

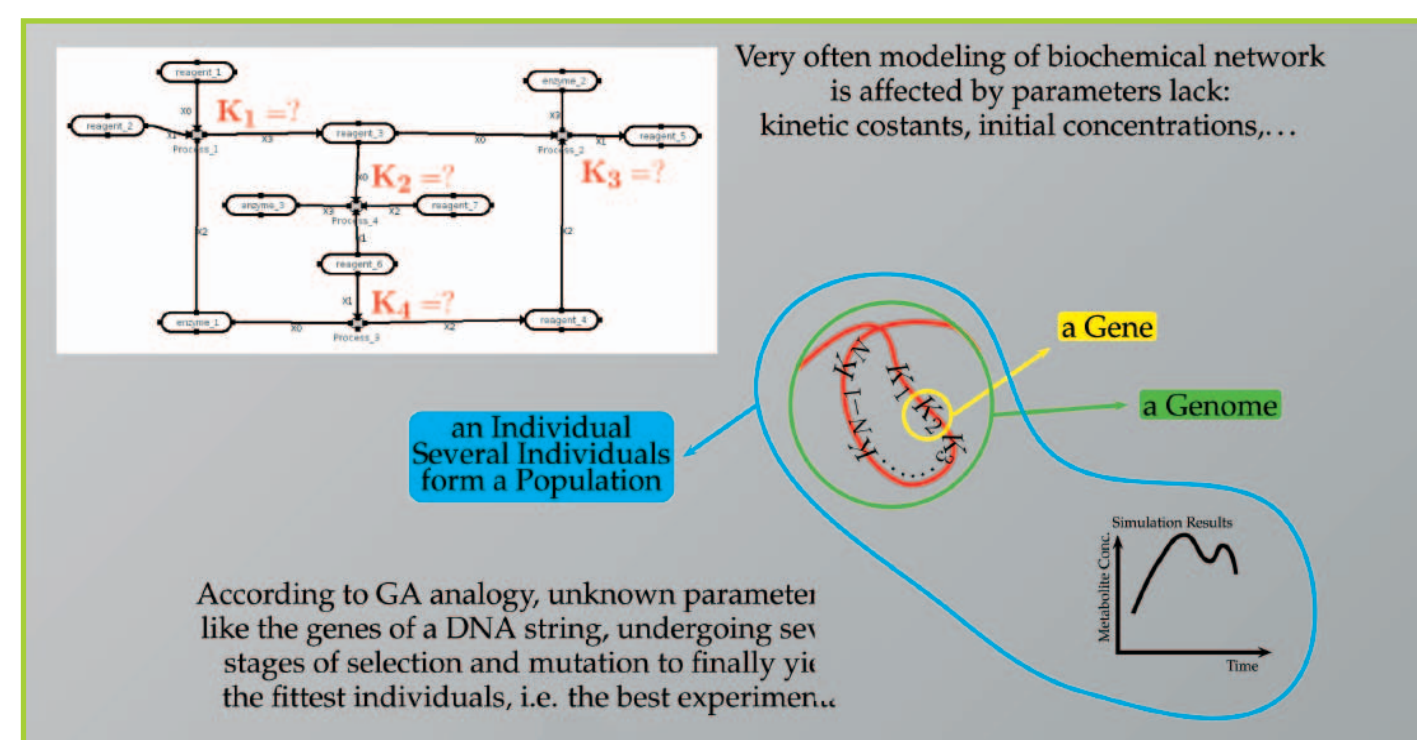
Single - and double - stage parallel implementations of genetic algorithms for parameters estimate of metabolic pathways

Advances in genomics and -omic sciences have begun to produce unprecedented amounts of data that await for analysis and interpretation. Reliable explanations of how processes are regulated require an accurate modeling approach at the systems level; quite often these models of biochemical networks rely on several unknown parameters which need to be estimated. Estimate of model parameters from experimental data remains a bottleneck for major breakthrough in Systems Biology; it consists in the solution of an inverse problem which requires the use of an efficient optimization algorithm. The genetic algorithm (GA) is a widely known method which yield reliable values for model parameters with a large computational demand.

