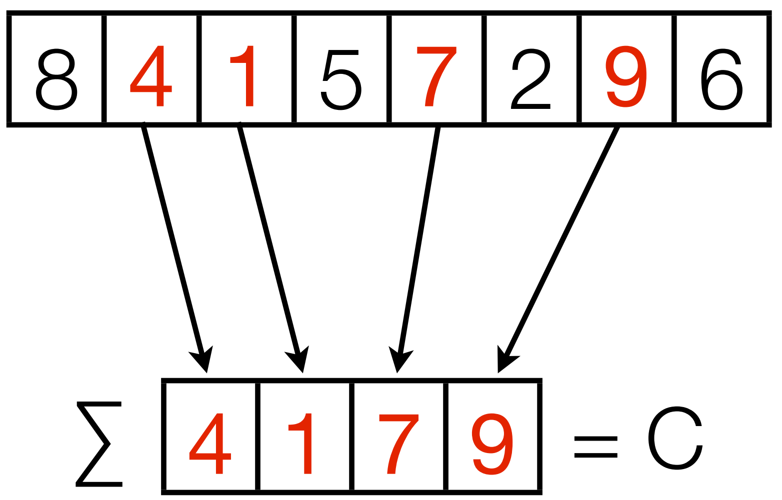


- With this schema the most of the communication falls on the central nodes, which are provided with the fastest hardware, moving some of the load out of the slowest nodes.
- From this inspiration it arises HydroCM as a generic model for a complete family of parallel heterogeneous metaheuristics based on the different structures of hydrocarbons.

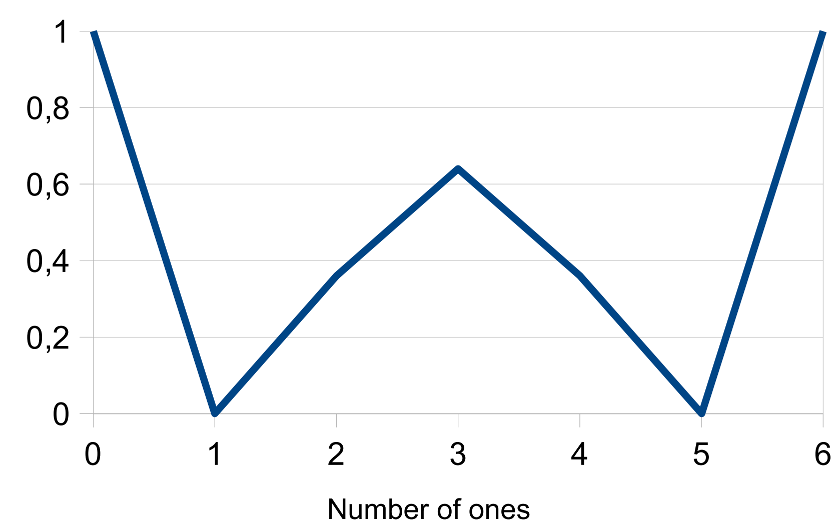
Benchmark Problems

- We have used two problems in the analysis: the Subset Sum Problem (SSP) and the Massively Multimodal Deceptive Problem (MMDP) with 6 bits.
  - SSP: consist in finding a subset of values from a set of integers such that their sum approaches a constant C without exceeding. Our instance consists of 2048 random integers in the range $[0...10^4]$ from a Gaussian distribution.
  - MMDP: is a deceptive problem (make the algorithm converge to wrong regions of the search space). A binary string encodes k 6 bit sub-problems which contribute with a partial fitness depending on its number of 1's (unitation). We have used an instance with 150 bits so that the global optimum is $k = 25$.

