1 Institutional Impact

Current and future national security applications at Los Alamos National Laboratory depend on a fundamental set of computer science problems, involving manipulation of large collections of interacting components and analysis of large data sets. Solving these requires effective algorithms that can operate in reasonable time spans. Additionally, many of the applications, such as organizing a communication network among distributed sensors, have constraints such as physical size, energy consumption, and the specific type of information that can be processed; we need efficient algorithms to optimize the systems under such constraints. Due to the scale of the data sets and the importance of the results, proposed methods must meet higher standards than those required in most industrial and commercial settings.

Faced with the need for better algorithms for large-scale discrete computational problems, scientists have increasingly turned to techniques from statistical physics, used to analyze large-scale complex systems and produce system-level predictions. Error probabilities in efficient, nearly optimal, error-correcting codes [1, 2] have been quantified by a mapping [3] to studies of disordered magnets. Predictive models for computational tractability in constraint satisfaction problems draw upon the phase transition between order and disorder [4]. These connections have led to novel algorithms such as *survey propagation* [5] that vastly outperform previously suggested methods.

We propose to develop algorithms for Laboratory applications using techniques from statistical physics. Our main prototype application will be the development of algorithms for wireless sensor networks. However, progress in this area will also have implications in other practical contexts, including social and biological networks, validation and verification, and data reconstruction.

LANL has been directly involved in the development of algorithms based on physical principles. The method of *extremal optimization* [6], developed at Los Alamos and motivated by the physics of self-organized criticality, has been successfully applied to community detection, image restoration, and sensor report pre-processing. Likewise, LANL scientists have devised a synchronization scheme for parallel discrete-event simulation [7] using the physics of surface growth. A variant of this scheme has been put to advantage in the *EpiSims* simulation of disease spread in social networks, leading to efficient containment strategies for targeted vaccination [8].

The impact of these individual contributions will be magnified with the development of an integrated, interdisciplinary Laboratory capability in algorithmic approaches drawn from physics methodologies. Our vision is a long-term program in the design and analysis of algorithms, drawing upon our strengths in the physical sciences and our recently developed capability in complex networks analysis. Our approach reflects the scientific priorities of the Laboratory's new Theory, Simulation and Computation directorate. Ultimately, our goal is to provide a dramatic reduction in the computing time and memory space required over a range of large-scale computational processes. We aim to find ways of solving *computationally hard* problems, whose complexity grows exponentially with problem size: if a problem with 100 components requires 2^{100} steps to solve, then even on a petascale machine it would take 10^7 years. Such tasks are prevalent in mission areas such as nonproliferation, homeland and infrastructure security, and weapons research. Efficient algorithmic solutions are indispensable.

Our focus application of wireless sensor networks has direct importance in non-proliferation and counter-terrorism. Distributed networks are rapidly proving to be an essential tool in detection challenges, whether biological, chemical, radiological, or image-related. Considerable efforts have been expended into the engineering aspects of implementation and deployment. Nevertheless, to obtain reliable performance of these networks under extreme limitations such as power constraints, numerous basic algorithmic issues still remain. These include routing, broadcast frequency assignment, quality of service and coding.

We will also consider other application challenges that are closely related — in many cases generalizations of problems in wireless sensor networks — and that may be addressed through similar algorithmic techniques. Community detection in adversary social networks and resource allocation in parallel computing architectures are issues of central importance to the Laboratory. Significantly, they all share basic graph partitioning problems that are impractical to address using classical methods but are excellent candidates for the methods we propose. Verification and validation is another key challenge for Los Alamos: formal verification is a state-of-the-art approach based on constraint satisfaction solvers to debug software. We expect that the survey propagation methodology discussed below will provide an algorithmic framework making formal verification possible on large-scale codes.

Los Alamos is well positioned for success in this area. The proposed research fits into a long-standing scientific tradition at the Laboratory of leveraging strengths in the physical sciences to innovate in information and computer science. Furthermore, our project team includes physicists, experts in network science, computer scientists and engineers already working in close collaboration. Our team's key accomplishments include the development of fidelity analysis in optical communications [9] and coding theory [10], the extremal optimization algorithm [6], and the establishment of LANL as a leading institution in the statistical physics of infrastructure networks.

2 Research Goals and Objectives

2.1 Prototype Application

Our team will focus on communication problems in Wireless Sensor Networks (WSN). These networks consist of many small, inexpensive, low-power devices ("motes") that communicate with each other and/or with a base station, and are deployed for very specific application such as video surveillance or seismic detection. The *MIT Technology Review* labels WSN the top emerging technology on the horizon and the National Research Council reports that the pervasive use of WSN may dwarf previous advances in information technology [11].

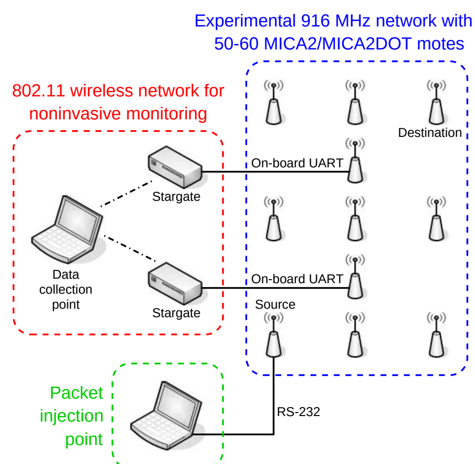

Figure 1: WSN built at LANL [12].

Unlike wireless communication devices in common use today, such as laptops or mobile phones, a WSN device cannot normally have its energy recharged once it is deployed. This severe energy limitation is critical to WSN, in the amount of information that each node can process and even more importantly in the amount of information that each node can transmit. The challenges of developing large-scale WSN require integrated network protocols including routing, media access control, and coding [13, 14, 15], while practical algorithms must be implemented in a distributed fashion across the network. Examples of significant algorithmic issues include:

• *Network Self-Configuration.* Consider a mobile, dynamic WSN. It is desirable for communication to be transmitted in a self-organized and hierarchical way. In an "ad-hoc" network of this kind, information for an individual sensor propagates to its ultimate destination ("sink") in multihops: sensor $\rightarrow$ head sensor $\rightarrow$ another head sensor $\rightarrow \cdots \rightarrow$ sink. For the network to provide optimal coverage, it must be self-reconfiguring: "communities" of individual sensors reporting to a head sensor need to be reassigned regularly, according to the changing nature of the signal. This reassignment involves the (computationally hard) problem of *community detection* or clustering: identifying an efficient partitioning of the sensor population in a distributed manner, so as to assign groups of sensors to head sensors while minimizing communication bandwidth.

