Probabilistic-based Circuit Design for Bio-Medical Applications

Jie Chen
Division of Engineering, Brown University, RI 20912, USA

Abstract

Bio-medical systems are expected to be much more prone to signal faults than conventional electrical systems. Recent development in bio-medical research require a new approach to circuit design that can tolerate noisy data. In this paper, we propose probabilistic-based circuits utilizing the Markov random field. Simulation results show that these circuits are more noise immune than conventional circuits, and thus are more suitable for bio-medical applications.

I. Introduction

Many important technological innovations have been developed throughout the history of medicine. The invention of the X-Ray machine, for example, revolutionized medicine and pioneered modern imaging. The invention of the microscope essentially redefined the fields of pathology and microbiology. Tremendous progress has been made with the aid of numerous new technologies, such as recombinant DNA methods, DNA sequencing, MRI (magnetic resonance imaging), PCR (polymerase chain reaction), monoclonal antibodies, etc. High-throughput measurement technologies for biological molecules have also been developed. In this process there has been a shift from an individual approach (genomics, proteomics, etc.) towards an integrated approach. Enabling technologies include *microarrays* [1] and *high content imaging* [2]. These new integrated approaches have led to the emergence of the field of systems biology, which focuses on achieving a system-level knowledge of biological units, as shown in Fig. 1.

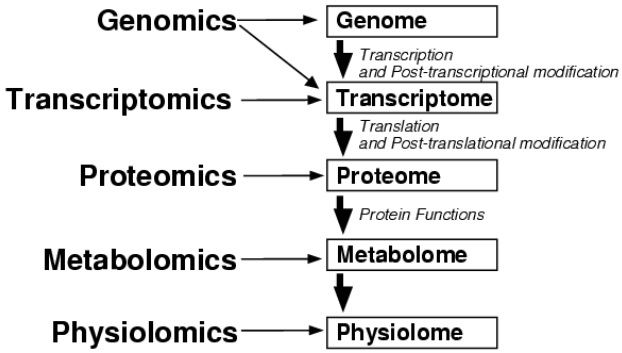


Figure 1: Schematic plot of what includes in system biology [3].

Microarray Technology Collecting large data sets via parallel processing is now a common practice in biological and drug research. Microarray technology has certain advantages over conventional approaches: these include decreased reagent usage, smaller sample requirements, and the robustness of using internal controls for comparisons. Microarrays can identify the functions and interactions of macromolecules in cells based on optical fluorescence. A microarray, or a "gene chip", is an orderly arrangement of oligonucleotide probes attached to a solid support. Some companies use 25-mer probes while others use 60-mer probes. A microarray measures the expression level of a gene by determining the amount of messenger RNA that is present [1]. Microarrays facilitate large-scale surveys of gene expression in which transcript levels can be determined for hundreds, thousands, or even tens of thousands of genes in a single experiment simultaneously. With the latest Affymetrix chip designs, the entire genome ($> 45,000$ genes) can be measured on a single array. All microarray experiments rely on the core principle that transcript abundance can be deduced by measuring the amount of labeled RNA that binds to a complementary probe. A microarray uses thousands of these probes; each probe represents the complement of at least a part of a transcript that might be expressed in tissue. Once the microarray has been constructed, the target mRNA population is labeled, typically with a fluorescent dye, so that hybridization to the probe can be detected when it is scanned with a laser. The intensity of the signal produced by 1,000 molecules of a particular labeled transcript could be twice as bright as the signal produced by 500 molecules. Similarly, the intensity produced by 10,000 molecules could be half as bright as one produced by 20,000 molecules.

