Bonsai: Compiling Queries to Pruned Tree Traversals

ALEXANDER J ROOT, Stanford University, USA
CHRISTOPHE GYURGYIK, Stanford University, USA
PURVI GOEL, Stanford University, USA
KAYVON FATAHALIAN, Stanford University, USA
JONATHAN RAGAN-KELLEY, Massachusetts Institute of Technology, USA
ANDREW ADAMS, Adobe Research, USA
FREDRIK KJOLSTAD, Stanford University, USA

Trees can accelerate queries that search or aggregate values over large collections. They achieve this by storing metadata that enables quick pruning (or inclusion) of subtrees when predicates on that metadata can prove that none (or all) of the data in a subtree affect the query result. Existing systems implement this pruning logic manually for each query predicate and data structure. We generalize and mechanize this class of optimization. Our method derives conditions for when subtrees can be pruned (or included wholesale), expressed in terms of the metadata available at each node. We efficiently generate these conditions using symbolic interval analysis, extended with new rules to handle geometric predicates (e.g., intersection, containment). Additionally, our compiler fuses compound queries (e.g., reductions on filters) into a single tree traversal. These techniques enable the automatic derivation of generalized single-index and dual-index tree joins that support a wide class of join predicates beyond standard equality and range predicates. The generated traversals match the behavior of expert-written code that implements query-specific traversals, and can asymptotically outperform the linear scans and nested-loop joins that existing systems fall back to when hand-written cases do not apply.

CCS Concepts: • **Software and its engineering** → **Compilers**; **Domain specific languages**; • **Computing methodologies** → *Boolean algebra algorithms*; Ray tracing; Collision detection.

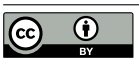
Additional Key Words and Phrases: compilation, data independence, acceleration structures, tree data structures

ACM Reference Format:
Alexander J Root, Christophe Gyurgyik, Purvi Goel, Kayvon Fatahalian, Jonathan Ragan-Kelley, Andrew Adams, and Fredrik Kjolstad. 2026. Bonsai: Compiling Queries to Pruned Tree Traversals. *Proc. ACM Program. Lang.* 10, PLDI, Article 178 (June 2026), 29 pages. https://doi.org/10.1145/3808256

1 Introduction

Augmented tree data structures accelerate queries over large collections of data. They achieve this by storing metadata, such as bounding boxes or aggregate values, at internal nodes that enable traversals to exclude or include entire subsets without examining each individual element. This mechanism, known as pruning or culling, is widely used: in database systems *indexes* are used to accelerate range and point queries [14, 29], in graphics systems *acceleration structures* are used to

Authors' Contact Information: Alexander J Root, Computer Science, Stanford University, Stanford, CA, USA, ajroot@cs.stanford.edu; Christophe Gyurgyik, Computer Science, Stanford University, Stanford, CA, USA, cpg@cs.stanford.edu; Purvi Goel, Computer Science, Stanford University, Stanford, CA, USA, pgoel2@cs.stanford.edu; Kayvon Fatahalian, Computer Science, Stanford University, Stanford, CA, USA, kayvonf@cs.stanford.edu; Jonathan Ragan-Kelley, Massachusetts Institute of Technology, Cambridge, MA, USA, jrk@mit.edu; Andrew Adams, Adobe Research, San Francisco, CA, USA, andrew.b.adams@gmail.com; Fredrik Kjolstad, Computer Science, Stanford University, Stanford, CA, USA, kjolstad@cs.stanford.edu.



ACM 2475-1421/2026/6-ART178
https://doi.org/10.1145/3808256

skip occluded geometry [38, 48] and to efficiently find collisions [21, 55], and in scientific computing *spatial trees* are used to limit computation to relevant regions [4, 34].

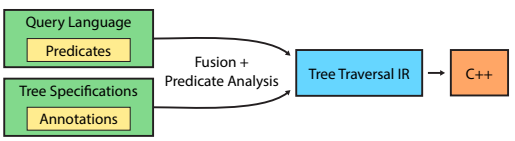
Despite their ubiquity, pruning logic is manually implemented. Traditional acceleration trees, such as Bounding Volume Hierarchies (BVHs) [21, 38, 48], B-trees [14], and R-Trees [29], encode pruning rules operationally through hand-written traversal logic. Generalized Search Trees [31] attempt to unify these structures under a common abstraction, but rely on user-specified search and consistency predicates. As a result, every new query requires a new hand-engineered traversal *for every data structure* [20, 69–71]. Modern query engines [52, 73, 79] must therefore treat these tree queries as opaque, manually optimized operators specialized to particular queries, rather than reusable mechanisms for accelerating a broad class of queries.

This work introduces a new direction: automatically generating tree traversals from separate specifications of the query and the tree data structure. Hand-written traversals across domains follow a single principle: traversals exploit *necessary* and *sufficient* conditions parameterized by tree metadata that determine whether a predicate is guaranteed to hold or fail for all elements in a subtree. This principle extends naturally to reductions: for associative reductions (e.g., sum, product), subtree metadata can be used to include entire subsets without visiting individual elements; for reductions of idempotent operators (e.g., min, max), subtree metadata can be used to skip subtrees that cannot impact the running aggregate. Our key insight is that these conditions can be automatically derived from high-level query predicates and annotations of tree metadata.

To enable the derivation of custom pruning logic, we adopt the perspective of *data independence* [13]: the query should be decoupled from the metadata used to accelerate it. Tree specifications provide this metadata as language-level annotations, allowing the compiler to derive pruning logic from the annotations and query operators. Mechanizing both the derivation of pruning conditions and the use of metadata for reductions enables fully automated generation of fused traversal code, eliminating the need for hand-written traversals.

Our approach is built on two core steps: First, a lowering algorithm fuses the operators of a high-level query into a tree traversal. The resulting traversal is expressed in terms of abstract necessary and sufficient predicates that guide pruning. Second, we use symbolic analysis to derive the implementation of the concrete necessary and sufficient conditions specific to the given query and tree's metadata. We use symbolic interval analysis to generate these conditions and provide a novel extension for analyzing spatial relations such as intersection and containment. Together, these techniques enable generation of specialized traversals for a broad class of search and non-equijoin algorithms, extending beyond what traditional systems support. Our technical contributions are:

- A lowering algorithm that fuses set filters and reductions into work-efficient tree traversals.
- A technique, termed *predicate analysis*, that extends symbolic interval analysis with rules for geometric operators, to derive pruning conditions from query predicates and tree metadata.
- Two generalized tree-based non-equijoin algorithms, enabled by the above techniques.



We implement these ideas in the Bonsai compiler, whose architecture is shown on the right. Bonsai compiles queries written in a simple functional query language (Section 3). Trees are separately declared as ADTs with metadata annotations (Section 4). Bonsai's lowering algorithm (Section 5) fuses query operations into a tree traversal by recursively rewriting a tree traversal intermediate representation (TTIR) defined in Section 5.1. Lowering filters and idempotent reductions to pruning traversals requires the generation of pruning and inclusion functions, which Bonsai derives via predicate analysis (Section 6). Finally, we demonstrate that the fusion algorithm can also generate code for joins (filters of products)

that coiterate multiple trees (Section 7). Across a range of benchmarks, our generated traversals match hand-written code and asymptotically outperform the linear scans and nested-loop joins that systems fall back to when hand-written cases do not apply.

Although Bonsai is a set operator compiler that produces efficient implementations of range queries, spatial joins, and similar classes of operations, it lacks features needed for it to be used as a backend for a relational database management system (DBMS). In particular, it would need a query planner and associated cost model, as well as efficient non-tree operator implementations (e.g., sort-merge join, hash join, deduplication, etc.). Bonsai is most accurately described as an *operator compiler*: it generates efficient implementations of individual and fused operators, but does not decide the order of the operators, which operators to fuse, or the allocation of temporaries. We discuss directions of future work for integrating our techniques into DBMSs in Section 11.

2 The Structure of Accelerating Queries with Trees

To illustrate the basic concepts of how trees can accelerate queries, consider a range query: finding all 1D points in a collection whose `x` coordinate lies between `lo` and `hi` (inclusive):

```
type Point = { x : f32; id : i32; };
func rangeq(ps : Set<Point>, lo hi : f32) = filter(|p : Point| lo ≤ p.x && p.x ≤ hi, ps);
```

Building a tree over the points can accelerate filtering. Consider the following tree, built on two variants: a leaf node that stores a single point; and an interior node with two children augmented to hold the lower bound and upper bound (`xl`, `xh`) of the `x` coordinates of all points stored below it.

```
type ITree = Leaf(p : Point) | Interior(left right : ITree, xl xh : f32);
```

The range query can be efficiently implemented on this tree by considering three cases on an `Interior` node's metadata:

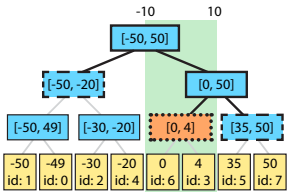
Always If the query range contains a subtree's range, all descendants are returned.

Maybe If the query range overlaps the subtree's range, some descendants may be in the query range, so the children are recursively visited and their subresults are unioned.

Never Otherwise, no descendants are included in the query, so the node is *pruned*.

We illustrate these cases in the below tree traversal code, where `yield` returns a singleton set and `scan t` returns all points in subtree `t`. We illustrate an example tree on the right. A range query with `lo=-10` and `hi=10` descends to the root's children, pruning the left subtree (first dashed box). It then descends on the right subtree, scanning the dotted orange box because its interval is contained in the query interval, and pruning its sibling (second dashed box).

```
func _rangeq(t : ITree, lo hi : f32) =
 match t
 | Leaf(p) → if lo ≤ p.x && p.x ≤ hi: yield p
 | Interior(left, right, xl, xh) →
    if lo ≤ xl && xh ≤ hi: scan t
    elif lo ≤ xh && xl ≤ hi: _rangeq(left, lo, hi) ∪
                             _rangeq(right, lo, hi)
    else: {};
```



These cases correspond to necessary and sufficient conditions on the query predicate. The *always* case requires a sufficient condition (one that guarantees the query predicate is true). The *never* case prunes a subtree when the necessary condition is false (the query predicate cannot be true). If neither of these cases is satisfied, the query predicate is *maybe* true, so the traversal must recurse.