• *Node Signature Assignment.* Multiple access protocols raise an important coverage question. In assigning a limited number of signatures to a network of sensors, one needs to avoid placing the same signatures on neighboring nodes, or more generally on nodes positioned within a direct communication reach. The task of finding a nonconflicting signature assignment is equivalent to *graph coloring*, a problem that is NP-complete (in the hardest computational complexity class) and is closely related to constraint satisfaction and automated reasoning.

• *Network Quality of Service.* The following questions are relatively easy to answer: (1) If each link in a network is characterized by unit capacity, what is the maximum flow of information achievable between any given sensor and the sink? (2) What is the shortest path connecting a sensor with the sink? These problems are solved in polynomial time, by the max-flow and Dijkstra algorithms [16]. However, the following *multi-objective* modification of (2) turns the problem into a hard one. Is there a path whose length is shorter than a given value, consuming less than a given amount of energy? This question is at the heart of energy-aware Quality of Service (QoS) algorithms. The ability to answer it efficiently will provide a solid foundation for designing routing protocols.

• *Coding and Data Reconstruction on Networks.* Correlations between data received by closely positioned sensors may provide an opportunity to combine data coming from many different sources enroute, thus eliminating redundancy, minimizing data transmission and saving energy. This idea is known as *data aggregation*. One approach is the Data-Centric (DC) protocol, where head sensors aggregate multiple input packets and subsequently perform joint data compression [17]. A complementary approach is Distributed Source Coding (DSC) [18], where correlated sensors already compress their data at the source, and head sensors perform joint decoding. Optimal coding is NP-hard (computationally hard) under both DC and DSC. Finally, closely positioned sensors can cause broadcast interference and data corruption. This leads to the difficult inference problem of determining the most likely original data recorded by each individual sensor.

2.2 Research and Development Approach

Our project's main application testbed will be WSN, but our vision is broader: a new method of algorithmic development using the emerging scientific discipline of statistical physics of algorithms [19, 20, 4]. We will address a suite of problems in networks and computer science for which current methods are not sufficiently powerful. These problems, typically involving large distributed systems, are tightly linked to each other through common algorithmic approaches.

Our overall strategy is to map the underlying problems onto physical systems of interacting spins or particles, and then apply statistical physics techniques to suggest new algorithms. These techniques, developed to study systems at the macroscopic rather than microscopic level, are well suited to deal with large data sets and make typical-case predictions, validated by simulation. This approach has proven successful in numerous industrial applications. The PageRank algorithm, used by Google to sort through the World Wide Web, is at its core an efficient algorithm for solving the diffusion equation on heterogeneous networks. Microsoft has established a focused research effort that exploits the theory of phase transitions to improve the reliability of their software.

There are three key steps in our approach:

• **State the problem in terms of optimization and/or inference.** Combinatorial optimization provides the general framework for our studies, as there is a powerful duality between optimization and inference. Community detection, where we search for the most likely assignment of communities, can be recast as optimization of a configuration likelihood over a large number of possibilities. Conversely, optimization can be viewed as the problem of finding the most probable pre-image, and thus can be cast as an inference problem. What makes these twin challenges hard is the phenomenon known to physicists as "frustration": the presence of conflicts that must be minimized but cannot be eliminated. In WSN, for example, reducing energy consumption conflicts with maintaining a given communication bandwidth. In the absence of frustration, one could minimize each term of the objective function separately. Statistical physics has a powerful set of tools for analyzing and solving problems exhibiting frustration. We will also employ graph theory, another formalism common to both statistical physics and computer science.

• **Develop and extend a heuristic algorithm.** A heuristic is an approximate, or *incomplete* algorithm: one that does not necessarily find the best solution. In emerging technologies such as wireless sensor networks, where high quality solutions must be found quickly, this is often the method of choice. However, as with asymptotic approximations that rely on the leading-order term in a series, one may view a heuristic algorithm as the first step in a "series" of systematically improvable approximations. This is especially effective if the first order heuristics become exact in some special cases or in certain limits. This approach provides us with a formal way of quantifying iterative improvements to algorithms. We will explore this idea of developing new approximate algorithms by starting with algorithms running in polynomial time, such as *belief propagation* [1], *minimal path* and *max-flow* [16] algorithms, and systematically extending them. We will also study and develop other physics-inspired modifications of the polynomial algorithms, such as generalized belief propagation [21] and survey propagation [5].

• **Analyze the algorithm.** The best heuristics find solutions that are close to optimal in all but a few cases. Thus, belief propagation fails when loops in the underlying graphical structures become important. We will analyze the cases where the algorithm performs poorly. Rare event and extreme deviation theory enable us to quantify the probability of rare failures. Additionally, phase transition methods such as scaling, renormalization, and universality allow us to characterize algorithms that work well in certain regimes but perform poorly in others where the problem becomes intractable. Many of the problems we consider, both in WSN and in other application areas, are defined on ad hoc networks that update dynamically. In order to validate our methods, we will model network and algorithm dynamics using computational methods based on Molecular Dynamics, including large-scale simulations.

3 Proposed Research

3.1 Community Detection

Community detection is the problem of dividing a network into communities, such that nodes are highly connected within a community but weakly connected between communities (see schematic in Fig. 2). Solving this problem in a rapid, distributed manner is key to efficient routing in ad hoc wireless networks. More generally, it is the central approach to detecting clustered structures on complex networks: cliques in adversary social networks, functional motifs in biological interaction networks, etc. [22]. The closely related graph partitioning problem is important to high-performance computing and parallel processing, for allocating assignments to different processors while minimizing inter-processor communication.

Existing algorithms, while often successful at revealing hidden structures in networks, use largely informal definitions of community. We propose the following systematic approach. Given a *hypothetical* community assignment for a network, assume that two nodes are linked with probability p_{in} if both lie within the same community, and p_{out} if not. The ratio p_{in}/p_{out} is then sufficient to compute the likelihood that the assignment produces the network in question. Define the optimal community assignment as the one maximizing the likelihood. This maximization is an NP-hard problem, equivalent in statistical physics to finding the ground state of a frustrated Potts model with conflicting short- and long-range interactions. Assuming $p_{in} > p_{out}$, the short-range interactions are ferromagnetic, favoring the assignment of neighboring nodes to the same community, while the long-range interactions are anti-ferromagnetic and prevent all nodes from simply lying within the same community.

