

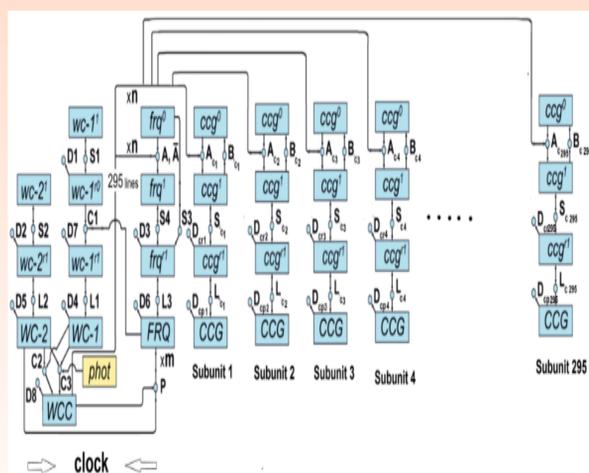
# Parallel strategies for identifying genetic networks describing the biological clock using GPUs

Ahmad Al-omari, Thiab Taha, Jonathan Arnold, Heinz-Bernd Schuttler

Institute of Bioinformatics, The University of Georgia, Athens, GA 30602, aomari@uga.edu

## Introduction

A graphics processing unit (GPU) offers a solution to a very important and fundamental problem of interest to many researchers, a problem that would be prohibitive to solve without the technology of GPUs. The problem is how the biological clock controls the rhythms of ~2400 genes in the genome (with 11,000 genes) of a model system, the filamentous fungus, *Neurospora crassa* (Dong, et al., 2008). Ultimately, we want to be able to predict and hence understand the dynamics of all of these genes and their products in a genetic network describing how the clock functions. The problem of identifying such large genetic networks is beyond the current capability of the largest and fastest serial computers in existence. In order to overcome this problem, a parallelization strategy that is described below will help us to begin to chip away at the entire circadian network, a network that is found from bacteria to humans (Lakin-Thomas, et al., 2011)



**Figure 1.** The genetic network to be solved by the GPUs. The subunits are independent from each other but on the clock. Solving this huge network is beyond the fastest computer in existence.

## Methods

### Family of Parallelization Strategies Flowchart

CPU uses Monte-Carlo simulation to:  
 Generate a set of parameters for solving the master module and the slave module.

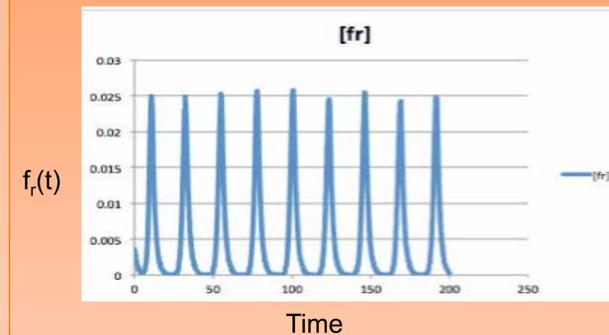
CPU sends the master module solution and the parameter set to each slave module

GPU solves the ODEs of the slave modules in parallel using the Adaptive Runge Kutta (ARK) method

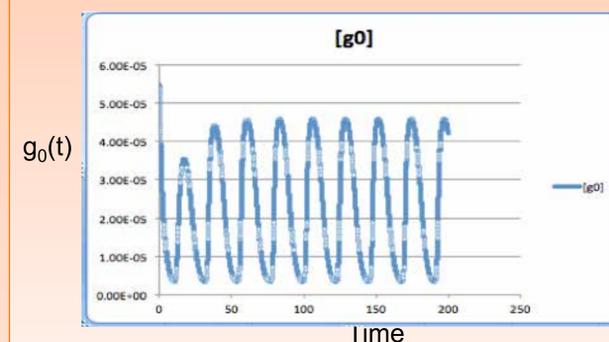
GPU sends the solution back to the CPU

CPU calculates a cost function (Chi-square using Metropolis procedure) that minimizes the difference between the experimental data and the fitted solution

## Results

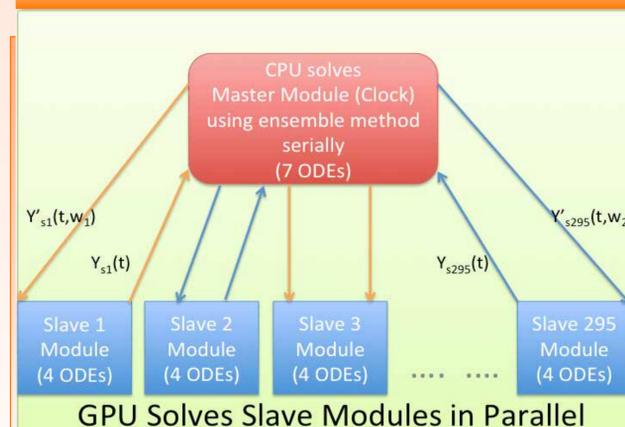


**Figure 2.** The solution of  $f_r(t)$  using the ensemble method over the time period [0 200] for the genetic network of the biological clock of *Neurospora crassa*.



**Figure 3.** The solution of  $g_0(t)$  using GPUs over the time (t) period [0 200] for the genetic network of a CCG subunit of *Neurospora crassa*.  $[g_0] = [ccg^0] = \text{concentration of } ccg^0$ .

## Model of solution procedure



**Figure 4.** Solution's Model

## Conclusion

We have demonstrated that we can implement the ensemble method (Yu et al., 2007). We have also demonstrated that we can solve the ODEs on GPUs for the slave module in **Figure 3**. We propose to:

- ◆ Implement an ensemble method on the CPU and GPUs for identifying the network in **Figure 1**. We have described a family of parallelization strategies on the GPUs for the ensemble method.

We propose to:

- ◆ Implement several parallelization strategies for the ensemble method and to examine their performance as a function of the number of slave modules.

We have more than 3000 data points for fitting the proposed large genetic network for the clock (Dong et al., 2008). We propose to:

- ◆ Fit the genetic network in **Figure 1** by the best parallelization strategy on the CPU and GPUs to our microarray data. The resulting network will enable us to examine how the clock controls ribosome biogenesis by a detailed molecular mechanism.

## References

- Dong, W., et al. (2008) Systems biology of the clock in *Neurospora crassa*, *PLoS one*, 3, e3105.  
 Huang, Y. (2007) Parameter estimation of chemical reaction networks. A Master thesis, Physics Department, The University of Georgia.  
 Lakin-Thomas, P.L., Bell-Pedersen, D. and Brody, S. (2011) 3 The Genetics of Circadian Rhythms in *Neurospora*, *Advances in genetics*, 74, 55.  
 Yu, Y., et al. (2007) A genetic network for the clock of *Neurospora crassa*, *Proceedings of the National Academy of Sciences*, 104, 2809-2814.

## Acknowledgment