Weighted Model Integration

Zhe Zeng
University of California, Los Angeles

Fanqi Yan

Paolo Morettin

Antonio Vergari

Andrea Passerini

Guy Van den Broeck

Oct 19th, 2023 - Probabilistic Circuits and Logic Workshop at Simons Institute

Outline

- Definition of Weighted Model Integration (WMI)
- Tractability Analysis of WMI Problem Class
- Application in Bayesian Deep Learning

Outline

- **Definition of Weighted Model Integration (WMI)**
- Tractability Analysis of WMI Problem Class
- Application in Bayesian Deep Learning

WMC vs. WMI

	Weighted Model Counting (WMC)
Domain	Discrete
Model Language	Propositional formula Δ
Formula Δ	$A \implies B$
Variable	Boolean: $A, B \in \{0, 1\}$
Weight	$w(A) = 1, w(\neg A) = 2$ $w(B) = 3, w(\neg B) = 5$

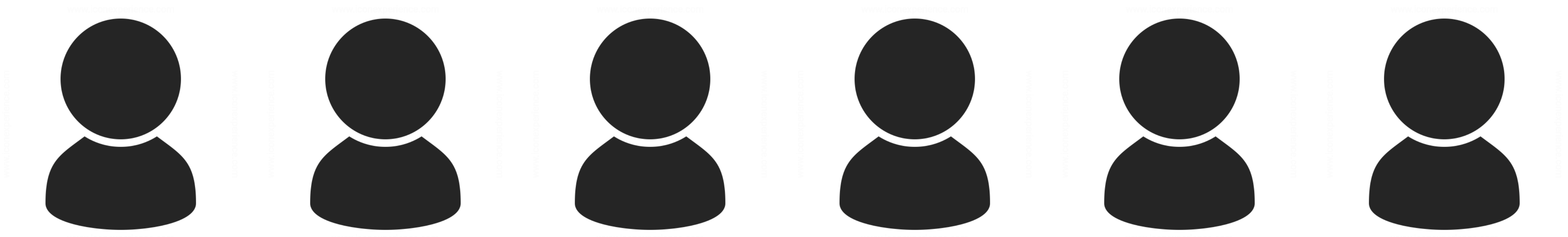
A	B	Model?	Weight
T	T	Yes	1 * 3 = 3
T	F	No	0
F	T	Yes	2 * 3 = 6
F	F	Yes	2 * 5 = 10

+ -----------

WMC = 19

Motivation

Example: Skill matching system



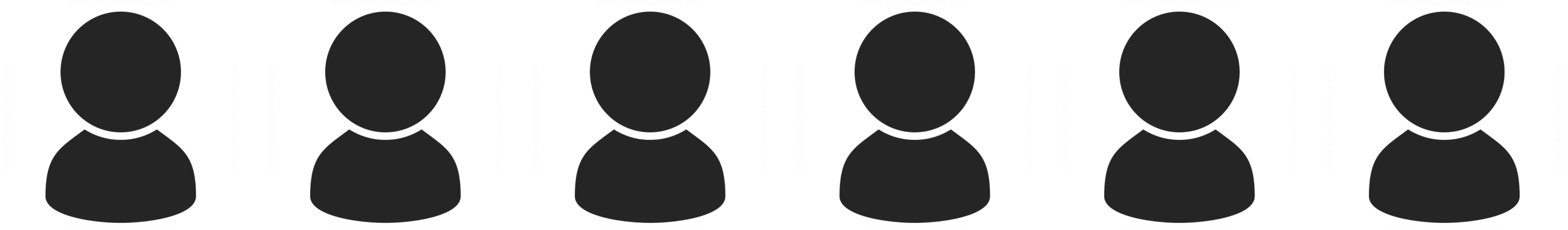
Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

Motivation

Example: Skill matching system

- Each ***player*** has a certain skill

  $\Rightarrow$ *continuous variables*

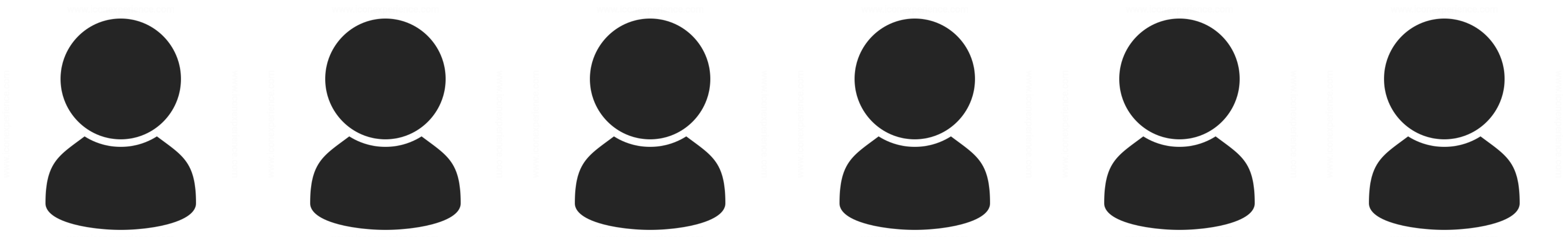


Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

Motivation

Example: Skill matching system

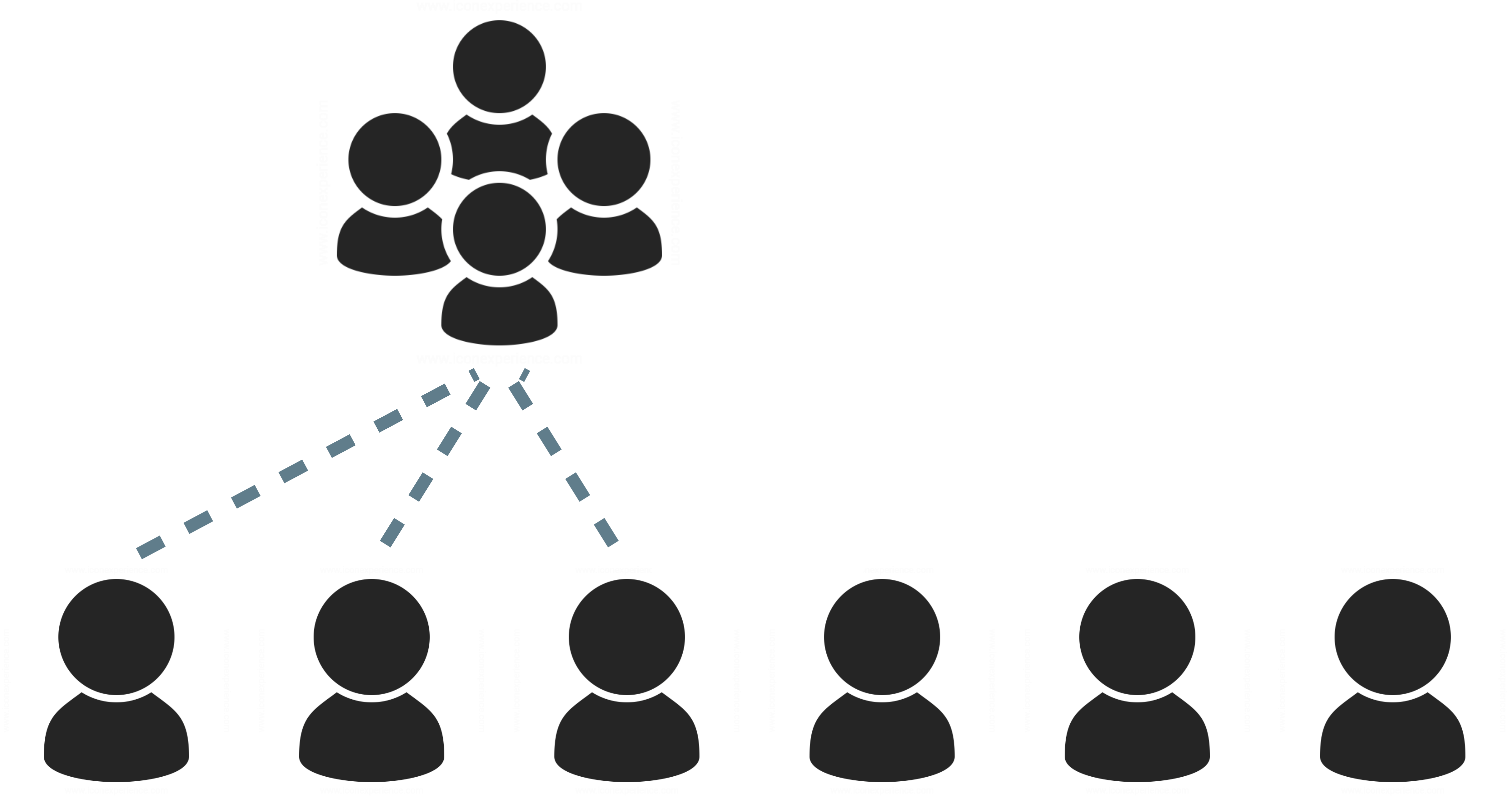
- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$



Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

Motivation

Example: Skill matching system



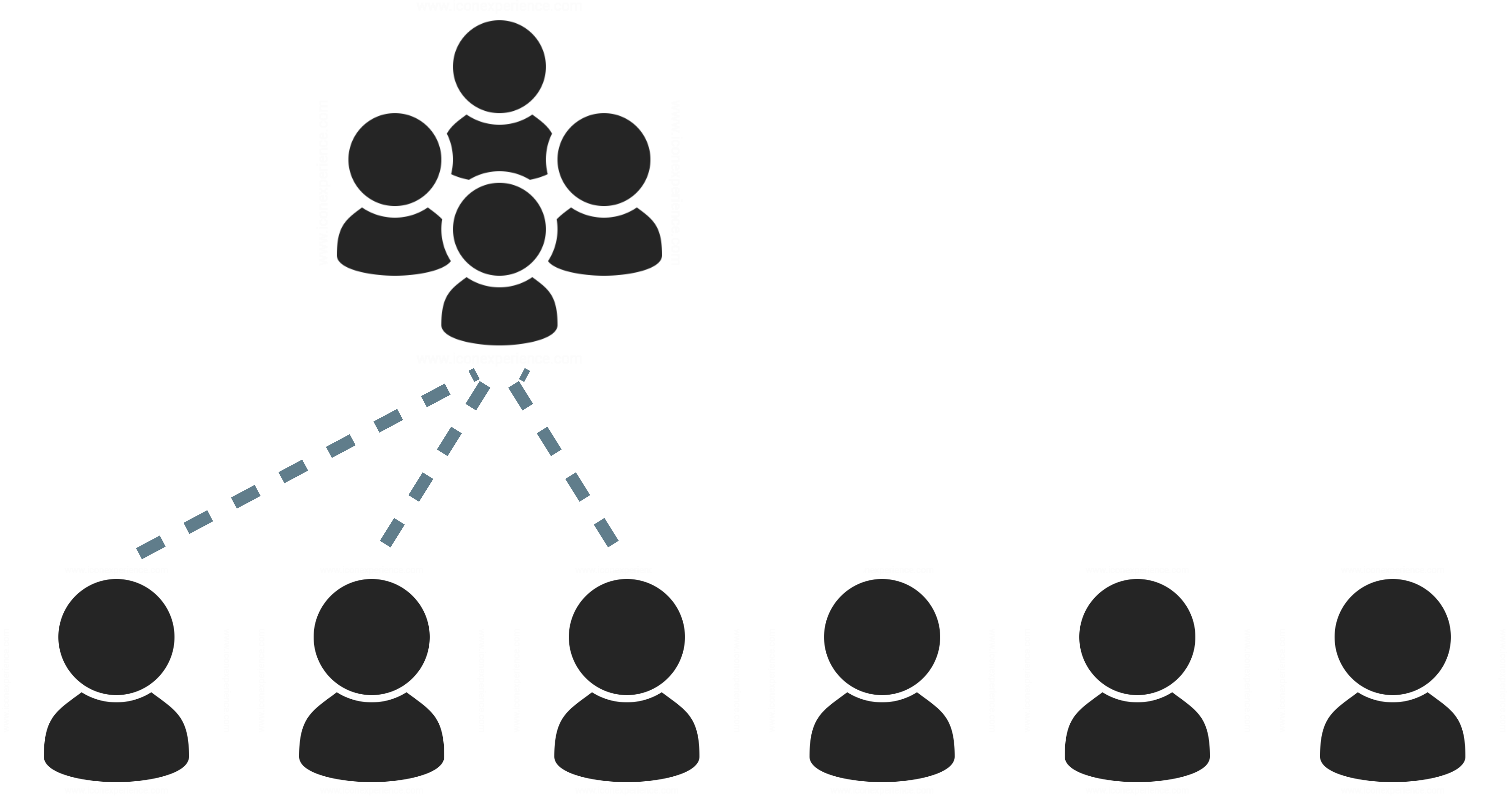
- $0 \le X_{P_i} \le 10$
  for $i = 1, \ldots, N$