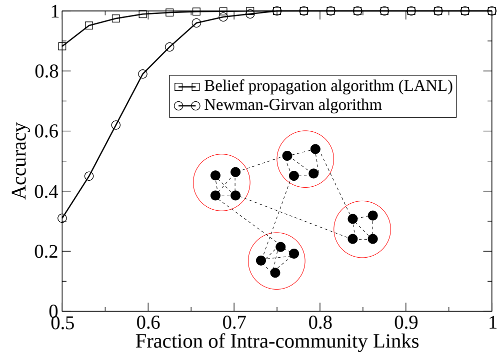

Figure 2: Schematic of community detection and results from our belief propagation algorithm and Newman-Girvan algorithm [23], over a set of computer-generated networks with 128 nodes split into 4 different communities.

We solve this problem using *belief propagation*, originally developed for inference problems in coding theory but surprisingly powerful here as well. We assign some basic probability (belief) that each node has a certain community assignment. We then update the probability for each node based on the current beliefs of its neighbors, using that information to determine its most likely new community assignment. The process continues iteratively, and can be readily implemented in a distributed fashion. Although there is no guarantee that the method will converge, our preliminary results (Fig. 2) suggest that this does indeed happen, within $O(n \log n)$ steps. Moreover, the community assignments it produces are significantly improved compared with existing algorithms.

To test the efficiency and reliability of our algorithms for problems in network science, we will analyze large-scale data sets with known community structure. One crucial example drawing upon our ongoing research in statistical analysis of bibliometric databases [24] is the collaboration network between scientists defined by co-authorship relations. The data for this network are extensive, highly accessible, and reliable. In parallel, we will consider routing in ad hoc networks, using testbed data from LANL's Distributed Sensor Networks project. Here, numerical modeling will be needed to validate our algorithmic results. We propose to use Molecular Dynamics simulations to develop networks with known community structure, testing our predictions against these numerics. The simulations will also enable us to study the effects of transitory community structures and community history, essential to understanding the dynamics of their formation.

3.2 Graph Coloring and Satisfiability

A basic combinatorial problem in computer science is coloring the nodes of a graph with a limited number of colors so that no two neighbors have the same color. An example application is the assignment of a limited number of frequency channels to broadcasting devices so that no two sufficiently close devices encounter a channel conflict. Further applications range from scheduling to register allocation in compilers. Since graph coloring is NP-complete, it may require exponentially long time to solve. Many practical instances of graph coloring, however, can be solved efficiently, in particular using algorithms based on methods from statistical physics. Recently it has been realized that the computationally difficult cases are often limited to certain special cases associated with a phase transition. This underlying phase space structure provides important clues to the existence of a solution and the appropriateness of particular algorithmic methods. For randomly generated graphs with a sufficiently low density of links, the problem is easy to solve. As the link density approaches a critical threshold, it remains solvable but becomes increasingly compuationally demanding to solve. Heuristically, the key in this region turns out to be a generalization of belief propagation known as *survey propagation* [5], operating on families of solutions rather than individual solutions. As the link density is further increased, a threshold is crossed and with high probability the problem is no longer solvable.

We will consider graph coloring instances that arise in the CDMA protocol for WSN communications, where a solution represents a conflict-free channel assignment. In a dynamically updating network, one needs a fast distributed method for reallocating broadcast channels as sensors are added, go into or out of power-saving "sleep" mode, or move to new positions. Survey propagation is well-suited to this purpose, and we aim to implement variants on the algorithm directly in TinyOS, for use in the Berkeley/Crossbow motes deployed in LANL's WSN testbed.

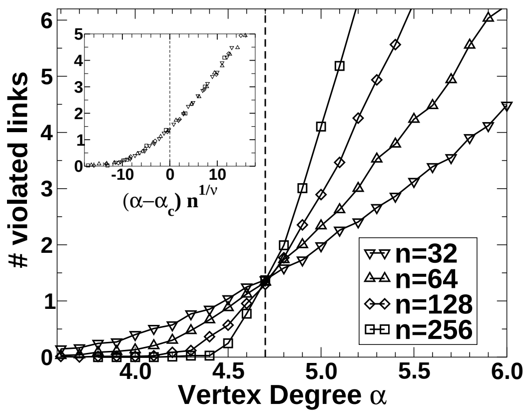

Figure 3: Finite size scaling in graph coloring with 3 allowed colors [27]. Average number of violated links is shown for random graphs of increasing link density (degree α), at various sizes. Colorable graphs have 0 violated links. Correct choice of phase transition location (vertical line at $\alpha_c \approx 4.70$) and scaling exponent ν result in data collapse onto single function (inset).

A related NP-complete problem with a similar phase structure that can again be addressed using survey propagation is *satisfiability* (SAT): given a set of logical clauses acting on Boolean variables, determine whether values can be assigned to the variables so as to satisfy all clauses. SAT is at the core of automated reasoning methods in computer science: model checking, theorem proving and software debugging all rely on solving large SAT instances [25]. We will consider large-scale SAT instances in the context of *formal verification*, representing the logical flow of a section of code in Boolean logic so that a satisfying assignment signals an *undesirable* event, typically a temporal exception or concurrency violation. An example is the assignment of process ID numbers in a Linux SMP kernel: one wants to avoid a "race condition" where two processors simultaneously try to assign the same number to two different processes. This maps to SAT instances that are often too large for present solvers, but within reach of survey propagation. We will develop such methods on this class of instances.

A major benefit of the phase diagram, however, is that lets us predict whether large instances are solvable without actually having to solve them. We intend to extend the theory developed on random graphs to the specific cases that arise in WSN and formal verification. Using our recent work on graph reduction heuristics [26], we will develop probabilistic models that generate ensembles of smaller instances while maintaining the desired structure. We will then locate the phase transition point using the *finite size scaling* numerical method from statistical physics on these ensembles, as we have successfully done for graph coloring on random graphs (Fig. 3). A further approach is to use a kinetic theory approach that we have developed recently [28]. We will represent in graphical form the process leading to the phase transition, modeling it as plaquette aggregation. Solving for the size distribution of these plaquettes gives a theoretical prediction for the transition location, which we have confirmed for a specific, tractable form of SAT.

