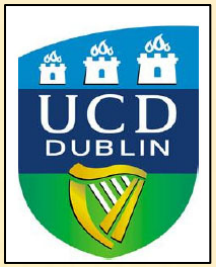


The EvoGrid

An Approach to Computational Origins of Life Endeavours

PhD Viva

July 13, 2011

Brief supporting presentation

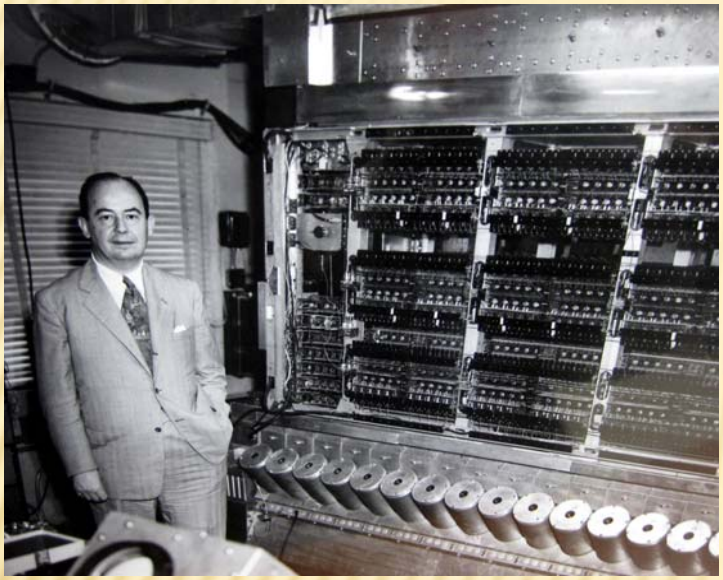
Presented by Bruce Damer
SMARTlab Research Institute, University College Dublin, Ireland
Contact email: bdamer@digitalspace.com

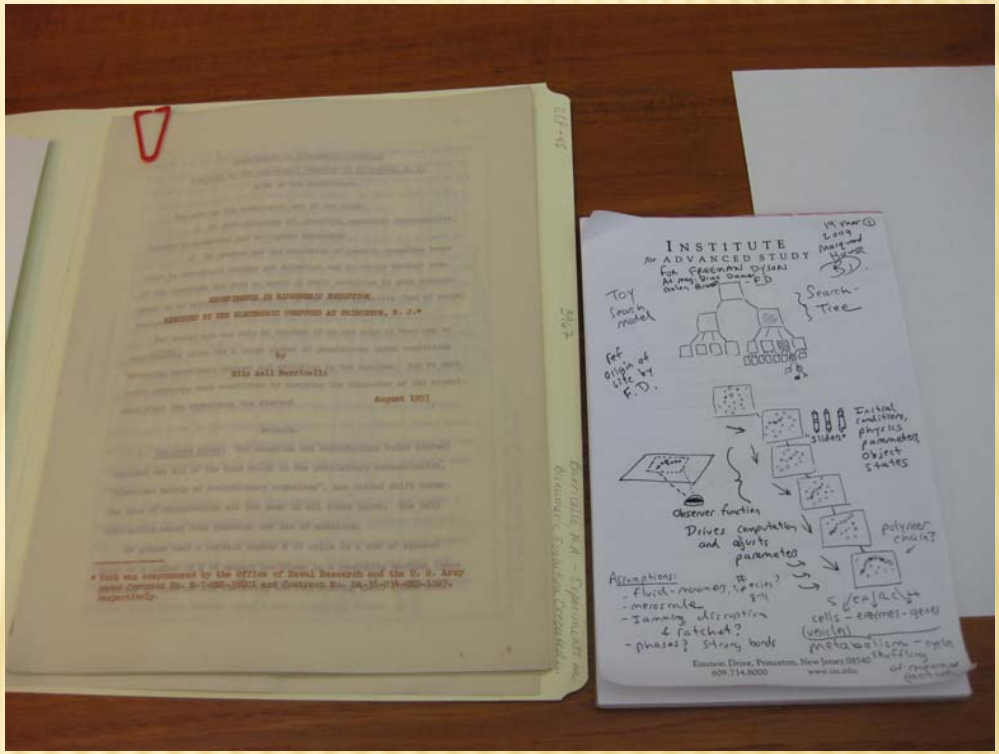
Brief Viva Presentation Roadmap

I. Framing the Challenge of Computational Origins of Life Endeavours, Historical Antecedents, Research Question, Hypothesis and Benefits to Science, and Cognate Fields
II. Design for the EvoGrid Simulation Framework and Optimizations
III. EvoGrid First Prototype Experiments and Analysis
IV. Limitations, Roadmap, Open Questions and Broader Considerations, Contributions to Knowledge

Resources and Acknowledgements

I. Framing the Challenge of Computational Origins of Life Endeavours, Historical Antecedents, Research Question, Hypothesis and Benefits to Science, and Cognate Fields





Historical antecedents informing the challenge and design of the digital simulation of evolution: Barricelli's *numerical symbioorganisms* (Barricelli, 1953).

Research Question: Hypothesis and Benefits to Science

Hypothesis

Distributed processing and search optimization employing stochastic hill climbing can produce significant performance improvements in the generation of emergent phenomena within small volume, short time frame molecular dynamics simulations over non-optimized solutions.

Benefits to Science

A method and platform for optimizing computation to select for pathways leading to de novo emergent structures and processes in simulated chemistry could benefit future systems supporting cyberbiogenesis computational origins of life endeavours.

The need for distributed, search-enabled small molecular dynamics and artificial chemistry simulations in origins of life endeavours

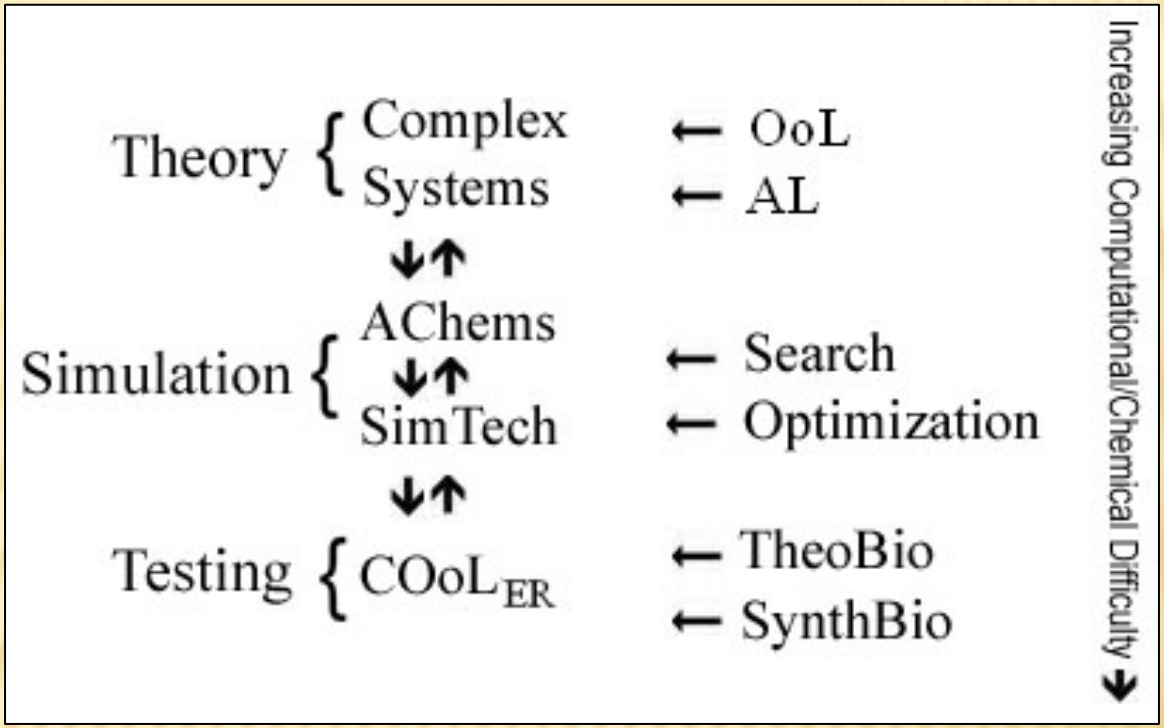
- Many models and operational systems exist (e.g. GARD[1])
- Off-the-shelf open and proprietary engines exist (GROMACS, LAMMPS, Desmond, & etc)
- Current efforts often use “human in the loop” inspection of data often running on single CPUs or small clusters
- Large MD systems are costly to build (Anton[2])
- Origin of life endeavours require emergent behavior (i.e. chemistry)
- Smaller “cameo” simulations may be less costly to run, especially in a distributed fashion
- Massively distributed scientific simulation has developed a good track record (BOINC[3])

[1]Shenhav B., Lancet D. (2004), Prospects of a computational origin of life endeavor, Origins of Life and Evolution of Biospheres, 34(1-2), 181-94.

[2]Shaw, D.E., Martin M. et al. (2008), Anton, a special-purpose machine for molecular dynamics simulation, Communications of the ACM, 51 (7), 91–97.