SSP



MMDP



Parameters and Platform

- For every ssGA population, the parameters are: a size of 64 individuals, a crossover probability of 0.8 and a mutation probability of 4 divided by the chromosome length.
- For SA we used the same mutation probability.
- We have chosen a migration frequency of 50 for all the models, migrating 1 individual in all cases.
- We have used 8 different machines:
  - 2 Intel Core 2 Quad Q9600 @ 2.66GHz with 4GB of RAM.
  - 6 Intel Pentium 4 @ 2.4GHz with 1GB of RAM.
- All of them are managed by a GNU/LINUX OS, but different distributions. Since the algorithms are implemented in JAVA, all of them run the 1.6.0_0 version of the Java Virtual Machine (JVM).

Tests and Analysis

- For our analysis we have used the presented instances of Ethane, a ssGA unidirectional ring (with the computers placed in a kind of MaxSumSort), and the panmictic versions of SA and ssGA.
- We have analyzed three performance measures, being numerical effort, total run time and speedup, as well as the evolution of the fitness.
- As we do not know the statistical distribution of the data, they have been statistically compared with Mann-Whitney U test.

Numerical Effort

Algorithm	SSP		MMDP6	
	Average	Std. Deviation	Average	Std. Deviation
Ethane G	**146418**	**174433**	1572735	919691
Ethane S	202815	198696	**708231**	**430353**
ssGA Ring	214824	239125	786583	805837
Panm. ssGA	179792	175177	*	*
Panm. SA	81737	93627	*	*

- Ethane performed better than the panmictic algorithms for both problems. For the SSP, both Ethane versions performed numerically better than the reference ssGA ring, and one of the instances (Ethane S) performed better even for the MMDP.
- Also the standard deviation is better in our algorithms and so their behavior is more robust.
- All the differences in behavior are statistically significant.

Total Run Time

Algorithm	SSP		MMDP6	
	Average	Std. Deviation	Average	Std. Deviation
Ethane G	**5318**	**6226**	9195	4942
Ethane S	7155	6922	**3052**	**1546**
ssGA Ring	7453	8107	3194	3380
Panm. ssGA	30008	29387	*	*
Panm. SA	13300	15443	*	*

- Ethane clearly reduced the time needed to find the solution with respect to the panmictic algorithms for both problems. For the SSP, both Ethane versions also performed better than the reference ssGA ring; for the MMDP Ethane S outperformed the reference model.
- Again, the standard deviation is better in our algorithms, showing a more stable behavior.
- All the differences in the behavior are statistically significant.

Speedup

- The speedup represents the ratio between sequential and parallel average execution times. The algorithms run in the single and multi-processor platform must be exactly the same, and the reference processor is the fastest one (standard defined for heterogeneous systems).

