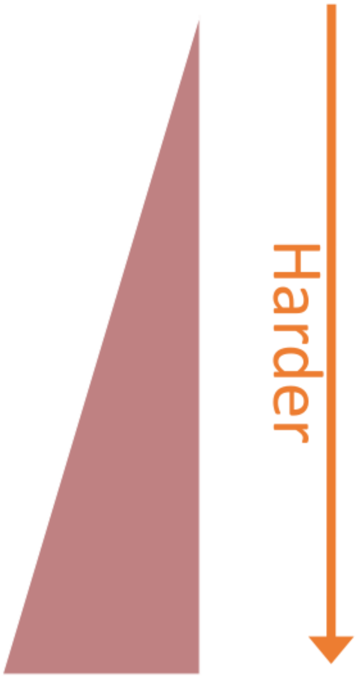
Stochastic Anytime Search for Bounding Marginal MAP

Radu Marinescu
IBM Research

Rina Dechter and **Alexander Ihler**
University of California, Irvine

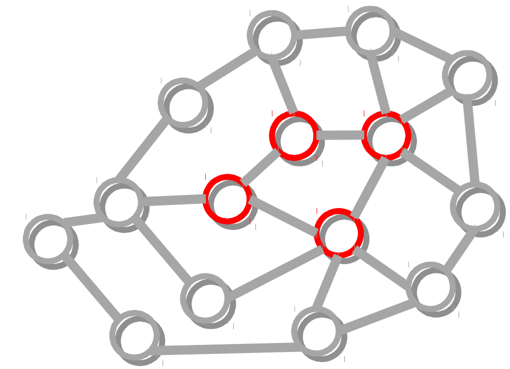
Types of queries

Max-Inference	$f(\mathbf{x}^*) = \max_{\mathbf{x}} \prod_{\alpha} f_\alpha(\mathbf{x}_\alpha)$
Sum-Inference	$Z = \sum_{\mathbf{x}} \prod_{\alpha} f_\alpha(\mathbf{x}_\alpha)$
Mixed-Inference	$f(\mathbf{x}_M^*) = \max_{\mathbf{x}_\mathrm{M}} \sum_{\mathbf{x}_S} \prod_{\alpha} f_\alpha(\mathbf{x}_\alpha)$



- **NP-hard**: exponentially many terms

Why Marginal MAP?

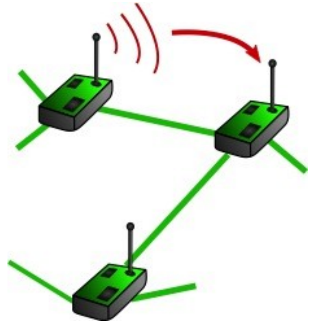


- Often, Marginal MAP is the right task
  - We have a model describing a large system
  - We care about predicting the state of some part

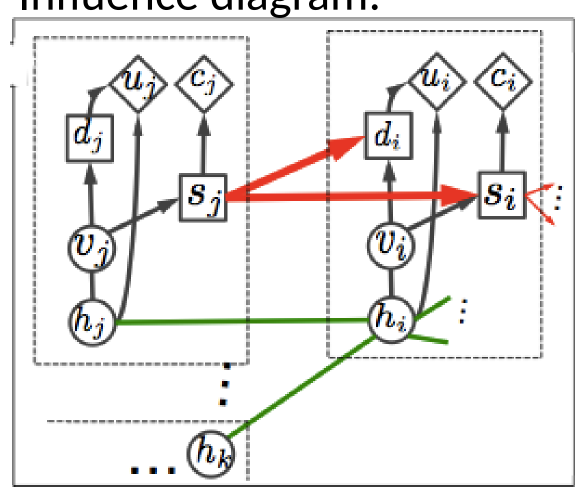
- Complexity: NP^{PP} complete
- Not necessarily easy on trees

- Example: decision making
  - Sum over random variables (random effects, etc.)
  - Max over decision variables (specify action policies)

Sensor network



Influence diagram:



Motivation and Contribution

- Marginal MAP Inference
  - Probabilistic inference query
    - Optimal partial configuration after marginalizing hidden/latent variables in a probability distribution
  - Complexity: NP^{pp} complete
  - Often it is the right task on various applications
    - Probabilistic conformant planning [Lee, Marinescu, Dechter, 2015]
    - Natural language processing task [Bird, Klein, Loper, 2009]
    - Image completion task [Xue, Li, Ermon, Gomes, Selman, 2016]
  - Conditioned summation subproblem is often intractable