High Content Cell Imaging The pharmaceutical industry is currently faced with an escalation in the amount of time and money needed for drug development. There is a growing need to increase confidence in early stage targets, improve lead selection, and reduce late-stage attribution. An equally important issue is how to choose the best therapies from an increasingly large drug arsenal for patients with various conditions. In addition to microarray technologies, another promising technology for improving such decision-making is cellular imaging-based assays. This technology enables functional analysis of target and pathway modulation in cells. High content screening methods, such as transcriptional and proteomic profiling, are discussed in [4], [5]. There are many types of high content cellular data, capturing DNA polymorphism, gene activation, mRNAs, proteins, metabolites, calcium flux, cytotoxicity, and proliferation. Processing this data has contributed substantially to our understanding of disease processes and drug actions. None of this data, however, can capture the important spatial and temporal information contained within functional cellular proteins.

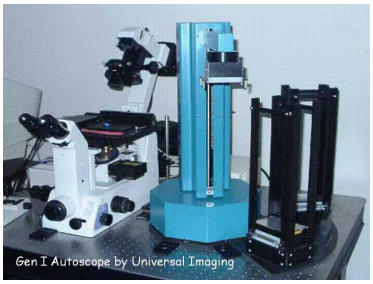


Figure 2: Gen I Autoscopy: An automated fluorescence imager in the Institute of Chemistry and Cell Biology, Harvard Medical School [2].

To harness the power of this new high content spatial and temporal information, we need to increase data acquisition throughput and to process and quantitate large amounts of image data effectively. By using automated microscopy, as shown in Fig. 2, combined with imaging techniques that show protein locations, cell structures and the physiological states of cells, researchers can increasingly observe systemic biological events such as multiple cellular events and intracellular events in individual cells [6]. Making quantitative cytological information available at early stages in the drug discovery process improves our understanding of drug targets and compound leads. However, the image informatics tools that are required to automate, quantitate and analyze cellular information on such large scales have not yet been developed. These are needed to complement the high throughput imaging technologies that have already been adopted. The availability of these informatics tools would enable automated microscopy and image analysis to be coupled with biostatistical and data mining techniques to provide an integrated systems biology approach to studying cells, the basic units of life. These informatics tools would also have many additional applications in the life and health sciences.

Compounds that affect the spatial arrangement of signaling proteins or cellular structures can provide important information about biological processes and therapeutic interventions. To identify such compounds using automated fluorescence microscopy is part of the "forward chemical genetics" approach [7]. Effort is mainly being put into high throughput image-based screening of large chemical libraries. Currently, about a dozen commercial vendors offer automated fluorescence imagers. Most, however, are assembled using off-the-shelf components. Their prices range from hundreds of thousands to over a million U.S. dollars for the high end products.

In the post-genomic era, researchers will access, analyze, and mine vast volumes and multiple types of data on a grand scale. New "computer-based approaches" or "algorithmic-based approaches" will tackle post-genomic challenges, such as functional genomics, comparative genomics, proteomics, metabolomics, pathway analysis, and systems biology. Microarrys and high-content imaging techniques are very promising for use in systems biology research. Signal faults, however, are unavoidable in data collected by these means, and their presence can cause prediction model failure. Uncertainty about signal faults is inherent in genomic signal processing systems. In this paper, we investigate circuit designs that operate in the presence of these signal faults. Probabilistic-based circuit design is the only way to deal with random and dynamic faults [8]. The reminder of the paper is organized as follows. In section II, we present our probabilistic-based circuit design. In section III, we present simulation results to verify our design. Finally, we conclude our paper in section IV.

II. Proposed Design

Error-free computing is assumed in current circuit design. This concept is, however, not applicable to designs in the bio-medical realm because signal errors in collected data via microarray and high-

content image techniques are *dynamic* in nature. *The fact that we cannot assume error-free computing requires new ways of thinking about designing circuits. Uncertainty about signal faults is inherent in bio-medical data processing systems; probabilistic design is the only way to deal with random and dynamic faults [8].* To deal with dynamic errors, we proposed to utilize the Markov random fields (MRF) [9] for designing probabilistic-based circuits.

