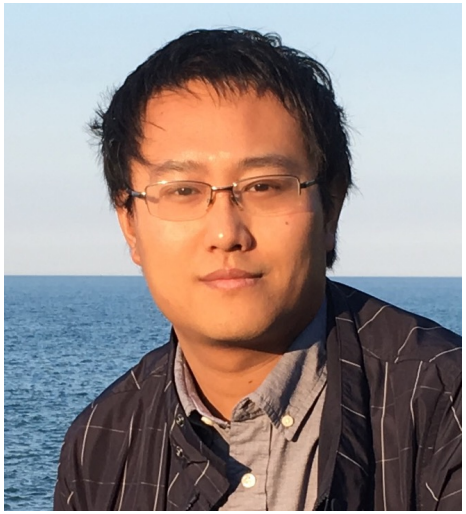
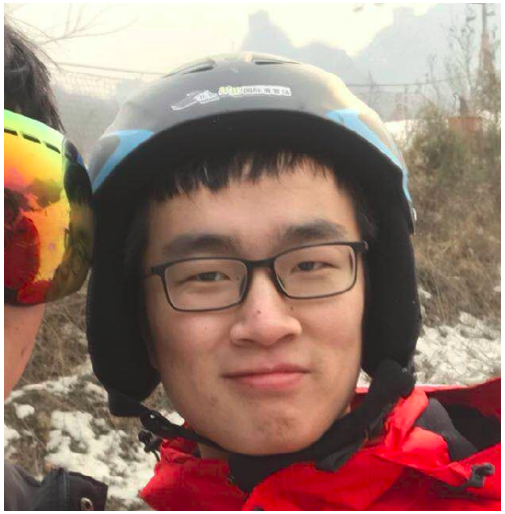
Imputing Missing Events in Continuous-Time Event Streams



Hongyuan Mei[1]



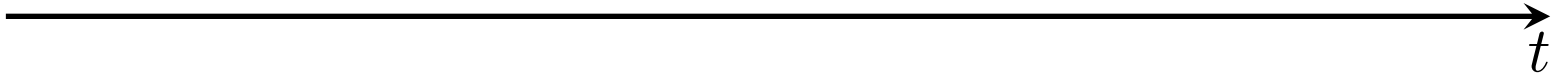
Guanghui Qin[2]

Jason Eisner[1]

[1]Department of Computer Science, Johns Hopkins University, USA
[2]Department of Physics, Peking University, China

Event Streams

Medical



Event Streams

Shopping

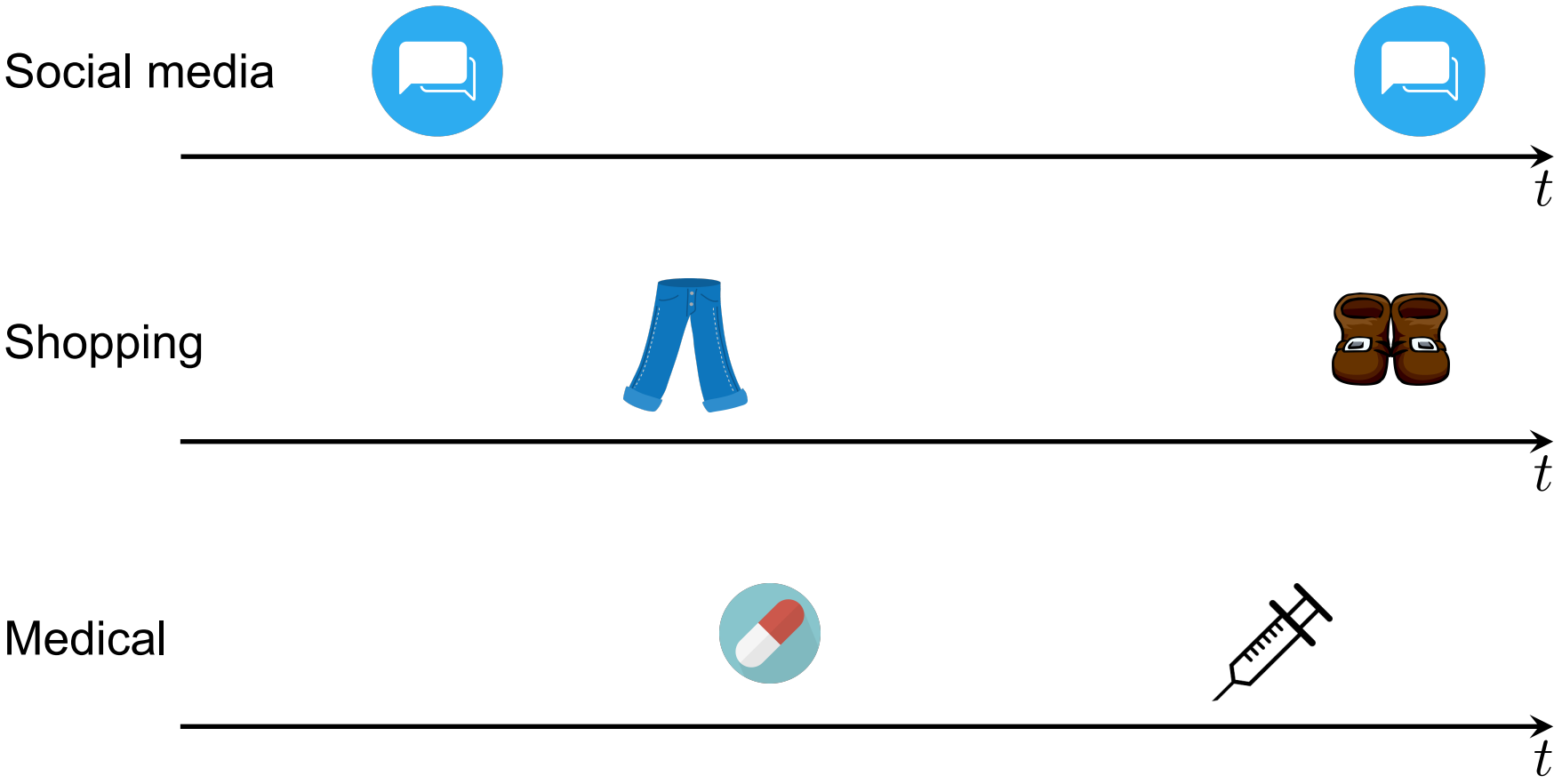
t

Medical

t

Event Streams

Event Streams



Event Streams

Let’s Play a Battle Game

Let’s Play a Battle Game

- Incomplete data

Let’s Play a Battle Game

- Incomplete data
- Opponent tanks come

0

Let's Play a Battle Game

- Incomplete data
- Opponent tanks come

0 t_1 t_2

Let's Play a Battle Game

- Incomplete data
- Opponent tanks come
- Maybe a new factory was built to produce them?

0 t_1 t_2

Let's Play a Battle Game

- Incomplete data
- Opponent tanks come
- Maybe a new factory was built to produce them?
- When was it built?

0 t_1 t_2

Let's Play a Battle Game

- Incomplete data
- Opponent tanks come
- Maybe a new factory was built to produce them?
- When was it built?

0 ? t_1 t_2

Let's Play a Battle Game

- Incomplete data
- Opponent tanks come
- Maybe a new factory was built to produce them?
- When was it built?

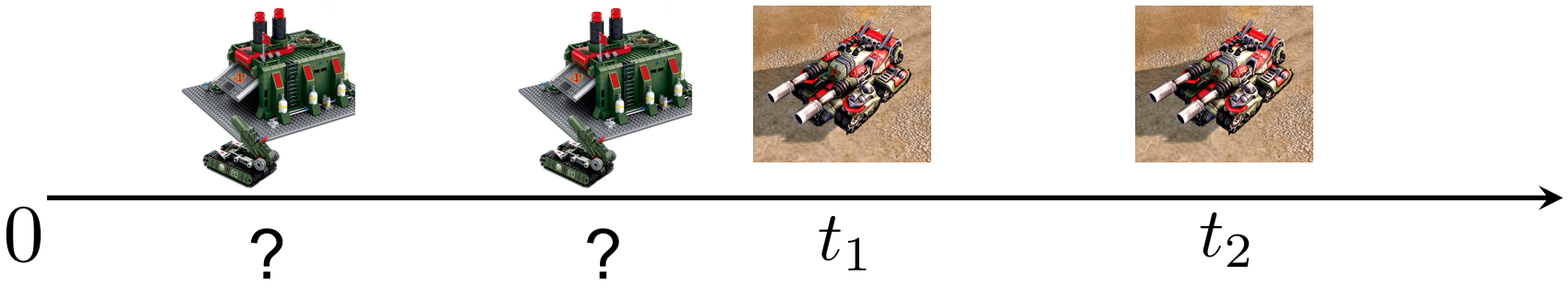
0 ? t_1 t_2

Let's Play a Battle Game

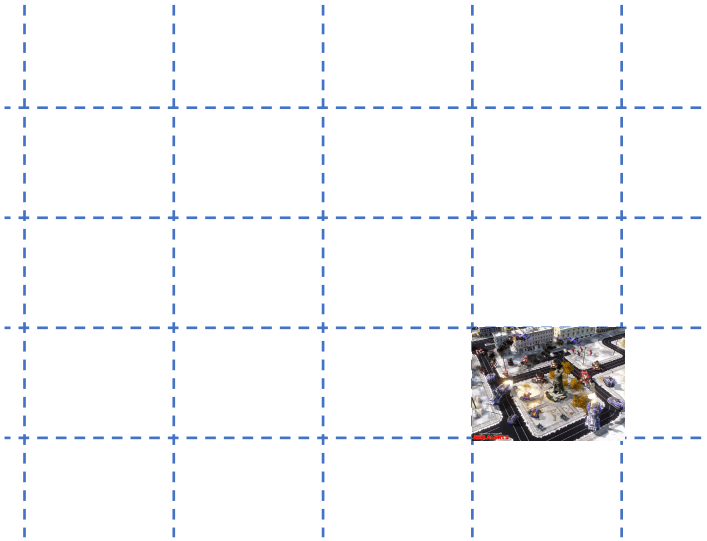
- Incomplete data
- Opponent tanks come
- Maybe a new factory was built to produce them?
- When was it built?

Let's Play a Battle Game

- Incomplete data
- Opponent tanks come
- Maybe a new factory was built to produce them?
- When was it built?
- Where was it built?



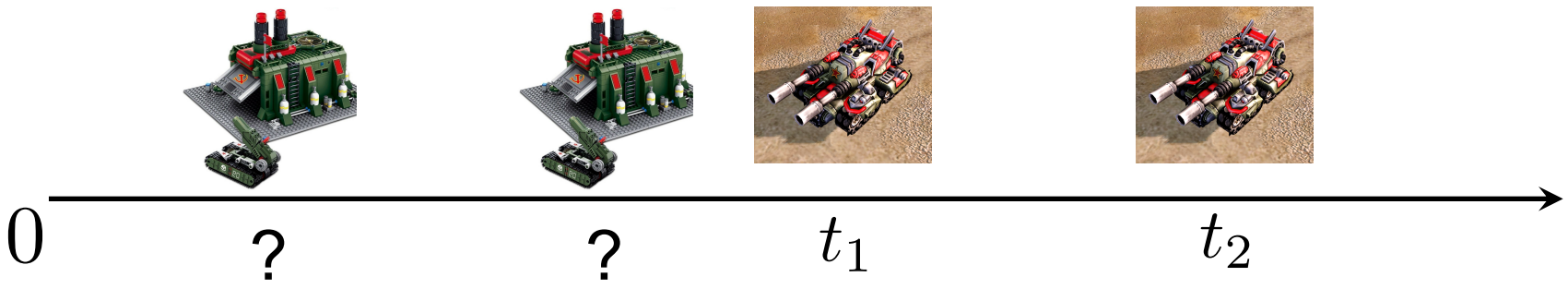
Let's Play a Battle Game



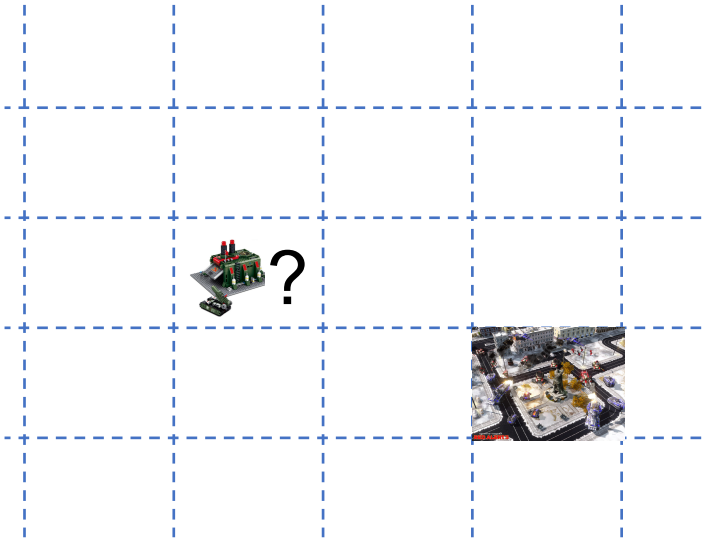
- Incomplete data
- Opponent tanks come
- Maybe a new factory was built to produce them?
- When was it built?
- Where was it built?

Let's Play a Battle Game

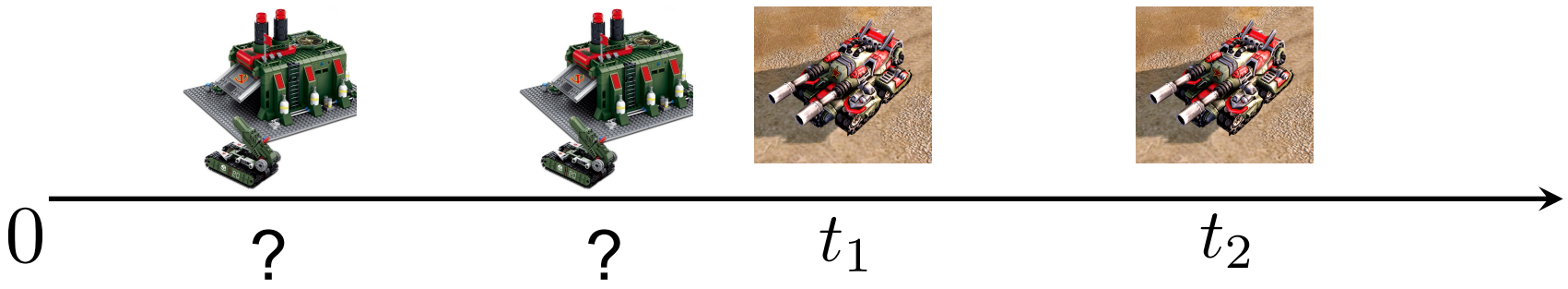
- Incomplete data
- Opponent tanks come
- Maybe a new factory was built to produce them?
- When was it built?
- Where was it built?