- Players can form ***teams***
  $\Longrightarrow$ *complex constraints*

Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

Motivation

Example: Skill matching system

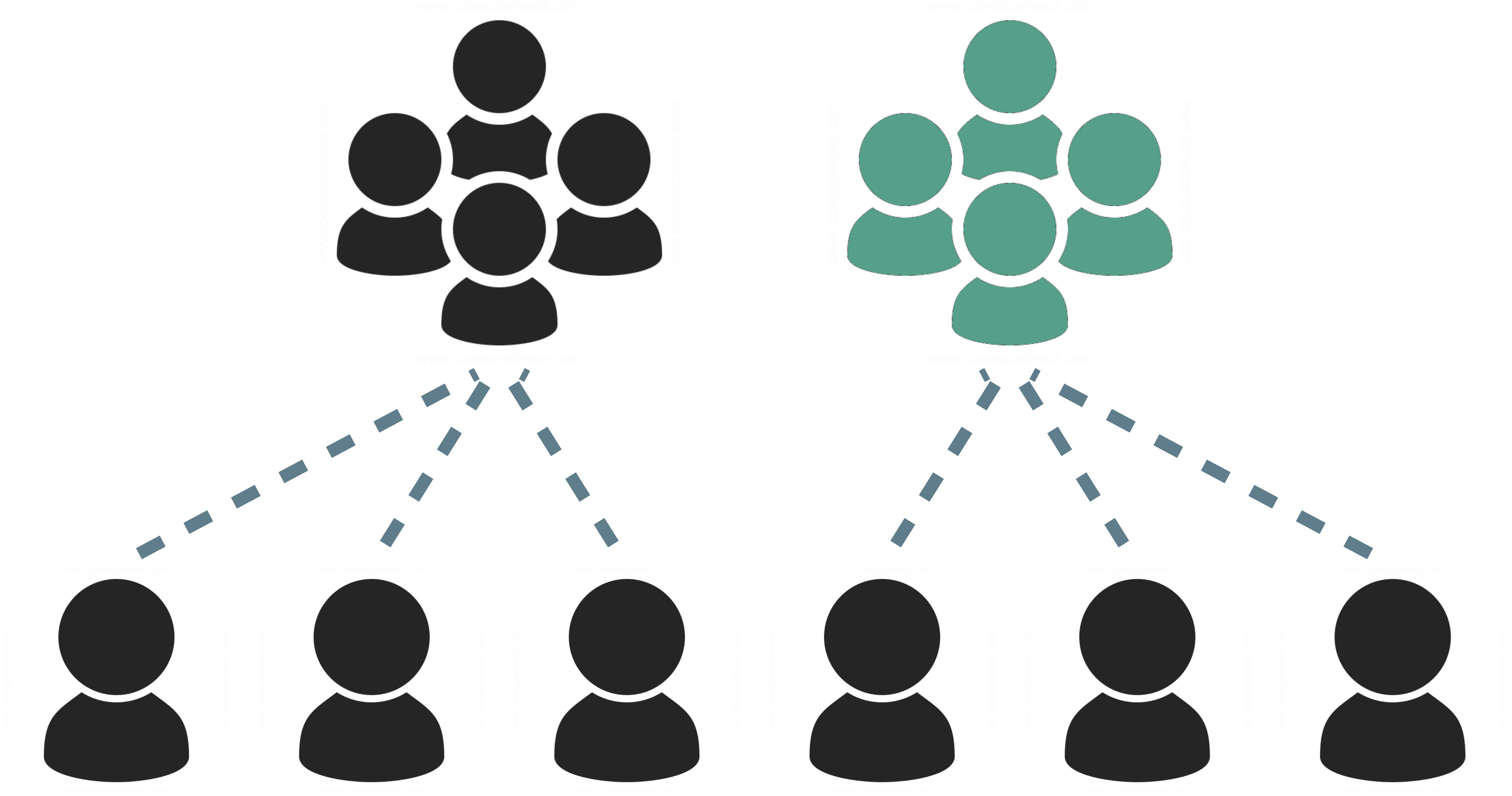


- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$
- $\mid X_{T_j} - X_{P_i} \mid < 1$
  for $j = 1, \ldots, M$

Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

Motivation

Example: Skill matching system

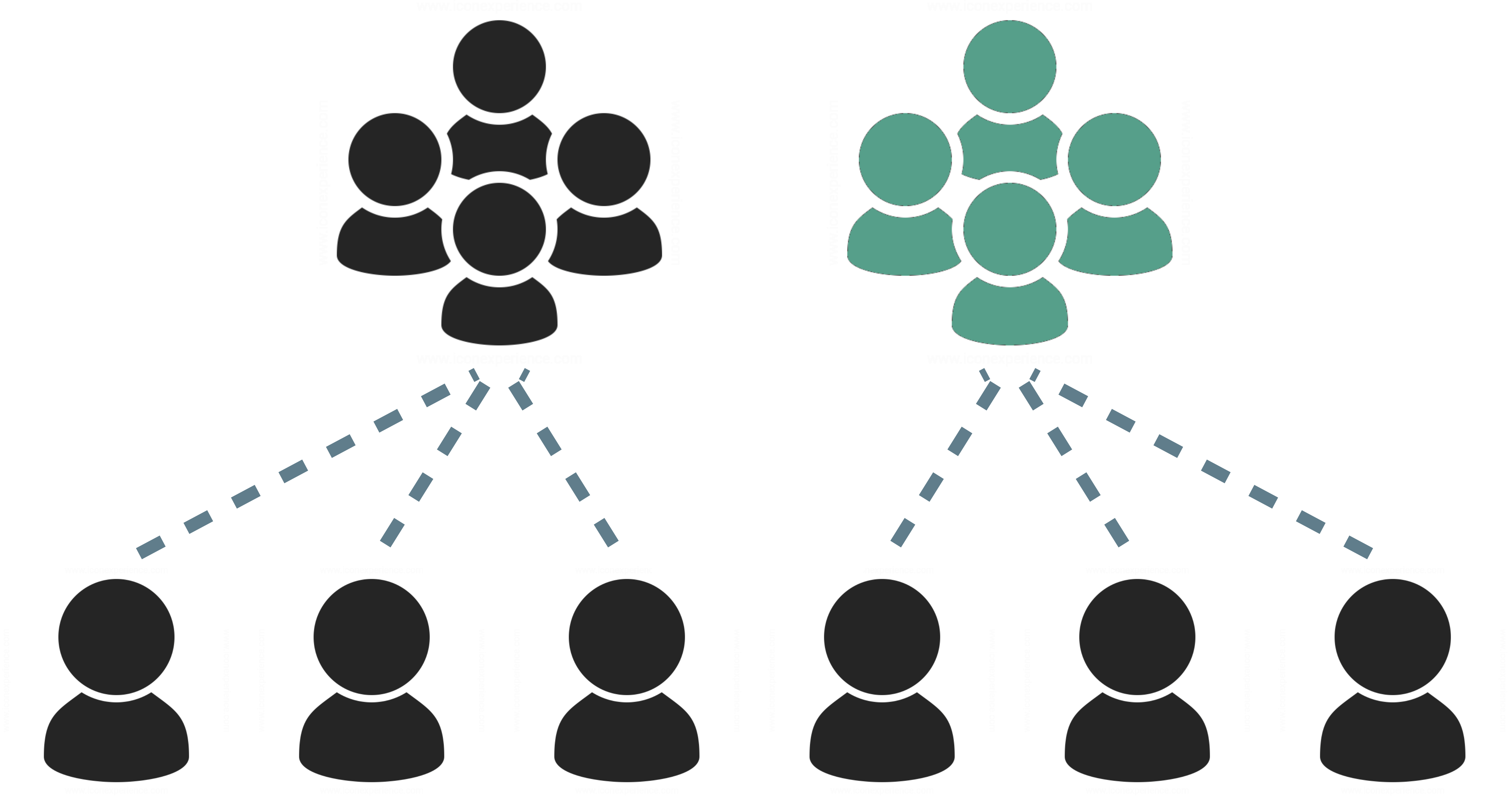


- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$
- $| X_{T_j} - X_{P_i} | < 1$
  for $j = 1, \ldots, M$
- Good teams form a ***squad***
  $\Longrightarrow$ *discrete variables*

Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

Motivation

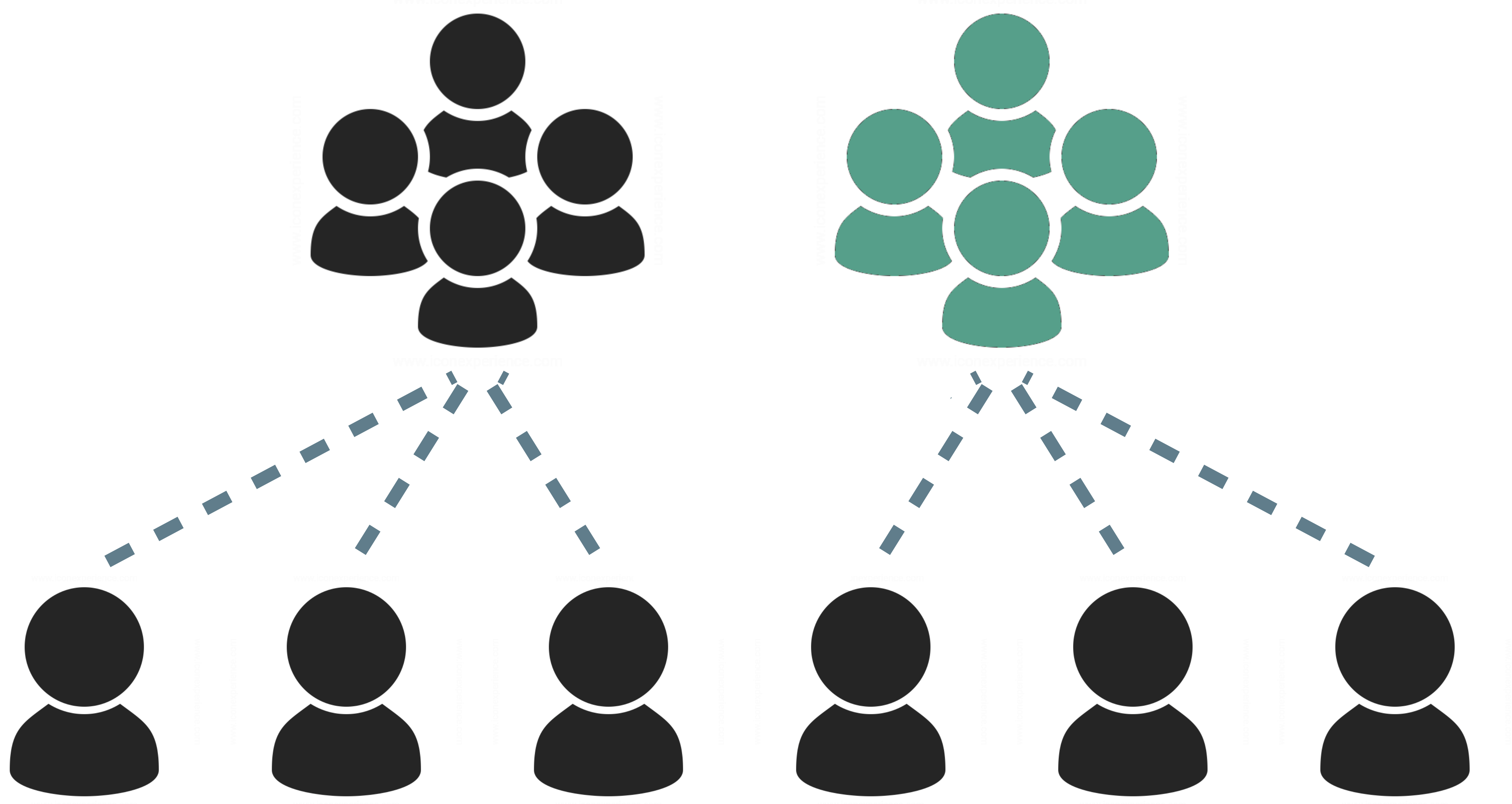
Example: Skill matching system



- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$
- $\mid X_{T_j} - X_{P_i} \mid < 1$
  for $j = 1, \ldots, M$
- $B_{S_j} \Rightarrow X_{T_j} > 2$
  for $j = 1, \ldots, M$

Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

Continuous + **discrete** + **constraints** = **SMT**



- $0 \leq X_{P_i} \leq 10$ for $i = 1, \ldots, N$
- $\mid X_{T_j} - X_{P_i} \mid < 1$ for $j = 1, \ldots, M$
- $B_{S_j} \Rightarrow X_{T_j} > 2$ for $j = 1, \ldots, M$

Barrett et al., "Satisfiability modulo theories", 2018

Continuous + **discrete** + **constraints** = **SMT**

$$\Delta = \bigwedge_i 0 \leq X_{P_i} \leq 10 \bigwedge_j \bigwedge_{i \in T_j} \mid X_{T_j} - X_{P_i} \mid < 1 \bigwedge_j (B_{S_j} \Rightarrow X_{T_j} > 2)$$

a single CNF SMT($\mathcal{LRA}$) formula Δ...

Barrett et al., "Satisfiability modulo theories", 2018

SMT + weights = Weighted Model Integration

$$\bigwedge_i 0 \leq X_{P_i} \leq 10$$

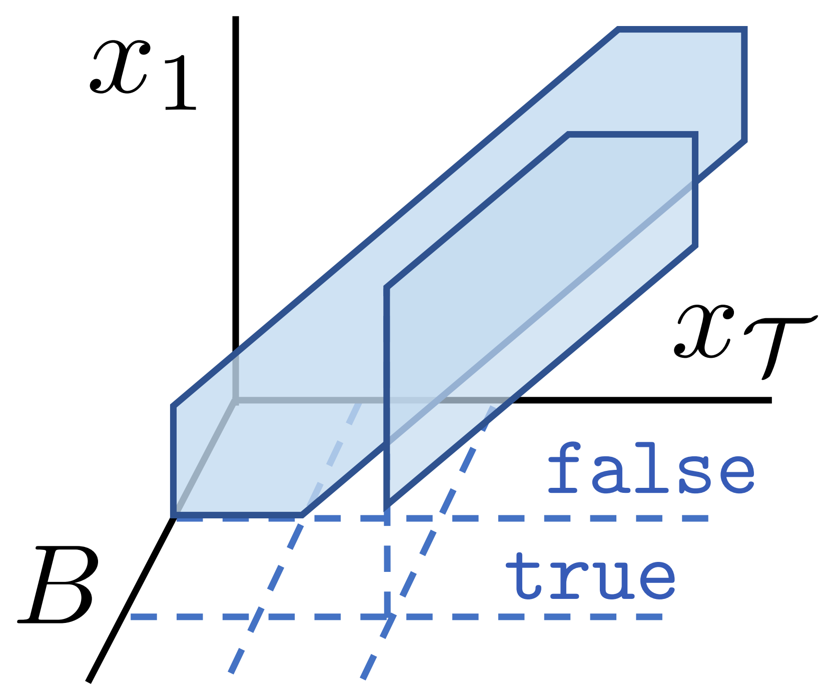
$$\bigwedge_j \bigwedge_{i \in T_j} | X_{T_j} - X_{P_i} | < 1$$

$$\bigwedge_j (B_{S_j} \Rightarrow X_{T_j} > 2)$$

+

$$\begin{cases} w(X_{P_i}), \\ \quad \text{if } 0 \leq X_{P_i} \leq 10 \\ w(X_{T_j}, X_{P_i}), \\ \quad \text{if } | X_{T_j} - X_{P_i} | < 1 \\ w(B_{S_j}, X_{T_j}), \\ \quad \text{if } B_{S_j} \Rightarrow X_{T_j} > 2 \end{cases}$$

=



complex support

densities

(*unnormalized*)

