Martin Ziegler

Syllabus

§8 Randomization

- Motivation: Reliability
- Sources of Randomness
- Las Vegas vs. Monte Carlo
- Primality Testing
- Errors and Amplification
- Blackbox Polynomial Test
- Schwartz-Zippel Lemma

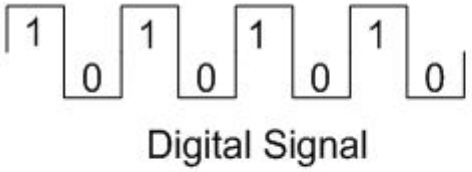
Motivation: *RELIABILITY* ?

$\approx 10^9$ gates in a PC, flipping $\approx 10^9$ times per second

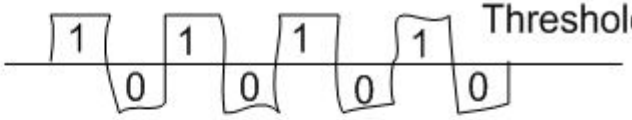
[Ziegler&Lanford'79] *Effect of Cosmic Rays on Computer Memories*

Table 1: Memory errors per year:

Platf.	Tech.	Per machine				
		CE Incid. (%)	CE Rate Mean	CE Rate C.V.	CE Median Affct.	UE Incid. (%)
A	DDR1	45.4	19,509	3.5	611	0.17
B	DDR1	46.2	23,243	3.4	366	–
C	DDR1	22.3	27,500	17.7	100	2.15
D	DDR2	12.3	20,501	19.0	63	1.21
E	FBD	–	–	–	–	0.27
F	DDR2	26.9	48,621	16.1	25	4.15
Overall	–	32.2	22,696	14.0	277	1.29