- **Contributions**
  - **Stochastic anytime search algorithms for bounding MMAP**
  - **Probabilistic bounds (lower/upper) on the optimal MMAP value**
  - **Applicable to MMAP with intractable conditioned summation**

Background: graphical models

- Graphical model $\mathcal{M} = \langle \mathbf{X}, \mathbf{D}, \mathbf{F} \rangle$
  - variables $\mathbf{X} = \{X_1, \ldots, X_n\}$
  - domains $\mathbf{D} = \{D_1, \ldots, D_n\}$
  - functions $\mathbf{F} = \{\psi_1, \ldots, \psi_r\}$

$$P(\mathbf{x}) = \frac{1}{Z} \prod_{\alpha \in F} \psi_\alpha(\mathbf{x}_\alpha)$$

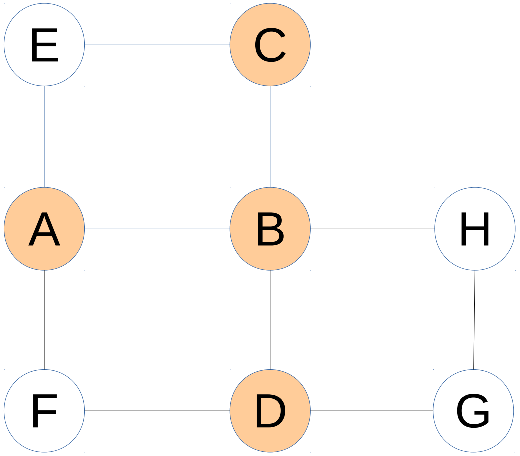
- Marginal MAP task

$$\mathrm{x}_M^* = \mathrm{argmax}_{\mathbf{x}_M} \sum_{\mathbf{x}_s} \prod_\alpha \psi_\alpha(\mathbf{x}_\alpha)$$

  - $\mathbf{x} = \mathbf{x}_M \cup \mathbf{x}_S$
  - Max and sum not commute

NP^PP-complete

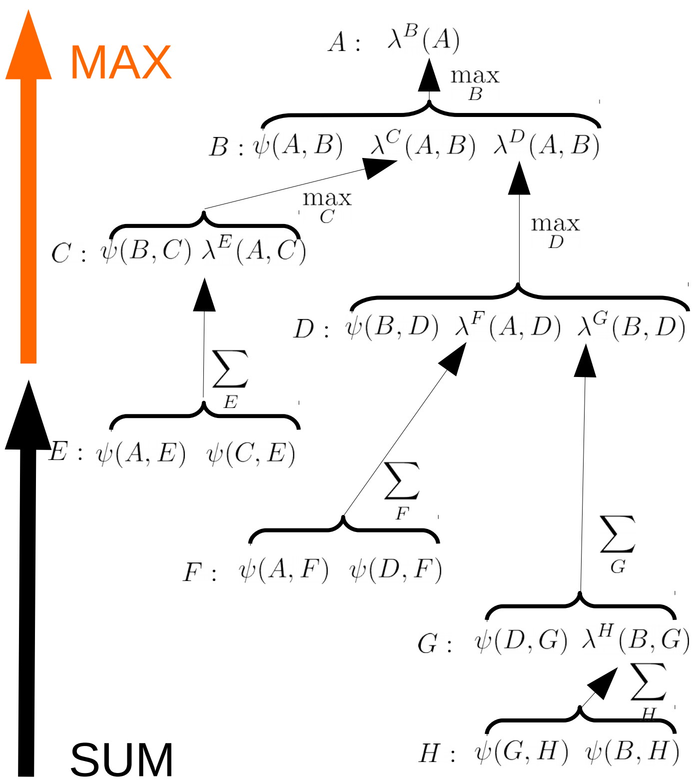
- Primal graph
  - nodes are variables
  - two nodes are connected if they appear in the same function