Similar reasoning can be used to accelerate *idempotent* and *associative* reductions. To illustrate, we extend our range query example to compute the minimum `id` of any point within the range:

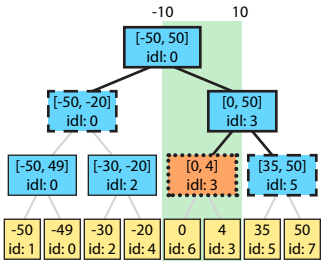
```
func minq(ps : Set<Point>, lo hi : f32) = min(|p : Point| p.id,
                                              filter(|p : Point| lo ≤ p.x && p.x ≤ hi, ps));
```

The `min` operation is both idempotent and associative. These properties enable two complementary optimizations: Associativity allows traversal to skip traversing fully-contained subtrees (in the *always* case) by directly composing a precomputed minimum stored in the tree instead of scanning for it; the idempotent property enables value-based pruning: if a subtree's minimum cannot improve the current best, we can skip it entirely (another *never* case).

To illustrate the use of these properties, we can extend the `ITree` to store the minimum `id` of its subtree as `idl`. This value can be used both to update the running minimum when a node is fully contained and in the pruning condition for the idempotent property. We provide the fused and lowered min-id-range query on the extended `ITree` below. When visiting the dotted orange box, the traversal uses its `idl` value to update the running minimum. The rightmost dashed box can then be pruned in one of two ways: it does not overlap the query interval, *and* it has a greater `idl` than the best found so far.

```
func _minq_impl(t : ITree, lo hi : f32, best : mut i32) =
 match t
 | Leaf(p) → if lo ≤ p.x && p.x ≤ hi: best = min(p.id, best)
 | Interior(left, right, xl, xh, idl, idh) →
    if lo ≤ xl && xh ≤ hi: best = min(idl, best) // inclusion
    elif lo ≤ xh && xl ≤ hi:
      if idl < best:                    // value-based pruning
        _minq_impl(left, lo, hi, best)
        _minq_impl(right, lo, hi, best);

func _minq(t : ITree, lo hi : f32) =
  let best : mut i32 = i32_max in _minq_impl(t, lo, hi, best);
  best;
```



This fusion provides asymptotic benefits by eliminating the allocation of an intermediate *and* enabling value-based pruning via the running minimum. Without fusion, the range query must construct the entire filter output before computing the minimum `id`.

To make such optimizations systematic, we now describe how queries and data structures can be specified in a way that makes their properties explicit. The next sections introduce a simple language for describing high-level set queries, alongside a language for expressing tree structures annotated with the metadata they store.

3 Query Language

Bonsai's query specification language consists of queries over unordered finite sets and multisets. Their element type can be any primitive type (integer, float, enum, fixed-length vectors of primitive types) or (often) a product type of primitive types. Geometry types are product types extended with spatial relationships defined on them (e.g., `intersects` and `contains`).

3.1 Set Operations

Bonsai supports operations on unordered sets and multisets: `filter`, `reduce` `map`, and `product` (the Cartesian product). In addition to these standard operations, Bonsai provides a number of idempotent reductions (e.g., `argmin`/`argmax`, `min`/`max`, and `any`/`all`). Bonsai operations have the same semantics for sets and multisets. For the sake of brevity, we refer to sets and multisets collectively as *sets* throughout the rest of the paper. Standard aggregations such as `count` or `avg` can be implemented by a `map` followed by a `reduce`: Bonsai's fusion algorithm ensures that this decomposition does not introduce inefficiencies. Reductions and filters are the primary operations accelerated by tree data structures, though filters or reductions of products can also be accelerated, as we illustrate in Section 7. We provide the grammar of Bonsai queries in Figure 1 and type signatures in Figure 2.

The set operators `any`, `all`, and `filter` accept a predicate that maps the set elements to boolean values. We support a grammar of scalar operations (comparators, conjunctions, disjunctions, mathematical operators, etc.) in addition to geometric predicates. The latter is discussed below.

3.2 Geometric Operations

In Bonsai, the type of a geometric object is a product type along with defined spatial relationships between pairs of geometric object types. Rather than restricting users to a hardcoded set of shapes, Bonsai supports arbitrary geometric entities (e.g., volumes, polygons, rays, points) provided the type exposes spatial relationships.

Our spatial operators follow the semantics of Egenhofer and Herring [18], and include topological operators (e.g., `intersects` and `contains`), ordering operators (e.g., $\leq_X$, which denotes ordering in the X dimension), and metric operations (e.g., `distmin`). Figure 3 gives visual examples of the topological and ordering relationships of the spatial operators and Figure 1 lists their grammar, which is part of the predicate subset of the query languages. While not exhaustive, these operators are expressive enough to encompass a broad class of spatial queries, including ray tracing, collision detection, and spatial SQL.

Bonsai provides a user-extensible library of geometric object types (e.g., `Ray`, `Triangle`, and `AABB`) with their spatial relationships following the implementations of Ericson [21].

q : Query	::= s	set variable
	| `filter`(P, q)	filtered set
	| `map`(F, q)	mapped set
	| `reduce`$(e_s, \oplus, q)$	aggregation
	| `product`(q_1, q_2)	Cartesian product
	| `min`(M, q) | `max`(M, q)	metric optimization
	| `argmin`(M, q) | `argmax`(M, q)	arg optimization
	| `any`(P, q) | `all`(P, q)	conditional logic
P : Predicate	::= $\|x : T\| \, e_b$	boolean lambda
M : Metric	::= $\|x : T\| \, e_r$	real-valued lambda
F : Map	::= $\|x : T\| \, e$	element transform
e_b : BoolExpr	::= $e_s \odot e_s$	scalar comparison $\odot \in \{<, \leq, =, \neq, \geq, >\}$
	| $e_b \wedge e_b$ | $e_b \vee e_b$ | $\neg e_b$	logical connectives
	| `contains`(e_g, e_g) | `covers`(e_g, e_g)	topological predicates [18]
	| `disjoint`(e_g, e_g) | `intersects`(e_g, e_g)	
	| `equals`(e_g, e_g) | `touches`(e_g, e_g)	
	| `within`(e_g, e_g)	
	| $e_g \leq_D e_g$ | $e_g <_D e_g$	ordering predicates on dimension D
e_r : RealExpr	::= x | n | $e_r \odot e_r$	variables, literals, arithmetic
	| `distmax`(e_g, e_g) | `distmin`(e_g, e_g)	distance metrics
e_g : GeoExpr	::= x | $G(e^*)$	geometric variable or constructor
T : Type	::= t | $T \times n$	scalar and vector types
	| `Set<T>`	set of elements
	| $(T_1, \ldots, T_k)$	product type (including geometric types)

Fig. 1. Context-free grammar of Bonsai's query language. The symbol s denotes a set variable, x a variable, n a numeric literal, $\oplus$ an associative and commutative binary operator, and G a geometric type constructor (e.g., `Ray()`).

```
filter(T → bool, Set<T>) : Set<T>
map(T → S, Set<T>) : Set<S>
reduce(T, T × T → T, Set<T>) : T
product(Set<T>, Set<S>) : Set<(T, S)>
min(T → ℝ, Set<T>) : ℝ
max(T → ℝ, Set<T>) : ℝ
argmin(T → ℝ, Set<T>) : T
argmax(T → ℝ, Set<T>) : T
any(T → bool, Set<T>) : bool
all(T → bool, Set<T>) : bool
```

Fig. 2. Type signatures for Bonsai's set operations. T and S are primitive types (e.g., integers, floats, or product types).

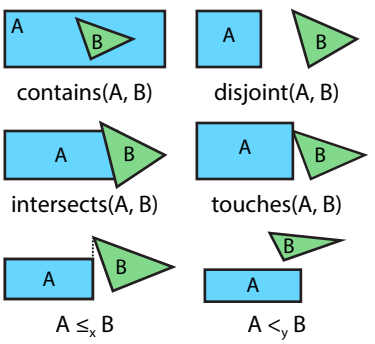


Fig. 3. Geometric predicate illustrations.

3.3 Examples

Bonsai's query language allows concise data-structure-agnostic representations of not just scalar queries, but also spatial queries, e.g., closest-hit ray tracing is expressible via:

```
closest(r : Ray, ts : Set<Triangle>) = argmin(|t : Triangle| distmin(r, t),
                                              filter(|t : Triangle| intersects(r, t), ts));
```

Shadow ray tracing, which only queries if there is a hit, is even more concise:

```
shadow(r : Ray, ts : Set<Triangle>) = any(|t : Triangle| intersects(r, t), ts);
```

Similarly, collision detection (a spatial join) is also concisely representable:

```
collisions(s0 : Set<Object>, s1 : Set<Object>) = filter(|a b : Object| intersects(a, b),
                                                        product(s0, s1));
```

4 Tree Specification Language

Bonsai provides a data modeling language based on Algebraic Data Types (ADTs) extended with augmentation *annotations* that enable the optimization of queries. Augmentations are declarative annotations over recursive data structures that specify useful metadata invariants, allowing Bonsai to accelerate certain operations. We focus on three primary types of augmentations:

Bounds augmentations Capture geometric bounds (e.g., intervals in 1D, bounding volumes in higher dimensions) over a subtree, attached to primitives or fields of sum-typed primitives. These can be used to accelerate filters and idempotent reductions.

Reduction augmentations Store partial aggregates over a subtree. These can be used to accelerate associative reductions.

Data tags Explicitly state that an ADT field should be interpreted as a member of the set that the tree implements.

