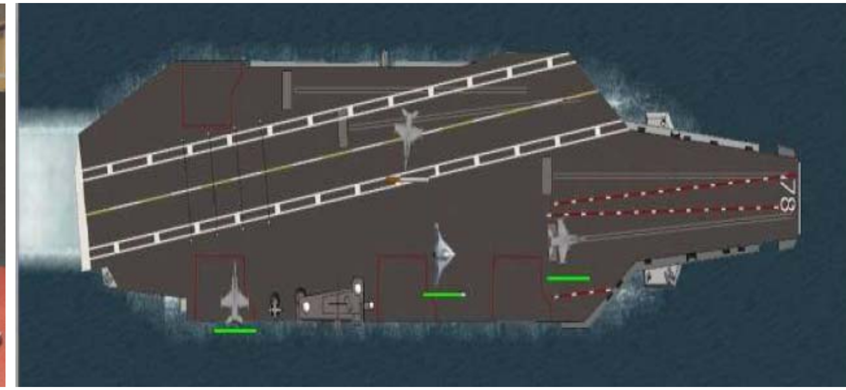
Planning in Information Space with Macro-actions

Nicholas Roy

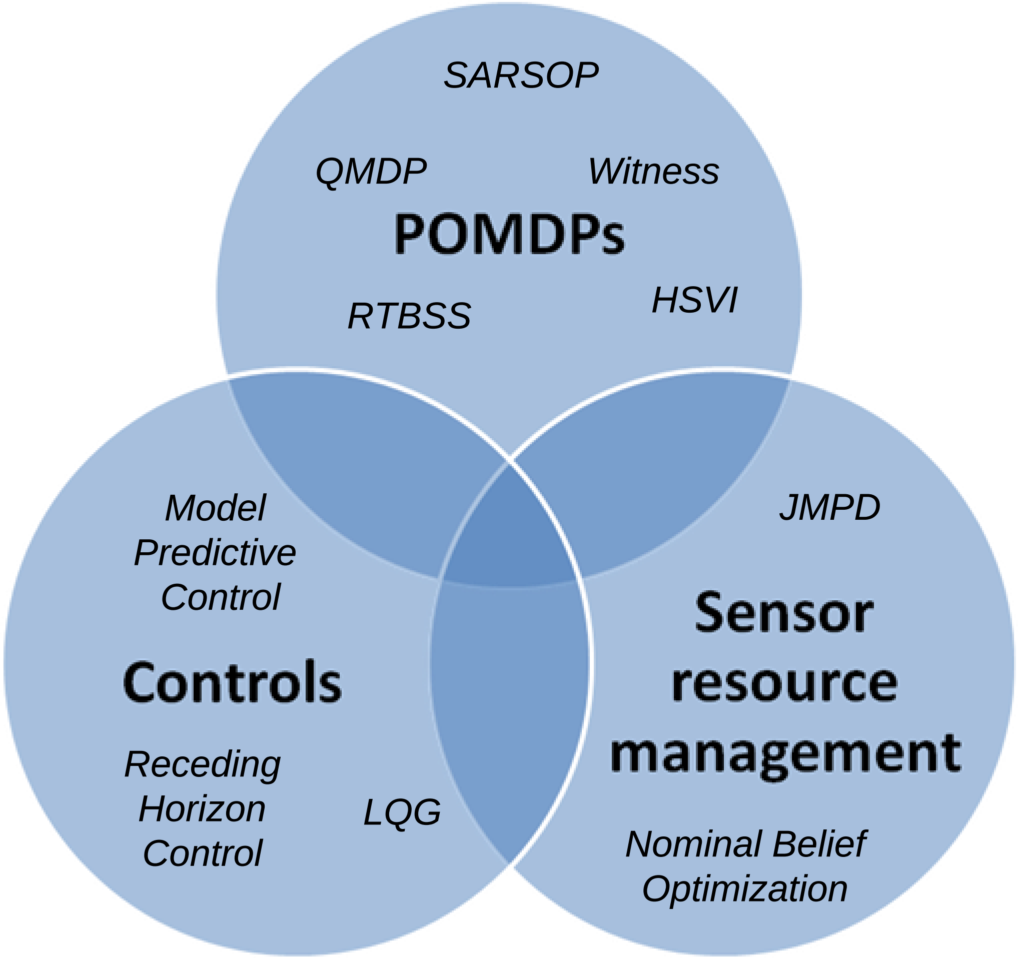
AERO ASTRO

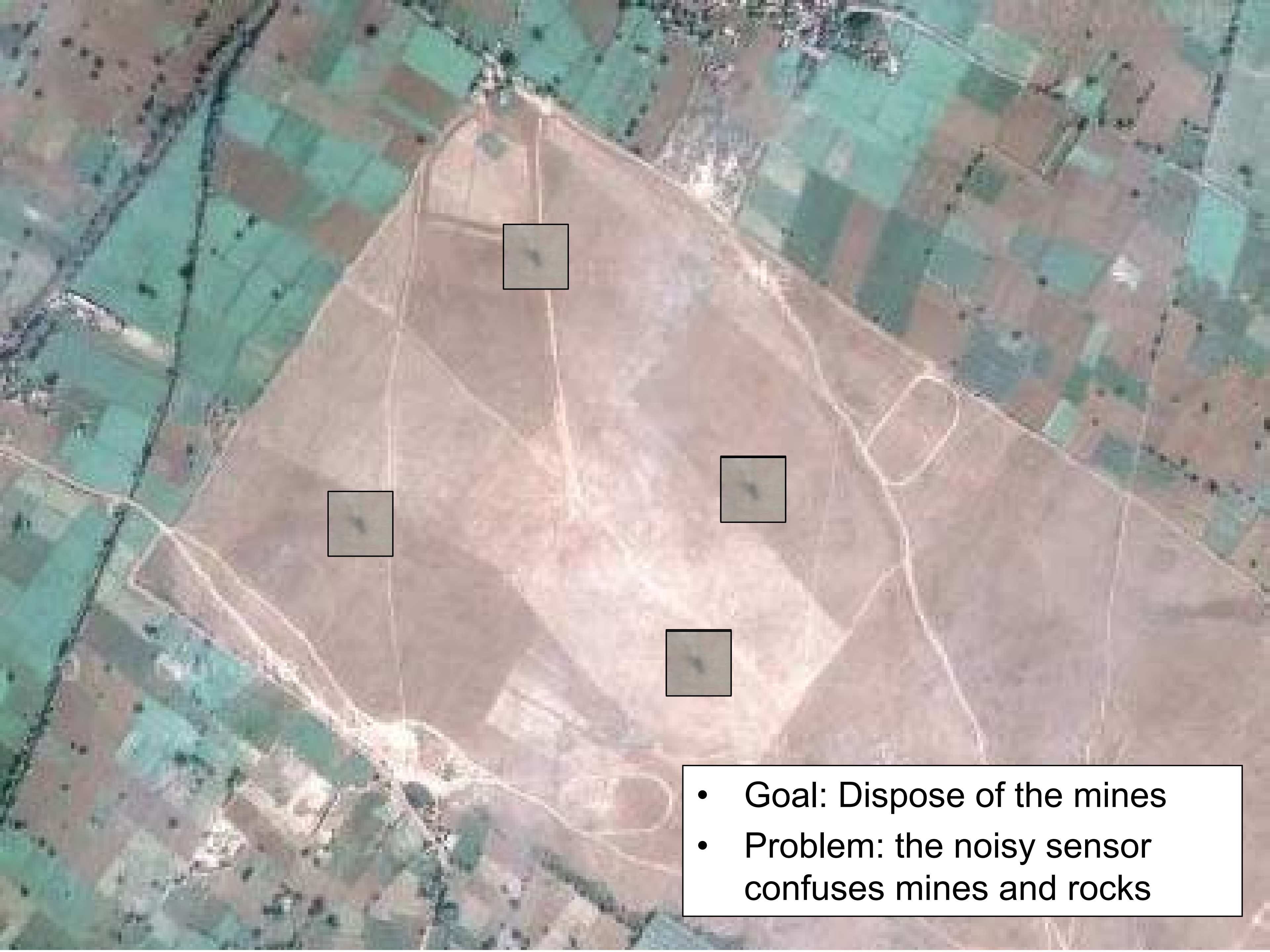

Autonomous
Navigation


Interactive



- What’s different about active learning on robots?
  - Much more complex objective.
    - Costs of learning must be explicitly modelled.
  - Must avoid failures while learning.
    - Want to behave reasonably while learning.
  - IID assumptions are deeply troubling.
    - Acquiring data can often change the underlying distribution.