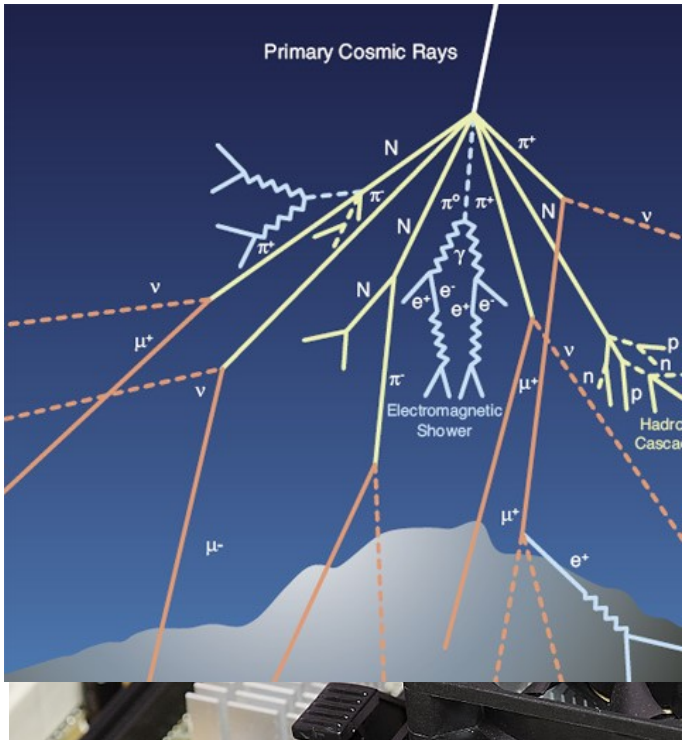
Digital Signal



Digital Signal Degradation

Analog Signal

Analog Signal Degradation



1. Efficiency
2. Elegance
3. Reliability: error probability $\leq 2^{-100}$

Sources of Randomness

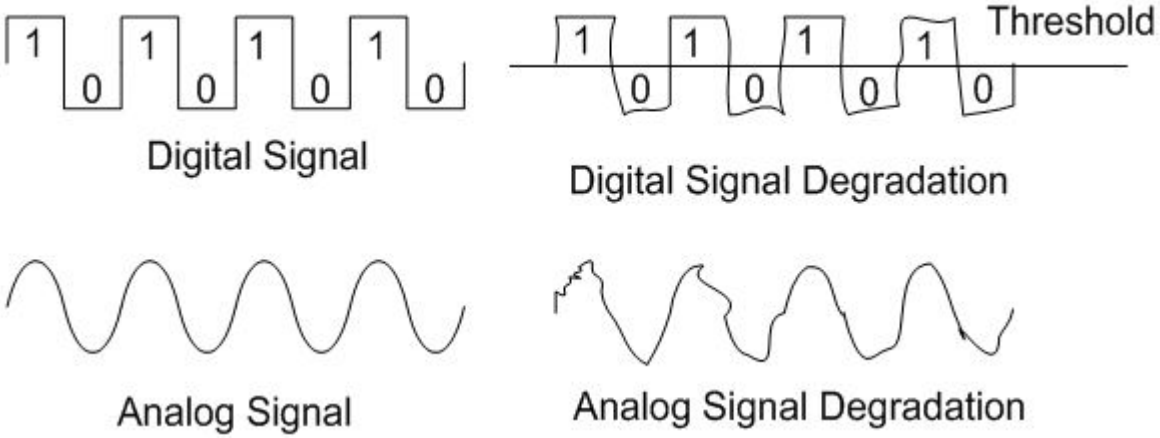
Entropy from: heat, user, quantum mechanics

- low rate, correlation, bias; (Martin-Löf) Tests

Pseudo-random sequence: deterministic!

Table 1: Memory errors per year:

Platf.	Tech.	Per machine				
		CE Incid. (%)	CE Rate Mean	CE Rate C.V.	CE Median Affct.	UE Incid. (%)
A	DDR1	45.4	19,509	3.5	611	0.17
B	DDR1	46.2	23,243	3.4	366	–
C	DDR1	22.3	27,500	17.7	100	2.15
D	DDR2	12.3	20,501	19.0	63	1.21
E	FBD	–	–	–	–	0.27
F	DDR2	26.9	48,621	16.1	25	4.15
Overall	–	32.2	22,696	14.0	277	1.29



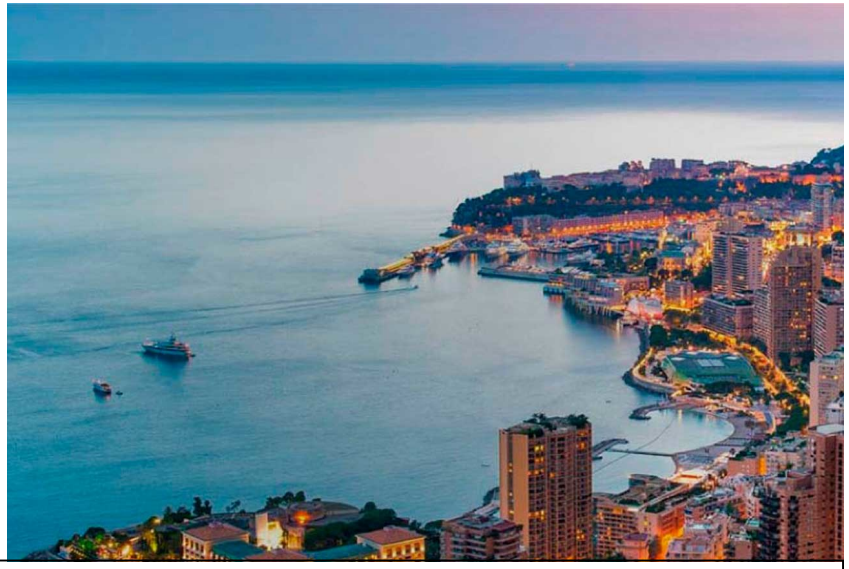
1. Efficiency
2. Elegance
3. Reliability: error probability $\leq 2^{-100}$

Las Vegas vs. Monte Carlo

- Result always correct
- *Expected* runtime

QuickSort?

- Result "*probably*" correct
- Guaranteed runtime



Recap: (Conditional) probability, random variable, (stochastic) independence, expected value, Bayes

Primality Testing

Decision problem $L \subseteq \mathbb{N}$: Input $x \in \mathbb{N}$, output **yes/no**

$x \in L$ $x \notin L$

For $L=\{2,3,5,7,11,\ldots\}$: test **isprime**(x)

$n := \log(x)$

Naïve idea: *Guess* and verify a proper factor of x.