The MRF represents the relationship between a set of random variables, $\mathbf{V} = \{V_1, V_2, \ldots, V_n\}$. Each V_i can take on values from a range set $\mathcal{S}$. The notation presented here follows the treatment in [10]. A *neighborhood* in the MRF, $\mathcal{N}_i$, is defined to represent the conditional dependence of a variable, V_i, on a subset of $\mathbf{V}$. The neighborhood can vary from complete dependence (the entire set $\mathbf{V}$) to complete independence (the null set). For instance, if the neighborhood of V_1 is $\mathcal{N}_1 = \{V_2, V_3\}$, we can compute its conditional probability as: $p(v_1|v_2, v_3, \ldots, v_n) = p(v_1|v_2, v_3)$. The probability of a given variable depends only upon a, typically small, neighborhood of other variables. A Markov random field can be formally defined as follows [11]:

$$p(v_i) > 0, \quad \forall \ V_i \in \mathbf{V} \quad (Positivity), \qquad \text{and}$$
$$p(v_i|\{\mathbf{V} - v_i\}) = p(v_i|\mathcal{N}_i) \quad (Markovianity) \qquad (1)$$

If and only if both previous conditions hold, $\mathbf{V}$ is said to be a Markov random field on a set of valuables $\mathcal{S}$, with respect to a neighborhood system $\mathcal{N}$.

The power of the MRF is that it can be used to describe any logic circuit. A logic circuit can be mapped onto a graph where nodes are random variables and edges are conditional dependencies between the variables. We take the M3 module of the ISCAS-85 C432 interrupt controller as an example circuit [12], and map it onto a MRF as shown in Fig. 3. In this application, the nodes correspond to logic signal terminals. The MRF neighborhoods of the MRF correspond to logic interactions. The objective is to maximize correct operational probability under structural and signal fault conditions.

There are many ways to design MRF architecture. We have applied an "auto-model" based approach to represent clique energy. A clique in a graph is a set of nodes where each node in the set has edges to the all other nodes [9]. For cliques up to order two, the energy function is given by: $E_c = \kappa + \Sigma_{i \in \mathcal{C}_0} \alpha_i v_i + \Sigma_{i,j \in \mathcal{C}_1} \beta_{ij} v_i v_j$. The constants α_i and β_{ij} are called *interaction coefficients*. The constant κ acts as an energy offset. Simple logic gates, such as XOR, have been simulated to demonstrate a MRF design [13]. The idea is that logic operation is achieved by maximizing the probability of state configurations in the logic network. Maximizing state probability is equivalent to minimizing a form of energy that depends on neighboring nodes in the network. For example, an exclusive-or (XOR) gate has inputs v_0 and v_1, and output v_2. The valid states of the XOR gate are $(v_0, v_1, v_2) = \{(000), (001), (101), (110)\}$. The clique energy can be computed by summing over the valid states based on the Boolean ring axiom as follows:

$$\begin{aligned} E_c &= -1 - (1 - v_0)(1 - v_1)v_2 - (1 - v_0)v_1(1 - v_2) \\ &\quad -v_0(1 - v_1)(1 - v_2) - v_0 v_1 v_2 \\ &= -1 + v_0 + v_1 + v_2 - 2v_0 v_1 - 2v_0 v_2 - 2v_1 v_2 \\ &\quad +4v_0 v_1 v_2. \end{aligned} \qquad (2)$$

To ensure the correct operation of an XOR gate, the energy E_c of valid state configurations, $(v_0, v_1, v_2) = \{(000), (011), (101), (110)\}$, has to be *less* than that of invalid state configurations $(v_0, v_1, v_2) = \{(001), (010), (100), (111)\}$. (please refer to our paper [14], [13] for a more detailed discussion).