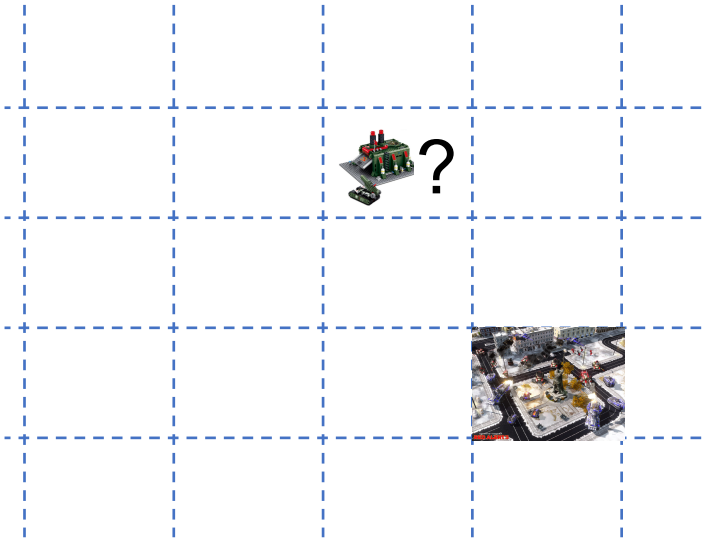
Let's Play a Battle Game



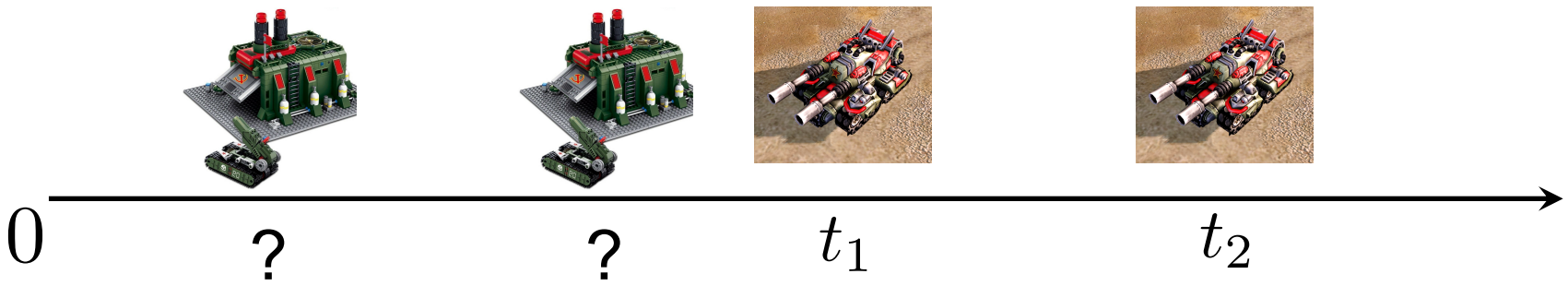
- Incomplete data
- Opponent tanks come
- Maybe a new factory was built to produce them?
- When was it built?
- Where was it built?



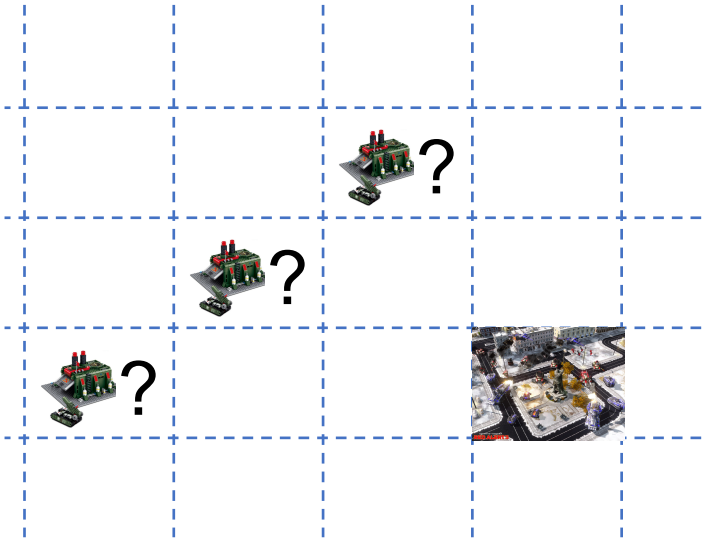
Let's Play a Battle Game

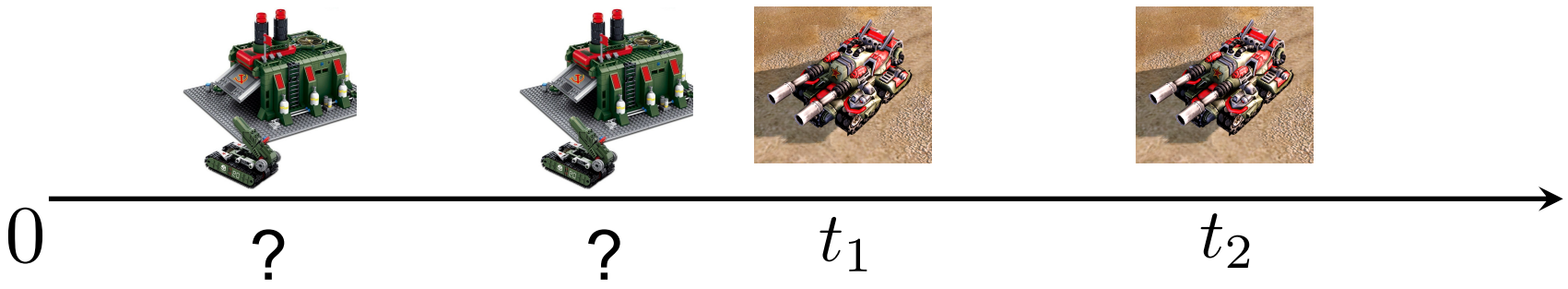


- Incomplete data
- Opponent tanks come
- Maybe a new factory was built to produce them?
- When was it built?
- Where was it built?

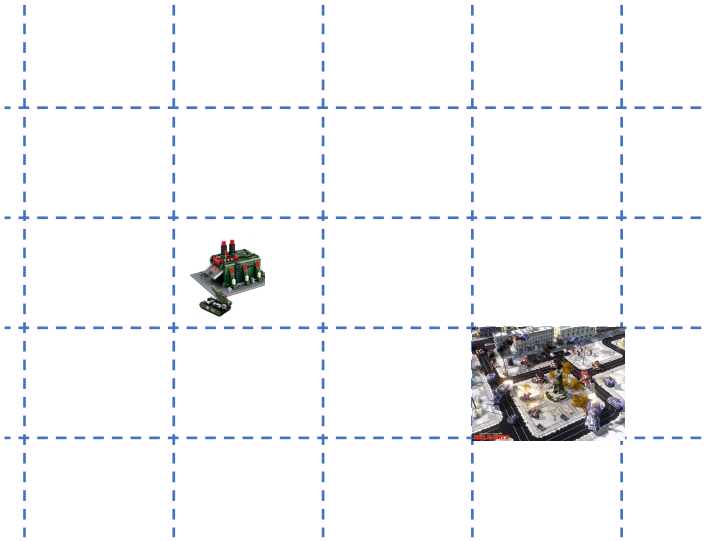


Why to Impute?





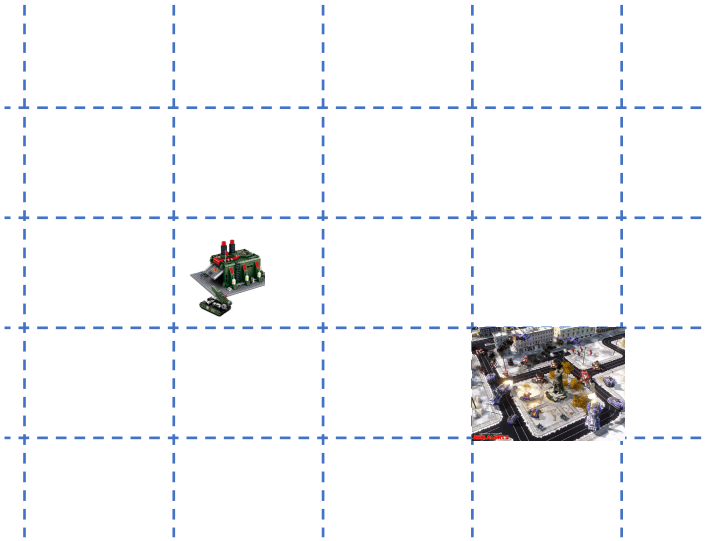
Why to Impute?



- If there is a factory over there
- We should go after it!

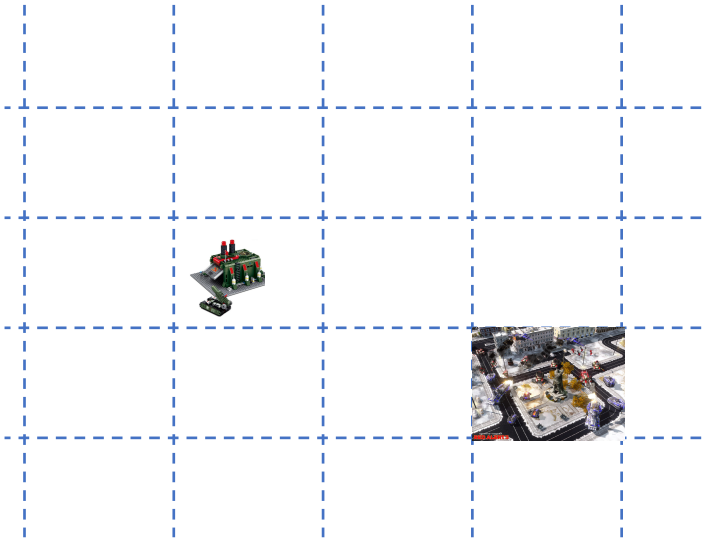
0 t_1 t_2

To Impute the Missing Events…



0 t_1 t_2

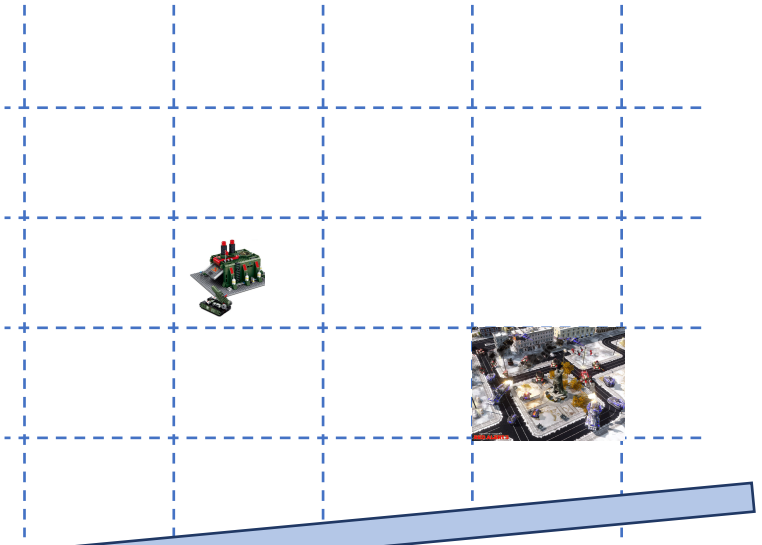
To Impute the Missing Events…



- We need to know the distribution of the complete sequences

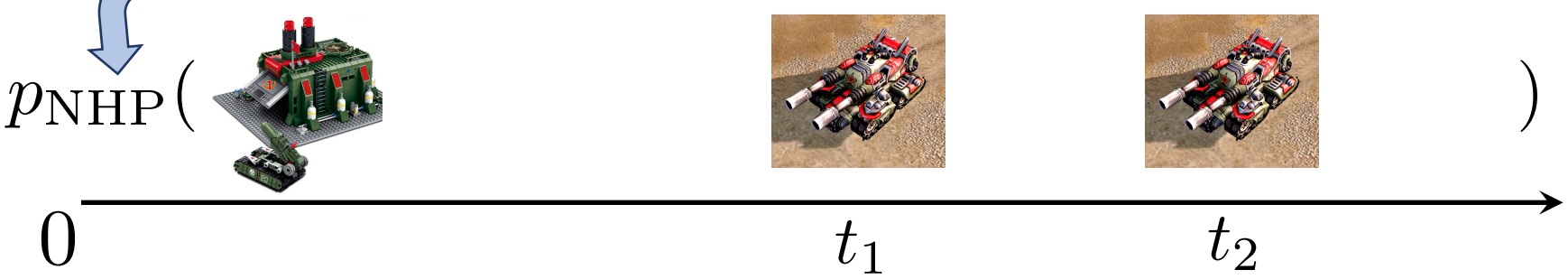
0 t_1 t_2

To Impute the Missing Events…

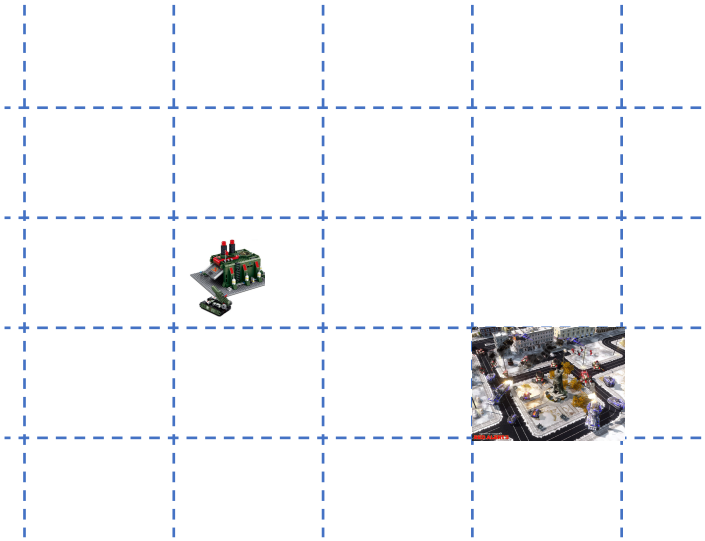


- We need to know the distribution of the complete sequences

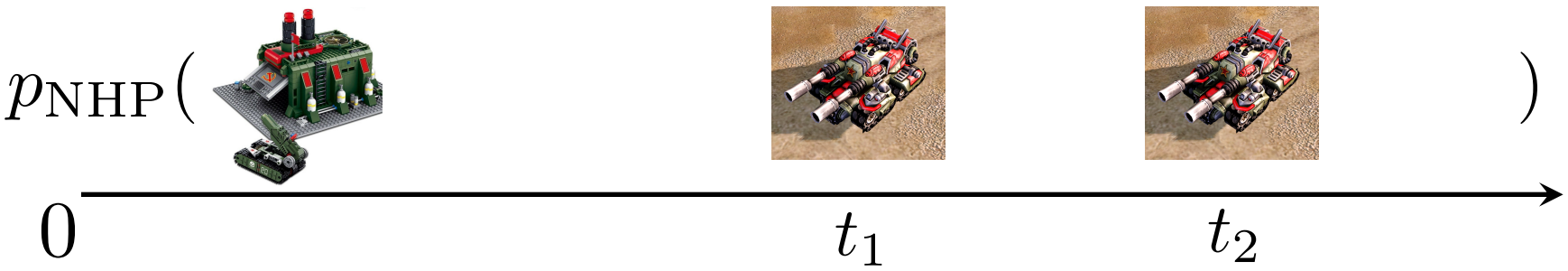
Neural Hawkes process (NHP)
(Mei & Eisner 2017 – NeurIPS)