Case x = product of two primes $\Rightarrow$ detect probability $\approx O(\sqrt{x})$ ↯

Correct-on-average algorithm: Blindly say "*composite*" ☺

Miller-Rabin Test: *Guess* $a<x$. Write $x-1=d\cdot 2^e$ with d odd.

If (*) holds, then say "x is *composite*".

(*) $a^d \not\equiv 1 \pmod{x}$ and $a^{d\cdot 2^r} \not\equiv -1 \pmod{x}$ for all $r<e$,

Theorem: a) If (*) holds, then x is composite.

b) If x is composite, then ≥¼ of all $a<x$ satisfy (*).

amplify!

Deterministic test [Agrawal/Kayal/Saxena'02]: time $O(n^7)$.

Errors and Amplification

Decision problem $L \subseteq \mathbb{N}$: Input $x \in \mathbb{N}$, output **yes**/**no**
$x \in L$ $x \notin L$

Algorithm $\mathcal{A}$ with one-sided error:

- false positives *only*: **yes** but $x \notin L$
- false negatives *only*: **no** but $x \in L$

Error probability $p^k \ll 1$

$\mathcal{A}'$: Repeat $\mathcal{A}$ k-times, report **no** (only) if *all* return **no**.

Example: $p=0.99$, $k=70 \cdot 100$

Algorithm $\mathcal{B}$ with two-sided error:

both false positive and false negative

Error probability $p<½$

$\mathcal{B}'$: Repeat $\mathcal{B}$ k-times and report the majority answer.

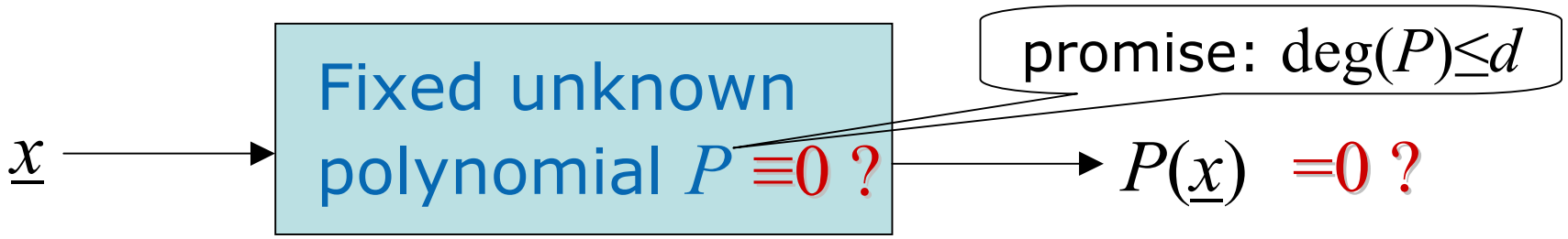
Example: $p=1/3$, $k=100$

Fact (Hoeffding): Let $X_1,..X_k$ independent random variables in $[0;1]$, $\underline{X} := (X_1+\ldots+X_k)/k$ and $t>0$.

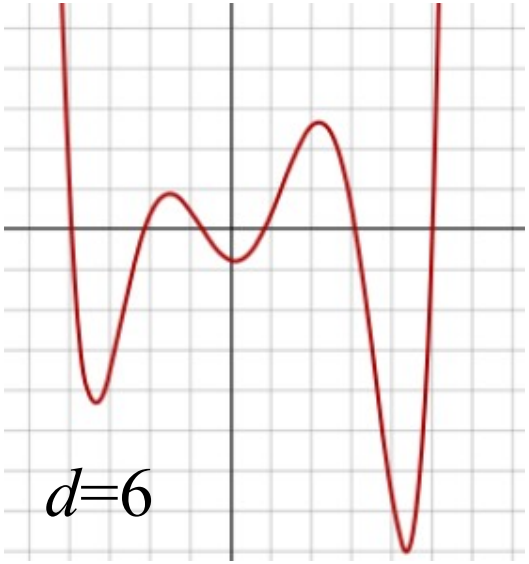
Then $\mathbb{P}\left[\ \underline{X} \geq \mathbb{E}[\underline{X}] + t\ \right] \leq e^{-2kt^2}$

X_n := n-th execution errs, $t := ½-p$

Blackbox Polynomial Test



Recap: A non-zero *uni*variate polynomial of $\deg \le d$ has at most d roots. *Total degree* (e.g. of $X^2 \cdot Y^3$ is 5).



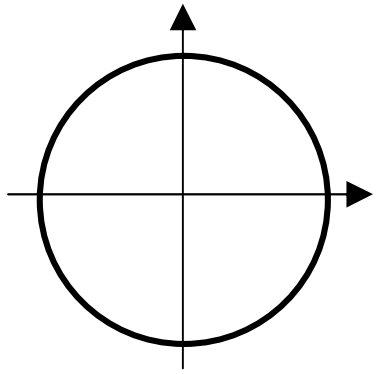
Generic blackbox polynomial test P:

Fix finite set D.

Sample $\underline{x} \in D$ uniformly at random.

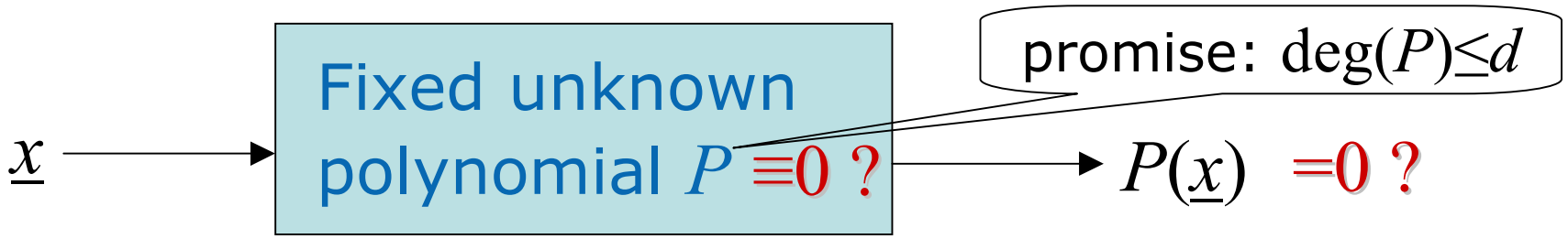
If $P(x)=0$ say "$P \equiv 0$".

*multi*variate?



X^2+Y^2-1

Schwartz-Zippel Lemma



[*Schwartz'80, Zippel'79, deMillo&Lipton'78*)**:** Fix finite $S \subseteq \mathbb{C}^n$.

Suppose $0 \not\equiv P \in \mathbb{C}[X_1,\ldots X_n]$ has total degree $\leq d$.

Sample $x_1,\ldots x_n$ from S independently uniformly at random.

Then $\mathbb{P}[\, P(x_1,\ldots x_n)=0\,] \leq d/|S|$. $|S|:=2d$, *then* amplify!

Proof: Write $0 \neq P(X_1,\ldots X_n) = \sum_{0\leq j\leq d} P_j(X_1,\ldots X_{n-1}) \cdot X_n^j$.

Let j be maximal s.t. $P_j \not\equiv 0$. Then $\mathbb{P}[\, P(x_1,\ldots x_n)=0\,] \leq$

$\leq \mathbb{P}[\, P_j(x_1,\ldots x_{n-1})=0] + \mathbb{P}[\, P(x_1,\ldots x_n)=0 \mid P_j(x_1,\ldots x_{n-1})\neq 0]$

$\leq (d\text{-}j)/|S|$ $\leq j/|S|$

$$\mathbb{P}[A] = \mathbb{P}[A\wedge B] + \mathbb{P}[A\wedge\neg B] \leq \mathbb{P}[B] + \mathbb{P}[A|\neg B] \cdot \cancel{\mathbb{P}[\neg B]}$$

§8 Summary

- Motivation: Reliability
- Sources of Randomness
- Las Vegas vs. Monte Carlo
- Primality Testing
- Errors and Amplification
- Blackbox Polynomial Test
- Schwartz-Zippel Lemma