Distribution-Aware Sampling and Weighted Model Counting for SAT

Supratik Chakraborty[1], Daniel J. Fremont[2], **Kuldeep S Meel[3]**,
Sanjit A. Seshia[2], Moshe Y Vardi[3]

[1]Indian Institute of Technology Bombay, India
[2]University of California, Berkeley
[3]Department of Computer Science, Rice University

July 31, 2014 AAAI 2014

Weighted Model Counting

Given:

- CNF Formula F, Solution Space: R_F
- Weight Function W(.) over assignments
  - $W(\sigma)$

Problem (WMC):

What is the sum of weights of satisfying assignments i.e. $W(R_F)$?

Example

$F = (a \vee b)$; $R_F = \{[0,1], [1,0], [1,1]\}$

$W([0,1]) = W([1,0]) = 1/3$ $W([1,1]) = W([0,0]) = 1/6$

$W(R_F) = 5/6$

Distribution-Aware Sampling

Given:

- CNF Formula F, Solution Space: R_F
- Weight Function W(.) over assignments
  - W(σ)

Problem (Sampling):

Pr (Sampling a solution y) = $W(y)/W(R_F)$

Example:

F = (a V b); R_F = {[0,1], [1,0], [1,1]}

W([0,1]) = W([1,0]) = 1/3 W([1,1]) = W([0,0]) = 1/6

Pr ([0,1] is generated] = (1/3) / (5/6) = 2/5

Exciting Applications

- Probabilistic Inference (Reduced to weighted model counting - Roth 1996)
- Probabilistic programming
- Constraint random verification (sampling)

