DTIC® has determined on 03 26 2009 that this Technical Document has the Distribution Statement checked below. The current distribution for this document can be found in the DTIC® Technical Report Database.

☑ DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

☐ COPYRIGHTED. U.S. Government or Federal Rights License. All other rights and uses except those permitted by copyright law are reserved by the copyright owner.

☐ DISTRIBUTION STATEMENT B. Distribution authorized to U.S. Government agencies only. Other requests for this document shall be referred to controlling office.

☐ DISTRIBUTION STATEMENT C. Distribution authorized to U.S. Government Agencies and their contractors. Other requests for this document shall be referred to controlling office.

☐ DISTRIBUTION STATEMENT D. Distribution authorized to the Department of Defense and U.S. DoD contractors only. Other requests shall be referred to controlling office.

☐ DISTRIBUTION STATEMENT E. Distribution authorized to DoD Components only. Other requests shall be referred to controlling office.

☐ DISTRIBUTION STATEMENT F. Further dissemination only as directed by controlling office or higher DoD authority.

Distribution Statement F is also used when a document does not contain a distribution statement and no distribution statement can be determined.

☐ DISTRIBUTION STATEMENT X. Distribution authorized to U.S. Government Agencies and private individuals or enterprises eligible to obtain export-controlled technical data in accordance with DoDD 5230.25.
Substantial progress was achieved for all objectives stated in this proposal. Several effective efficiency enhancement techniques were proposed from specialized model-directed hybrids to methods capable of learning from prior runs on similar problems. Substantial speedups of an order of magnitude or more were obtained with several of these techniques on difficult classes of problems and evidence was provided of supermultiplicative speedups resulting from a combination of multiple efficiency enhancements. Theoretical models were developed to capture some of the facets of these techniques, providing inputs useful for a better use these techniques in practice. Progress in efficiency enhancement of EDAs resulted in several important large-scale applications, from the solution of a noisy optimization problem with over one billion variables to the analysis of flexible protein structures for proteins of over 900 amino acids. Model-based optimization techniques and other metaheuristics were also successfully applied to the quadratic assignment problem, potential function design in physics, and cancer chemotherapy treatment optimization with one or multiple drugs.
Progress Report
Supermultiplicative Speedups of Probabilistic Model-Building Genetic Algorithms
AFOSR Grant No. FA9550-06-1-0096
February 1, 2006 to November 30, 2008

David E. Goldberg, Kumara Sastry, Martin Pelikan
IlliGAL Report No. X0024 (Limited Circulation)
February 2009

Illinois Genetic Algorithms Laboratory
Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign
117 Transportation Building
104 S. Mathews Avenue Urbana, IL 61801

20090325302
1 Objectives

The objectives of AFOSR grant no. FA9550-06-1-0096 are as follows:

1. Integrate probabilistic model building and evaluation relaxation.
2. Integrate probabilistic model building and time continuation.
3. Integrate probabilistic model building and local search procedures.
4. Integrate probabilistic model building and parallel evaluation.
5. Combine the above sources of supermultiplicative speedup to investigate upper limits of integrated efficiency enhancement.
6. Extend facetwise and little models to predict integrated solution quality and speedup.
7. Demonstrate speedups using the developed procedures in two problem areas of Air Force interest.

Significant progress has been made on each goal. The progress is briefly reviewed in the next section.
2 Status of Effort

Efficient genetic algorithm optimizes problems with over billion variables. An efficient, fully parallelized genetic algorithms was implemented to achieve efficient, scalable solutions on difficult optimization problems containing over a billion variables. The results demonstrated scalable solution of a difficult test function—a noisy, blind problem over a vector of binary decision variables—on instances over a billion variables using a specially programmed parallel genetic algorithm. The GA is able to find the optimum in the presence of noise quickly, reliably, and accurately, and the solution scalability follows known convergence theories. These results on noisy problem together with other results on problems involving varying modularity, hierarchy, and overlap foreshadow routine solution of billion-variable problems across the landscape of complex systems.