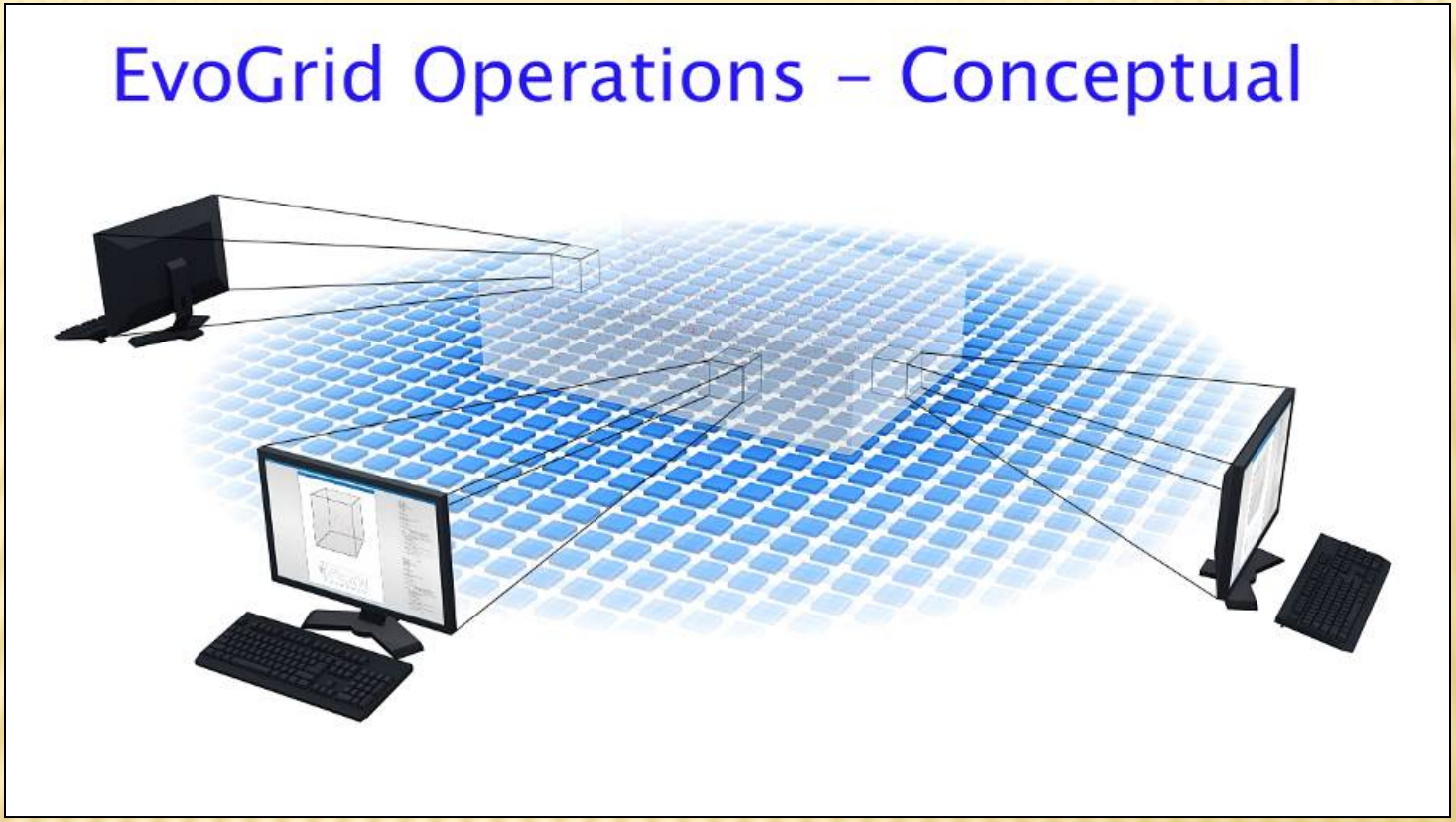
[3]Anderson, D.P. (2004), BOINC: A system for public-resource computing and storage, Proceedings of the 5th IEEE/ACM international Workshop on Grid Computing (November 08 - 08, 2004).

Map of cognate fields



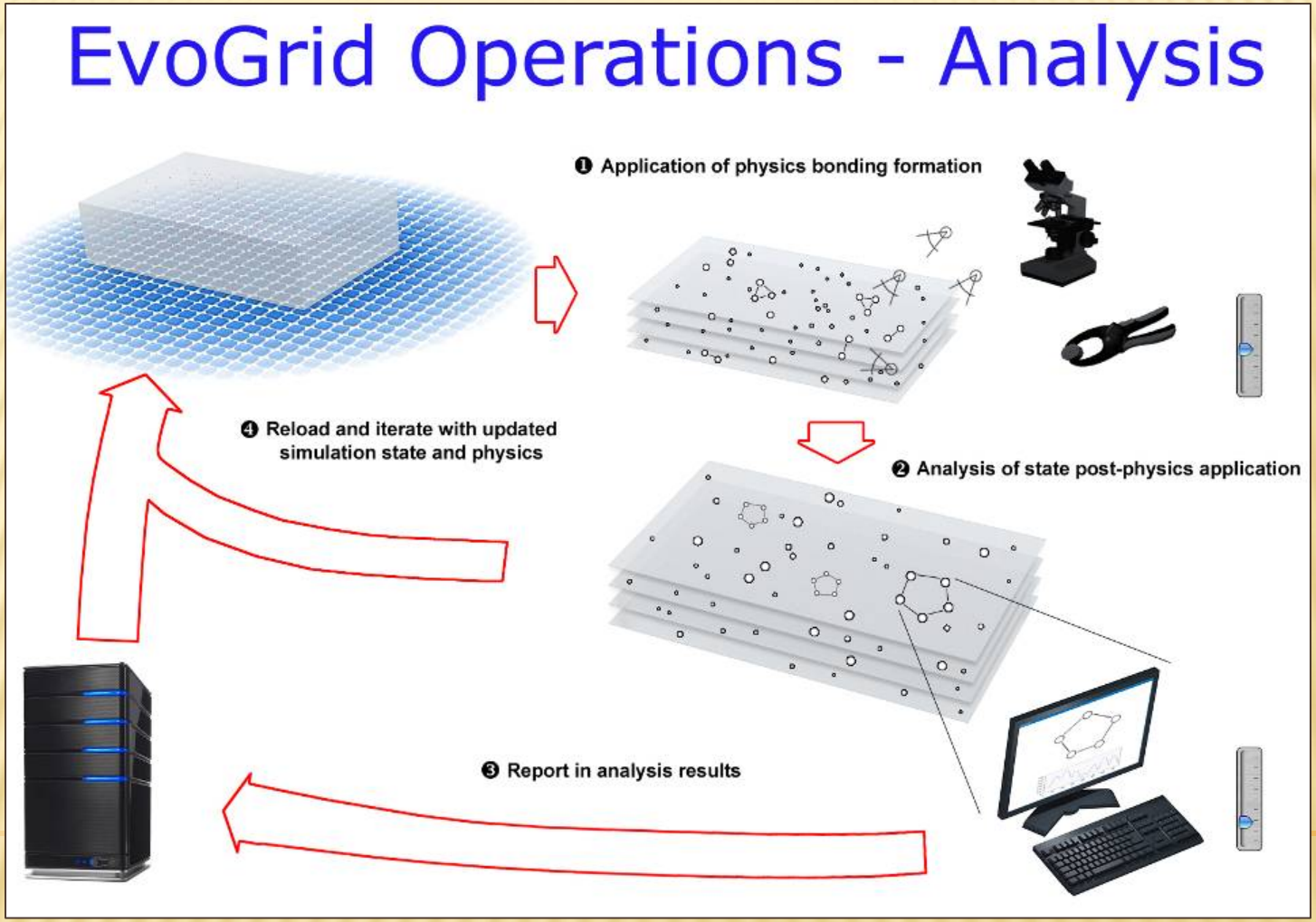
Interrelationships between cognate fields and supporting fields which underlie computational origins of life endeavours in artificial chemistries.

II. Design for the EvoGrid Simulation Framework and Optimizations

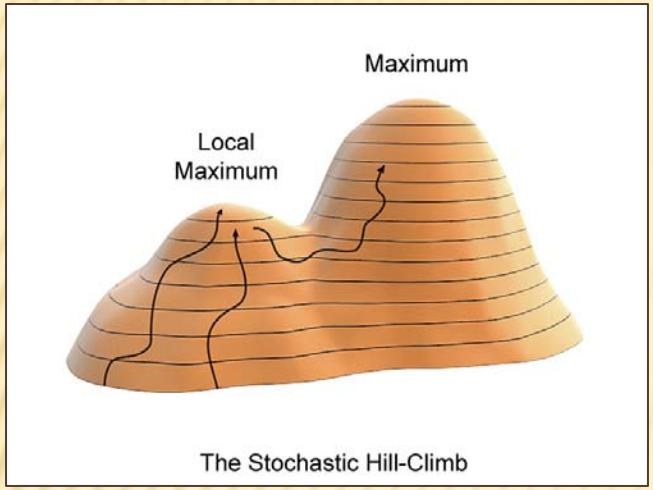


EvoGrid: Concept of multiprocessor support of distributed chemical simulation and multiple client analysis