$\text{Pr}_\Delta(\mathbf{X}, \mathbf{B})$

Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

SMT + densities = Weighted Model Integration

Given an SMT($\mathcal{LRA}$) formula Δ over continuous vars $\mathbf{X}$ and discrete ones $\mathbf{B}$, and weight function $\mathcal{W}$, the ***weighted model integral*** (WMI) is

$$\text{WMI}(\Delta, \mathcal{W}; \mathbf{X}, \mathbf{B}) \triangleq \sum_{\boldsymbol{b} \in \mathbb{B}^{|\mathbf{B}|}} \int_{(\mathbf{x}, \boldsymbol{b}) \models \Delta} w(\mathbf{x}, \boldsymbol{b})\, d\mathbf{x}.$$

Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

SMT + densities = Weighted Model Integration

Given an SMT($\mathcal{LRA}$) formula Δ over continuous vars $\mathbf{X}$ and discrete ones $\mathbf{B}$, and weight function $\mathcal{W}$, the ***weighted model integral*** (WMI) is

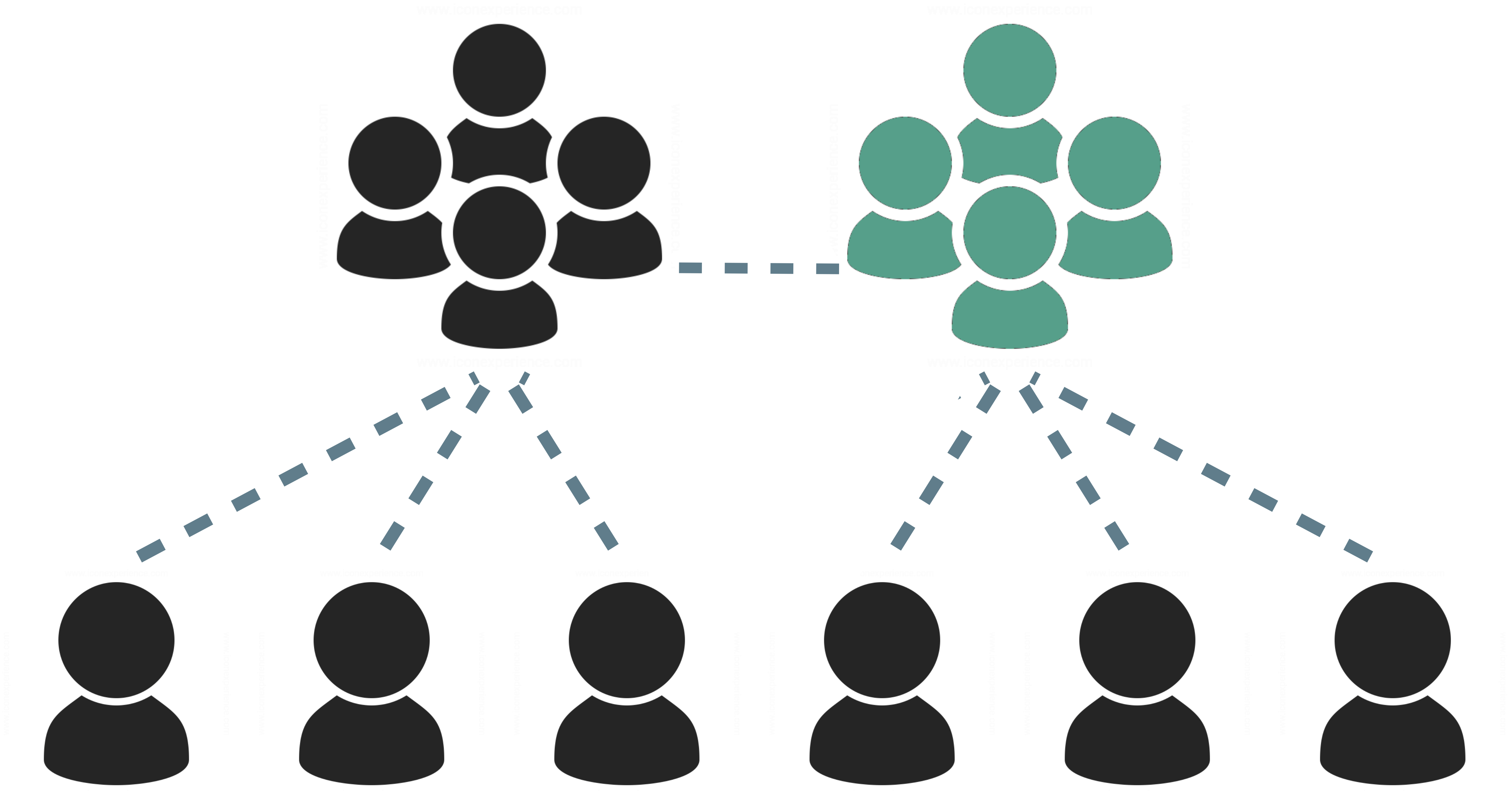
$$\text{WMI}(\Delta, \mathcal{W}; \mathbf{X}, \mathbf{B}) \triangleq \sum_{\boldsymbol{b} \in \mathbb{B}^{|\mathbf{B}|}} \int_{(\mathbf{x}, \boldsymbol{b}) \models \Delta} w(\mathbf{x}, \boldsymbol{b}) \, d\mathbf{x}.$$

$\Longrightarrow$ *integrating the* ***densities*** *of the* ***feasible regions*** *of* Δ*!*

i.e., computing the ***partition function*** of the unnormalized distribution $\Pr_{\Delta}$

Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

SMT + **weights** = **Weighted Model Integration**



"What is the probability of team T_1 outperforming team T_2, if T_1 is a squad but T_2 is not?"

Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

WMI

Advanced probabilistic reasoning

"What is the probability of team T_1 outperforming team T_2, if T_1 is a squad but T_2 is not?"

Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

WMI

Advanced probabilistic reasoning

$$\Phi_S : (B_{S_1} = 1 \land B_{S_2} = 0) \implies T_1 \text{ is a squad, } T_2 \text{ is not}$$

$$\Phi_T : (X_{T_1} > X_{T_2}) \implies T_1 \text{ outperforms } T_2$$

Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

WMI

Advanced probabilistic reasoning

$$\Phi_S : (B_{S_1} = 1 \wedge B_{S_2} = 0) \implies T_1 \text{ is a squad, } T_2 \text{ is not}$$

$$\Phi_T : (X_{T_1} > X_{T_2}) \implies T_1 \text{ outperforms } T_2$$

$$\mathsf{Pr}_{\Delta}(\Phi_T \mid \Phi_S) = \frac{\mathsf{WMI}(\Delta \wedge \Phi_T \wedge \Phi_S, \mathcal{W})}{\mathsf{WMI}(\Delta \wedge \Phi_S, \mathcal{W})} = \frac{4,206}{7,225} \approx 58.22\%$$

$\Longrightarrow$ *conditional probabilities as a ratio of two weighted model integrals*

Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

Outline

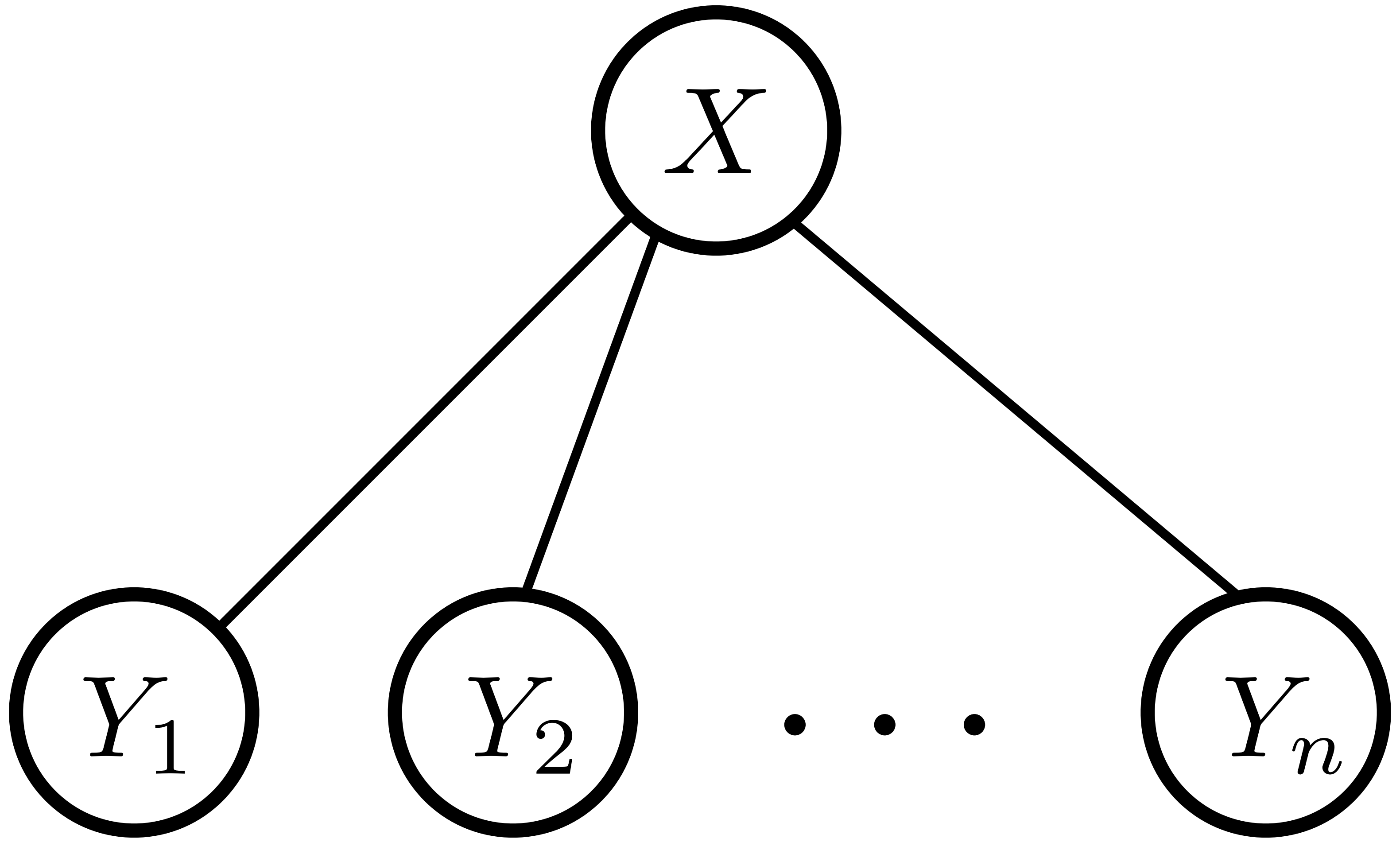
- Definition of Weighted Model Integration (WMI)
- **Tractability Analysis of WMI Problem Class**
- Application in Bayesian Deep Learning

WMI Solvers

- Predicate abstraction *[Morettin et al. 2017, 2019]*
- Compilation-based (XADD, XSDD) *[Kolb et al. 2018; Zuidberg Dos Martires et al. 2019]*
- Search-based (And/Or Search) *[Zeng et al. 2019]*
- Message passing *[Zeng et al. 2020]*

`https://github.com/weighted-model-integration/pywmi`

Search-based WMI Solver



Context-specific independence!

One n+1-dim integration *O(exp(n))*

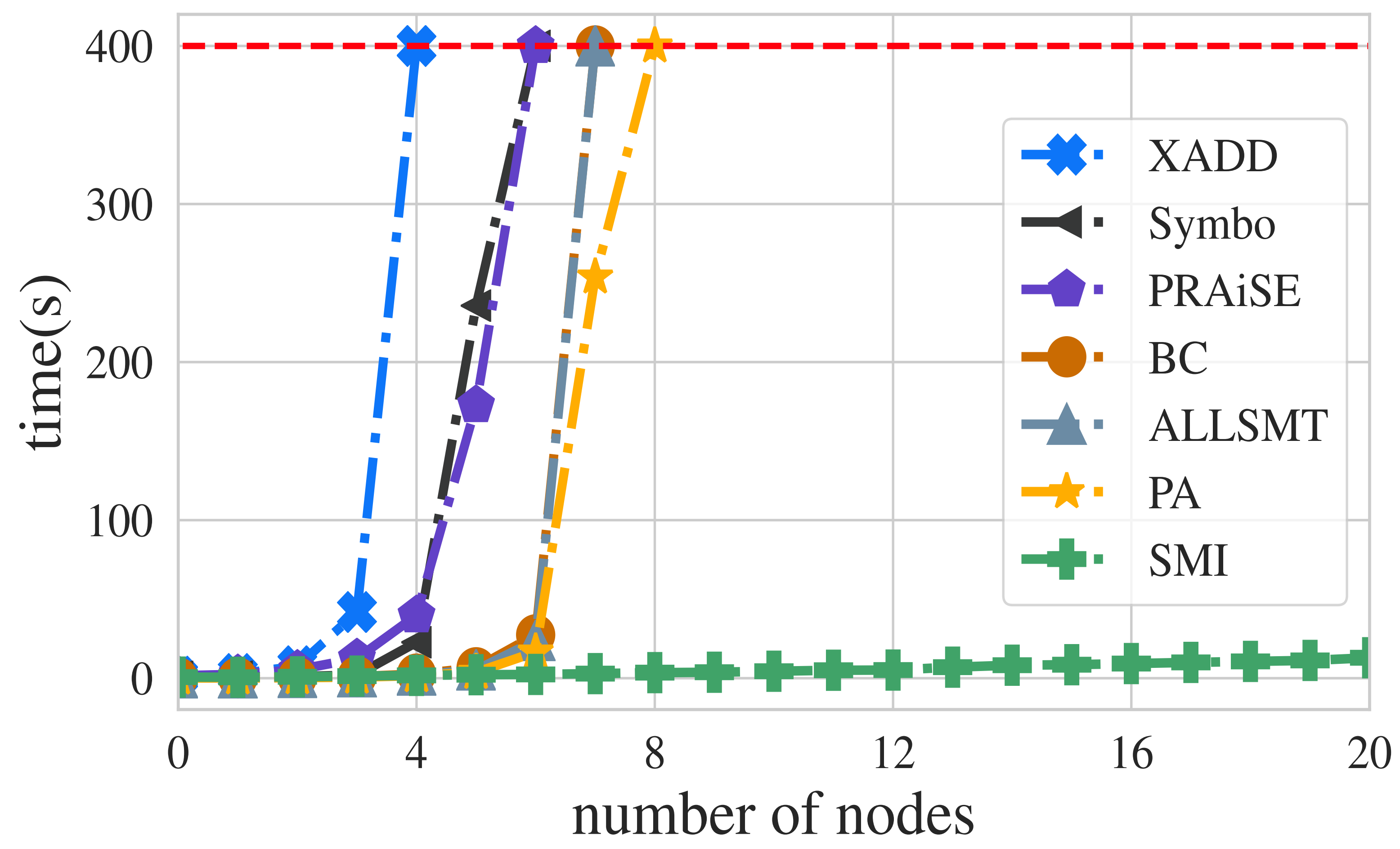
$$\int f(x) \prod_{i=1}^{n} f_i(x, y_i)\, dy_1, \ldots, y_n, x$$

n+1 one-dim integration *O(n)*

$$\int f(x) \prod_{i=1}^{n} \left(\int_{l_i(x)}^{u(x)} f_i(x, y_i)\, dy_i \right) dx$$