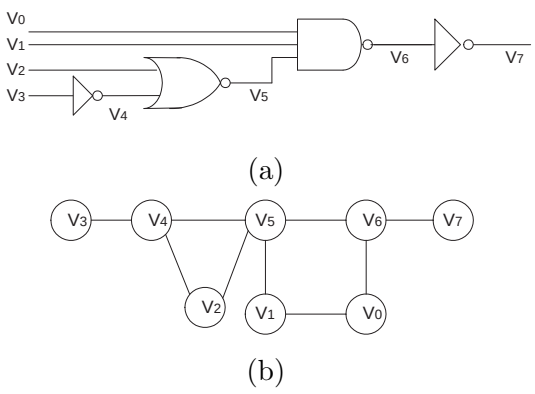
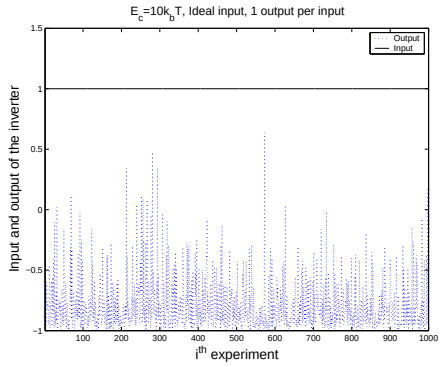


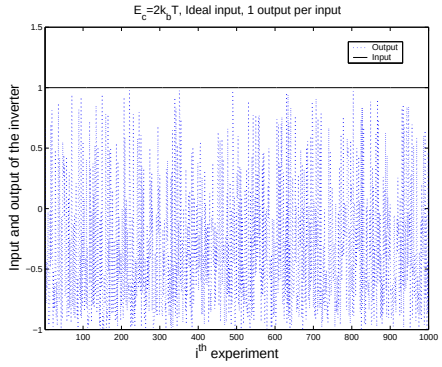
Figure 3: (a) An example circuit (the M3 of the ISCAS-85 C432 interrupt controller) (b) mapping the circuit onto a MRF. Here, lines between nodes indicate neighborhood relations.

Exploring Temporal Redundancy to Achieve Signal Fault Tolerance A general MRF is equivalent to the Gibbs form, which was established by the Hammersley and Clifford theorem [15] as: $p(x) = \frac{1}{Z} e^{-\frac{E_c}{k_b T}}$. The thermal energy is $k_b T$, and E_c represents the potential energy of a given logic state. If $E_c \gg k_b T$, the error probability $p(x)$ is negligible. Thermal energy fluctuations can easily cause logic errors when $E_c \simeq k_b T$. If the limit of $\frac{E_c}{k_b T} \to 0$, all logic states are equally probable. Let us take an inverter with input v_0 and output v_1 as

an example. Its clique energy can be expressed as $E_c = 2v_0v_1 - v_0 - v_1$ [13]. We want inputs to be distributed symmetrically around zero, and this can be achieved by mapping input values from $\{0, 1\}$ to $\{-1, 1\}$. A joint probability can then be obtained as follows: $p(v_0, v_1) = \frac{1}{Z} e^{-\frac{1}{2k_bT}(v_0v_1 - 1)}$. This Gibbs distribution can be used to model signal distribution over time from a previous logic stage, which can be simulated by drawing random samples from the Gibbs distribution. If we keep the input at logic "1", $v_0 = 1$, we can get a series of outputs as shown in Fig. 4 (a) and (b) when $E_c = 10k_bT$ and $E_c = 2k_bT$, respectively. To reduce signal error, we can simply increase clique energy $E_c \gg k_bT$ to resemble conventional silicon gates that operate at energy levels much higher than the thermal limit, resulting in very stable MRF logic states. We propose to exploit temporal redundancy to reduce signal faults or "soft errors" by taking the ensemble average of a series of output samples. However, low temperature operation (below room temperature) is not being advocated here.



(a)



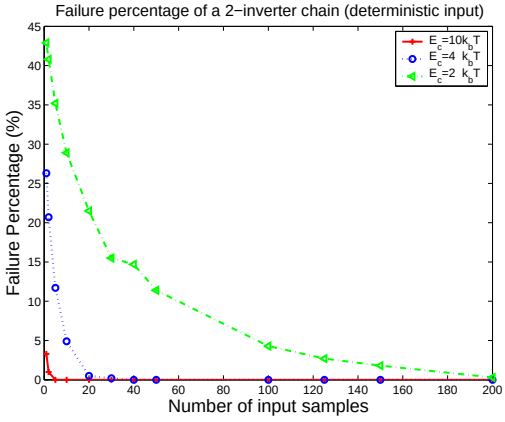
(b)

Figure 4: Simulation results of an inverter working under different energies (a) many orders of magnitude above the thermal limit $10k_bT$; and (b) around the thermal limit. Inputs are distributed symmetrically around zero, and this can be achieved by mapping input logic values from $\{0, 1\}$ to the real measurements of $\{-1, 1\}$ (after [14]).