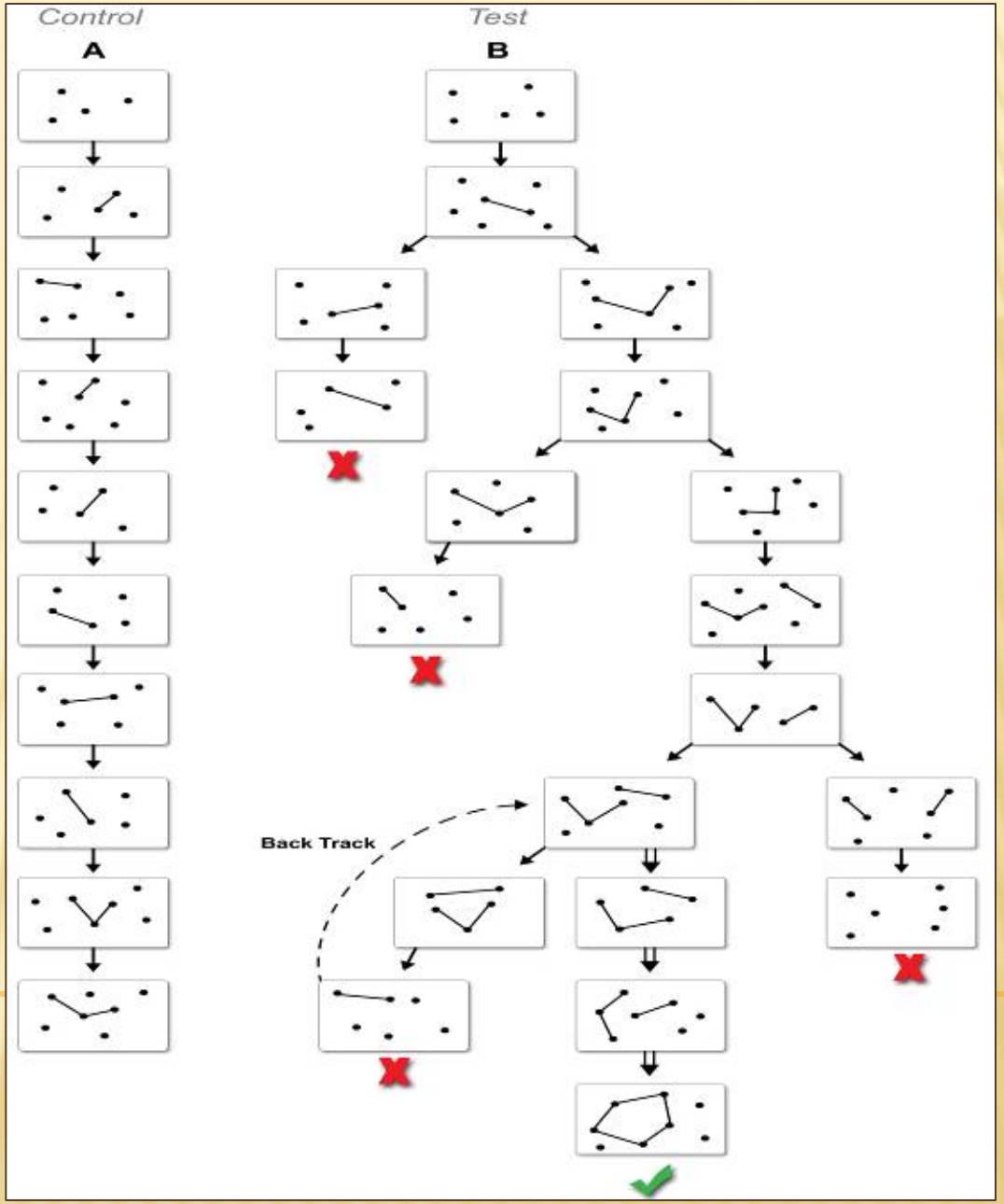
EvoGrid: Concept of Operations & Analysis



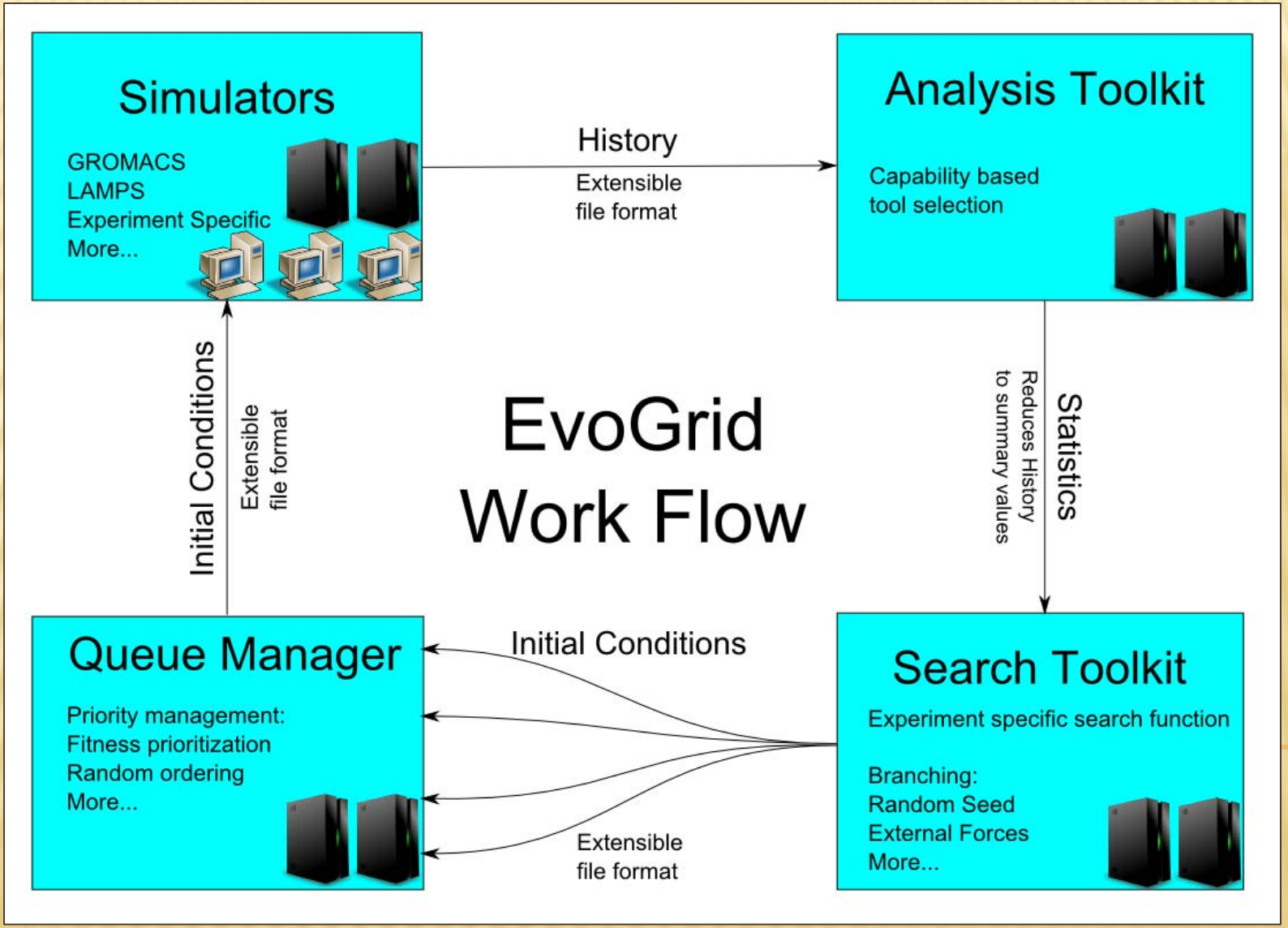
EvoGrid Optimization



Concept of Search (fitness function) implementing a Stochastic hill-climbing algorithm utilizing analysis, feedback and temporal backtracking



EvoGrid: Global Work Flow

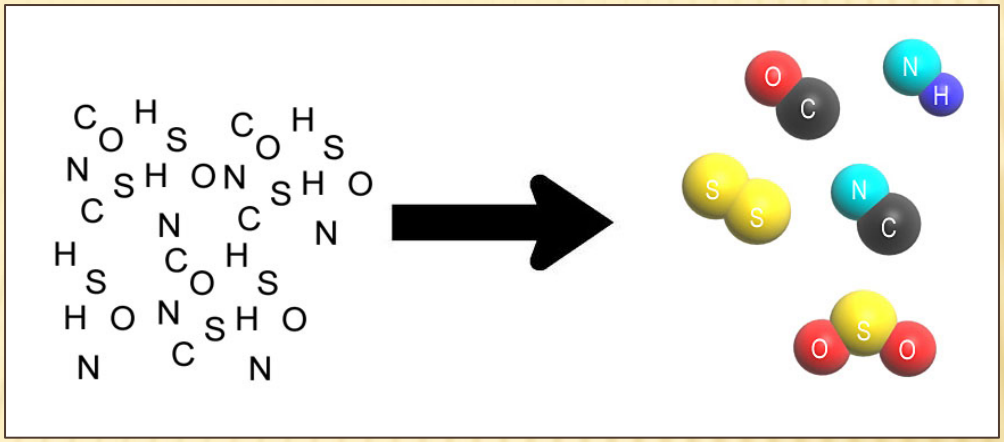


Hardware configuration for EvoGrid, first and second grids: DigiBarn (2010) and U.C. San Diego (2011)

