Scalable FPGA Solutions for Learning Bayesian Networks

Systems Biology: Signaling Networks
U.C. Berkeley & Stanford University

Chris Fletcher, Narges Asadi, Greg Gibeling, Eric Glass
Dan Burke, Farzad Fard, Ilia Lebedev, Mingjie Lin
RAMP Undergraduates
John Wawrzynek, Wing Wong, Garry Nolan
http://ramp.eecs.berkeley.edu

Network Discovery & MCMC

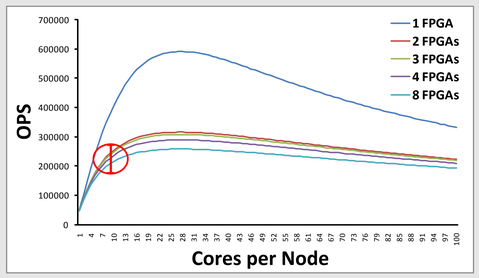
Motivation: Bayesian networks can encode cell protein structure and interactions between different proteins
→ Protein structure analysis allows biologists to study the cause of human disease and drug effectiveness

Problem: Learning Bayesian network structure is NP-Hard
→ Search space grows super-exponentially with network size
→ Solution space might have local (not global) "best solutions"

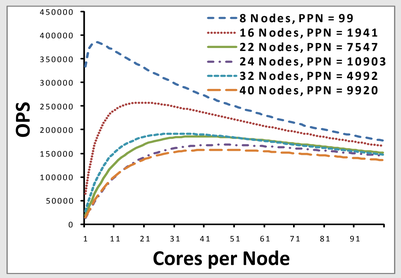
Solution: Map the algorithm (Markov chain Monte Carlo - MCMC) to a network of FPGAs
→ "Bayesian scoring" algorithm is highly parallel
→ Algorithm requirements map very well to on-chip memory

Results

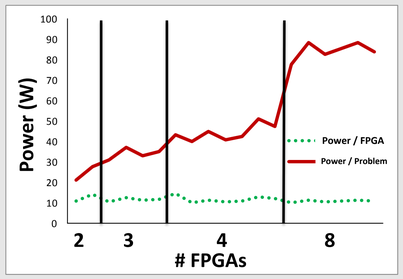
Network: 22 Nodes
4 "In-Degree" → 7547 Parents Per Node
Vary hardware, fix network



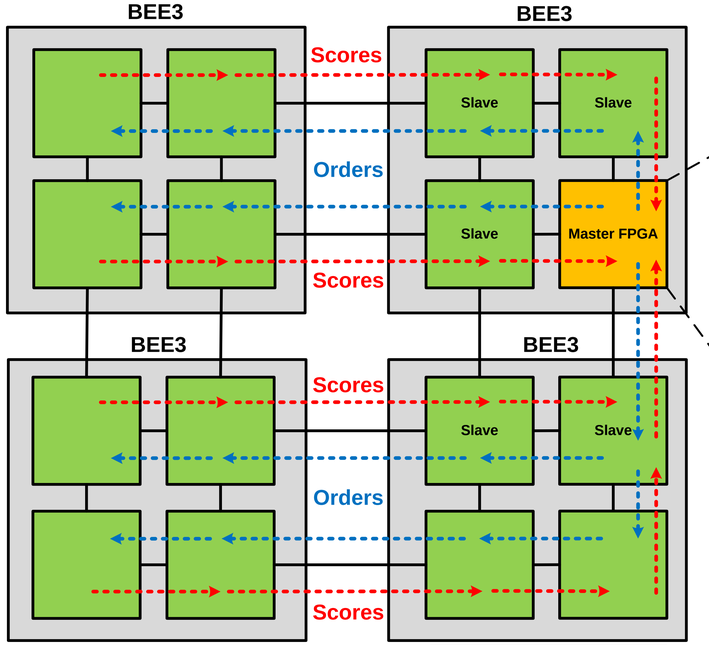
Setup: 10,000 iterations / restart
50 random restarts
Fix hardware, vary network

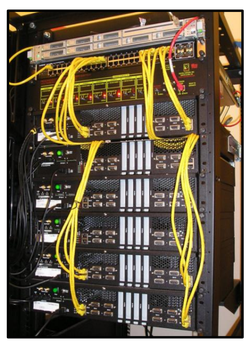


GPP: 3 Ghz quad-core, 12 GB DRAM
FPGA: Xilinx Virtex-5 LX155T (-2)
Per {problem, FPGA} power