Efficiency enhancement from sporadic model building. Sporadic model building was proposed to speed up model building in the hierarchical Bayesian optimization algorithm and other advanced estimation of distribution algorithms. The proposed technique led to significant speedup of the algorithm without sacrificing solution quality or reliability. The speedups grow with problem size and, thus, the bigger the problem, the higher speedups can be obtained. For difficult problems, speedups of at least one order of magnitude can be expected. Theoretical model was developed and verified to provide a robust approach to setting the parameters of sporadic model building.

Evaluation relaxation for interactive GAs invented. One of the bottlenecks of interactive GAs—where users evaluate the relative quality of candidate solutions—is user fatigue. A principled method for eliminating user fatigue was invented and significant speedups were reported. In essence, the method relies on a surrogate induced from a partial ordered graph of user preferences. The surrogate is used for optimization to partially replace expensive user evaluations.

Evaluation relaxation used to speedup machine learning. Tight integration of surrogate and structural models were carried over to genetics based machine learning. Specifically, evaluation relaxation scheme was used to alleviate the computational cost of evaluating fitness of a rule against large data sets. First results indicate that the fitness estimates of the surrogates are of high-quality and the computational cost is significantly lower than the true fitness evaluations. These initial results strongly suggest possibilities of super-multiplicative speedups in genetics-based machine learning domain as well.

7 orders of magnitude speedup in reaction chemistry simulations. We (Todd Martinez (2005 MacArthur fellow), Duane Johnson, Kumara Sastry and David E. Goldberg) have applied multiobjective GAs and model building to enable fast and accurate quantum reaction chemistry simulations. This work is providing fresh insights and important research avenues toward fast, accurate chemistry calculations which will dramatically accelerate both the science (biophysical basis of vision, air pollution, and photosynthesis) and synthesis of new pharmaceuticals, chemicals and materials.

9 orders of magnitude speedup in materials modeling. We (Duane Johnson, Pascal Bellon, Kumara Sastry, and David E. Goldberg) have applied genetic programming to multi-scale modeling of alloys using a hybrid of the molecular dynamics (MD) and kinetic Monte Carlo (KMC) procedures. In 2D surface modeling our calculations project speedups of 9 orders of magnitude at 300 degrees Kelvin. Two papers have been published (International Journal for
Multiscale Computational Engineering and Physics Review B) and was chosen as a focused article of frontier research by the editors of American Institute of Physics. Simply stated, genetic programming is used to perform customized statistical mechanics by bridging the different time scales of MD and KMC quickly and well.

Random additively decomposable problems. A class of additively decomposable problems was proposed to test and analyze performance of optimization techniques for nearly decomposable and hierarchical problems. A method was proposed to generate random instances of the proposed class of problems and experimental results were obtained to confirm scalability theory and various properties of simple and advanced evolutionary algorithms.

Extended compact classifier system invented. The idea of an estimation of distribution algorithm has become commonplace in optimization problems, but these ideas have not been widely explored in machine learning applications. First results and publications were prepared on the extension of compact classifier system to learn and discover interactions between attributes by datamining a population of promising rules.

Critical components for handling modularity, hierarchy, and overlap discovered. Techniques using dependency structure matrices (DSMs) have been used by large corporations and governmental agencies in organizational design. Work on DSM clustering led to the creation of a GA called the DSMGA using DSM clustering to do building-block decomposition. This year, work on DSMGA has led to the discovery of key components required to handle modularity, hierarchy, and overlap—common ingredients of complex systems. The key components for conquering modularity, hierarchy, and overlap include accurate decomposition, niching, chunking, proper sequencing, and well-informed decision making.

Application of genetic algorithms to prediction from NIR Spectra. An implementation of a genetic algorithm was developed for reliable prediction of chemical properties from near-infrared (NIR) spectra. From experimental data, a model is constructed that both (1) fits experimental data well and (2) generalizes well to unseen samples. The prediction is done by a simple linear model in combination with a genetic algorithm used to select important attributes for reliable prediction. Promising results have been obtained on several data sets based on the experiments on old documents.

Structural analysis of flexible macromolecular systems. An implementation of a genetic algorithm was developed for the analysis of flexible macromolecular systems such as intrinsically disordered or multidomain proteins with flexible linkers. Data obtained at the Lawrence Livermore National Lab in Berkeley are used as input to a genetic algorithm, which determines subsets of theoretical structures that approximate empirical results best. The results are then analyzed to provide insight into the flexibility of the studied system. The first results were tantalizing, and the method is currently being applied in practice.