SARSOP
QMDP
Witness
POMDPs
RTBSS
HSVI
Model Predictive Control
Controls
Receding Horizon Control
LQG
JMPD
Sensor resource management
Nominal Belief Optimization

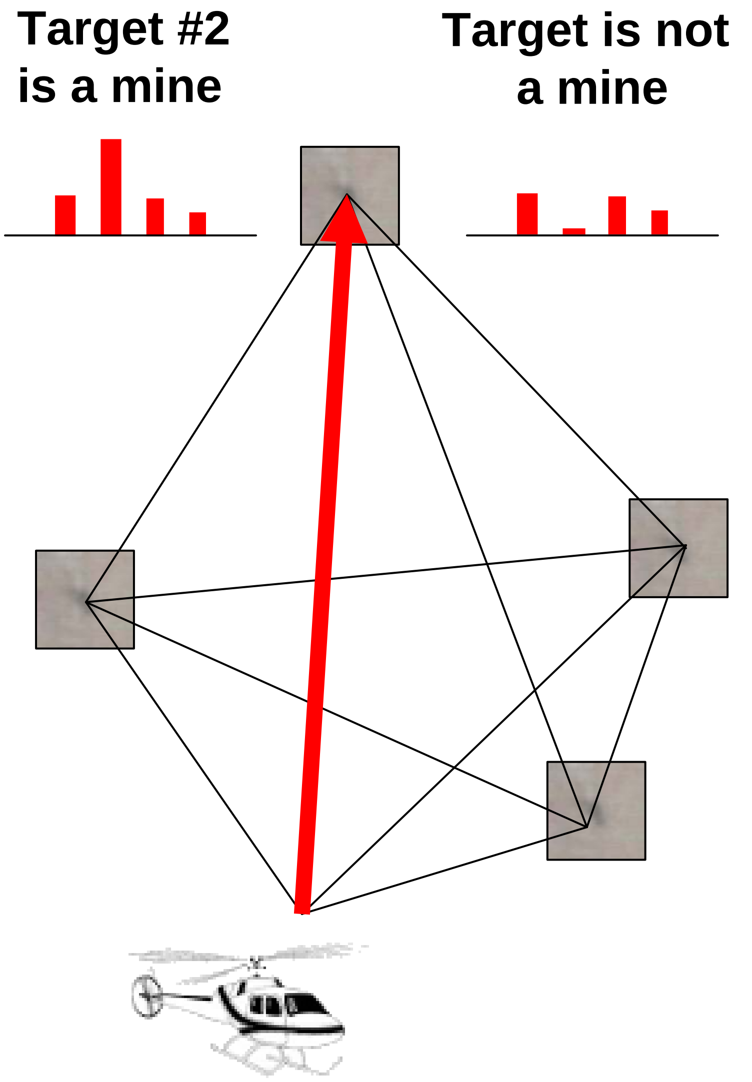


- Goal: Dispose of the mines
- Problem: the noisy sensor confuses mines and rocks

RockSample

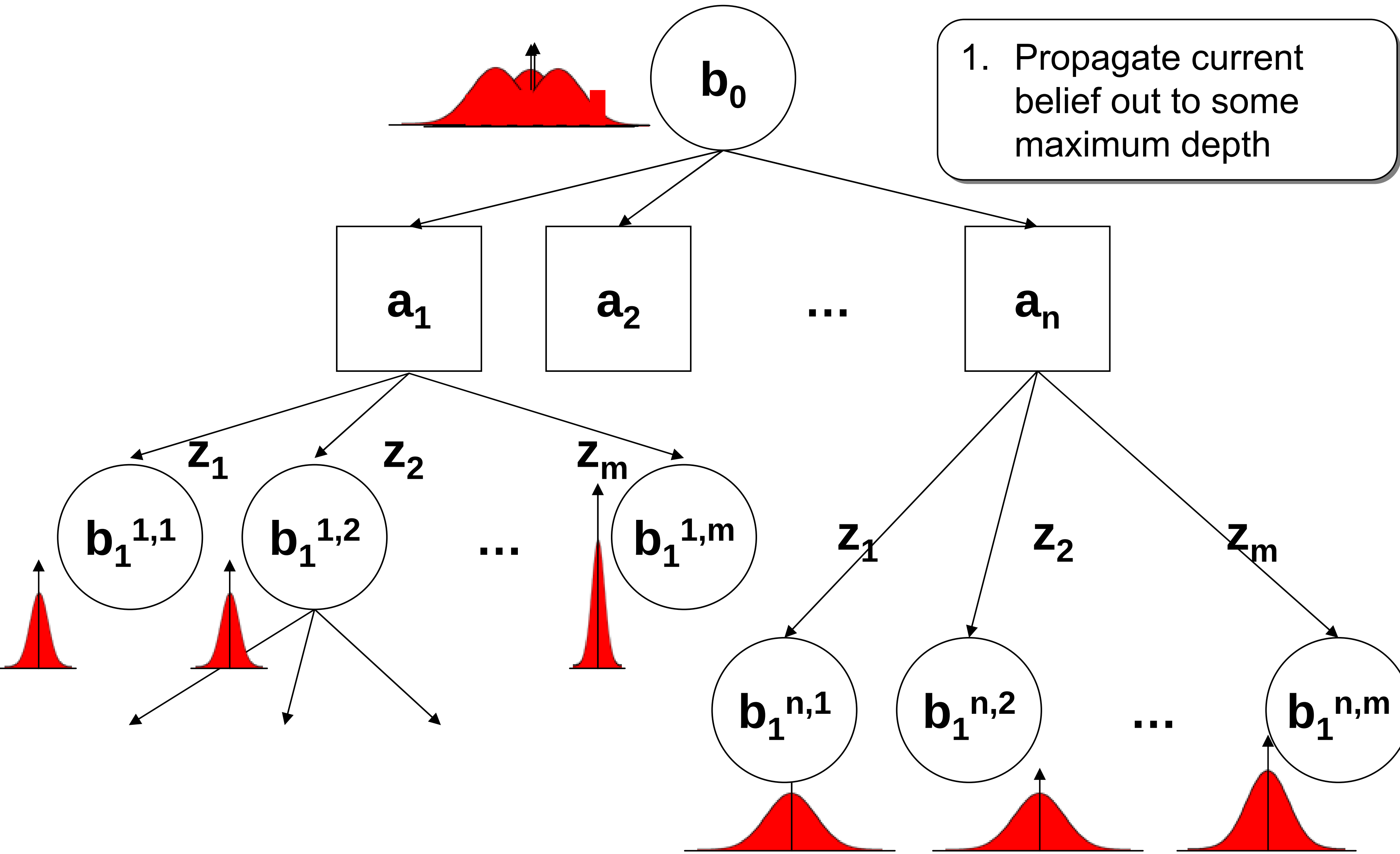
- Given cost of flight, reward of disposing of actual mines…
- Search for a sequence of paths through the graph that maximize expected reward