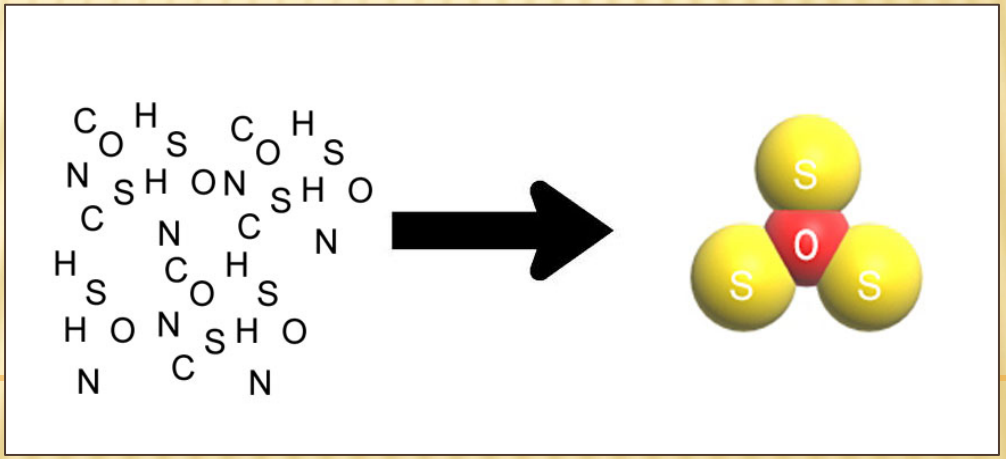
Grid1: 2 months operation
(4 cores average)
Grid2: approx 5 months operation
(15-30 cores, distribution of daemons)

III. The EvoGrid First Prototype Experiments and Analysis

Meta-experiment:
Lots of molecules
(directed search)



Meta-experiment:
Large molecules
(directed search)

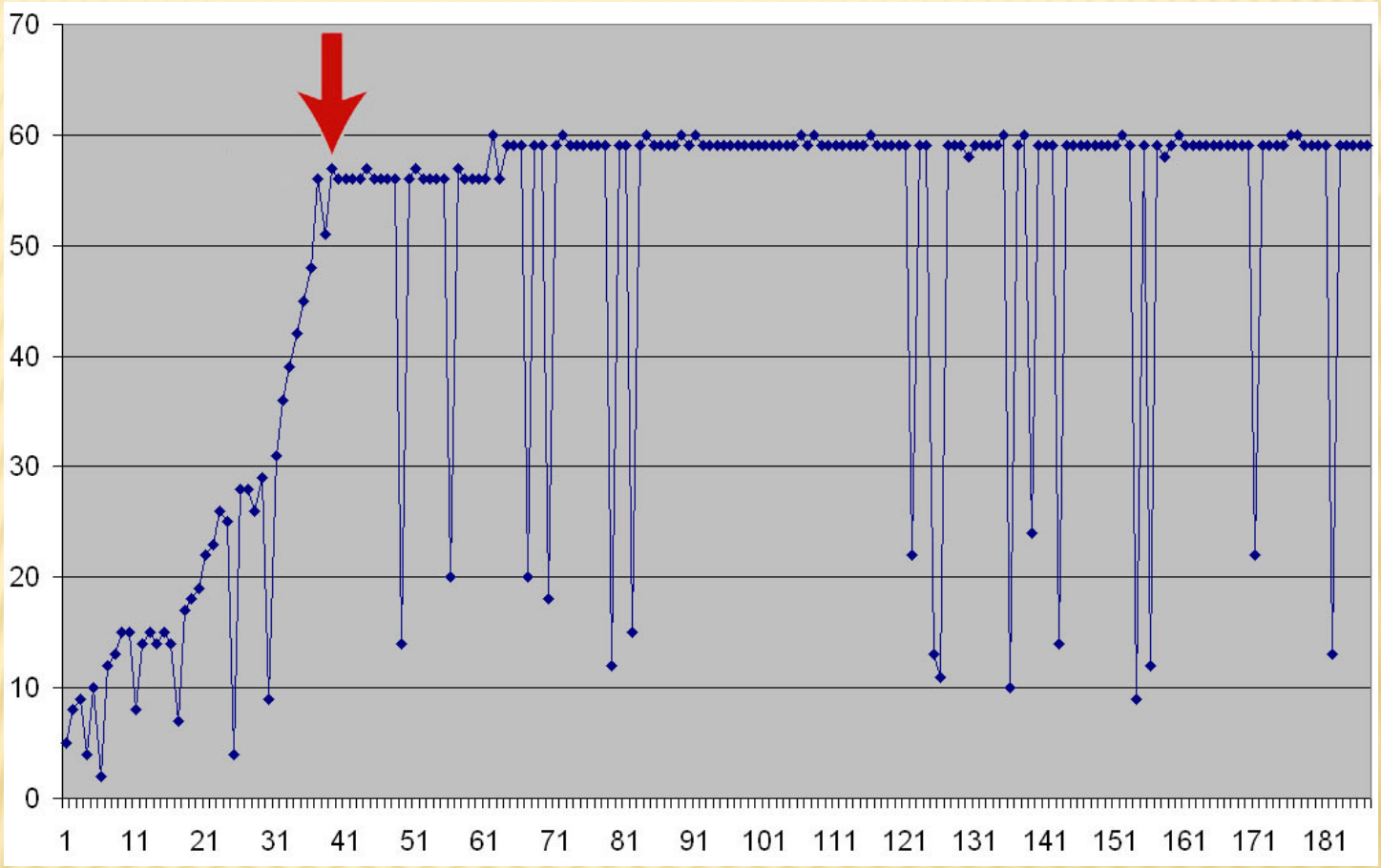


Stochastic Hill Climbing Pseudocode, All Experiments

```
Top:
   Calculate Fitness Score of completed simulation

   For exp 1,2, priority simple function:
     Set Priority = Fitness Score

   For exp 4,5,6,7 priority generation function:
     selection daemon On
     If experiment 4,5
        Set selection daemon Off

     Retrieve Parent Priority
     Retrieve Parent Fitness Score

     Set Priority = Parent Priority *
                    (0.9 * exp( Fitness Score -
                                Parent Fitness Score ) )

   For all simulations
      Simulate
      Branch
      If selection daemon On
         Randomly select highest priority branch
         Simulate

Goto Top
```

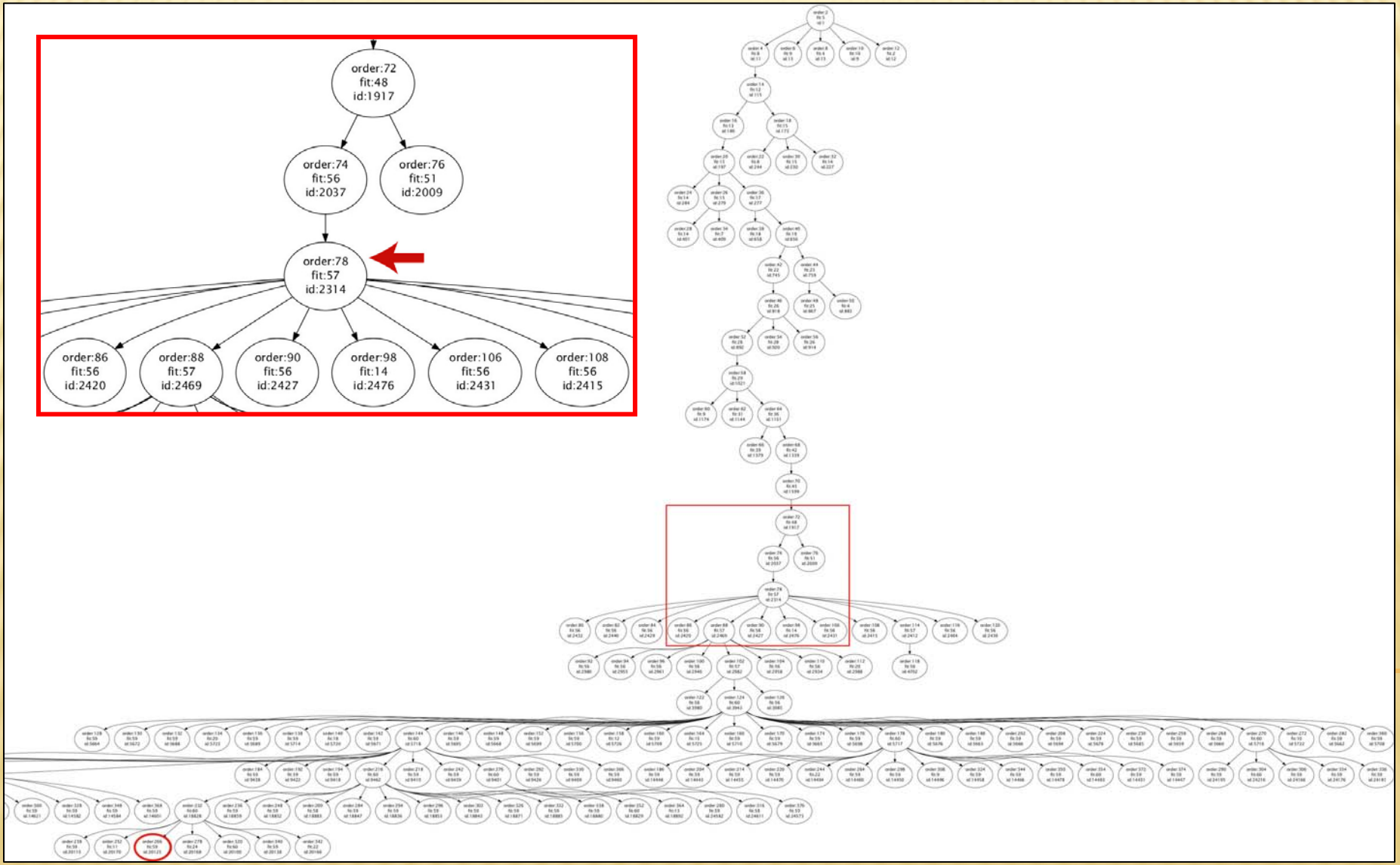
Abstract high level pseudocode mechanism to calculate priority, branch generation, random selection, followed by simulation for all seven experiments.

