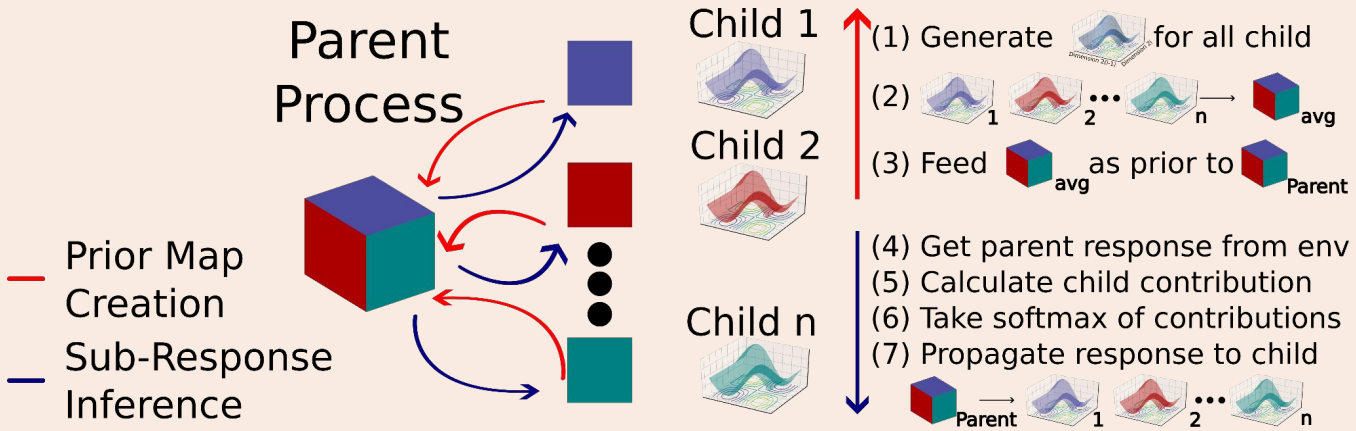
Bidirectional Information Flow (BIF): A Novel Hierarchical Gaussian Process for Bayesian Optimization

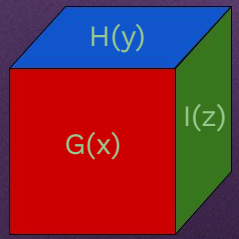
Juan D. Guerra, Thomas Garbay,
Numa Dancause, Guillaume Lajoie, Marco Bonizzato



https://classroomclipart.com/image/vector-clipart/baseball-pitcher-throwing-ball-clipart-10060.htm



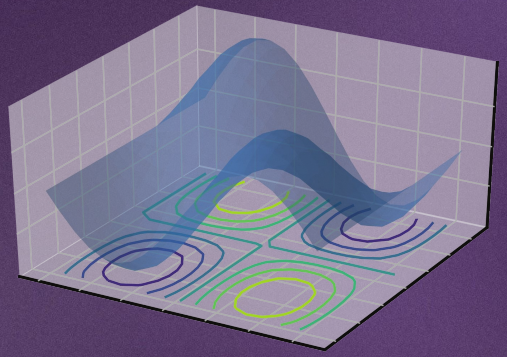
$$F(x, y, z) = \phi[G(x), H(y), I(z)]$$



Problem Statement

- **Objective**: learn how to pitch
  - This is an optimization problem

- Pitching is one motion
- Can state this formally as a function
- Break down pitch into three components
  - Wrist
  - Elbow
  - Shoulder
- Each component interacts to produce a throw



Gaussian Process Bayesian Optimization

- Bayesian Optimization (BO) provides a sample efficient optimization method
- Maximize or minimize a target function through trial-and-error
- Gaussian processes (GP) are highly expressive functions
    - Provide mean and uncertainty predictions