- Posterior distribution is not deterministic



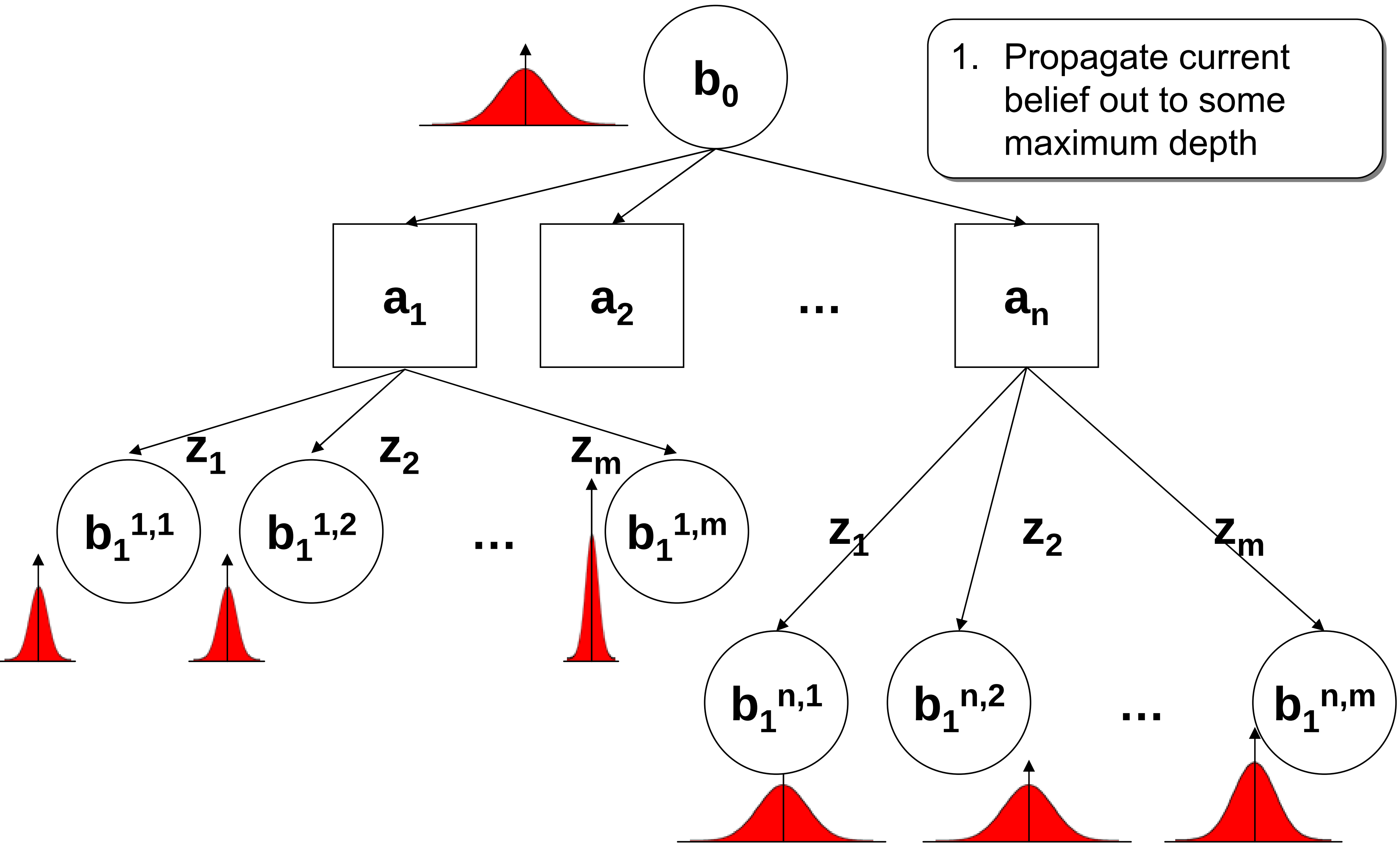
Forward Search in Information Space

1. Propagate current belief out to some maximum depth

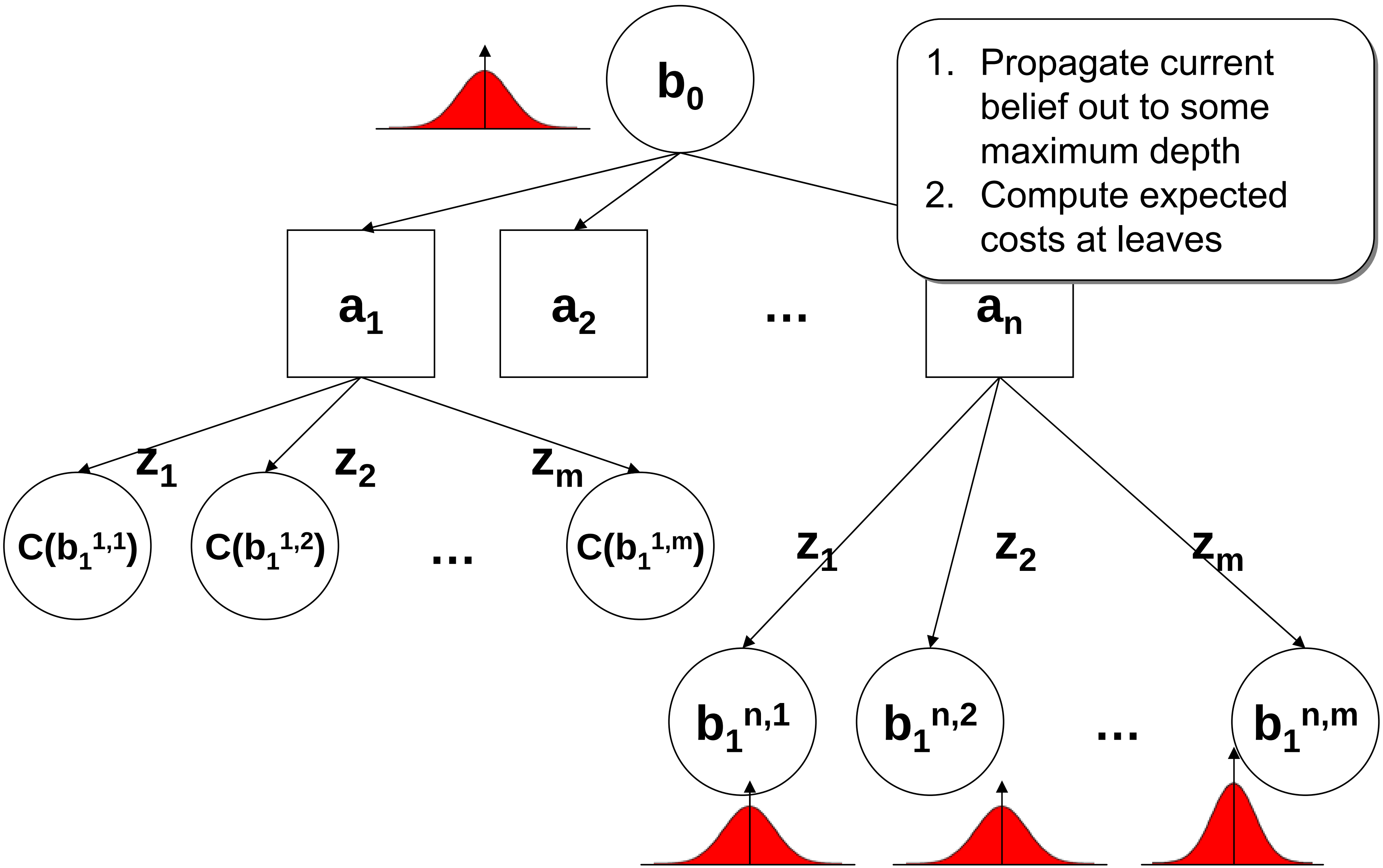


Forward Search in Information Space

1. Propagate current belief out to some maximum depth

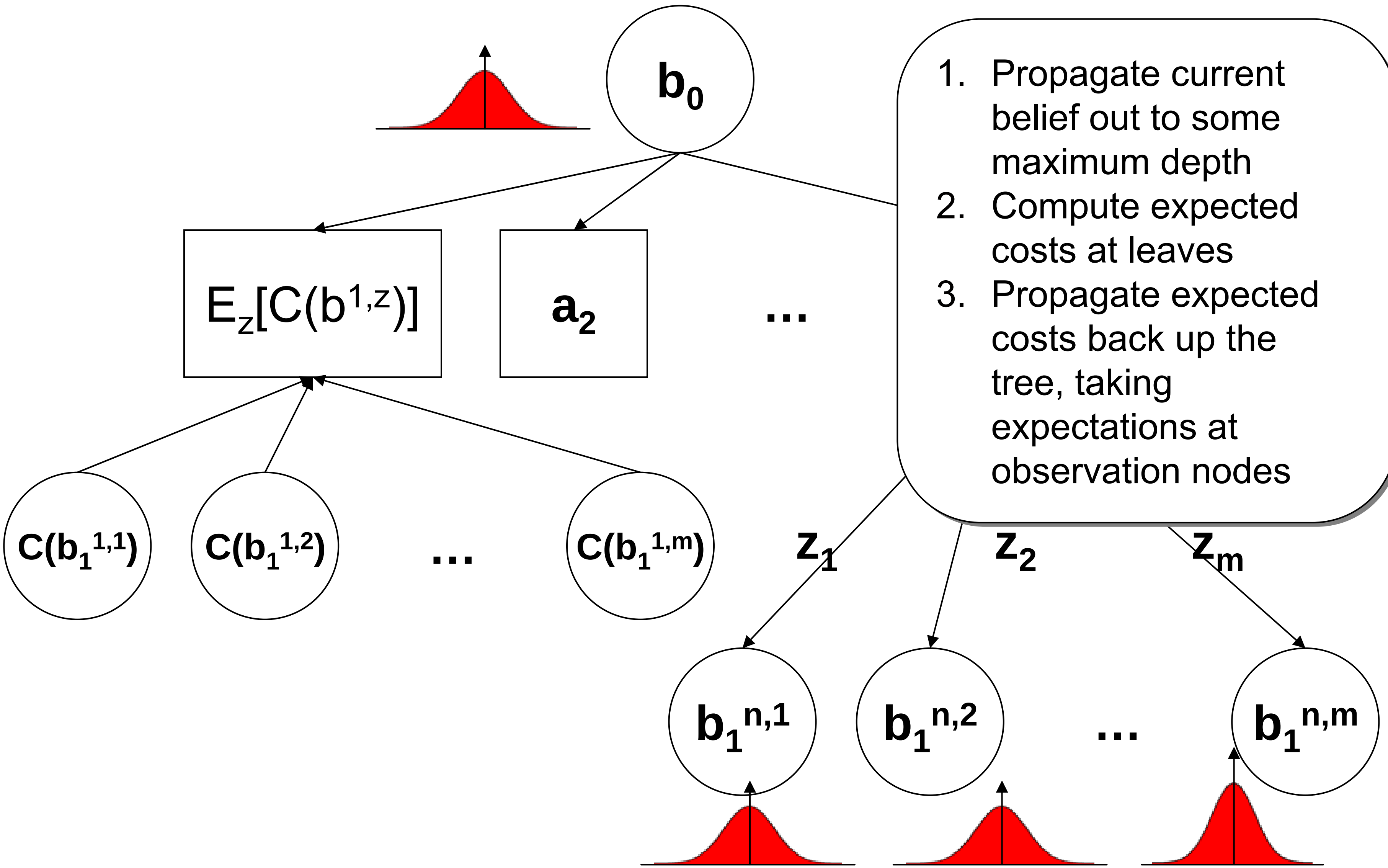


Forward Search in Information Space