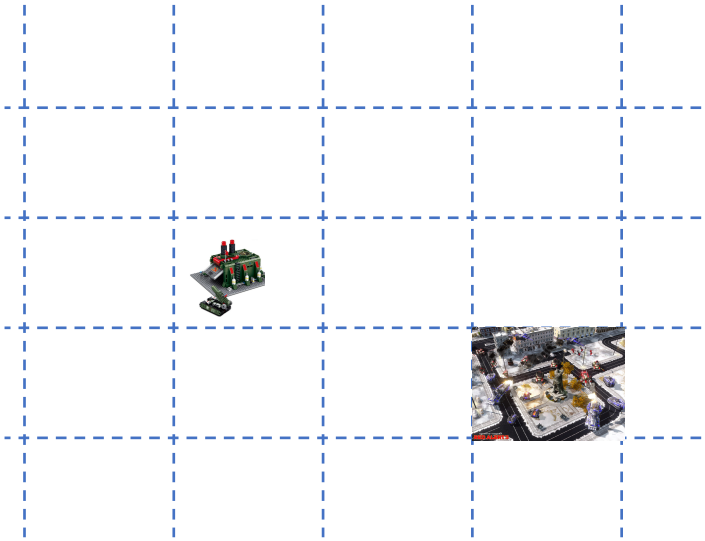
To Impute the Missing Events…



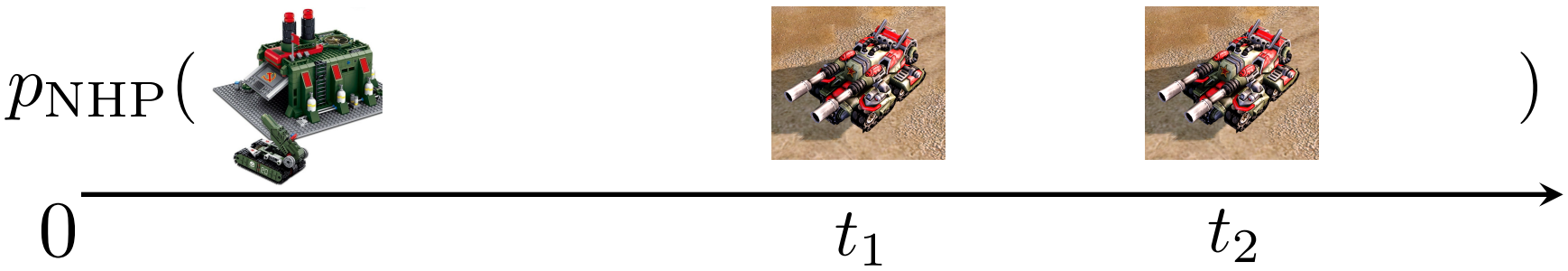
- Where do we get the complete data to train its parameters?



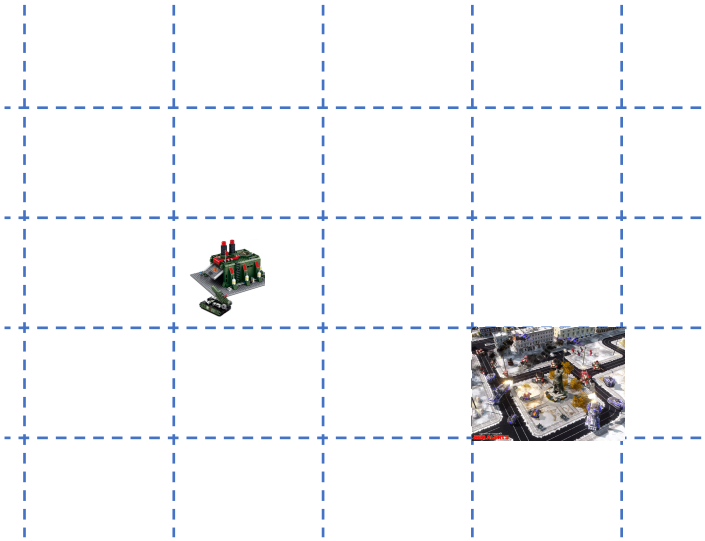
To Impute the Missing Events…



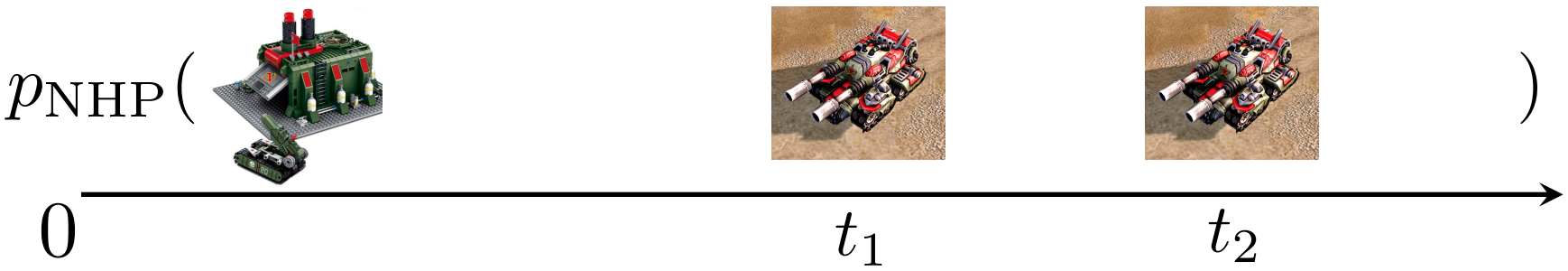
- Where do we get the complete data to train its parameters?
- Game logs shown to us after the game (e.g. replay)



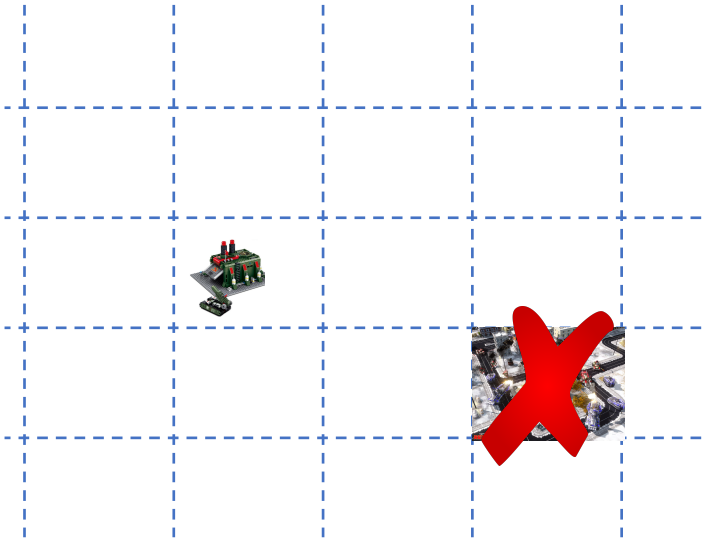
To Impute the Missing Events…



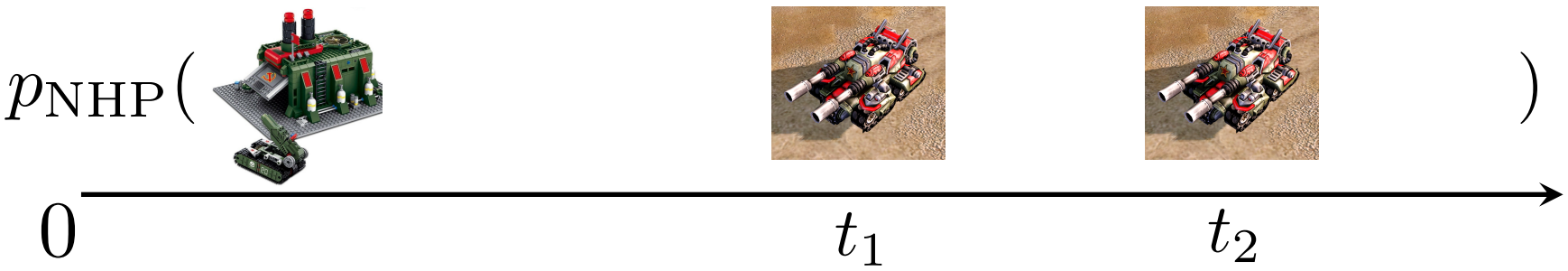
- We also need to know the missingness mechanism
- Don't propose anything which we know won't be missing



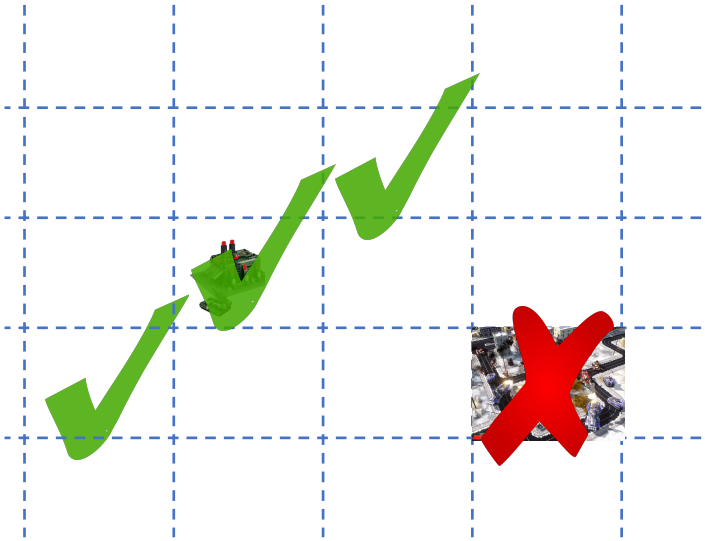
To Impute the Missing Events…



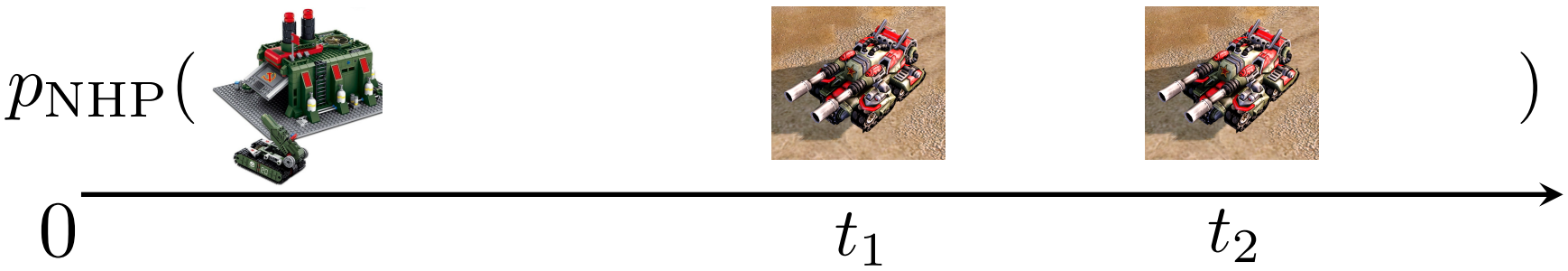
- We also need to know the missingness mechanism
- Don't propose anything which we know won't be missing
- In-view events won't be missing