Annotations are attached using a `with` clause on standard recursive ADT definitions. For example, the interval tree from Section 2, implementing a set of `Point`s with fields `x` and `id`, is encoded as:

```
tree ITree implements Set<Point> =
| Interior(left right : ITree, xl xh : f32, idl idh : i32)
  with x in [xl, xh]    // implicit forall expands to: forall p in subtree: p.x in [xl, xh]
  with id in [idl, idh] // implicit forall expands to: forall p in subtree: p.id in [idl, idh]
  with min(id) = idl    // implicit forall expands to: min(forall p in subtree: p.id) = idl
| Leaf(p : Point) with data = p;
```

These annotations directly correspond to the augmentations described above. The first and second `Interior` annotations are scalar bounds on the fields `x` and `id`, respectively, of all points stored in the subtree. The third annotation is a reduction augmentation: it essentially marks the lower bound of `id` as *tight*, meaning it can be used to accelerate a `min` operation over `id` on a subtree. For most interval tree implementations, all bounds would be implemented as tight bounds; for brevity, we do not add these annotations to `x` and `idh` in the code snippet above. The `Leaf` annotation simply tags `p` as a set element.

Bonsai allows geometric data to be labeled via the same notation as scalar data. Our specification language explicitly models *object hierarchies*, where nodes bound discrete subsets of geometric primitives (e.g., BVHs [21, 38, 48] or R-trees [29]), rather than *spatial hierarchies* that recursively partition the underlying space itself (e.g., k-d trees [7] or octrees). Because spatial partitions allow primitives to span multiple nodes, they yield weaker structural invariants (overlap rather than strict set containment) and require complex, query-specific deduplication techniques [19], which we discuss further in Section 11. Just as Bonsai's query language supports any geometry-typed object (Section 3), any geometry-typed object may serve as a bounding volume annotation.

Consider a standard binary Axis-Aligned Bounding Box (AABB) tree[1] [59], which has a leaf and interior node that both store low and high vectors representing an AABB in 3D space. Bonsai's data modeling language represents this structure as:

```
tree TriBVH implements Set<Triangle> =
| Leaf(low high : vec<f32, 3>, prims : Triangle[]) with data = prims with AABB(low, high)
| Interior(low high : vec<f32, 3>, left right : TriBVH) with AABB(low, high);
```

This bounds annotation states that all geometries beneath the node lie within the volume. Note that while AABBs can be decomposed into per-dimension scalar annotations, other bounding volumes [21, 84] cannot. This motivates Bonsai's support for geometry-typed bounds annotations.

We note that there are some tree data structures, such as Benthin et al. [6]'s, that store augmentations for the child nodes in the parent node. This subtly shifts where pruning occurs, but not in a meaningful way; for brevity of explanation, this paper describes lowering machinery only for the case where a node stores its own augmentations, not its children.

5 Lowering Queries onto Trees

Our algorithm for fusing queries into tree traversals is a recursive bottom-up rewrite-based technique, in the spirit of StreamFusion [15]. The key observation is that a tree traversal is structured as a case analysis over node variants: leaves yield data and interior nodes recurse. Each set operator applies uniformly over this structure: modifying how data is yielded, whether recursion proceeds, or how results are combined. This uniformity allows lowering to be expressed as local rewrites at leaves and recursive nodes, which naturally compose into a single fused traversal. To make these rewrites precise, we introduce a compact intermediate representation (TTIR) that isolates exactly the constructs needed for tree traversals. We first show how to lower and fuse set operations onto trees, abstractly assuming a technique for generating the pruning and inclusion functions `always`, `maybe`, and `never` for a given predicate; Section 6 shows how we derive these functions. Algorithm 1 summarizes our lowering procedure, whose rewrite rules are presented and explained in the following subsections.

5.1 Tree Traversal Intermediate Representation (TTIR)

TTIR is a small fusion calculus for tree traversals, analogous to StreamFusion's calculus for lists [15]. It is not intended as a general-purpose language; its role is to expose the structure on which operators act: producing data at leaves, recursing at internal nodes, and aggregating results. To that end, TTIR is a functional IR with `match` and `if` for control flow, extended by a handful of domain-specific constructs:

- `yield` and `iter` return data items from leaves, for singletons and sub-collections (e.g., vectors, arrays) respectively.
- `scan` aggregates the results of subtrees; it performs a set union by default, but can apply other reductions (e.g., `sum`). It can also apply a function to subtree elements before aggregation.
- `from` recurses on subtrees.
- `upd` modifies the accumulator of a running reduction.

This design enables fusion: operator-specific rewrites can be defined locally on each construct and composed recursively. Note that this section discusses `scan` and `from` being applied to a single tree (recursing on its children). Section 7 illustrates an extension to multiple arguments, which is necessary for coiterating multiple trees, e.g., in lowering `product`.

[1]Referred to as an R-tree [29] in the database community.

The base case of lowering produces a direct traversal of a tree by `yield`ing all singleton data, `iter`ating all sub-collection data, and `scan`ning all interior nodes. This is the GenerateTreeIterator method referenced on Line 5 of Algorithm 1. To illustrate, consider the lowering of iteration on the example tree on the left, into the traversal code on the right:

```
tree ExampleTree implements Set<i32> =
| Leaf(i : i32) with data = i
| LargeLeaf(is : vec<i32, 4>) with data = is
| Interior(left right : ExampleTree);
```

```
func traverse(t : ExampleTree) = match t
 | Leaf(i) → yield i
 | LargeLeaf(is) → iter is
 | Interior(left, right) → scan t;
```

The `scan` is applied to `t` itself rather than its children. This reflects the fact that augmentations are associated with the current node, and later compilation steps will exploit them. If instead augmentations were stored on the children (as in Woop et al. [84]), the traversal would `scan` the children directly, with their results implicitly unioned.

5.2 Lowering Filters

Filtering refines which leaves yield an element, and whether recursion continues at interior nodes. LowerFilter in Algorithm 2 generates code with this behavior. At leaves, data is tested against the predicate P before being yielded or iterated. Recursive nodes are scanned if P is always true for the subtree, are recursed on if P might be true, and are pruned otherwise. Figure 4a shows the lowered filter for predicate P on `ExampleTree`.

This rewrite allows naturally fusing chained filters, as each simply inserts local conditions on `yield`, `iter`, and `scan` constructs. We show the result of such fusion in Figure 4b, which `scan`s if both predicates P and Q can be proven always true and recursively evaluates otherwise. Crucially, this means the algorithm does not attempt to enumerate all possible truth tables of compound predicates (e.g., for `filter(Q, filter(P, x))` considering every combination of `always(P)`, `always(Q)`, `maybe(P)`, `maybe(Q)`). This favors avoiding exponential code explosion and the need to produce filters specialized for simplified versions of predicates. It is possible that doing so could yield performance improvements; we simply avoid it to prevent exponential code blow-up.

```
func filter_P(t : ExampleTree) = match t
 | Leaf(i) → if P(i): yield i
 | LargeLeaf(is) → filter(|i| P(i), is)
 | Interior(left, right) →
    if always(P, t): scan t
    elif maybe(P, t): from t
```

(a) Lowering a single filter P.

```
func filter_PQ(t : ExampleTree) = match t ...
 | Interior(left, right) →
    if always(P, t):
     if always(Q, t): scan t
     elif maybe(Q, t): from t
    elif maybe(P, t) && maybe(Q, t): from t
```

(b) Interior case for filters P and Q

Fig. 4. Examples of lowered `filter`s on the `ExampleTree`.

5.3 Lowering Associative Reductions

Associative reductions can be computed hierarchically, allowing reuse of intermediate results across subtrees. Each is defined by an idempotent identity value and a commutative,[2] associative binary operator for combining subresults. These algebraic properties make them amenable to acceleration via tree augmentations that store precomputed values over subsets/subtrees.

Associative reductions on their own can be evaluated directly from augmentations stored at the root node, but also integrate naturally with filters. A reduction that wraps a filter can be fused into an efficient traversal that uses precomputed values only when the filter predicate is proven always

[2]Commutativity is required because our sets are unordered.

Algorithm 1 Recursive Query Lowering

```
1: Input: Query expression Q
2: Output: TTIR that iterates the query result
3: function LowerQ)
4:   match Q with
5:   | s ⇒ GenerateTreeIterator(s)
6:                ▷ Apply yield and iter to tagged data, and scan to nodes.
7:   | filter(P, S) ⇒ LowerFilter(P, S)                    ▷ Algorithm 2
8:   | map(F, S) ⇒
9:       rewrite Lower(S) with
10:        | yield x ⇒ yield F(x)
11:        | iter xs ⇒ iter map(F, xs)
12:        | scan tr ⇒ scan F(tr)
13:    | reduce(id, C, S) ⇒                  ▷ Lower associative reductions
14:      WrapWithAccumulator(a, id,
15:        rewrite Lower(S) with
16:        | yield x ⇒ upd a C(a, x)
17:        | iter xs ⇒ upd a reduce(a, C, xs)
18:        | scan tr ⇒ if tr has C(tr) then
19:                       upd a C(a, tr.C(tr))
20:                     else
21:                       scan<C> tr    )                ▷ End wrapper
22:   | product(S0, S1) ⇒ LowerProd(S0, S1)                ▷ Algorithm 7
23:   | min(M, S) ⇒ LowerMin(M, S)                         ▷ Algorithm 2
24:   | max(M, S) ⇒ LowerMax(M, S)
25:   | argmin(M, S) ⇒ LowerArgMin(M, S)                   ▷ Algorithm 3
26:   | argmax(M, S) ⇒ LowerArgMax(M, S)
27:   | any(P, S) ⇒ LowerAny(P, S)                         ▷ Algorithm 4
28:   | all(P, S) ⇒ LowerAll(P, S)
29: end function
```

Algorithm 2 Filter and Min Lowering

```
1: Input: Query predicate P and set expression S
2: Output: TTIR that iterates the query result
3: function LowerFilter(P, S)
4:   rewrite Lower(S) with
5:    | yield x ⇒ if P(x): yield x
6:    | iter xs ⇒ iter filter(P, xs)
7:    | scan tr ⇒ if always(P, tr): scan tr
8:                 elif maybe(P, tr): from tr
9:    | from tr ⇒ if maybe(P, tr): from tr
10: end function
11: Input: Query metric M and set expression S
12: Output: TTIR that stores the result in accumulator a
13: function LowerMin(M, S)
14:   WrapWithAccumulator(a, ∞,
15:     rewrite Lower(S) with
16:      | yield x ⇒ upd a minb(a,M(x))
17:      | iter xs ⇒ upd a minb(a, min(M, xs))
18:      | scan tr ⇒ if tr has min(M, tr) then
19:                    upd a minb(a, min(M, tr))
20:                  else if tr has max(M, tr) then
21:                    upd a minb(a, max(M, tr))
22:                    if maybe(min(M(tr)) < a):
23:                      from tr
24:                  else
25:                    if maybe(min(M(tr)) < a):
26:                      from tr
27:      | from tr ⇒ if maybe(min(M(tr)) < a):
28:                    from tr    )          ▷ End accumulator wrapper
29: end function
```

true, recursively evaluating the query predicate otherwise, as illustrated in Section 2. Such fusion is asymptotically useful, avoiding the need to store the filter result before aggregation.