Experiment #1 number of molecules

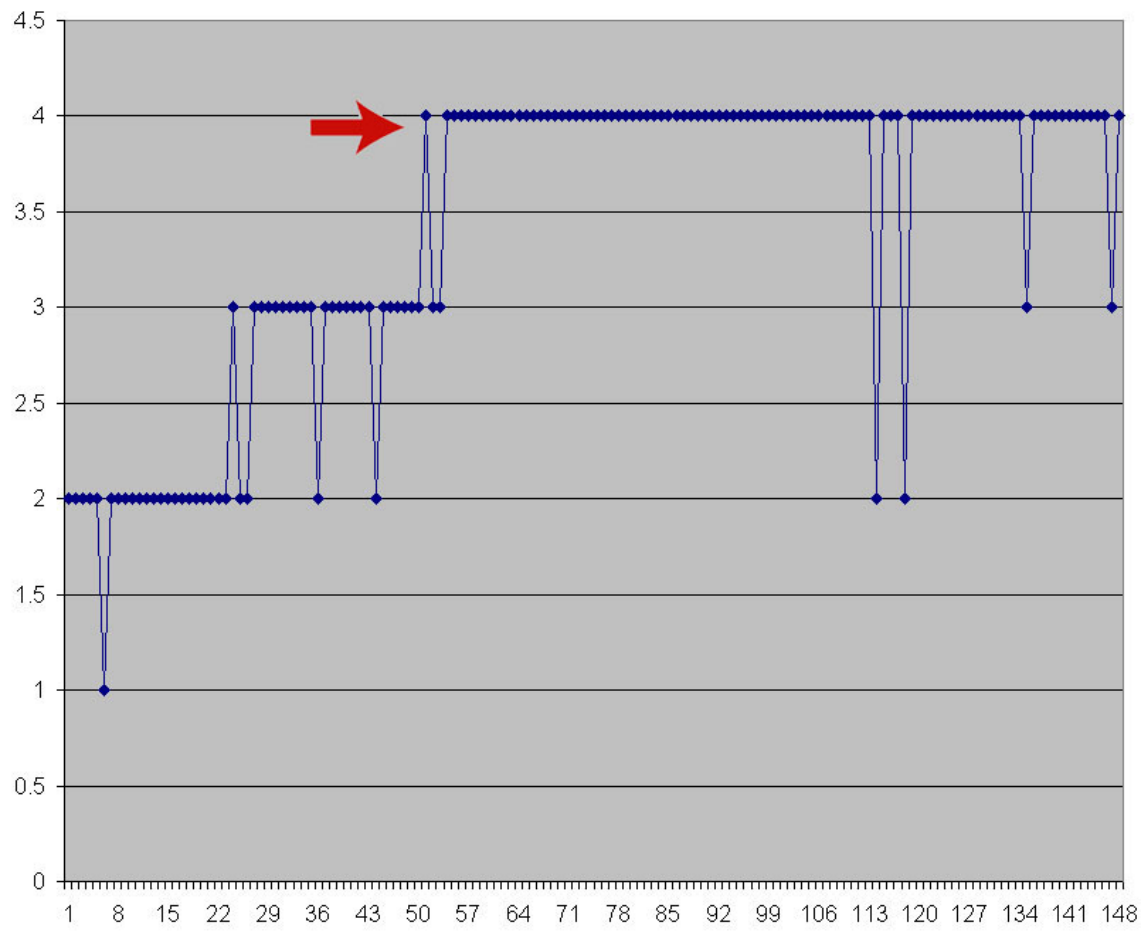


Point of attainment of first stable (long term) maximum.
Y-axis: number of molecules, X-axis: processed simulations (39).