The proposed design can be illustrated by sampling the output distribution of a two-stage inverter chain. The input is at logic "1", and the inverter's output distribution follows the Gibbs distribution, $p(v_0, v_1) = \frac{1}{Z} e^{-\frac{1}{2k_bT}(v_0v_1 - 1)}$. The table in Fig. 5 shows how temporal redundancy can help reduce signal faults. For instance, without temporal redundancy ($N = 1$), the failure percentage is 3.3% for $E_c = 10k_bT$, while it is 42.9% at $E_c = 2k_bT$ (refer to the first row in the table of Fig. 5 (a)). If we utilize temporal redundancy ($N = 200$), the failure percentage can drop to 0.3% (the third entry of the last row in the table).

N (samples)	E_C=10K $_b$T	E_C=4K $_b$T	E_C=2K $_b$T	E_C=K $_b$T
1	3.3	26.3	42.9	46.5
2	1	20.7	40.8	44.9
5	0	11.7	35.2	44.5
10	0	4.9	28.9	43.7
20	0	0.5	21.5	41.9
30	0	0.2	15.5	40.6
40	0	0	14.7	39.1
50	0	0	11.4	38.2
100	0	0	4.3	33.1
125	0	0	2.7	31.6
150	0	0	1.8	30.8
200	0	0	0.3	30.1

(a)



(b)

Figure 5: The effect of ensemble averaging for a chain of two inverters. The input to the first inverter is fixed at ideal logic "one". The output of this first inverter follows the Gibbs distribution. (a) A table showing percentage of failure vs. the number of samples in the ensemble average of the final output. Blank table entries indicate the failure rate was zero. (b) A graph showing the same information from the table in (a).

III. Proposed Circuit Designs

Our next step is to map MRF designs onto proof-of-concept silicon devices. In this section, we show

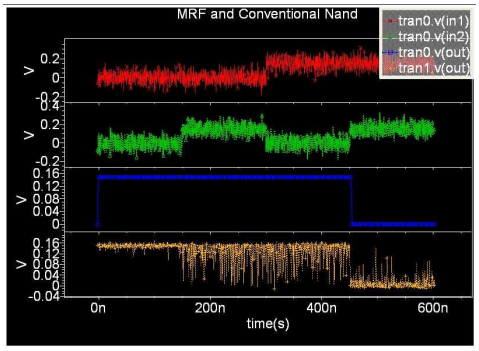
how to design basic logic gates and small circuits. We can then extend these building blocks to make more complicate circuits in the future.

The appropriate noise sources for bio-medical applications are far from being well understood. For this reason, we have chosen Gaussian noise for preliminary investigation. To simulate circuits working in noisy environment, we reduce circuits' supply voltage to the sub-threshold region. The rational behind this design is shown in Fig. 4: *noise becomes significant when the supply voltage drops*. We can utilize this phenomena to simulate circuits working under a noisy environment. We show a standard CMOS NAND gate and a MRF NAND gate operating at sub-threshold supply voltage in Fig. 6 (a). These circuits work under the sub-threshold condition, where the supply voltage is 0.15v while the threshold is 0.22v. The top two waves (read and green) in Fig. 6 (a) show the two noisy inputs of both NAND gates. The MRF output (second to the bottom in blue) is stable, whereas the standard CMOS output (shown at the bottom) switches between correct and incorrect output values. From this simple simulation, we observe that the probabilistic-based design can function well even under very noisy condition at the cost of extra delay and power consumption. Table 1 shows delay comparison between these designs.