Lowering an associative reduction is also done through local rewrites on TTIR constructs, illustrated in Lines 15–22 of Algorithm 1: `yield` and `iter` apply the operator on the running accumulator and the yielded set elements, `scan` incorporates subtree augmentations when available, and `from` continues to recurse. For `scan`s, there are two (compile-time) cases: Lines 19–20 handle the case that a tree node stores the subresult, and Line 22 applies a reduction scan if the tree node does not. Note that associative reductions that are also idempotent (e.g., `min`/`max` and `any`/`all`) exhibit extra pruning potential; they incorporate value-based pruning when possible.

5.4 Lowering Idempotent Reductions

Idempotent reductions, including `min`, `max`, `argmin`, `argmax`, `any`, and `all`, form reductions over a lattice structure. They allow subtree pruning whenever it can be proven that a subtree cannot affect the final result. For example, consider the `min-id` query from Section 2. During traversal, if it can be determined that no value in a subtree has a smaller `id` than the current best, the entire subtree can be skipped. This observation generalizes to all idempotent reductions: whenever it can be proven that a subtree does not affect the final result, it need not be visited.

LowerMin in Algorithm 2 illustrates our lowering rewrite for `min` with a metric applied. `yield` and `iter` simply reduce on the leaf data, and `scan` and `from` are rewritten locally to exploit subtree metrics. Note that `scan` lowering first applies pruning via the associative reduction property, using stored metadata if available, but otherwise falls back to value-based pruning that the idempotent property enables. This lowering leverages minimum-value metadata if available, but can otherwise use maximum-value metadata to prune the search space conservatively. Such optimizations are in line with state-of-the-art minimum-distance queries [22, 70].

To illustrate how different stored metadata enable distinct optimizations, consider a min-reduction on a filtered set with filter predicate P. When lowered on a tree that stores the minimum metric (`MinTree` traversal on the left, below), that stored subtree's value can be used to update the accumulator any time the predicate P is proven true (inclusion-based skipping). When P is *maybe*

true, the value can still aid further pruning: if it exceeds the running minimum, the subtree cannot contribute to the result. The same query on a tree that only stores the maximum metric (`MaxTree` traversal on the right, below) offers no inclusion-based skipping, but the stored maximum can still induce a tighter bound on the minimum value.

```
func min_wmin(t : MinTree, acc : i32&) =
 match t
 | Leaf(i) → if P(i): upd acc minb(acc, i)
 | Interior(left, right, min_i) →
    if always(P, t): upd acc minb(acc, min_i)
    elif maybe(P, t):
      if min_i < acc: from t
```

```
func min_wmax(t : MaxTree, acc : i32&) =
 match t
 | Leaf(i) → if P(i): upd acc minb(acc, i)
 | Interior(left, right, max_i) →
    if always(P, t):
      upd acc minb(acc, max_i); from t
    elif maybe(P, t): from t
```

Lowerings for `argmin` and `argmax` mirror those for `min` and `max`, but also track the element achieving the extremum (Algorithm 3). `any` and `all` can early-return when the predicate is proven always or never true on a subtree, as shown in Algorithm 4.

Algorithm 3 Argmin Lowering

```
Input: Query metric M and set expression S
Output: TTIR that stores the result in accumulator a
function LowerArgMin(M, S)
  WrapWithAccumulator(a, {∞, null},
   rewrite Lower(S) with
    | yield x ⇒ upd a argminb(a, {M(x), x})
    | iter xs ⇒ upd a argminb(a, argmin(M, xs))
    | scan tr ⇒ if tr has max(M, tr) then
                  upd a minb(a, max(M, tr))
                  if maybe(min(M(tr)) < a): from tr
                 else
                  if maybe(min(M(tr)) < a): from tr
    | from tr ⇒ if maybe(min(M(tr)) < a): from tr  )
end function
```

Algorithm 4 Any Lowering

```
Input: Query predicate P and set expression S
Output: TTIR that stores the result in accumulator a
function LowerAny(P, S)
  WrapWithAccumulator(a, false,
   rewrite Lower(S) with
    | yield x ⇒ upd a (a ∨ P(x))
    | iter xs ⇒ upd a (a ∨ any(P, x))
    | scan tr ⇒ if always(P, tr):
                  upd a true
                elif ¬a ∧ maybe(P, tr):
                  from tr
    | from tr ⇒ if ¬a ∧ maybe(P, tr):
                  from tr   )          ▷ End accumulator wrapper
end function
```

6 Predicate Analysis

Lowering filters and idempotent reductions requires generating necessary (`maybe`) and sufficient (`always`) conditions, a process we call *predicate analysis*. Our implementation uses symbolic interval analysis. A key insight is that symbolic interval analysis [61] is sufficient for deriving necessary and sufficient conditions in linear time through a simple AST traversal of the query predicate, in contrast to general necessary/sufficient condition synthesis (e.g., inductive invariant generation [17, 64]), which often requires exponential search. Our evaluation shows that interval-derived conditions are tight enough for real-world applications and match the state-of-the-art systems that manually implement query-specific pruning. Symbolic analysis thus provides a practical and efficient way to derive pruning conditions directly from the structure of query predicates.

To clarify the relation between interval analysis and predicate analysis: a necessary condition (`maybe`) is a condition implied by the predicate, and is thus an *upper bound* of the predicate. Likewise, a sufficient condition (`always`) implies the predicate, and is thus a *lower bound* of the predicate. For scalar expressions, it is therefore sufficient to derive these conditions by applying standard symbolic interval analysis to generate the bounds on a boolean expression. Many geometric relationships can be similarly bounded by necessary and sufficient conditions (see Section 6.3).

Notably, pruning works best when bounds are *tight*: this means that a lower bound should be the weakest (most frequently true) sufficient condition, and the upper bound should be the strongest (least frequently true) necessary condition. These correspond to scanning as often as possible and pruning as often as possible, respectively. This is the ideal goal of generating such bounds, but we make no guarantees that our algorithm derives the weakest and strongest bounds (notably, for some correlated expressions, e.g., $x - x$, interval analysis is known to produce non-tight bounds).

We first describe our notation, then provide minor background on scalar interval analysis, and lastly illustrate our extension to handle geometric predicates such as intersection and containment.

6.1 Notation and Terminology

We denote the lower bound of an expression E (either boolean or numeric) as $\lfloor E \rfloor$, and the upper bound as $\lceil E \rceil$. Upper bounds and lower bounds are equivalent to `always` and `maybe` by:

$$\texttt{always}(E) = \lfloor E \rfloor \qquad \texttt{maybe}(E) = \lceil E \rceil$$

In interval analysis, it is important to note the difference between *varying* parameters and *uniform* parameters: *varying* parameters are values in a predicate that can take multiple values, bounded by either an interval (in the scalar case) or a volume (in the geometric case); *uniform* parameters are constant with respect to the queried data. For example, in a standard range query that searches for all x such that $low \leq x \leq high$, x is a varying parameter and low and $high$ are uniform parameters. Interval analysis rules (including ours in Section 6.3) are frequently defined differently depending on which operands of an expression are varying or uniform.

If an expression cannot be bounded, its bounds default to the limits of its type, e.g., $[\texttt{false}, \texttt{true}]$ for boolean expressions, $[0, \texttt{UINTMAX}]$ for unsigned integers, and $[-\infty, \infty]$ for floats.

6.2 Background: Scalar Interval Analysis

Symbolic interval analysis derives bounds recursively in a bottom-up traversal of the predicate's AST. Varying parameters are replaced with their intervals, and operators are evaluated on the intervals themselves by considering the monotonicity of an operator. We illustrate the reasoning behind comparisons and boolean combinators here; additional operators are described in Appendix A.

Comparisons. The comparison of two numbers can be bounded by a comparison of the ranges that bound each number. Consider the expression $x < y$: if x and y are varying, then the expression $x < y$ *may* be true if x's lower bound is less than y's upper bound (otherwise, all values that y can be are less than all values that x can be). Conversely, $x < y$ is *always* true if the upper bound of x is less than the lower bound of y. This reasoning produces the following bounds:

$$\lceil x < y \rceil \mapsto \lfloor x \rfloor < \lceil y \rceil \qquad \lfloor x < y \rfloor \mapsto \lceil x \rceil < \lfloor y \rfloor$$

The $\leq$ operator has the same monotonicity of operands as $<$, and can be bounded in the same way. Likewise, similar reasoning applies to equality, though the lower bound requires that the intervals of the arguments each contain a single value:

$$\lceil x = y \rceil \mapsto \lfloor x \rfloor \leq \lceil y \rceil \wedge \lfloor y \rfloor \leq \lceil x \rceil \qquad \lfloor x = y \rfloor \mapsto \lfloor x \rfloor = \lfloor y \rfloor \wedge \lceil x \rceil = \lceil y \rceil \wedge \lfloor x \rfloor = \lceil x \rceil$$

Boolean combinators. Boolean *and*, $\wedge$, and boolean *or*, $\vee$, are both monotonically increasing in their arguments, and are therefore bounded by the bounds of their arguments. Boolean *negation*, $\neg$, is monotonically decreasing, and therefore is upper bounded by the negation of the lower bound of its argument, and lower bounded by the negation of the argument's upper bound.

$$\lceil a \wedge b \rceil \mapsto \lceil a \rceil \wedge \lceil b \rceil \qquad \lceil a \vee b \rceil \mapsto \lceil a \rceil \vee \lceil b \rceil \qquad \lceil \neg a \rceil \mapsto \neg \lfloor a \rfloor$$

$$\lfloor a \wedge b \rfloor \mapsto \lfloor a \rfloor \wedge \lfloor b \rfloor \qquad \lfloor a \vee b \rfloor \mapsto \lfloor a \rfloor \vee \lfloor b \rfloor \qquad \lfloor \neg a \rfloor \mapsto \neg \lceil a \rceil$$

While such bounds are well-established [75], applying them to generate tree-pruning functions is, to our knowledge, novel. We further show that this reasoning naturally extends to spatial operators.