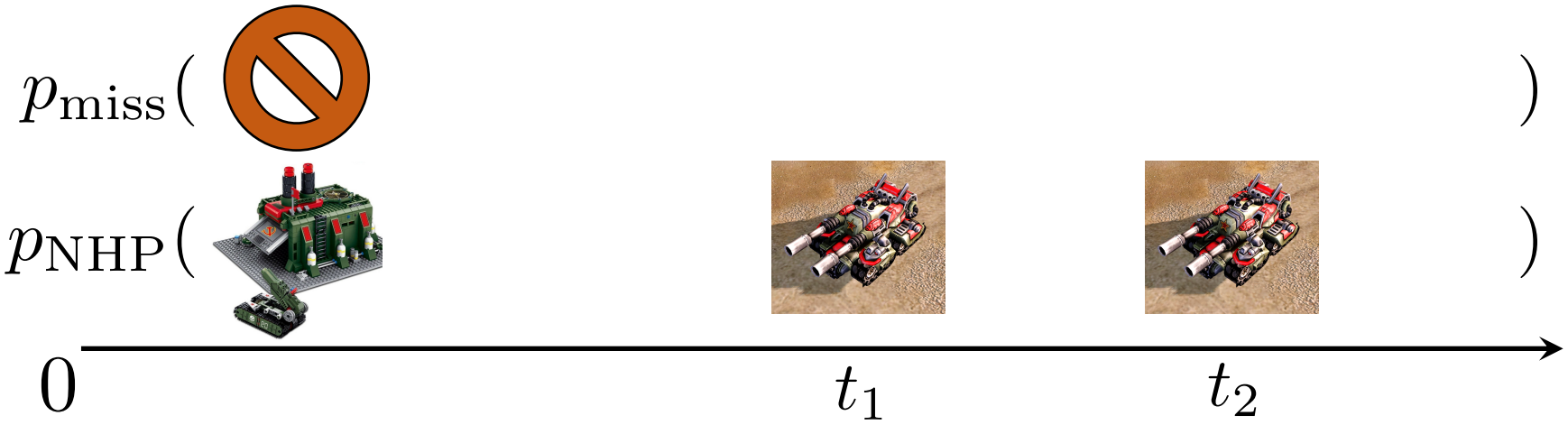
To Impute the Missing Events…



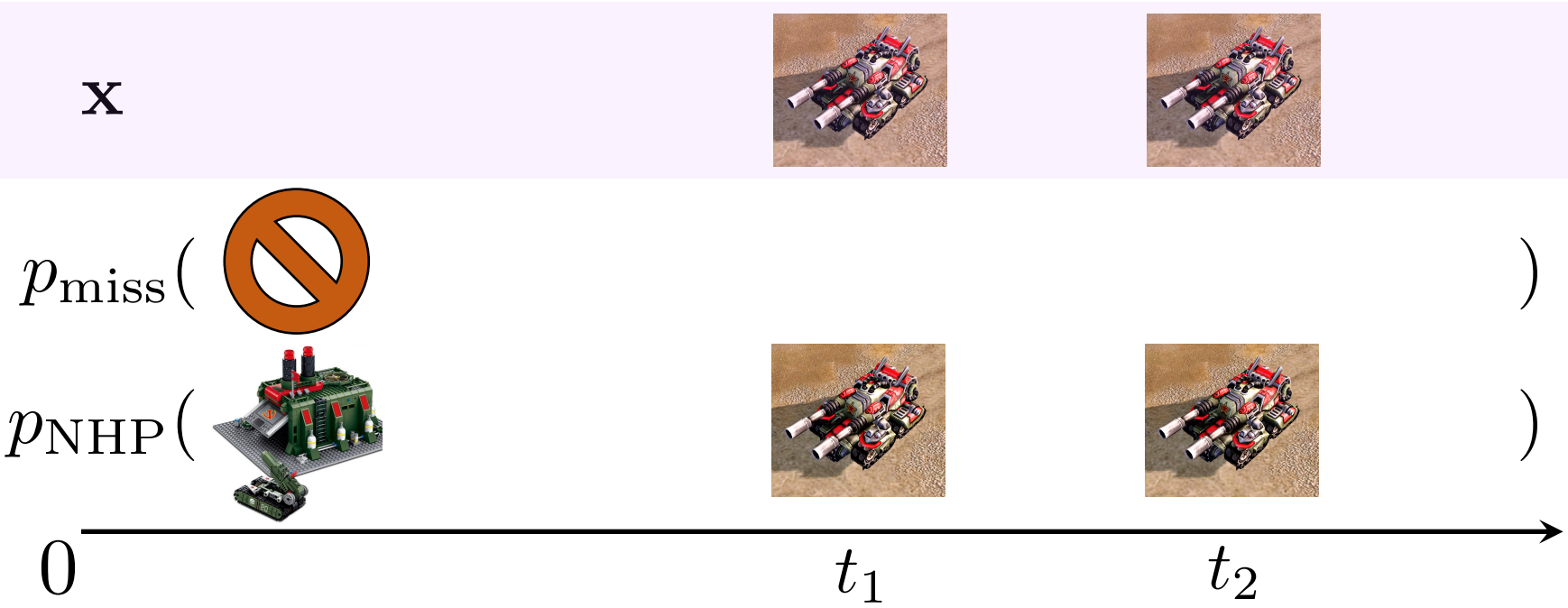
- We also need to know the missingness mechanism
- Don't propose anything which we know won't be missing
- In-view events won't be missing
- Out-of-view events must be missing



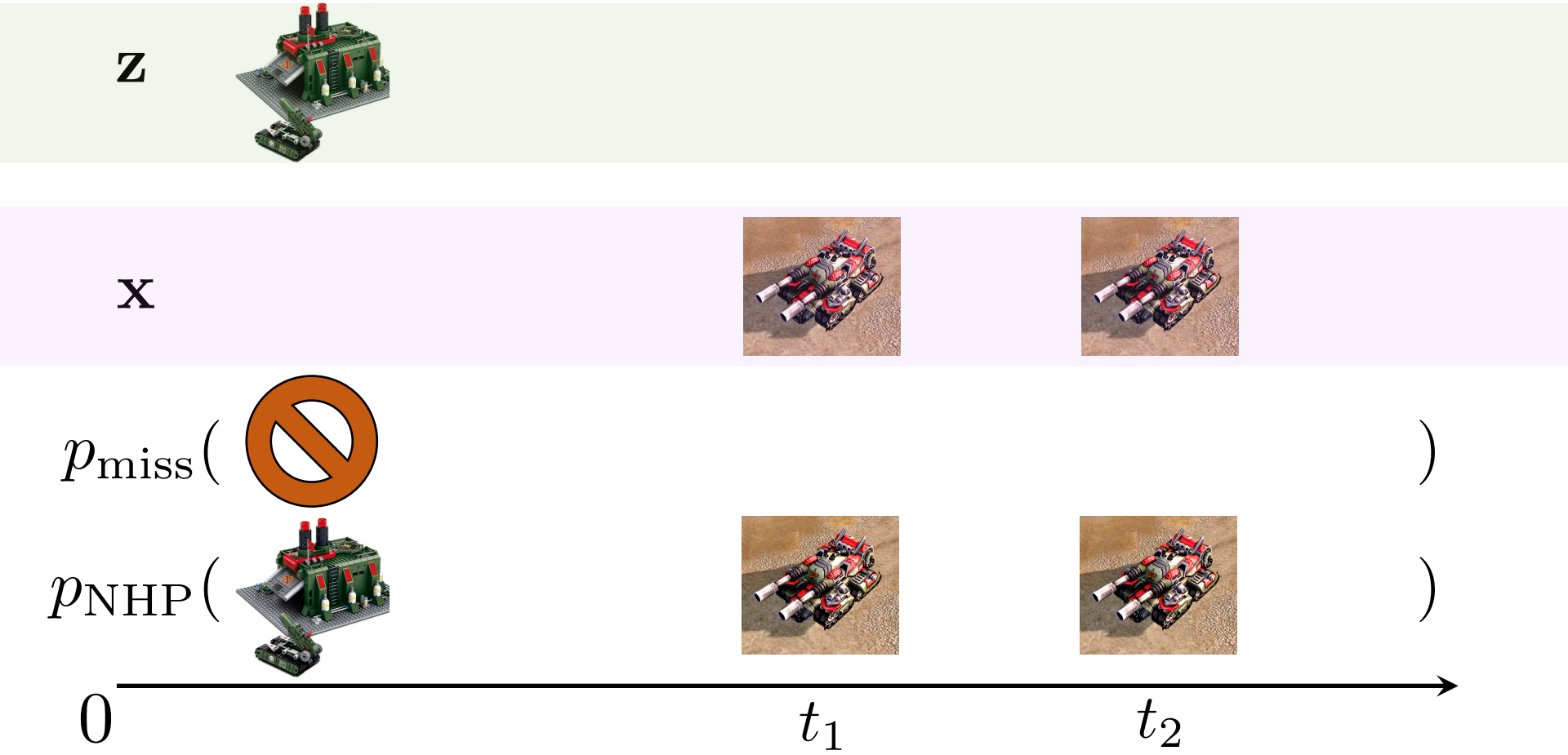
To Impute the Missing Events…



To Impute the Missing Events…



To Impute the Missing Events…



To Impute the Missing Events…

$\mathbf{z}$? ? ? ?

$\mathbf{x}$

0 t_1 t_2

To Impute the Missing Events…

$$p(\mathbf{z} \mid \mathbf{x}) \propto p_{\mathrm{NHP}}(\mathbf{x} \sqcup \mathbf{z}) p_{\mathrm{miss}}(\mathbf{z} \mid \mathbf{x} \sqcup \mathbf{z})$$

$\mathbf{z}$? ? ? ?

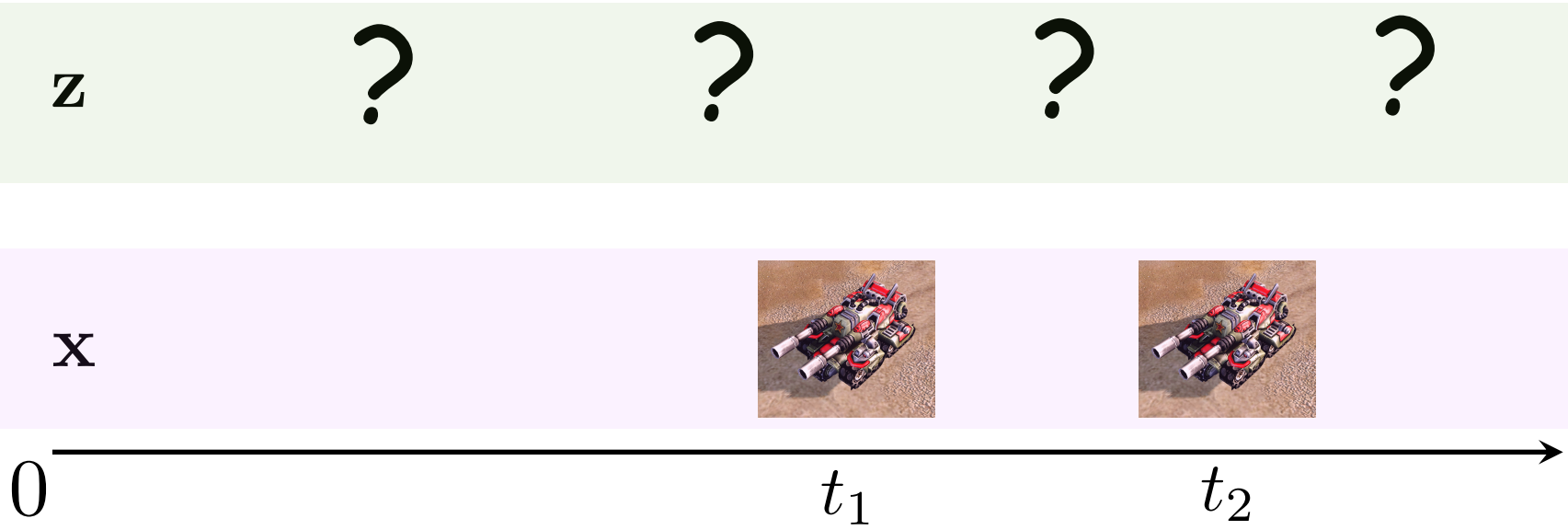
$\mathbf{x}$

0 t_1 t_2

Challenge

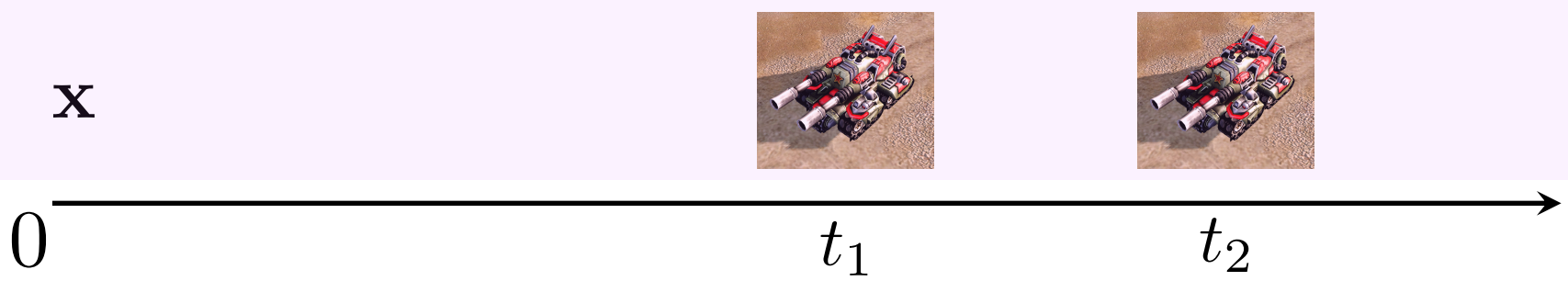
$$p(\mathbf{z} \mid \mathbf{x}) = ?$$

- Complicated p_{NHP}
- Exact inference is intractable
- How about Monte Carlo?



