Non-equilibrium statistical mechanics of Noisy Computing

Alexander Mozeika[1], David Saad[1] and Jack Raymond[2]

[1]Aston University, UK
[2]Hong Kong University of Science and Technology, China

July 24, 2010

Outline

- **Background:** Noise in Computing, von Neumann's model.
- **Motivation**
- **Random Boolean functions:** Savicky's growth process.
- **Model:** Layered variant of growth process.
- **Method:** Generating functional.
- **Results:** Phase transitions, functions computed, evolution of error.
- **Ongoing work:** Probability of a Boolean function.
- **Summary**

Noise in computing

- Noise is inherent in most forms of computing.
- Classical computing circuits suffer from thermal noise and production errors.
- Impact of noise is more dramatic with increasing circuit complexity and scale.
- Quantum computers suffer from decoherence.
- Neural networks and biological systems are inherently noisy.

von Neumann's model of noisy computation

- von Neumann, 1952: For each gate in circuit assign $0 < \epsilon < 1/2$ which is probab. for the gate to operate incorrectly. Can circuit perform desired operation with error (in the output) $\delta < 1/2$ (reliable)?
- ϵ-**noisy gate** - $\alpha : \{-1,1\}^k \to \{-1,1\}$, but for each $\mathbf{S} \in \{-1,1\}^k$: $\alpha(\mathbf{S}) \to -\alpha(\mathbf{S})$ with prob. ϵ.
- **Circuit/Formula** - directed acyclic graph/rooted tree.
  - Pippenger, 1988: k-ary formulas which tolerate noise: (i) must be deeper (slower) and (ii) compute any Boolean functions with $\delta < 1/2$ only when $\epsilon < (k-1)/2k$.
  - Evans et. al., 1998: NAND-gate formulas: $\epsilon < (3-\sqrt{7})/4$.
  - Evans et. al., 2003: k-ary formulas: $\epsilon < \frac{1}{2} - \frac{2^{k-2}}{k\binom{k-1}{(k-1)/2}}$.

Motivation

- Rigourous results correspond to *worst* cases.
- Our aim is to analyze the *typical* behavior of noisy Boolean random circuits/formulas:
  - noise thresholds for different gates/distributions over gates
  - level of error expected at any depth/layer
  - functions computed
  - sensitivity of functions
  - convergence-rate dependence on input and noise

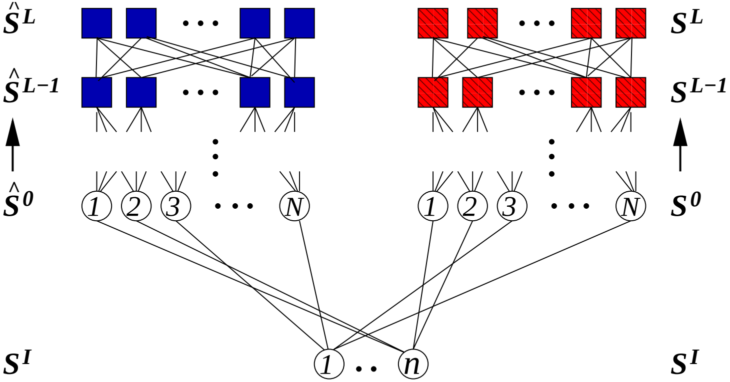
Random Boolean functions

- **Savicky's growth process**:

$$A_0 = \{1, -1, S_1, \ldots, S_n, -S_1, \ldots, -S_n\}$$
$$A_{\ell+1} = \{\alpha(\phi_1, \ldots, \phi_k); \phi_j \in A_\ell \text{ for } j = 1, 2, \ldots, k\}$$

- Savicky, 1990: For all *balanced* and *nonlinear* α (MAJ-k,etc.) and for all f: $\lim_{\ell \to \infty} P_\ell(f) = \left(\frac{1}{2}\right)^{2^n}$.
- Brodsky et. al., 2005: Depending on A_0 and α process converges to uniform distribution on some set of functions or to single Boolean function.