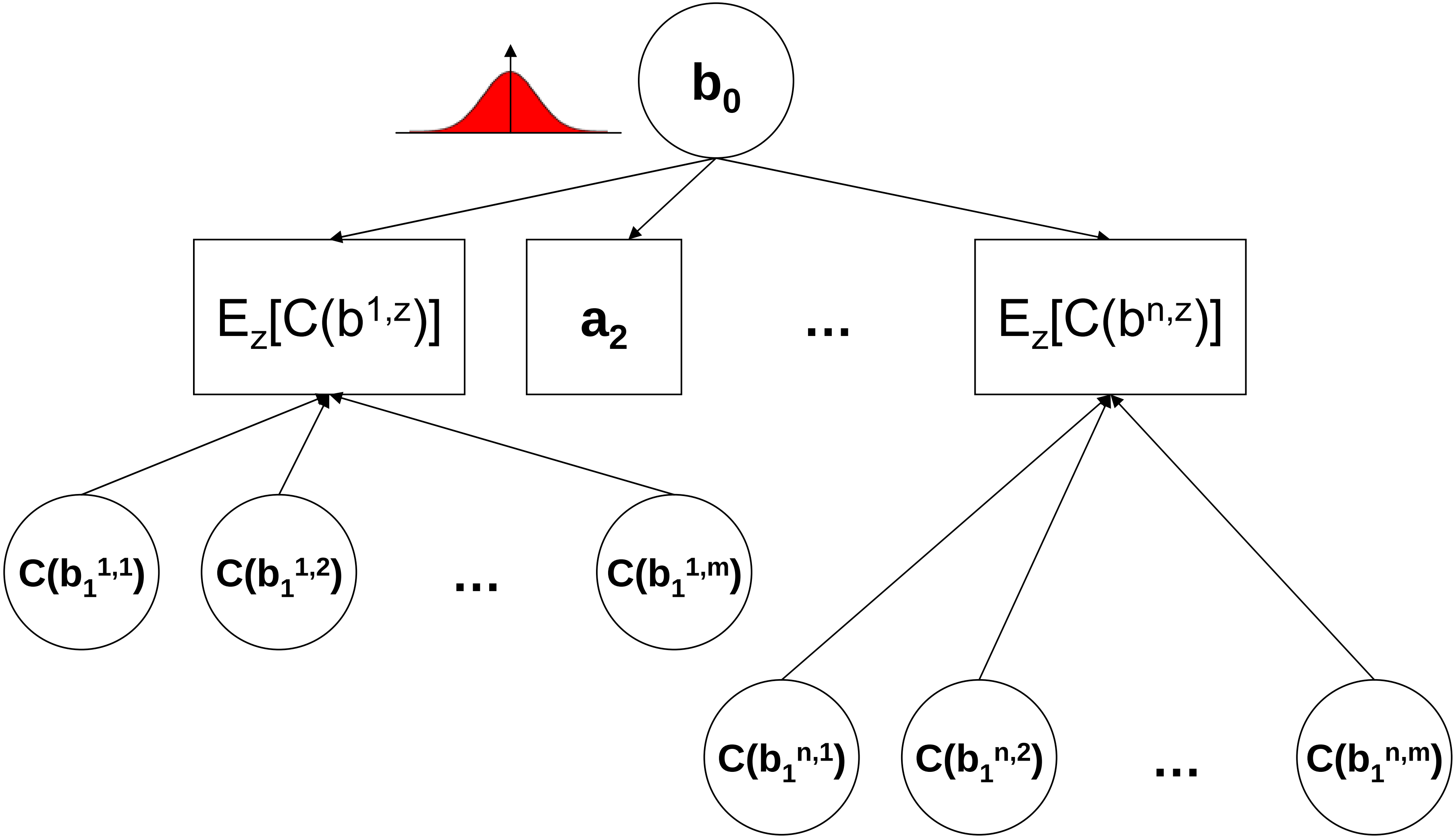
1. Propagate current belief out to some maximum depth
2. Compute expected costs at leaves

Forward Search in Information Space



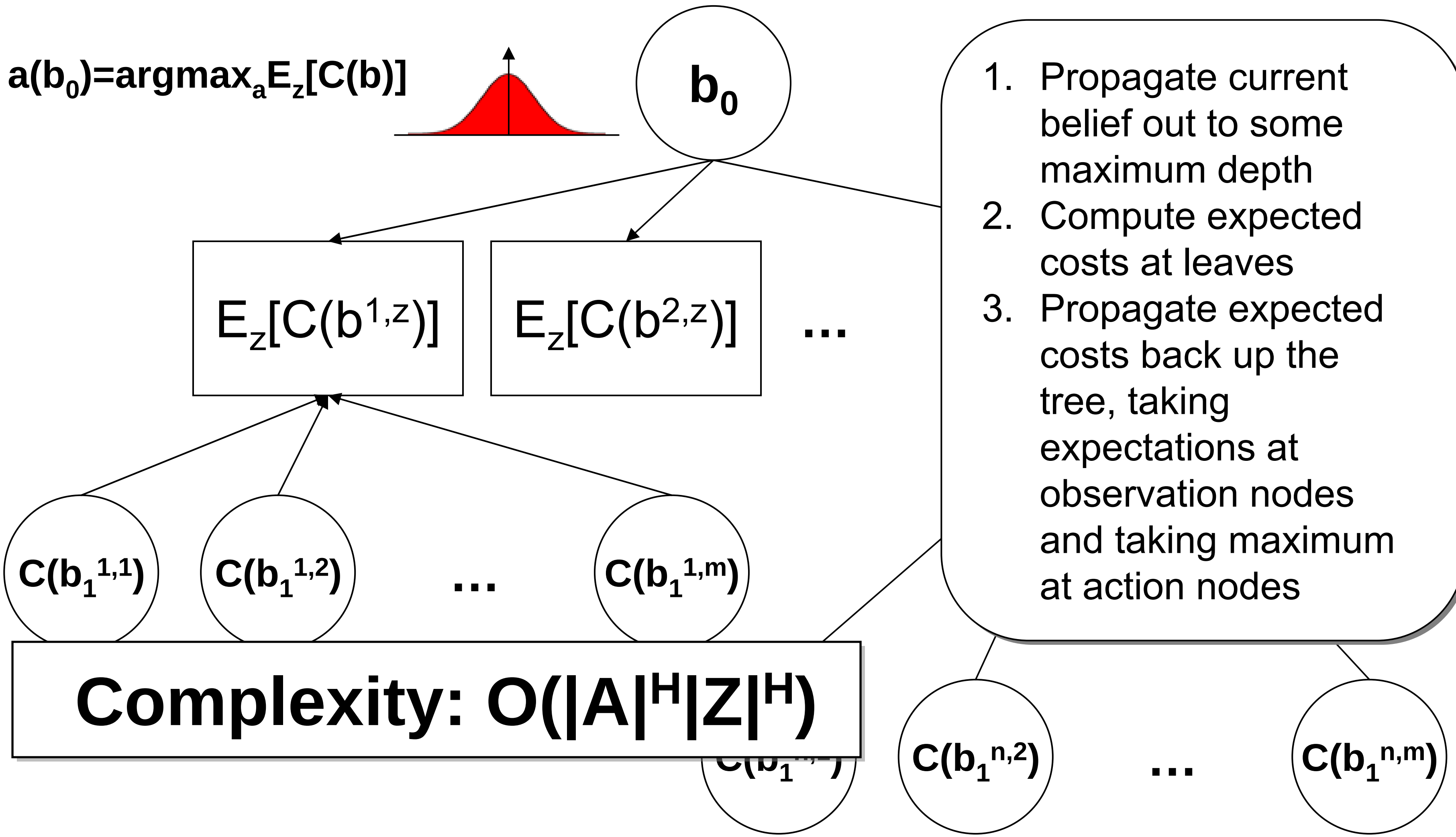
1. Propagate current belief out to some maximum depth
2. Compute expected costs at leaves
3. Propagate expected costs back up the tree, taking expectations at observation nodes