3.3 Multi-Objective Shortest Path

The preceding problems are readily translated into the statistical mechanics framework of minimizing a Hamiltonian function — a scalar energy quantity — over a large state space. In the *multi-objective shortest path* (MOSP) problem, however, we must minimize two or more functions simultaneously: links on a graph now have associated weights that are vector quantities. The MOSP problem arises in a variety of applications. Quality-of-Service routing for WSN communications requires minimizing with respect to both energy consumption and network delay [29]. Problems in military logistics involve optimizing troop movements with respect to travel time and concealment costs.

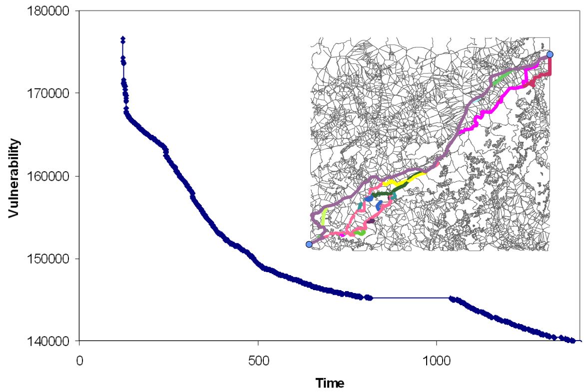

Figure 4: Image shows non-dominated paths joining two points in a road network with 25,000 troop mobility segments, each having a travel time and concealment cost. All paths pass through a single checkpoint, indicating strategically significant positions. Plot shows concealment cost vs. travel time among non-dominated paths.

MOSP is qualitatively more difficult than the equivalent single-cost problem: whereas Dijkstra's shortest-path algorithm runs in polynomial time and there are many efficient implementations, MOSP is NP-hard. The reason is that in general, no single path is optimal with respect to all objectives. Solutions consist of *non-dominated* paths, called the *efficient set*, such that no other path can improve on all objectives. In principle, a variant on Dijkstra's algorithm can generate the efficient set, but this could involve tracking an exponentially growing number of intermediate solutions as the algorithm progresses.

As with graph coloring and satisfiability, however, in many cases MOSP does not exhibit worst-case complexity [30]. For instance, when the costs on non-dominated paths lie on a convex hull, there are efficient search strategies for solving the problem by solving repeated instances of the single cost shortest path problem. We will determine the general conditions leading to qualitative changes in complexity, which we anticipate are associated with phase transitions. Based on our analysis of real instances of troop mobility problems (Fig. 4), it appears that certain classes of graphs allow bounds on the solution set size that significantly limit growth of intermediate solutions. Some of these bounds can be generated by solving a set of related single-cost shortest-path problems, enabling efficient decomposition strategies for solving the MOSP problem.

We will also consider the *structure* of the efficient set. In the case of troop mobility, we have identified discontinuities in concealment costs vs. travel time (plot in Fig. 4), representing points in solution space where the tradeoff between the two changes significantly. We expect that these points divide the efficient set into subsets with distinctly different properties. Finally, we will study the effect of spatial inhomogeneity on the solution structure: large variations in the cost structure of the underlying graph that are present in both the troop mobility and WSN problems can cause the properties of the efficient set to change greatly under a small change in the network such as the loss of a single key node. We will determine the statistical features of these fluctuations and study how they scale for large networks.

3.4 Distributed Coding and Data Reconstruction

The goal of coding is to transmit data across a communications channel as efficiently and as accurately as possible. A transmitted image often contains enough redundancy to be compressed, reducing the amount of data sent. This is *source coding*, especially important for WSN where power sharply limits transmission bandwidth. On the other hand, one might explicitly include some redundancy in the signal, to mitigate the effect of noise and interference in the transmission channel. Wireless networks operate on multiple-access channels, and even with channel assignments that in principle are non-conflicting, there is interference between communications on nearby channels. This is *channel coding*, improving accuracy but at the cost of additional data transmission. A related and more general problem is that of *data reconstruction*: inferring the most likely original data given the redundancy available. This can involve data corrupted not only by transmission but also by recording. An important application at Los Alamos is tracking turbulent flows from high-speed videos of fluid dynamics experiments at Los Alamos, where there is large frame-to-frame redundancy but also a significant noise level.

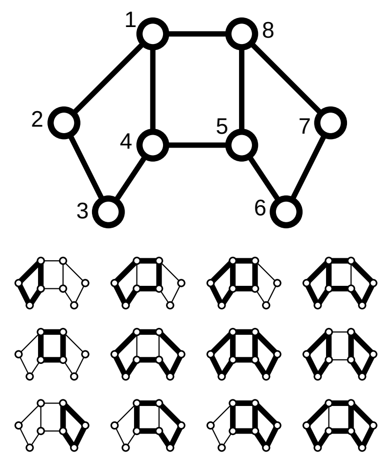

Figure 5: A graph and twelve generalized loops.

The traditional framework for source and channel coding involves a single pair of nodes: transmitter and receiver. New developments are needed when large number of components communicate, as with WSN. *Distributed Source Coding* is a modern proposal to compress data from multiple correlated sources [18]. Thanks to a recently discovered *duality* between channel and source coding, there are now opportunities to construct schemes capable of joint channel-source coding on these networks. Our work will begin at the physical level of the network, by understanding the correlations between sources on realistic models for practical applications such as bio-, chem- and radiological sensing. We will then design and analyze improved decoding algorithms based on a graphical representation in which bits are neighbors if they are both involved in a given parity check, a basic operation in error-correction. This inference graph will be sparse, making belief propagation a viable choice for gaining information about a given bit from statistical information on neighboring bits. To analyze the performance of these new schemes, we will adopt and extend our *instanton* approach, characterizing the most probable noise configuration using rare-event analysis [10].