Zeng et al., "Efficient Search-Based Weighted Model Integration", 2019

Search-based WMI Solver



One n+1-dim integration *O(exp(n))*

$$\int f(x) \prod_{i=1}^{n} f_i(x, y_i)\, dy_1, \ldots, y_n, x$$

n+1 one-dim integration *O(n)*

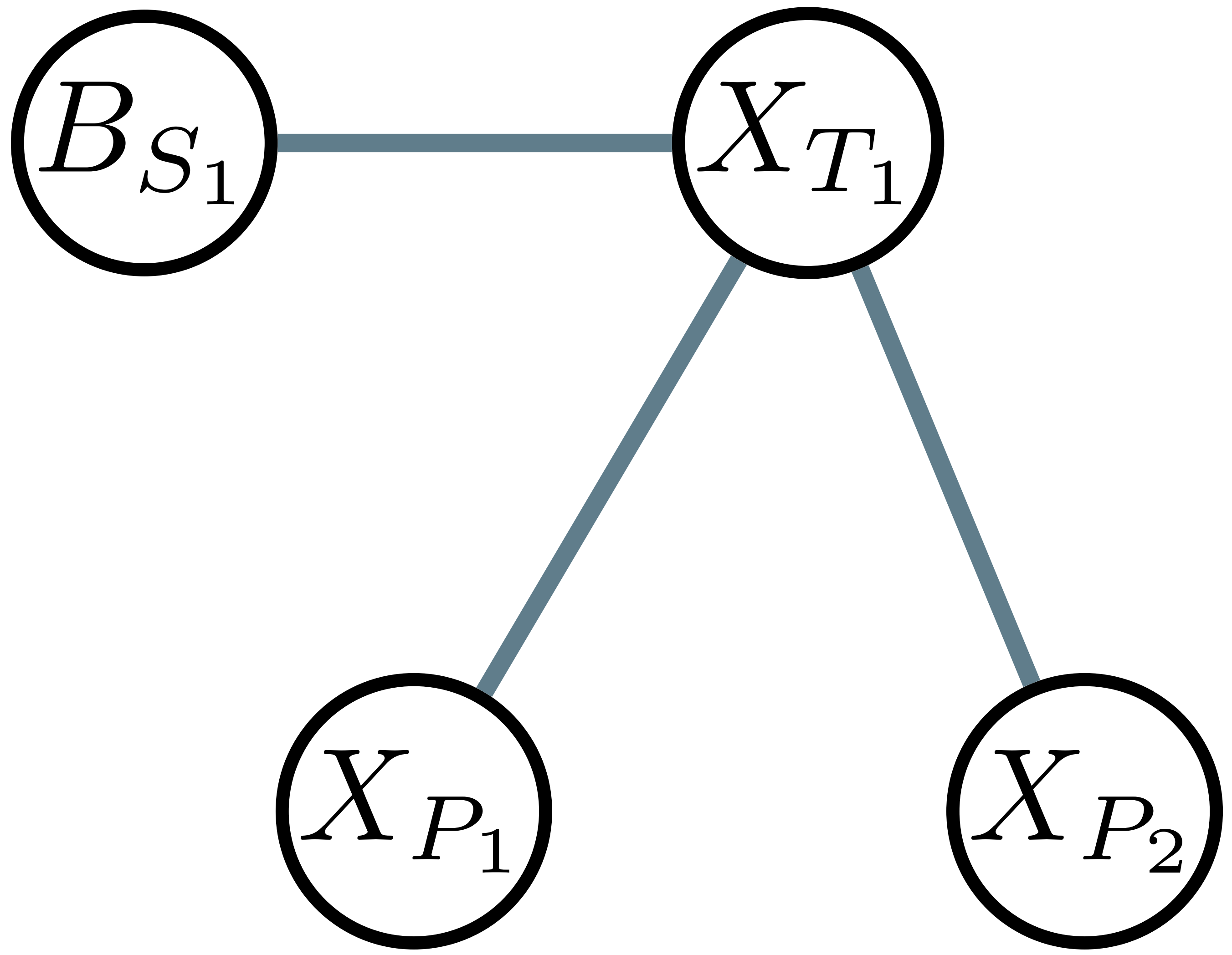
$$\int f(x) \prod_{i=1}^{n} \left(\int_{l_i(x)}^{u(x)} f_i(x, y_i)\, dy_i \right) dx$$

Zeng et al., "Efficient Search-Based Weighted Model Integration", 2019

MP-WMI

We frame tractable WMI inference at scale as a ***message passing*** scheme…

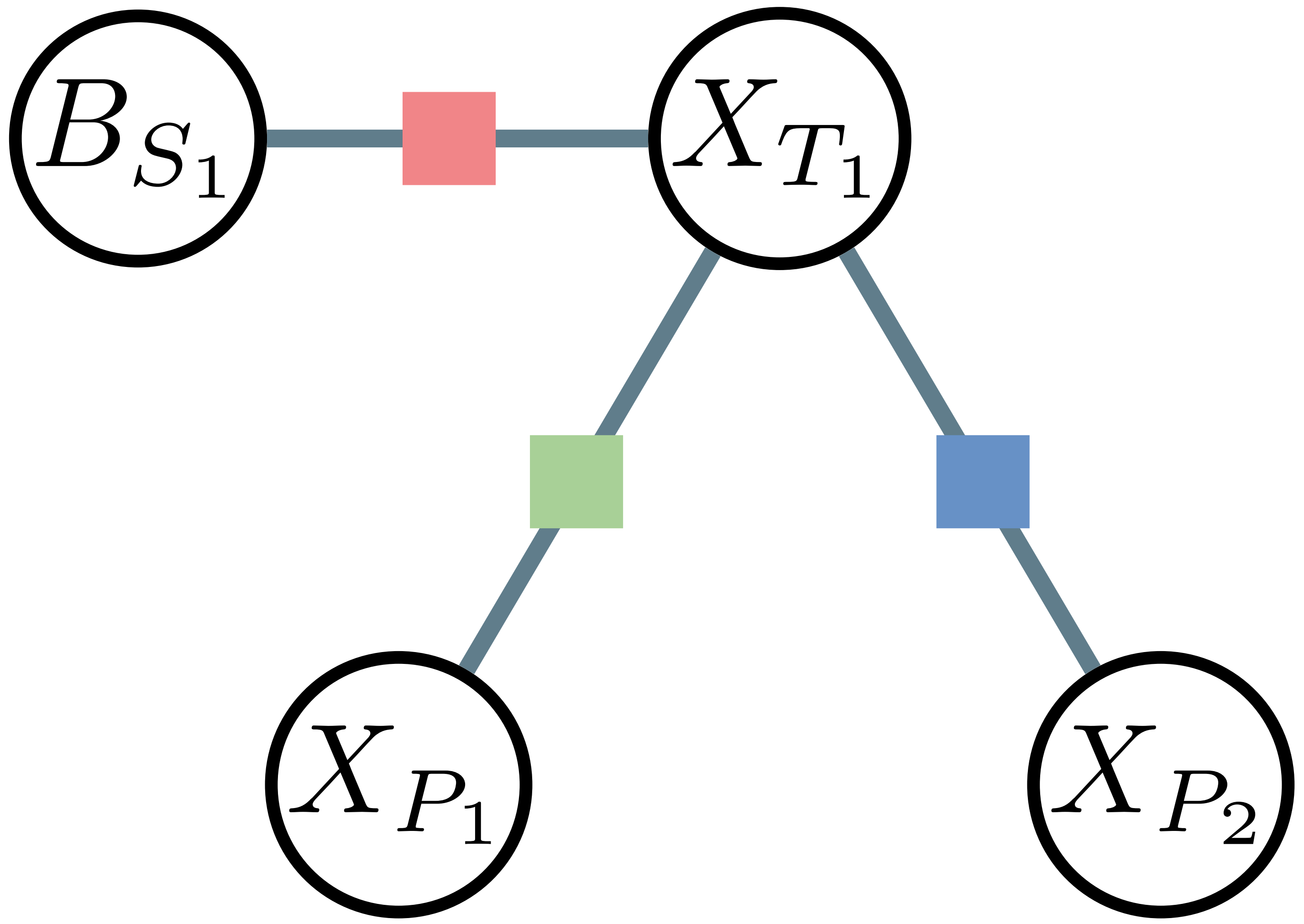
…on primal graphs…



MP-WMI

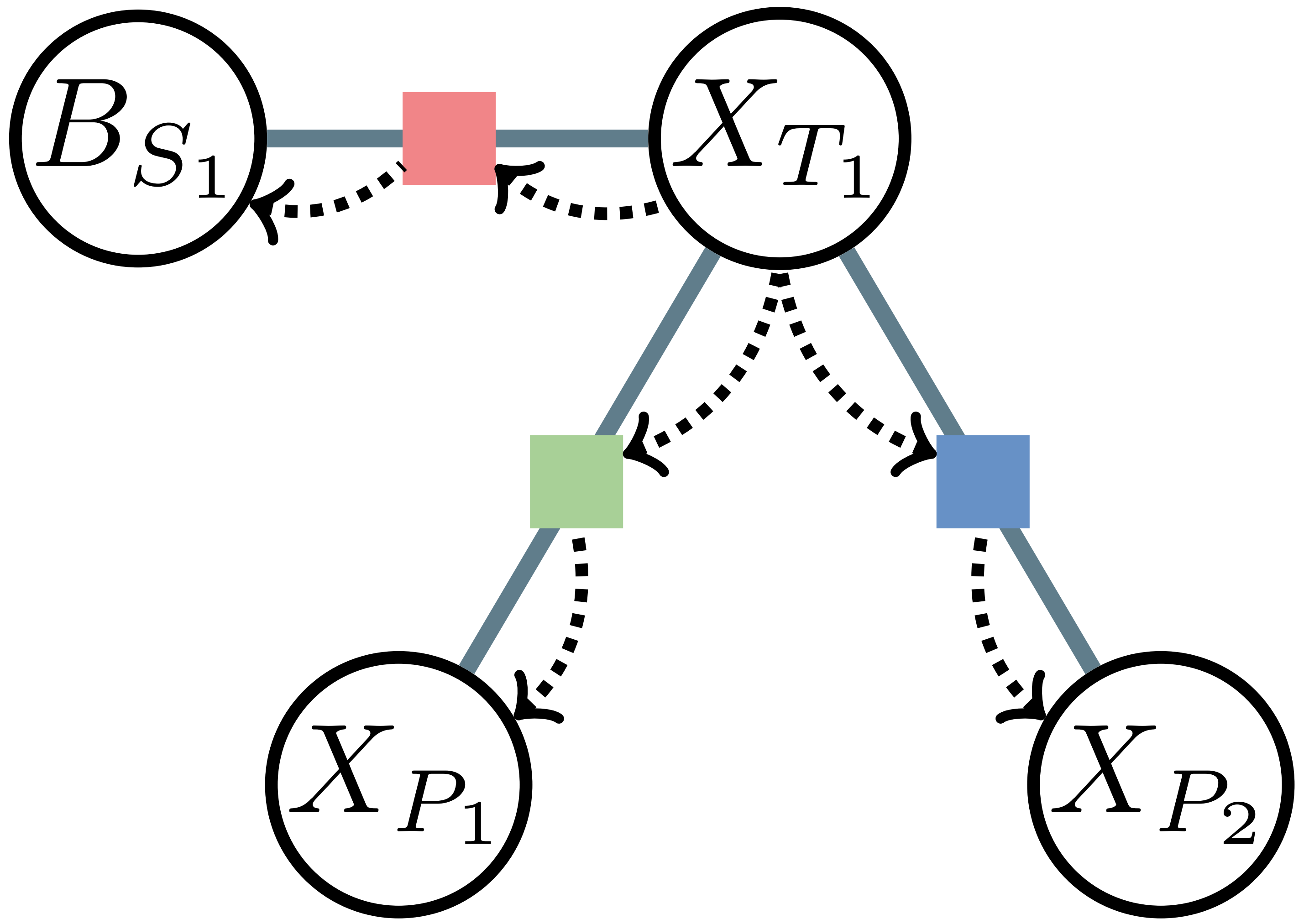
We frame tractable WMI inference at scale as a ***message passing*** scheme...

...on primal graphs turned into ***factor graphs***



MP-WMI

We frame tractable WMI inference at scale as a ***message passing*** scheme...