Forward Search in Information Space



Forward Search in Information Space

$a(b_0)=\text{argmax}_a E_z[C(b)]$

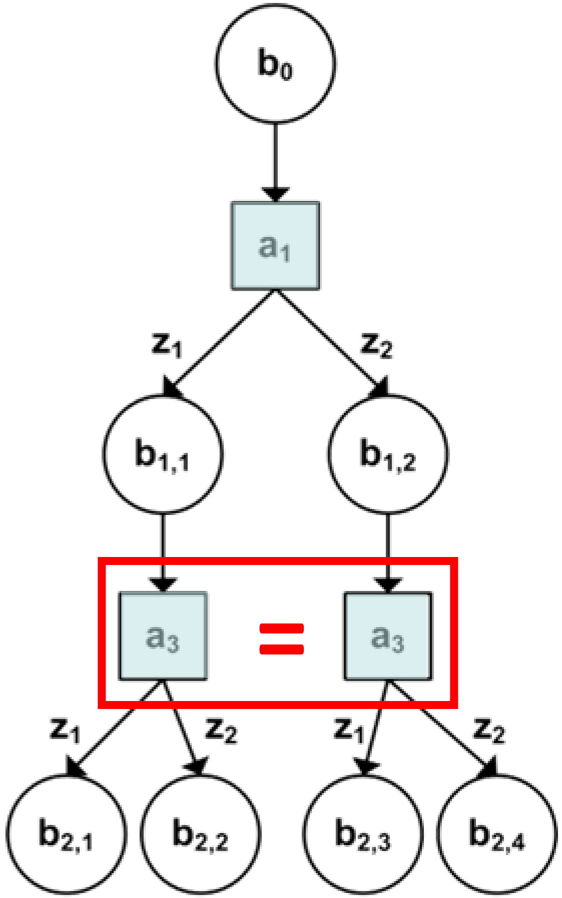


1. Propagate current belief out to some maximum depth
2. Compute expected costs at leaves
3. Propagate expected costs back up the tree, taking expectations at observation nodes and taking maximum at action nodes

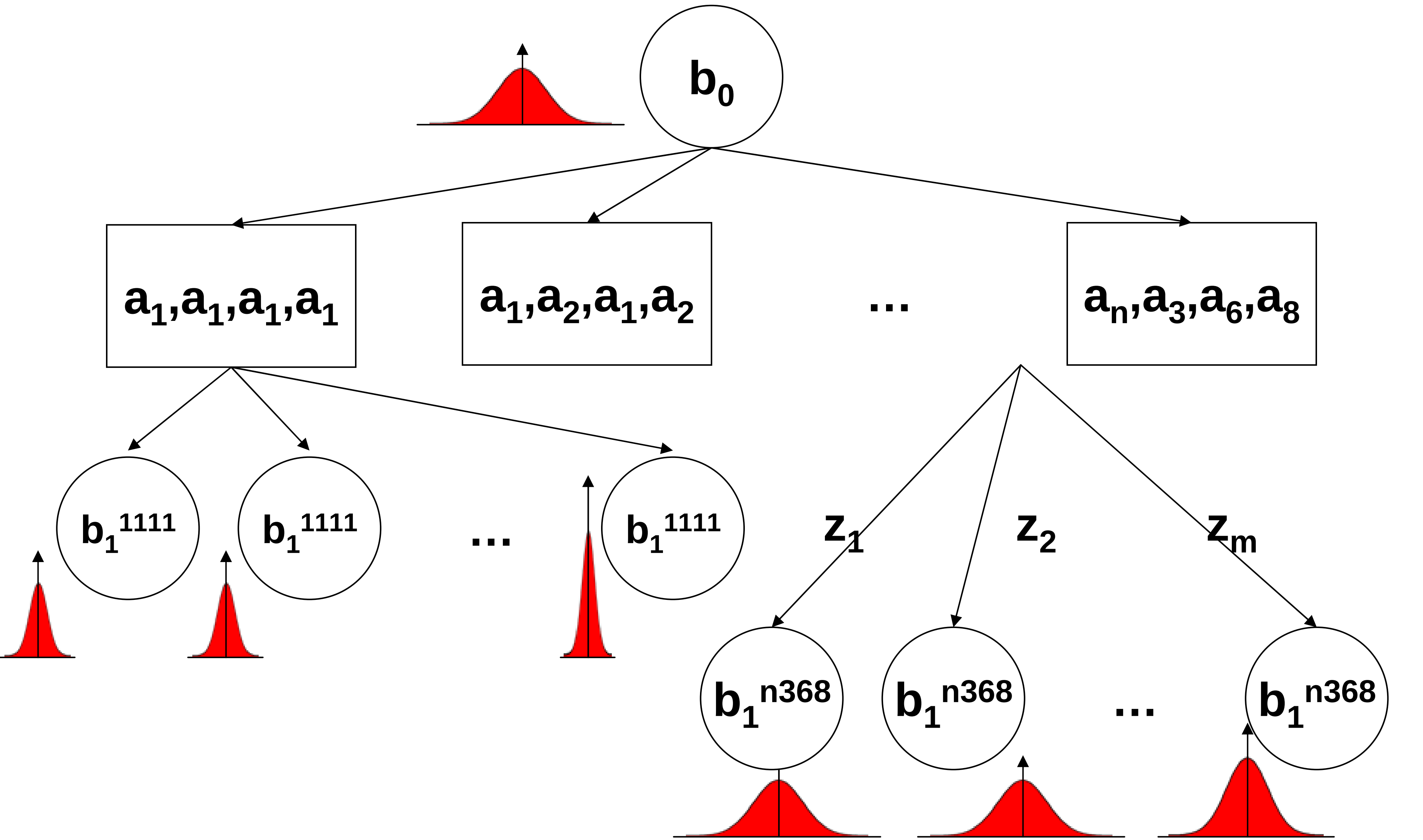
Complexity: $O(|A|^H|Z|^H)$

Macro Actions

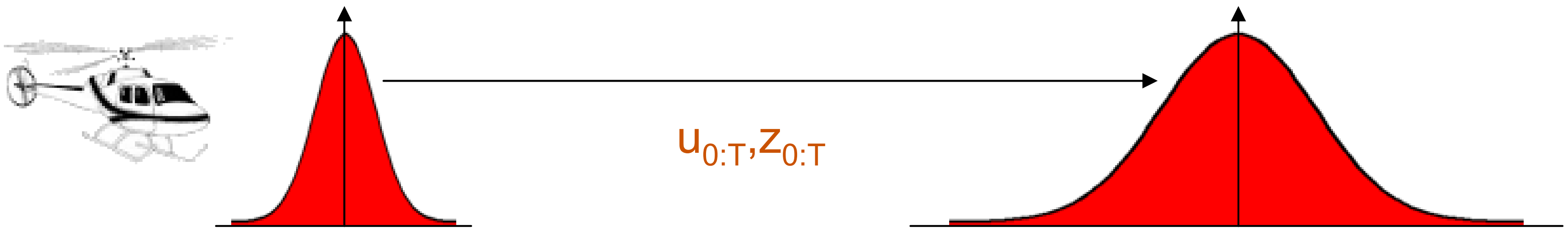
- Condition only at key points
- Macro-actions
  - Fixed-length, open-loop action sequences