6.3 Geometric Bounds

In the same way that scalar boolean operators are bounded by their necessary and sufficient conditions parameterized by their bounding intervals, geometric boolean operators can be bounded by necessary and sufficient conditions parameterized by their bounding volumes.

While this analogy is conceptually straightforward, applying scalar interval analysis directly to implementations of geometric predicates such as `intersects` or `contains` is ineffective in practice. Such implementations are typically hundreds of lines of specialized geometric code [59], and interval propagation through this code rarely exposes the high-level spatial relationships necessary for pruning. Instead, we aim to derive these bounds directly from the semantics of each predicate.

Each geometric predicate in Figure 1 is binary, and can be analyzed by considering three cases: when the first argument is varying (contained by a bounding volume) and the second is uniform, when the first argument is uniform and the second is varying, and when both arguments are varying. Note that symmetric operators like `intersects` only have two cases, and in the case that both arguments are uniform, the expression itself is uniform and is a singular-valued interval.

An upper bound is implied by the predicate, and a lower bound implies the predicate. Thus, the upper bound and lower bound of a predicate P satisfy:

$$P(X) \wedge \mathsf{bounded}(X) \rightarrow \lceil P \rceil \qquad \lfloor P \rfloor \wedge \mathsf{bounded}(X) \rightarrow P(X)$$

where $\mathsf{bounded}(X)$ asserts that all varying parameters are bounded by their respective bounding volumes, and both the upper and lower bounds refer only to the uniform parameters of P and the bounding volumes of the varying parameters.

As a guiding example, consider searching for all objects contained within a query sphere. If a subtree is ever fully contained within the query sphere, then all objects in the subtree must be contained within the query sphere. Alternatively, if the subtree's bounding volume intersects the query sphere, it is possible that some objects in the subtree may be contained within the query sphere. These give rise to the bounds on the containment predicate with a uniform first parameter, u (`sphere`), and a varying second parameter, v bounded by a bounding volume V_v:

$$\lceil \texttt{contains}(u, v) \rceil \mapsto \texttt{intersects}(u, V_v) \qquad \lfloor \texttt{contains}(u, v) \rfloor \mapsto \texttt{contains}(u, V_v)$$

When the first argument is varying and the second is uniform (e.g., in a query which searches for all objects that *contain* a query object), the predicate can be upper-bounded, but a bounding volume is not enough to prove objects in a bounding volume *always* contain a query object, so there is no lower bound based on this augmentation.

$$\lceil \texttt{contains}(v, u) \rceil \mapsto \texttt{intersects}(V_v, u) \qquad \lfloor \texttt{contains}(v, u) \rfloor \mapsto \texttt{false}$$

Likewise, when both arguments are varying, there is no lower bound, and the upper bound is simply the intersection of the two bounding volumes (if the bounding volumes do not intersect, it is impossible for any object in one to contain any object in the other).

We generally see this pattern in geometric predicates: almost all can be upper bounded, which means they benefit from pruning subtrees where the predicate can never be true, but many do not have lower bounds, so cannot often be proven to be *always* true in a given subtree. We provide the rest of our lower bound and upper bound rules for geometric predicates in Algorithm 5 and Algorithm 6 below. Note that distances and ordering predicates are monotonic functions, and therefore are bounded by simply replacing any varying arguments with their bounding volumes. u denotes uniform geometric values, v, v_0, and v_1 denote varying geometric values with corresponding bounding volumes V_v, V_0, and V_1. Ordering relationships (e.g., $\leq$ and $<$) are monotonic and are therefore bounded by rewriting varying arguments to their bounding volumes (not included for brevity). Metric relationships, `distmin` and `distmax`, are always lower bounded by `distmin` and

upper bounded by `distmax` applied to the same arguments but replacing varying arguments with their bounding volumes (uniform parameters are bounded by themselves for the sake of brevity).

Algorithm 5 Geometric Upper Bounds

Input: Geometric predicate E
Output: Upper bound of E
function $\lceil E \rceil$
  match E **with**
  | $\texttt{contains}(u, v) \mapsto \texttt{intersects}(u, V_v)$
  | $\texttt{contains}(v, u) \mapsto \texttt{intersects}(V_v, u)$
  | $\texttt{contains}(v_0, v_1) \mapsto \texttt{intersects}(V_0, V_1)$
  | $\texttt{covers}(u, v) \mapsto \texttt{intersects}(u, V_v)$
  | $\texttt{covers}(v, u) \mapsto \texttt{covers}(V_v, u)$
  | $\texttt{covers}(v_0, v_1) \mapsto \texttt{intersects}(V_0, V_1)$
  | $\texttt{disjoint}(u, v) \mapsto \neg\texttt{contains}(u, V_v)$
  | $\texttt{disjoint}(v, u) \mapsto \neg\texttt{within}(V_v, u)$
  | $\texttt{within}(u, v) \mapsto \texttt{within}(u, V_v)$
  | $\texttt{within}(v, u) \mapsto \texttt{intersects}(V_v, u)$
  | $\texttt{within}(v_0, v_1) \mapsto \texttt{intersects}(V_0, V_1)$
  | $\texttt{equals}(u, v) \mapsto \texttt{within}(u, V_v)$
  | $\texttt{equals}(v, u) \mapsto \texttt{contains}(V_v, u)$
  | $\texttt{equals}(v_0, v_1) \mapsto \texttt{intersects}(V_0, V_1)$
  | $\texttt{intersects}(u, v) \mapsto \texttt{intersects}(u, V_v)$
  | $\texttt{intersects}(v, u) \mapsto \texttt{intersects}(V_v, u)$
  | $\texttt{intersects}(v_0, v_1) \mapsto \texttt{intersects}(V_0, V_1)$
  | $\texttt{touches}(u, v) \mapsto \texttt{intersects}(u, V_v)$
  | $\texttt{touches}(v, u) \mapsto \texttt{intersects}(V_v, u)$
  | $\texttt{touches}(v_0, v_1) \mapsto \texttt{intersects}(V_0, V_1)$
  | $\_ \mapsto \texttt{true}$
end function

Algorithm 6 Lower Bounds and Metric Bounds

Input: Geometric predicate E
Output: Lower bound of E
function $\lfloor E \rfloor$
  match E **with**
  | $\texttt{contains}(u, v) \mapsto \texttt{contains}(u, V_v)$
  | $\texttt{covers}(u, v) \mapsto \texttt{covers}(u, V_v)$
  | $\texttt{disjoint}(u, v) \mapsto \texttt{disjoint}(u, V_v)$
  | $\texttt{disjoint}(v, u) \mapsto \texttt{disjoint}(V_v, u)$
  | $\texttt{disjoint}(v_0, v_1) \mapsto \texttt{disjoint}(V_0, V_1)$
  | $\texttt{intersects}(u, v) \mapsto \texttt{contains}(u, V_v)$
  | $\texttt{intersects}(v, u) \mapsto \texttt{within}(V_v, u)$
  | $\texttt{within}(v, u) \mapsto \texttt{within}(V_v, u)$
  | $\_ \mapsto \texttt{false}$
end function
Input: Geometric metric E **Output**: Lower bound on E
function $\lfloor E \rfloor$
  match E **with**
  | $\texttt{distmin}(v_0, v_1) \mapsto \texttt{distmin}(V_0, V_1)$
  | $\texttt{distmax}(v_0, v_1) \mapsto \texttt{distmin}(V_0, V_1)$
end function
Input: Geometric metric E **Output**: Upper bound on E
function $\lceil E \rceil$
  match E **with**
  | $\texttt{distmin}(v_0, v_1) \mapsto \texttt{distmax}(V_0, V_1)$
  | $\texttt{distmax}(v_0, v_1) \mapsto \texttt{distmax}(V_0, V_1)$
end function

7 Joins as Tree Traversals

The previous two sections introduced the machinery required to compile filters and reductions over sets. Now we show how this machinery can also be used to generate efficient code for non-equijoins, including spatial joins. The key insight is that a join predicate can be analyzed like a filter predicate with multiple varying parameters, enabling pruning of both sides of the join.

Traditional database systems implement joins using strategies like hash join (build a hash table on one side, probe from the other) or sort-merge join (sort both sides, then merge). These strategies work well for equijoins (equality predicates), but generally do not extend to non-equijoins with arbitrary predicates, particularly geometric predicates like intersection or containment. Non-equijoins are often lowered to a *nested join*, i.e., quadratic enumeration of all pairs followed by predicate evaluation.

Spatial databases [29] and collision detection algorithms [21] have long used tree-based indexes to accelerate spatial queries. Our compilation approach generalizes these techniques to arbitrary predicates by treating join predicates the same way we treat filter predicates: analyzing them to generate pruning conditions and compile efficient tree traversals. We extend two well-known strategies: a single-index join analogous to a hash join, and a dual-index join analogous to sort-merge join and inspired by dual-tree traversals in collision detection and spatial joins.