Hierarchical generation tree for Experiment #1

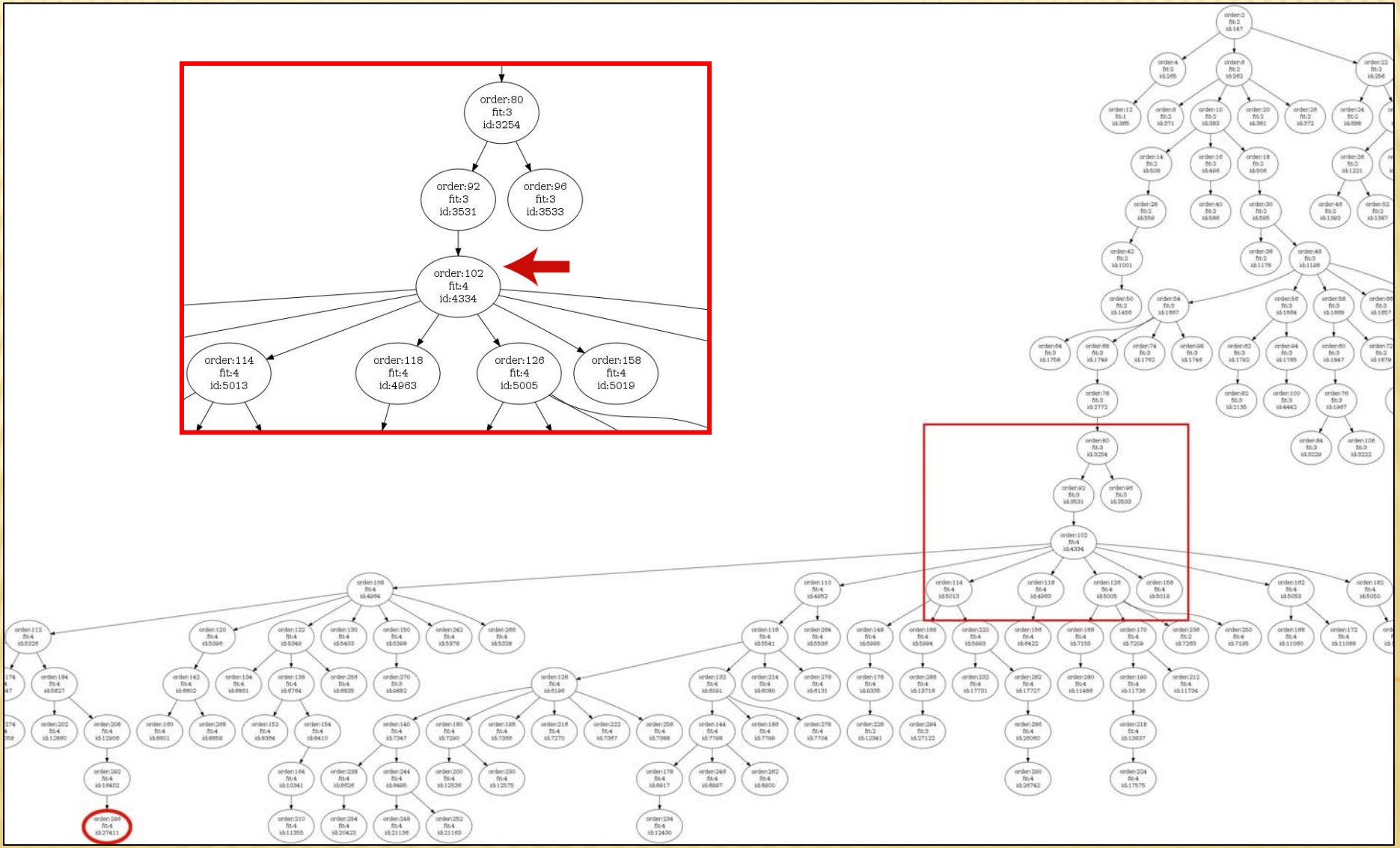


Experiment #2 maximum molecular size

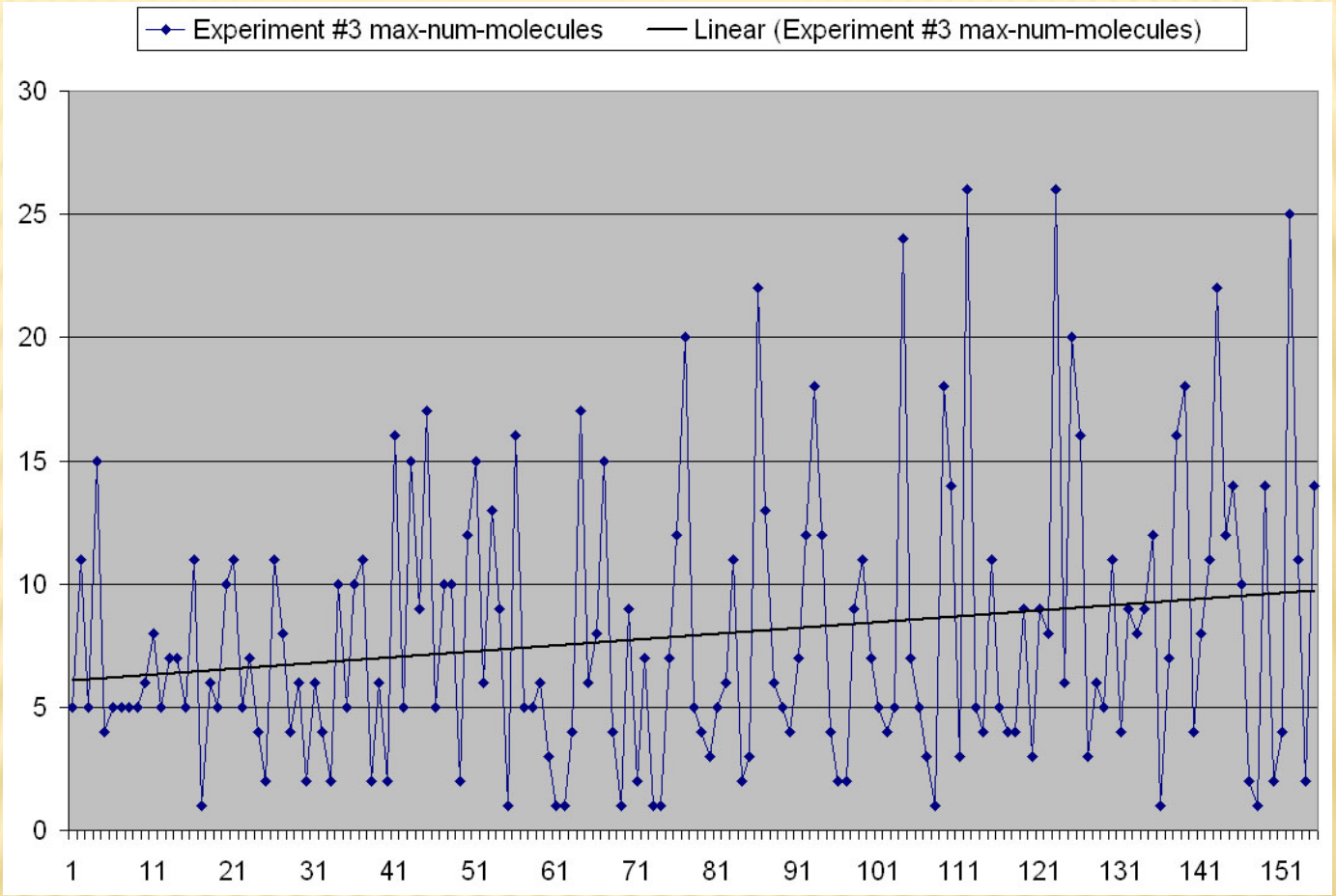


Point of attainment of first stable maximum.
Y-axis: size of molecules, X-axis: processed simulations (51)

Hierarchical generation tree for Experiment #2



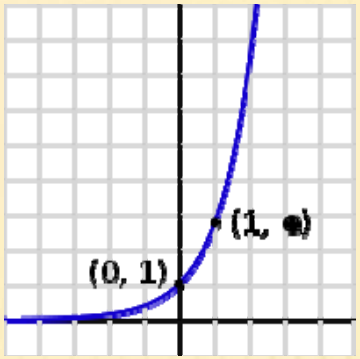
Experiment #3 Control (no directed search)



Linear trendline through Control (for number of molecules) shows only slight growth from 6 to 10 vs. 60 molecules for Test Experiment #1.

priority_generation function pseudocode operation in practice

exponential function $y = e^x$



Set Priority =
Parent Priority * (0.9 * exp(Fitness Score – Parent Fitness Score))

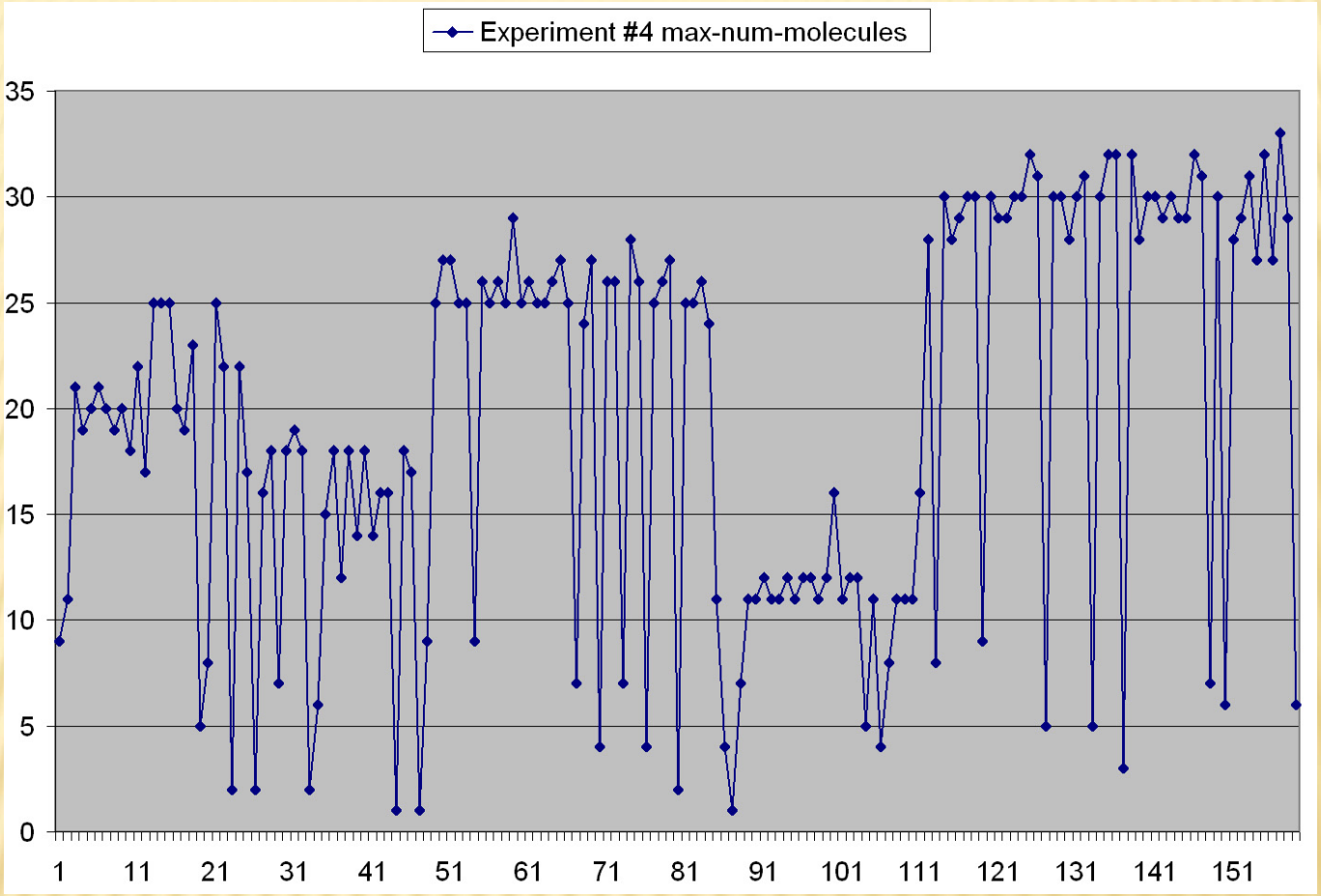
For equal fitness scores (current and parent) exp(0) = 1 so the degradation factor is 0.9 or 10%

For a decline in fitness -n (current fitness < parent fitness): exp(-n) = small nonzero positive result, ie: 0.9 * exp (-1) is 0.9 * 0.367879441 = 0.331091497 or small nonzero degradation factor

For an increase in fitness n (current fitness > parent fitness): exp(n) = large and increasing positive result, ie: 0.9 * exp(1) is 0.9 * 2.71828183 = 2.44645365