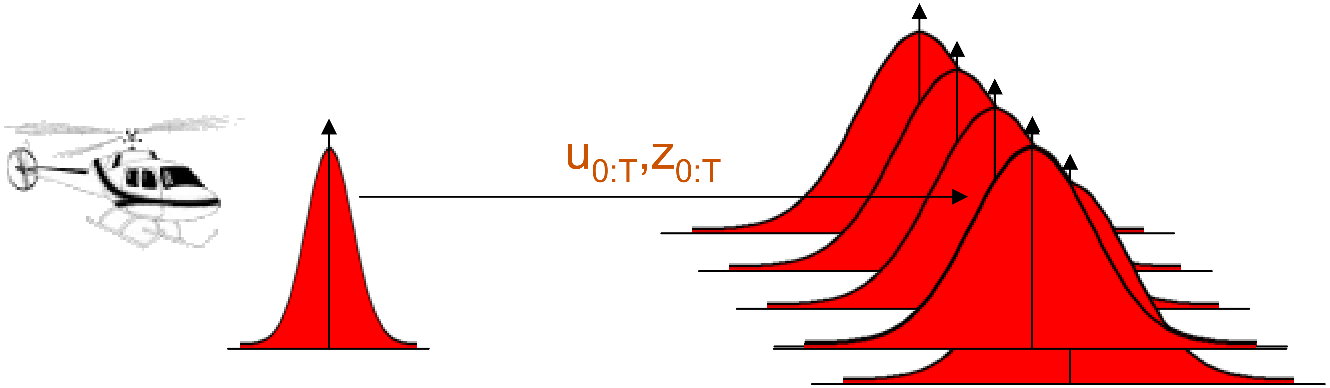
Macro Actions



How to generate posteriors

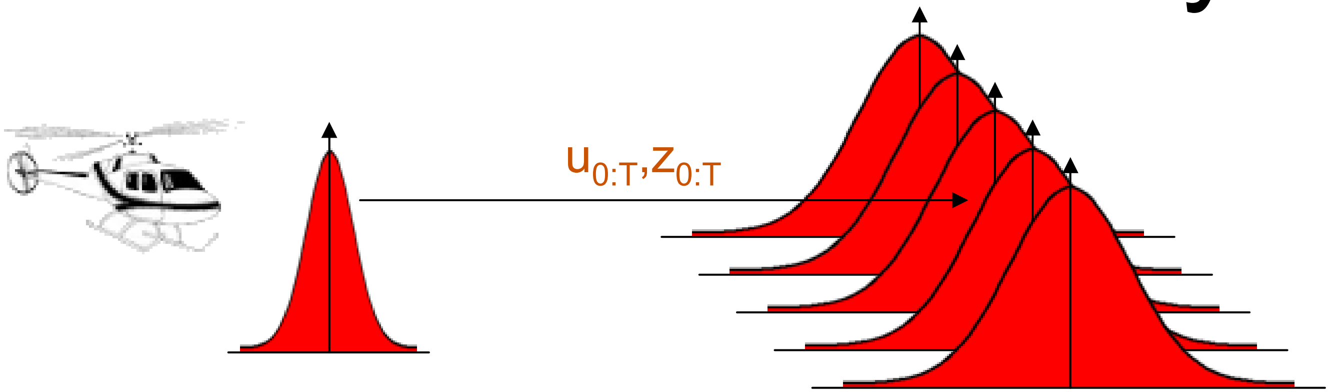


Posterior Belief Distribution



- Posterior belief not deterministic
- Action sequence leads to a distribution over posterior beliefs
- Compute expected reward over distribution of distributions
- Compare ∫R(b'|u'0:T)db' > ∫R(b|u0:T)db

Exact Linear Gaussian Systems



- Analytic solution exists for linear Gaussian systems : O(n)

$$\mu_{t-1} \sim N\left(m_{t-1}, \Sigma^{\mu}{}_{t-1}\right)$$

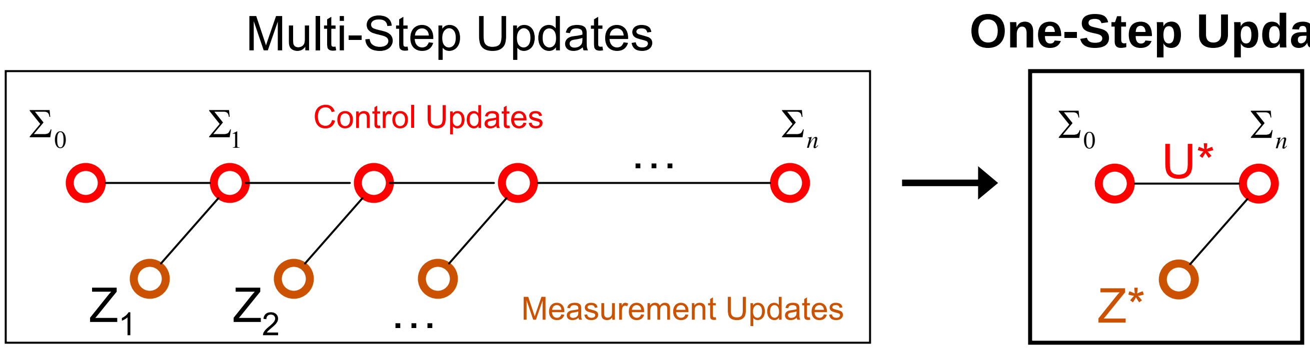
$$\mu_t \sim N\left(A_t m_{t-1} + B_t u_t, \Sigma^{\mu}{}_{t-1} + \overline{\Sigma}_t C_t K_t{}^T\right)$$

Multi-Step Update as One-Step

EKF Covariance Update

Control: $\bar{\Sigma}_t = G\Sigma_{t-1}G^T + R$

Measurement: $\Sigma_t = \left(\bar{\Sigma}_t^{-1} + HQ^{-1}H^T\right)^{-1}$



Solution: Decomposition

- Key idea: factor the covariance matrix

$$\Sigma = BC^{-1}$$

- Motion update

$$\bar{\Sigma}_t = \bar{E}_t \bar{D}_t^{-1}$$

$$\begin{bmatrix} \bar{D}_t \\ \bar{E}_t \end{bmatrix} = \begin{bmatrix} 0 & G_t^{-T} \\ G_t & R_t G_t^{-T} \end{bmatrix} \begin{bmatrix} B_{t-1} \\ C_{t-1} \end{bmatrix}$$

Solution: Decomposition

- Key idea: factor the covariance matrix

$$\Sigma = BC^{-1}$$

- Measurement update

$$\Sigma_t = B_t C_t^{-1}$$

$$\begin{bmatrix} B_t \\ C_t \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & M_t \end{bmatrix} \begin{bmatrix} \overline{D}_t \\ \overline{E}_t \end{bmatrix}$$

Solution: Decomposition

- One-step transfer function for the covariance:

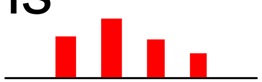
$$\zeta_t = \begin{bmatrix} 0 & I \\ I & M_t \end{bmatrix} \begin{bmatrix} 0 & G_t^{-T} \\ G_t & R_t G_t^{-T} \end{bmatrix}$$

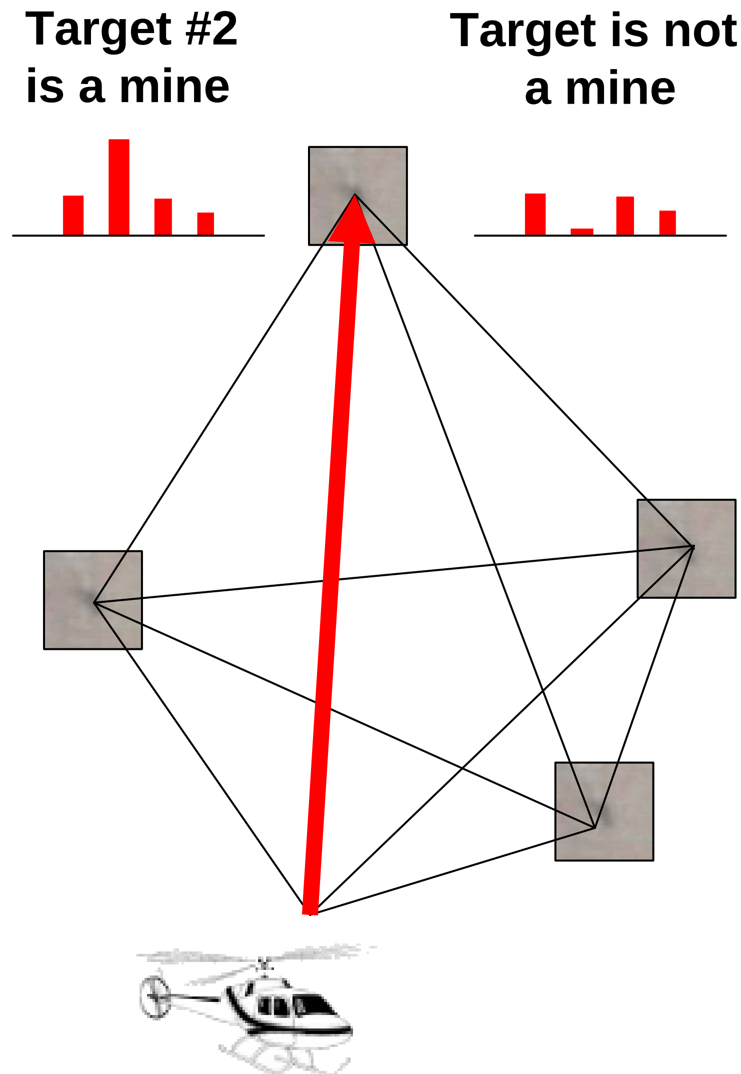
$$\Rightarrow \begin{bmatrix} B_T \\ C_T \end{bmatrix} = \left(\prod_{t=0}^{T} \zeta_t \right) \begin{bmatrix} B_0 \\ C_0 \end{bmatrix}$$

- (To recover covariance, $\Sigma = BC^{-1}$)
- This trick is not new.
  - Kaileth et al., Linear State Estimation.
  - Mourikis and Roumeliotis, 2006.