7.1 Single-Index Join

A straightforward non-equijoin implementation with our techniques is an iterate-locate pattern akin to hash-join: a tree (index) is built on one side; the other side iterates through its elements, locating values in the tree that satisfy the join predicate. In functional notation, this is simply:

```
single(set0 : Set<T>, set1 : Set<S>) = map(|t : T| (t, filter(|s : S| P(t, s), set1)), set0);
```

This expression iterates over each element `t` in `set0` (the outer loop), and for each `t`, filters `set1` to find all `s` that satisfy the join predicate `P(t, s)`. Note that for the purpose of predicate analysis, `t` is a uniform parameter and `s` is a varying parameter.

Output. The return type of this function is a set of tuples where each tuple contains an element of type `T` and a *set* of elements of type `S`: `Set<(T, Set<S>)>`. This represents an implicit groupby operation, grouping matching `S` elements by their corresponding `T` element. If the desired output is a flat list of pairs `Set<(T, S)>`, a final flattening step is required.

Performance Characteristics When n is the size of the outer set and m is the size of the inner set, the worst-case runtime is $\Theta(n \cdot m)$, when no (or limited) pruning is possible, and the tree traversal takes linear time. For highly-selective filters that turn the filter logarithmic (e.g., a range predicate), the complexity is $O(n \cdot \log |m| + m \cdot \log |m|)$, where the first term corresponds to n tree traversals and the second term corresponds to the tree build complexity (assuming standard tree construction algorithms [38]). As in databases, the choice of which side to index can significantly impact performance: indexing the smaller set minimizes build and materialization costs, but indexing the set with better spatial locality may enable more effective pruning. The outer loop is also trivially parallelizable.

7.2 Dual-Index Join

In some cases, we can further improve performance with a coiterative join strategy akin to sort-merge join, inspired by the dual-tree traversals in collision detection [21] and spatial joins [29]. By building trees on *both* sides of the join, we can prune subtrees from both input sets simultaneously. Conceptually, a dual-index join computes a filtered Cartesian product:

```
dual(set0 : Set<T>, set1 : Set<S>) = filter(|t : T, s : S| P(t, s), product(set0, set1));
```

Algorithm. The dual tree traversal computes the filtered Cartesian product by recursing on a pair of nodes, one from each side of the join, and checking if the join predicate can be true in those two subsets. Abstractly, the traversal is lowered to the traversal in Figure 5a, which is the result of applying the lowering rules for `product`, provided in Algorithm 7, composed with the `filter` lowering in Algorithm 2, onto a simple binary tree with `Leaf` and `Interior` variants.

Example. For collision detection (where the filter predicate is `intersects`), this abstract code is lowered to Figure 5b, which matches the standard dual tree collision detection algorithms [21].

Output. The return type is a set of tuple pairs.

Performance Characteristics. Let n and m be the sizes of the outer and inner sets. The worst-case runtime is $\Theta(n \cdot m)$ when no pruning occurs. The runtime for this class of algorithms depends heavily on the predicate and the data distributions. For example, if the root nodes are disjoint w.r.t the query predicate, the traversal terminates immediately in $O(1)$ time. In general, it is not possible to give a meaningful asymptotic bound without a specific predicate or data distribution, but runtime is always lower-bounded by construction of both trees: $\Omega(n \cdot \log |n| + m \cdot \log |m|)$. Parallelization is nontrivial (there is no trivially parallelizable outer loop), but the recursive traversal can be parallelized via work stealing [9].

Algorithm 7 Product Lowering

```
1: Input: Set expressions S_0 and S_1
2: Output: TTIR that iterates the product of the two sets
3: function LowerProduct(S_0, S_1)
4:   rewrite Lower(S_0) with
5:   | yield x ⇒
6:     rewrite Lower(S_1) with
7:     | yield y ⇒ yield (x, y)
8:     | iter ys ⇒ iter map(|y| (x, y), ys)
9:     | scan t1 ⇒ scan {x} t1
10:    | from t1 ⇒ from {x} t1
11:  | iter xs ⇒
12:    rewrite Lower(S_1) with
13:    | yield y ⇒ iter map(|x| (x, y), xs)
14:    | iter ys ⇒ iter product(xs, ys)
15:    | scan t1 ⇒ scan {xs} t1
16:    | from t1 ⇒ from {xs} t1
17:  | scan t0 ⇒
18:    rewrite Lower(S_1) with
19:    | yield y ⇒ scan t0 {y}
20:    | iter ys ⇒ scan t0 {ys}
21:    | scan t1 ⇒ scan t0 t1
22:    | from t1 ⇒ from t0 t1
23:  | from t0 ⇒
24:    rewrite Lower(S_1) with
25:    | yield y ⇒ from t0 {y}
26:    | iter ys ⇒ from t0 {ys}
27:    | scan t1 ⇒ from t0 t1
28:    | from t1 ⇒ from t0 t1
29: end function
```

```
func dual_traversal(t0 : Tree, t1 : Tree) =
 match t0, t1 with
 | Leaf(d0), Leaf(d1) →
  if P(d0, d1): yield (d0, d1)
 | Leaf(d0), Interior(l, r) →
  if always(P, d0, t1): scan t0, t1
  elif maybe(P, d0, t1): from t0, t1
  | Interior(l, r), Leaf(d1) →
  if always(P, t0, d1): scan t0, t1
  elif maybe(P, t0, d1): from t0, t1
  | Interior(l0, r0), Interior(l1, r1) →
  if always(P, t0, t1): scan t0, t1
  elif maybe(P, t0, t1): from t0, t1
```

(a) Generic dual traversal join of two trees under predicate P. `scan` and `from` recurse on each pair of children (the product of children).

```
func collisions(t0 : Tree, t1 : Tree) =
 match t0, t1 with
 | Leaf(d0), Leaf(d1) →
  if intersects(d0, d1): yield (d0, d1)
 | Leaf(d0), Interior(l, r) →
  if contains(d0, t1): scan t0, t1
  elif intersects(d0, t1): from t0, t1
  | Interior(l, r), Leaf(d1) →
  if contains(d1, t0): scan t0, t1
  elif intersects(t0, d1): from t0, t1
  | Interior(l0, r0), Interior(l1, r1) →
  if intersects(t0, t1): from t0, t1
```

(b) Dual tree traversal with an `intersects` join predicate. When both arguments vary, no sufficient condition exists, so the final `always` is false.

Fig. 5. Dual-tree traversal specialization. The generic dual traversal **(a)** lowers to collision detection **(b)** by instantiating predicate P as `intersects`. `always` and `maybe` specialize to `contains` and `intersects`.

8 Code Generation and Implementation

Bonsai compiles to C++. Tree layouts are specified using Scion [30] to provide compact layouts for self-comparisons and match other systems in cross-system comparisons in Section 9.

Outputs. Every Bonsai tree traversal produces some accumulated value, the output type. Every generated function has a reference to an accumulator of this type as the output parameter.

Generating sets. If the output type of a query is a set, then the accumulator is our custom C++ `Set<T>` type: `yield`s become appends and `iter`s become grouped appends.

Producing scalars. If the output type of a query is a scalar (e.g., `min`), then the accumulator is just the C++ version of the scalar type, and `upd` is lowered to a mutation.

Lowering `from`. `from` is always lowered into recursive calls applied to all children of the argument type. For `from`s applied to multiple parameters (from lowering a `product`), these are lowered to the Cartesian product of children nodes, as in standard dual tree traversals [21].

Lowering `scan`. Every scan is lowered into the standard tree traversal on the base tree type, with any `map` function applied to the leaf primitives, and any reduction operator lowered last. Note that in the case of multiple tree parameters (from lowering a `product`), `scan`s again become Cartesian products, and aggregate functions are applied on the products.

Example. Consider lowering a collision detection join that counts[3] the number of collisions:

```
f(s0 : Set<T>, s1 : Set<S>) = count(filter(|t : T, s : S| intersects(t, s), product(s0, s)));
```

Bonsai fuses the `count` into the traversal in Figure 5b. The lowered C++ (Figure 6) passes the accumulator as the final parameter to both traversal functions (querying, scanning). While specialized scans could be generated (e.g., when the first parameter is always a leaf), we avoid this to prevent a combinatorial number of specialized variants. Such specialization could be profitable.

9 Evaluation

We evaluate two primary claims for both regular filters and our generalized join algorithms:

(1) Bonsai achieves pruning efficiency and runtime performance comparable to hand-written tree traversals; and

[3] A `count` aggregate is a map that maps all elements to 1 followed by a sum reduction.

```
void f(Tree t0, Tree t1, u64 &c) {
 if (is_leaf(t0)) {
  if (is_leaf(t1)) {
    if (intersects(t0.prim, t1.prim)) c++;
  } else { // t1 is interior
   if (contains(t0.prim, t1.vol)) {
    f_scan(t0, t1, c);
   } else if (intersects(t0.prim, t1.vol)) {
    f(t0, t1.left, c);
    f(t0, t1.right, c);
   }
  }
 } else { // t0 is interior
  if (is_leaf(t1)) {
   // flip the t0 leaf t1 interior case above
  } else { // t1 is interior
  if (intersects(t0.vol, t1.vol)) {
   f(t0.left, t1.left, c);
   f(t0.right, t1.left, c);
   f(t0.left, t1.right, c);
   f(t0.right, t1.right, c); }}}}
```

(a) The count of collisions, lowered to C++.

```
void f_scan(Tree t0, Tree t1, u64 &c) {
 if (is_leaf(t0)) {
  if (is_leaf(t1)) {
   c++;
  } else {
   // t1 is interior
   f_scan(t0, t1.left, c);
   f_scan(t0, t1.right, c);
  }
 } else {
  // t0 is interior
  if (is_leaf(t1)) {
   f_scan(t0.left, t1, c);
   f_scan(t0.right, t1, c);
  } else {
   // t1 is interior
   f_scan(t0.left, t1.left, c);
   f_scan(t0.right, t1.left, c);
   f_scan(t0.left, t1.right, c);
   f_scan(t0.right, t1.right, c);
}}}
```

(b) Lowered scan with a count accumulator.

Fig. 6. Fused, final lowered C++ of a coiterating two trees, and counting collisions over leaf and interior variants. The user supplies intersects/contains code (in C++ or in Bonsai's kernel language).