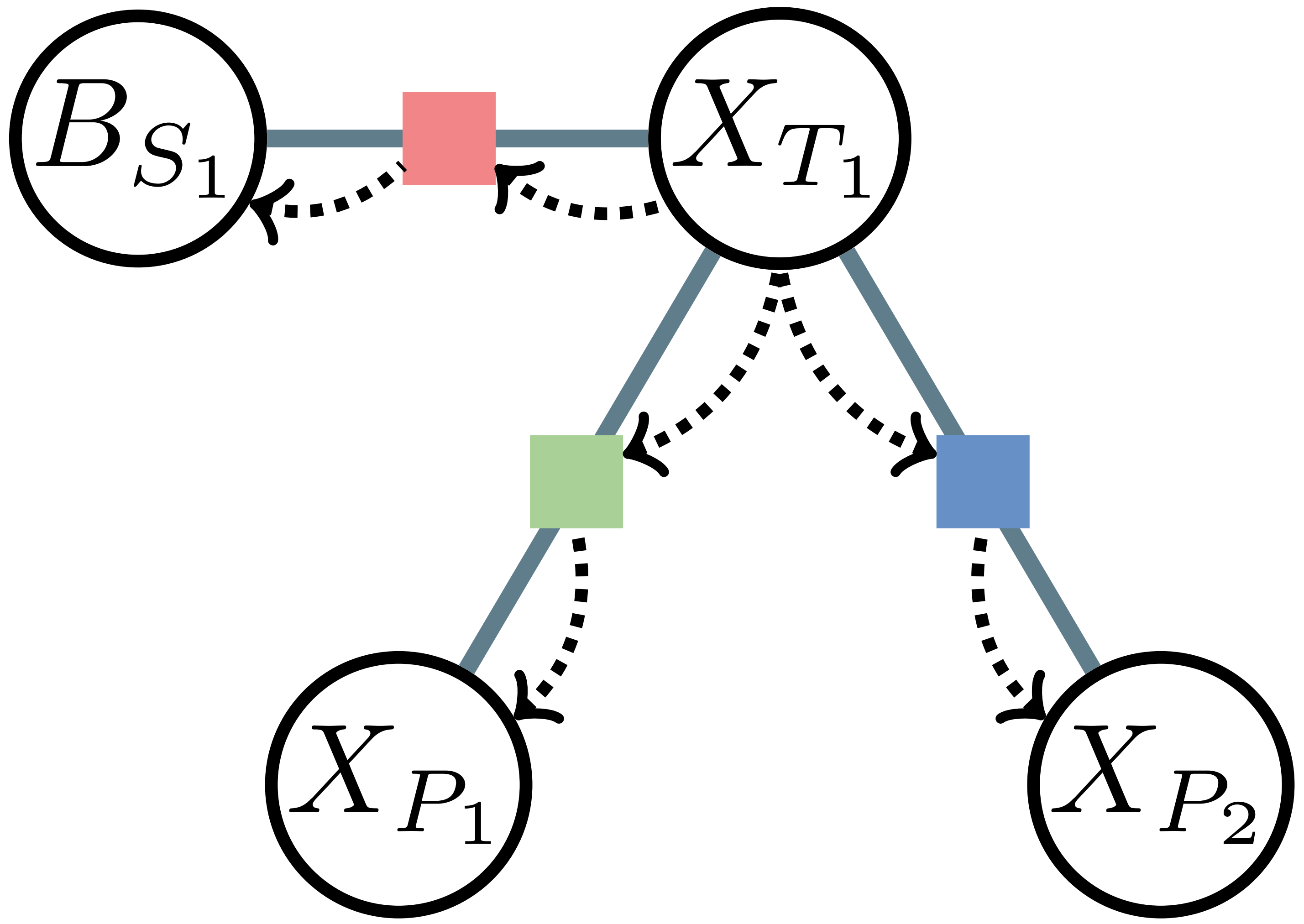


...on primal graphs turned into ***factor graphs***

- comprising an ***upward*** and a ***downward*** pass

MP-WMI

We frame tractable WMI inference at scale as a ***message passing*** scheme…



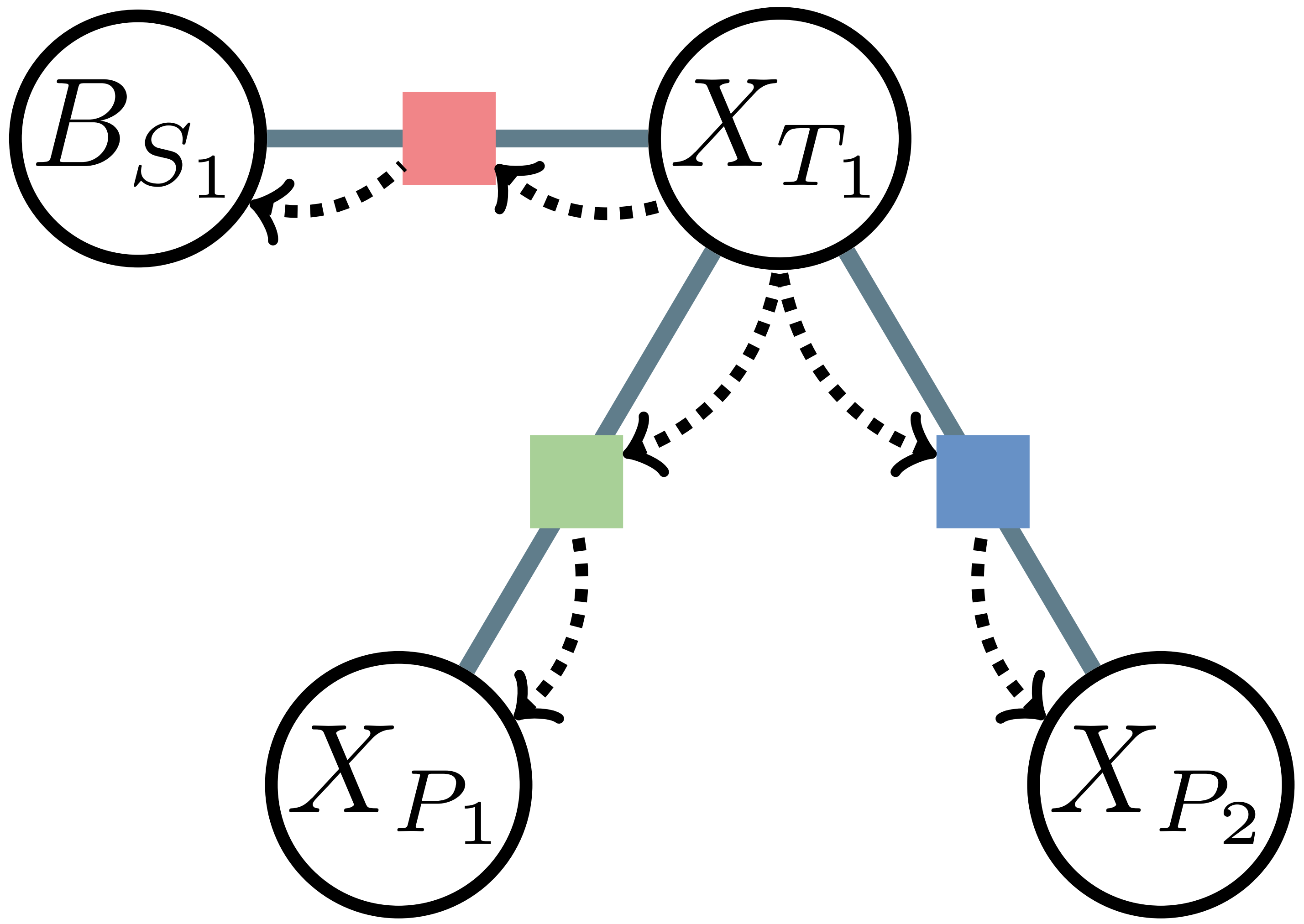
…on primal graphs turned into ***factor graphs***

- comprising an ***upward*** and a ***downward*** pass
- exchanging messages from ***node to factors***

$$\mathsf{m}_{x_i \to f_{\mathcal{S}}}(x_i) = \prod\nolimits_{f_{\mathcal{S}'} \in \mathsf{neigh}(x_i) \setminus f_{\mathcal{S}}} \mathsf{m}_{f_{\mathcal{S}'} \to x_i}(x_i)$$

MP-WMI

We frame tractable WMI inference at scale as a ***message passing*** scheme...



...on primal graphs turned into ***factor graphs***

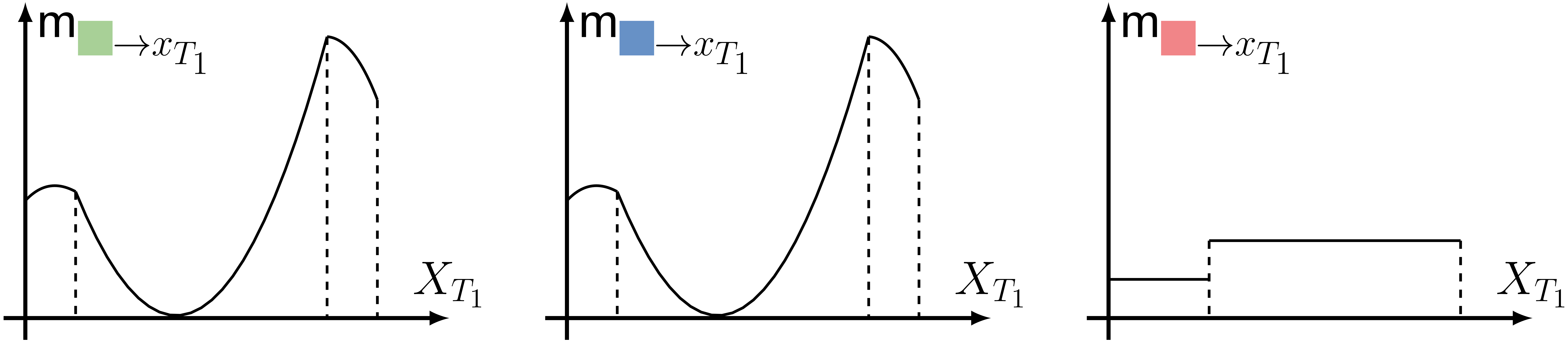
- comprising an ***upward*** and a ***downward*** pass
- exchanging messages from ***node to factors***
- and from ***factors to nodes***

$$\mathsf{m}_{f_{ij}\to x_i}(x_i) = \int f_{ij}(x_i, x_j) \cdot \mathsf{m}_{x_j \to f_{ij}}(x_j)\, dx_j$$

MP-WMI

An SMT formulation induces a ***piecewise weight representation***

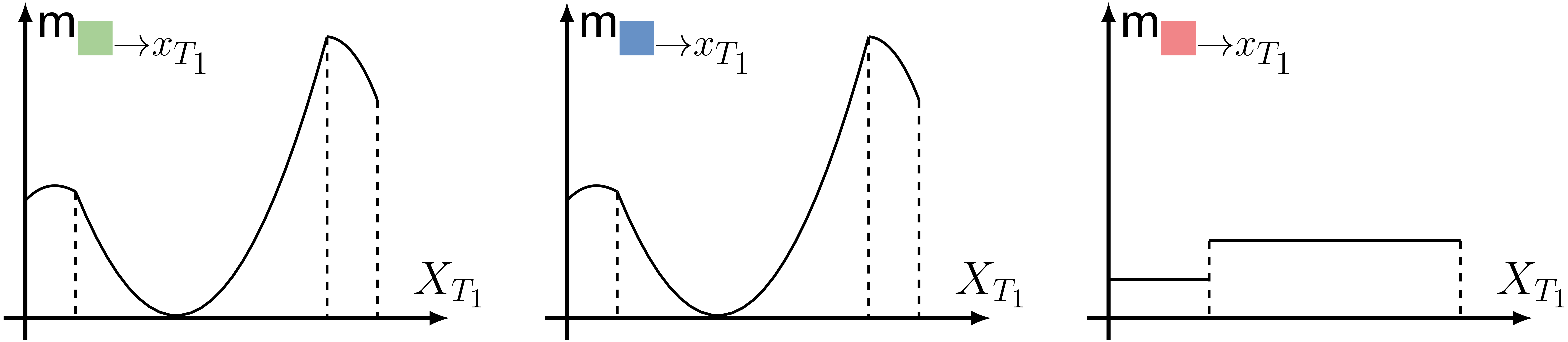
$\Rightarrow$ *strikingly different from message passing for classical PGMs!*



MP-WMI

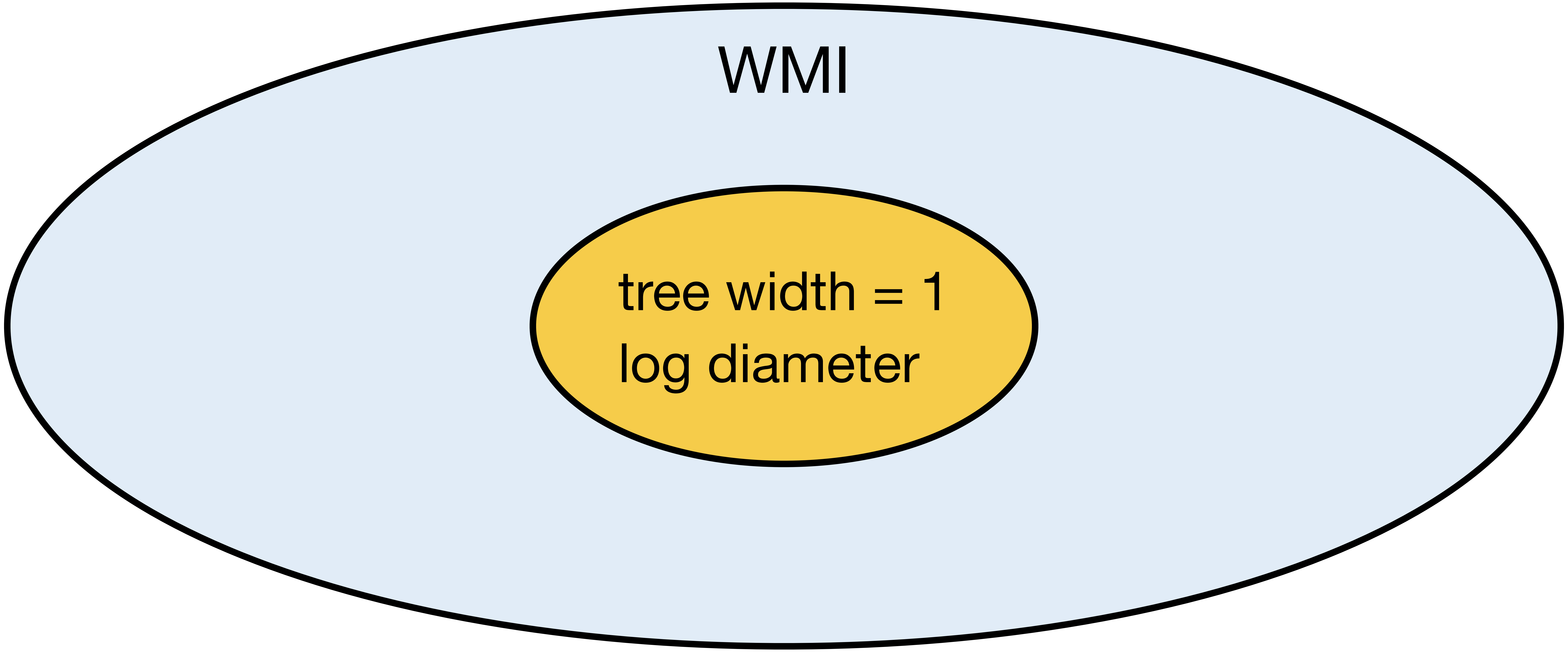
An SMT formulation induces a **_piecewise weight representation_**

