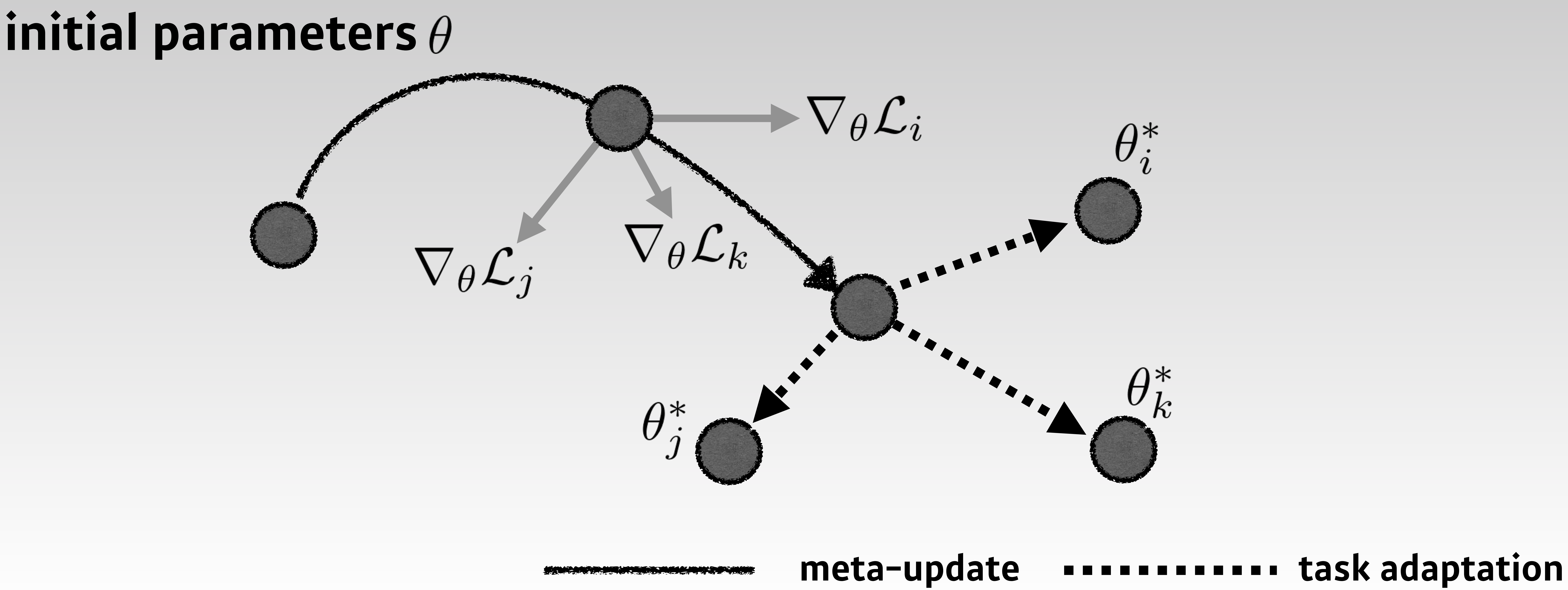
Bayesian Model-Agnostic Meta-Learning

Taesup Kim* (presenter), Jaesik Yoon*
Ousmane Dia, Sungwoong Kim, Yoshua Bengio, Sungjin Ahn

Model-Agnostic Meta-learning (MAML)

“gradient-based meta-learning framework”



Model-Agnostic Meta-learning (MAML)

For each task in a batch:

Task Model ← **Task adaptation**

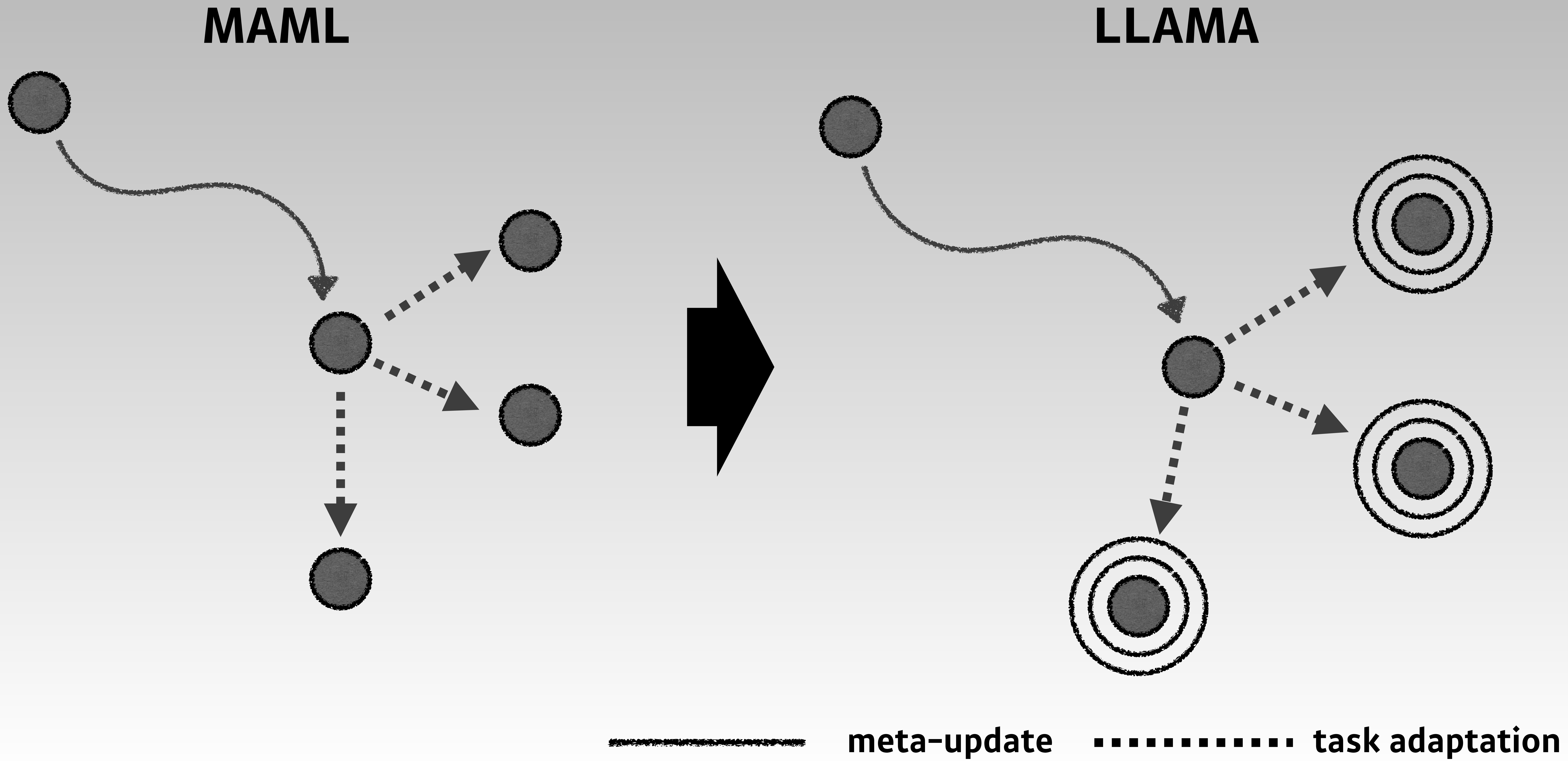
Initial Model ← **Meta-update**

Gradient-Based Meta-Learning + "Bayesian"