Model



- We use *layered* variant of Savicky's growth process (feed-forward $N \times (L+1)$ circuit).
- $S^I = \{-1, 1, S_1, \ldots, S_n, -S_1, \ldots, -S_n\}$, $|S^I| \in O(N^0)$.

Probabilistic framework

- The i-th gate in the ℓ-th layer operates according to:

$$S_i^\ell = \eta_i^\ell \alpha_i^\ell (S_{i_1}^{\ell-1}, \ldots, S_{i_k}^{\ell-1}),$$

where $P(\eta_i^\ell = -1) = \epsilon$, which leads to the microscopic law

$$P(S_i^\ell | S_{i_1}^{\ell-1}, \ldots, S_{i_k}^{\ell-1}) = \frac{\mathrm{e}^{\beta S_i^\ell \alpha_i^\ell (S_{i_1}^{\ell-1}, \ldots, S_{i_k}^{\ell-1})}}{2\cosh[\beta \alpha_i^\ell (S_{i_1}^{\ell-1}, \ldots, S_{i_k}^{\ell-1})]},$$

where $\tanh\beta = 1 - 2\epsilon$.

- **Joint probability of microscopic states** in two systems:

$$P[\{\mathbf{S}^\ell\}; \{\hat{\mathbf{S}}^\ell\}] = P(\mathbf{S}^0, \hat{\mathbf{S}}^0 | S^I) \prod_{\ell=1}^{L} P(\mathbf{S}^\ell | \mathbf{S}^{\ell-1}) P(\hat{\mathbf{S}}^\ell | \hat{\mathbf{S}}^{\ell-1}),$$

where $P(\mathbf{S}^\ell | \mathbf{S}^{\ell-1}) = \prod_{i=1}^N P(S_i^\ell | S_{i_1}^{\ell-1}, \ldots, S_{i_k}^{\ell-1})$.

Generating functional analysis I

- **Generating function**:

$$Z[\psi;\hat{\psi}] = \left\langle \mathrm{e}^{-\mathrm{i}\sum_{\ell,i}\{\psi_i^\ell S_i^\ell + \hat{\psi}_i^\ell \hat{S}_i^\ell\}} \right\rangle,$$

  where $\langle \ldots \rangle$ denotes average w.r.t. $P[\{\mathbf{S}^\ell\};\{\hat{\mathbf{S}}^\ell\}]$.

- **Moments**: $\langle S_i^\ell \hat{S}_j^{\ell'} \rangle = -\lim_{\boldsymbol{\psi},\hat{\boldsymbol{\psi}}\to\mathbf{0}} \frac{\partial^2}{\partial_{\psi_i^\ell}\partial_{\hat{\psi}_j^{\ell'}}} Z[\psi;\hat{\psi}]$.

- For $N \to \infty$ the system is self-averaging: $\overline{Z[\ldots]} = \int\{\mathrm{d}\mathbf{P}\mathrm{d}\hat{\mathbf{P}}\}\mathrm{e}^{N\Psi[\mathbf{P},\hat{\mathbf{P}}]}$, where $\overline{\cdots}$ average over disorder.

- **Order parameter**: $P^\ell(S,\hat{S}) = \frac{1}{N}\sum_{i=1}^N \overline{\langle \delta_{S_i^\ell;S}\delta_{\hat{S}_i^\ell;\hat{S}}\rangle|_{S^I}}$.

- **Observables**: $m(\ell) = \frac{1}{N}\sum_{i=1}^N \overline{\langle S_i^\ell \rangle}$, $\hat{m}(\ell)$ and $C(\ell) = \frac{1}{N}\sum_{i=1}^N \overline{\langle S_i^\ell \hat{S}_i^\ell \rangle}$.