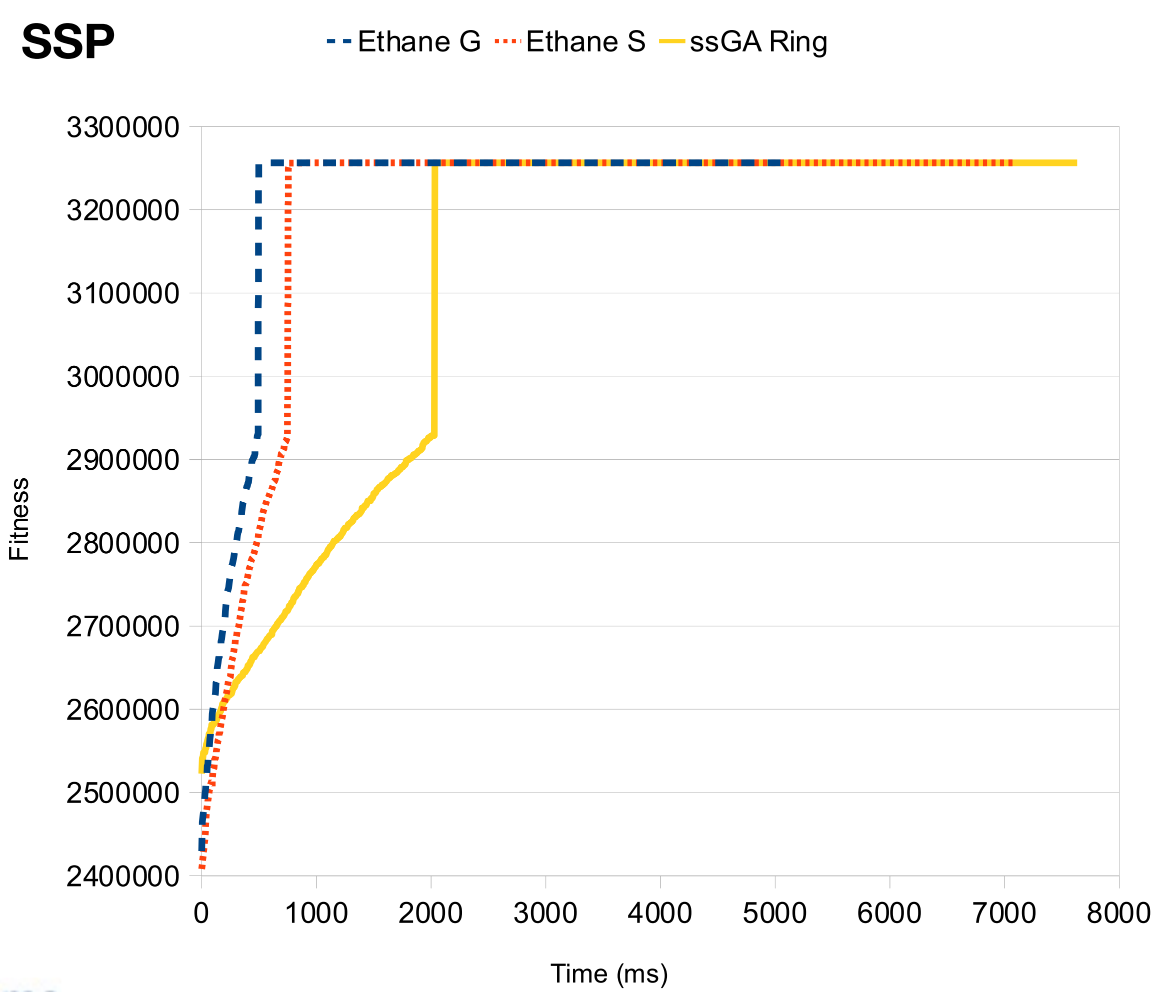
Algorithm	SSP		MMDP6	
	Avg. time	Speedup	Avg. time	Speedup
Ethane G	**15995**	**3.00x**	41943	4.56x
Ethane S	17817	2.49x	**20627**	**6.76x**
ssGA Ring	18137	2.43x	21227	6.64x

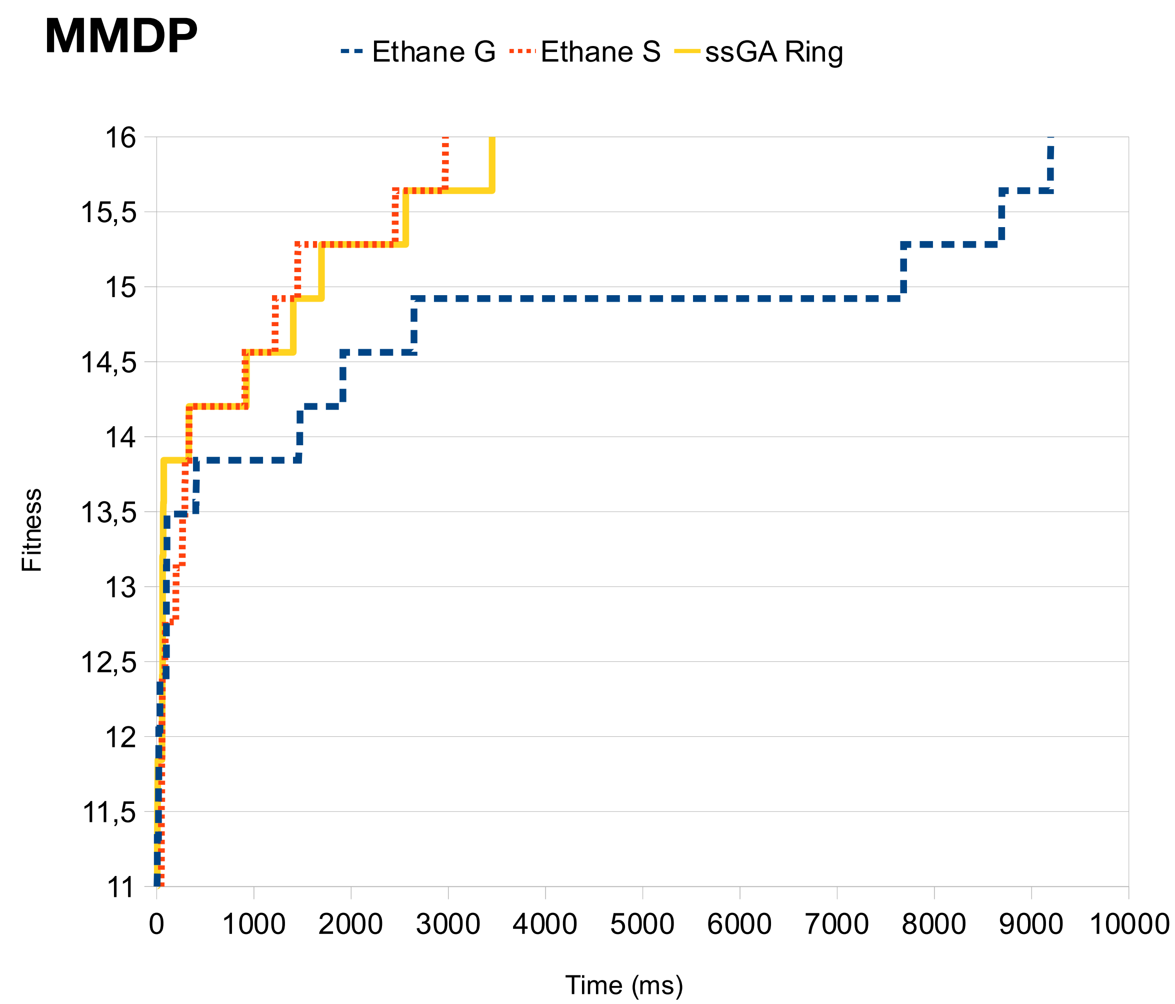
- Both versions of Ethane have obtained a better speedup than the ssGA for the SSP, but only Ethane S has achieved a better speedup for the MMDP.

- The speedup for the MMDP was quite good although lineal speedup was not achieved. In the case of SSP, Ethane G and S have not showed a very good speedup, and ssGA has showed even a worse speedup. This fact could be explained by the huge difference among the power of the different hardware configurations used.

Evolution of the Fitness

- In the case of SSP both versions of Ethane clearly outperforms the ssGA ring.
- For MMDP, Ethane S performed better than Ethane B, and was even able to converge faster than the ssGA ring.





Conclusions and Future Work

- We have presented a new heterogeneous parallel search algorithm based on the structure of ethane.
- We have also shaped a general heterogeneous model based on the structure of hydrocarbons.
- The tests have shown that our model outperforms the reference ssGA ring in terms of run time and numerical effort, and is even able to find solutions in a more robust manner.
- As future work we propose to extend the HydroCM model, as well as to assess its performance with different configurations and real-life applications.
- Our goal is to turn an often undesired heterogeneous platform into a useful platform for (specialized) algorithms.

http://neo.lcc.uma.es

Thanks for your attention. **Questions?**