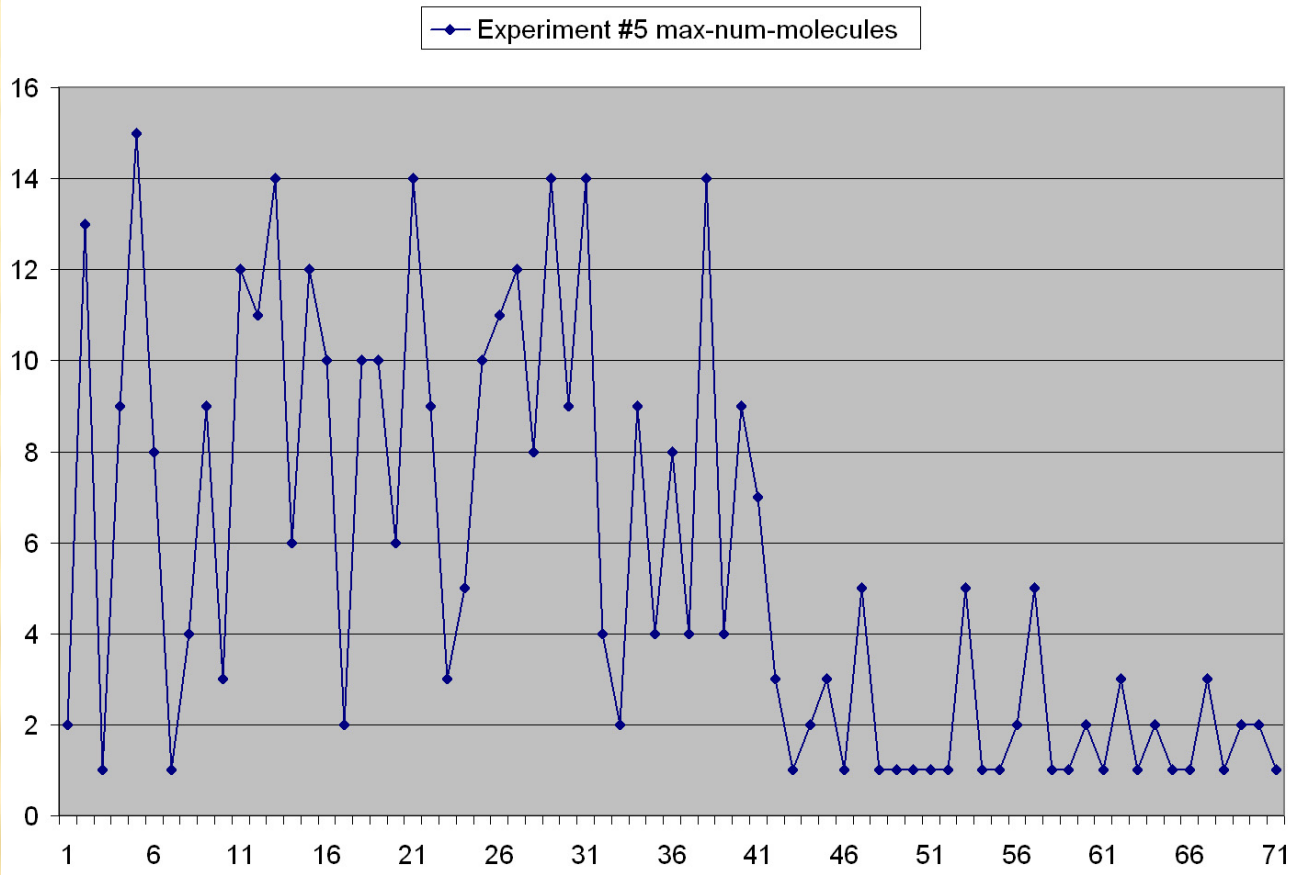
So search may gently move off existing maxima but aggressively travel up new maxima, avoiding entrapment on ridges or saddles

Experiment #4 (*priority_generation*, no random selection)



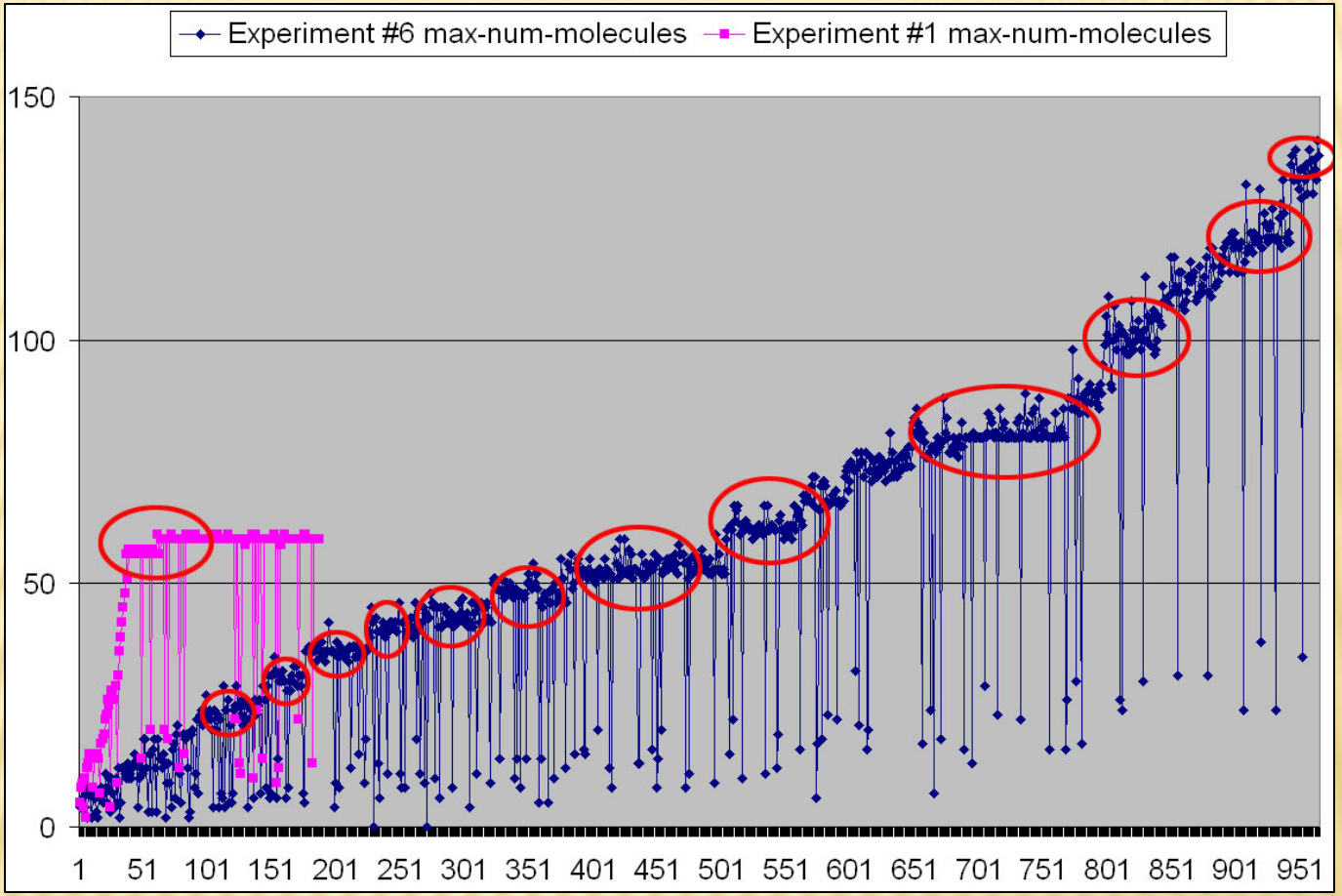
Attainment of local maxima for num-molecules but not sustained (irreversible) trend. Mol-size achieved sizes of 3 but not sustained trend.

Experiment #5 (scoring on molecular size *priority_generation*, no random selection)



Inability to find any maxima in num-molecules, mol-size generated no trend (all molecules of size 2).

Comparison of key experiments and analysis



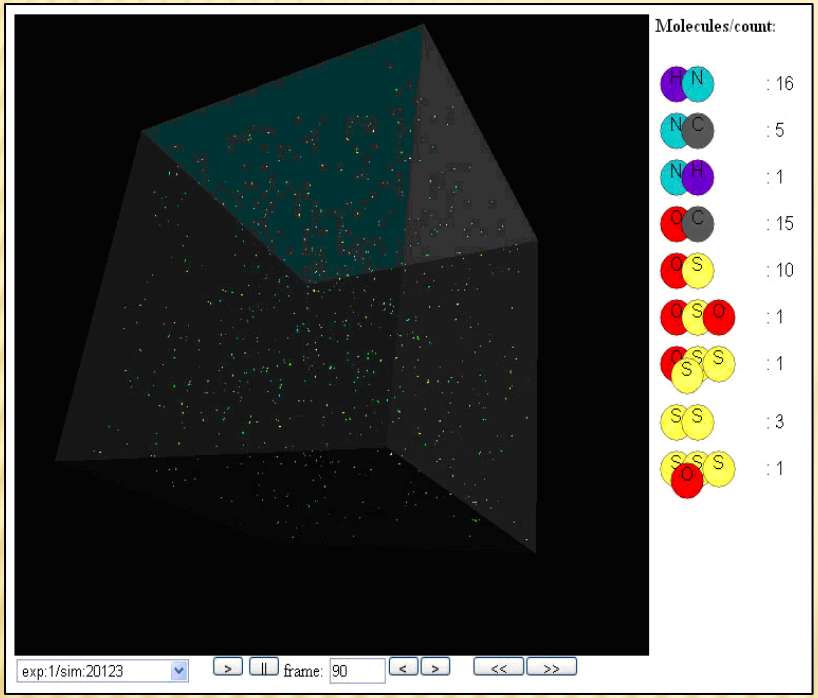
Experiment #1: plateau maximum of 60 molecules
Experiment #6: surmounts serial maxima, eventual plateau at 141 molecules (189 at termination of experiment)