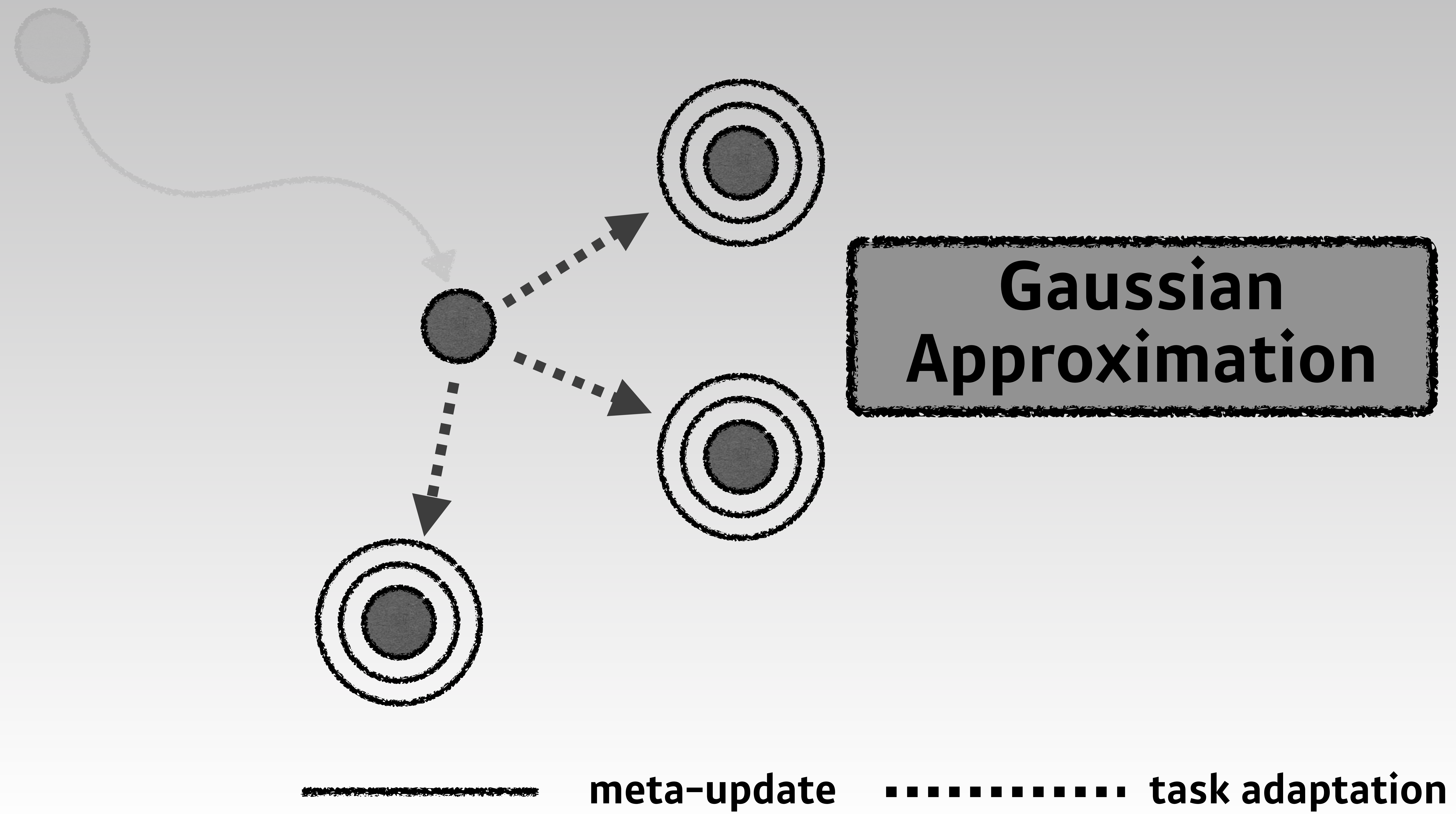
Uncertainty

- ☑ Robust to overfitting
- ☑ Safe/efficient exploration
- ☑ Active learning

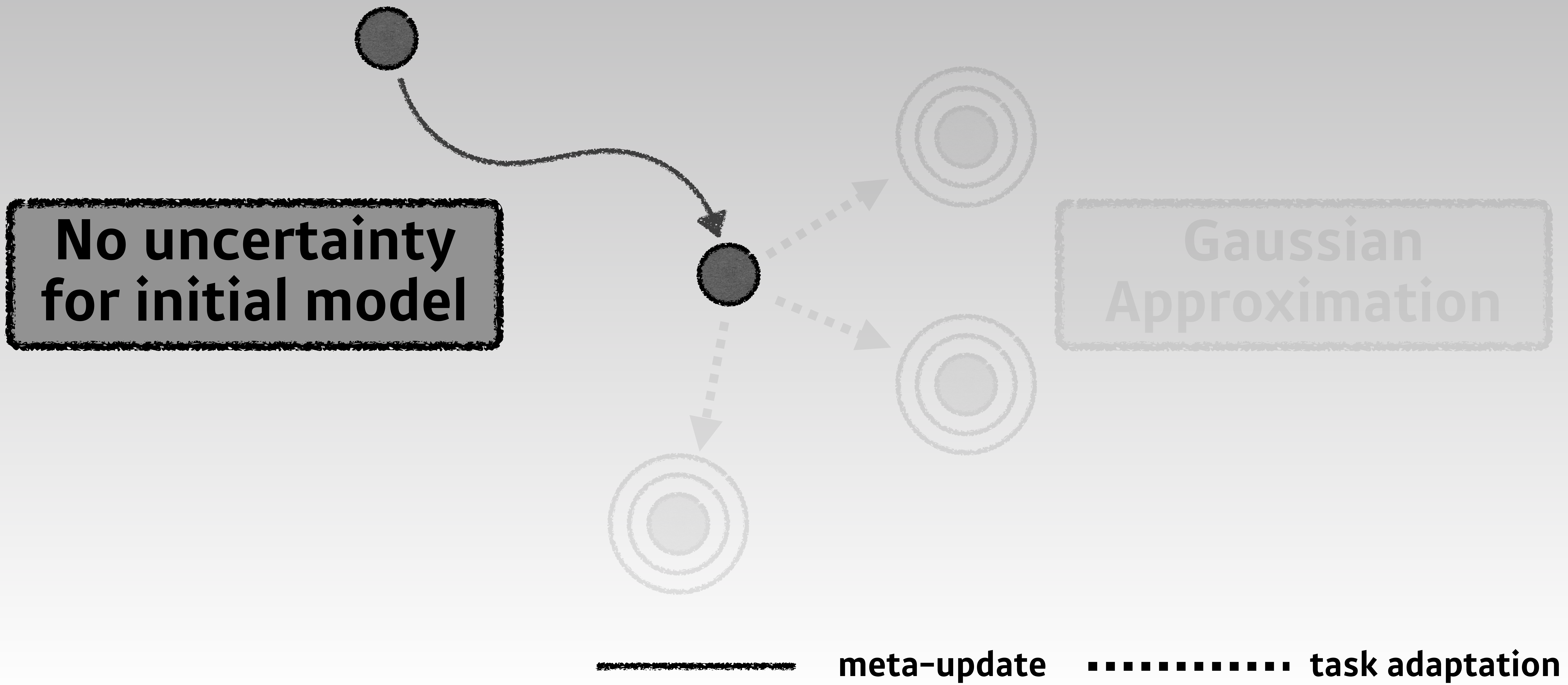
Lightweight Laplace Approximation for Meta-Adaptation (LLAMA)



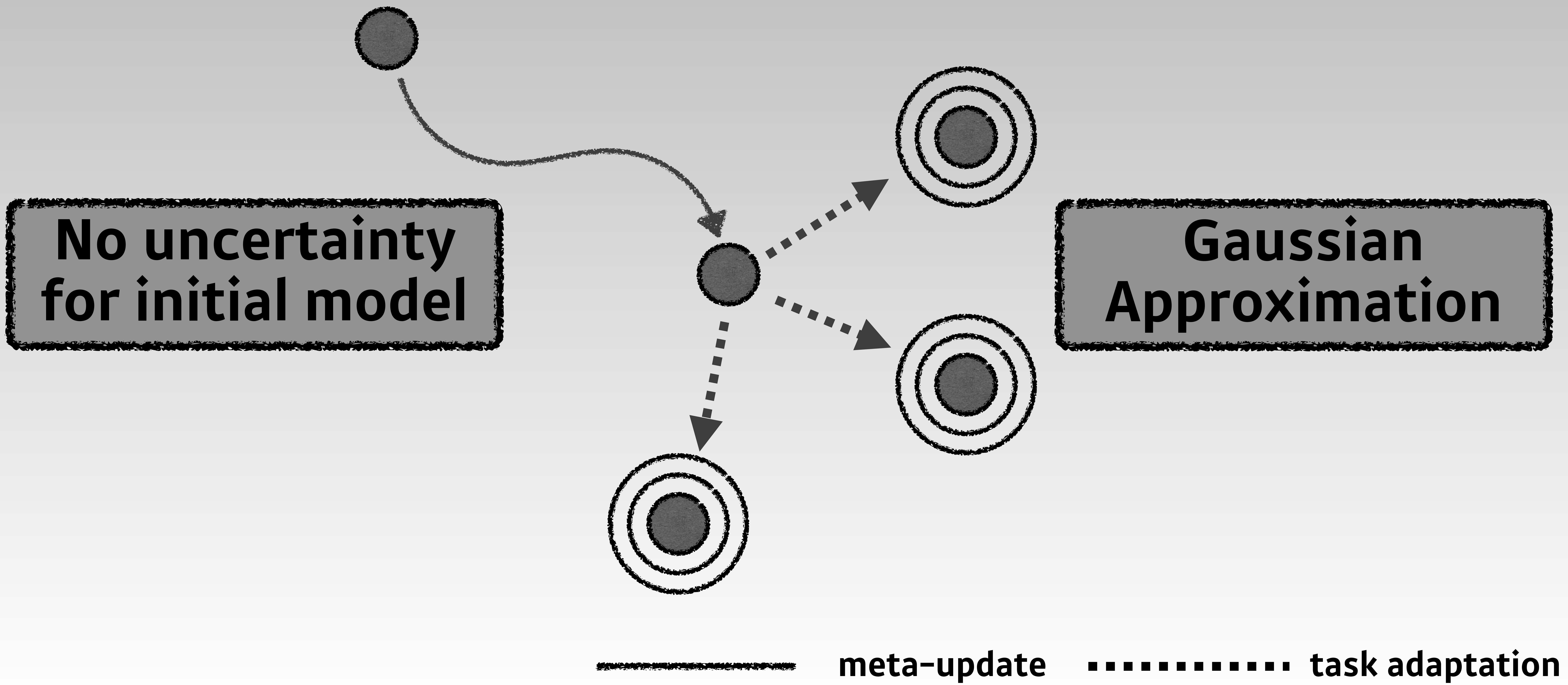
Lightweight Laplace Approximation for Meta-Adaptation (LLAMA)