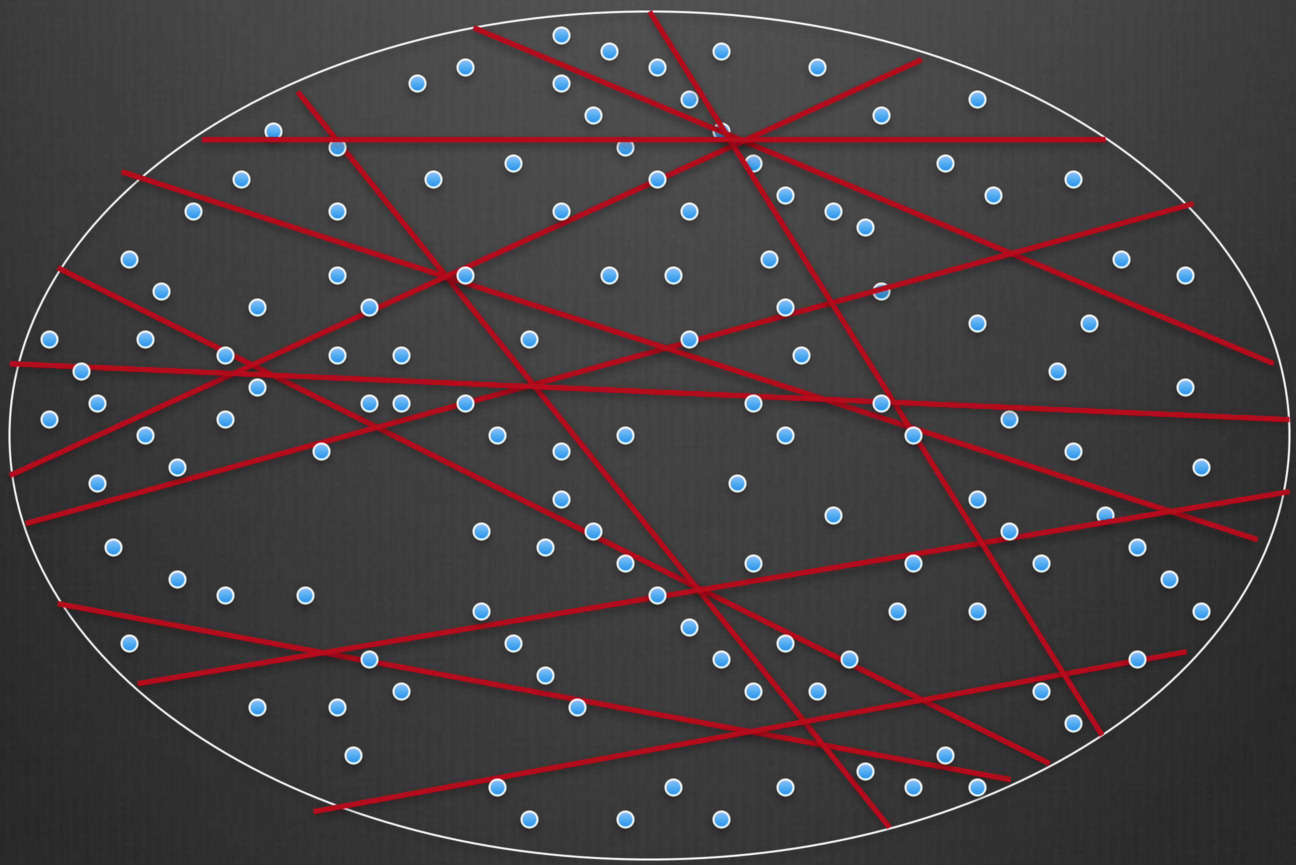
Prior Work

- Exact Methods (Cachet, SDD)
  - Poor Scaling
- Guarantee-less Techniques (MCMC)
  - No Guarantees
- Approximate methods with Guarantees
  - Requires MPE oracle

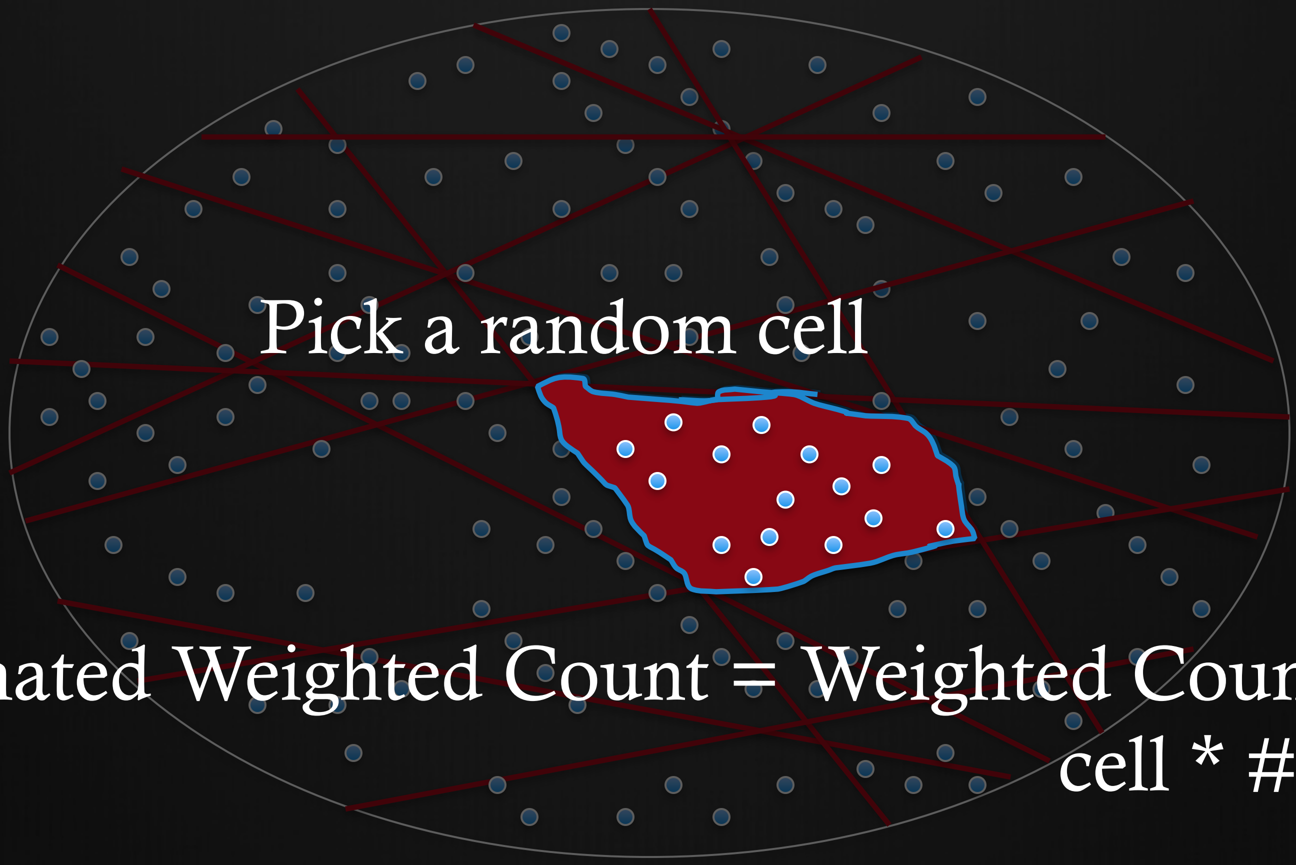
Main Contributions

- Novel parameter tilt (ρ) to characterize complexity
  - $\rho = W_{max} / W_{min}$ over satisfying assignments
- Small Tilt (ρ)
  - Efficient hashing-based technique requires only SAT oracle (no need for MPE oracle)
- Large Tilt (ρ)
  - Framework with access to PB solver

Partitioning into equal “small” cells



Partitioning into equal "small" cells



Estimated Weighted Count = Weighted Count of the cell * # of cells

How to Partition?

How to partition into roughly equal (weighted) small cells of solutions without knowing the distribution of solutions?

3-Universal Hashing [Carter-Wegman 1979, Sipser 1983]

Strong Theoretical Guarantees

- **Weighted Counting:**

$$Pr[\frac{1}{(1+\varepsilon)w(R_F)} \leq C \leq \frac{1+\varepsilon}{w(R_F)}] \geq 1-\delta$$

- **Sampling:**

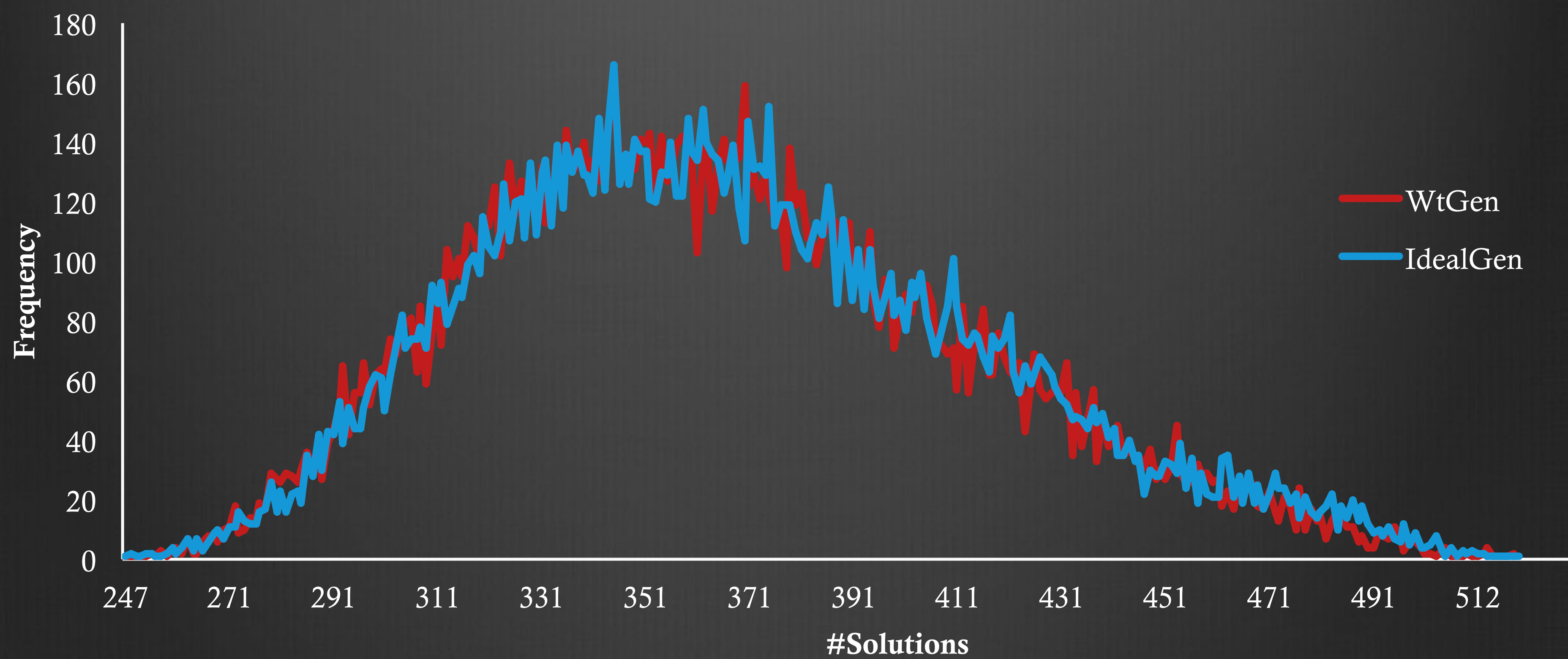
$$\frac{w(y)}{(1+\varepsilon)w(R_F)} \leq \Pr[\text{y is Sampled}] \leq (1+\varepsilon)\frac{w(y)}{w(R_F)}.$$

- **Complexity:** # of calls to SAT solver is linear in ρ

Experimental Comparison

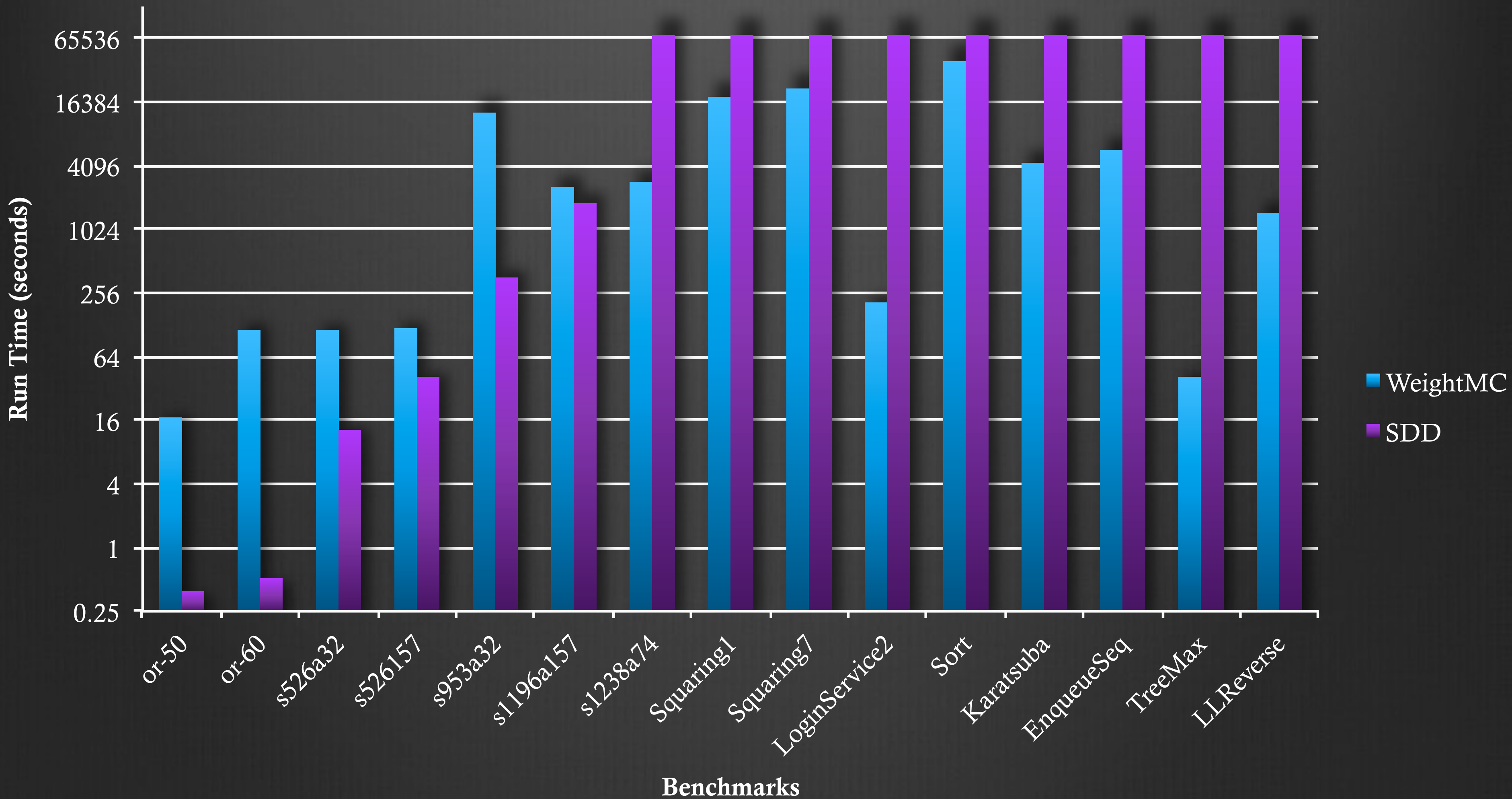
- Benchmarks
  - Grid networks, Plan recognition, ISCAS89, Bounded model checking
- WeightMC: $\rho = 3$, $\varepsilon = 0.8$, $\delta = 0.2$
- WeightGen: $\rho = 3$, $\kappa = 16$
- Objectives:
  - Distribution quality v/s Ideal Sampler
  - Runtime performance v/s SDD

Sampling Distribution



- Benchmark: case110.cnf; #var: 287; #clauses: 1263
- Total Runs**: $4x10^6$;** Total Solutions **: 16384**

Significantly Faster than SDD



Takeaways

- Distribution-Aware sampling and weighted model counting are important problems
- A novel parameter to characterize complexity
- Efficient scheme for problems with low tilt
- Significantly faster and practically close to the real distribution in practice