For the more general problem of data reconstruction, our approach is to develop a statistical model that characterizes the properties of data corruption. In the case of fluid experiment data, this involves both random noise and geometrically structured interference. We will map the inference problem of finding the most likely pre-image to an optimization problem on a spin system, just as for community detection. Although this is computationally hard, in some cases we can develop efficient heuristics. If the underlying structure of the graph is tree-like, e.g., head sensors in a WSN, belief propagation is the method of choice: it is provably exact when the graph has no loops, and still highly effective when the graph has no small loops. However, even for video pixels, structured on a lattice with a high density of small loops, there is a systematic approach to improving on belief propagation. We have recently developed a loop calculus method [31], relating belief propagation to the leading term in a finite diagrammatic series. The addition of the terms in the series (Fig. 5) yields a *complete* algorithm on the graph with loops: one that is guaranteed to find the optimal solution. We will modify belief propagation to account for the most relevant terms in the loop series, starting with the case of a perfect lattice relevant to video images. This will extend and develop a methodology, *generalized belief propagation*, that has already been used successfully for multi-user detection problem in cellphone networks [32].

4 Schedule of Deliverables and Milestones

FY07:

• Develop algorithmic methods based on belief propagation. For community detection algorithm, determine robustness, scaling and convergence properties. Test on social networks generated from co-authorship database and EpiSims data. For graph coloring, develop distributed implementation of survey propagation appropriate for WSN channel assignment. Improve upon belief propagation method using loop series, and test on structured graphs.

• Define and develop realistic statistical models describing: spatial configurations of sensors in LANL's WSN testbed (Distributed Sensor Network project); geographic instances of troop movements; source correlations in biological and radiological sensing applications. Define specific satisfiability instances arising in software verification problems.

• Describe solution set structure for specific instances of multi-objective shortest path (MOSP) corresponding both to troop mobility and to energy-aware Quality of Service (QoS) on WSN. Analyze scaling and extreme statistics. Compare with single-objective solution structure in random graph coloring and satisfiability. Adapt techniques for studying phase structure from those problems.

FY08:

• Apply community detection and graph coloring algorithms to WSN testbed. Scale community detection up to full bibliometric database. Test performance when algorithm is extended to include survey propagation. Apply belief propagation with loop improvements to reconstruction of fluid dynamics experimental data from noisy high-speed video frames. Develop systematic survey propagation approach for satisfiability instances from software verification and evaluate effectiveness on race condition identification.

• Formulate duality between distributed source and channel coding given source correlation models developed in FY07.

• Describe phase structure for channel assignment, energy-aware QoS and troop mobility, based on statistical models developed in FY07. Develop kinetic theory model for random satisfiability.

• Determine bounds on intermediate solution set size for troop mobility problem, and apply to QoS. Develop improved MOSP algorithms that take advantage of solution and phase structure.

FY09:

• Test and implement algorithms on WSN testbed. Validate community detection for WSN using Molecular Dynamics simulations. Develop efficient distributed source-channel coding schemes, using specific duality formulated in FY08. Analyze algorithmic performance using "instanton" approach to quantify effects of error-producing rare events. Implement rapid QoS algorithms that employ solution set bounds. Develop reduced versions of distributed graph coloring survey propagation algorithm for implementation in Crossbow motes using TinyOS.

• Perform comprehensive experimental study of new algorithms based on testbed implementations, for main problem areas studied in project. Develop theory predicting quantitative performance of different extensions of belief propagation, explaining which specific extensions are well-suited for which problems, and why.

• Develop further algorithmic and analysis methods: apply community detection algorithms to additional social and biological networks; extend survey propagation methods in formal verification; analyze kinetic theory using exact methods and scaling techniques.

5 Budget Justification

The total budget is $1,249K/ $1,311K/ $1,377K for FY07/08/09. M&S is $63K/ $66K/ $69K per year for visitors, external collaborators, computers and travel. Personnel include 3 postdocs (funded 0.5-0.6 FTE each), $161K/ $169K/ $177K and one GRA (0.5FTE, joint with ISR Distributed Sensor Networks project), $32K/ $33K/ $35K. The allocation of funds by division is: T-Div $725K/ $762K/ $800K, CCS-Div $428K/ $450K/ $472K, D-Div $95K/ $100K/ $105K.

References

[1] R. G. Gallager, *Low density parity check codes* (MIT Press, Cambridge, 1963).
[2] C. Berrou, et. al Proc. IEEE Conf. , 1993, Geneva; vol. 2, p.1064–70.
[3] N. Sourlas, Nature **339**, 693 (1989).
[4] A. Percus, et. al *Computational Complexity & Statistical Physics* (Oxford Press, 2006).
[5] M. Mezard, G. Parisi, R. Zecchina, Science **297**, 812 (2002).
[6] S. Boettcher, & A. Percus, Artificial Intelligence **119**, 275 (2000).
[7] G. Korniss, M. A. Novotny, H. Guclu, Z. Toroczkai, and P. A. Rikvold, Science **299**, 677 (2003).
[8] S. Eubank, et. al, Nature **429**, 180 (2004).
[9] M. Chertkov, et al. PNAS **98**, 14208 (2001); JOSA B 19, 2538 (2002); (2003); JLT **22**, 1155 (2004).
[10] M. G. Stepanov, et. al Phys. Rev. Lett. **93**, 198702 (2004); **95**, 228701 (2005).
[11] National Research Council, Computer Science and Telecommunications Board. *Embedded, Everywhere: A Research Agenda for Networked Systems of Embedded Computers.* (Washington, DC, 2001).
[12] A. M. Mielke, et al, IEEE Instrumentation & Measurement Magazine **8**, 33 (2005).
[13] J. Stankovic, ACM SIGBED Review, July 2004.
[14] K. Akkaya, M. Younis, Ad Hoc Networks **3**, 325 (2005).
[15] W. Ye, J. Heidemann, D. Estrin, IEEE/ACM Transactions on Networking, **12** 493 (2004).
[16] T.H. Cormen, C.E. Leiserson, and R.L. Rivest, Introduction to Algorithms, (MIT York 1998).
[17] B. Krishnamachari, D. Estrin, S. Wicker, Distributed Event-Based Systems ’02, Vienna, July 2002.
[18] Z. Xiong, A. D. Liveris, and S. Cheng, IEEE Signal Processing Magazine, Sep. 2004, p.80.
[19] C. Gomes and B. Selman, Science **297**, 784–785 (2002).
[20] M. Mezard, Science **301**, 1685–1686 (2003).
[21] J.S. Yedidia, W.T. Freeman, Y. Weiss, IEEE IT**51**, 2282 (2005).
[22] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature **435**, 814(2005).
[23] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
[24] E. Ben-Naim, *LANL publications 1993-2003 a statistical study*, LA-UR-03-9087.
[25] E. Clarke, A. Biere, R. Raimi, Y. Zhu, *Formal Methods in System Design* **19**, 7(2001).
[26] V. Krishnamurthy, M. Faloutsos, M. Chrobak, L. Lao, J.-H. Cui and A.G. Percus, Proceedings of NETWORKING 2005. Lecture Notes in Computer Science 3462, 328.
[27] S. Boettcher and A.G. Percus, “Extremal optimization at the phase transition of the 3-coloring problem,” Phys. Rev. E **69**, 066703 (2004).
[28] E. Ben-Naim & P. L. Krapivsky, Phys. Rev. E **71**, 026129 (2005).
[29] S. Sichalou, L. Georgiadis, Computer Networks **43**, 351 (2003).
[30] A. Warburton, Operations Research **35**, 70 (1987).
[31] M. Chertkov & V. Chernyak, cond-mat/0601487 & cond-mat/0603189.
[32] O. Shental, et. al, 2004 IEEE IT Workshop, San Antonio, 2004, p.225.

