Based on the reviewers’ comments we have made and number of alterations to our submission, the details of which are described below:

1. **Reviewer 2 comments:**
   
   1.1. The use of “which” (and its associated punctuation) is mistaken in many, many places. This needs a thorough review.
       
       _Answer:_ We have been through the paper and removed any instances of “which” we deemed unnecessary. We feel that the remaining instances are grammatically correct and reflect the writing style of the authors.

   1.2. Section 1, line 1: By allowing a developer to introduce design steps incrementally, ...
       
       _Answer:_ Done.

   1.3. page 2, item 3: proof-failure analysis
       
       _Answer:_ All instances of “proof failure” have been changed to “proof-failure” throughout the paper.

   1.4. p2,l33: limitations -¿ challenges
       
       _Answer:_ Done.

   1.5. p3,l20: Rephrase the sentence starting “We restrict the application of our ... ” This is an odd comment, and it only becomes clear what you much later.
       
       _Answer:_ Rewritten to: “In order to cope with the possible occurrence of an incorrect model, the invariant discovery process terminates when no failed PO is discharged by the set of proposed invariants. This also addresses cases in which the required invariant(s) cannot be generated through our technique – examples of this case will be given in Sections 7.1.1 and 7.2.”

   1.6. S2.2.1 is not really needed, but anyway: font of the quote is too small, l18: represent, rather than representS
       
       _Answer:_ The size of the quote could not be changed since it is the size specified by the style format of the journal.

   1.7. p6,l39: consisting -¿ it consists
       
       _Answer:_ Changed to “this consists”.

   1.8. 1st paragraph of S2.2.2: it is heavy, and has sentences that go over four lines with seven verbs. The same happens at the end of this section. Rewrite.
       
       _Answer:_ Done.

   1.9. p12,l41: traces; it
       
       _Answer:_ Done.

   1.10. p14,l23: explain why the other ones are not relevant; l24: 7 PRs AND the ...; l52: some -¿ there
       
       _Answer:_ An explanation has been added: “The remaining PRs are mainly domain specific. For instance, HR contains PRs specialised to generate integer sequences and to handle operations over floating point numbers. We disable these specific purpose PRs from the invariant discovery process.”

   1.11. p15,l46: Moreover -¿ For example; l50: relate TO the
       
       _Answer:_ Done.

   1.12. p16,l41: the characterisation of non-core concepts is not clear here. Give an example, or improve explanation.
       
       _Answer:_ An example was added: “For instance, if a PO contains the expression a ∨ b, where a and b are core concepts, the disjunct PR can be applied, with the two core concepts as inputs, to replicate the expression within HR. Therefore, a ∨ b represents a non-core concept.”

   1.13. FH5: explain better, at this point, why overcoming a proof failure potentially leads to new failures.
       
       _Answer:_ An explanation has been added: “That is because in order to ensure that the system is consistent with a new invariant, new POs are generated. This may give rise to proof failures when additional properties are required in order to prove the model does not violate the new invariant.”
1.14. p18,l27-29: explain how the parameters are changed in the definition of filters in each iteration.
   Answer: An explanation has been added: “In other words, in the first iteration, the parameters of the filtering heuristics are the prioritised core and non-core concepts that appear within the goal of the failed POs and the conjectures associated to them. In the second iteration, the heuristics are parameterised with the core and non-core concepts that appear within the hypotheses of the failed POs and the conjectures associated to them.”

1.15. p21,l17: analysis -¿ analyse; l37: is A combination
   Answer: Done.

1.16. p23,item 2: give an example of the problem
   Answer: An example has been added illustrating the problem.

1.17. p24,i3: discuss any limitations of the prover.
   Answer: Done.

1.18. Make the data related to all experiments available and point to where it can be found, so that a reader can reproduce the experiments.
   Answer: A footnote has been added in page 27 pointing out where the files related to the experiments can be found.

1.19. Discuss the effort of applying FH4 in the experiments, since it is not implemented.
   Answer: An explanation has been added in page 27, item 4: “This involves manually translating each conjecture generated up to heuristic FH3 into Event-B and introducing one by one to the Event-B model in the Rodin toolset. Then, a record is kept of the failures addressed by each conjecture and when all the conjectures have been analysed, the recorded information is compared to select the final set of invariants”.

1.20. p33,l23: incomplete sentence; l56; reduce in a loss of information?
   Answer: Done.

1.21. The problems identified in Section 7.1.1 could be better discussed. For each of them, it would be interesting to know whether you are suggesting to extend the technique, and what would be the difficulty of achieving that. In particular, FH2 seems a problem.
   Answer: The addition of new PRs and the extension of the types accepted by HR’s concepts, two of the challenges mentioned in section 7.1.1, represent extensions of HR’s core features and not of the technique itself. We have pointed this out throughout section 7.1.1. So far we have not modified the core of HR; therefore, it is not possible at this stage to measure the impact of implementing these extensions as this requires in-depth knowledge of HR’s architecture. However, as we mention in section 9, we plan to implement these extensions in our future work. Regarding heuristic FH2, we have observed from our experiments that this heuristic works well in most cases; therefore, we don’t plan to modify it at the moment.

1.22. Table 9: the use of colour is not going to be possible in the final version. Use another way of distinguishing the red rows.
   Answer: An extra column was added to mark the rows that were previously highlighted in red.

1.23. Section 7.3.1: consider making this a separate discussion section.
   Answer: A new section has been added. That is: Section 7.4 Discussion: Beyond current applications.

1.24. p36,l55: not -¿, no
   Answer: Done.

1.25. p38,l54: provable; a retrieve
   Answer: Done.

1.26. p39,l28: code; these
   Answer: Done.

1.27. Explain why you generated two Java implementations for each example. I still think there is too much about Daikon in this paper, but no doubt the Daikon community will take the bait. :-) Make the data for these experiments available as well.
   Answer: An explanation has been added in the third paragraph of section 8.1.1: “Moreover, two alternative java representations were generated in both experiments in order to illustrate one of Daikon characteristics—that the invariant discovery process is influenced by the syntax of the code. We come back to this later in this section”.

1.28. p43,i4 invariants; pre- and ...
   Answer: Done.

2. Reviewer 3 comments:

2.1. p3 13: Bolton writes about retrieve *relations* (in her abstract at least), probably better to stick with that. If they do all turn out to be functions, make a point of that only if it matters.
   Answer: Done.
2.2. p3 42 extra space after HREMO
   Answer: Done.

2.3. p20 Fig 17: please use standard terminology for shapes ("squares" and "rectangles" are rectangles and parallelograms).
   Answer: Done.

2.4. p21 17: analysis \textendash, analyse
   Answer: Done.

2.5. p26 and on: no good reason why the elements of the Event-B specification are typeset in italic in the text, but not in the figures and tables.
   Answer: For clarity, we use italics to emphasise a reference to a model element in the text. This is not necessary in the figures, so these are typeset without italics.

2.6. p39 26 such *an* execution
   Answer: Done.

2.7. p40 15 extra space after HREMO
   Answer: Done.

2.8. p42 49 "not high" hmmmm, "low"?
   Answer: Done.

2.9. p45 [Bol05] fix bibtex entry to give capitalisation for alloy and z
   Answer: Done.

2.10. References [CBW00x] in pairs with identical titles, really different papers?
     Answer: We are grateful to the reviewer for pointing this out and we have now corrected this.