Sequential Monte Carlo

- Sample from $q(\mathbf{z} \mid \mathbf{x})$



Sequential Monte Carlo

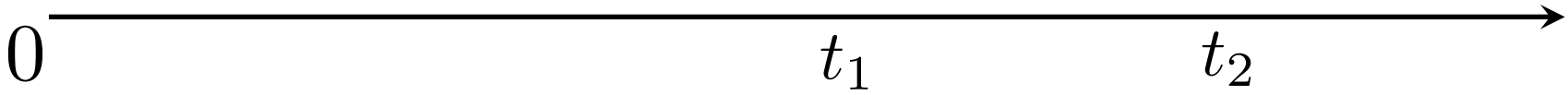
$\mathbf{z}_1$

$\mathbf{z}_2$

$\mathbf{z}_3$

$\mathbf{x}$

- Sample from $q(\mathbf{z} \mid \mathbf{x})$



Sequential Monte Carlo

$\mathbf{z}_1$

$\mathbf{z}_2$

$\mathbf{z}_3$

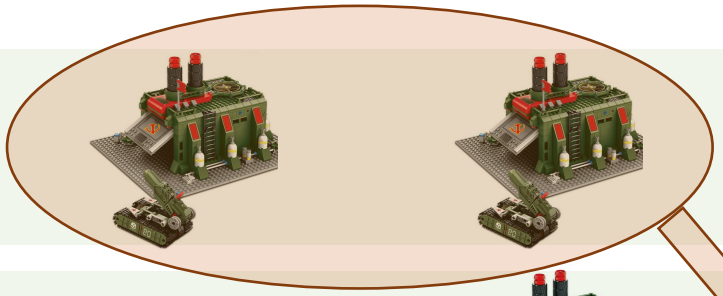
$\mathbf{x}$

- Sample from $q(\mathbf{z} \mid \mathbf{x})$
- Weight them by w

$$w \propto p(\mathbf{z} \mid \mathbf{x}) / q(\mathbf{z} \mid \mathbf{x})$$

0 t_1 t_2

Sequential Monte Carlo



$\mathbf{z}_1$

$\mathbf{z}_2$

$\mathbf{z}_3$

- Sample from $q(\mathbf{z} \mid \mathbf{x})$
- Weight them by w

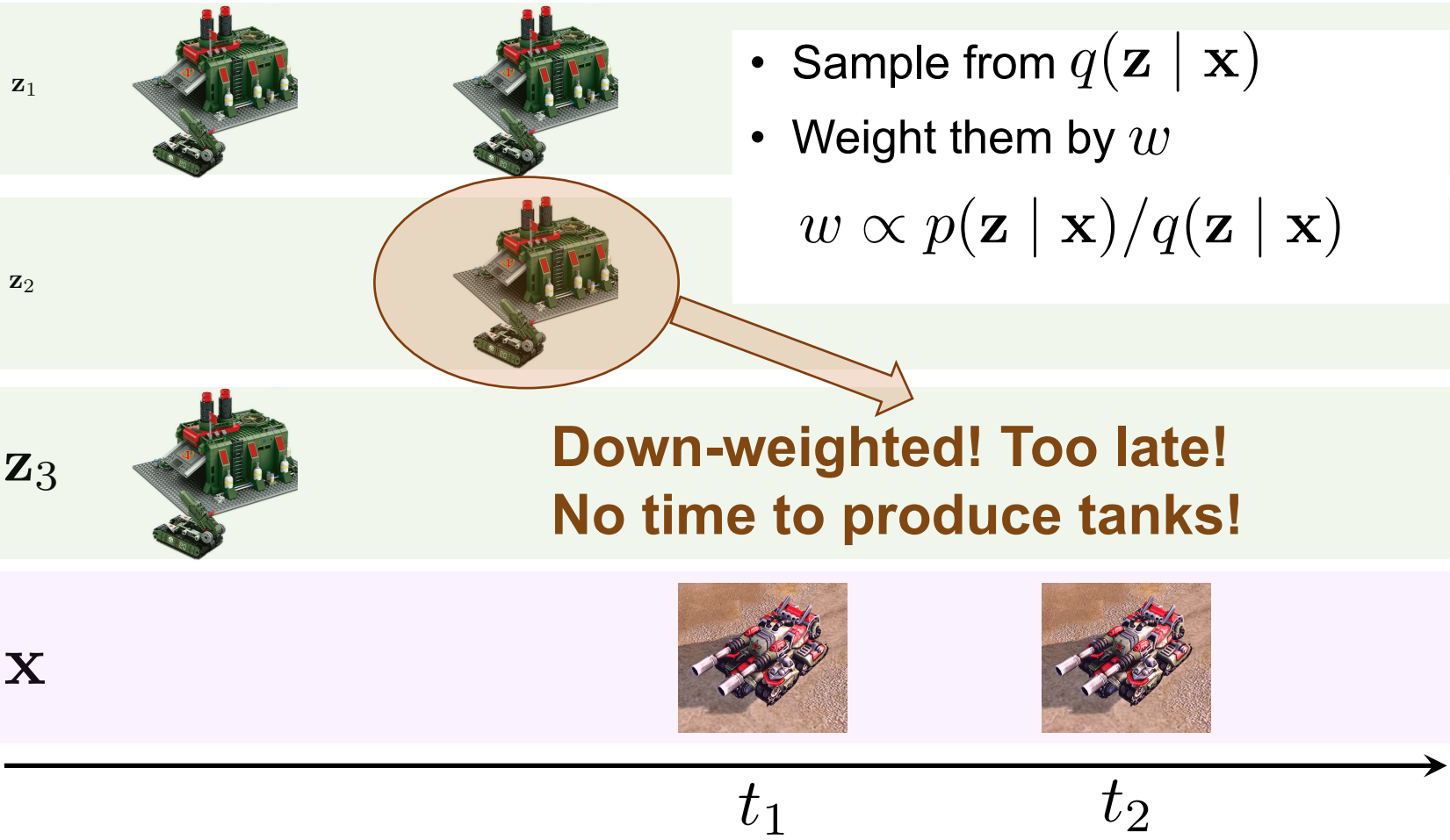
$$w \propto p(\mathbf{z} \mid \mathbf{x}) / q(\mathbf{z} \mid \mathbf{x})$$

Rare to have two factories in a row!

$\mathbf{x}$

0 t_1 t_2

Sequential Monte Carlo



- Sample from $q(\mathbf{z} \mid \mathbf{x})$
- Weight them by w

$$w \propto p(\mathbf{z} \mid \mathbf{x}) / q(\mathbf{z} \mid \mathbf{x})$$

Down-weighted! Too late!
No time to produce tanks!

Sequential Monte Carlo

$\mathbf{z}_1$

$\mathbf{z}_2$

- Sample from $q(\mathbf{z} \mid \mathbf{x})$
- Weight them by w

$$w \propto p(\mathbf{z} \mid \mathbf{x}) / q(\mathbf{z} \mid \mathbf{x})$$

$\mathbf{z}_3$

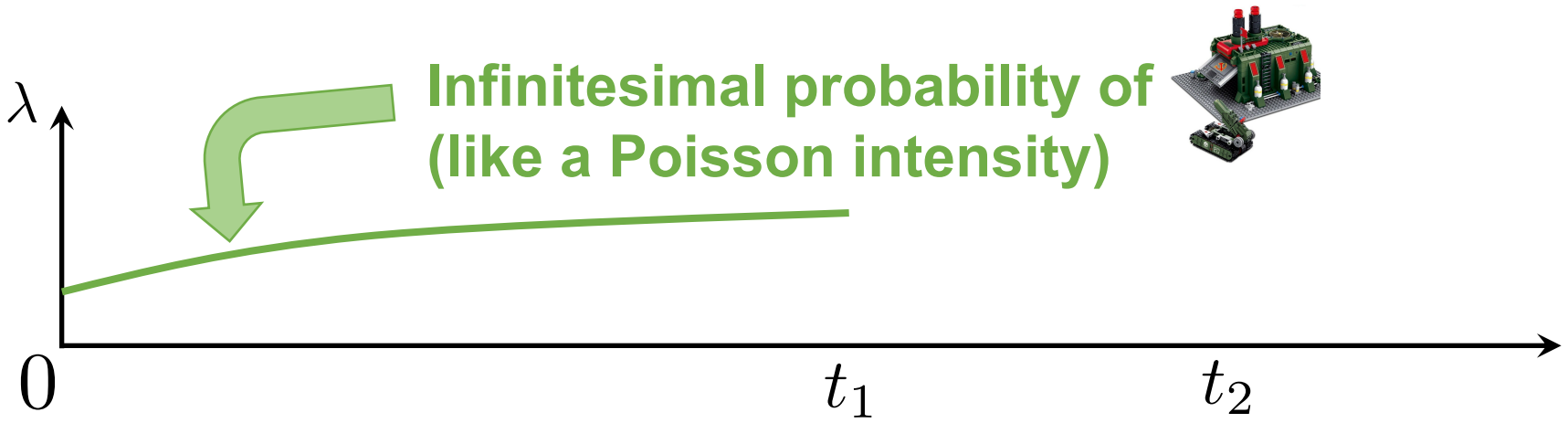
Upweighted!

