

# Rewrite System Showdown: Stochastic Search vs. EqSat

Qiantan Hong\*  
qthong@cs.stanford.edu  
Stanford University  
USA

Yihong Zhang\*  
yz489@cs.washington.edu  
University of Washington  
USA

Rupanshu Soi\*  
rsoi@cs.stanford.edu  
Stanford University  
USA

Alex Aiken  
aaiken@stanford.edu  
Stanford University  
USA

## Abstract

Equality saturation has become a dominant paradigm for equational program optimization. However, it has never been rigorously compared to another approach to the same problem, even though several exist, the most notable being stochastic search. In this paper, we compare equality saturation to stochastic search over five benchmarks to answer the question: are e-graphs actually good?

## 1 Introduction

Equality saturation (EqSat) is an equational program optimization technique, with powerful, ready-to-use libraries such as *egg* [27], *egglog* [21, 30], and many others [4, 9, 16, 25]. It has seen success in many domains, from floating-point expressions [18] to 3D printing [17], from synthesizing better hardware [5] to faster GPU kernels [29]. For many, equality saturation has become the go-to tool for building domain-specific program optimizers.

At a high-level, EqSat takes an input program and a set of semantics-preserving rewrite rules  $R = \{l_1 \rightarrow r_1, l_2 \rightarrow r_2, \dots\}$ , searches the space of equivalent programs defined by the rewrite rules, and returns the best program it finds. Importantly, EqSat uses the rewrite rules to grow an *e-graph* of equivalent programs. E-graphs compactly represent equivalence relations between programs, which allows EqSat to efficiently explore the search space.

Recently, a number of extensions to EqSat have been proposed, including DAG-costed extractions [3, 24, 28], effectful programs [2, 22], theories [31], optimistic analyses [1], and contextual reasoning [6]. The complexity of these extensions is often engendered by the compressed representation of terms in an e-graph.

Before we expend more effort in developing extensions, let us take a step back and ask: are e-graphs actually good? Somewhat surprisingly, EqSat has not been compared against other approaches to equational program optimization, not even in the original *egg* paper, but EqSat is certainly not the only approach to equational program optimization. Because of the lack of a comparison, practitioners rely on popularity

or word of mouth when picking an approach to program optimization.

To fill in this gap in the literature, we conduct the first extensive comparison of EqSat against another well-established technique: stochastic search. Stochastic search, based on Markov chain Monte Carlo (MCMC) sampling [7], has also found successful applications in a number of domains [10–12, 19, 20]. Unlike equality saturation, stochastic search always manipulates concrete terms.

We implement stochastic search with an interface similar to the *egg* library, including two features that many EqSat applications rely on, namely, per-node analyses, and rules beyond purely syntactic rewrites. This allows us to port existing EqSat applications to stochastic search with only mild effort. We note that *egg*'s interface is well-suited to host other equational program optimization techniques and is not limited to EqSat and stochastic search.

We collect a set of benchmarks for equational program optimization. Our benchmarks range over different domains and from small, self-contained problems to full-blown applications taken from prior work. The benchmarks are:

1. Matrix chain multiplication
2. Trigonometric simplification
3. Indefinite integration
4. Decompiling CAD into structured programs [17]
5. Proving inequalities in the Halide compiler [14]

Preliminary results indicate that neither technique dominates the other, suggesting that it may be possible to develop simpler alternatives to EqSat that are equally powerful for equational program optimization but avoid the complexity of e-graphs.

## 2 Stochastic Search

In this section, we give a brief overview of stochastic search. Starting from an initial term  $t$ , stochastic search explores equivalent terms via Markov Chain Monte Carlo (MCMC) sampling. The search is governed by a proposal function  $P(t)$ , which induces a set of candidate terms through the application of the ruleset to any sub-term of  $t$ . A successor term  $t' \in P(t)$  is sampled with a probability proportional

\*Equal contribution.

**Algorithm 1** Stochastic Search

---

**Require:** Proposal function  $P$ , cost function  $C$ , initial term  $t_0$ , budget  $B$ , inverse temperature  $\beta$ , soft restart period  $n_{\text{soft}}$ , exploration steps  $E$ , hard restart period  $n_{\text{hard}}$

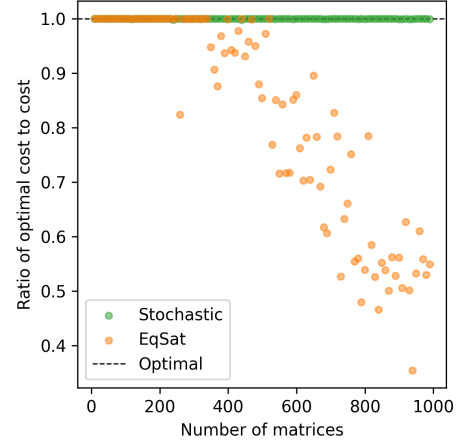
- 1: **parallel for**  $i = 1$  **to**  $B$  **do**
- 2:      $n \leftarrow 0, n_{\text{stall}} \leftarrow 0, t \leftarrow t_0, t_i^* \leftarrow t_0$
- 3:     **while**  $n_{\text{stall}} < n_{\text{hard}}$  **do**
- 4:         **if**  $n \bmod n_{\text{soft}} < E$  **then**
- 5:              $\beta_i \leftarrow 0$                       $\triangleright$  Exploration phase
- 6:         **else**
- 7:              $\beta_i \leftarrow \beta$                       $\triangleright$  Exploitation phase
- 8:         Sample  $t' \in P(t)$  with probability  $\propto$   
 $\exp\left(-\frac{\beta_i}{2} (C(t') - C(t))\right)$
- 9:          $t \leftarrow t', n \leftarrow n + 1$
- 10:        **if**  $C(t) < C(t_i^*)$  **then**
- 11:             $t_i^* \leftarrow t, n_{\text{stall}} \leftarrow 0$
- 12:        **else**
- 13:             $n_{\text{stall}} \leftarrow n_{\text{stall}} + 1$
- 14: **return** term with lowest cost from  $\{t_1^*, \dots, t_B^*\}$

---

to  $\exp\left(-\frac{\beta}{2}(C(t') - C(t))\right)$  where  $C$  is a cost function and  $\beta \in \mathbb{R}$  is a hyperparameter called the inverse temperature. The successor term replaces the current term and this process repeats until a computational budget is exhausted.

Restarting is crucial for the efficacy of stochastic search. The distribution of solution quality across runs is often heavy-tailed, so many independent runs can give much faster search times than a few long runs [13]. Moreover, restarts help in getting a run out of local minima. For these reasons, our implementation supports two kinds of restart. In a soft restart, we set  $\beta = 0$  periodically to help the search get out of local minima. In a hard restart, we reset the state of the run to the very beginning; hard restarts only occur if the run has failed to make progress for a given number of steps. Algorithm 1 gives pseudocode for stochastic search, showing sampling and both kinds of restart.

A central advantage of stochastic search is its embarrassingly parallel nature: we can spawn many threads, each with a different seed, to explore a different part of the underlying program space. In fact, this scaling is critical to the performance of stochastic search (Section 3.2). Moreover, stochastic search only uses constant memory per thread as it only maintains a single term at a time. This makes stochastic search scale well with time. In contrast, not only is the parallelization of EqSat non-trivial and the subject of several research efforts [26, 30], the memory usage of EqSat can also blow up as more equalities are added to the e-graph, so any scaling of EqSat will always be constrained by available memory, unless mitigation measures like pulsing [14], guidance [15], or garbage collection are used.



**Figure 1.** Showdown for matrix chain multiplication.

### 3 Showdown

In this section, we present detailed results comparing EqSat and stochastic search for 5 applications. Experiments are run on servers with a 64-Core CPU with hyperthreading and 500 GB of memory. We give 10 seconds of wall-clock time to both techniques, except Section 3.5 where the time limit is set to 3 seconds. We use egg’s default BackOffScheduler for EqSat. Unless noted otherwise, stochastic search is run using 128 threads, while EqSat is run single-threaded. While this setup gives stochastic search 128× CPU time (but not wall-clock time), we believe it still constitutes a fair comparison due to the embarrassingly parallel nature of stochastic search.

#### 3.1 Matrix Chain Multiplication

This benchmark concerns the classic optimization problem of reassociating a chain of matrix products to reduce the total number of scalar multiplications. To illustrate, consider a product of three matrices  $ABC$ , where the dimensions of the matrices are  $2 \times 3$ ,  $3 \times 4$  and  $4 \times 5$  respectively. There are two ways to associate this product,  $(AB)C$  and  $A(BC)$ . Recall that the number of scalar multiplications required to multiply two matrices of size  $m \times n$  and  $n \times k$  is  $mnk$ . Using this formula, we can compute the number of multiplications needed in the two associations, which come out to be 64 and 90 respectively, making  $A(BC)$  a much faster implementation of the chain multiplication. This problem can be formulated as equational program optimization with two rules for the associativity of matrix multiplication. We randomly generate tests with a number of matrices between 10 and 1,000. For each test, we also compute the optimal association using the standard  $O(N^3)$  dynamic programming algorithm.

Results are given in Figure 1. The y-axis denotes the ratio of the optimal cost to the costs of associations produced by stochastic search and EqSat. Stochastic search is always within 0.3% of the optimal solution. EqSat is close to optimal when the number of matrices is moderate (e.g.,  $\leq 300$ ),

but quickly becomes suboptimal as the number of matrices increases.

It is not surprising that, below a certain limit at least, both techniques fare equally well on this problem, despite the large program space. It was recently proven that the mixing time of a random walk on the state space for stochastic search<sup>1</sup> is  $O(n^3 \log^3 n)$ , which is also the time for stochastic search to converge to its stationary distribution at high temperature [8]. The space complexity of EqSat on this problem happens to be  $O(n^3)$ —while not exactly the same, the two techniques are in the same ballpark regarding saturation or convergence. However, as it takes EqSat  $n$  iterations to saturate, saturation becomes intractable once  $n$  gets larger, which explains the poor scalability.

We also tried pulsing for EqSat (omitted due to lack of space), where we run three iterations of EqSat, extract the best term from the e-graph, and use it to seed a fresh e-graph until the 10 s time limit is reached. Pulsing significantly improves EqSat’s performance: for  $n \leq 800$ , it consistently remains within 1% of the optimal cost. However, as  $n$  exceeds 800, performance begins to degrade again, with 5 out of 20 tests deviating by up to 10% from optimal, and the worst case deviating by 65%.

### 3.2 Trigonometric Simplification

For this benchmark, we collected a set of 35 trigonometric simplification exercises, such as proving that  $\sin^4 x - \cos^4 x + 1 = 2 \sin^2 x$ . The ruleset consists of trigonometric identities (e.g.,  $\sin^2 x + \cos^2 x = 1$ ) and purely algebraic ones (e.g.,  $x^2 - y^2 = (x + y)(x - y)$ ). The cost of a term is the number of function symbols in the term, i.e.,  $\text{AstSize}$ . There is an intended solution for each test, but it may not have the minimal cost. We consider a test solved if a solution of cost less than or equal to the intended solution was found.

One difficulty in creating this benchmark is soundness. In particular, two rules in our ruleset may cause unsoundness:

$$\frac{b}{a} \Rightarrow \frac{1}{\frac{a}{b}} \quad \text{if } b \neq 0 \quad (\text{RECIP})$$

$$\frac{ab}{cb} \Rightarrow \frac{a}{c} \quad \text{if } b \neq 0 \quad (\text{CANCEL})$$

To understand why, we must reason about the domain of the LHS and the RHS of both rules. By domain, we mean the set of values of the variables for which the LHS or RHS is defined, e.g.,  $\text{dom}(\text{LHS}(\text{CANCEL})) = \{(a, b, c) \mid a, c \in T, b \in T \setminus \{0\}\}$  where  $T$  is the set of all ground terms under consideration. RECIP can cause unsoundness in both EqSat and stochastic search as  $\text{dom}(\text{RHS}(\text{RECIP})) \subsetneq \text{dom}(\text{LHS}(\text{RECIP}))$  and the guard only checks syntactic equality with zero. Therefore, the rule will apply and result in unsoundness if, for instance,  $b = x - x$  and 0 is not yet in the e-class of  $x - x$  (for EqSat). CANCEL can cause unsoundness in EqSat but not in stochastic

<sup>1</sup>The state space for a matrix chain of length  $n$  is the 1-skeleton of associahedron  $K_n$ .

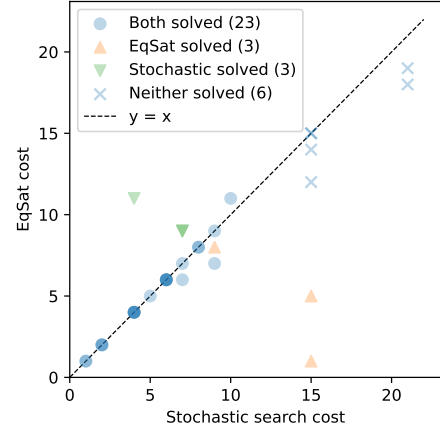


Figure 2. Showdown for trigonometric simplification.

search as  $\text{dom}(\text{LHS}(\text{CANCEL})) \subsetneq \text{dom}(\text{RHS}(\text{CANCEL}))$ , so it can cause unsoundness only if applied right to left, which happens in EqSat because EqSat forgets the direction in which a rule was applied. Consequently, there is a proof in EqSat of  $0 = 1$  by relying on CANCEL and a few other (sound) rules:

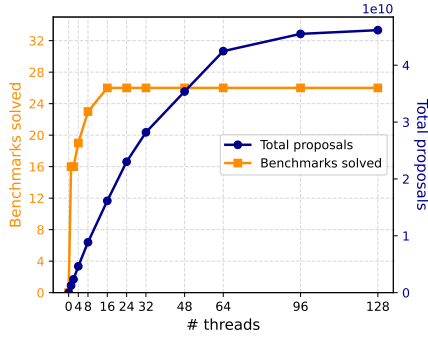
$$1 \Leftarrow \frac{x - x}{x - x} \Rightarrow \frac{0}{x - x} \Rightarrow 0$$

To ensure the validity of solutions in the presence of unsoundness, we adopt two mitigation strategies. For EqSat, we follow a checkpointing strategy inspired by Herbie: a copy of the e-graph is made at the end of each iteration, and if unsoundness is detected, the run is aborted and the last checkpointed e-graph is used for extraction. For stochastic search, the run undergoes a hard restart upon detection of unsoundness. Across our benchmark suite, we ensure that solutions found by both techniques are always valid.

Figure 2 shows the results. Both EqSat and stochastic search solved 26/35 tests, although the set of tests they solved are not the same. We also measured how stochastic search scales with the number of threads, and the results are shown in Figure 3. We observe that the number of proposals (see Algorithm 1) scales in proportion with the number of threads, and that stochastic search solves significantly more benchmarks with more threads, up to 16 threads.

### 3.3 Indefinite Integration

Also taken from homework exercises, this benchmark contains 10 problems of indefinite integration, such as proving that  $\int x \cos x \, dx = x \sin x + \cos x$ . The ruleset includes many rules found in the trigonometric benchmark, as well as rules specific to integration, such as for the integral of  $\sin x$ , or for integrating by parts. Like trigonometric simplification, this ruleset can also cause unsoundness, and we deal with it by using the mitigation strategies listed in Section 3.2. Here we use different cost functions for EqSat and stochastic search.



**Figure 3.** Number of proposals and benchmarks solved vs. number of threads for stochastic search on trigonometric simplification.

**Table 1.** Showdown for indefinite integration.

Category	Count
Both solved	6
Only EqSat solved	1
Only Stochastic solved	2
Neither solved	1

For EqSat, we use `AstSize` with the exception that the function symbols for integration and differentiation (needed for integration by parts) are weighed by 100 (other function symbols are weighed by 1). However, this cost function is unsuitable for stochastic search as it makes applying linearity of integration too expensive:  $\int (a + b) dx \Rightarrow \int a dx + \int b dx$  would increase cost by 100, making this often fruitful step very unlikely to be taken. To remedy this problem, we design a cost function that prefers integration of smaller terms. The cost function for stochastic search weighs integration and differentiation nodes by the square of the costs of their children. This function makes the linearity of integration a cost-decreasing step as  $(x + y)^2 > x^2 + y^2$ .

Another difficulty with constructing this benchmark was that many integration problems require a  $u$ -substitution, but neither technique supports  $u$ -substitutions, at least not without modifications. In our benchmark, we did not include any problems that require a  $u$ -substitution. To illustrate the problem, consider the integral  $\int x^2 \sin x^3 dx$ . The standard way to proceed is to substitute  $u = x^3 \Rightarrow du = 3x^2 dx$ . The integral becomes  $\int \frac{1}{3} \sin u du = -\frac{1}{3} \cos u = -\frac{1}{3} \cos x^3$ . There are two difficulties with implementing this process. First is the choice of sub-term to replace—this is a creative step and there are often numerous choices out of which only one or two will lead to the solution. The wrong choice will further complicate the integral and it is difficult to recover once in this state, so an efficient method of undoing incorrect substitutions is desirable. The situation is simpler for stochastic

search, as threads that make the wrong substitution could simply restart, but for EqSat, it is unclear how incorrect substitutions could be undone efficiently. The second difficulty is in the mechanical aspects of differentiating with respect to the chosen sub-term and backsubstituting into the integral. In our example, the latter can be viewed as two rewrites that must happen at the same time:  $x^3 \Rightarrow u$ ,  $dx \Rightarrow \frac{1}{3} du$ . Such a “double rewrite” is not supported by standard implementations of EqSat or stochastic search, though it can be added with some effort.

Results are shown in Table 1. For this benchmark we do not give the costs as they are not particularly enlightening, due to the different weights of integration and differentiation function symbols. Interestingly, of the 2 tests that stochastic search solved but EqSat did not, EqSat managed to carry out the integration but failed to simplify the resulting expression fully. For the test that EqSat solved but stochastic search did not, stochastic search failed to carry out the integration fully.

### 3.4 Decompiling CAD into Structured Programs

Next, we consider a decompilation task. Szalinski [17] is a tool for compressing low-level unstructured CAD programs by finding equivalent programs that use functional primitives such as maps and folds. Szalinski is especially attractive as a benchmark because it was also included as a benchmark in the egg paper. The input to Szalinski is a flat CAD program like the following, which denotes 5 cubes placed in a straight line:

```
(Union
  (Translate (0 0 0) Cube)
  (Translate (2 0 0) Cube)
  (Translate (4 0 0) Cube)
  (Translate (6 0 0) Cube)
  (Translate (8 0 0) Cube))
```

Using rewrite rules, Szalinski is able to identify repetitive structure in this program and synthesize a more compact equivalent, which is more amenable to user editing:

```
(Fold Union
  (Tabulate (i 5)
    (Translate ((* 2 i) 0 0) Cube)))
```

We reimplement Szalinski using stochastic search and use the same set of rules as the original with small tweaks<sup>2</sup>. We evaluate the two implementations on the Thingiverse dataset. As in the Szalinski paper, we consider both the original programs from Thingiverse and their perturbed variants; the latter are meant to test robustness of the decompilation.

Results are shown in Figure 4. The left two figures show the evaluation results on the normal programs, and the right two figures show results on perturbed programs. The scatter

<sup>2</sup>We simplified the right-hand sides of two rules to make stochastic search’s hill climbing easier. We added pushdown rules for `Sort` and `Unsort`. Finally, we added rules for rewriting the intersection/union of a singleton into itself.

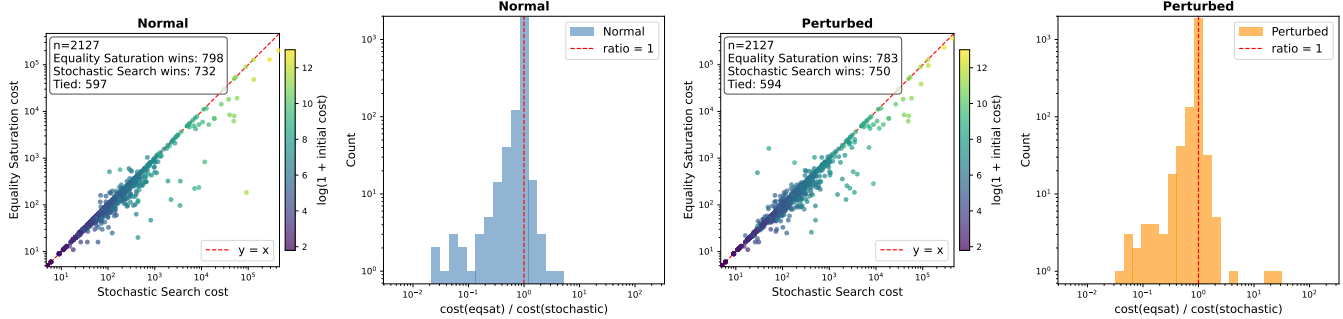


Figure 4. Showdown on compressing CAD expressions.

plots use a similar format as before: the x-coordinate of a point gives the cost of the program found by stochastic search, and the y-coordinate gives the cost found by EqSat. The histograms show the distribution of the ratios between costs found by EqSat and by stochastic search.

Overall, neither technique dominates the other—there are programs on which EqSat finds a better solution than stochastic search, and programs on which the opposite is true. However, we do find that on programs on which stochastic search wins, it often wins by a small margin, whereas on programs on which EqSat wins, the margin is usually larger. We suspect this is because some of Szalinski’s rewrite rules are ill-suited to stochastic search, which causes it to miss certain critical rewrite paths.. To illustrate Szalinski’s workload, consider an abstract rewrite system with three rules:  $\{a \Rightarrow b, b \Rightarrow a, f(b, \dots, b) \Rightarrow g(b, \dots, b)\}$ , where  $f$  and  $g$  have some large arity  $N$ . We start with the term  $f(a, \dots, a)$  and our goal is to prove it equal to  $g(b, \dots, b)$ . Further assume that the cost function is simply  $C(g(b, \dots, b)) = 0$  and 1 otherwise. EqSat can prove this within 2 iterations. On the other hand, because the cost function provides little guidance, stochastic search must bounce around this high-dimensional space until it hits the term  $f(b, \dots, b)$  purely by chance, which requires, in expectation,  $O(2^N)$  time. Only then can the rule  $f(b, \dots, b) \Rightarrow g(b, \dots, b)$  apply, opening up a path to the solution. Though a simplified example, essentially the same thing happens in Szalinski, as all the Translate function symbols need to be in a certain form before they can be folded together. It is possible that the performance of stochastic search on this benchmark can be improved with rule engineering and a better cost function, but we leave that as future work.

### 3.5 Proving Inequalities in the Halide Compiler

The last benchmark is from Caviar [14], an EqSat reimplementa-tion of the rewriting system inside the Halide compiler. Halide generates many inequalities during its bounds inference, and proving them is critical for Halide to generate efficient code. These inequalities concern algebraic expressions, similar to our benchmarks on trigonometry and integration,

Table 2. Showdown for proving inequalities from Halide.

Category	Count
Both solved	2298
Only EqSat solved	361
Only Stochastic solved	184
Neither solved	2943

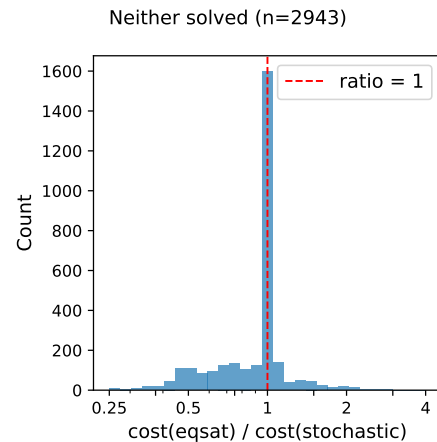


Figure 5. Cost distribution of unproved Halide inequalities.

as well as max and min operators. We implemented the same set of rules with stochastic search, and evaluate both systems on Caviar’s “hard” dataset. Each test consists of an inequality over symbolic variables to be proved or disproved, e.g.,  $\max(i, 2) < \max(i + 3, 3)$ .

Results are shown in Table 2 and Figure 5. Of the 5,786 tests, 40% were solved by both techniques, 6% were solved by EqSat only, 3% were solved by stochastic search only, and a large fraction, about 50%, were not solved by either. Note that Caviar’s ruleset is relatively weak, and therefore, not all tests can be decided using it. Overall, the results on Caviar align with the other benchmarks: while there is a core set of tests passed by both techniques, the tests passed by EqSat

are neither a superset nor a subset of the tests passed by stochastic search, and vice versa.

## 4 Discussion

Finally, we discuss the differences between EqSat and stochastic search in light of our results, draw some conclusions, offer some ideas for future work, and answer the question raised in the introduction.

**E-graphs good?** Anecdotally, while working on this project, we were often surprised by how well EqSat performed on some benchmarks. For matrix chain multiplication in particular (Section 3.1), we expected stochastic search to comfortably outperform EqSat for medium to large problems, as a long sequence of rewrites may be required to reassociate the chain into the optimal form and finding this chain ought to be difficult without any cost guidance. However, the results indicate that EqSat performs almost optimally for problem sizes up to 300, which is already large. This can be attributed to the compression provided by the e-graph—although the size of the search space is exponential in the number of matrices, it can be represented by the e-graph in just  $O(n^3)$  space.

While we still believe that EqSat will perform poorly for problems that require long sequences of rewrites, and some evidence to the affirmative is provided in Section 3.1, we had a difficult time coming up with a realistic example of such a problem. It may be the case that EqSat works well for exactly the kinds of equational program optimization problems that occur in practice.

**Is EqSat Pareto optimal over stochastic search?** A common theme across our 5 benchmarks is that neither technique comes out to be Pareto optimal over the other—there are tests on which EqSat performs better than stochastic search, and vice-versa. This result suggests that a combination of the two techniques, perhaps EqSat imbued with some stochasticity, or a variant of stochastic search that uses an e-graph to do “local” exploration while relying on randomness to jump to far parts of the space, could be the ultimate equational program optimization engine.

**Unsoundness and directionality of rules.** One difference between the two techniques is that a directed rule induces an equivalence relation over terms in the case of EqSat and a refinement relation over terms in the case of stochastic search. As an upshot, some rules that cause unsoundness for EqSat work just fine for stochastic search, such as  $\frac{x}{x} \Rightarrow 1$ . Section 3.2 discusses how we mitigate this issue for some benchmarks. Anecdotally, unsoundness was a larger issue for EqSat than for stochastic search—the latter had to be run for much longer before it exploited unsoundness. It is unclear to us whether this is an advantage or disadvantage of EqSat, because one could argue that quick exploitation of unsoundness indicates better ability at exploiting the ruleset.

**Hyperparameters.** A disadvantage of stochastic search is the requirement of hyperparameter tuning, namely, of tuning the restart schedule and values of  $\beta$ . While it was not difficult to find good values of these hyperparameters in our experiments, no such tuning had to be done for EqSat. It is however worth noting that several applications of EqSat do require hyperparameter tuning in the form of rule scheduling. Rule scheduling is used when applying the entire ruleset at the same time would lead to a blow-up of the e-graph, so the ruleset is divided into disjoint sets and applied to the e-graph in phases. `BackoffScheduler` in `egg` is a simple example of rule scheduling. Another hyperparameter in EqSat is the concept of top-k e-nodes, which is used in incremental EqSat [23] to handle very large e-graphs.

**Scalability.** An advantage of stochastic search is scalability owing to its embarrassingly parallel nature. Figure 3 provides some evidence to this effect by showing how the number of passing tests scales with the number of threads given to stochastic search. In principle, we could scale stochastic search up to thousands or even tens of thousands of threads by using multi-node execution, whereas parallel or distributed execution at this scale remains challenging for EqSat.

**Non semantics-preserving rewriting.** An advantage of stochastic search that we did not explore in this paper is that, unlike EqSat, it can benefit from non-semantics-preserving rules. Perhaps the most well-known application of this is `STOKE` [19], a superoptimizer based on stochastic search for x86 programs. Most of the rewrites in `STOKE` were not semantics-preserving; instead, its cost function had a correctness component that guided its search to correct programs. `STOKE` often found completely different assembly-level implementations of a given algorithm, and this was possible only because its rewrite rules allowed it to travel to very different parts of the space from which it started. It would be very difficult to match these results using only semantics-preserving rewrites.

## References

- [1] Russel Arbore, Alvin Cheung, and Max Willsey. 2025. Optimism in Equality Saturation. *arXiv preprint arXiv:2511.20782* (2025).
- [2] Anonymous Authors. 2025. Better E-Graphs for Imperative Programs (This ICFP Submission).
- [3] Yaohui Cai, Kaixin Yang, Chenhui Deng, Cunxi Yu, and Zhiru Zhang. 2025. SmoothE: Differentiable E-Graph Extraction. In *Proceedings of the 30th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 1* (Rotterdam, Netherlands) (*ASPLOS '25*). Association for Computing Machinery, New York, NY, USA, 1020–1034. doi:10.1145/3669940.3707262
- [4] Alessandro Cheli. 2021. Metatheory.jl: Fast and Elegant Algebraic Computation in Julia with Extensible Equality Saturation. *Journal of Open Source Software* 6, 59 (2021), 3078. doi:10.21105/joss.03078
- [5] Chen Chen, Guangyu Hu, Dongsheng Zuo, Cunxi Yu, Yuzhe Ma, and Hongce Zhang. 2024. E-Syn: E-Graph Rewriting with Technology-Aware Cost Functions for Logic Synthesis. In *Proceedings of the 61st ACM/IEEE Design Automation Conference* (San Francisco, CA, USA)

- (DAC '24). Association for Computing Machinery, New York, NY, USA, Article 124, 6 pages. doi:10.1145/3649329.3656246
- [6] Samuel Coward, George A. Constantinides, and Theo Drane. 2023. Combining E-Graphs with Abstract Interpretation. In *Proceedings of the 12th ACM SIGPLAN International Workshop on the State Of the Art in Program Analysis* (Orlando, FL, USA) (SOAP 2023). Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3589250.3596144
- [7] Persi Diaconis. 2009. The Markov Chain Monte Carlo Revolution. *Bull. Amer. Math. Soc.* 46, 2 (2009), 179–205. doi:10.1090/S0273-0979-08-01238-X
- [8] David Eppstein and Daniel Frishberg. 2023. Improved Mixing for the Convex Polygon Triangulation Flip Walk. In *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 261)*, Kousha Etesami, Uriel Feige, and Gabriele Puppis (Eds.). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 56:1–56:17. doi:10.4230/LIPIcs.ICALP.2023.56
- [9] Kiran Gopinathan. 2025. Ego: An Extensible E-Graph Library for OCaml. <https://github.com/verse-lab/ego>. Accessed: 2026-04-06.
- [10] Stefan Heule, Eric Schkufza, Rahul Sharma, and Alex Aiken. 2016. Stratified synthesis: automatically learning the x86-64 instruction set. *SIGPLAN Not.* 51, 6 (June 2016), 237–250. doi:10.1145/2980983.2908121
- [11] Qiantan Hong and Alex Aiken. 2024. Recursive Program Synthesis using Paramorphisms. *Proc. ACM Program. Lang.* 8, PLDI, Article 151 (June 2024), 24 pages. doi:10.1145/3656381
- [12] Zhihao Jia, Oded Padon, James Thomas, Todd Warszawski, Matei Zaharia, and Alex Aiken. 2019. TASO: optimizing deep learning computation with automatic generation of graph substitutions. In *Proceedings of the 27th ACM Symposium on Operating Systems Principles* (Huntsville, Ontario, Canada) (SOSP '19). Association for Computing Machinery, New York, NY, USA, 47–62. doi:10.1145/3341301.3359630
- [13] Jason R Koenig, Oded Padon, and Alex Aiken. 2021. Adaptive restarts for stochastic synthesis. In *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*. 696–709.
- [14] Smail Kourta, Adel Abderahmane Namani, Fatima Benbouzid-Si Tayeb, Kim Hazelwood, Chris Cummins, Hugh Leather, and Riyadh Baghdadi. 2022. Caviar: an e-graph based TRS for automatic code optimization. In *Proceedings of the 31st ACM SIGPLAN International Conference on Compiler Construction* (Seoul, South Korea) (CC 2022). Association for Computing Machinery, New York, NY, USA, 54–64. doi:10.1145/3497776.3517781
- [15] Thomas Kundenedhler, Andrés Goens, Siddharth Bhat, Tobias Grosser, Phil Trinder, and Michel Steuwer. 2024. Guided Equality Saturation. *Proc. ACM Program. Lang.* 8, POPL, Article 58 (Jan. 2024), 32 pages. doi:10.1145/3632900
- [16] Rodrigo Mesquita. 2022. hegg: Fast Equality Saturation in Haskell. <https://hackage.haskell.org/package/hegg>. Hackage package, BSD-3-Clause license. Accessed: 2026-04-06.
- [17] Chandrakana Nandi, Max Willsey, Adam Anderson, James R. Wilcox, Eva Darulova, Dan Grossman, and Zachary Tatlock. 2020. Synthesizing Structured CAD Models with Equality Saturation and Inverse Transformations. In *Proceedings of the 41st ACM SIGPLAN Conference on Programming Language Design and Implementation* (London, UK) (PLDI 2020). Association for Computing Machinery, New York, NY, USA, 31–44. doi:10.1145/3385412.3386012
- [18] Pavel Panchekha, Alex Sanchez-Stern, James R. Wilcox, and Zachary Tatlock. 2015. Automatically Improving Accuracy for Floating Point Expressions. *SIGPLAN Not.* 50, 6 (June 2015), 1–11. doi:10.1145/2813885.2737959
- [19] Eric Schkufza, Rahul Sharma, and Alex Aiken. 2013. Stochastic super-optimization. In *Proceedings of the Eighteenth International Conference on Architectural Support for Programming Languages and Operating Systems* (Houston, Texas, USA) (ASPLOS '13). Association for Computing Machinery, New York, NY, USA, 305–316. doi:10.1145/2451116.2451150
- [20] Eric Schkufza, Rahul Sharma, and Alex Aiken. 2014. Stochastic optimization of floating-point programs with tunable precision. *SIGPLAN Not.* 49, 6 (June 2014), 53–64. doi:10.1145/2666356.2594302
- [21] Saul Shanabrook. 2024. Egglog Python: A Pythonic Library for E-graphs. arXiv:2305.04311 [cs.PL] <https://arxiv.org/abs/2305.04311>
- [22] Stan Seibert Siu Kwan Lam, Stuart Archibald. 2025. Numba v2: Towards a SuperOptimizing Python Compiler. In *Proceedings of the Python in Science Conference (SciPy)*. fncj2446. doi:10.25080/fncj2446
- [23] Rupanshu Soi, Benjamin Driscoll, Ke Wang, and Alex Aiken. 2025. Incremental Equality Saturation. In *EGRAPHS*. Seoul, South Korea. <https://rupanshusoi.github.io/pdfs/egraps-25.pdf>
- [24] Glenn Sun, Yihong Zhang, and Haobin Ni. 2024. E-Graphs as Circuits, and Optimal Extraction via Treewidth. arXiv:2408.17042 [cs.DS] <https://arxiv.org/abs/2408.17042>
- [25] Jonathan Van der Cruyssen, Abd-El-Aziz Zayed, Mai Jacob Peng, and Christophe Dubach. 2026. Parallel and Customizable Equality Saturation. In *Proceedings of the 35th ACM SIGPLAN International Conference on Compiler Construction* (Sydney, NSW, Australia) (CC '26). Association for Computing Machinery, New York, NY, USA, 94–105. doi:10.1145/3771775.3786266
- [26] Jonathan Van der Cruyssen, Abd-El-Aziz Zayed, Mai Jacob Peng, and Christophe Dubach. 2026. Parallel and Customizable Equality Saturation. In *Proceedings of the 35th ACM SIGPLAN International Conference on Compiler Construction* (Sydney, NSW, Australia) (CC '26). Association for Computing Machinery, New York, NY, USA, 94–105. doi:10.1145/3771775.3786266
- [27] Max Willsey, Chandrakana Nandi, Yisu Remy Wang, Oliver Flatt, Zachary Tatlock, and Pavel Panchekha. 2021. egg: Fast and extensible equality saturation. *Proc. ACM Program. Lang.* 5, POPL, Article 23 (jan 2021), 29 pages. doi:10.1145/3434304
- [28] Yichen Yang, Phitchaya Mangpo Phothilimthas, Yisu Remy Wang, Max Willsey, Sudip Roy, and Jacques Pienaar. 2021. Equality Saturation for Tensor Graph Superoptimization. arXiv:2101.01332 [cs.AI]
- [29] Yihong Zhang, Derek Gerstmann, Andrew Adams, and Maaz Bin Safer Ahmad. 2026. Pushing Tensor Accelerators beyond MatMul in a User-Schedulable Language. In *2026 IEEE/ACM International Symposium on Code Generation and Optimization (CGO)*. 242–254. doi:10.1109/CGO68049.2026.11395214
- [30] Yihong Zhang, Yisu Remy Wang, Oliver Flatt, David Cao, Philip Zucker, Eli Rosenthal, Zachary Tatlock, and Max Willsey. 2023. Better Together: Unifying Datalog and Equality Saturation. *Proc. ACM Program. Lang.* 7, PLDI, Article 125 (jun 2023), 25 pages. doi:10.1145/3591239
- [31] Philip Zucker. 2025. Omelets Need Onions: E-graphs Modulo Theories via Bottom-up E-matching. arXiv:2504.14340 [cs.PL] <https://arxiv.org/abs/2504.14340>