(2) Bonsai can generate pruned traversals that existing systems are missing, resulting in improved performance for certain queries.

We also present an ablation study quantifying the impact of fusion on compound filter-reduction queries, and analyze how data distribution impacts the performance of different join strategies.

9.1 Methodology

We evaluate on an Intel Core i9-14900K (3.2 GHz, 24 cores) with 196 GB DDR4 RAM, 896 KiB of L1 data cache, 32 MiB of L2 cache, and 36 MiB of L3 cache. Benchmarks run single-threaded to isolate asymptotic behavior and bound to performance cores via `numactl`. Generated kernels are compiled with `clang++ 21.1.3`. Each benchmark runs 7 times; we discard the fastest and slowest and report the mean of the remaining 5 runs. Runs timeout after 30s.

For graphics queries, we compare against Fastest Closest Points in the West (FCPW) [70] and the Flexible Collision Library (FCL) [55]. For scalar queries, we compare to SQLite [33] and DuckDB [60], both configured to run entirely in memory. Graphics models are drawn from the McGuire archive [46]; scalar benchmarks use synthetic uniformly random data, except the salary join, which follows Khayyat et al. [40]. All generated data uses a fixed seed (42) for reproducibility.

Unless otherwise stated, trees are built using a standard recursive spatial median-split algorithm [38]. Optimizing tree construction is not a focus of this work; reported runtimes for joins include both with and without tree build times for transparency.

9.2 Compile Times

Compilation times are interactive, i.e., 1–5 ms per query for all queries. Compiling the generated C++ code with `-O3` takes approximately 50 ms per query. This could be reduced with lighter compiler optimizations or less template metaprogramming in our benchmarking framework.

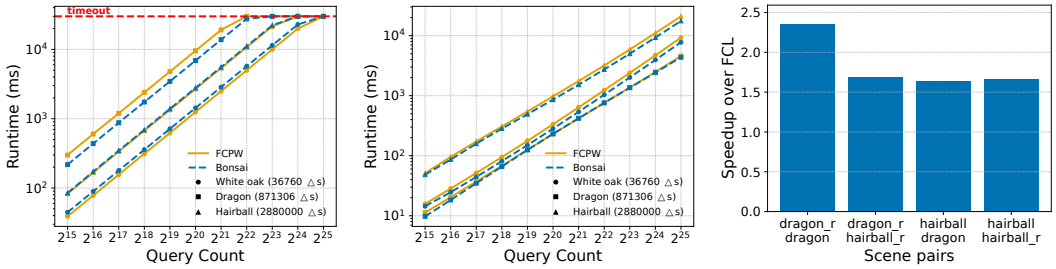
9.3 Comparison to Hand-Written Traversals

9.3.1 Graphics Queries. To demonstrate that Bonsai's geometric pruning and lowering match state-of-the-art performance, we compare to three representative and optimized graphics queries: closest point queries, ray tracing, and collision detection. Closest point queries (`min` over `distmin`) and ray tracing (`argmin` on `filter`) use a single-index join, iterating over an array of points or rays and querying a tree built on the scene geometry (e.g., triangles). We directly copy FCPW's tree topology and layout for fair comparison. For collision detection, we compare our lowered `filter` of a `product` against FCL's hand-written dual-index traversal, matching FCL's tree layout, and made a best effort at duplicating their build algorithm.

Closest point queries: Figure 7a shows runtimes of Bonsai and FCPW [70] closest point queries over three scenes and varying numbers of randomly generated query points within each scene's bounding box. Bonsai is on par with FCPW: it is on average (arithmetic mean) 0.87× the throughput of FCPW on the White Oak scene; 1.38× on the Dragon scene; and 0.97× on the Hairball scene. We see two differences between FCPW's code and Bonsai-generated code that could explain the performance differences: although FCPW has better vector instruction usage (via the Eigen library [28]), it records information beyond just the closest point to the query point. We believe the speedup on Dragon could be due to Bonsai's query specialization removing such metadata.

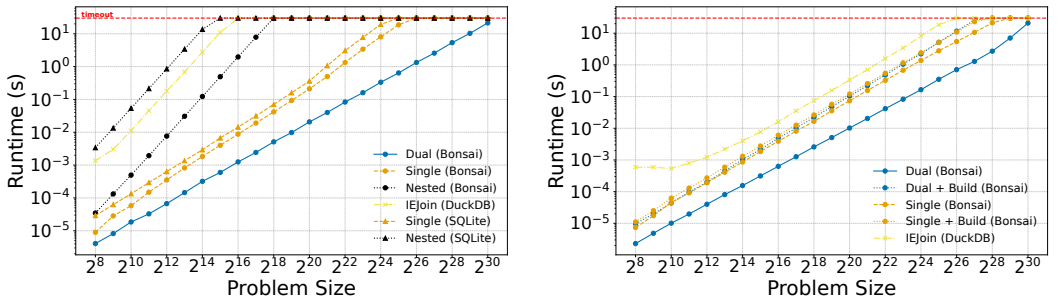
Closest hit ray tracing: Figure 7b compares Bonsai with FCPW on ray tracing performance. Again, Bonsai is on par with FCPW: the range of speedups for the White Oak scene is 1.09×–1.19× (avg. of 1.17×); for the Dragon scene: 0.99×–1.16× (avg. of 1.04×); and the Hairball scene: 1.06×–1.19× (avg. of 1.13×). As before, we believe any speedup comes from query specialization, as manual inspection confirms the tree traversals generated by Bonsai are identical to FCPW's hand-written traversals. Because ray-bounding box and ray-triangle intersection are less vectorizable than the point-box and point-triangle distance queries used for closest point queries, FCPW's improved vector instruction support offers less advantage here. However, extending Bonsai with improved vectorization support remains a valuable future direction.

Collision detection: Figure 7c shows a speedup plot of Bonsai versus FCL [55] across scene pairs from FCL's benchmarking suite. Bonsai-generated code is consistently faster—not due to improved pruning, but because FCL relies heavily on virtual function dispatch for geometry intersection and includes profiling hooks that cannot be disabled. Bonsai achieves a 2.36× speedup on dragon–dragon rotated scenes (8,688 collisions); 1.69× on dragon rotated–hairball rotated (48,238 collisions); 1.64× on hairball–dragon (123,055 collisions); and 1.66× on hairball–hairball rotated (5,118,441 collisions). These results indicate that Bonsai matches the pruning efficiency of hand-written



(a) Closest point queries. (b) Ray tracing (primary rays). (c) Collision detection.

Fig. 7. Comparison of Bonsai versus SotA closest point queries and ray tracing in FCPW [70], and collision detection versus FCL [55]. Lower is better for runtimes **(a)** and **(b)**, and higher is better for speedups **(c)**.



(a) Range Join (b) Salary Join

Fig. 8. Runtime comparisons with state of the art (SotA) database management systems. Lower is better. In **(a)**, we plot the runtimes of a 2D range join (join predicate $x_0 \in [x_1 - k, x_1 + k] \wedge y_0 \in [y_1 - k, y_1 + k]$). Without an index, SQLite produces a nested join, but with an index on (x_1, y_1), it will perform a single-index join; DuckDB performs an IEJoin regardless of the presence of an index. Bonsai's nested and single-index joins are on-par with SQLite, but the dual index join out-performs all of them. In **(b)** we directly compare to DuckDB's IEJoin on a benchmark from the original paper [40], counting the number of employees in a database who make less money but pay higher taxes than a peer. Even incorporating Bonsai's unoptimized tree construction times, both Bonsai generated joins out-perform DuckDB's custom join.

collision systems, while the speedup shows the benefit of specializing code at compile time instead of relying on runtime dynamic dispatch.

9.3.2 Range and Inequality Joins. Relational DBMSs like SQLite [33] use hand-optimized tree traversals (e.g., B-trees) to accelerate range queries of the form `x in [low, high]`. To evaluate Bonsai in this context, we test *range joins*, where each row in one table performs a range query over the other. Figure 8a shows Bonsai 's single-index join matches SQLite's native traversal, while its dual-index join is faster by leveraging both indexes. Our nested joins also outperform SQLite's, confirming them as valid baselines. Bonsai's gains stem from SQLite's interpreter overhead.

There also exist specialized (non-tree-based) algorithms for *inequality joins* [40], a class of open-ended range predicates. We evaluate Bonsai-generated code against DuckDB's IEJoin on one of its benchmarks (Figure 8b), and find that it outperforms the native implementation, even when including index build time. Overall, these findings establish that Bonsai can reproduce the performance of expert-written join algorithms while generalizing to new join types, making it a strong foundation for exploring advanced query operators beyond current database capabilities.

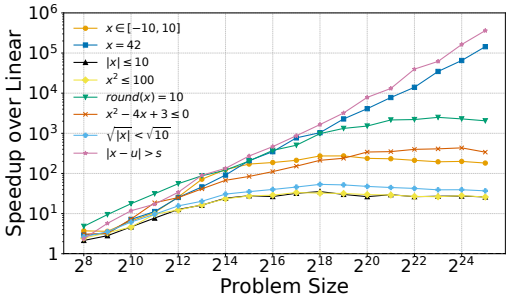
9.4 Comparison to Non-Pruning Code

We evaluate Bonsai on queries for which state-of-the-art systems perform linear or quadratic scans. We include filters, reductions, and joins, highlighting cases where Bonsai generates pruned traversals that traditional systems do not map to tree queries.

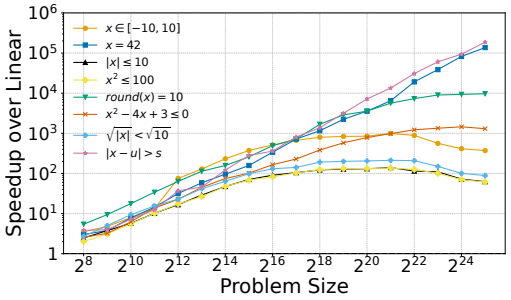
9.4.1 Filters and Reductions. Table 1 lists 12 linear queries that we use to highlight the asymptotic benefits of tree traversals. The first two queries (a range and a point query) are accelerated by most database systems we examined. None of the systems perform index scans for the remaining ten, despite seven being algebraically reducible to standard range queries that they already support.