RockSample

- Given cost of flight, reward of disposing of actual mines…
- Search for a sequence of paths through the graph that maximize expected reward

- Posterior distribution is not deterministic
- Distribution is multinomial 



Non-Gaussian Distributions

- Approximate version available for exponential family distributions
  - e.g., Approximate parameter **x** of a Bernoulli with Gaussian

$$x_{t-1} \sim N(\mu_{t-1}, \Sigma_{t-1})$$

$$x_t = A_t x_{t-1} + B_t u_t + \varepsilon_t$$

Non-Gaussian Distributions

- Approximate version available for exponential family distributions
  - e.g., Probability that a rock is actually an IED

$$x_{t-1} \sim N(\mu_{t-1}, \Sigma_{t-1})$$

$$x_t = A_t x_{t-1} + B_t u_t + \varepsilon_t$$

Observation model is also Bernoulli, e.g., chemical detector

$$p(z_t \mid \theta_t) = \exp\left(z_t{}^T \theta_t - \beta_t(\theta_t) + \kappa_t(z_t)\right)$$

$$\theta_t = W(x_t)$$

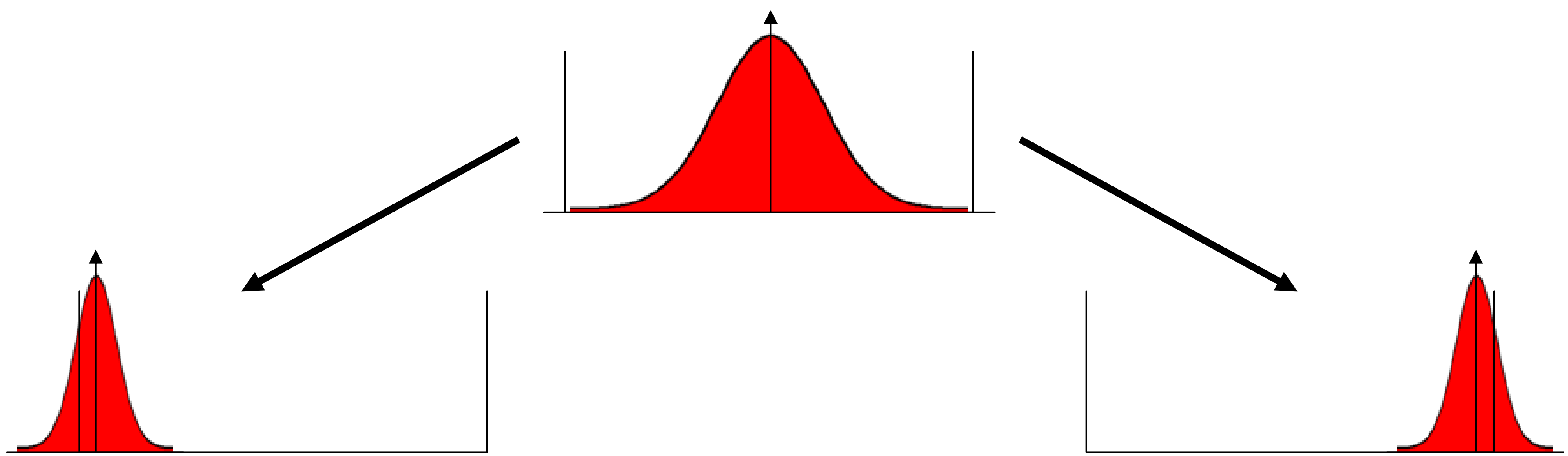
$$Y_t = \left.\frac{\partial W(x_t)}{\partial x_t}\right|_{x_t = \overline{\mu}_t}$$

Kalman Filtering for Non-Gaussian Distributions

$$\tilde{z}_t = \left(\bar{\theta}_t - \ddot{\beta}_t^{-1}\left(\dot{\beta}_t - z_t\right)\right)$$

$$\mu_t = \bar{\mu}_t + K_t\left(z_t - W\left(\bar{\mu}_t\right)\right)$$

$$K_t = \bar{\Sigma}_t Y_t \left(Y_t \bar{\Sigma}_t Y_t + \ddot{\beta}_t^{-1}\right)^{-1}$$



Experimental Performance

Problem	Algorithm	Ave. rewards		Online time(s)	Offline time (s)
ISRS (8,5)	SARSOP	12.10	±0.26	0.00	10000
	Naïve FS	9.56	±1.08	3.36	0.00
	Hand-coded SCP	19.71	±0.63	0.74	0.00
	PUMA	17.38	± 1.41	162.48	0.00

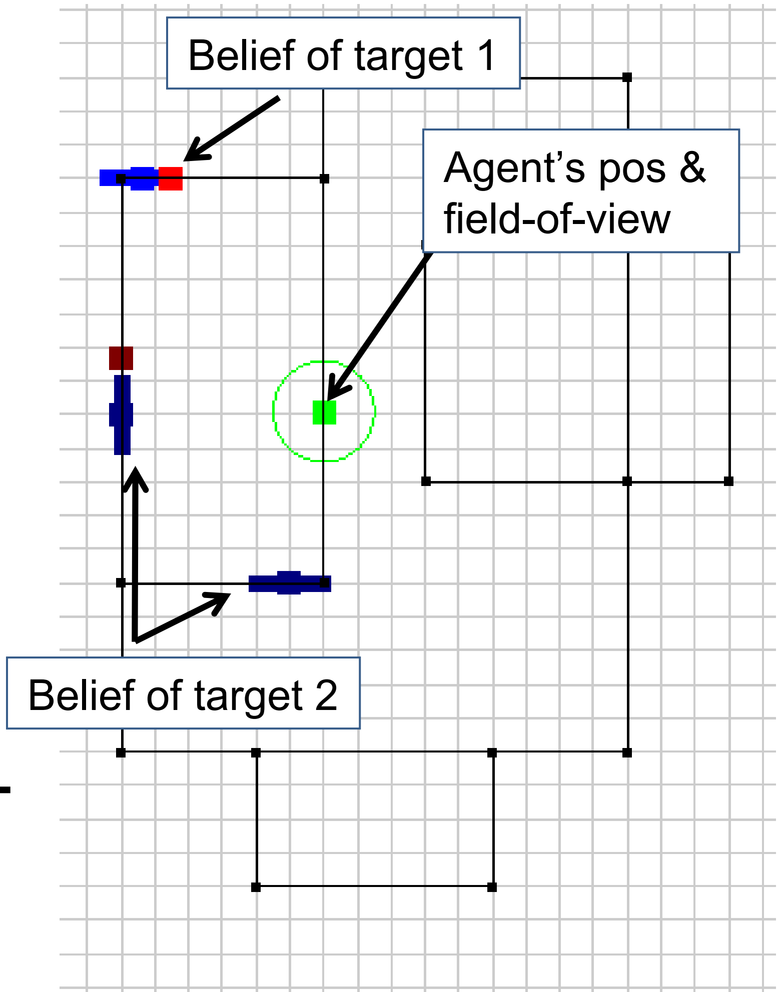
- ISRS: 2048 states
- Largest version of this problem solved so far: $10^4 \times 2^{30}$ states, 1-2 minutes per step