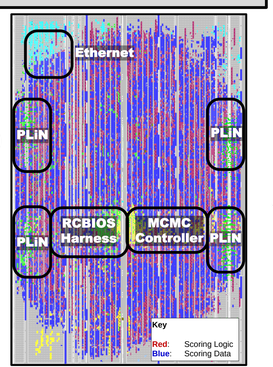


Scalability



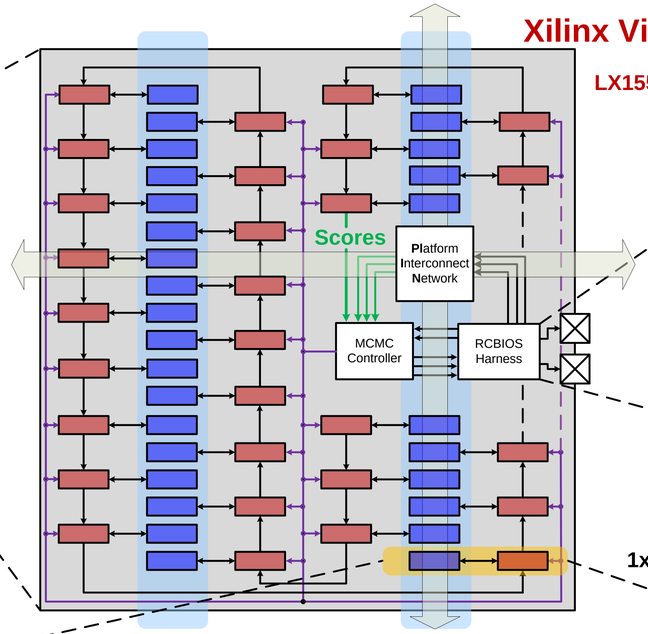


FPGA (BEE3) Rack

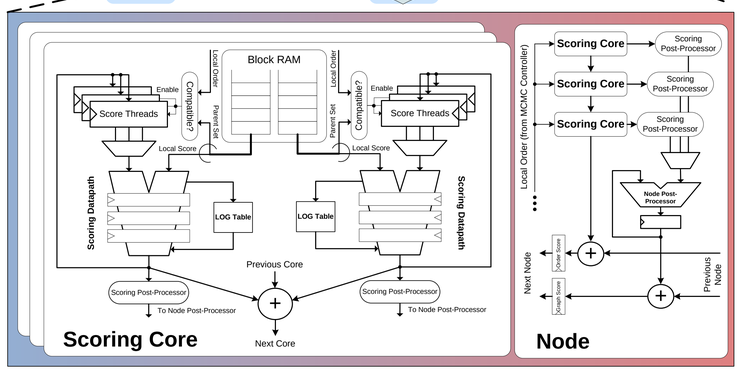


1x FPGA

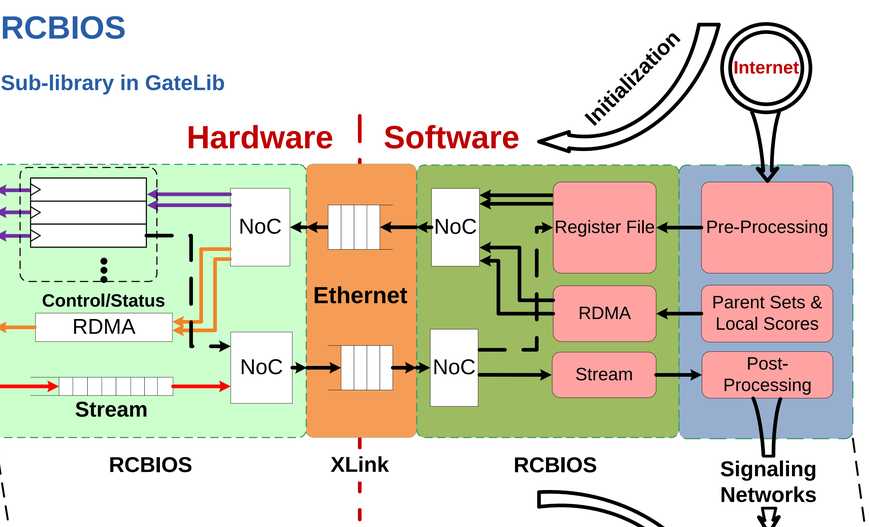
Implementation



1x Node



Infrastructure



RCBIOS Hosted on EmCon

RCBIOS: **R**econfigurable **C**luster **B**asic **I**/**O** **S**ystem
→ Framework enabling communication between front-end PC and FPGA
→ Scalable in performance and size
→ Physical-link independent
→ Standard interfaces to other GateLib components

EmCon: **BEE3 Embedded Controller**
→ Intel Atom GPP (serves as MCMC front-end)
→ Ethernet, Serial, USB interfaces to each FPGA
→ Enables "MCMC in a box"