Accelerated Adaptive Markov Chain for Partition Function Computation

Stefano Ermon*, Carla Gomes*, Ashish Sabharwal[+], and Bart Selman*

*Department of Computer Science, Cornell University

[+]IBM Watson Research Center

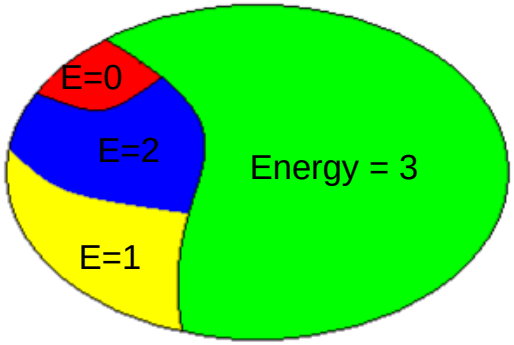
Partition Function Computation

- Normalization constant in factored probabilistic models (e.g., MRFs, MLNs with soft probabilistic constraints)

  Z = Sum over exponentially many configurations

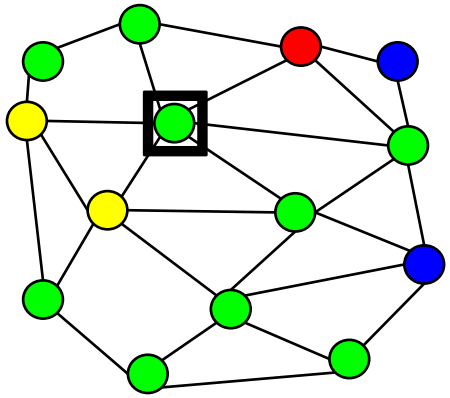
  ⟹ *Hard to compute. Can we approximate it?*

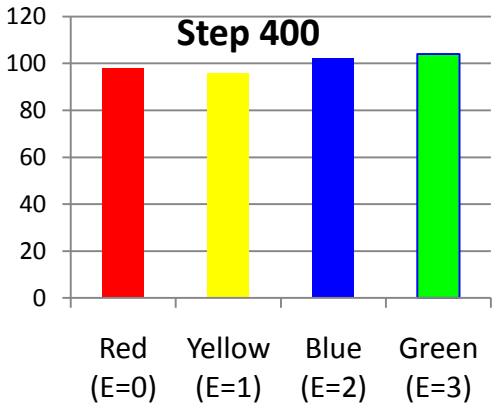
- *Flat Histogram* method (Wang-Landau)



Partition of the set of all possible configurations (according to energy)

Adaptive MCMC will eventually visit all subsets (= colors = energy levels) equally often (Contrast: Metropolis/SA, according to Boltzmann weight)

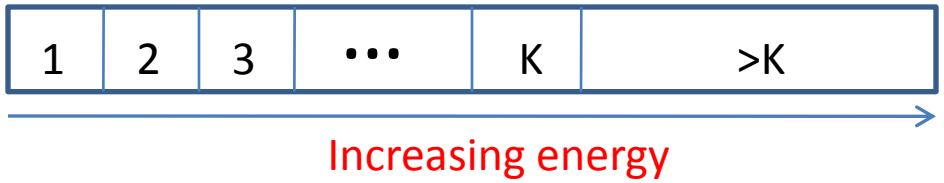




Estimates the size of the subsets (*density of states*), which also gives the partition function Z