Ground states of large Sherrington-Kirkpatrick spin glasses. Several approaches were developed for reliable detection of ground states of large instances of the Sherrington-Kirkpatrick (SK) spin-glass model. To achieve this goal, hybrid techniques were proposed that combine advanced genetic and evolutionary algorithms with specialized local search. The results should be important for both physicists as well as practitioners in evolutionary computation. The project was later extended to the one-dimensional SK spin glass with power-law interactions.

Learning from experience in advanced estimation of distribution algorithms. Estimation of distribution algorithms (EDAs) are stochastic optimization techniques
that explore the space of potential solutions by building and sampling explicit probabilistic models of promising candidate solutions. While the primary goal of applying EDAs is to discover the global optimum or at least its accurate approximation, besides this, any EDA provides us with a sequence of probabilistic models, which in most cases hold a great deal of information about the problem. We proposed several approaches to enhancing the efficiency of EDAs by analyzing information provided by the sequences of models from previous EDA runs, leading to substantial speed-ups on future problem instances of similar type. This line of research also led to better understanding of the structure of probabilistic models discovered by advanced EDAs, such as the hierarchical Bayesian optimization algorithm (hBOA).

3 Personnel Supported

This section details the individuals supported on this project. 

Faculty supported. Professor David E. Goldberg, the principal investigator, was supported during the summers of 2006 and 2007. Professor Martin Pelikan, CO-PI, was supported during the summers of 2006 and 2007.

Other affiliated visitors, postdoctoral personnel. The following is a list of visiting faculty or postdocs affiliated with the project. Unsupported affiliates may have had some travel or incidental expenses paid by the project:

1. Professor Pier Luca Lanzi (Politecnico di Milano, Italy)
2. Dr. Xavier Llorà (National Center for Supercomputing Applications, UIUC)
3. Dr. Noriko Imafuji Yasui (University of Illinois at Urbana Champaign)
4. Professor Minqiang Li (Tianjin University, China)
5. Claudio F. Lima (University of Algarve, Portugal)
6. Luis de la Ossa (University of Castilla la Mancha, Spain)
7. Qin Kai (Nanyang Technological University, Singapore)
8. Albert-Orriols Puig (Ramon Llull University, Barcelona, Spain)
9. Thiago Duque (Universidade de Sao Paulo, Brasil)
10. Yuji Sato (Hosei University, Japan)
11. Shunsuke Saruwatari (Tokyo University, Japan)
12. Jian-Hung Chen (Feng Chia University, Taiwan)

Graduate student affiliates. The following is a list of graduate students supported or affiliated with the project (source of funding in parentheses). Unsupported affiliates may have had some travel or incidental expenses paid by the project:

1. Osvaldo Gomez (Fullbright)
2. Kazuhisa Inaba
3. Mark Hauschild (NSF)
4. Rajiv Kalapala (NSF)
5. Onur Pekecan
6. Elizabeth Radetic (NSF)
7. Kumara Sastry (MatSE)
8. Kenneth P. Turvey (none)
9. Paul Winward (AFOSR)
10. Tian-Li Yu (AFOSR)
11. Thyago Sellmann Pinto Cesar Duque (AFOSR)

Undergraduate student affiliates. The following is a list of graduate students supported or affiliated with the project. Unsupported affiliates may have had some travel or incidental expenses paid by the project:

1. Martin Hanlon (AFOSR)
2. Takaoki Ueda (AFOSR)
3. Hana Roh (AFOSR)

4 Publications for 2005

Published


5 Publications for 2006

Published


### 6 Publications for 2007

**Submitted**


**Published**


## 7 Publications for 2008

### Submitted

Csefalvayova, L., Pelikan, M., Kralj Cigic, I., Kolar, J., M. Strlic, Use of genetic algorithms with multivariate regression for determination of gelatine in historical papers from FT-IR and NIR spectral data. *Analytical and Bioanalytical Chemistry*.
Hammel, M., Yu, Y., Mahaney, B. L., Cai, B., Ye, R., Phipps, B.M., Pelikan, M., Rambo, R.P.,
Hura, G.L., Lees-Miller, S. P., J. A. Tainer, Structural Insights into the Ku70/Ku80/DNA-PKcs
Complex and Consequences for Nonhomologous End Joining. Molecular Cell.