$$\mathbf{X}_M = \{A, B, C, D\}$$
$$\mathbf{X}_S = \{E, F, G, H\}$$

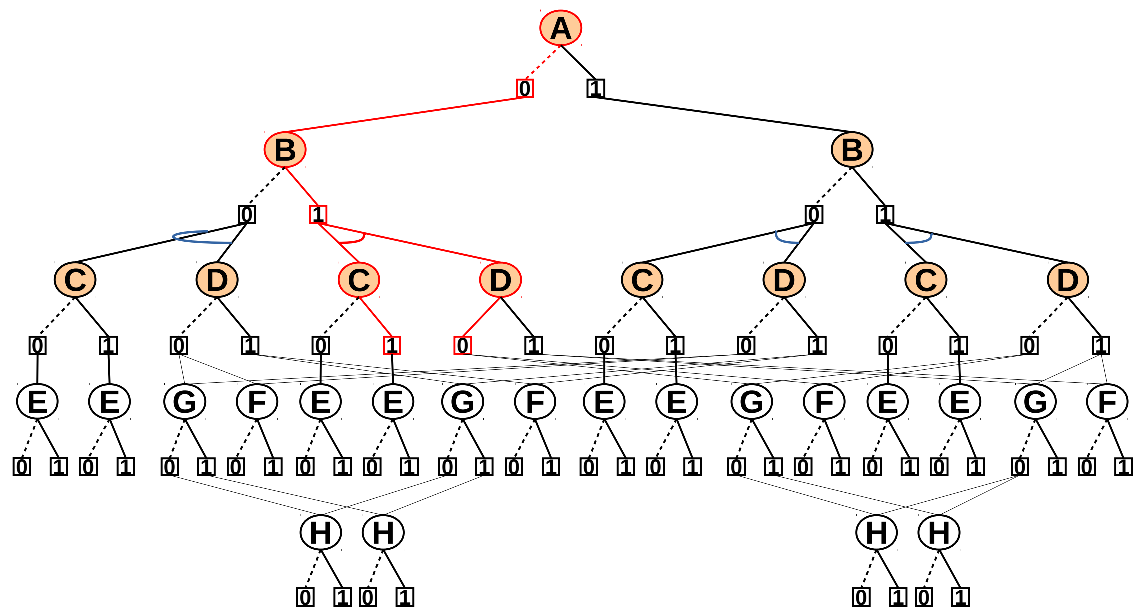
Background: AND/OR search spaces

- Bucket elimination [Dechter, 1999]



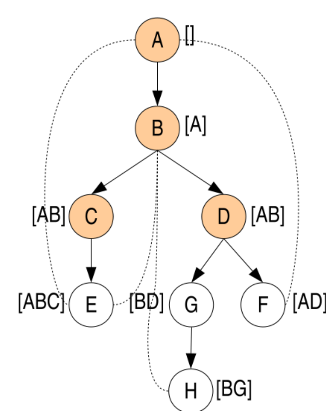
O(exp(w*))

- AND/OR search graph [Mateescu & Dechter, 2007]



- Pseudo tree

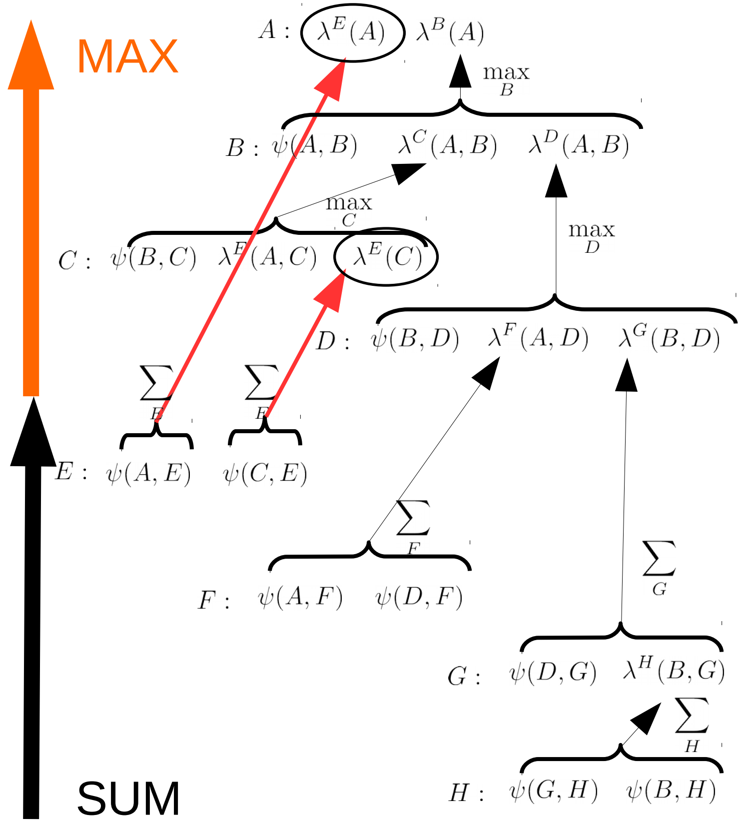
[Freuder & Quinn, 1985]