Generating functional analysis II

- Evolution from layer to layer:

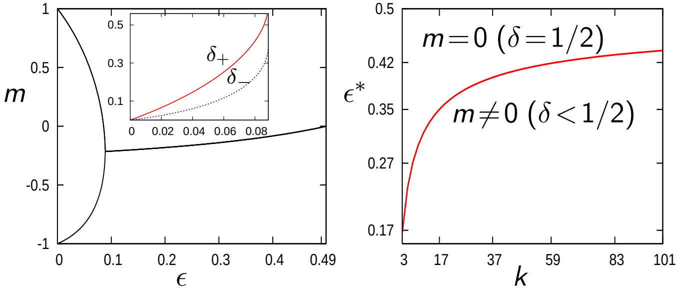
$$m(\ell+1) \;=\; (1-2\epsilon)\sum_{\{S_j\}}\prod_{j=1}^{k}\left[\frac{1+S_j m(\ell)}{2}\right]\langle\alpha(S_1,..,S_k)\rangle_\alpha$$

$$C(\ell+1) \;=\; (1-2\epsilon)\sum_{\{S_j,\hat{S}_j\}}\prod_{j=1}^{k}\left[\frac{1+S_j m(\ell)+\hat{S}_j\hat{m}(\ell)+S_j\hat{S}_j C(\ell)}{4}\right] \times\left\langle\alpha(S_1,..,S_k)\alpha(\hat{S}_1,..,\hat{S}_k)\right\rangle_\alpha,$$

where $m(0)=\hat{m}(0)=\frac{1}{|S'|}\sum_{S\in S'} S,\;\; C(0)=1.$

- **Sensitivity**: $\Delta(\ell)=\lim_{\epsilon\to 0}\frac{1}{2}\left[1-\frac{1}{N}\sum_{i=1}^{N}\overline{\langle S_i^\ell\hat{S}_i^\ell\rangle}\right].$
- **Error**: $\delta(\ell)=\max_{S'}\frac{1}{2}\left[1-\frac{1}{N}\sum_{i=1}^{N}\overline{\langle S_i^\ell\hat{S}_i^\ell\rangle|_{S'}}\right].$

Phase transitions



- MAJ-k gate: $\epsilon^* = 1/2 - 2^{k-2}/k\binom{k-1}{(k-1)/2}$.
- NAND gate: $\epsilon^* = (3-\sqrt{7})/4$.
- $\epsilon^* > \epsilon$ information about S^I (input) preserved for arbitrarily many layers.
- Complicated computational task = significant number of layers $\Rightarrow$ only relatively simple operations when $\epsilon > \epsilon^*$.

Boolean functions computed

When $\epsilon = 0$ typical formula on layer $\ell \to \infty$ computes:

- in MAJ-k circuit

$$F = \left\{ \begin{array}{lll} +1 & & \text{if } m(0) > 0 \\ \pm 1 & \text{with prob. } {}^{1}/_{2} & \text{if } m(0) = 0 \\ -1 & & \text{if } m(0) < 0 \end{array} \right\}$$

where $m(0) = \frac{1}{|S^l|} \sum_{S \in S^l}$.

- in NAND circuit

$$F = \left\{ \begin{array}{ll} -1 & \text{if } m(0) < 2-\sqrt{5} \\ +1 & \text{if } m(0) > 2-\sqrt{5} \end{array} \right.$$

Sensitivity

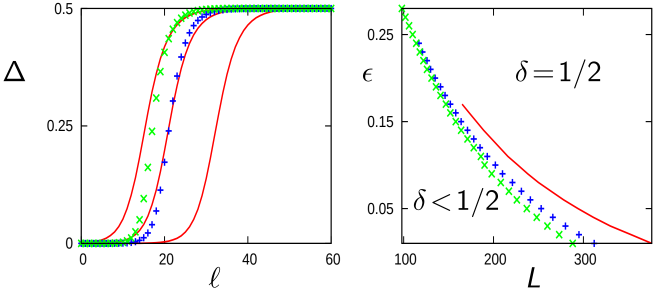


Figure: Left: Hamming distance for $k = 3$ (solid line), with input mismatch $\Delta(0) = 10^{-3}, 10^{-4}, 10^{-6}$ (left to right), and for $k = 5$ ($+$), $k = 7$ ($\times$) with $\Delta(0) = 10^{-6}$. Right: Reliability boundaries.

- For $\epsilon = 0$ and $m(0) = 0$ the MAJ-k circuit is sensitive.
- For $\epsilon > 0$ circuit amplifies noise and $\delta(L)$ grows but remains limited for small ϵ.