$\mathbf{x}$

0 t_1 t_2

Sequential Monte Carlo

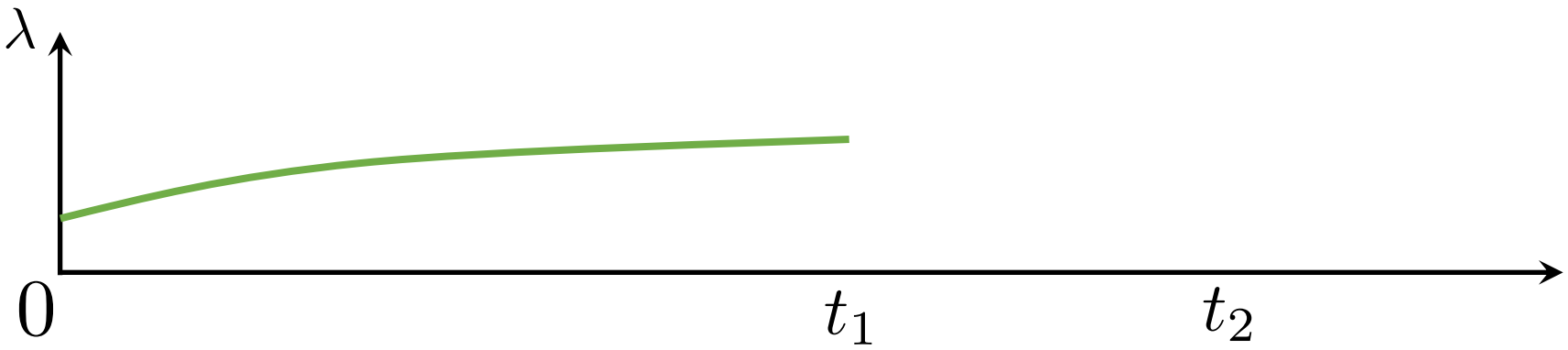
- Use the trained NHP



Sequential Monte Carlo

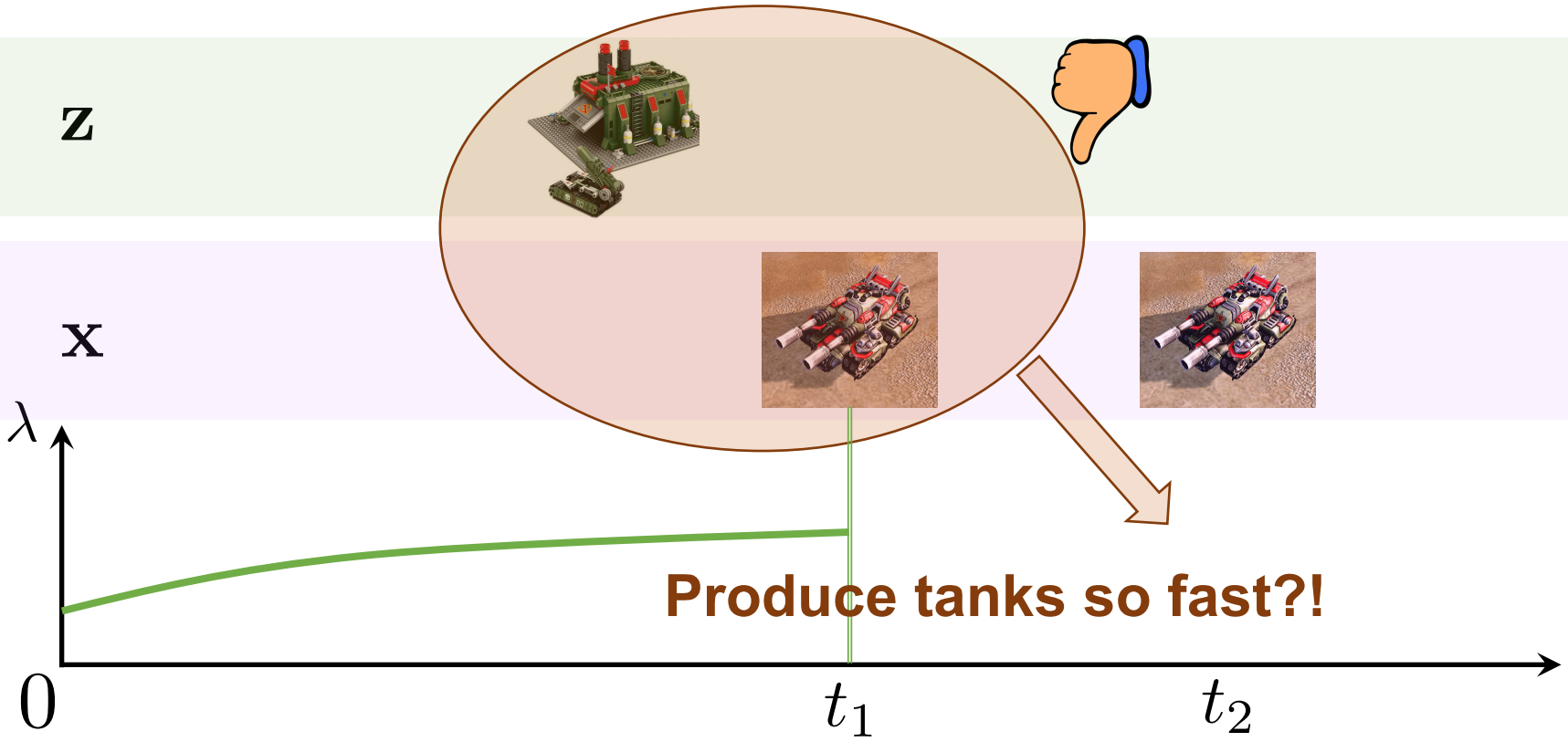
- Use the trained NHP

$\mathbf{z}$



Sequential Monte Carlo

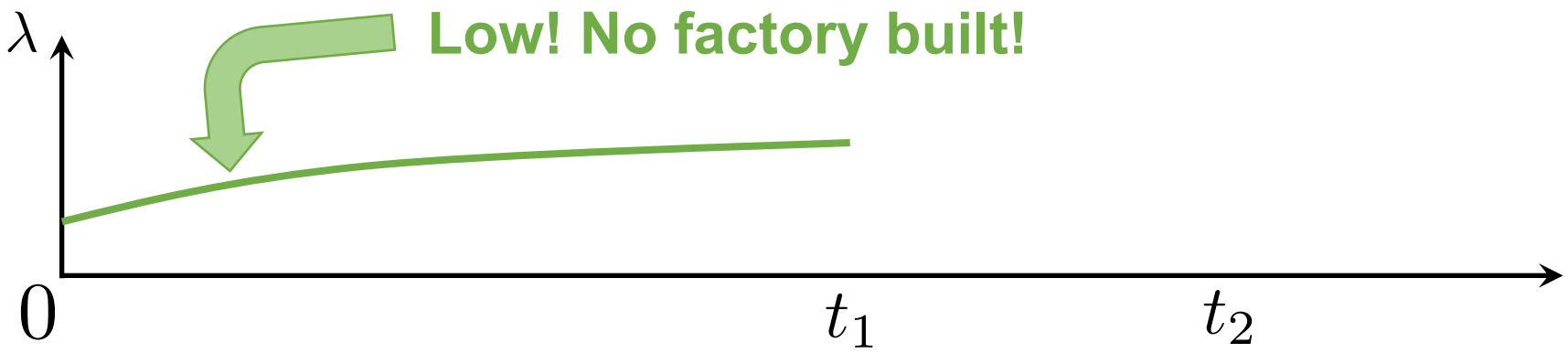
- Use the trained NHP



Sequential Monte Carlo

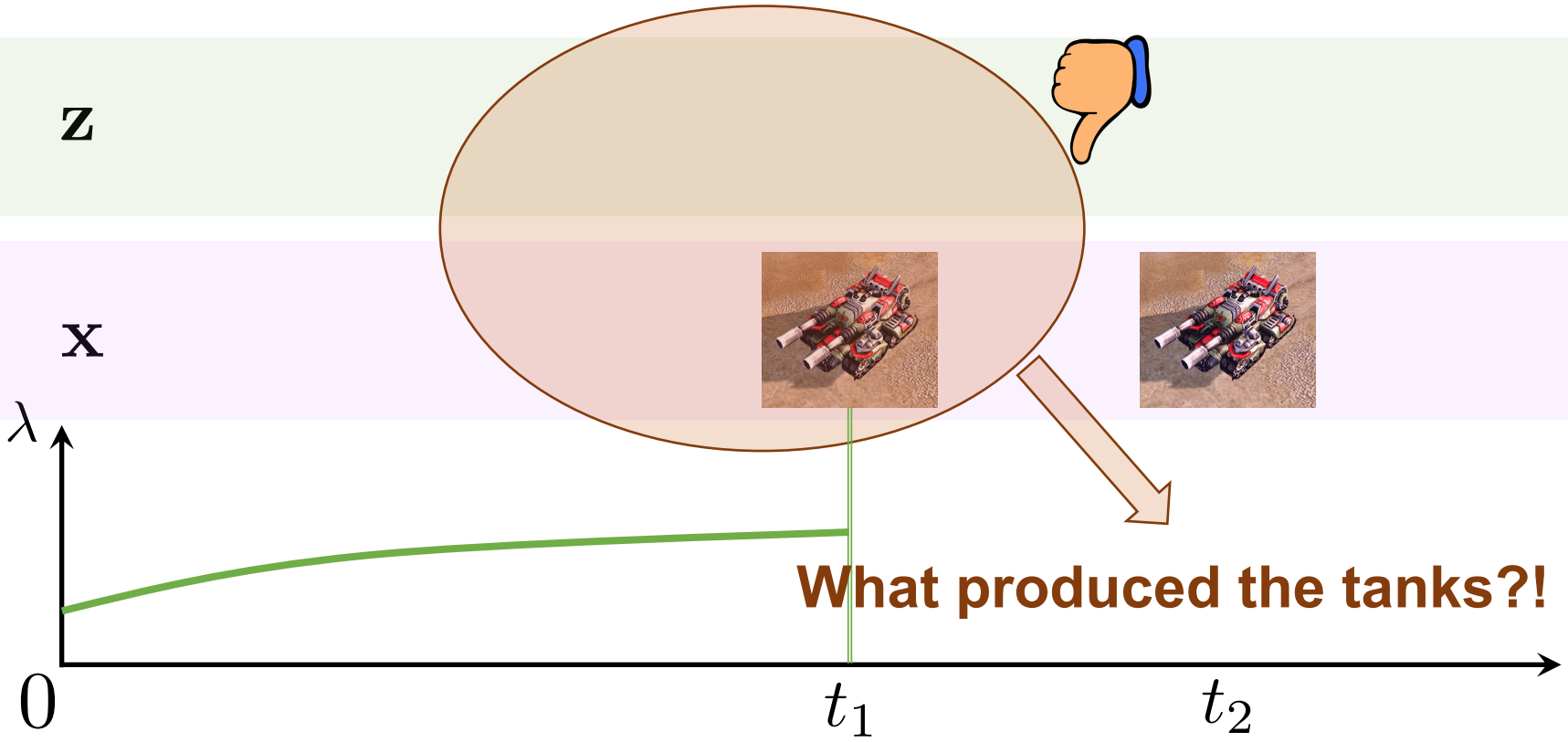
- Use the trained NHP

$\mathbf{z}$



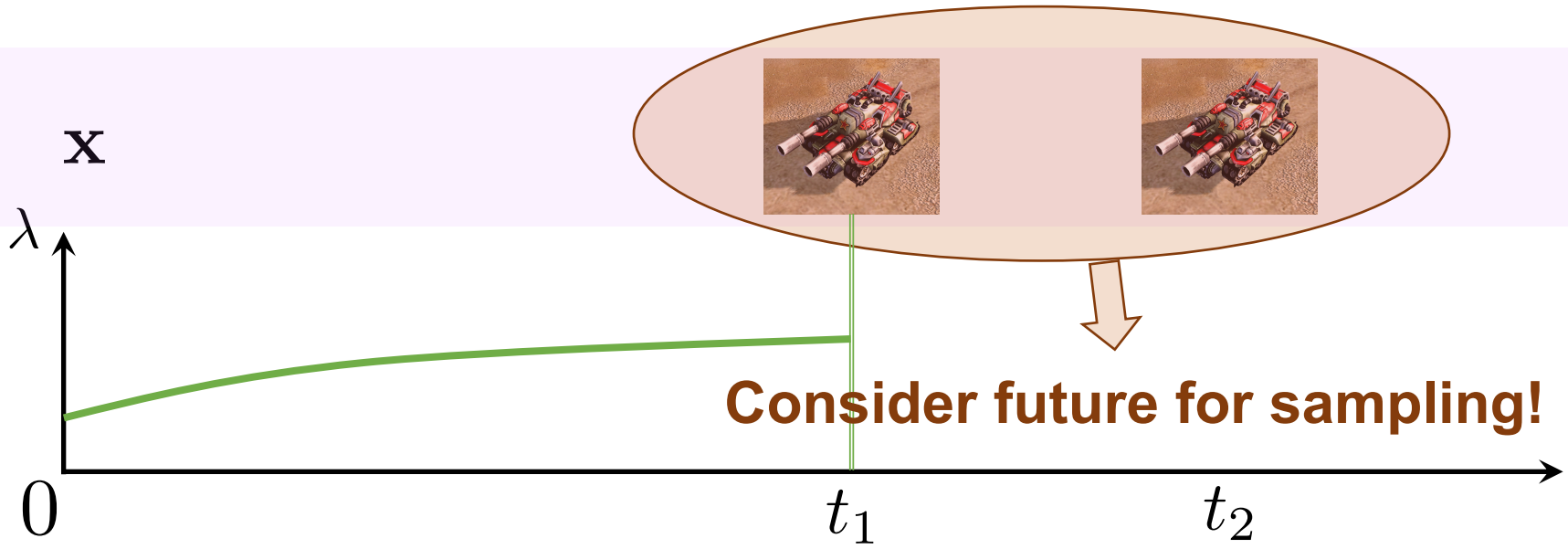
Sequential Monte Carlo

- Use the trained NHP