Lightweight Laplace Approximation for Meta-Adaptation (LLAMA)



Lightweight Laplace Approximation for Meta-Adaptation (LLAMA)



Bayesian Model-Agnostic Meta-Learning (BMAML)

Bayesian Model-Agnostic Meta-Learning (BMAML)

For each task in a batch:

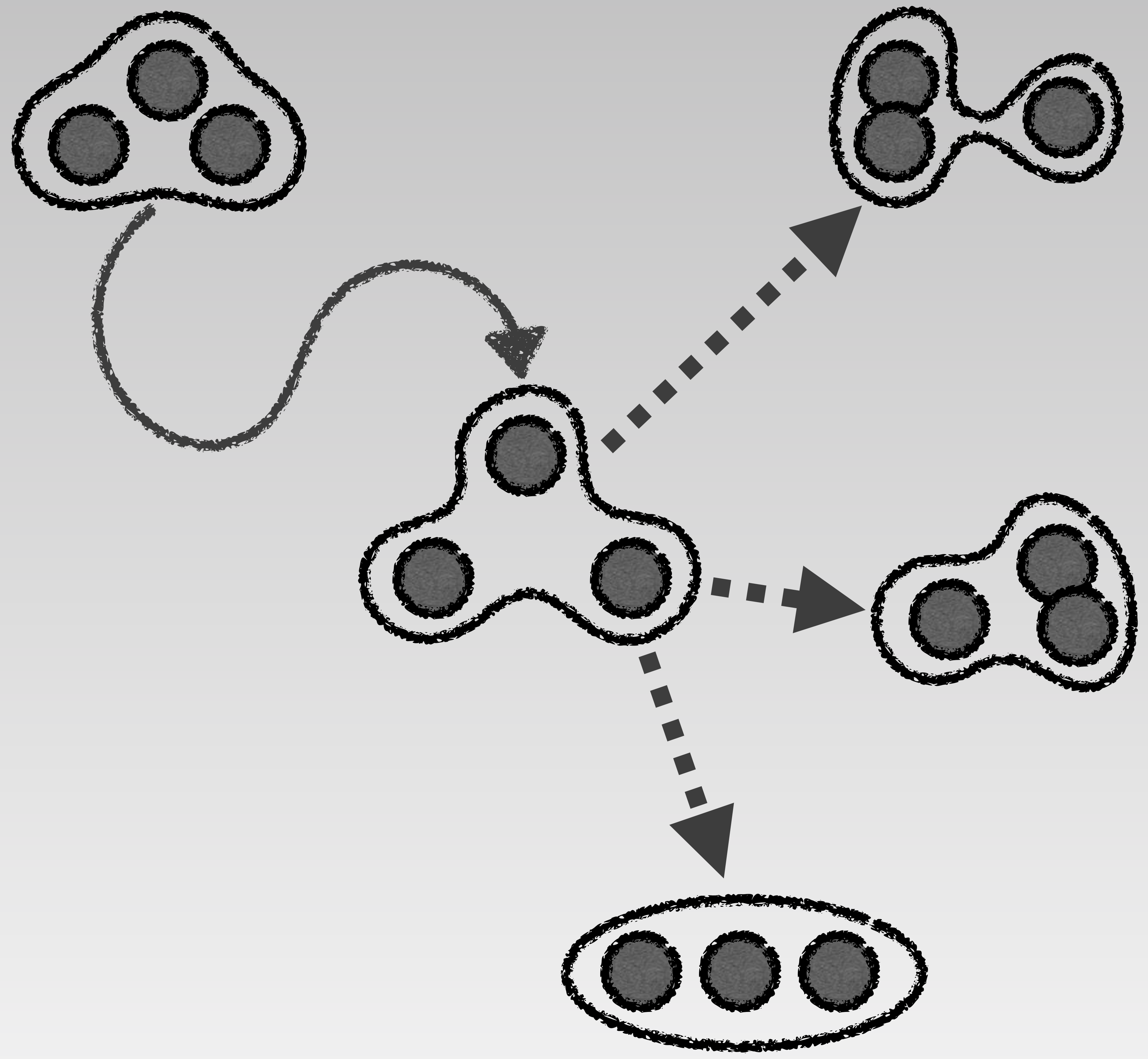
~~Task Model~~ ← ~~Task adaptation~~

Task Distribution ← Bayesian fast adaptation

~~Initial Model~~ ← ~~Meta-update~~

Initial distribution ← Bayesian meta-update

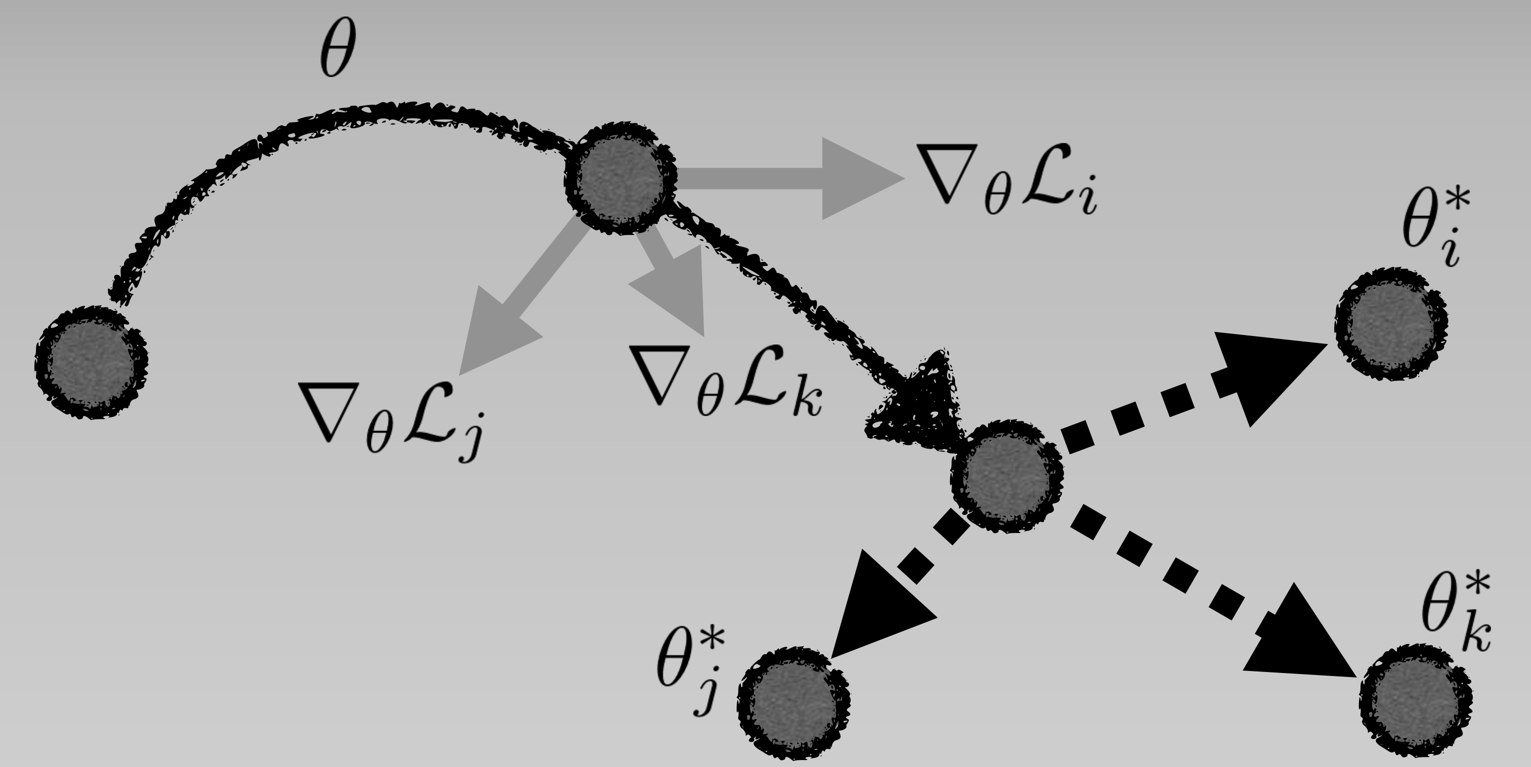
BMAML



complex multimodal

meta-update   task adaptation

Bayesian Fast Adaptation (BFA)



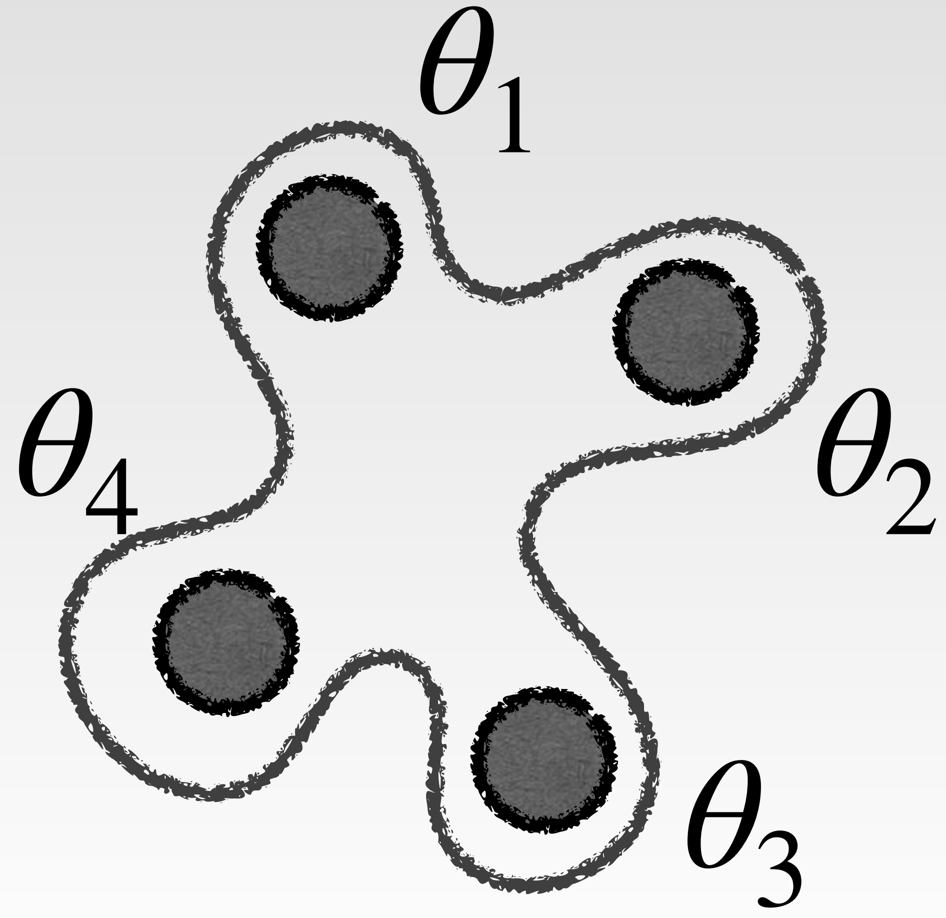
Model-Agnostic Meta-Learning (MAML)

"**gradient-based** meta-learning framework"

+

Stein Variational Gradient Descent (SVGD)

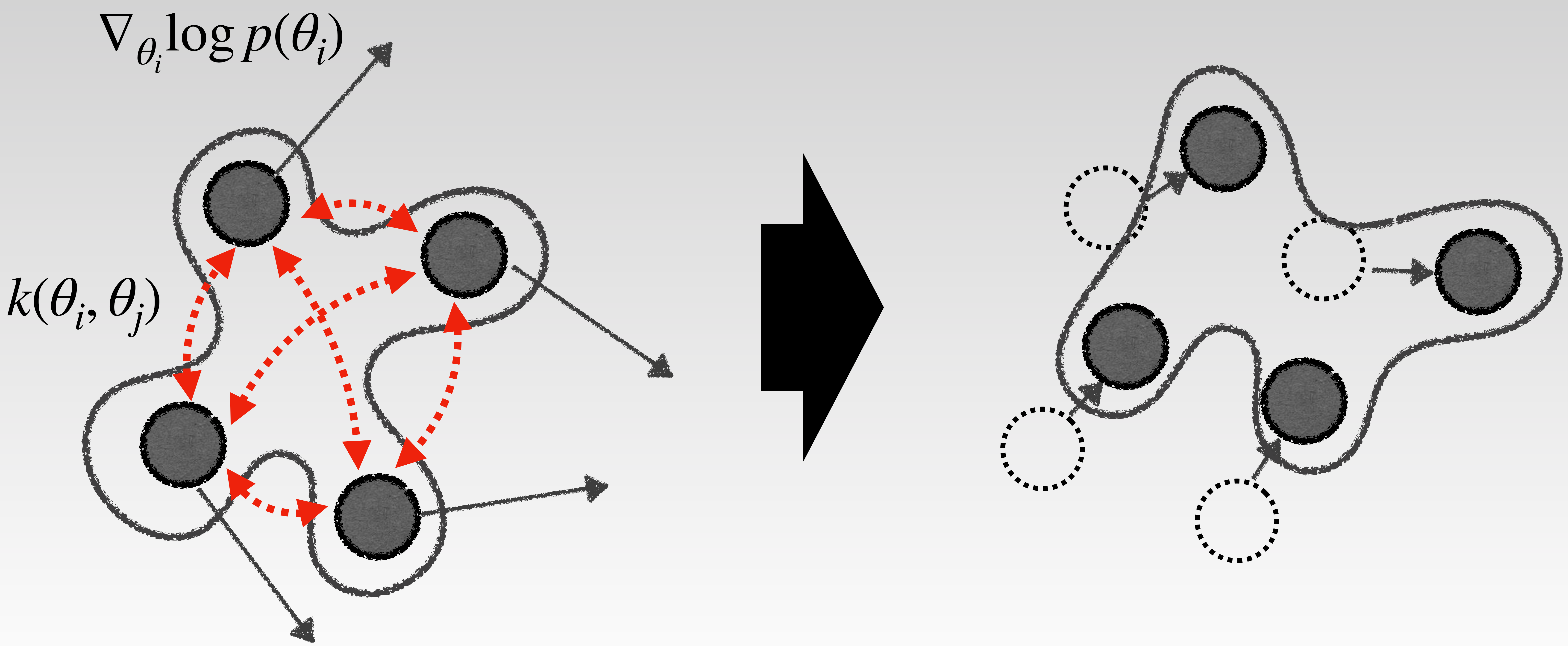
"**particle-based** posterior approximation"



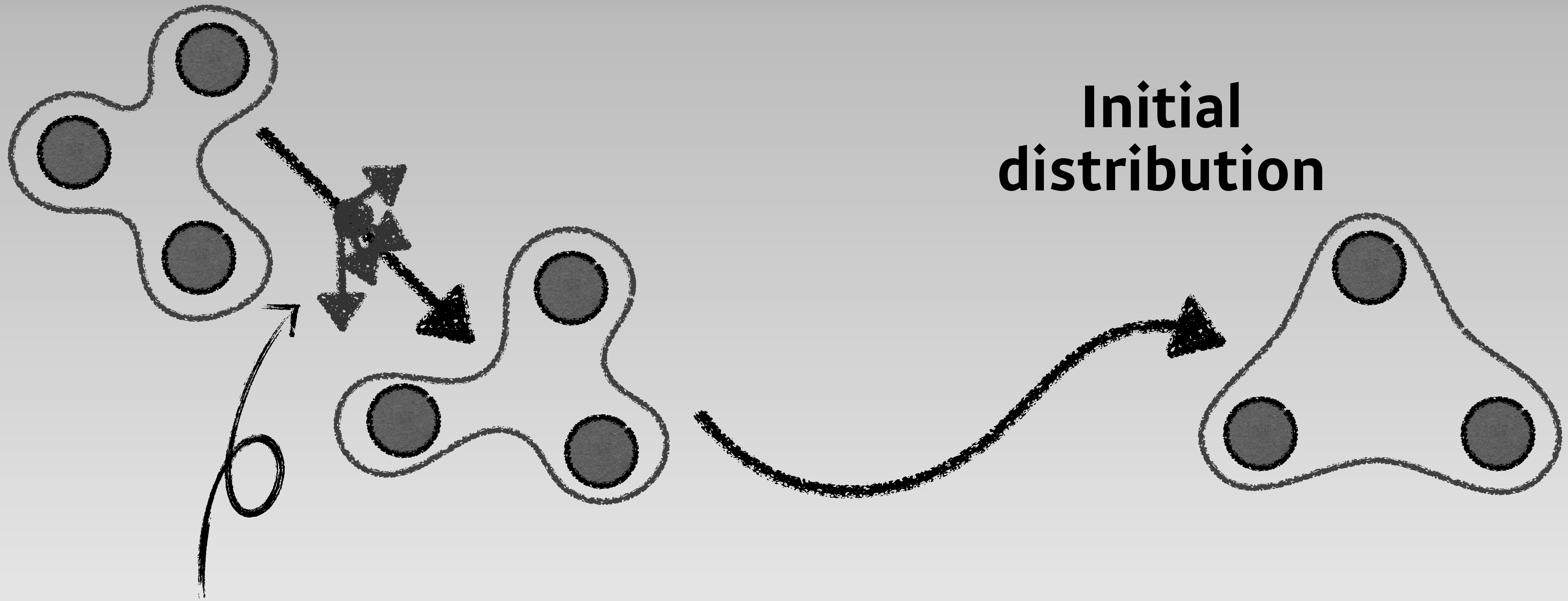
Stein Variational Gradient Descent (SVGD)

“particle-based posterior approximation”

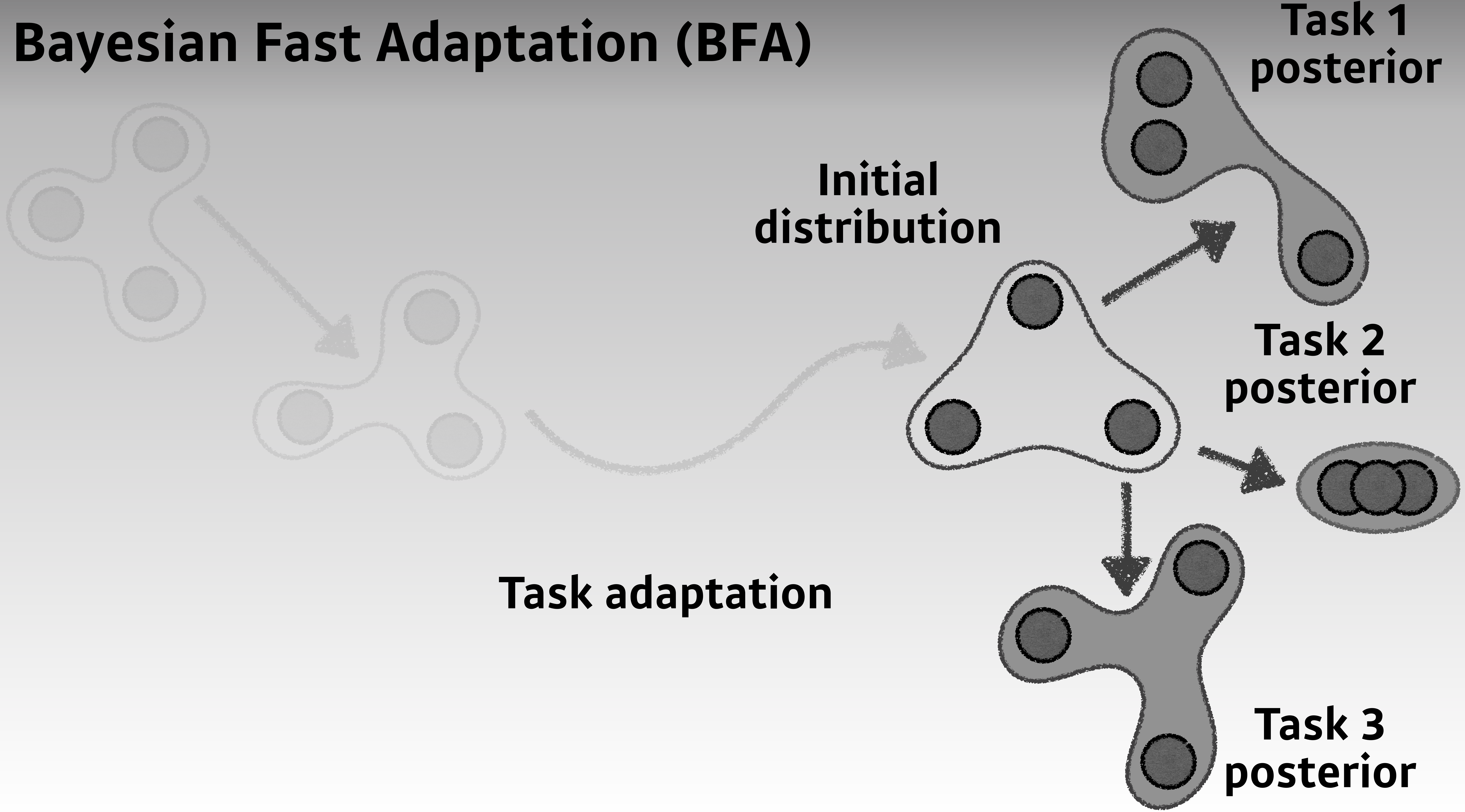
“backprop to initial model through deterministic SVGD particles”



Bayesian Fast Adaptation (BFA)

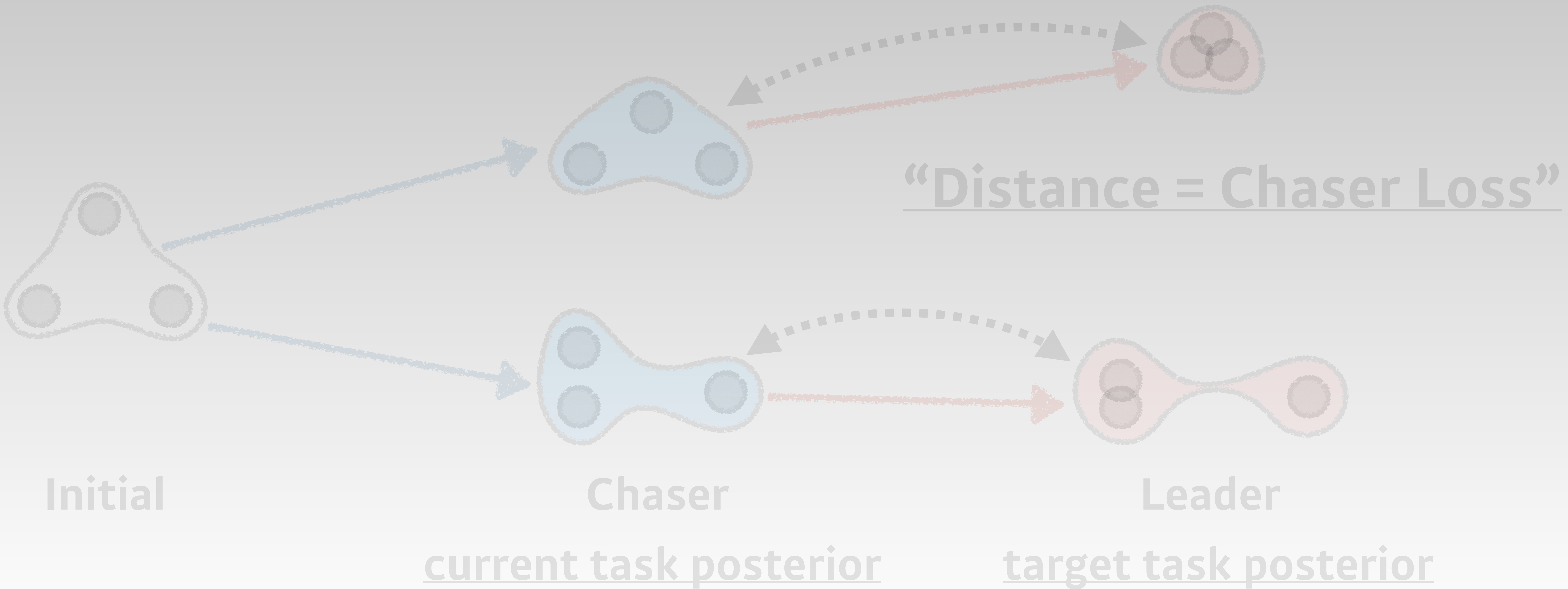


Bayesian Fast Adaptation (BFA)

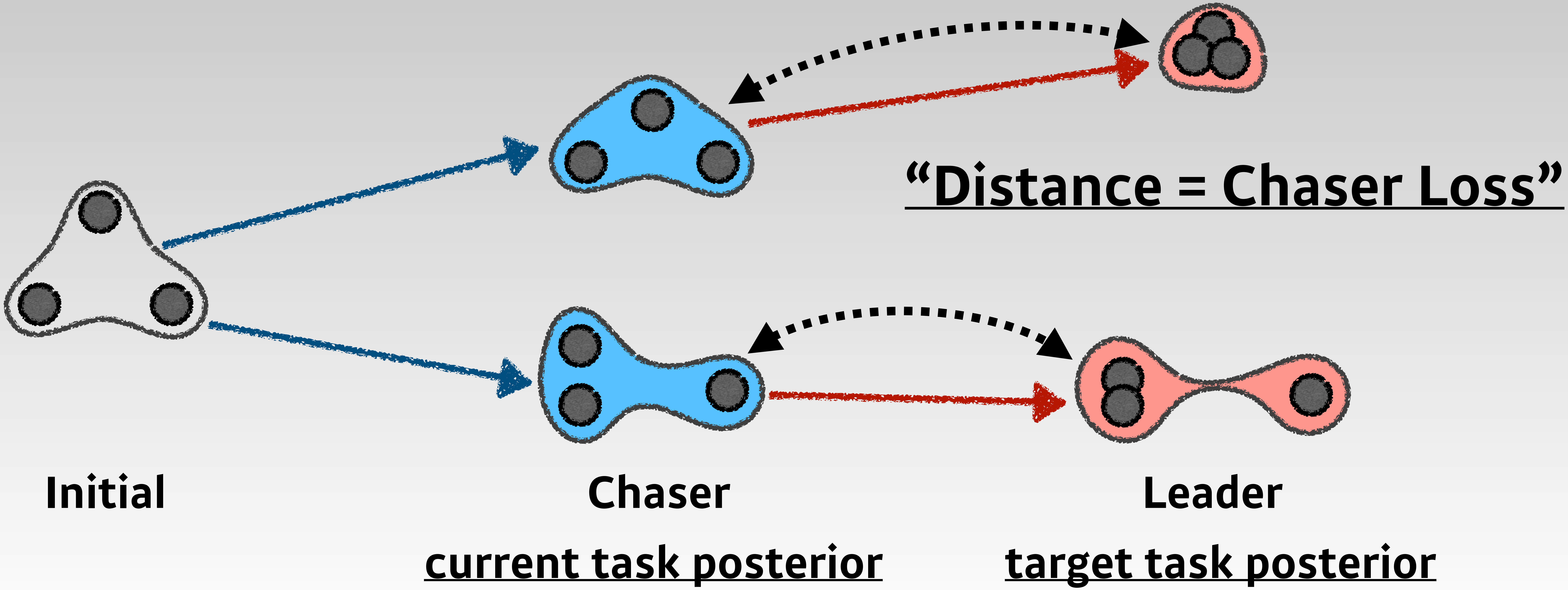


Bayesian Meta-Update with Chaser Loss

"extend uncertainty-awareness to meta-update"

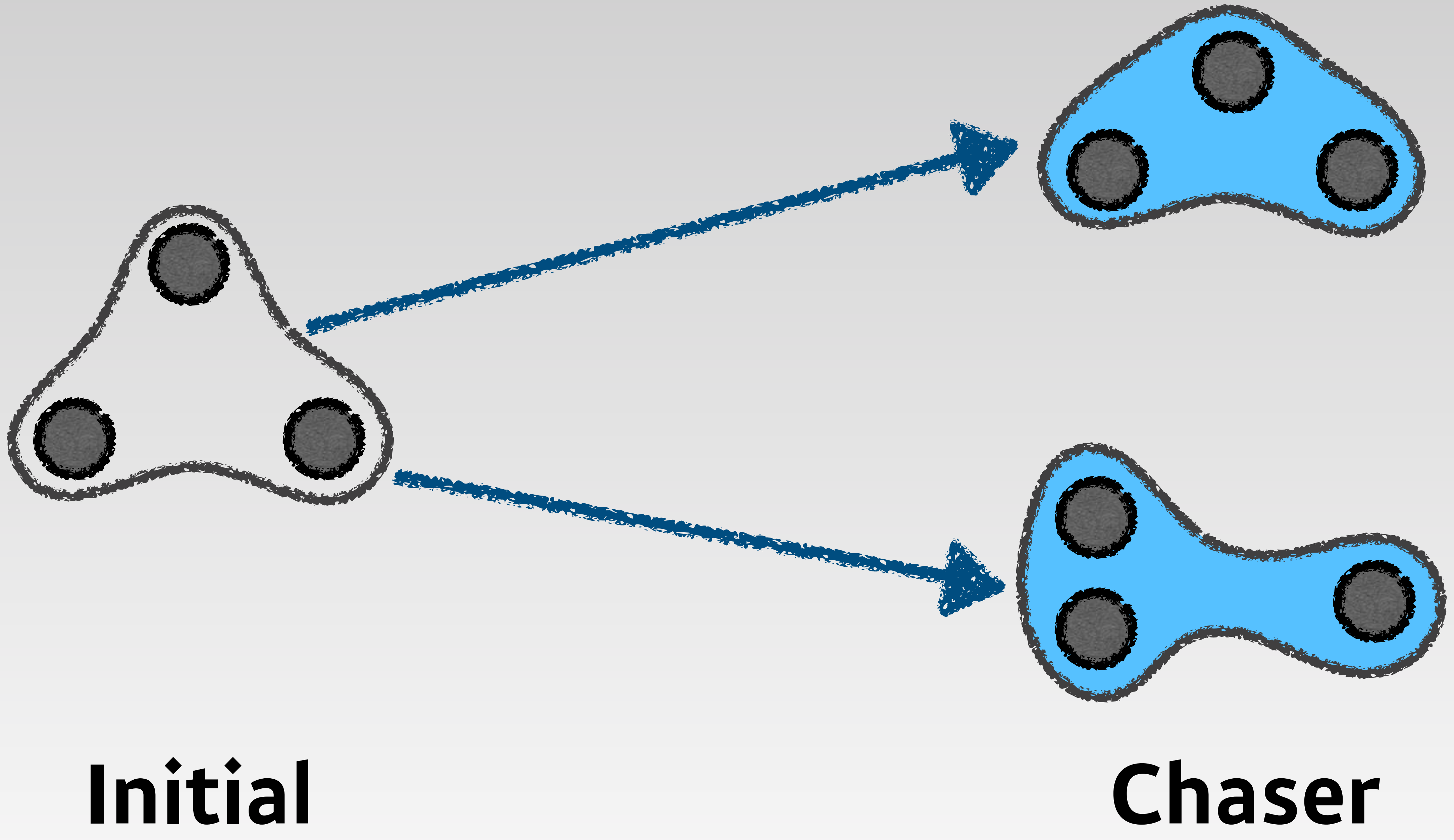


Bayesian Meta-Update with Chaser Loss



Bayesian Meta-Update with Chaser Loss

For each task,
- Compute <u>CHASER PARTICLES</u>

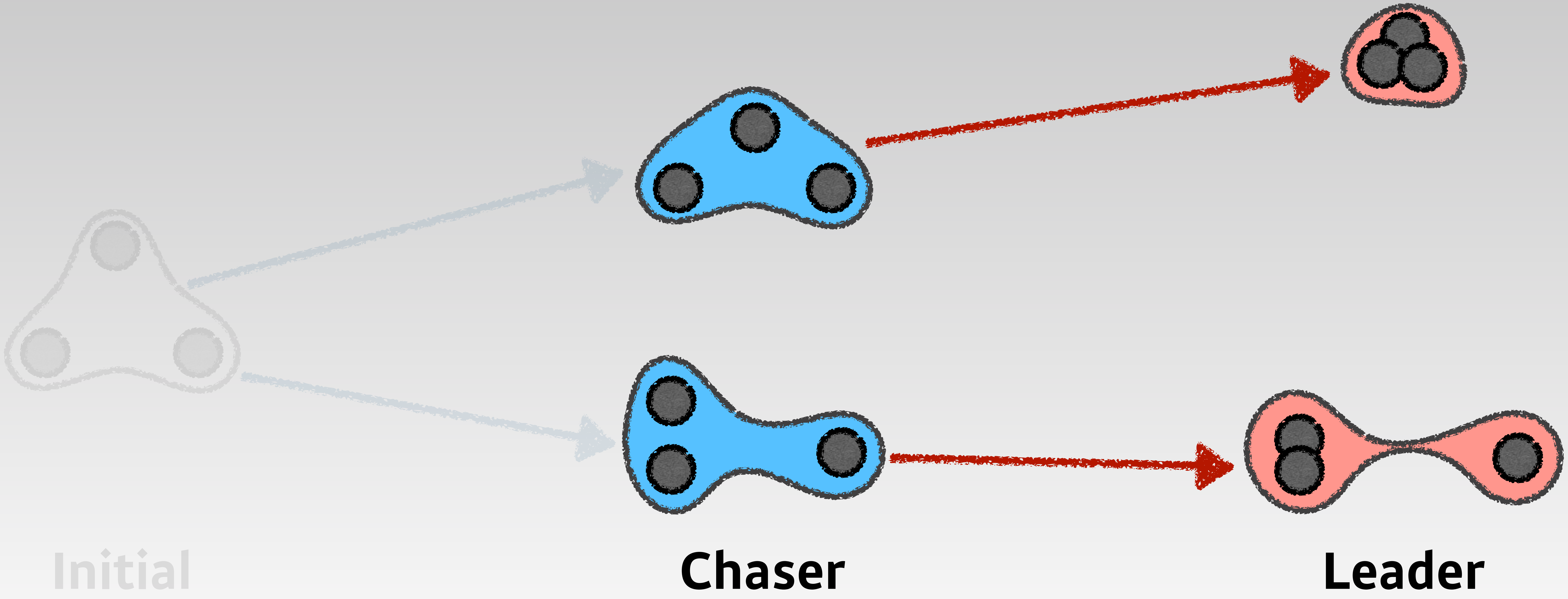


$$\text{chaser } \Theta_\tau^n(\Theta_0) = \mathrm{SVGD}_n(\Theta_0; \mathcal{D}_\tau^{\mathrm{trn}}, \alpha)$$

Bayesian Meta-Update with Chaser Loss

For each task,

- Compute <u>CHASER PARTICLES</u>
- Compute <u>LEADER PARTICLES</u>

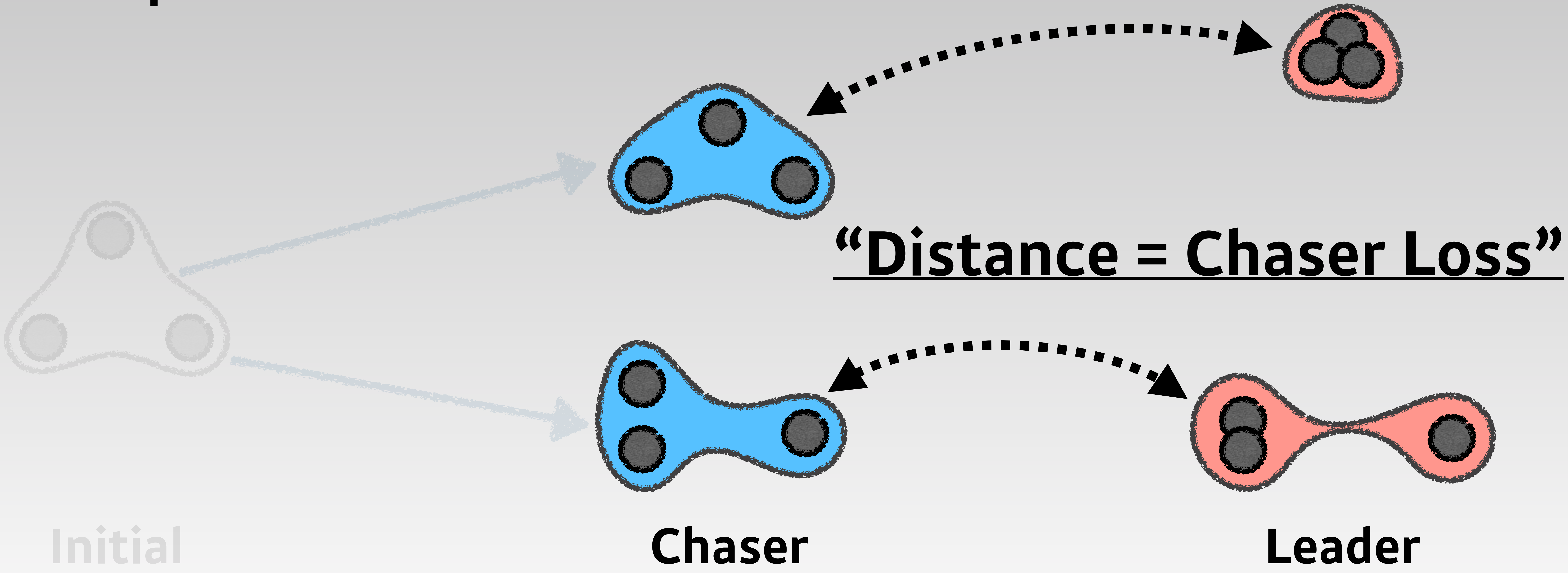


$$\text{leader } \Theta_\tau^{n+s}(\Theta_0) = \text{SVGD}_s(\Theta_\tau^n(\Theta_0); \mathcal{D}_\tau^{\text{trn}} \cup \mathcal{D}_\tau^{\text{val}}, \alpha)$$

Bayesian Meta-Update with Chaser Loss

For each task,

- Compute <u>CHASER PARTICLES</u>
- Compute <u>LEADER PARTICLES</u>
- **Compute <u>CHASER LOSS</u>**



$$\mathcal{L}_{\text{BMAML}}(\Theta_0) = \sum_{\tau \in \mathcal{T}_t} d_s(\Theta_\tau^n \parallel \Theta_\tau^{n+s}) = \sum_{\tau \in \mathcal{T}_t} \sum_{m=1}^{M} \|\theta_\tau^{n,m} - \theta_\tau^{n+s,m}\|_2^2$$

Experiments

Regression

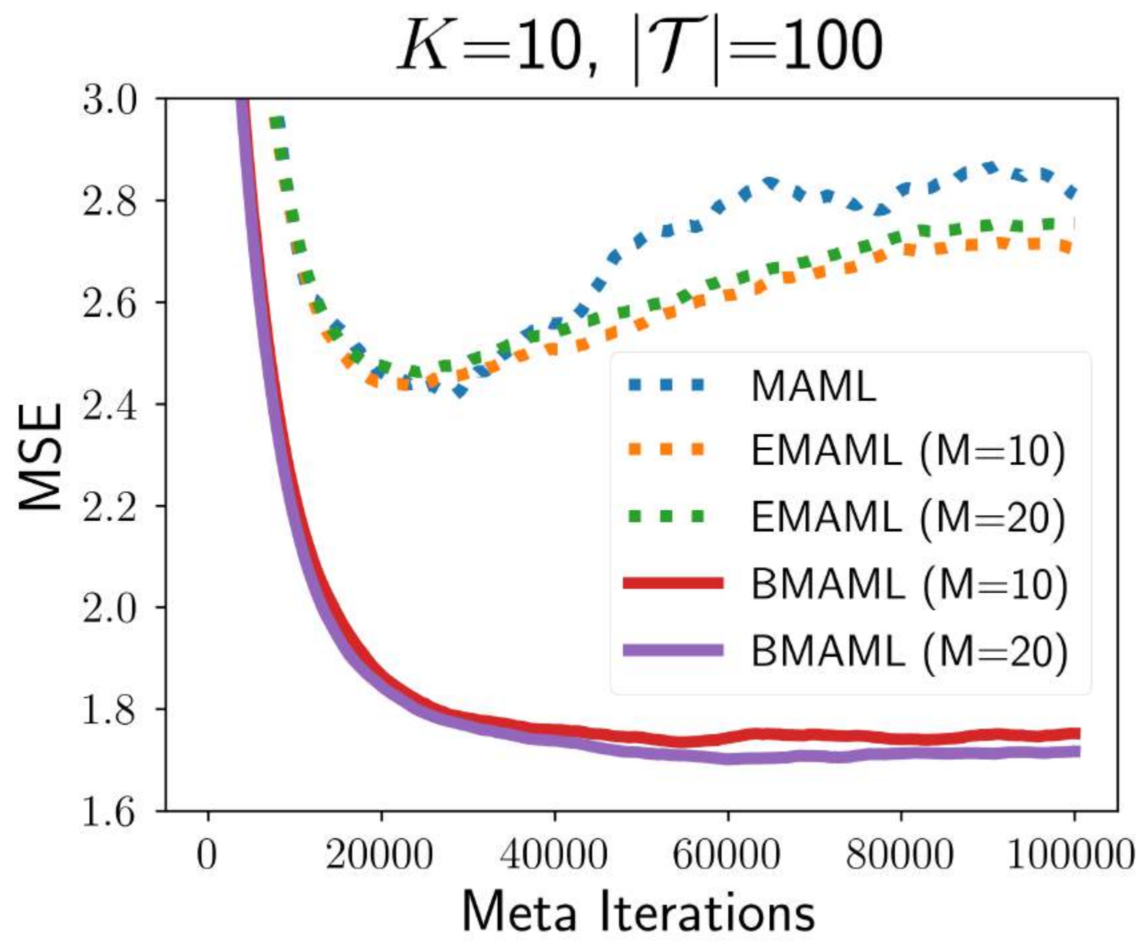
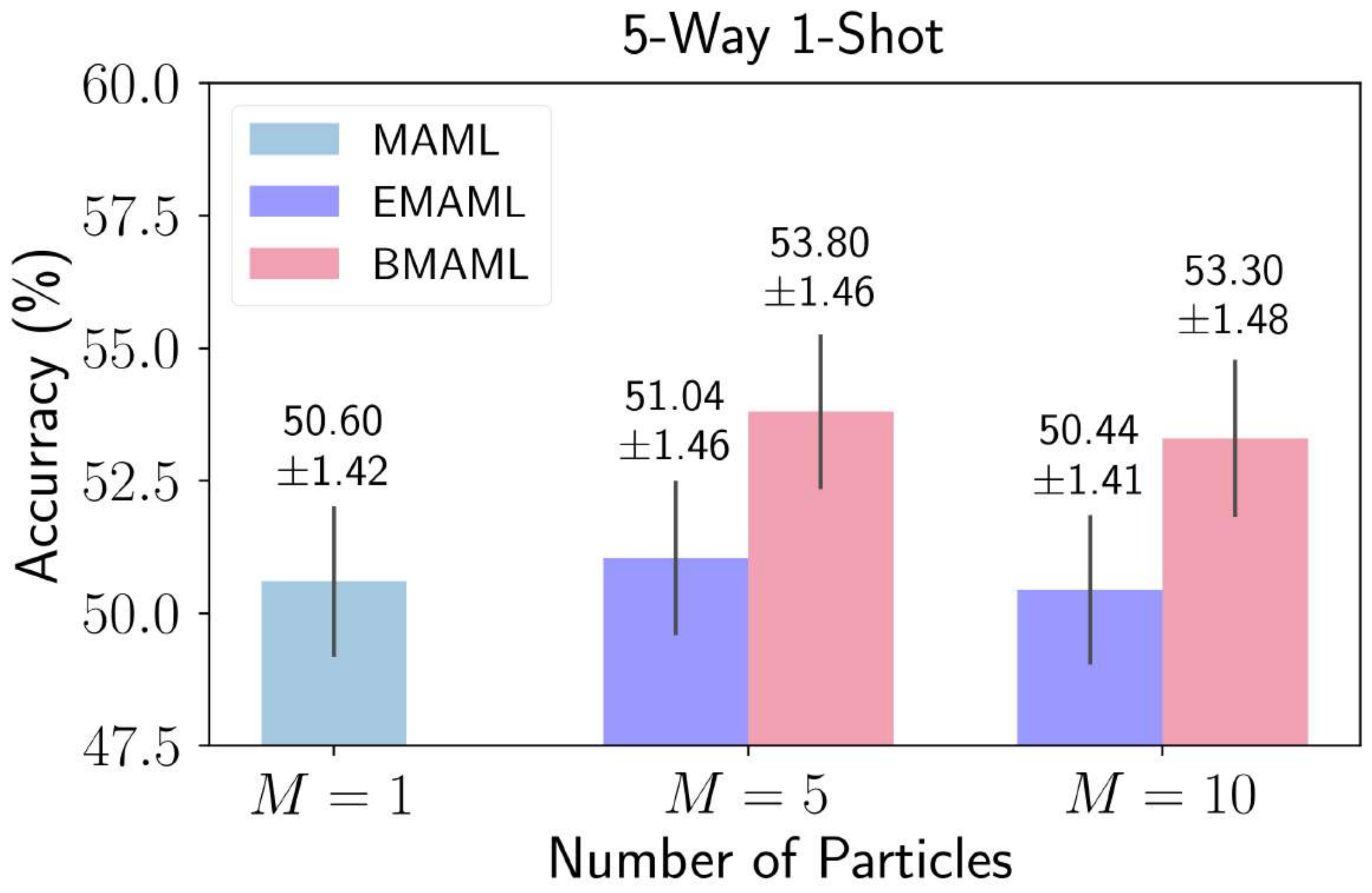
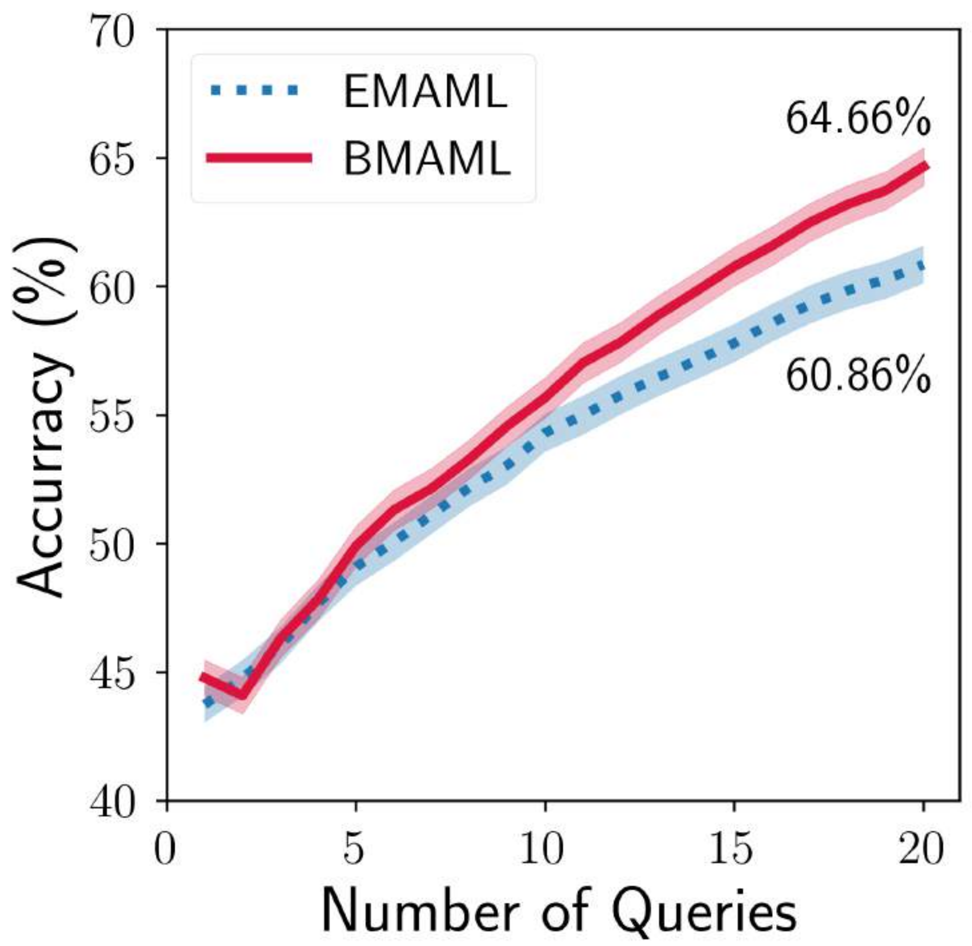


Image Classification

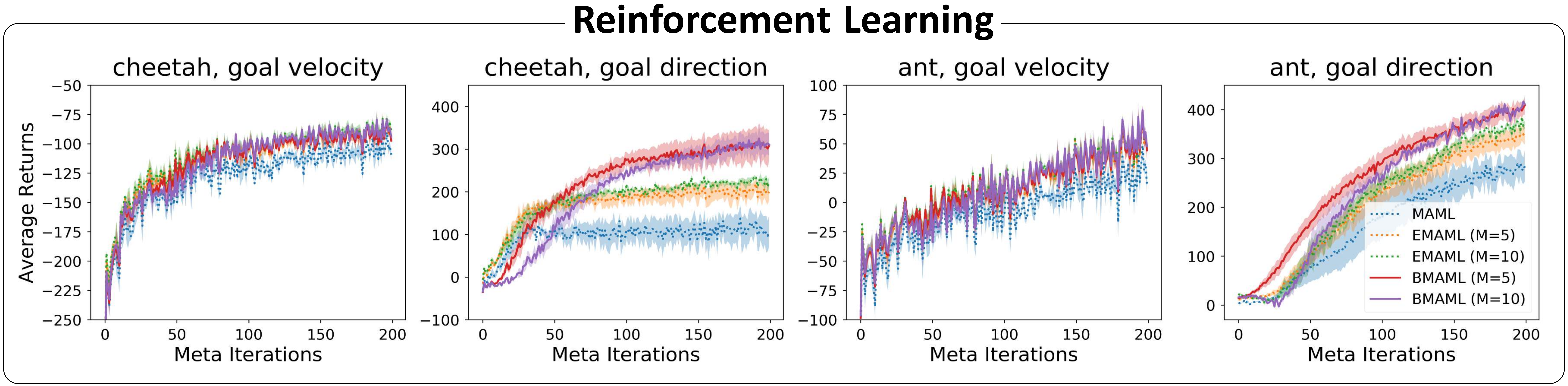


Active Learning



- prevent overfitting with better performance
- evaluate effectiveness of measured uncertainty

Experiments



- better policy exploration

See you at Poster “AB #15”
(room 210 & 230)