Node value v(n):
- max for MAP vars
- sum for SUM vars

Background: WMB heuristics

- Mini-bucket elimination [Dechter& Rish 2001]
  - "i-bound", limit on the number of variables in a single mini-bucket



  - Mini-bucket upper bound

$$\sum_E [\psi(A,E)\psi(C,E)] \le [\sum_E \psi(A,E)][\sum_E \psi(C,E)]$$

- Weighted Mini-bucket [Liu & Ihler, 2012]
  - Holder's inequality

$$\sum^{w} [\prod_r (\psi_r)] \le \prod_r [\sum_x^{w_r}]$$

$$\sum_x^{w} f(x) \triangleq [\sum_x f(x)^{\frac{1}{w}}]^w \qquad w = \sum_r w_r$$

- WMB Moment Matching [Liu & Ihler, 2011] [Marinescu,Ihler,Dechter, 2014]
  - MAP variables

$$\psi_{kr} = \psi_{kr} \left(\frac{\mu}{\mu_r} \right)$$

$$\mu_r = \max_{\mathbf{Y}_r} \psi_{kr}; \; \mu = \left(\prod_r \mu_r \right)^{1/R}$$

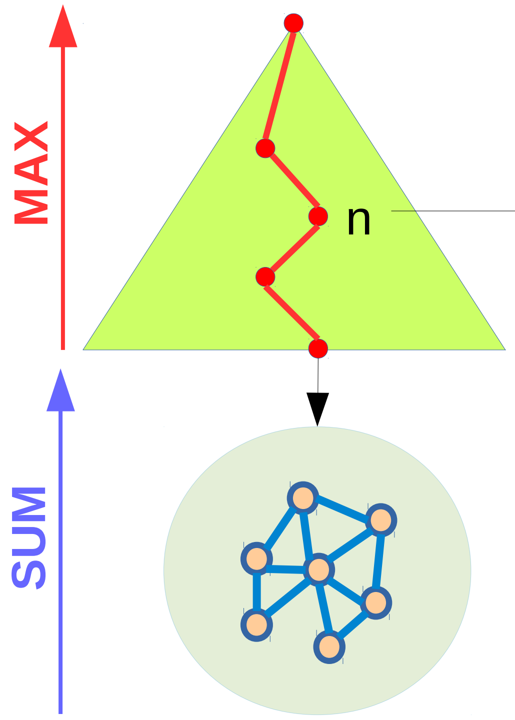
  - SUM variables

$$\psi_{kr} = \psi_{kr} \left(\frac{\mu}{\mu_r} \right)^{w_{kr}}$$

$$\mu_r = \sum_{\mathbf{Y}_r} (\psi_{kr})^{1/w_{kr}}; \; \mu = \prod_r (\mu_r)^{w_{kr}}$$

AND/OR search with mini-bucket heuristics

- Explore the AND/OR search space
  - Heuristic estimates for nodes: WMB scheme
  - Conditioned SUM suproblems: exact search