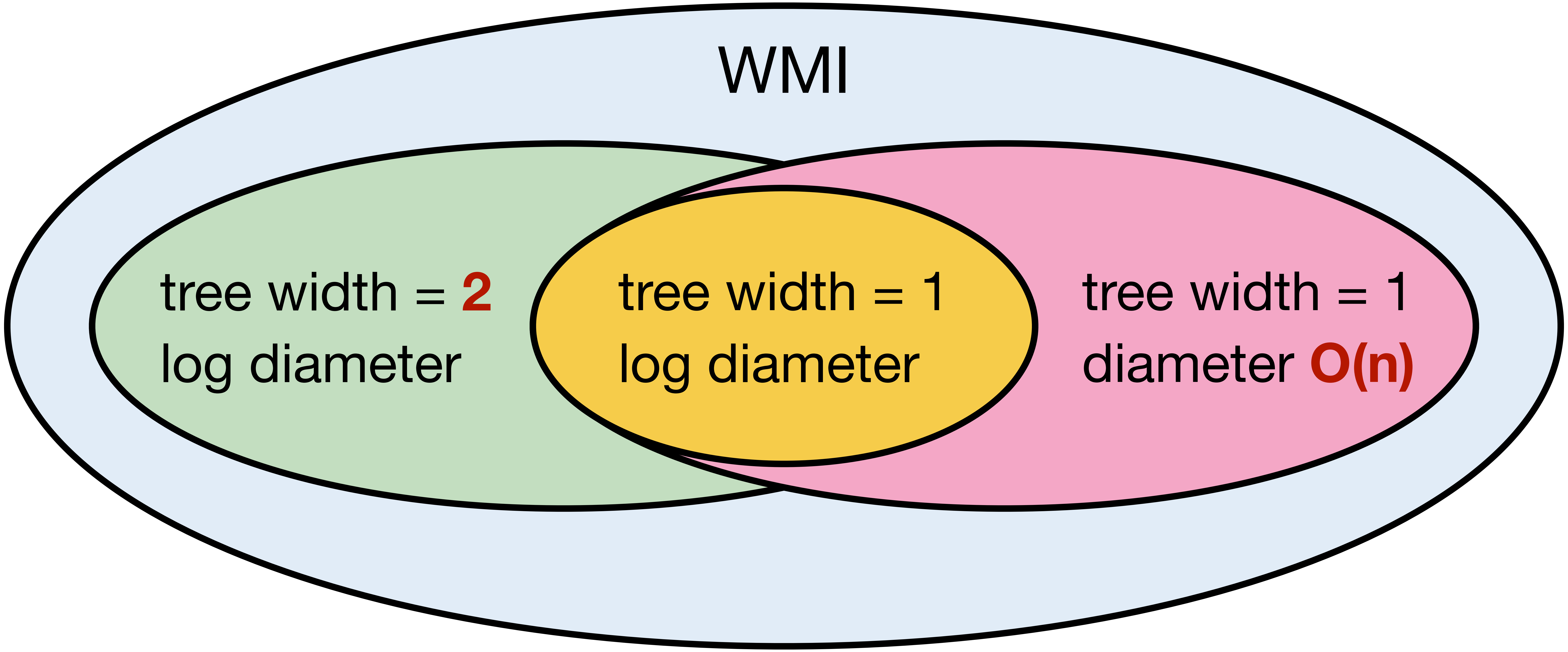
$\Rightarrow$ *strikingly different from message passing for classical PGMs!*



The number of all pieces in MP-WMI is $\mathcal{O}(4nc)^{2d+2}$, where d is the graph diameter

$\Rightarrow$ *the primal graph should have a* ***bounded diameter!***


WMI
tree width = 1
log diameter



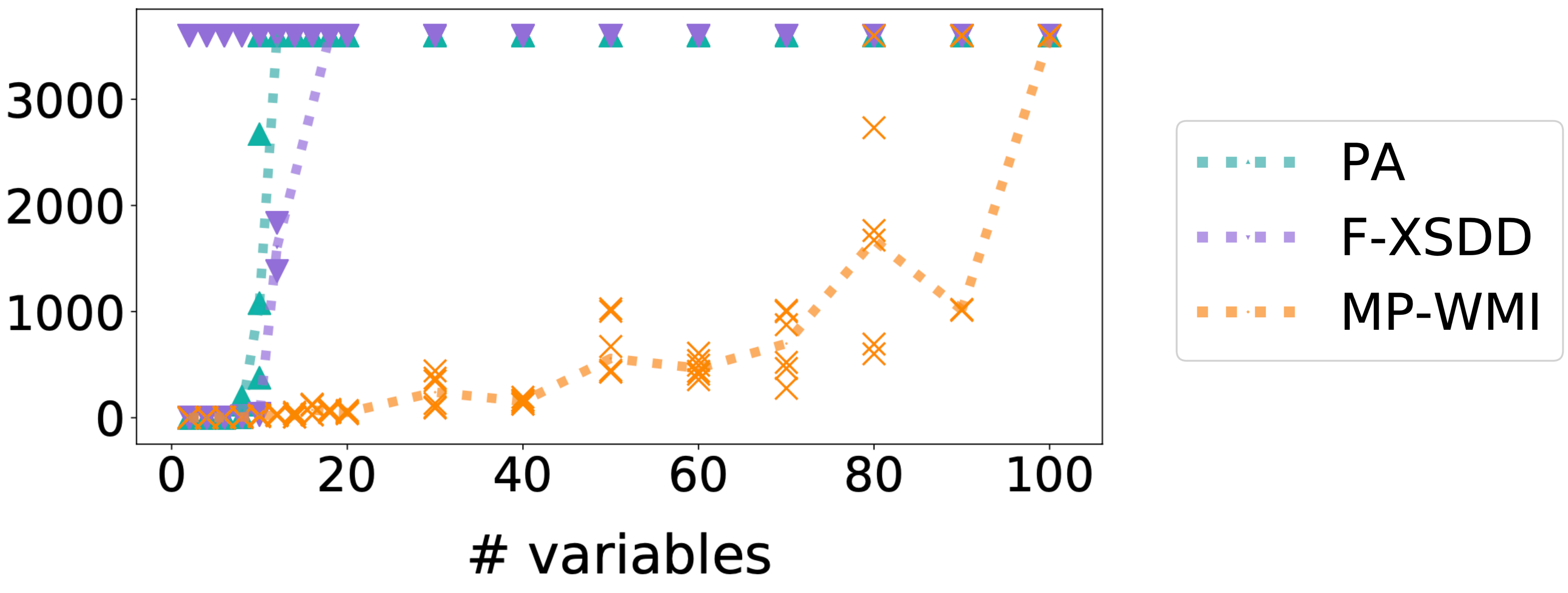
Theorem *inference in* $\mathcal{WMI}(\mathbf{\Omega}, \log(n), 2)$ *is #P-hard.*

Theorem *inference in* $\mathcal{WMI}(\mathbf{\Omega}, n, 1)$ *is #P-hard.*

Z. Zeng et al., "Probabilistic Inference with Algebraic Constraints: Theoretical Limits and Practical Approximations", 2020

Open Questions

- What circuit representation can we have for WMI?
- How to approximate WMI?



Outline

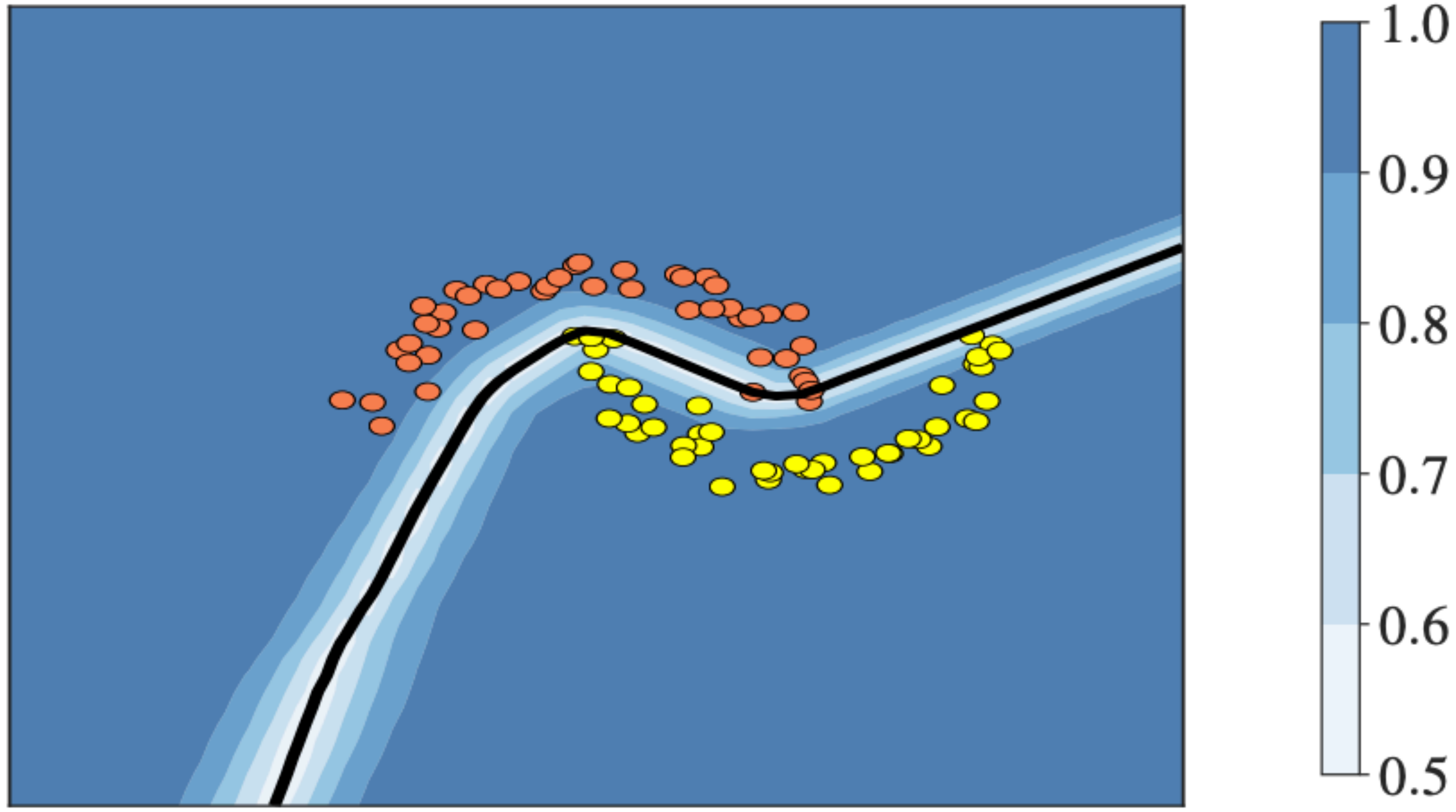
- Definition of Weighted Model Integration (WMI)
- Tractability Analysis of WMI Problem Class
- **Application in Bayesian Deep Learning**

Collapsed inference for Bayesian deep learning

Zhe Zeng and Guy Van den Broeck. Collapsed inference for Bayesian deep learning. NeurIPS, 2023.

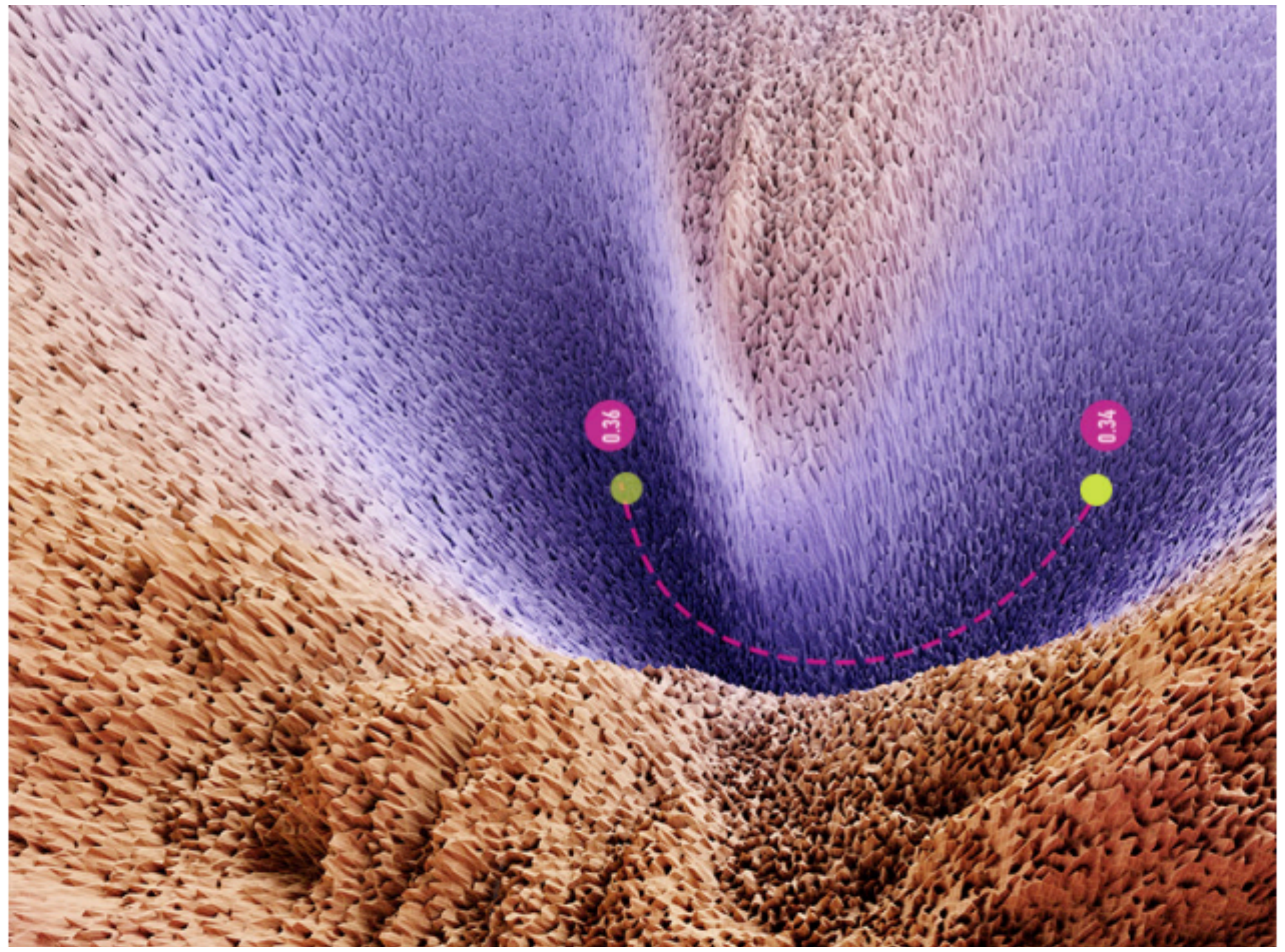
Motivation

Bad Uncertainty Estimation



Confidence by a ReLU neural network [1]