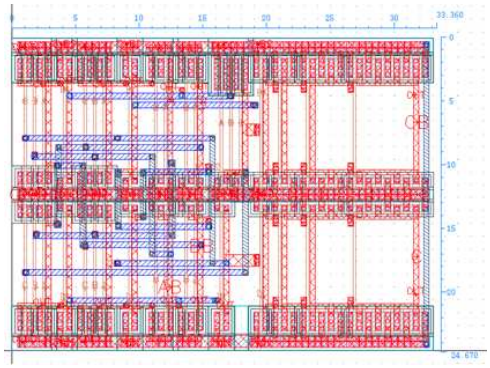
Table 1: Delay of conventional NAND and MRF NAND. Threshold is 0.22v.

Vdd	MRF NAND (in second)	Conventional NAND (in second)
0.15	1.54E-07	2.11E-10
0.2	1.53E-07	3.60E-11
0.25	1.67E-09	1.26E-11
0.3	1.09E-09	7.08E-12

All circuits can be built based on basic logic gates, like inverter and NAND gates. we therefore can make simple circuits on the top of these building blocks. Next, we design a probabilistic-based 8-bit adder to show the circuit's noisy immunity capability. By injecting noise into 8 inputs, as shown in Fig. 7 (a), we observe that the MRF adder, shown in Fig. 7 (c), can still perform normal while the conventional CMOS adder cannot function properly, shown in Fig. 7 (b). The chip layout is shown in Fig. 8 based on 0.18μm CMOS technology. The maximum frequency is 2.5MHz and transistor count is 5040.



(a)



(b)

Figure 6: (a) Standard CMOS NAND and MRF NAND, (b) The layout of MRF NAND.

IV. Conclusion

The error-free computing is not appropriate for bio-medical applications. In this paper, we proposed probabilistic-based circuit design to deal with the uncertainty involved in microarray and high-content imaging data. We proved that the MRF circuits are more robust than conventional the CMOS circuits. Research into more appropriate noise models are carried out over the course of this project.

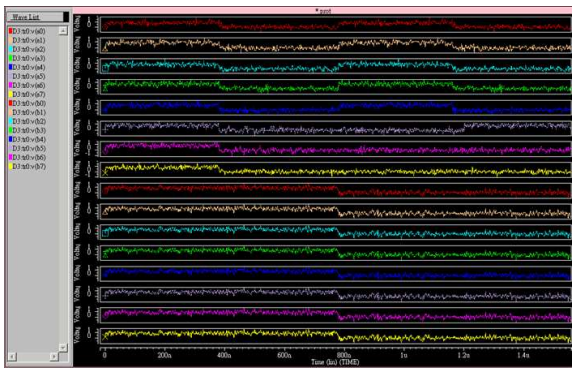
Acknowledgment The author would like to thank Professor Joe Mundy, Andy Wu, and Ruth Bahar for valuable suggestions.

References

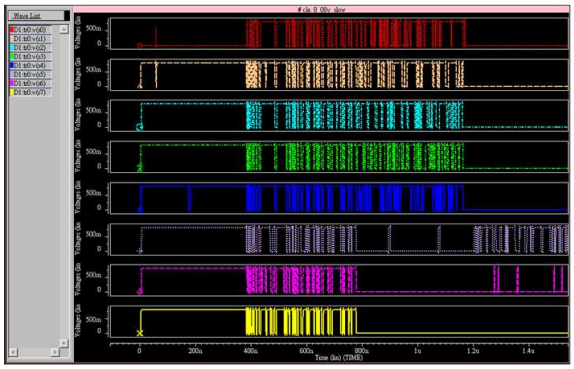
[1] P.O Brown and D. Botstein, "Exploring the new world of the genome with dna microarrays", *Nat Genet*, vol. 21, pp. 33–37, 1999.

[2] Stephen Wong and Jie Chen, "Emerging biomedical technologies at the micro and nano levels", *IEEE Signal Processing Magazine*, vol. 7, 2005.