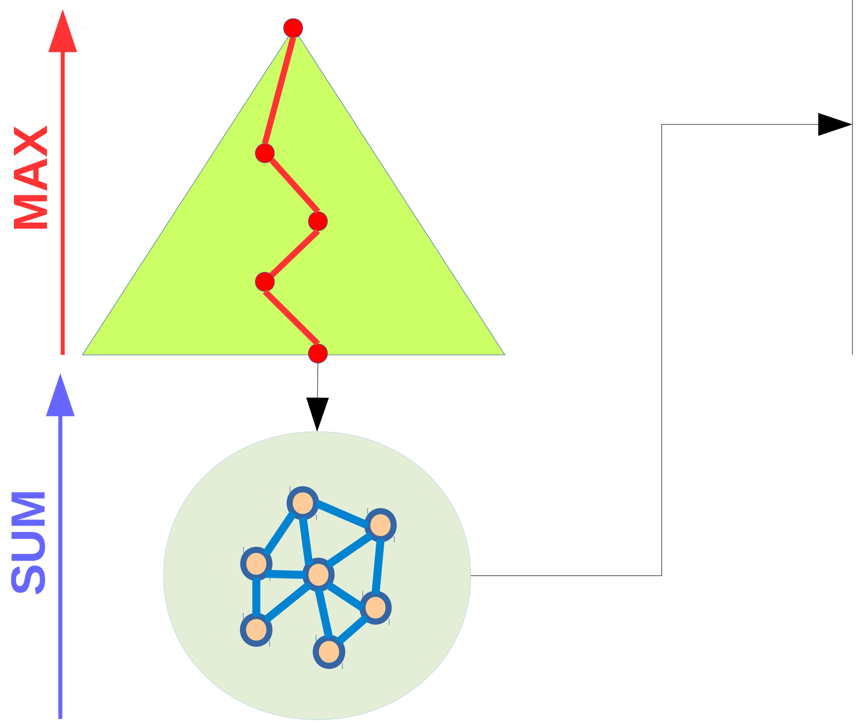
$$h(n) = WMB(\bar{x}_n) \geq v(n)$$

$$Z = \sum_{\mathbf{X}_S} P(X)|_{\bar{x}_M}$$

Probabilistic Lower Bounds

[Liu *et al*. 2015]

$$\max_{\mathbf{X}_M} \sum_{\mathbf{X}_S} P(\mathbf{X})$$



$$Z = \sum_{\mathbf{X}_S} P(X)|_{\bar{x}_M}$$

Solving the conditioned SUM subproblem is hard!

$$\#P - complete$$

Compute a (probabilistic) lower bound on the conditioned sum subproblem

$$Pr(\hat{Z} - \Delta(n,\delta) \leq Z) \geq (1-\delta)$$

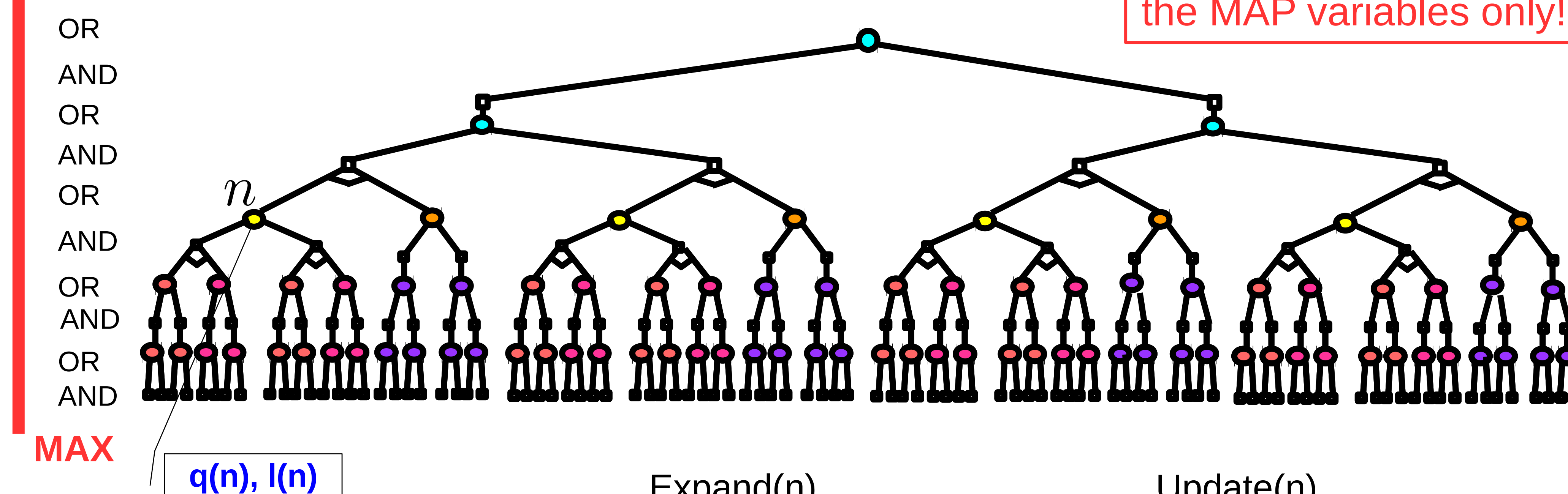
WMB based importance sampling scheme:

- n - number of samples
- δ - confidence value
- Z_{wmb} - result of WMB
- $\hat{Z}$ - Importance Sampling estimate

$$\Delta(n,\delta) = \sqrt{\frac{2\hat{var}(w(x))\log(2/\delta)}{n}} + \frac{7Z_{wmb}\log(2/\delta)}{3(n-1)}$$

AND/OR Search

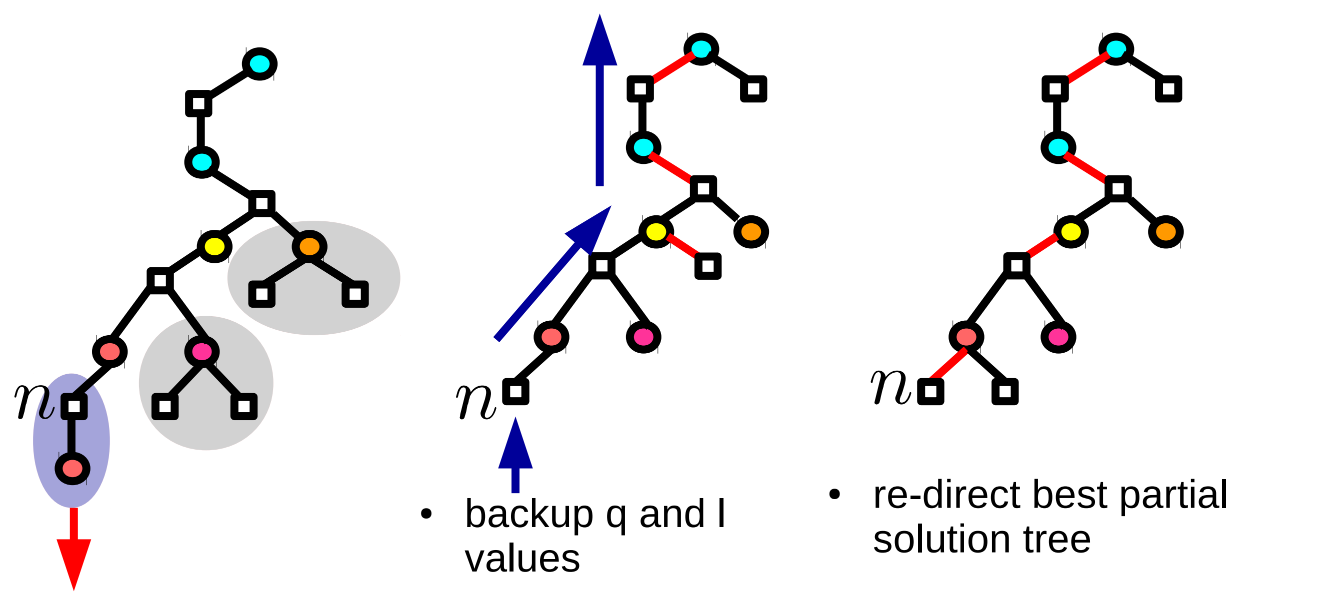
AND/OR search over the MAP variables only!



- **q(n)** : upper bound at n
- **l(n)** : lower bound at n

Expand(n)

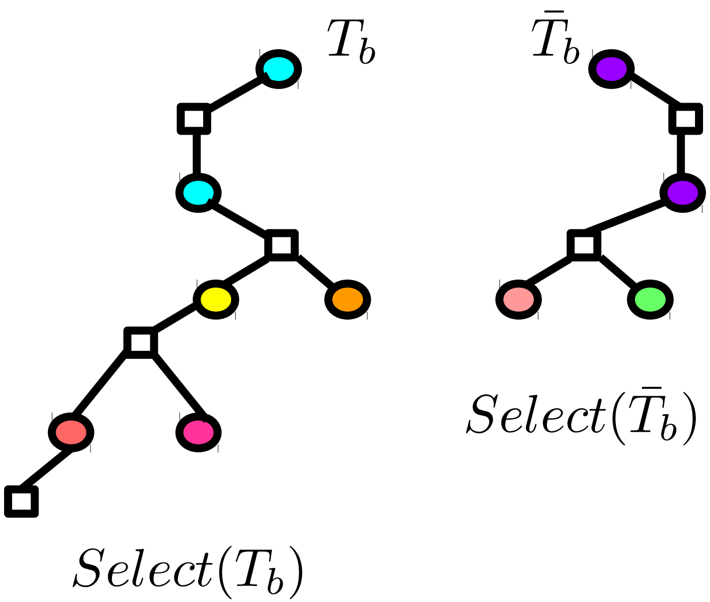
Update(n)



