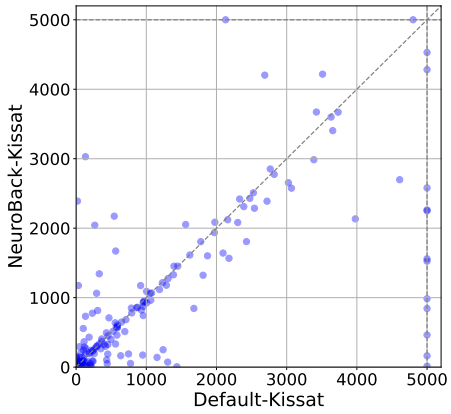
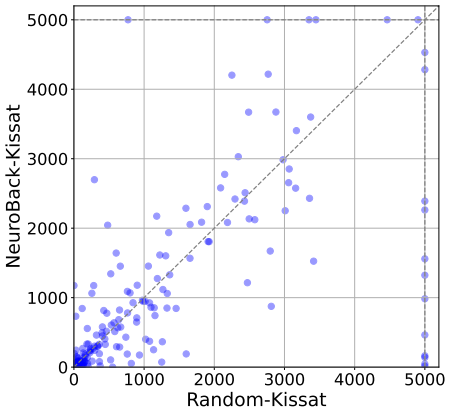
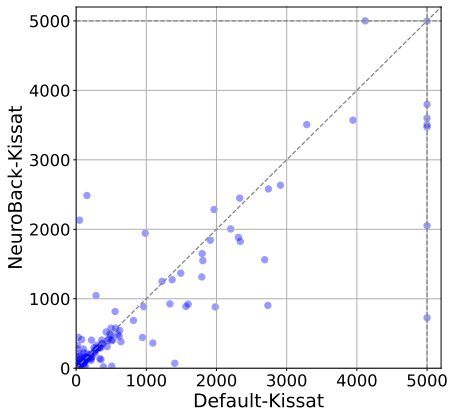


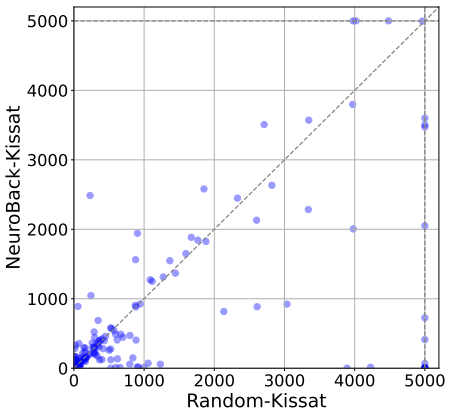
(a) Default-Kissat vs. NeuroBack-Kissat on SATCOMP-2022.



(b) Random-Kissat vs. NeuroBack-Kissat on SATCOMP-2022.



(c) Default-Kissat vs. NeuroBack-Kissat on SATCOMP-2023.



(d) Random-Kissat vs. NeuroBack-Kissat on SATCOMP-2023.

Figure 1: Time taken by `Default-Kissat` vs. `NeuroBack-Kissat` on SATCOMP-2022 (a), `Random-Kissat` vs. `NeuroBack-Kissat` on SATCOMP-2022 (b), `Default-Kissat` vs. `NeuroBack-Kissat` on SATCOMP-2023 (c), and `Random-Kissat` vs. `NeuroBack-Kissat` on SATCOMP-2023 (d) to solve each test problem in seconds (for problems that are solved by at least one solver). Each problem is represented by a dot whose location indicates the solving time of each method. The dots on the dashed lines at 5,000 seconds indicate failures. It is evident that more dots are present in the lower triangular areas, indicating that there are more problems on which `NeuroBack-Kissat` outperforms both `Default-Kissat` and `Random-Kissat`.

1 Additional Experimental Results

The scatter plots is another commonly used plot in the SAT community for comparing the solving effectiveness of two solvers on each problem. Fig. 1 shows the scatter plots of `NeuroBack-Kissat` and its two baseline solvers, `Default-Kissat` and `Random-Kissat`. It is evident that more dots are present in the lower triangular area, indicating that there are more problems on which `NeuroBack-Kissat` outperforms both `Default-Kissat` and `Random-Kissat`. Specifically, `NeuroBack-Kissat` outperforms `Default-Kissat` on 43 and 40 additional problems in SATCOMP-2022 and SATCOMP-2023, respectively, reducing solving time by 117 and 36 seconds per problem. Similarly, `NeuroBack-Kissat` outperforms `Random-Kissat` in SATCOMP-2022 and SATCOMP-2023 on 22 and 29 more problems, respectively, leading to a reduction in solving time of 98 and 246 seconds per problem.

2 PERFORMANCE ON SOLVED SAT AND UNSAT PROBLEMS

Upon detailed analysis, for 661 problems from both SATCOMP-2022 and SATCOMP-2023 testing sets, there are 194 unsat problems and 216 sat problems that are solved by either `Default-Kissat` or `NeuroBack-Kissat`. For the 194 solved unsat problems, `NeuroBack-Kissat` outperformed `Default-Kissat` in 121 cases (62.4%) while `Default-Kissat` outperformed `NeuroBack-Kissat` in only 61 problems (31.4%). For the 216 solved sat problems, `NeuroBack-Kissat` outperformed `Default-Kissat` in 110 problems (50.9%), while `Default-Kissat` outperformed `NeuroBack-Kissat` in 87 problems (40.3%). While `NeuroBack-Kissat` showed a higher improvement rate in unsat problems compared to sat ones (62.4% vs 50.9%), the extent of improvement was more significant in sat problems. On average, `NeuroBack-Kissat` enhanced the performance of sat problems by 53.2%, compared to an average improvement of only 14.6% in unsat problems. These trends were similarly observed when comparing `NeuroBack-Kissat` with `Random-Kissat`.

The experimental results highlight two key aspects. First, they demonstrate that NeuroBack's predicted variable phases can enhance the efficiency in solving unsat problems. Our explanation is that NeuroBack's phase predictions can aid in directing the search towards the unsatisfiable part of the search space. While NeuroBack cannot satisfy all components of a given SAT problem, it may predict phases that satisfy certain components, thereby allowing the solver to concentrate on the unsat part. Furthermore, in modern SAT solvers such as `Default-Kissat` Biere & Fleury (2020), an assignment that falsifies the fewer clauses is often preferred in the searching loop, allowing the solver to specifically target the unsat portions of the clause set. Consequently, the phases predicted by NeuroBack can facilitate identifying an assignment that reduces clause falsification, thereby enhancing solving unsat problems.

Second, the experimental results also show that NeuroBack achieves a more pronounced improvement in solving sat problems than in solving unsat problems. This distinction stems from the inherent nature of these problems. In sat problems, a complete satisfying assignment exists, where each variable is assigned a phase that leads to a solution. Conversely, in unsat problems, only partial satisfying assignments exist, with phases assigned to just a subset of variables. Consequently, the phases predicted by NeuroBack have a generally greater impact in resolving sat problems. This is because, for these problems, the predicted phases can contribute directly to finding a satisfying assignment. In contrast, for unsat problems, the utility of predicted phases is somewhat restricted to identifying partial solutions or refining the search scope. This fundamental difference in the nature of sat versus unsat problems underpins the varying degrees of effectiveness observed in NeuroBack's performance.

3 SETTING UP THE MEMORY LIMIT FOR NEUROBACK-KISSAT

In our experimental setup, which includes a machine equipped with 256GB of memory running 64 solver instances in parallel, we have conservatively set the SAT formula size threshold at 135 MB. This ensures that the memory usage of each solver instance does not exceed our specified memory threshold of 10GB. This threshold setting is based on our practical experience. Increasing this threshold could potentially lead to memory contention issues. Users might choose to adjust the formula size threshold based on their machine's memory capacity. Alternatively, they might simply establish a memory threshold for each solver instance based on their machine's memory capacity and allow model inference to proceed until this threshold is reached, which typically incurs an overhead of no more than a few seconds.

REFERENCES

Armin Biere and Mathias Fleury. Chasing target phases. In *Workshop on the Pragmatics of SAT*, 2020.