Hauschild, M., M. Pelikan, K. Sastry, D. E. Goldberg. Using Previous Models to Bias Structural
Learning in the Hierarchical BOA. Evolutionary Computation, MIT Press.

Pelikan, M., Hammel, M., G. L. Hura, Structure and Flexibility within Proteins as Identified
Through Small Angle X-ray Scattering. General Physiology and Biophysics.

Published

Brownlee, S., J. McCall, M. Pelikan, S. Shakya, An Application of a Multivariate Estimation
of Distribution Algorithm to Cancer Chemotherapy. Genetic and Evolutionary Computation

Butz, M. V., P.-L. Lanzi, X. Llorá, and D. Loiacono, An Analysis of Matching in Learning Classifier
Systems. Genetic and Evolutionary Computation Conference (GECCO-2008), pp. 1349-1356,
ACM Press.

Butz, M. V., P. Stalph, and P.-L. Lanzi, Self-Adaptive Mutation in XCSF. Genetic and Evolutionary

Duque, T. S. P. C., D. E. Goldberg, and K. Sastry, Enhancing the Efficiency of ECGA. Parallel

Duque, T. S. P. C., D. E. Goldberg, and K. Sastry, Improving the Efficiency of the Extended
Compact Genetic Algorithm. Genetic and Evolutionary Computation Conference (GECCO-

Fernandes, C. M., C. F. Lima, and A. C. Rosa, UMDAs for Dynamic Optimization Problems.

Hauschild, M., and M. Pelikan, Enhancing Efficiency of Hierarchical BOA Via Distance-Based

Hauschild, M., M. Pelikan, K. Sastry, D. E. Goldberg, Using Previous Models to Bias Structural
Learning in the Hierarchical BOA. Genetic and Evolutionary Computation Conference

Howard, G. D., L. Bull, and P.-L. Lanzi, Self-Adaptive Constructivism in Neural XCS and
XCSF. Genetic and Evolutionary Computation Conference (GECCO-2008), pp. 1389-1396,
ACM Press.

Genetic Algorithm Based on Mixtures of Models. In Chen, Y.-P. and Lim M.-H. (Eds.) Linkage
in Evolutionary Computation, pp. 335-358, Springer.

Real-Valued Optimization Problems. In Chen, Y.-P. and Lim M.-H. (Eds.) Linkage in
Evolutionary Computation, pp. 61-86, Springer.


Lima, C. F., C. M. Fernandes, and F. G. Lobo. Investigating Restricted Tournament Replacement
in ECGA for Non-Stationary Environments. Genetic and Evolutionary Computation Conference


8 Publications for 2009

Submitted


Accepted, In Press


9 Interactions and Transitions

This section lists meeting participation, presentations, and transitions.

Meeting Participation and Presentation

All conference papers above represent presentations by Professor Goldberg, Professor Martin Pelikan, Kumara Sastry, their affiliates, or Professor Goldberg's and Professor Pelikan's students. Professor Goldberg gave the following keynote speeches:


*Little Models, Big Results.* 18th Joint Australian Conference on Artificial Intelligence, and the Australian Conference on Artificial Life, Sydney, Australia (2005).

*Solving Larger, Faster, and Harder: The Incredible Story of Supermultiplicative Speedups.* Frontiers of Computational Science, Nagoya, Japan, 2005.

Additionally, Professor Pelikan gave the following keynote speeches and tutorials:


Transitions

hBOA US utility patent issued. The hierarchical Bayesian optimization algorithm patent, _A Method for Optimizing a Solution Set_, was issued as US patent no. 7,047,169.

hBOA licenses in the works. The hierarchical Bayesian optimization algorithm is being under evaluation license for applications in stock market decision making. Test results are promising and a decision on a full license of the technology is imminent. A license of hBOA to a firm in electronic design automation is in the final stages of negotiation.

Nextumi releases ShareThis. With venture capital provided by IllinoisVentures and Blue Chip Venture Capital, Nextumi Inc., a web 2.0 startup company that simplifies sharing of photos, videos, urls and other content between people across different devices released its first product ShareThis. Nextumi licensed AFOSR sponsored competent GA research (patents) to help to create consumer-adaptive technology.