Experimental Performance

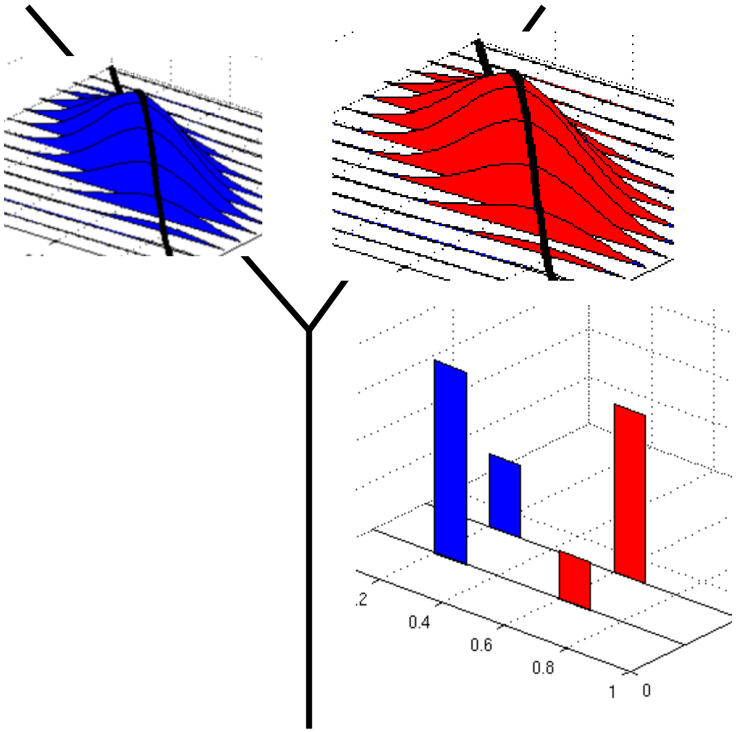
Problem	Algorithm	Ave. rewards		Online time(s)	Offline time (s)
ISRS (8,5)	SARSOP	12.10	±0.26	0.00	10000
	Naïve FS	9.56	±1.08	3.36	0.00
	Hand-coded SCP	19.71	±0.63	0.74	0.00
	PUMA	17.38	± 1.41	162.48	0.00
Tracking	Naïve FS	19.58	± 0.42	0.023	0.00
	Hand-coded SCP	27.48	± 0.49	1.010	0.00
	PUMA	35.55	± 1.28	28.52	0.00

Multi-Target Tracking

- Helicopter tracking multiple targets
- Limited field-of-view, noisy sensor
- Goal: Minimize uncertainty + distance travelled

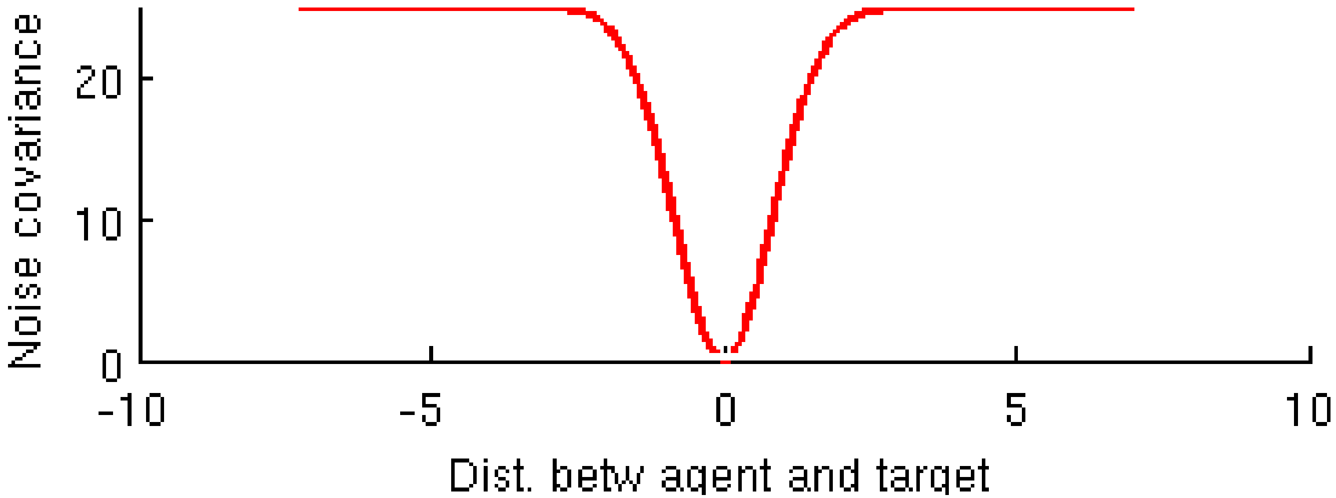


Multi-Modal Gaussian Posterior



Sensor Model

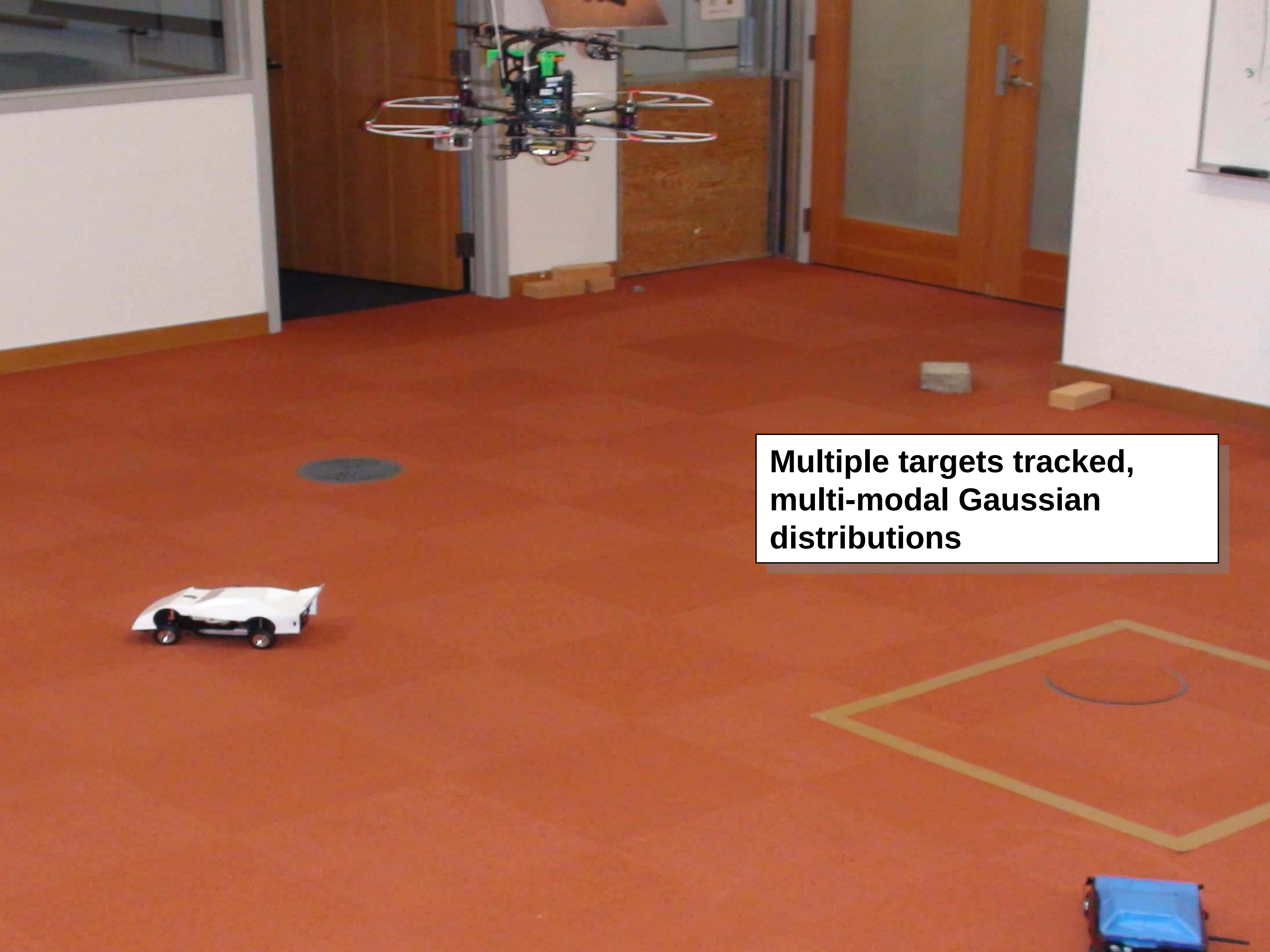
- Sensor covariance is a Gaussian function of range



Comparisons

	Dist. traveled	Ave. modes	Total cost
MMPBD	138.76	1.081	-51.76
Greedy	133.52	1.524	-61.25
Naïve FS	112.20	1.775	-85.11

- Greedy strategy
  - Localize target with largest covariance
- Naïve Forward Search
  - Primitive actions
  - “Macro-observations”


Multiple targets tracked, multi-modal Gaussian distributions

Summary

- Robust, long-term autonomy in large-scale environments
- Planning algorithms for worlds in which we have limited knowledge of the state, model of the system, or a map of the world
- Key Issue: Control of Information
- Technical approaches:
  - Understanding how information propagates
  - Machine learning