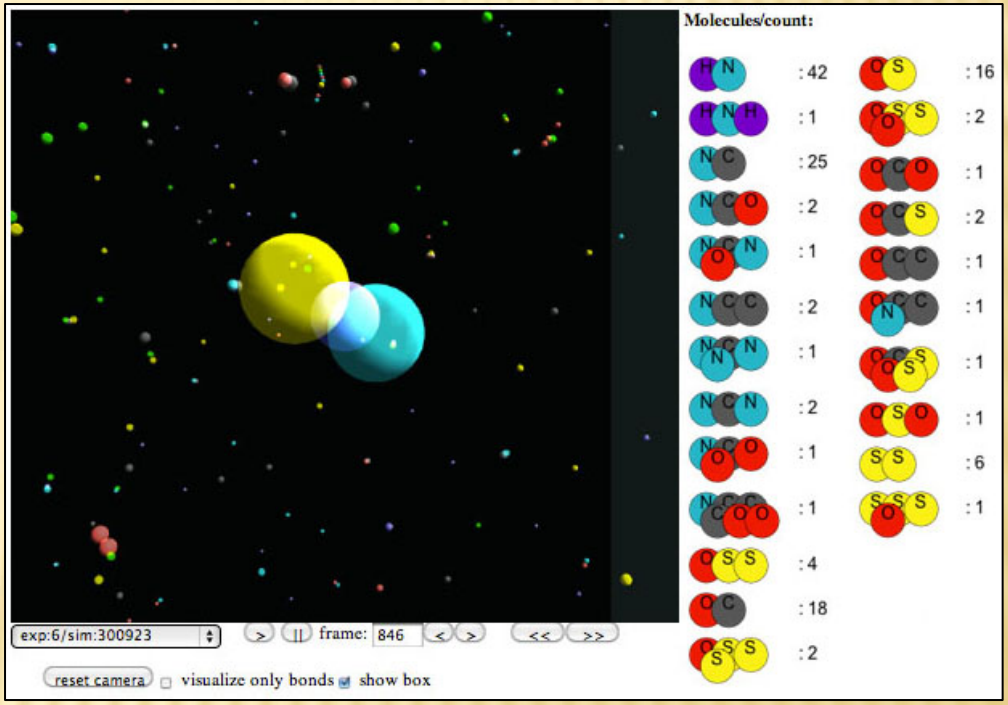
Summary of all Experiments (May 4, 2011)

Experiment	#Processed (#Failed) Simulations	#Incomplete Simulations	Complete: Incomplete Ratio	Highest Score
1- num mol	187 (26)	11015	1:51.713615023474	60
2- mol size	148 (39)	40654	1:217.40106951872	4
3 -control	155 (21)	16178	1:91.920454545455	N/A
4-num mol	158 (39)	5884	1:29.868020304569	33
5-mol size	71 (42)	38820	1:343.53982300885	2
6-num mol	966 (134)	61415	1:50.422824302135	141
7-mol size	212 (45)	4391	1:16.759541984733	3

1,2 succeeding in hill climbing but reached long term plateaux
3 (control) provided no-search comparative case for 1,2
4,5 implemented degradation factor without random selection, failed to achieve significant maxima
6 used degradation and selection, achieved multiple serial maxima
7 used degradation and selection, coarse granularity of mol-size score led to failure to achieve trend

Sample results by “molecular” products





Experiment #1 (max=60)

Experiment #6 (max=141)

WebGL 3D viewer depicting snapshot of simulations with current virtual molecular products (yields).

Successes of EvoGrid prototype experiments

A stochastic hill climb through levels of complexity (ratchets) was observed in several experiments. Performance of early test experiments #1 and #2 over control (#3) showed a significant (6x sustained in num-molecules and 2x sustained in molecular-size) improvement in complexity (yield in chemical terms). Experiments #4 and #5 implemented a version of *priority_generation* and failed to show sustained complexity. Experiment #6 implemented the full version of *priority_generation* and showed a significant improvement over Experiments #4, #5, and #1.

Computational cost savings in terms of processed vs. abandoned frames was:
1:51.71 (187 processed and 11,015 frames abandoned) for Experiment #1)
1:50.42 (966 frames processed and 61,415 framed abandoned for Experiment #6)

Surprise result: for search function tuned to provide a pathway for larger molecules (Experiment #2), more molecules were produced than the experiment which was meant for large molecule generation. Explanation: multiple local maxima explored in quest for search for larger molecules produces more molecular products or “tries”. This expressed the emergence of nondeterministic (non teleologically driven) phenomena in the system.

Key comparative experiments and analysis

- The simple hill climbing of Experiment #1 produced a single sustained maximum
- Experiment #6's mechanism utilizing degradation produced a climb through multiple maxima, although at an additional cost of time/computing resources
- The fitness landscape is not random, rather it is correlated (ref Kauffman)
- Kauffman's observation of exponential growth of hill climbing cost born out
- Underlying objective function is altering landscape but this is not characterized

Future work would benefit from:

- Generation of histogram of the occurrence of the types of molecules
- Mathematical treatment of the formulae and rates of creation and destruction of molecules for several experiments based on Lotka–Volterra equations
- Examination of simulations to extract the chemical reaction graph (molecules formed and broken over time) to serve as valuable input for hypopopulated reaction graph experiment of Kauffman[1]

[1]Damer and Kauffman, Private Conversation (2011)

IV. Limitations, Roadmap, Open Questions and Broader Considerations, Contributions to Knowledge

Limitations: naïve bond formation and branching operations yielding unrealistic chemistry. Performance of commodity cluster is insufficient for many bio-relevant phenomena over significant time scales.

Implementation road map: more capable molecular dynamics simulation engines (MOPAC, LAMMPS, GARD etc) with multi-scale, multi-physics, support for GPUs, BOINC network. More realistic physical models for quantum dynamical bond interactions (covalent, ionic) and chemical physicodynamics: fluctuations and periodicity, dissipative systems, fluxes, phases of matter.

Societal considerations: origin of life inside computer networks and realization in chemical experiments: major hallmark of science, EvoGrid Turing Test, religious thought and philosophy. Societal impact includes possible explanation of origin of life without a creator, and lens on where life could exist in the universe (exobiology).

Next steps: grant funding for continued development, book proposal, Experiment #8 and expansion of grid at U.C. San Diego.

Experimental road map

Experiment #1: Astrochemistry Model (Allamandola)

Experiment #2: FLiNT Nanocell Model (Fellermann, Rasmussen et al.)

Experiment #3: Riboyzyme Selection Experiment Model (Szostak, Bartel et al.)

Experiment #4: Model of a Hypopopulated Reaction Graph for the Study of Autocatalytic Sets and the Adjacent Possible (Kauffman et al)

Experiment #5: Model for RNA-Making Reactors in Hydrothermal Vents (Russell et al.)

Experiment #6: Model for Encapsulation of Polymers in Multilamellar Structures through Wet/Dry Cycles (Deamer et al.)

Experiment #7: Model of the FLiNT Protocell Life Cycle (Rasmussen et al.)

Experiment #8: Complex Free Encapsulation Origin of Life Model (CREATR-Damer et al.)

Intrinsic contributions to knowledge

Addressing the Hypothesis

Distributed processing and search optimization employing stochastic hill climbing can produce significant performance improvements in the generation of emergent phenomena within small volume, short time frame molecular dynamics simulations over non-optimized solutions. Given careful tuning of the core algorithms for search, scoring, prioritization and branching, these optimization methods did produce significant performance improvements in terms of time saved and computational products produced over a non-optimized solution, in one case generating a full order of magnitude more molecular bonds. This class of computation can be characterized as adaptive search on a rugged but correlated landscape, is highly sensitive to initial conditions and yet is capable of engaging in a significant climb to a series of maxima.

Benefits to Science

A method and platform for optimizing computation to select for pathways leading to de novo emergent structures and processes in simulated chemistry could benefit future systems supporting cyberbiogenesis computational origins of life endeavours.

Extrinsic contributions to knowledge

A history of artificial life and chemical simulation endeavours as well as a literature review of informative fields contributing to the design of viable computational origin of life frameworks.

A vision and definition of a new term cyberbiogenesis which captures the marriage of in silico computer simulation and in vitro chemical experiment for origin of life endeavours.

A map of the major cognate fields that illustrate how these fields inform cyberbiogenesis enterprises.

A listing of current limitations and a technical roadmap for the improvement of the current EvoGrid prototype and a roster of experiments in origins of life research to which future EvoGrid platforms may be applied.

A series of open questions both for the EvoGrid and for an emerging field of computational origins of life simulation.

An illustration and discussion of scientific, philosophical, religious, and general societal conundrums posed by this line of research.

Thank You!
Resources and Acknowledgements

EvoGrid project pages at: http://www.evogrid.org (code available-LGPL)

I would like to thank SMARTlab Research Center DOS/supervisory team/UEL & UC Dublin, DigitalSpace and Elixir Technologies Corporation for funding support. Individuals and groups supporting this work include: Peter Newman, Ryan Norkus and Galen Brandt (DigitalSpace), Miroslav Karpis, Richard Gordon (University of Manitoba), Tom Barbalet (NobleApe, Biota.org), David Deamer (UC Santa Cruz), John Graham, Calit2/U.C. San Diego, all informal external advisors, and Basit Hamid (founder and CEO, Elixir Technologies Corporation).

Bruce Damer, *bdamer@digitalspace.com*
DigitalSpace Corporation
343 Soquel Ave, Suite 70
Santa Cruz CA 95062, Ph: 831 338 9400