Prior Work

Additive GP

- Arbitrary number of children
- Structural bias baked in
- Outputs are strictly added

$$\sum_{i=1}^{d} f_i(x_i)$$

Duvenaud et al., 2011

Deep GP

- Kernels are nested
- Aimed as a GP alternate to NNs
- Highly expressive
- Not interpretable

$$y_{nd} = f_d^Y(\mathbf{x}_n) + \epsilon_{nd}^Y$$
$$x_{nq} = f_q^X(\mathbf{z}_n) + \epsilon_{nq}^X$$

Damianou and Lawrence, 2023

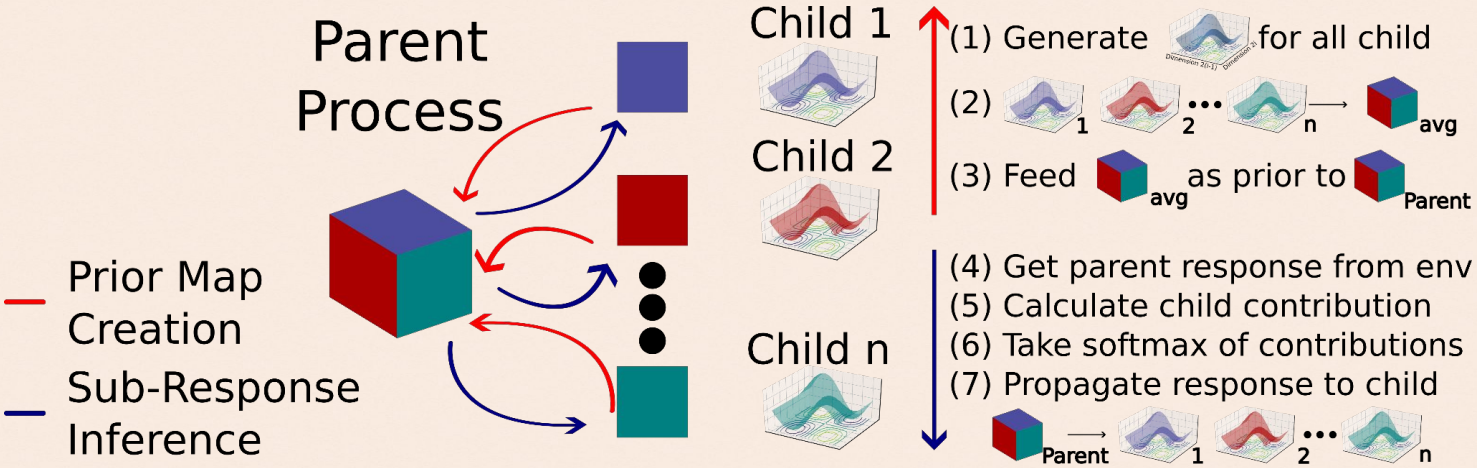
Multi-Task GP

- Use task similarity kernels
- Share information for performance gains
- Require response signals for each subfunction

$$K_{lk}^f \cdot k_x(x, x')$$

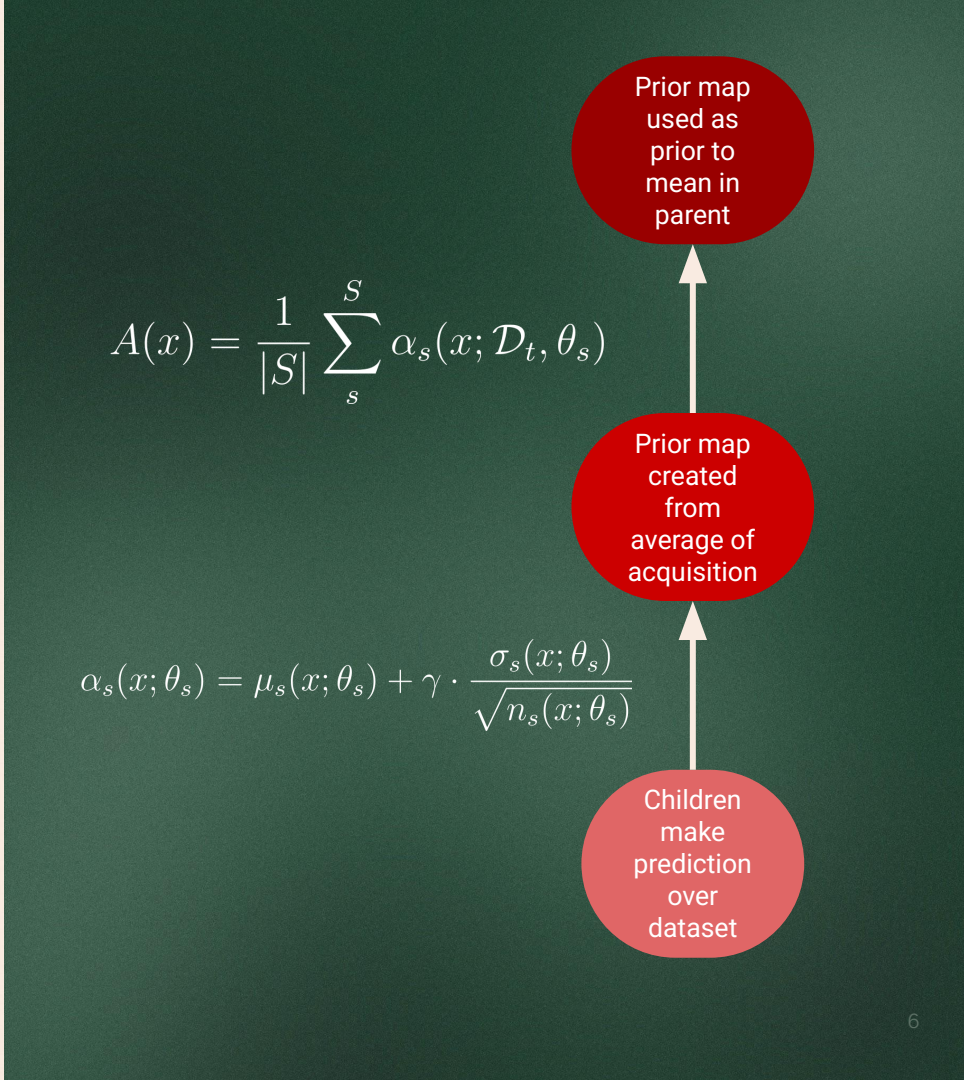
Bonilla et al., 2014

Methodology
Bidirectional Information Flow



Child to Parent Information

- Leverage children's environment knowledge to **boost parent's knowledge**
- Child information encoded as **prior maps**
- Take **average of child beliefs** using **acquisition function** and posterior distribution



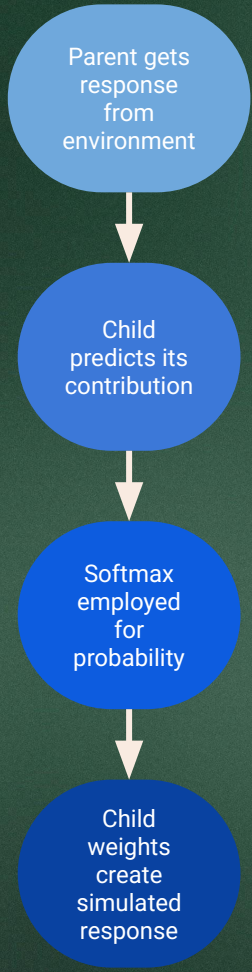
Parent to Child Information

- Want to **decode** the combined signal
- Creates **approximate training points** for children
- Allows us to **train children** without needing to explicitly obtain individual stimuli

$$c_s(x) = \frac{\alpha_s(x)}{\max_{x'} \alpha_s(x')}$$

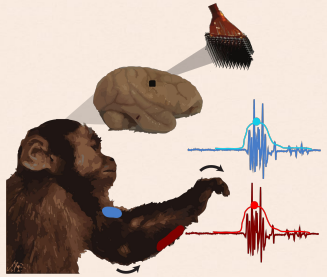
$$w_s(x) = \frac{\exp(c_s(x))}{\sum_{j \in S} \exp(c_j(x))}$$

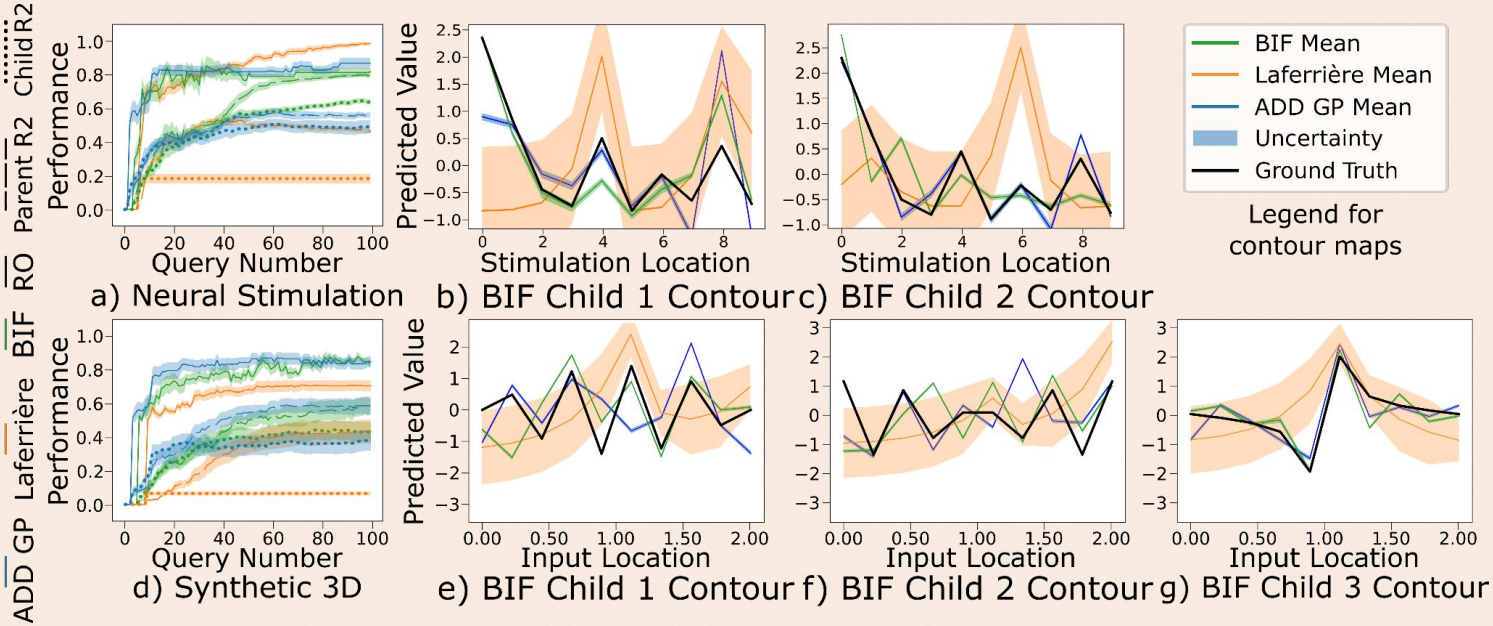
$$y_s(x) = w_s(x) \cdot y_p(x)$$



Results

- BIF 1st or 2nd in 38/40 metrics
- Child R^2 achieved is sufficient for reconstruction





Conclusion

BIF is a robust BO surrogate with applications to high-risk and data scarce problems

01 Sample efficiency proven empirically

02 BIF supported by proven regret guarantees

03 Interpretability improved

Thank you!

Code available at

https://github.com/JuanDGuerra20/Bidirectional-Information-Flow

Or contact us at

juan.guerra@mila.quebec, guillaume.lajoie@mila.quebec, or
marco.bonizzato@polymtl.ca

POLYTECHNIQUE MONTRÉAL
Université de Montréal
Mila
UNIQUE CENTRE
NSERC CRSNG
Fonds de recherche Nature et technologies Québec
New Frontiers in Research Fund
Fonds Nouvelles frontières en recherche
NEURAL INFORMATION PROCESSING SYSTEMS