A Overview of key Participants

Our interdisciplinary proposal is based on a very broad but highly expert team. This expertise has been established over numerous LDRD and programmatic efforts, where members of the team have worked together in close collaboration. Our team will be instrumental in building a strong Laboratory capability in the statistical physics of algorithms for national security applications.

In the past 3 years, the participants in this proposal have authored over 100 papers, including publications in Nature, PNAS, and over 30 Physical Review Letters. They have edited books and hold a patent on the subject of the project. The work has been featured and highlighted in media sources such as Nature, the New York Times, BBC News, Abq Journal, MSNBC, Discover Magazine, Scientific American, Physical Review Focus, and Technology Research News. Visit **http://cnls.lanl.gov/˜chertkov/alg.htm** for detailed CVs and the full list of relevant references.

Also within this time frame, they have given over 80 invited colloquia and seminars at such institutions as Harvard, MIT, Caltech, Yale, UCSB, U of Chicago, U of Illinois, Columbia, JHU, KITP and at conferences such as the APS march meeting, Newton Institute, Trieste, and SIAM. Additionally the participants have been organizers of seven conferences during this time.

A.1 Key Participants' Role:

• Chertkov, Ben-Naim, Hastings, Percus and Stepanov are responsible for developing a unifying set of physics-based approaches to the design, analysis, and improvement of algorithms. This sub-team will be working across the problems to cross-fertilize the theoretical physics ideas central for the project's success. The external collaborators Chernyak, Krapivsky, Mezard and Toroczkai, all internationally known physicists, will help with the theoretical methods development. Chertkov will lead this effort.

• Hansson, Alexander, Chertkov, Daniel, Izraelevitz and Percus are most directly responsible for implementing the algorithmic solutions on working prototypes of Wireless Sensor Networks. This group of researchers will be working in close contact with experimental researchers at LANL and elsewhere, building and testing these prototypes. Krishnamachari will be an external collaborator, bringing his expertise in algorithms and dynamics for Wireless Sensor Networks. Hansson will lead this effort.

• Hastings, Ben-Naim, Istrate, Percus and Reichhardt will work on improving existing algorithms and designing new algorithms for community detection. Hastings will lead this effort.

• Percus, Ben-Naim, Istrate and Alexander are responsible for developing computational complexity theory, phase transition and kinetic theory approaches for graph coloring and satisfiability. Moore is an external collaborator helping on the mathematical aspects of the project. This group will also focus on posing and efficiently solving the formal verification problem. Percus will lead this effort.

• Izraelevitz, Hastings, Ben-Naim, Daniel and Percus will be developing new approaches the to evaluation, analysis, and improvement of algorithms for the multi-objective shortest path problem. Izraelevitz will lead this effort.

• Stepanov, Chertkov and Hansson are responsible for developing novel information theory schemes for distributed coding and data reconstruction. Koetter is an external collaborator who brings outstanding expertise in information theory and electrical engineering to this problem. Stepanov will lead this effort.

• Ben-Naim, Reichhardt, Alexander, Percus and Stepanov have the role of performing large-scale distributed simulations in data clustering, optimization and error-analysis. Ben-Naim will lead this effort.

A.2 Brief Bio of Key Participants:

Michael Chertkov (PI, T-13, 0.5FTE) received his Ph.D. in theoretical physics in 1996 from Weizmann Institute. He was an R. H. Dicke Fellow at Princeton University, Physics Department and moved to LANL in 1999 as J.R. Oppenheimer Fellow in T-13/CNLS. He joined T-13 in 2002 as a Staff Member. His area of expertise includes statistical hydrodynamics, statistical and non-linear optics, information theory, error-correction theory, and soft condensed matter. He has one patent and over 50 publications in refereed journals including a PNAS and 13 Physical Review Letters. He has organized 5 CNLS conferences, including one in January 2005 on *Statistical Physics Approach to Coding Theory*. His achievements most relevant to this project are (1) discovery of an explicit relation between the belief propagation algorithm and a loop series representing the complete algorithm on a graph with loops (co-authored with V. Chernyak); (2) development of novel physics inspired approach solving the problem of the error-floor analysis of the coding theory (co-authored with M. Stepanov, V. Chernyak and B. Vasic).

Allon Percus (Co-PI, CCS-3, 0.5FTE) received his A.B. in physics from Harvard in 1992 and his Ph.D. from Orsay in 1997. In addition to his LANL duties, he is currently Associate Director of the Institute for Pure and Applied Mathematics at UCLA. His main research interests are in discrete optimization and statistical physics. Much of his work is on the interface between these two fields. In his research on the stochastic traveling salesman problem, he produced the most precise numerical estimate to date for large n asymptotic tour lengths. He was a co-founder of the Extremal Optimization method, which has been used successfully on several hard combinatorial optimization problems. He has led several previous LDRD/ER efforts, including *Extremal Optimization* and *Improving Local Search*. He has organized numerous conferences and workshops on combinatorics, phase transitions and algorithmic complexity, including a symposium on *Phase Transitions in Computer Science* at the 2004 AAAS Annual Meeting that highlighted breakthroughs in statistical physics, computational threshold phenomena and coding theory. Together with Gabriel Istrate and Cristopher Moore, he has edited a book on *Computational Complexity and Statistical Physics*, published by Oxford University Press in 2006.