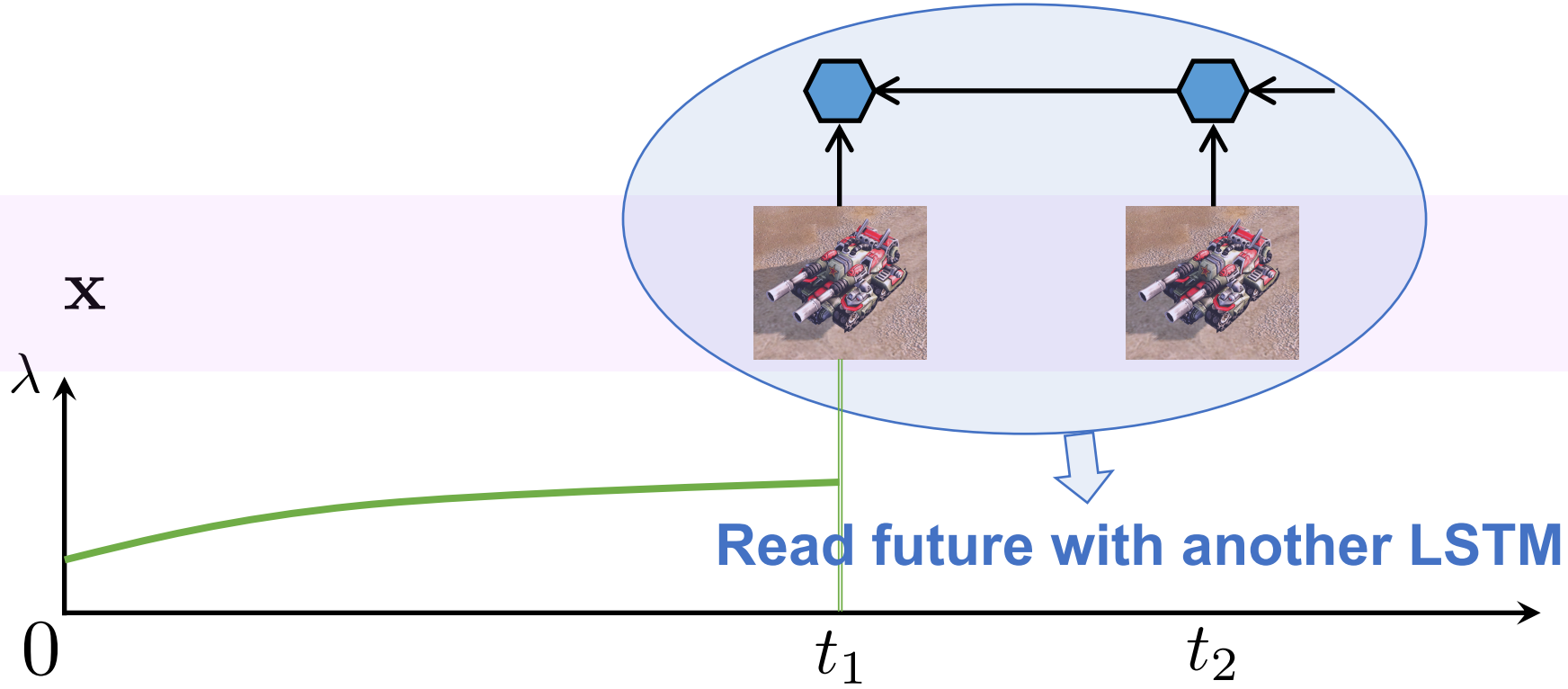
Sequential Monte Carlo

- Use the trained NHP
- Take future into account



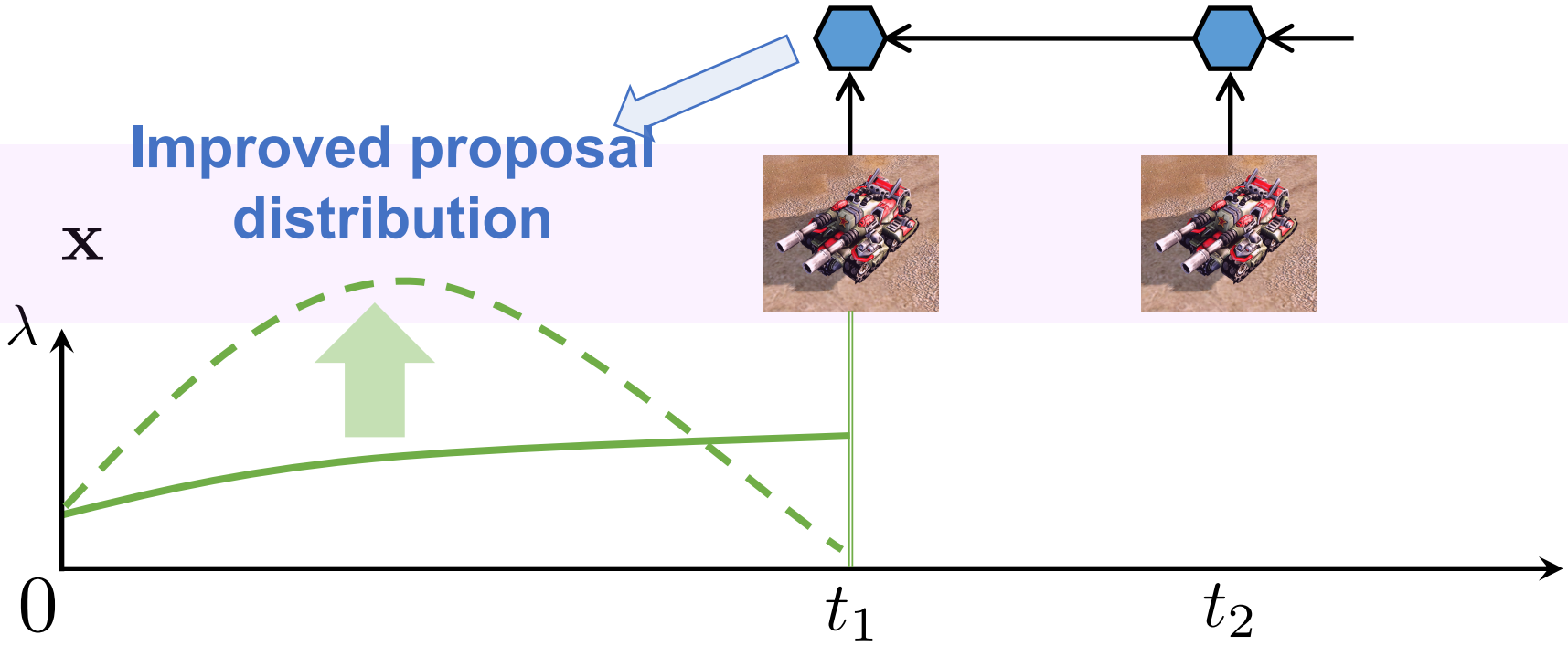
Sequential Monte Carlo

- Use the trained NHP
- Take future into account



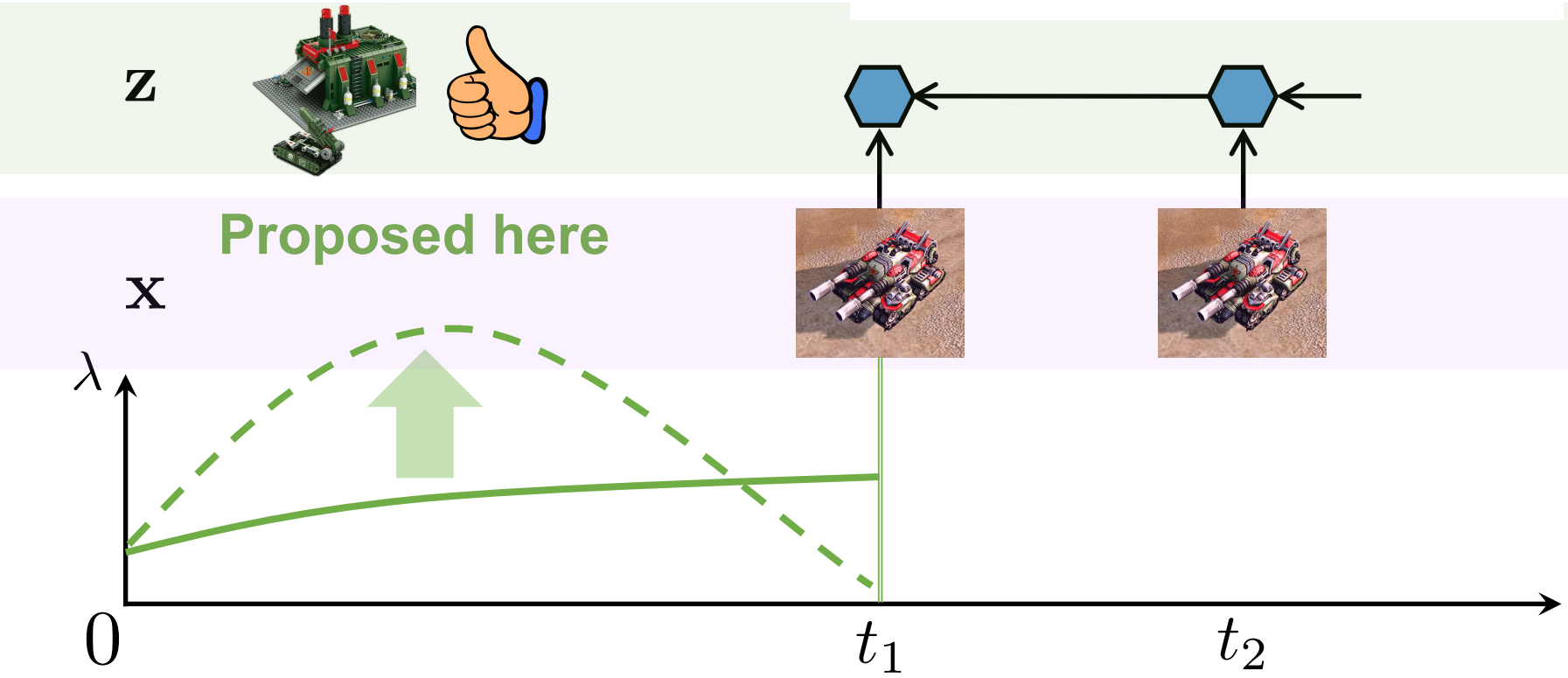
Sequential Monte Carlo

- Use the trained NHP
- Take future into account

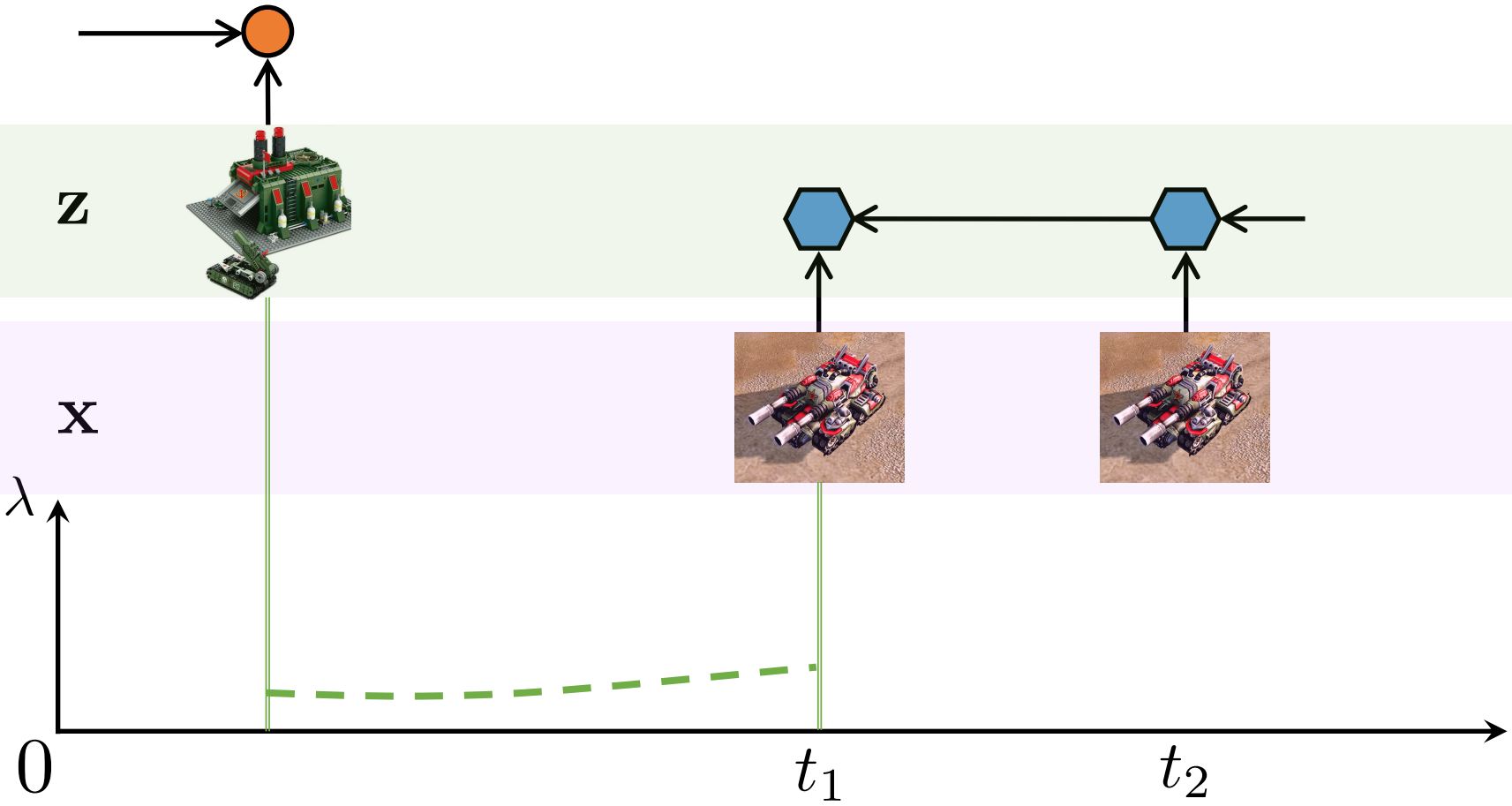


Sequential Monte Carlo

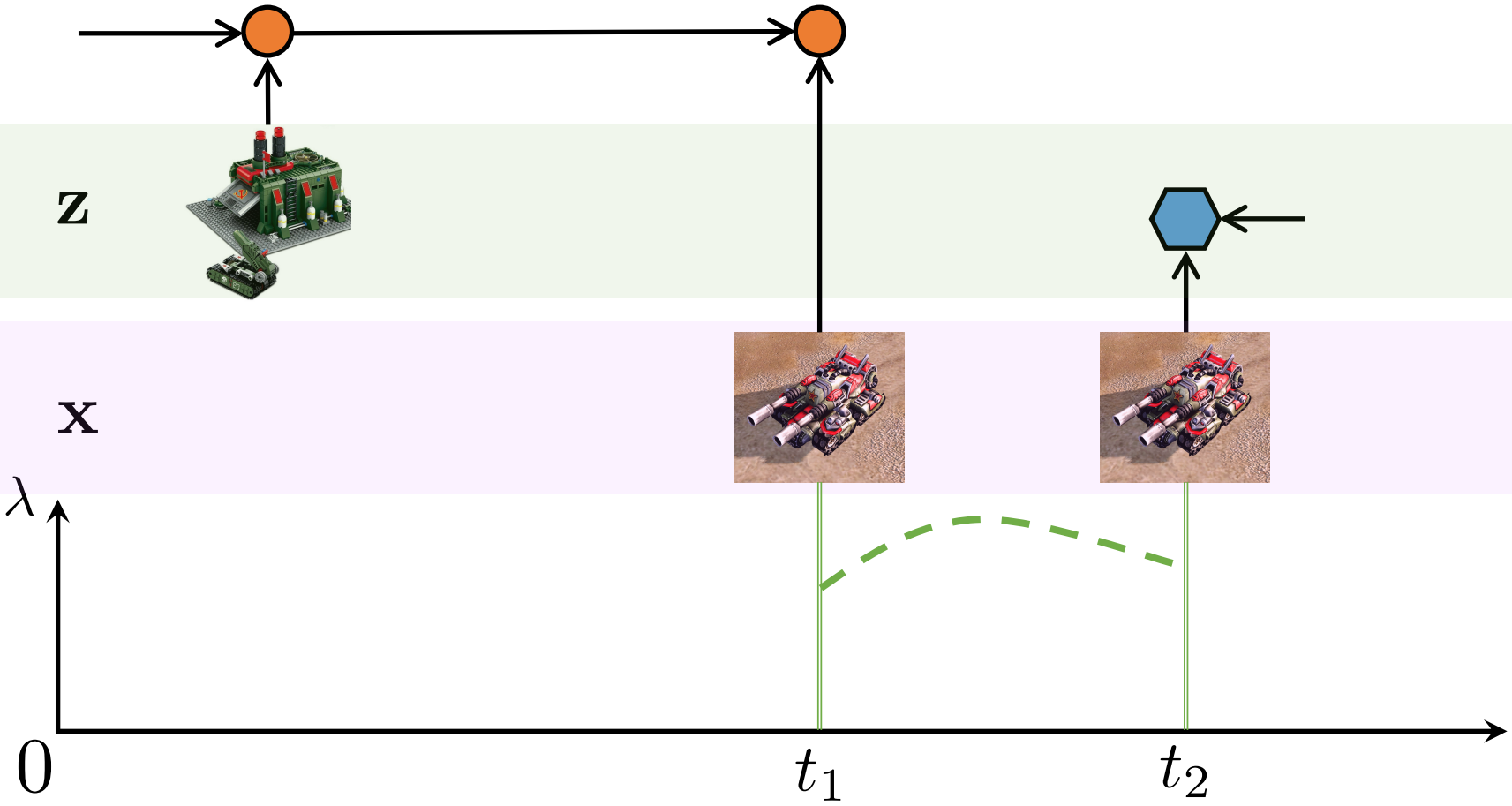
- Use the trained NHP
- Take future into account