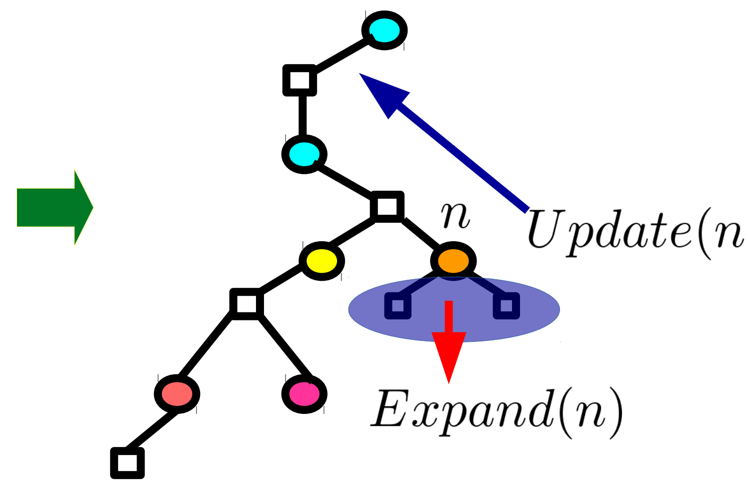
- T_b : best partial solution tree (partial solution tree where OR nodes direct the child m with best q(m)

AnySBFS (anytime stochastic best-first search)

Stochastic node selection



Best-first expansion & update



- Select a tip node n
- Expand and Update n

T_b solution tree

- **Lower bound SUM evaluation based on the WMB-IS scheme**

- Probability p:
  - select tip node in best partial solution tree T_b
- Probability (1-p):
  - select tip node NOT in best partial solution tree $\bar{T}_b$

Search is conduced over the MAP variables only!

AnyLDFS (anytime learning depth-first search)

Best-first update

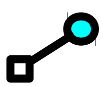
Best-first selection

Depth-first expansion

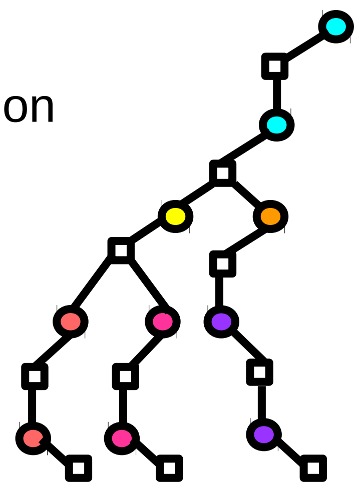
Best-first selection

T_b

$Expand(T_b)$



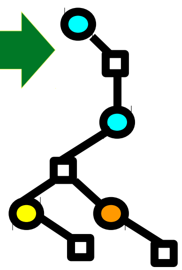
$Select(T_b)$



T_b solution tree

$Update(T_b)$

T_b



$Select(T_b)$

- Greedy depth-first expansion to a solution tree

T_b solution tree

- **Lower bound SUM evaluation based on the WMB-IS scheme**

Search is conduced over the MAP variables only!

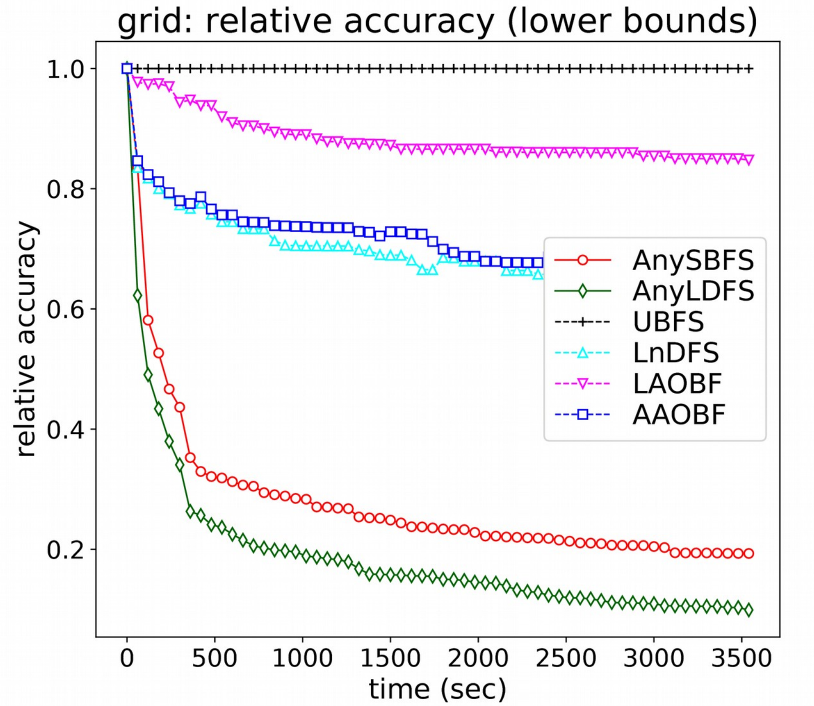
Experiments

- Anytime Algorithms
  - State-of-the-art
    - LAOBF [Marinescu, Lee, Dechter, Ihler, 2017]
    - AAOBF [Marinescu, Lee, Dechter, Ihler, 2017]
    - LnDFS [Marinescu, Lee, Dechter, Ihler, 2017]
    - UBFS [Qi, Ihler, 2018]

    (LAOBF, AAOBF, LnDFS: Hybrid best+depth-first search)
  - **Proposed schemes**
    - **AnySBFS (p = 0.5)**
    - **AnyLDFS**
- Benchmarks
  - Grid, pedigree, promedas, planning (UCIrvine graphical models repo)
  - 10% of variables selected randomly as MAP variables
    - Hard (intractable) conditioned summation subproblems
  - Parameters: confidence 0.05

Results

$$acc_{lb} = \frac{|l_t - l^*|}{l^*}$$

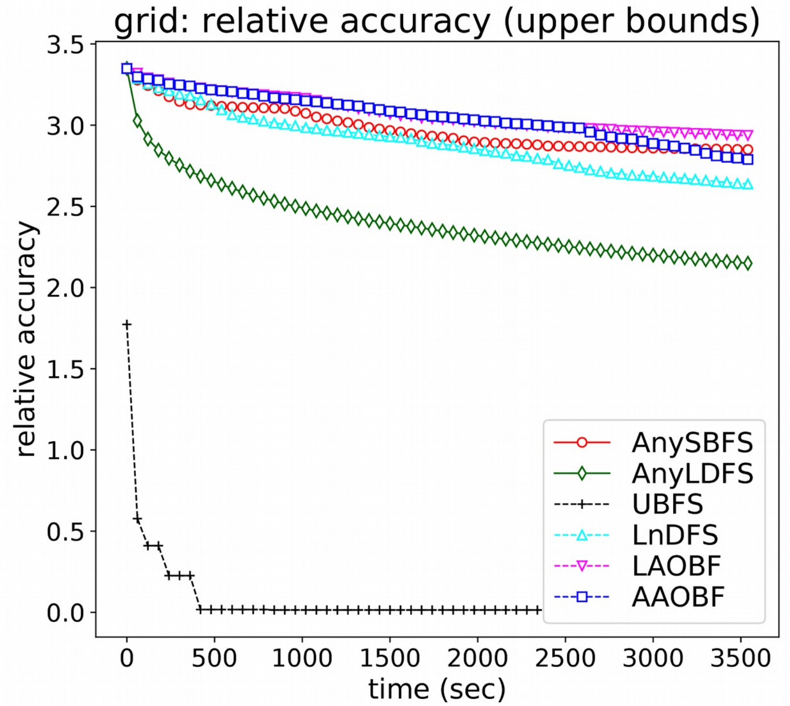


l_t – lower bound at time t

l^* – tightest lower bound found

Average over 150 instances

$$acc_{ub} = \frac{|u_t - u^*|}{u^*}$$



u_t – upper bound at time t

u^* – tightest upper bound found

Average over 150 instances

Conclusion

- Stochastic anytime search algorithms improved upon the state-of-the-art algorithms, especially upon hybrid search
  - higher quality anytime solutions with high probability
  - tighter anytime upper bounds
  - effective use of memory

- Applicable to MMAP with intractable summation subproblems where existing schemes could not produce a single solution