Entrepreneurial engineer published. Professor Goldberg published a new book, _The Entrepreneurial Engineer_ (Wiley, 2006) for engineering students and practicing engineers to help develop those personal, interpersonal, and organizational skills necessary today.

Scalable optimization via probabilistic modeling. Professor Pelikan, Kumara Sastry, and Erick Cantu-Paz edited a book _Scalable Optimization via Probabilistic Model Building_ which includes subject matter ranging from state-of-the art probabilistic model-building algorithms to the application of PMBGAs on real-world problems from some of the leading researchers in the area.

Missouri Estimation of Distribution Algorithms Laboratory (MEDAL). Professor Pelikan founded the Missouri Estimation of Distribution Algorithms Laboratory (MEDAL) at the Department of Mathematics and Computer Science at the University of Missouri in St. Louis. MEDAL focuses on design, application, enhancement and dissemination of robust and scalable optimization methods. More information can be found at [http://medal.cs.umsl.edu/](http://medal.cs.umsl.edu/).

10 New Discoveries, Inventions, or Patent Disclosures

Patent applications have been filed on new discoveries/inventions including:

- The hierarchical Bayesian optimization algorithm.
- A robust and scalable GA, called design structure matrix GA, inspired from organizational theory.
- Collaborative systems design and innovation support over the web using interactive GAs, human-based GAs, and chance discovery.
- Supermultiplicative speedups via a tight integration of probabilistic models and evaluation relaxation.
- Supermultiplicative speedups via adaptive time continuation, where surrogate and structural models are used to adaptively determine the appropriate choice between crossover- and mutation-dominated search.
• Evaluation relaxation for avoiding user fatigue and user inconsistency in interactive GAs.
• Multiobjective GAs for fast and accurate quantum chemistry simulations
• Efficient GA implementation for solving hard problems with millions and billions of variables
• Scalable methods for solving design, optimization, and modeling problems of complex systems with modularity, hierarchy, and overlap

In all, nine patents have been filed, with one US patent issued, in connection with AF research in the recent past:


11 Honors and Awards

**NSF CAREER award.** Professor Pelikan received NSF CAREER Award (2006-2011) for the project *Design and Application of Scalable Hierarchical Optimization Algorithms by Combining Evolutionary Computation, Machine Learning and Statistics.*

**JSD Professor.** Professor Goldberg was named Jerry S. Dobrovolsky Distinguished Professor in Entrepreneurial Engineering in May 2003. The investiture was held September 23, 2003.

**ISGEC Fellow.** Professor Goldberg was named part of the inaugural class of Senior Fellows for the International Society for Genetic and Evolutionary Computation. (2003).

Student Honors

GECCO-2006 Best Paper Awards. Two papers from lab members won best paper awards at the 2006 Genetic and Evolutionary Computation conference organized by ACM SIGEVO.

Silver "Humies" award. IlliGAL teams entry, "Multiobjective Genetic Algorithms for Multiscale Excited-State Direct Dynamics in Photochemistry," was awarded the silver "Humies" medal at the human competitive results competition at the 2006 Genetic and Evolutionary Computation Conference. The teams entry was presented by Kumara Sastry. Other team members included Duane D. Johnson, Alexis L. Thompson, David E. Goldberg, Todd J. Martinez, Jeff Leiding, and Jane Owens.

GECCO-2007 Best Paper Nominations. Four papers from lab members are nominated for best paper at the 2007 Genetic and Evolutionary Computation conference organized by ACM SIGEVO. One paper was awarded.

Finalist, Lemelson-Illinois Prize. Kumara Sastry was a finalist for $30,000 Lemelson-Illinois Prize for the most inventive student at the University of Illinois.

Focused article of frontier research. Paper Genetic programming for multi-timescale modeling was chosen by the AIP editors as focused article of frontier research in Virtual Journal of Nanoscale Science & Technology, 12(9), 2005.

Two Bronze "Humies" Awards. Two projects from lab members were awarded bronze medal in the competition for the best human competitive results Humies (Human Competitive Results) at GECCO-2007.

GECCO-2008 Best Paper Awards. One paper from lab members won the best paper award at the 2008 Genetic and Evolutionary Computation conference organized by ACM SIGEVO.