Convergence rates and evolution of error

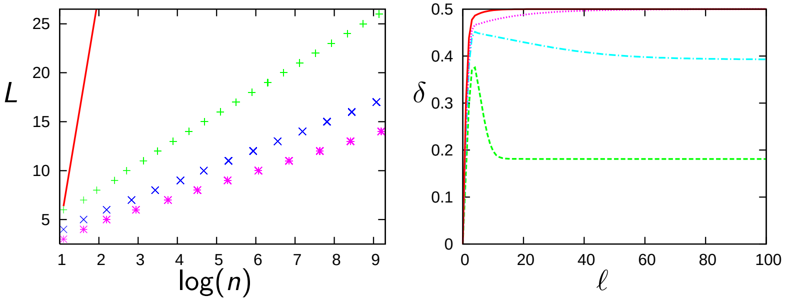


Figure: Left: Number of layers in the noiseless MAJ-k based circuit computing MAJ-n function. Theoretic data-points are represented by symbols $+$ $(k=3)$, $\times$ $(k=5)$ and $*$ $(k=7)$. The straight line corresponds to the bound $k2^k \log(n)$, plotted here for $k=3$ only. Right: Evolution of error in MAJ-5 circuit.

- For $\epsilon = \epsilon^*(k) \pm \Delta\epsilon$ and k large: $|m(\ell) - m(\infty)| \approx \mathrm{e}^{-\mathrm{const}\sqrt{k}\Delta\epsilon\ell}$.
- The error can not be reduced when $\epsilon > \epsilon(0)$.

Ongoing work

- Consider 2^n copies of noisy system with the same random *layered* toplolgy but for *different* inputs.
- Probability of a Boolean function on layer $\ell + 1$

$$P^{\ell+1}(\mathbf{S}) = \sum_{\{\mathbf{S}_j\}} \prod_{j=1}^{k} \left[P^{\ell}(\mathbf{S}_j) \right] \left\langle \prod_{\sigma_1,..,\sigma_n} \frac{\mathrm{e}^{\beta S(\sigma)\alpha(S_1(\sigma),..,S_k(\sigma))}}{2\cosh \beta\alpha(S_1(\sigma),..,S_k(\sigma))} \right\rangle_{\alpha}$$

  where $\mathbf{S} \in \{-1,1\}^{2^n}$ and $\sigma = (\sigma_1,..,\sigma_n) \in \{-1,1\}^n$.
- For $\beta \to \infty$ and single gate α we recover equation used to study Savicky's growth process.
- Furthermore, the same equation is valid for random *recurrent* topology.

Evolution of Entropy

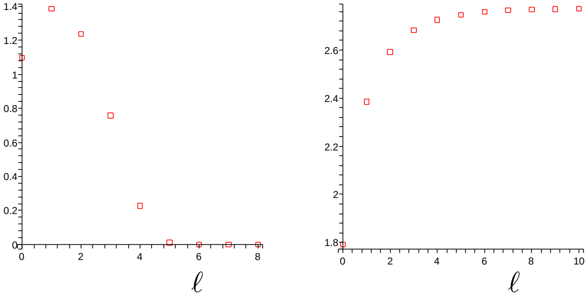


Figure: Evolution of entropy for $\beta \to \infty$ in MAJ-3 circuit. Left: $S^I = \{1, S_1, S_2\}$. Right: $S^I = \{-1, 1, S_1, S_2, -S_1, -S_2\}$

- Shannon entropy: $H^\ell = -\sum_{\mathbf{S}} P^\ell(\mathbf{S}) \log P^\ell(\mathbf{S})$

Summary

- The models of random formulae can be mapped onto a physical framework which allows one to analyze the typical behavior of noisy Boolean random formulae (balanced).
- The typical-case macroscopic behavior observed corresponds straightforwardly to the bounds obtained in the theoretical computer science.
- The framework enables one to explore further properties: the level of error and function-bias expected at any depth, the sensitivity to input perturbations, the convergence rate depending on the input bias.
- We also studied the effects of hard and threshold noises.
- *Phys. Rev. Lett.* 2009; **arXiv:0908.3981v1**.