Risky Point Estimation



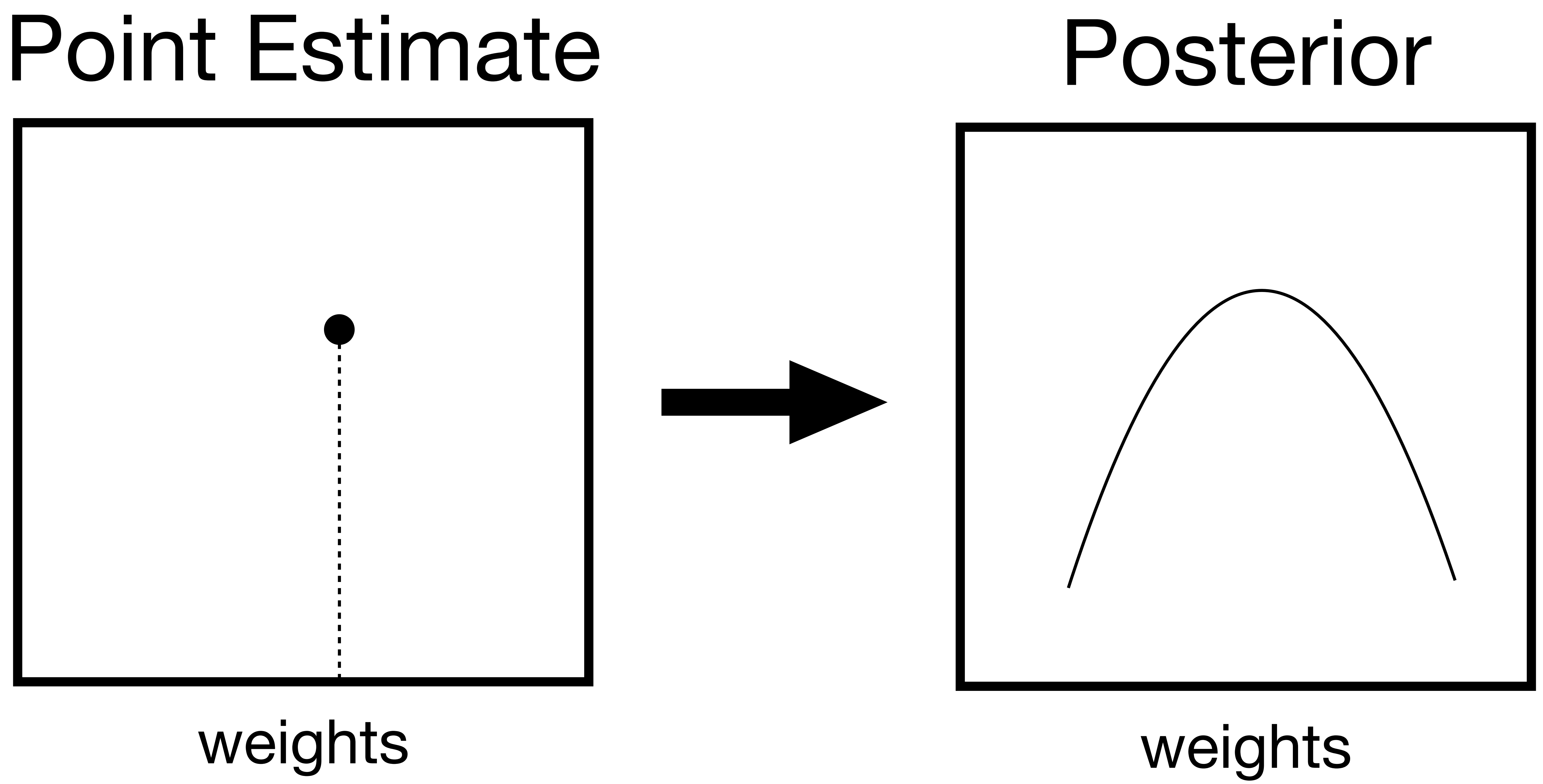
Loss surface [2]

➡**Bayesian Deep Learning** for *robust* and *reliable* predictions

[1] Kristiadi, Agustinus, Matthias Hein, and Philipp Hennig. "Being bayesian, even just a bit, fixes overconfidence in relu networks." International conference on machine learning. PMLR, 2020.
[2] Garipov, Timur, et al. "Loss surfaces, mode connectivity, and fast ensembling of dnns." Advances in neural information processing systems 31 (2018).

Bayesian Model Average (BMA)

Key idea



$$p(y \mid x, w) \qquad\qquad p(y \mid x) = \int p(y \mid x, w)\, p(w)\, dw$$

Motivation

- **Goal:** Bayesian model average

  Predictive posterior $p(y \mid x) = \int p(y \mid x, w) p(w) \, dw$

  Expected prediction $\mathbb{E}[y] = \int y \; p(y \mid x) \, dy$

- **Challenge:** DNNs are too big!

  ➡ *Costly* to maintain too many samples

  ➡ *Low sample efficiency* given the complex integrand

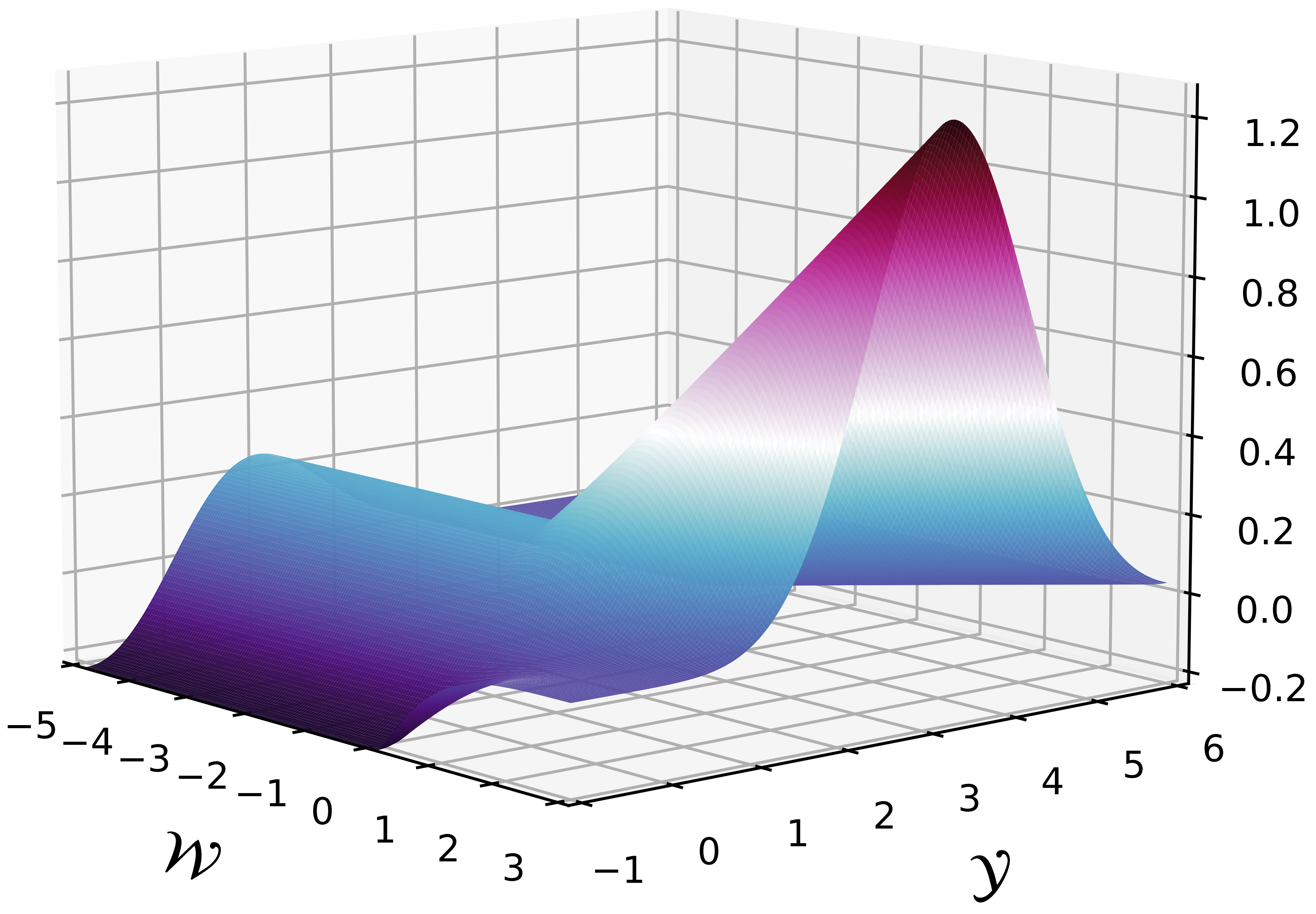
How complex? 🤔

How *complex* is the integrand? Simplest case!

Expected prediction $\mathbb{E}[y] = \int y \, \underline{p(w \mid D)} \, \underline{p(y \mid \underline{f(x)}, w)} \, dw \, dy$

Uniform Gaussian

Single ReLU



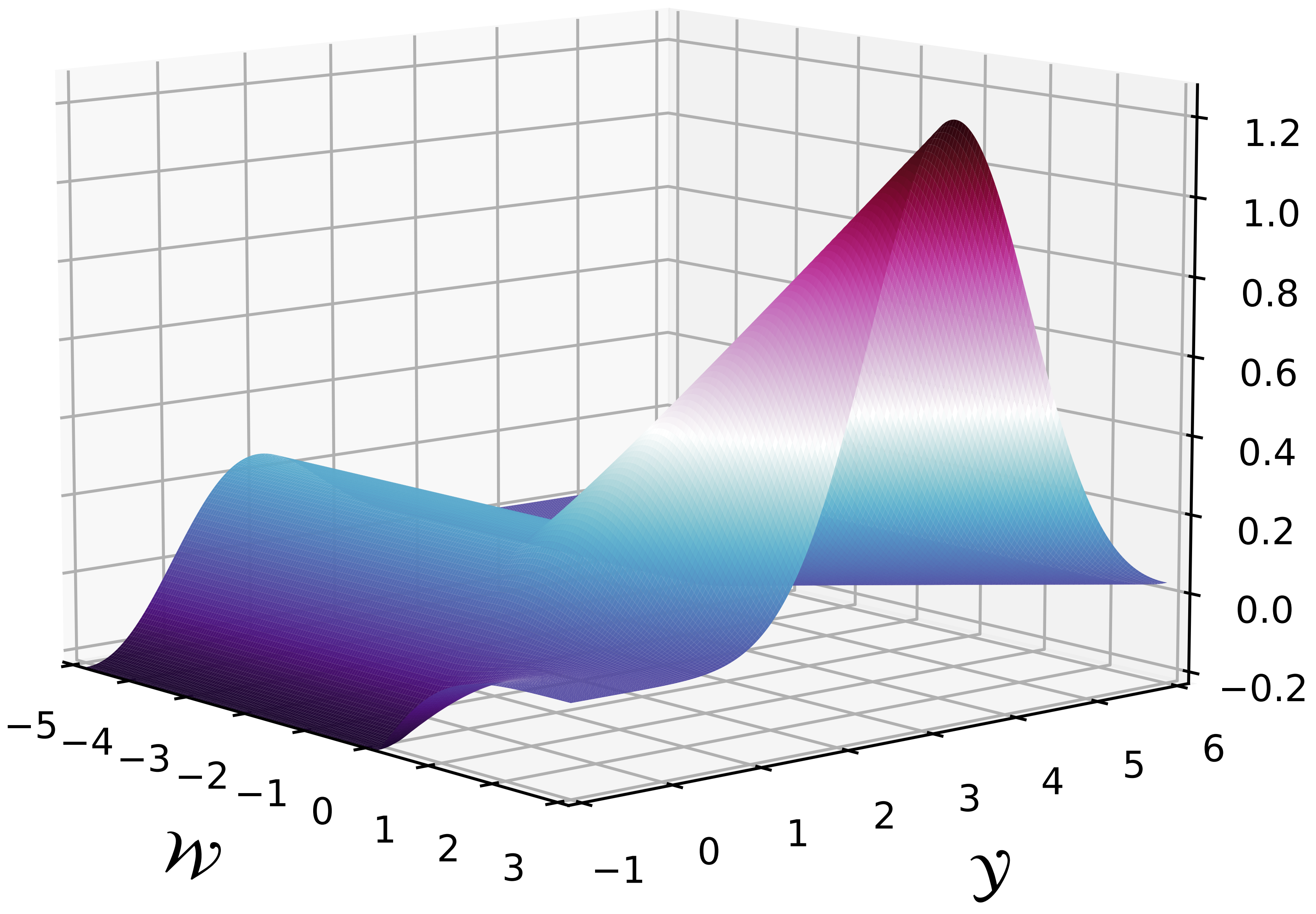
Non-convex, multi-modal, no closed form 🤯