Our aim is to develop a parallel implementation for parameter estimate based on GA to fully deploy the large computational power of a modern grid such as that set up at the Enea Portici site. Our implementation relies on:

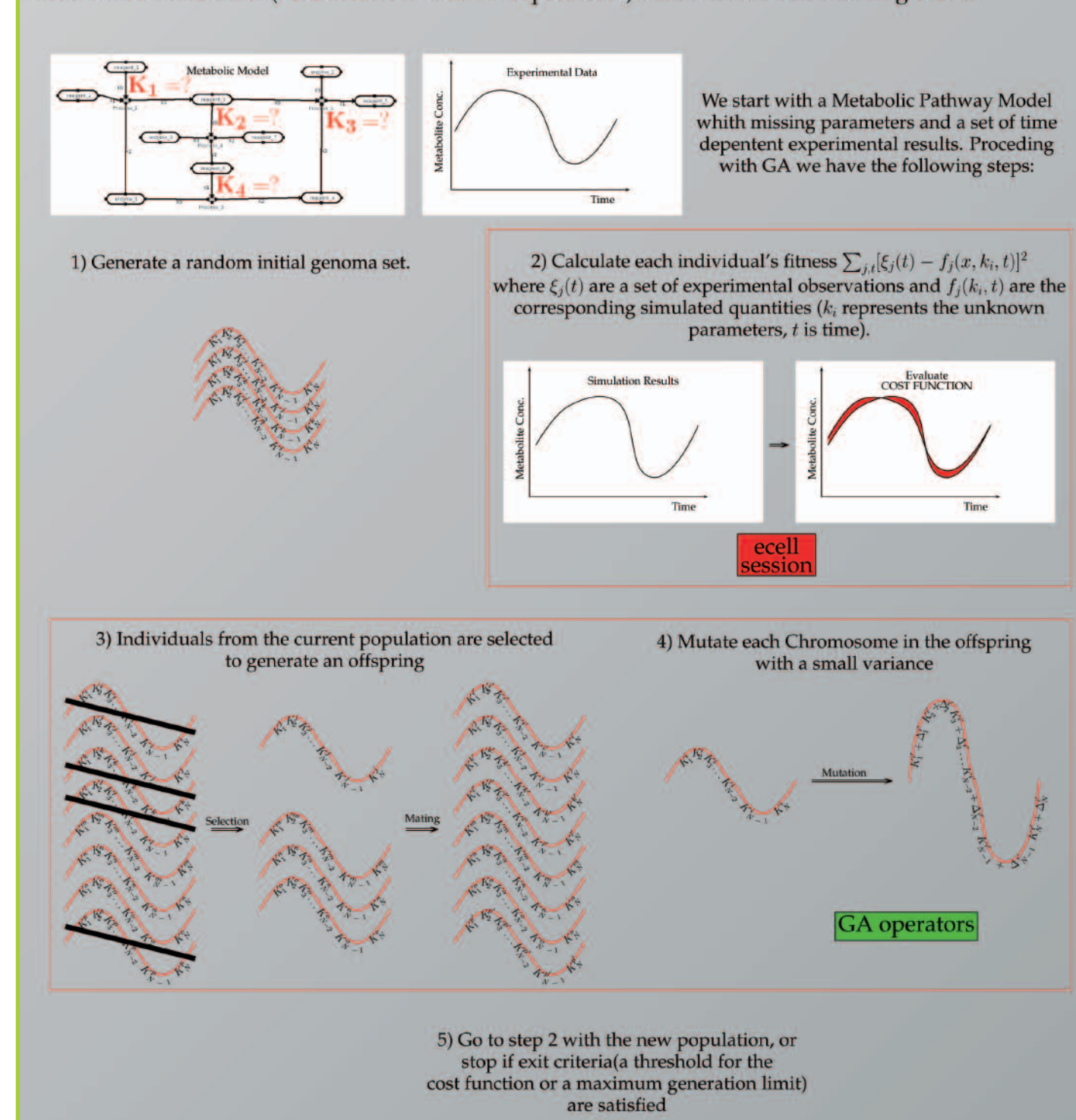
- Ecell software from Keio University (Japan) for simulations of biochemical networks
- LSF - load sharing facility - a job scheduler for (multi) cluster (SGE and Globus are supported as well)

The genetic algorithm: Introduction

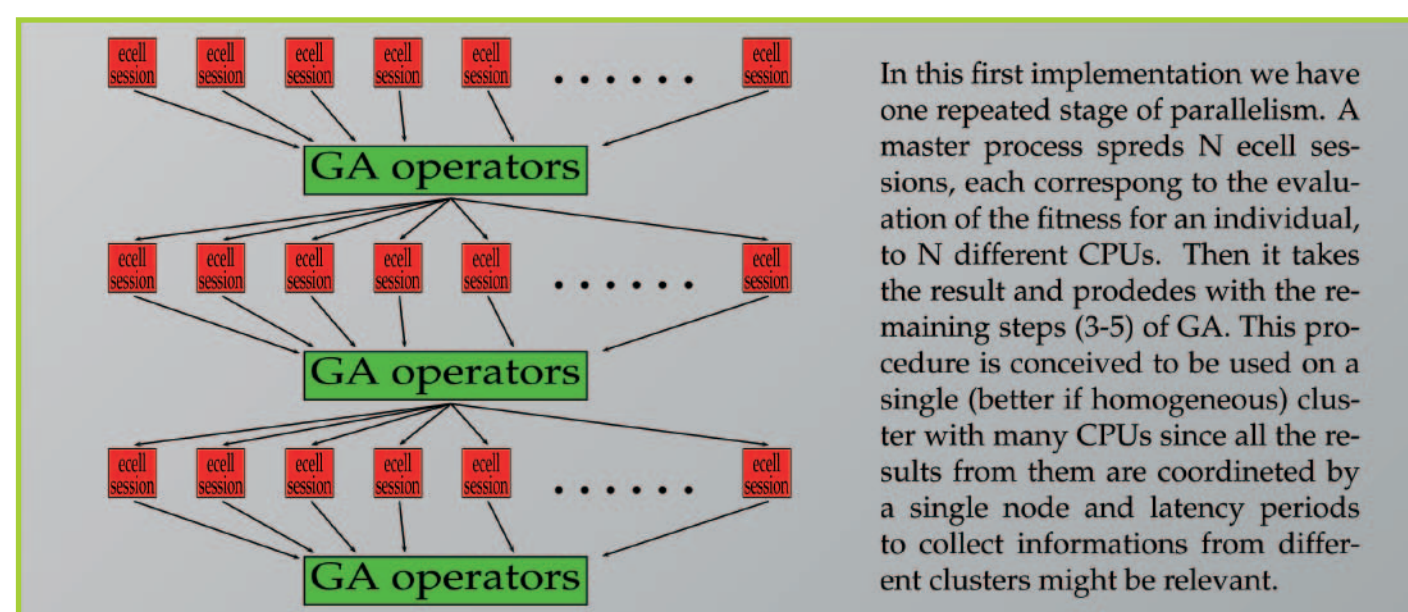


The genetic algorithm: Standard Operating Procedure

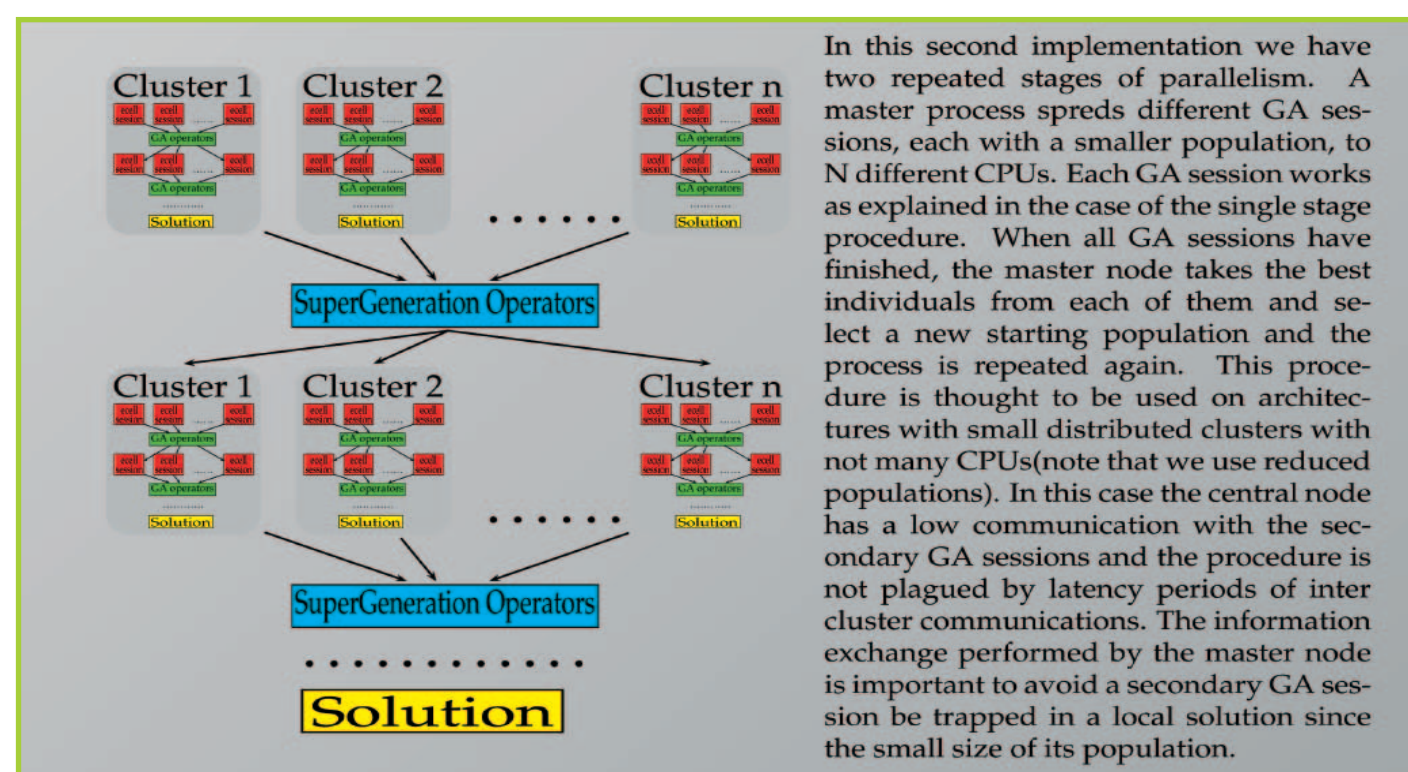
In the following picture the standard GA procedure is shown. In order to efficiently parallelize GA we determined some units ("ecell session" and "GAoperators") which can be run on a single CPU:



Parallel implementation (single stage)

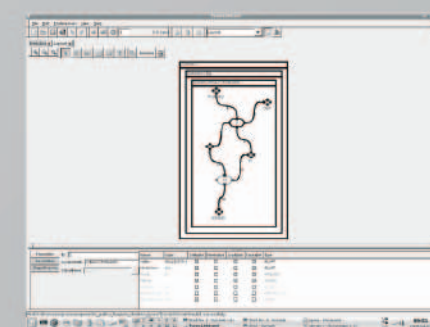


Parallel implementation (double stage)

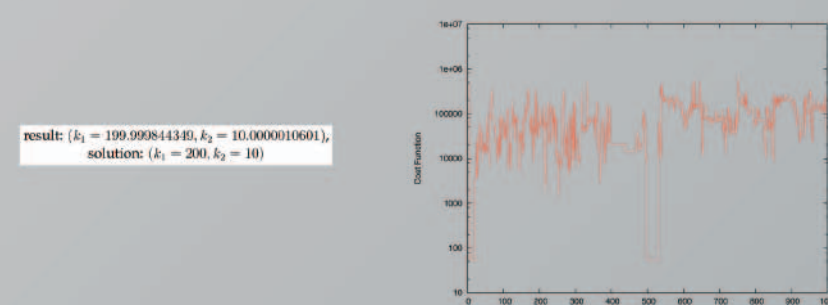


Examples

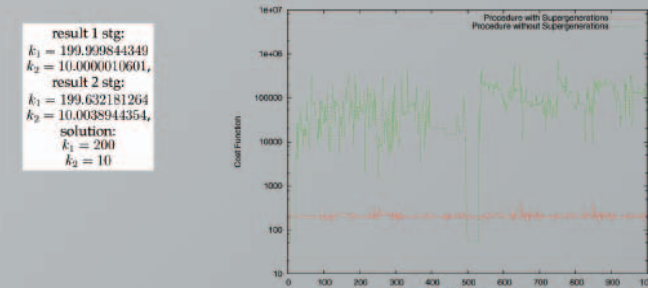
We used our implementations for GA in the case of the Circadian Cycle metabolism. We used ecell3-model-editor, which is an Ecell tool, for drawing the biochemical network model:



This is the evolution of the cost function (≈ 100 CPUs, generations=1000, population=100, 10^5 total individuals):



Comparison between evolution of the cost function of single and double stages procedures:



Although the double stage procedure yields good results, the single stage one usually performs better. That is mainly due to the use of a larger population which is a very important parameter in GA.