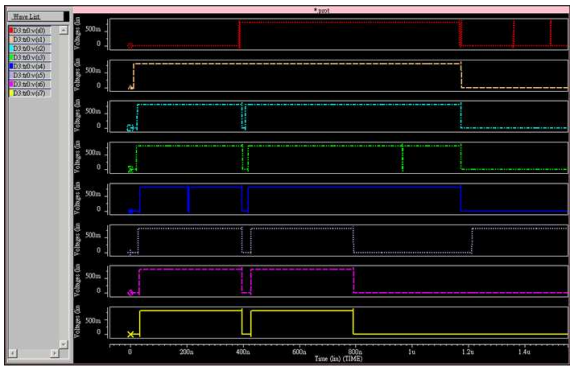
[3] Edward Dougherty, Ilya Shmulevich, Jie Chen, and Jane Wang, *Genomic Signal Processing and Statistics*, European Applied Signal Processing Inc., 2005.



(a)



(b)



(c)

Figure 7: (a) Noise inputs to 8-bit adder, (b) The outputs of conventional 8-bit adder, (c) The outputs of MRF 8-bit adder.

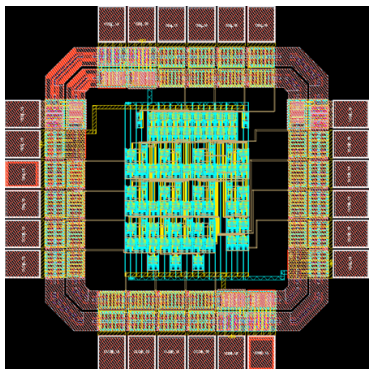


Figure 8: The layout of 8-bit adder (courtesy to Prof. Andy Wu's group).

[4] A.C. Gavin, M. Bosche, R. Krause, and et al., "Functional organization of the yeast proteome by systematic anlaysis of protein complexes", *Nature*, vol. 415(6868), pp. 141–147, 2002.

[5] M.B. Eisen, P.T. Spellman, P.O. Brown, and D. Bostein, "Cluster analysis and display of genome-wide expression patterns", *Proc Natl Acad Sci*, vol. 95(25), pp. 14863–8, 1998.

[6] R.F. Murphy, M.V. Boland, and M. Velliste, "Towards a systematics for protein subcellular location: quantitative description of protein localization patterns and automated analysis of fluorescence microscope images", *Proc Int Conf System Mole Biology*, vol. 8, pp. 251–259, 2000.

[7] J.C. Yarrow, Y. Feng, Z.E. Perlman, and et al., "Phenotypic screening of small molecule libraries by high throughput cell imaging", *Comb Chem High Throughput Screen*, vol. 6(4), pp. 279–286, 2003.

[8] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.

[9] S. Z. Li, *Markov random field modeling in computer vision*, Springer-Verlag, 1995.

[10] S. Geman and K. Kochanek, "Dynamic programming and the graphical representation of error-correcting codes", *IEEE Transactions on Information Theory*, vol. 47, no. 2, 2001.

[11] R. Chellappa and A. Jain, *Markov Random Fields: Theory and Applications*, Academic Press, 1993.

[12] F. Brglez and H. Fujiwara, "A neural netlist of 10 combinational benchmark circuits and a target translator in fortran", in *Proceedings of IEEE Intl. Symp. on Circuits and Systems*, June 1985, pp. 663–698.

[13] J. Chen, J. Mundy, Y. Bai, S.M.C. Chan, P. Petrica, and R. I. Bahar, "A probabilistic approach to nano-computing", in *In Proceedings of the Second Workshop on Non-Silicon Computing, San Diego, CA*, June 2003.

[14] I. R. Bahar, J. Chen, and J. Mundy, *Nano, Quantum, and Molecular Computing: Implications to High Level Design and Validation: A Probabilistic-based design for nanoscale computation*, Kluwer Academic Publishers, 2004.

[15] J. Hammersley and P. Clifford, "Markov fields on finite graphs and lattices", Tech. Rep., University of California, Berkeley, 1968.