Frank Alexander (CCS-3, 0.1FTE) received his Ph.D. in Physics from Rutgers Univ. in 1991. He was a post-doctoral fellow at CNLS/LANL in 1991-1993 and at LLNL in 1993-1995. At Livermore he developed algorithms for computational kinetic theory and hybrid numerical methods. From 1995 to 1998 Frank was a research assistant professor at the Center for Computational Science at Boston University. While in Boston, he developed methods to improve computation in nonequilibrium physics. In 1998 Frank moved back to Los Alamos as a staff member in CIC-19 (now CCS-3). Since his return, he has been working on a variety of problems in the above areas as well as in applying methods of statistical physics to modeling complex systems and time series analysis. Frank became team leader in 2000 and he is Deputy Group Leader since 2002. Among the many conferences he has organized is the March 2006 CNLS Workshop on *Challenges and Opportunities in Distributed Sensor Networks*.

Eli Ben-Naim (T-13, 0.5FTE) received his Ph.D. in Physics from Boston University in 1994, was postdoctoral research associate at the University of Chicago (1994-1996), Director's postdoctoral fellow at Los Alamos (1996-1998), and a technical staff member at the Complex Systems group (1998-present). He is now the Complex Systems (T-13) Group Leader. He is an expert in statistical physics, nonlinear physics, and random processes with fundamental contributions to the kinetic theory of granular materials and traffic flows. His recent works on kinetic theory of random structures including random graphs and random trees are directly relevant for the proposed research. He authored over 85 publications in peer-reviewed journals. He serves on the editorial board of Physical Review E (the leading statistical and nonlinear physics journal) and the advisory board of Journal Physics A.

Brent Daniel (D-3, 0.1FTE) received a Ph.D. in physics from Ohio State University before going on to a postdoctoral position at Los Alamos National Laboratory's Center for Nonlinear Studies. His efforts, in collaboration with experimentalists in MST-10, were focused on investigating transfer mechanisms in two- and three-dimensional turbulence. A substantial amount of image analysis and particle tracking software was written as part of this work. Following completion of his postdoc, he became a Technical Staff Member in LANL's Decision Applications Division. Recent work has focused on developing statistical methods for siting bioaerosol collectors; using Bayesian techniques to reconstruct bioterrorist attacks; and describing the dispersion of contaminants in water distribution networks. He has also led comparative live-agent tests of deployable bioaerosol collection platforms and a study of the soil-borne pathogens that can lead to low-level nuisance alarms on bioaerosol monitoring systems.

Anders Hansson (CCS-DO, 0.5FTE) received his Ph.D. in communication theory from Chalmers University of Technology, Göteborg, Sweden, in 2003. He spent the 2000-2001 academic year as Sweden-America Foundation Fellow at the Communication Sciences Institute of USC. During 2003-2005, he was a Postdoctoral Research Associate at LANL, and since January 2005, he is a Technical Staff Member in CCS-5. Dr. Hansson's area of expertise includes spatio-temporal fading, receiver front-end processing, continuous phase modulation, and adaptive soft-input soft-output algorithms. He collaborates with the ISR project on *Distributed Sensor Networks,* and is currently supervising a student, Matthew Nassr, working part-time on that project. Parts of his research have been implemented by TrellisWare Technologies, San Diego, CA, for deployment in U.S. government communication systems. His recent interests include simulation and modeling of large-scale socio-technical systems, acceleration of high-performance computing algorithms through reconfigurable supercomputing, and wireless networks. He has published 25 papers, including four papers in IEEE Trans. on Communications.

Matthew Hastings (T-13, 0.75FTE) received his B.A. in physics and mathematics magna cum laude from Yale in 1994, and his Ph.D. in theoretical physics in 1997 from the Massachusetts Institute of Technology. His seminal Ph.D. thesis developing the conformal mapping algorithm for diffusion-limited aggregation led to his receiving the Lockett Award given to the best Ph.D. work in theoretical physics at MIT. He was an R. H. Dicke Fellow at Princeton University and then an R.P. Feynman Fellow in T-13/CNLS. He joined T-13 as a staff member in 2004. His area of expertise includes statistical physics, networks, non-equilibrium dynamics, strongly correlated electrons, and soft condensed matter. He has published 16 papers in Physical Review Letters. His work has been featured in Nature News and Views, Physical Review Focus, the Bell Labs Condensed Matter Journal Club, and Technology Research News. He received the LANL postdoctoral distinguished award for his work on Ratchet Cellular Automata.

David Izraelevitz (D-6, 0.25FTE) received the B.S. magna cum laude in Computer and Systems Engineering from R.P.I. and the S.M. and Ph.D. degrees from M.I.T. His current work involves development of systems dynamics models of critical metropolitan infrastructure segments to support homeland security decision-making, and multisource fusion algorithms for military intelligence applications. Prior to joining LANL he developed a variety of graph algorithms for military terrain analysis applications at SRI International. His areas of interest are statistical estimation techniques for data mining and signal processing applications and network optimization algorithms, especially under multiple objectives.

Gabriel Istrate (CCS-5, 0.25FTE) received his Ph.D. in Computer Science from the University of Rochester in 1999. His Ph.D. thesis deals with exactly-solved models of phase transition in combinatorial problems. He moved to LANL in 1999 as a Director Funded Postdoc, and became a staff member in 2001. His area of expertise includes computational complexity, probabilistic analysis of algorithms, combinatorial optimization and simulation of sociotechnical systems. He has authored 27 papers, 10 of which dealing with phase transitions in combinatorial problems. He served on the Program Committee for ICALP (the most prestigious European conference in Theoretical Computer Science) in 2005 and is co-editor (together with Allon Percus and Cris Moore) of the book *Computational Complexity and Statistical Physics*.

Charles Reichhardt (T-13, 0.25 FTE) started as a staff member in T-13 in 2004. Previously he was an R.P. Feynman Fellow, and before that a postdoctoral researcher at the physics department at the University of California, Davis. He received his Ph.D. in physics from the University of Michigan in 1998. As an undergraduate studied physics and mathematics at the University of California Irvine in 1993. His research is in non-equilibrium physics of in soft condensed matter and solid state physics. Particular techniques he uses and develops are large scale computational methods for many-particle interacting systems to study phase transitions and critical behavior. He has published over 75 papers including 25 in Physical Review Letters. He is the winner of the 2003 LANL distinguished postdoctoral award and a 2004 LANL research achievement award.

