CATEGORY: **BIOINFORMATICS & GENOMICS - BG07**

POSTER **P4220**

CONTACT NAME
Raghuram Thiagarajan: raghut@umich.edu

GPU TECHNOLOGY CONFERENCE

Reverse Engineering of Genome-Scale Biological Networks

Raghuram Thiagarajan, Jason N. Bazil and Daniel A. Beard

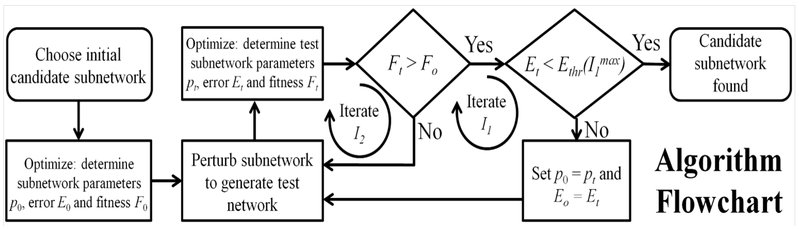
Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, Michigan

VPR The Virtual Physiological Rat Project

Abstract

Availability of genome-scale data sets in biology present a great opportunity as well as a challenge for computational biologists. Simulation and model based analysis on such large-scale dynamical systems pose compute-intensive problems. A reverse-engineering algorithm optimized for parallel architectures has been developed to study these dynamical systems. The parallel architecture and processing power of Graphics processing units(GPUs) provide a platform to carry out genome-scale simulations. We show that genome-scale networks can be inferred using this reverse-engineering algorithm in a matter of days on a single Tesla K20 GPU.

Algorithm Overview



- The network inference model[1] uses a non-linear dynamic model to approximate gene regulation.
- The algorithm is validated using data simulated from networks designed to possess biologically relevant motifs[2,3].

mRNA Transcription and Translation Model

Differential equation model: $\dot{x}_j = r_j - d_j x_j$

For the *j*th gene in a network

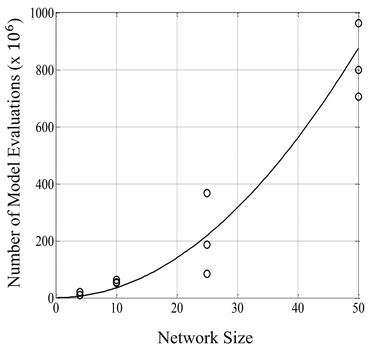
x – mRNA concentration

r – mRNA transcription rate

d – mRNA degradation rate

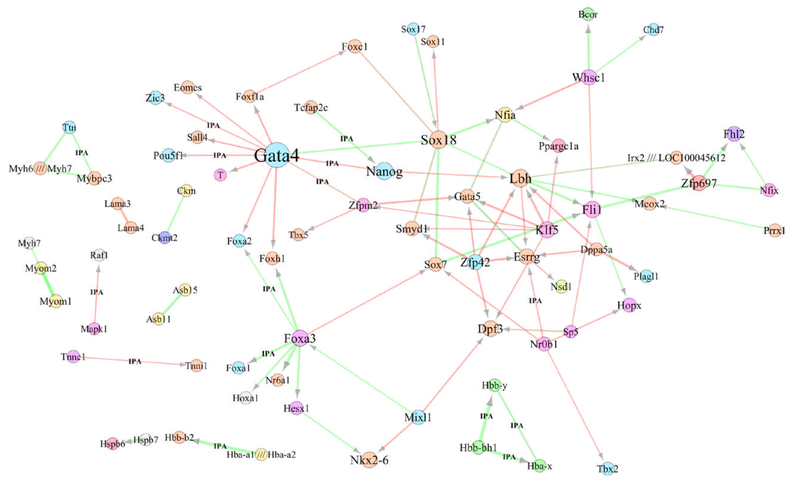
where, $r_j = r_{0j} \dfrac{\sum_i \left(x_i(t-\tau)A_{i,j}\right)^n + e_j^n}{1 + \sum_{\forall i \neq j} \left(x_i(t-\tau)A_{i,j}\right)^n + \sum_{\forall k \neq j} \left(x_k(t-\tau)I_{k,j}\right)^n}$

Scaling Relative to Network Size



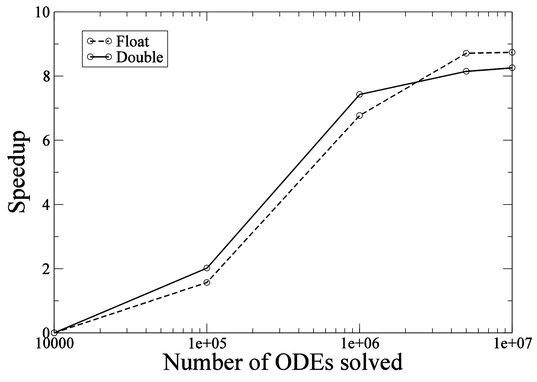
- The number of model evaluations are found to scale as $\sim N^2$ for *in-silico* networks.
- Significant improvement over deterministic model-based inference ($\sim N^3$) and information theory based approaches ($\sim N^2 \log N$).
- The number of model evaluations are almost a billion for a network size of 50.