Motivation

- **Goal:** Bayesian model average

Predictive posterior $p(y \mid x) = \int p(y \mid x, w) p(w) \, dw$

Expected prediction $\mathbb{E}[y] = \int y \; p(y \mid x) \, dy$

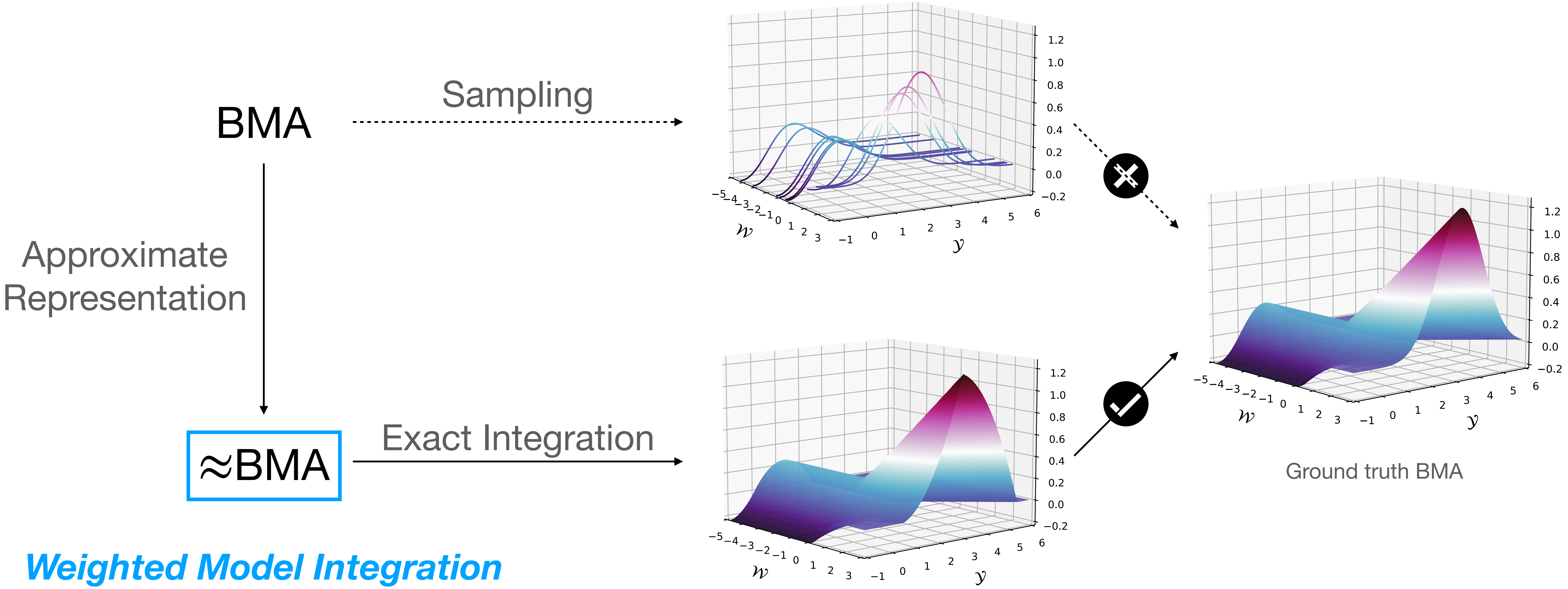


Is there a *better* way to estimate the integral than *sampling*?

Yes! 🤩

Idea

A reduction from BMA to Exact Integration

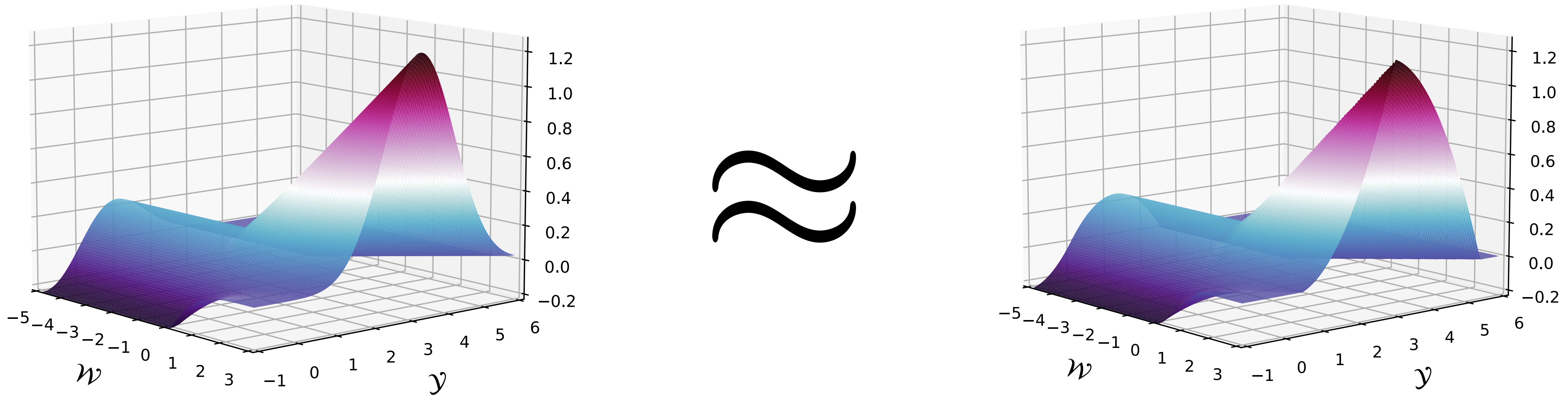


Idea

A reduction from BMA to ~~Exact Integration~~ *WMI*

ReLU NN $\xrightarrow{=}$ SMT formula

Posterior $\xrightarrow{\approx}$ Weight functions



Accurate approximation!

… but scalability?

Limitations

	Sampling	BMA via WMI
Accuracy	❌	✅
Flexibility	✅	❌*
Scalability	✅	❌**

* Limited to fully connected layers

** Integration over polytopes in arbitrarily high dimensions is #P-hard

How to combine good from both worlds? 🤔

Limitations

	Sampling	BMA via WMI	Collapsed Inference
Accuracy	❌	✅	✅
Flexibility	✅	❌*	✅
Scalability	✅	❌**	✅

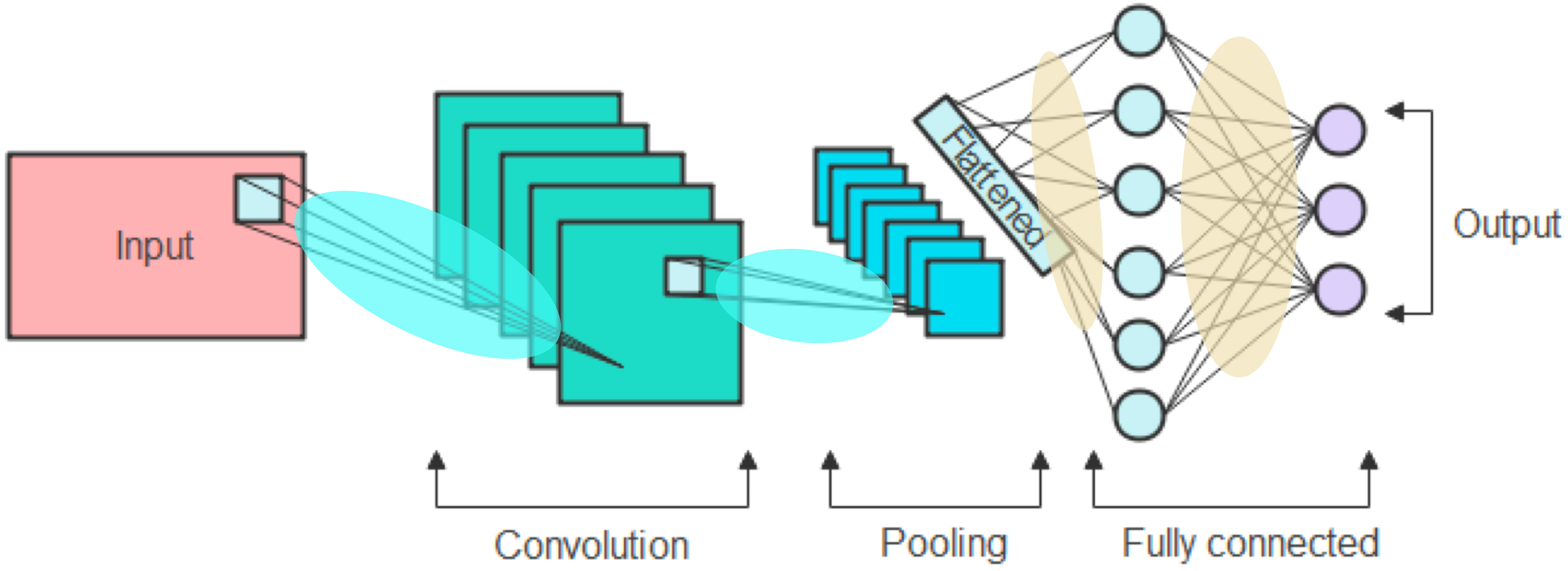
* Limited to fully connected layers

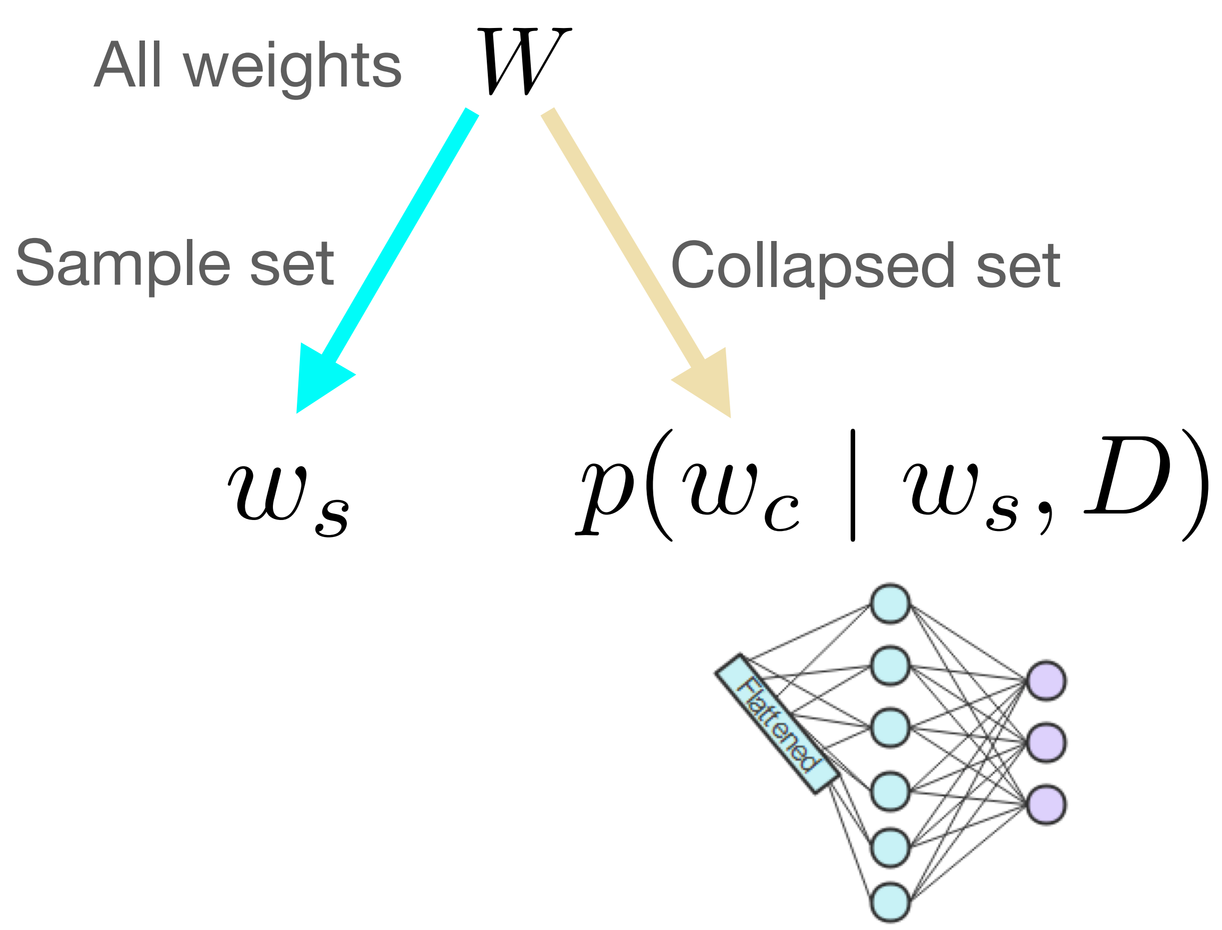
** Integration over polytopes in arbitrarily high dimensions is #P-hard

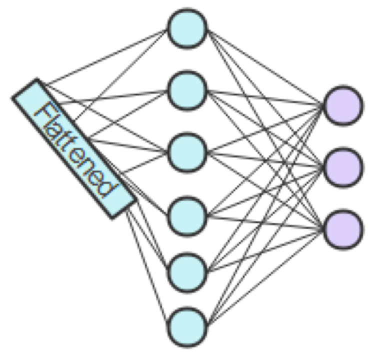
How to combine good from both worlds? 🤔

➡ *Collapsed inference scheme!* 💪

Collapsed Inference[5]





Expected prediction in BMA $\mathbb{E}[y] = \frac{1}{n} \sum_{w_s} \texttt{WMI}($  $)$

Accuracy + Flexibility, Scalability! 🥳

Image Classification

Metric	NLL		ACC		ECE	
Dataset	CIFAR-10	CIFAR-100	CIFAR-10	CIFAR-100	CIFAR-10	CIFAR-100
CIBER	**0.1927 ± 0.0029**	**0.9193 ± 0.0027**	**93.64 ± 0.09**	**74.71 ± 0.18**	0.0130 ± 0.0011	**0.0168 ± 0.0025**
SWAG	0.2503 ± 0.0081	1.2785 ± 0.0031	93.59 ± 0.14	73.85 ± 0.25	0.0391 ± 0.0020	0.1535 ± 0.0015
SGD	0.3285 ± 0.0139	1.7308 ± 0.0137	93.17 ± 0.14	73.15 ± 0.11	0.0483 ± 0.0022	0.1870 ± 0.0014
SWA	0.2621 ± 0.0104	1.2780 ± 0.0051	93.61 ± 0.11	74.30 ± 0.22	0.0408 ± 0.0019	0.1514 ± 0.0032
SGLD	0.2001 ± 0.0059	0.9699 ± 0.0057	93.55 ± 0.15	74.02 ± 0.30	**0.0082 ± 0.0012**	0.0424 ± 0.0029
KFAC	0.2252 ± 0.0032	1.1915 ± 0.0199	92.65 ± 0.20	72.38 ± 0.23	0.0094 ± 0.0005	0.0778 ± 0.0054

- Our approach is applicable to large NNs
- It achieves accurate estimations of uncertainty
- It boosts predictive performance

Conclusion

- WMI can be a powerful tool for many!

Open Questions

- What circuit representation can we have for WMI?
- How to approximate WMI?
- ……

Thank you!