Mikhail Stepanov (T-13, 0.75 FTE) was “Soros” graduate student for three years in a row and received his Ph.D. in physics in 1999 from the Institute of Automation and Electrometry, Russia. He was a postdoc at the Weizmann Institute (1999-2001), member of the Institute for Advanced Studies, Princeton (2002-2003), and joined CNLS/T-13 as a postdoc in 2004. His area of expertise includes nonlinear spectroscopy, fiber optics, turbulence, pattern growth, data clustering and error correcting codes. He has 19 publications in refereed journals, including 1 in Nature and 5 in Physical Review Letters. His expertise most relevant to this project is in numerical analysis of clustering algorithms and instantons in statistical physics and information theory.

A.3 External Collaborators:

Our advisory board consists of the following seven external collaborators. These collaborators are all internationally well-known scientists and cover a broad range of expertise. They have all agreed to serve on the board, and will convene yearly to review progress and plan future work.

Vladimir Chernyak is Professor at Wayne State University. He is an internationally leading expert in statistical field theory, quantum optics, physical chemistry and fiber optics, publishing more then 180 papers cited more than 2,600 times total. He has four years of industrial experience in optics communication as a Senior Research Scientist at Corning, Inc.

Ralf Koetter is Professor at the Coordinated Science Laboratory, UIUC. In 2000, he started a term as associate editor for coding theory of the IEEE Transactions on Information Theory. He received an IBM Invention Achievement Award in 1997 and an IBM Partnership Award in 2001. He is a member of the Board of Governers of the IEEE Information Theory Society. He was recently awarded the 2004 Best Paper Award by the IEEE Information Theory Society. For further information see: http://ww.comm.csl.uiuc.edu/~koetter .

Paul Krapivsky, Professor at BU, is an internationally recognized figure in non-equilibrium statistical physics with over 150 publications in the past 10 years. His recent work on the theory of random growing structures is seminal. His expertise relevant to this proposal includes rate equation analysis of data structure and phase transitions in computational complexity.

Bhaskar Krishnamachari is Philip and Cayley MacDonald Early Career Chair Assistant Professor at the Viterbi School of Engineering, University of Southern California. His research interests are focused on performance analysis and design of algorithms for information routing, and self-configuration in large-scale wireless sensor networks. He received the NSF CAREER award in 2004 and the USC Viterbi School of Engineering Junior Faculty Research Award in 2005. He has published extensively on sensor networks and phase transition phenomena. His book ”Networking Wireless Sensors” was released by Cambridge University Press in 2005.

Marc Mezard is a *Directeur de Recherche* in CNRS (Orsay). He has received the golden medal of CNRS in 1990 and the Ampere prize of the French Academy of Sciences in 1996. He has been PI of a European network of 13 European laboratories on ”Statistical physics of collective behavior in disordered systems and information processing” in 1993-1997 and is presently the PI of a European network of 10 laboratories on ”Statistical physics of information processing and combinatorial optimization”. He is the inventor of the *survey propagation* algorithm.

Cristopher Moore is Professor at UNM. He has over 80 publications. Recent work relevant to this proposal includes studying the structure of the internet, the phase transition in satisfiability, and the computational complexity of physical simulations.

Zoltan Toroczkai is former Deputy Director of CNLS, who recently accepted a position as Professor at the University of Notre Dame. He is the author of over 50 publications, including 2 Nature articles (on networks), 1 Science article (on parallel computing) 2 PNAS article (on chemical reactions and mixing) and 1 article in Physics Today (surface growth). His areas of expertise include complex networks, agent-based systems, statistical mechanics, fluid flow, dynamical systems and chaos theory.

B Commitments of key personnel to other projects

Proposed projects are indicated by *

Name	Project title	Relation of project to this proposal	Time and effort commitment
M. Chertkov M. Stepanov	ER06-08: Novel physics inspired approach to error-correction	The ER develops new approach to analysis of modern error-correction codes used in point-to-point noisy communication. The DR is devoted to developing algorithms and it does not address the problem of point-to-point communication at all, however progress in one project can be beneficial for the other.	0.3FTE 0.5FTE
M. Chertkov M. Stepanov	ER07-09: Reconstructing Statistics of Rare/Catastrophic Events in Fluid Mechanics: Numerics, Experiment and Theory*	none	0.1FTE 0.5FTE
M. Chertkov	DR05-07: Lagrangian measurements of fluid mixing	none	0.25FTE
M. Chertkov	WSR06: Prediction of Mixing Induced by Rayleigh-Taylor Instability	none	0.25FTE
A. Percus G. Istrate	WSR06: New Approaches to Fault Tolerance	The WSR studies how percolation properties affect reliability of computing. Most of the work for this lies "upstream" of algorithmic development, but each project could benefit substantially from the results of the other.	0.5FTE 0.5FTE
F. Alexander	ER07-09: Rigorous coarse-graining algorithms for predictive cellular modeling*	none	0.3FTE
E. Ben-Naim	ER06-08: Energy Distributions in Granular Flows	none	0.5FTE
B. Daniel	DR07-09: Building Scale Pathogen Exposure Modeling*	none	0.33FTE
B. Daniel	DHS: BioWatch	Much of our work for DHS relies heavily on solving tough optimization and inference problems. New approaches and fundamental improvements to the algorithms developed in the DR may help to increase the quality and timeliness of our BioWatch response.	0.5FTE
D. Izraelevitz	ER07-09: New Approach to Bayesian Inference Under Modeling Uncertainty*	none	0.5FTE

D. Izraelevitz	DHS06: Critical Infrastructure Protection Decision Support System	none	0.5FTE
D. Izraelevitz	DNDO06-07: Los Alamos Research, Innovation, and Analysis Team (LARIAT)	none	0.25FTE
A. Hansson G. Istrate	DR07-09: Damaged Transportation Infrastructure: Assessment and Flow Optimization*	none	0.4FTE 0.4FTE
A. Hansson	ER07-09: Speeding up Contingency Studies*	none	0.25FTE
A. Hansson	ER07-09: A Framework for Adaptive, Scalable, and Robust Communication in Nanoscale Networks-on-Chip*	none	0.25FTE
M. Hastings	ER07-09: Series Expansions for Processes on Networks*	none	0.5FTE
M. Hastings	ER07-09: Accurate Divide-and-Conquer Schemes for Quantum Systems*	none	0.5FTE
C. Reichhardt	ER07-09: Critical Behavior at the Jamming Transition*	none	0.5FTE