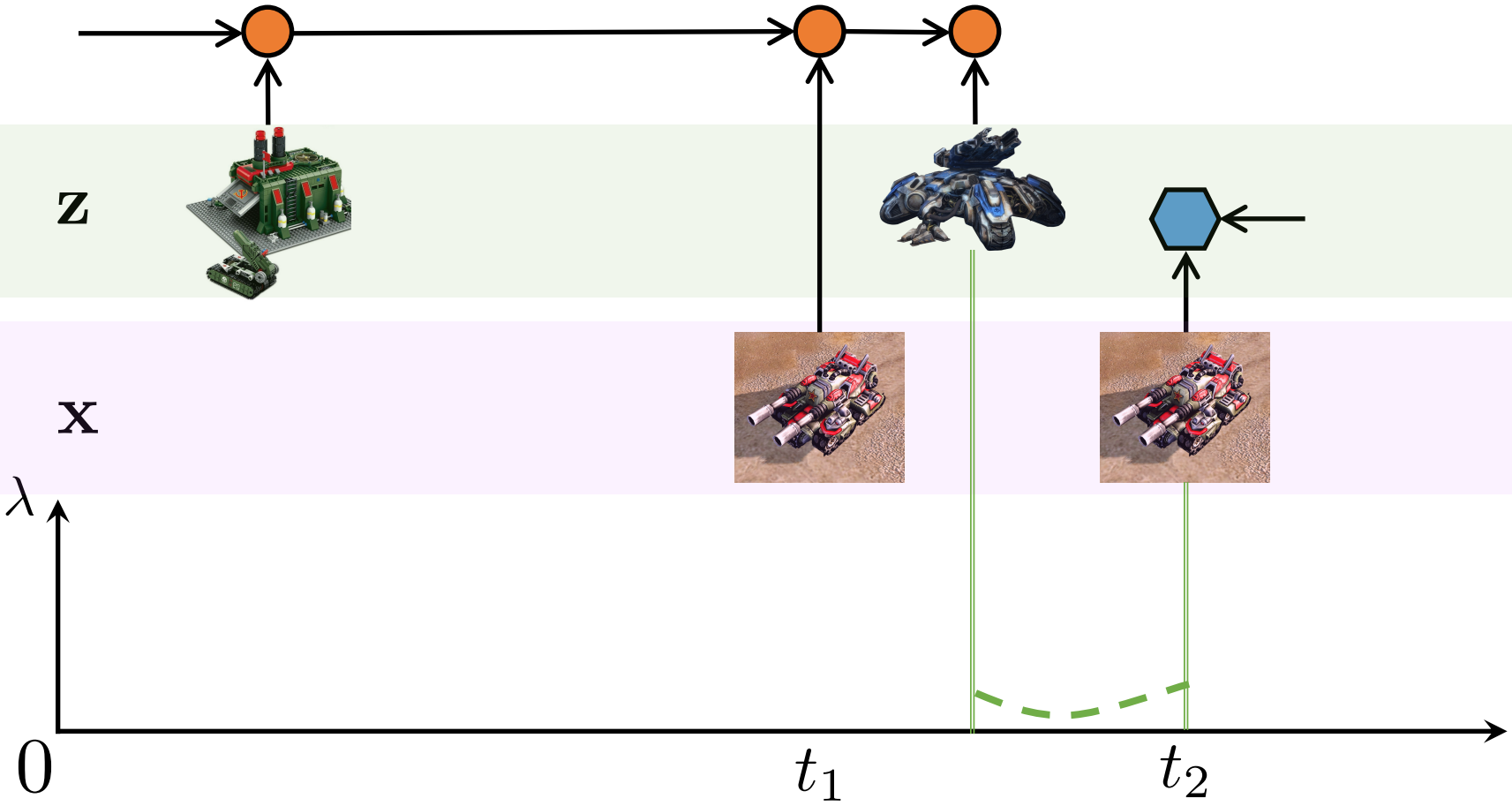
Sequential Monte Carlo



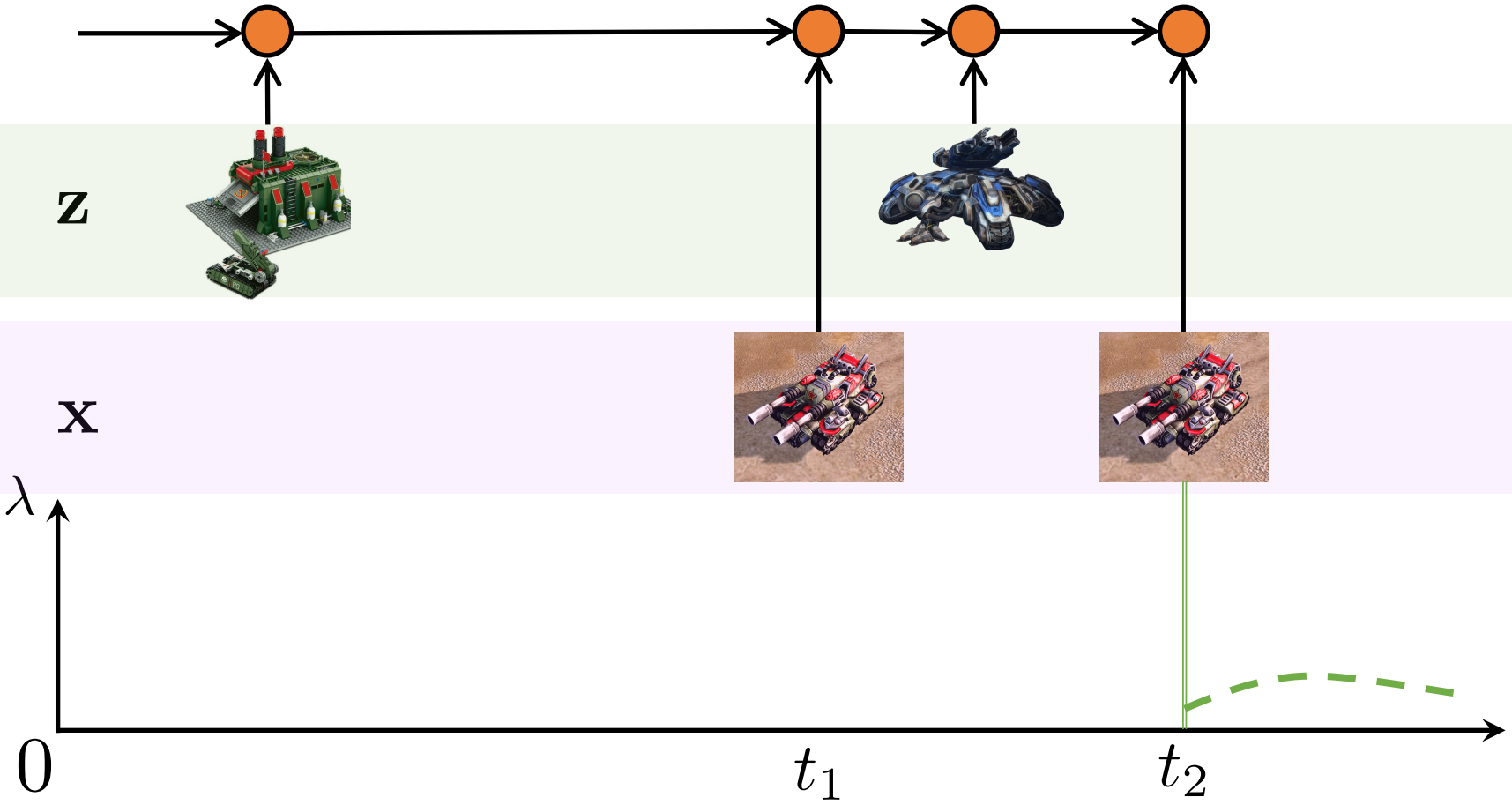
Sequential Monte Carlo



Sequential Monte Carlo

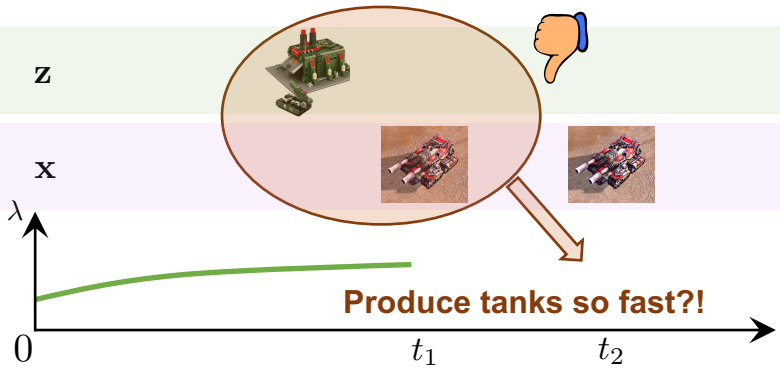


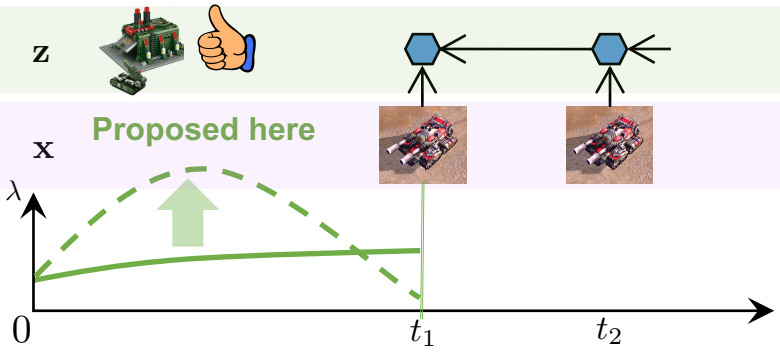
Sequential Monte Carlo



Sequential Monte Carlo

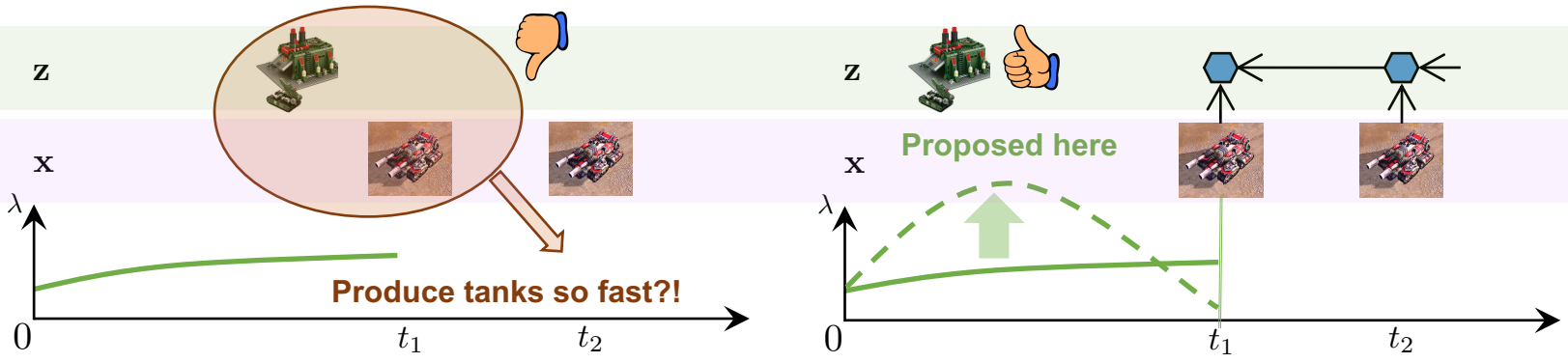
- Particle filtering
- Particle smoothing





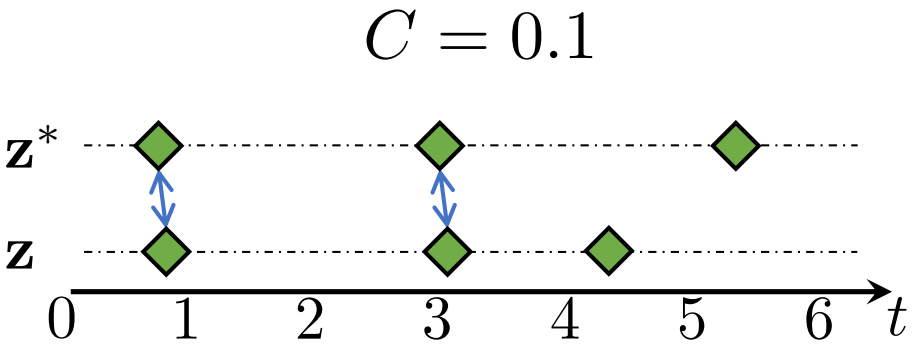
More in Our Paper

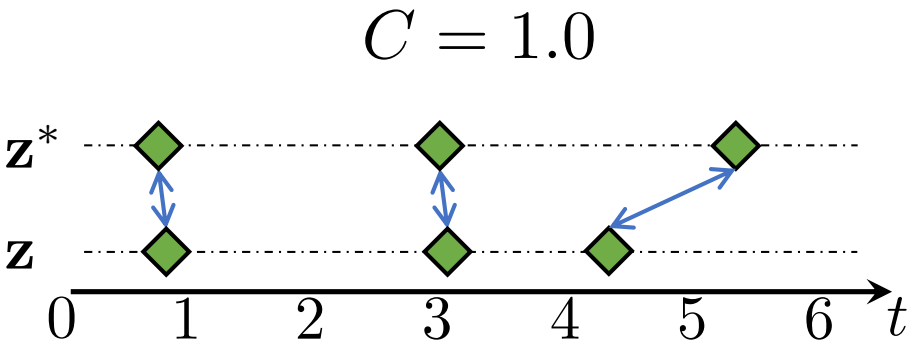
- How to train $q(\mathbf{z} \mid \mathbf{x})$ for particle smoothing



More in Our Paper

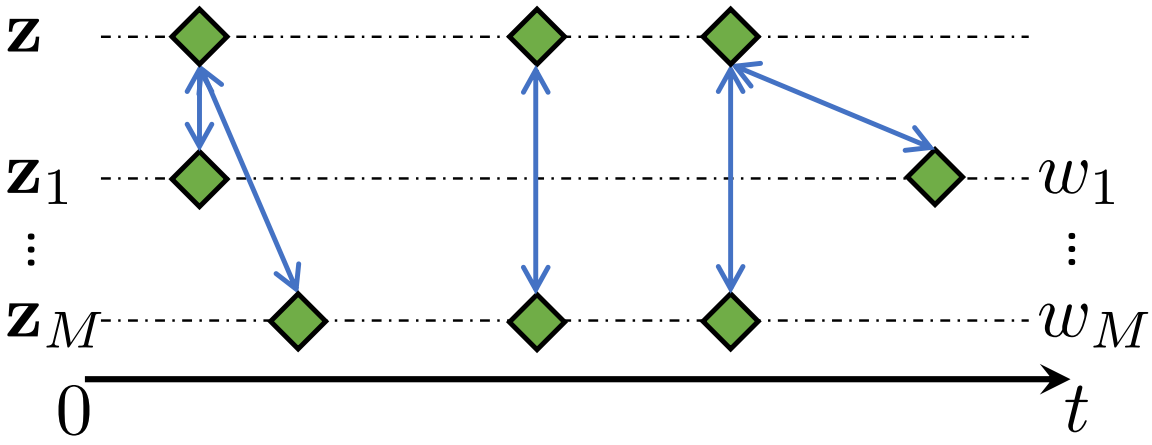
- How to train $q(\mathbf{z} \mid \mathbf{x})$ for particle smoothing
- Optimal transport distance





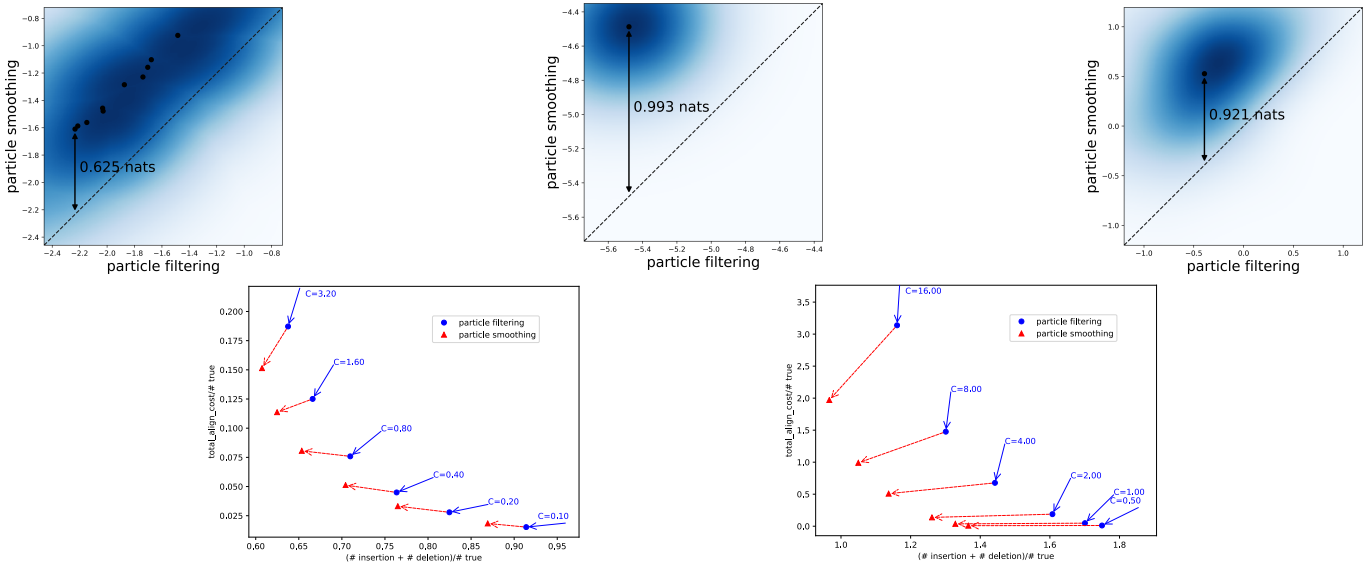
More in Our Paper

- How to train $q(\mathbf{z} \mid \mathbf{x})$ for particle smoothing
- Optimal transport distance
- Minimum Bayes risk decoding



More in Our Paper

- How to train $q(\mathbf{z} \mid \mathbf{x})$ for particle smoothing
- Optimal transport distance
- Minimum Bayes risk decoding
- Positive experimental results



Imputing Missing Events in Continuous-Time Event Streams

Tue Jun 11th 18:30 – 21:00
Pacific Ballroom #257

Hongyuan Mei[1], Guanghui Qin[2], Jason Eisner[1]

[1]Department of Computer Science, Johns Hopkins University, USA
[2]Department of Physics, Peking University, China

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING