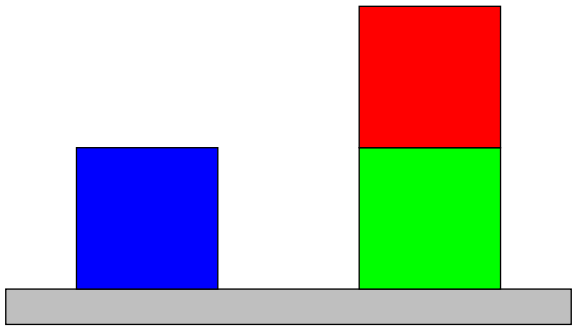
Mechanically Proving Guarantees of Generalized Heuristics: First Results and Ongoing Work

Mohammad Abdulaziz[1, 2]
Florian Pommerening[3] Augusto B. Corrêa[3]

[1]King's College London, United Kingdom,
[2]Technische Universität München, Germany,
[3]University of Basel, Switzerland

GenPlan'22

Description Logics for Planning



Primitive Concepts & Roles

- $ontable = \{$  ,  $\}$
- $on = \{($  ,  $)\}$
- $holding = \emptyset$
- $clear = \{$  ,  $\}$
- $clear_G = \{$  $\}$

Example: Clearing a Block

Definition (Generalized Potential Heuristic)

Linear combination of features well-defined over all tasks:

$$h(s) = \sum_{f \in \mathcal{F}} w(f) \cdot f(s)$$

Generalized Potential Heuristic for Blocksworld

$\rightarrow$ Blocksworld tasks where the goal is to clear a set of blocks

$$h(s) = 2 \cdot |C_1| + |C_2|$$

- $C_1 \equiv \exists on^+.clear_G$:
  "Set of blocks above some block that needs to be cleared"
- $C_2 \equiv holding$:
  "Set of blocks being held"

Generalized Potential Heuristics with Guarantees

generalized potential heuristics with performance guarantees:

- heurisistics that lead search directly to goal state (i.e., DDA)
- exist for many domains
- can be learned from given examples
- pen-and-paper proofs of generalization

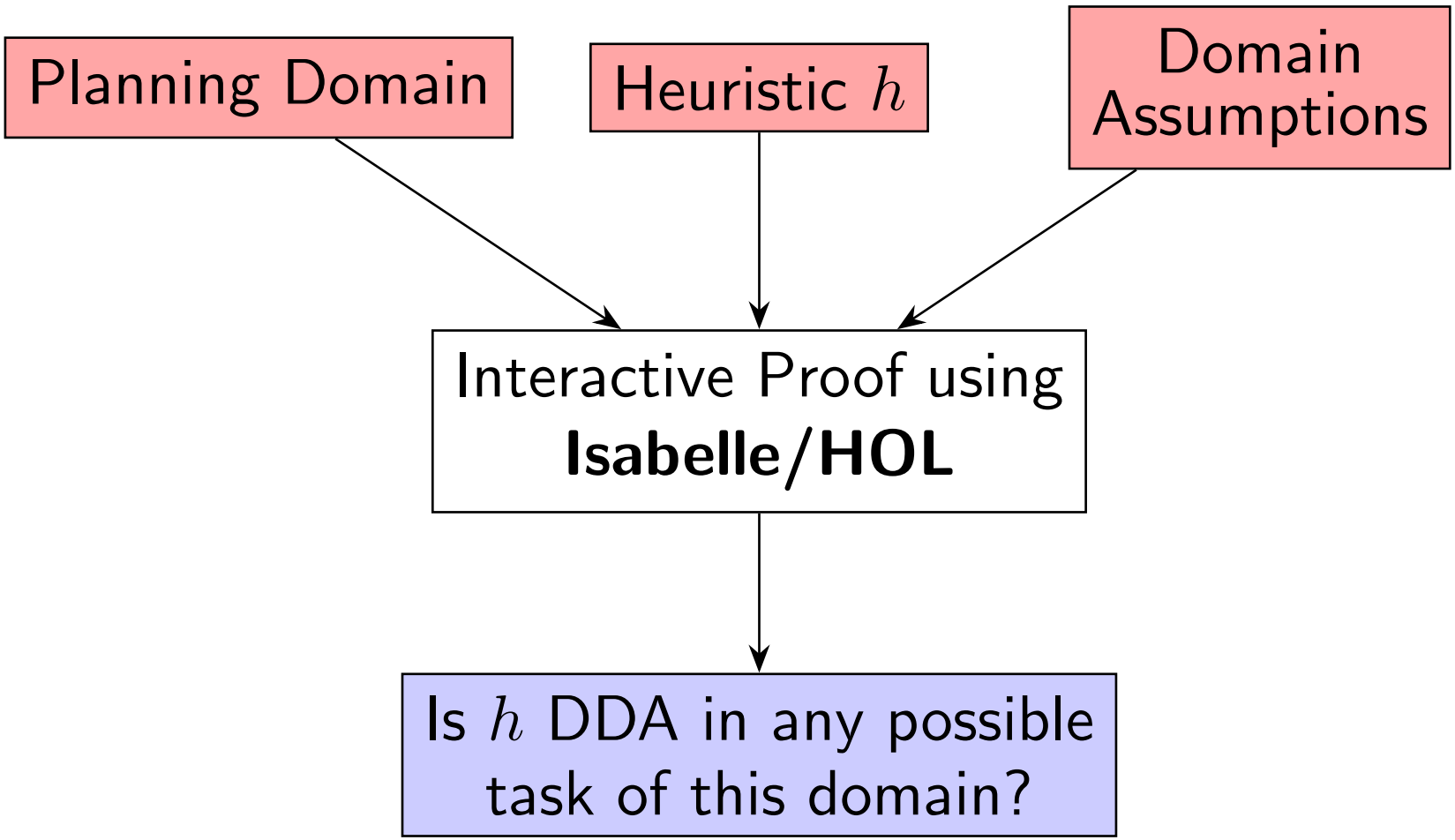
Generalized Potential Heuristics with Guarantees

generalized potential heuristics with performance guarantees:

- heurisistics that lead search directly to goal state (i.e., DDA)
- exist for many domains
- can be learned from given examples
- pen-and-paper proofs of generalization

our goal: automate these proofs as much as possible

Main Idea


Planning Domain
Heuristic h
Domain Assumptions
Interactive Proof using **Isabelle/HOL**
Is h DDA in any possible task of this domain?

Current Limitations

heuristics representing tiered-measures of progress:

- order between concepts used, from "best" to "worst"
- object can be at (max.) one concept
- move object to better concept = make search progress

invariants are given:

- assume they are provided
- prove that they are invariants
- related: Bonet et al. (IJCAI 2019)

Current Limitations

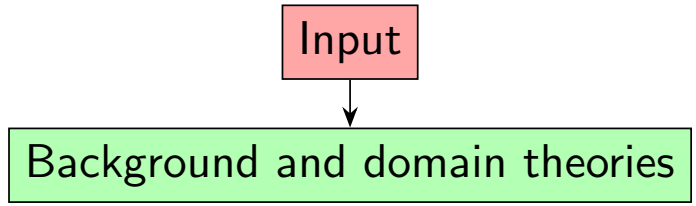
heuristics representing tiered-measures of progress:

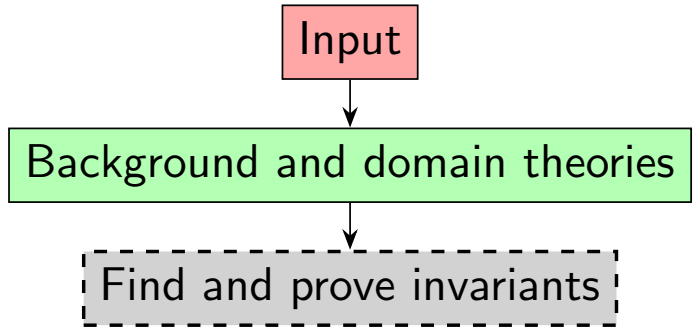
- order between concepts used, from "best" to "worst"
- object can be at (max.) one concept
- move object to better concept = make search progress

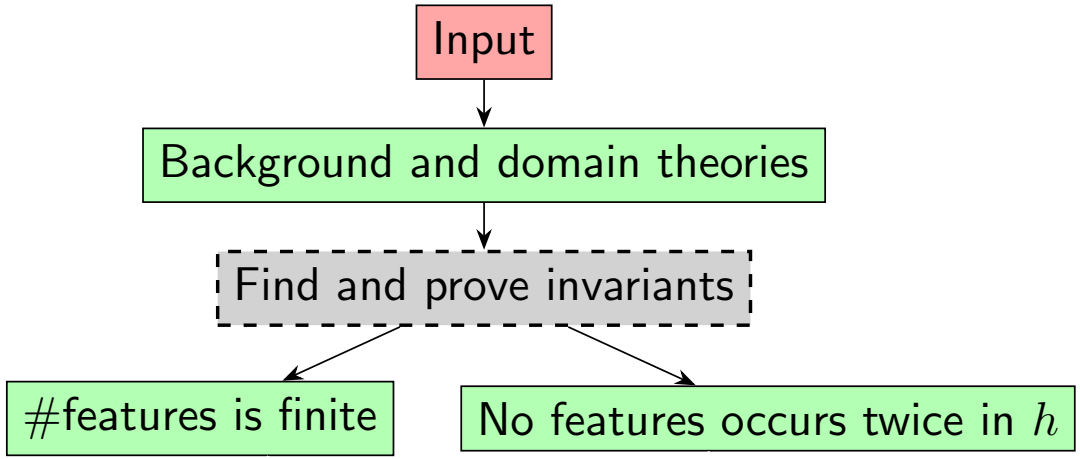
invariants are given:

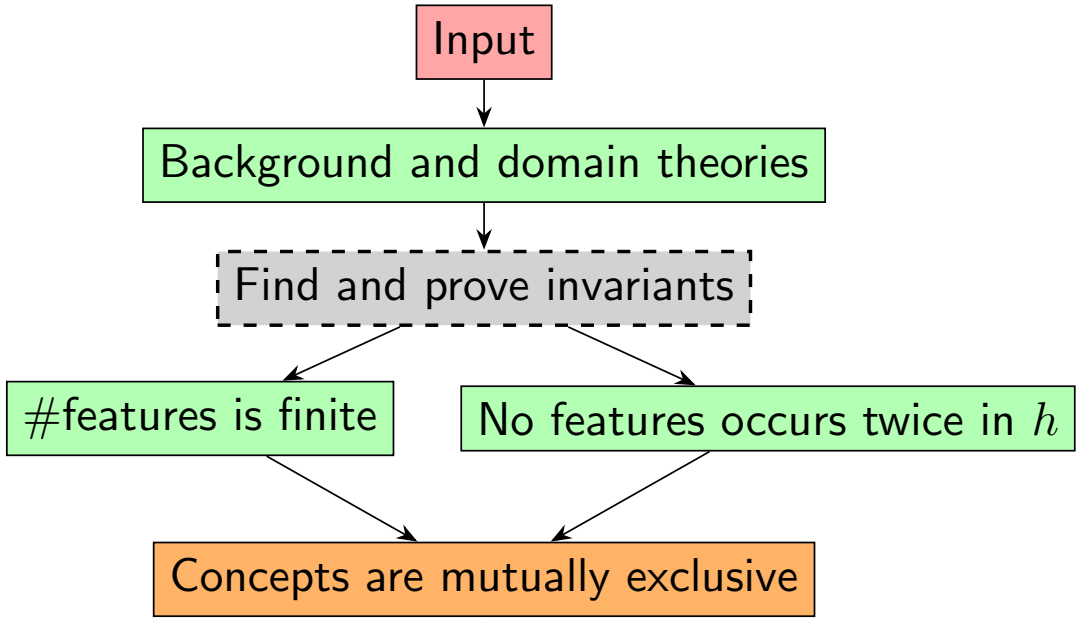
- assume they are provided
- prove that they are invariants
- related: Bonet et al. (IJCAI 2019)

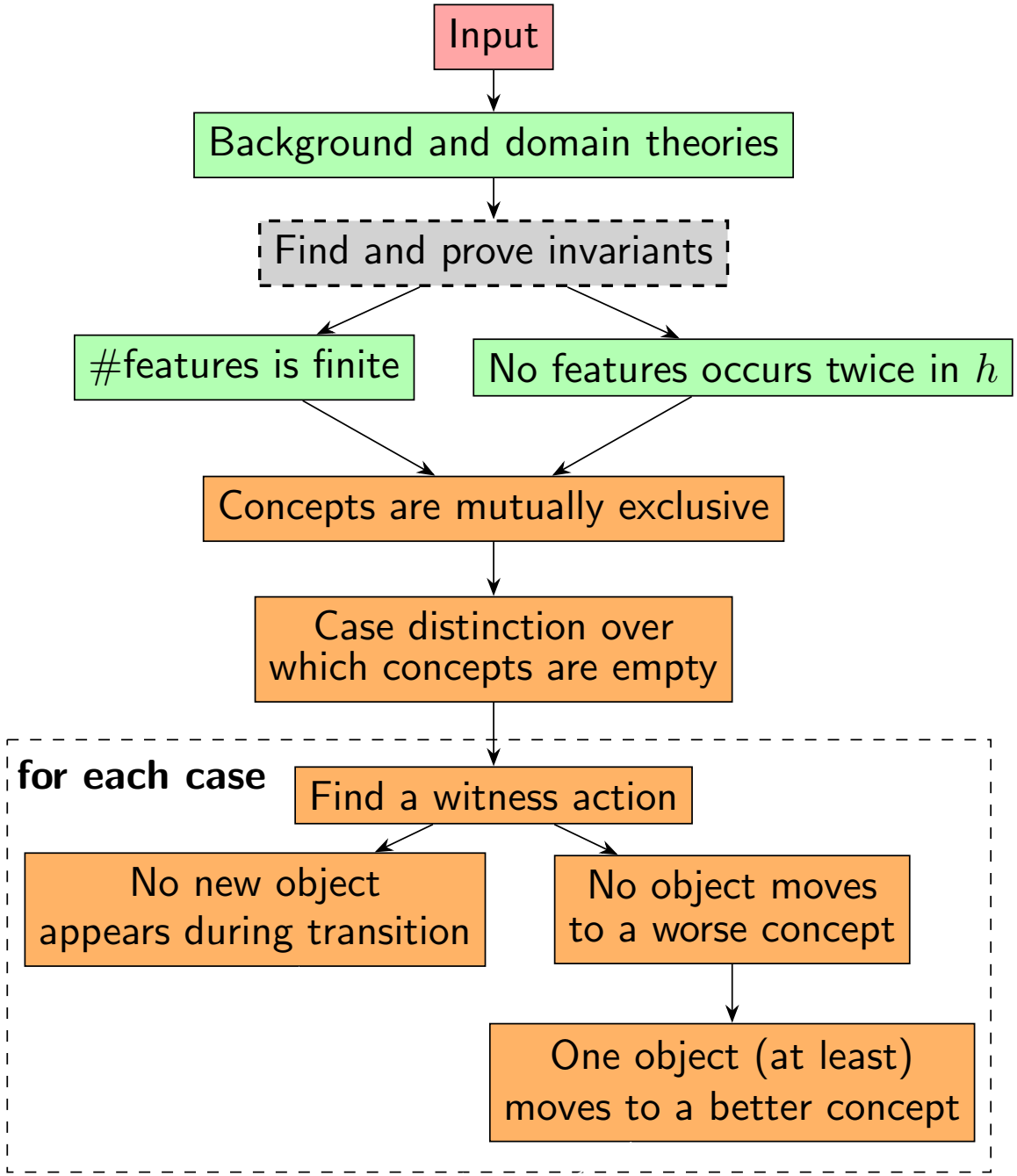
most importantly: so far we only implemented Miconic

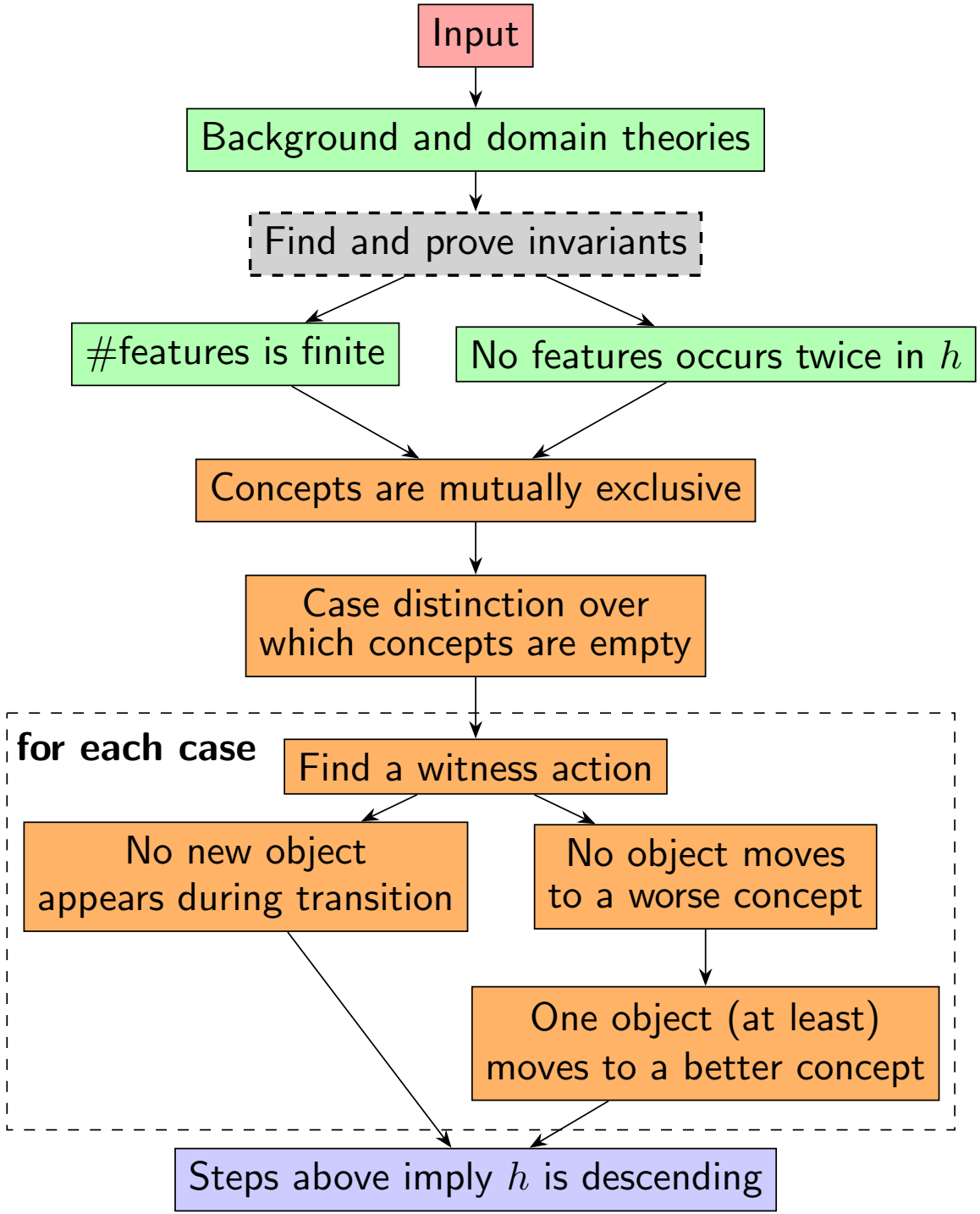

Input
Background and domain theories


Input
Background and domain theories
Find and prove invariants


Input
Background and domain theories
Find and prove invariants
#features is finite
No features occurs twice in h


Input
Background and domain theories
Find and prove invariants
#features is finite
No features occurs twice in h
Concepts are mutually exclusive


Input
Background and domain theories
Find and prove invariants
#features is finite
No features occurs twice in h
Concepts are mutually exclusive
Case distinction over which concepts are empty
for each case
Find a witness action
No new object appears during transition
No object moves to a worse concept
One object (at least) moves to a better concept


Input
Background and domain theories
Find and prove invariants
#features is finite
No features occurs twice in h
Concepts are mutually exclusive
Case distinction over which concepts are empty
for each case
Find a witness action
No new object appears during transition
No object moves to a worse concept
One object (at least) moves to a better concept
Steps above imply h is descending

Conclusion

what we have:

- interactive proof for Miconic domain
- parts of the proof are already fully automated
- thousands of lines of Isabelle/HOL theory
  - great part of it is generated automatically by scripts
  - in a new domain, we expect the user to do very little

next steps:

- finish automation of the proof as much as possible
- use for other domains
- other types of generalized potential heuristics