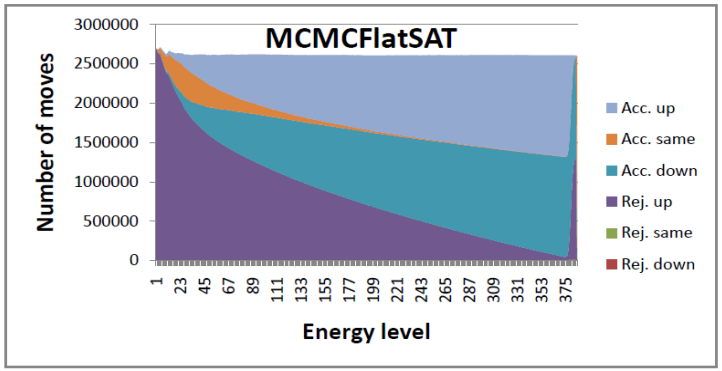
Our Contributions

1a. **Energy Saturation**: single bucket for high energy states

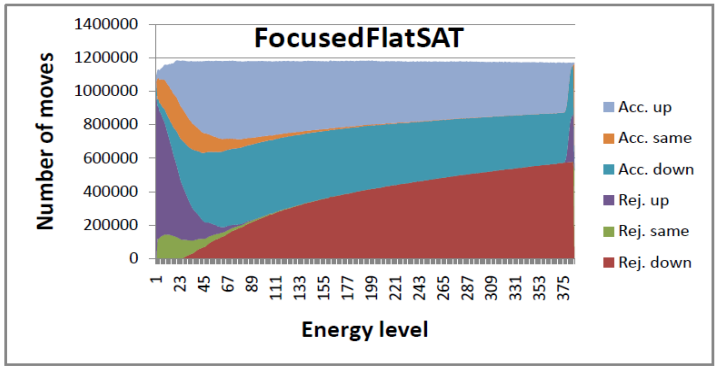


- Fewer buckets ⟹ faster
- (Tight) Upper bound on Z

1b. **Focused moves**: variables occurring in violated constraints are flipped more frequently (preserving detailed balance)



Greatly reduced number of null moves (purple) ⟹



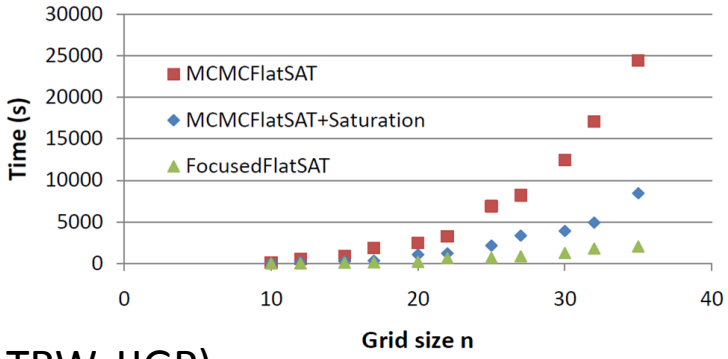
2. **New application**: density of states gives *parameterized partition function*

e.g. at all temperatures, all weights of the soft constraints ⟹ **learning**

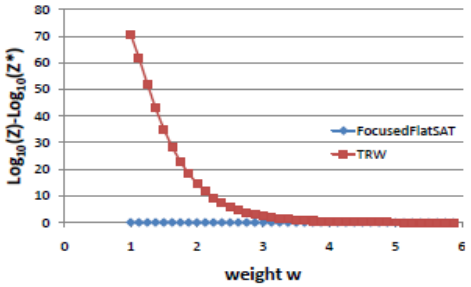
Experimental Results

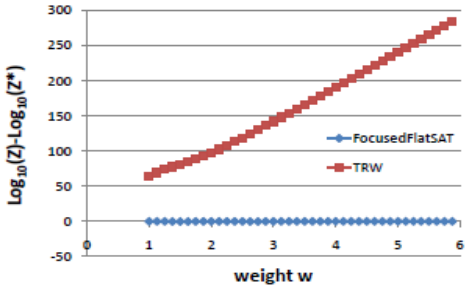
- **Improved Scaling**

  Focused Moves + Saturation
  outperforms standard flat histogram



- **Better Accuracy** (vs. Gibbs Sampling, TRW, IJGP)





- Hard Constraints (model counting)
- Soft Constraints
- Hard & Soft constraints

- **Improved Weight Learning**

  Close to optimal likelihoods for the trained weights in synthetic Markov Logic Networks