Applications of the Algorithm to Cardiogenesis

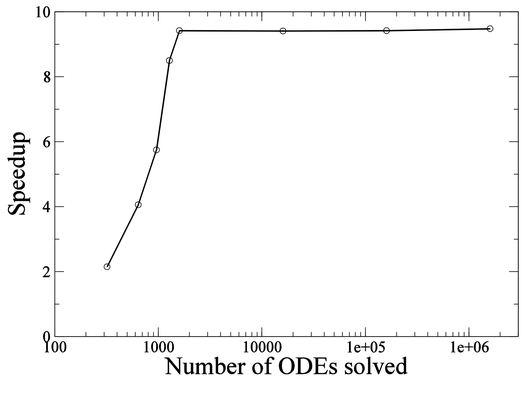


An example of inferred network generated for left ventricle using approximately 200 genes involved in cardiogenesis. The "IPA" symbol signifies IPA validated interactions. The current implementation is limited to ~100s of genes. Hence, we expect implementing this algorithm on a GPU platform is expected to facilitate genome-scale network analysis in tractable timescales.

ODE solvers on GPU

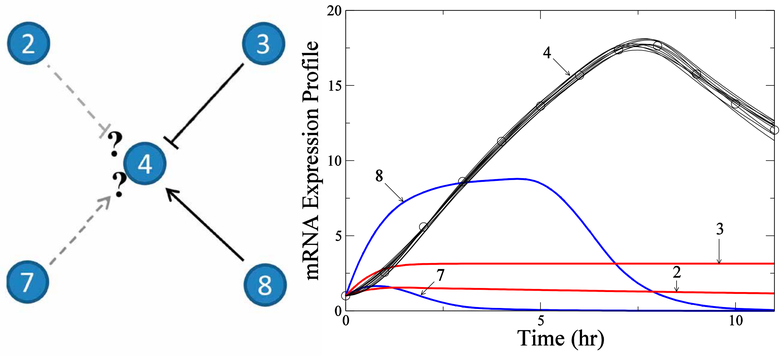


- ODEINT, which uses Thrust to solve ODEs on the GPU, yields a speedup of 10x
- The non-linear differential equation can be stiff, but ODEINT doesn't have a stiff solver for GPU computing

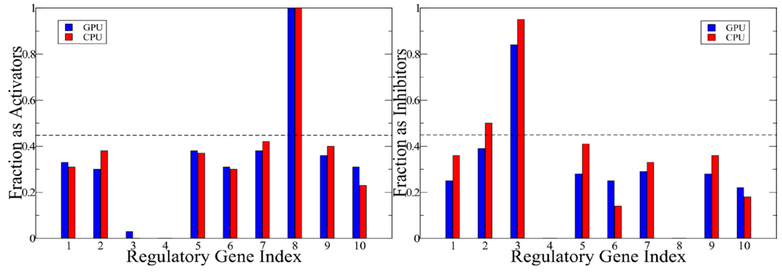


- LSODA, a stiff ODE solver for GPU computing, provides an order of magnitude speedup

Network Inference using LSODA



Subnetwork topology and mRNA expression profiles with target gene as gene 4. Solid black lines represent edges recovered by the algorithm, which are present in the true network and dashed gray lines represent edges not recovered by the algorithm, but which are present in the true network. The mRNA expression profile described by the regulator genes determined by the algorithm is presented for the target gene 4. The number against each profile represents the gene index.



Histograms of activator and inhibitor edges for this system are illustrated. These histograms indicate the relative frequency that each edge appeared in a total of 2,051 trial networks for the GPU and 500 trial networks for the CPU are found to effectively match the time-course data for this variable

- Genes 8 and 3 have been found as true positives
- Gene 2 is a close second, which appears in approximately 50% of subnetworks and gene 7 comes is second highest on the activator list
- The algorithm returned 10 acceptable subnetworks by testing 1,000 subnetworks in 3s on a Tesla K20 GPU
- A 1,000 node network would require assessment of 50,000, which is a compute time of at least 40 hours on a single Tesla K20 GPU.

Conclusions

- The algorithm identified candidate subnetworks capable of simulating the expression data.
- Scale-free, hierarchical networks emerge from the most significant edges for the cardiogenesis network.
- Estimation of genome-scale networks is feasible with use of multiple GPUs

This work was supported by the Virtual Physiological Rat Project funded through NIH grant P50-GM094503. The authors thank NVIDIA® for their donation of two Tesla K20 GPU cards through their academic research grant.

[1] Bazil, JN, F Qi and DA Beard. 2011. "A parallel algorithm for reverse engineering of biological networks", Integrative Biology, 2011, **3**, 1215-1223.
[2] D. Marbach, R. J. Prill, T. Schaffter, C. Mattiussi, D. Floreano and G. Stolovitzky. "Revealing strengths and weaknesses of methods for gene network inference", Proc. Natl. Acad. Sci. U. S. A., 2010, **107**, 6286–6291.
[3] D. Marbach, T. Schaffter, C. Mattiussi and D. Floreano. "Generating Realistic in silico Gene Networks for Performance Assessment of Reverse Engineering Methods", J. Comput. Biol., 2009, **16**, 229–239.