Table 1. Filter predicates used for our evaluation.

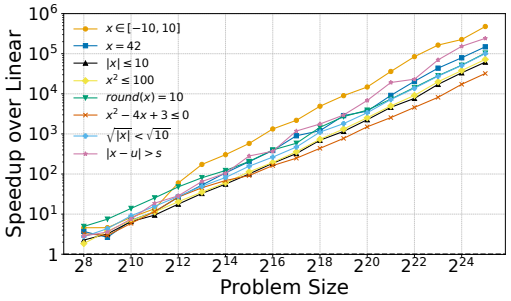
Query Predicate	Postgres	MySQL	DuckDB	SQLite	Bonsai
$x \in [-10, 10]$	**Index**	**Index**	Linear	**Index**	**Index**
$x = 42$	**Index**	**Index**	**Index**	**Index**	**Index**
$\lvert x\rvert \leq 10$	Linear	Linear	Linear	Linear	**Index**
$x^2 \leq 100$	Linear	Linear	Linear	Linear	**Index**
$round(x) = 10$	Linear	Linear	Linear	Linear	**Index**
$x^2 - 4x + 3 \leq 0$	Linear	Linear	Linear	Linear	**Index**
$\sqrt{\lvert x\rvert} < \sqrt{10}$	Linear	Linear	Linear	Linear	**Index**
$\lvert x - u\rvert > s$	Linear	Linear	Linear	Linear	**Index**
$\lvert x - y\rvert < 1$	Linear	Linear	Linear	Linear	**Index**
$\lvert x\rvert + \lvert y\rvert < 1$	Linear	Linear	Linear	Linear	**Index**
$x^2 + y^2 < 10$	Linear	Linear	Linear	Linear	**Index**



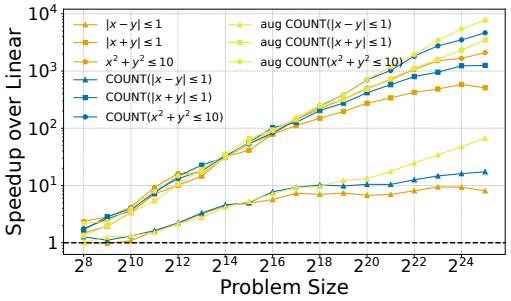
(a) Accelerated filters on interval trees.



(b) Accelerated aggregation (count) of filters.



(c) Accelerated aggregation with count aggregation.



(d) Multidimensional filter predicates.

Fig. 9. Speed-up plots over default linear scans for uniformly sampled data in $[-1000, 1000]$. Higher is better.

These seven contain predicates on a single variable and highlight the benefits of using tree traversals. Figure 9a illustrates performance gains achieved by computing these filters as tree traversals. However, many queries plateau when they become write-bound, except for the highly selective point and standard deviation queries. Figure 9b extends these filters with a `count` aggregation. These traversals are now read-bound, and many simply perform scans to aggregate the count. We therefore extend the queried tree with a count augmentation in Figure 9c, resulting in massive asymptotic improvements.

Although a sufficiently powerful rewrite system could convert some of these queries into range queries, others, particularly those with multi-variable predicates, cannot. The final three filter predicates in Table 1 fall into this category: a diagonal band, a diamond-shaped region, and a circular filter. Figure 9d plots speedups for filters (orange), a fused count-filter reduction (blue), and the same reduction on a tree augmented with subtree count metadata (yellow). All outperform linear scans; more selective queries (diamond and circle) achieve the largest speedups.

9.4.2 Fusion Ablation. Fusion is particularly beneficial when filters are not very selective, as fusion of a reduction on a filter *both* removes the need for an expensive intermediate data structure, and leverages reduction metadata available in the tree. To highlight these benefits, Figure 10 compares fused `count(filter())` queries to their unfused counterparts. Unfused variants perform a tree traversal to compute the filtered set, and then return its size. Because the reduction is an $O(1)$ operation on the

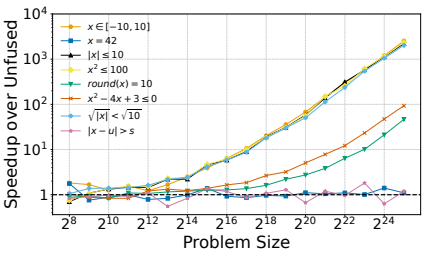
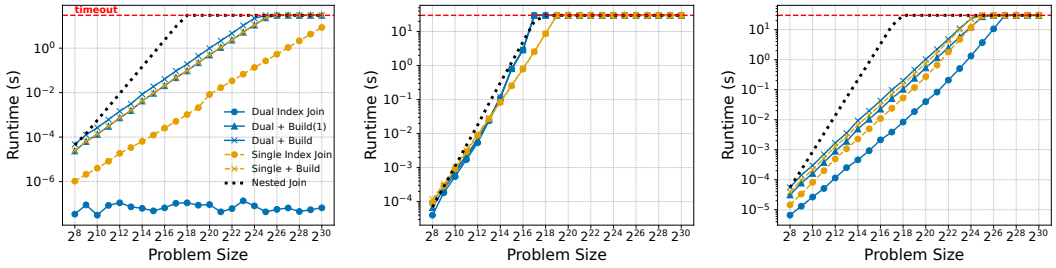


Fig. 10. Ablation of count-filter fusion.



(a) Radius 1. (b) Radius 100. (c) Radius 100000.

Fig. 11. Torus join results, with join predicate $\sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2} \in [10, 20]$. Both tables are uniformly randomly sampled within circles of varying radii, to illustrate the impact of data distribution on join choice.

intermediate data structure, any slowdown is a direct result of either (a) allocation and deallocation of the intermediate or (b) scanning subtrees that could otherwise return the count augmentation. In line with this observation, Figure 10 shows that fusion yields the greatest speedups for less selective queries, which benefit from reduced reading and writing overhead.

9.4.3 Joins. Beyond range and salary joins (Figures 8a and 8b), we evaluate a *torus join*, a hybrid of range and distance joins that retrieves all pairs of points within a distance range. No database we tested, including spatial systems, accelerated this join. Figure 11 shows its performance under varying data distributions. Despite substantial differences in absolute performance, both single and dual-index joins scale better than the nested join, even when accounting for tree build time.

In Figure 11a, points are sampled from a tight circle; *no* pairs satisfy the join predicate. This highlights the strength of dual joins: the algorithm terminates almost immediately after detecting that the trees are disjoint w.r.t the predicate. When tree-build costs are included, the single-index join outperforms the dual join, as the latter must build both trees. Thus, for unindexed tables, a single-index join is preferable, but dual joins are asymptotically superior when both inputs are indexed. In Figure 11b, sampling from a larger circle yields little pruning for either join type. Within the measured range (before timeout), the single-index join (with or without build times) appears to scale better than the dual-index join. Figure 11c considers an even larger circle. Pruning improves over Figure 11b but is less extreme than in Figure 11a, as many distant subtrees are pruned.

Overall, Bonsai-generated joins can outperform nested joins. The best join strategy depends on the predicate and data distribution; we leave cost models to automate this choice for future work.

10 Related Work

Significant research has been conducted on the design of tree data structures. For graphics, we refer the reader to Ericson [21] and Meister et al. [48] for in-depth surveys of spatial acceleration structures in collision detection and ray tracing, respectively. For databases, we refer the reader to the spatio-temporal access methods surveys [45, 50, 54] and indexing survey [24].

Generalized Search Trees (GiSTs) [31] attempt to unify database index structures under a common abstraction, but are not a compilation technique: they still require users to write the pruning function (termed "Consistent" in their model). GiSTs also do not support a notion of `always` functions (for scanning entire subtrees), nor subtree aggregates. We believe our model could serve as a useful extension of GiSTs to allow them to function for a larger variety of queries.

Term Rewriting Systems. Rewriting systems aim to canonicalize queries into a small number of queries for which pruning is possible, e.g., a range query. Rewriting systems are challenged

by local minima and non-termination, motivating both extensive efforts to synthesize them from real-world data [53, 65], and alternatives like e-graphs [51, 83, 85, 87, 88]. Our technique simplifies the task of targeting tree traversals by greatly expanding the space of valid targets: rather than rewriting solely to a limited set of query operators, a system needs only to produce predicates with derivable necessary or sufficient conditions. While no rewriting was needed for our benchmarks, such systems remain valuable in practice for simplifying predicates before bounds analysis (e.g., by removing correlated terms that yield suboptimal interval bounds).

Query Compilation. Query compilers [52, 73, 79] largely treat index queries and tree traversals as external black-box operators, relying on pattern matching to invoke hand-optimized implementations when a cost model deems them profitable. The specific filters or database join algorithms (e.g., hash, sort-merge, and nested-loop joins [25] and their spatial or multidimensional variants [69]) generally depend on specialized traversal code for specific predicates or data types. These techniques accelerate common queries but do not generalize to arbitrary predicates or reductions. Our work complements these systems by automatically generating predicate-aware tree traversals, enabling efficient filters and joins without requiring hand-written specialization.

Interval Analysis in Computer Graphics. While our algorithm for generating pruning functions was inspired by Halide's symbolic interval analysis [61, 62], there has long been use of symbolic and numeric interval analysis in computer graphics [39, 49, 75, 80, 81]. Interval analysis enables hierarchical reasoning about whether equations have solutions over an interval, allowing optimizers to prune solution spaces [75]. We extend this idea to compile tree queries and to handle spatial operators.

Pruning in Databases. Similarly, the databases community has long used pruning to reject partitions of data that cannot satisfy a query predicate [26, 76, 77]. This idea resembles our generation of the `maybe` pruning function, though it is typically used at runtime, not compile time. Recently, Zimmerer et al. [90] proposed computing an equivalent to the `always` function via the relationship $\texttt{always}(P) = \neg\texttt{maybe}(\neg P)$, but is limited to accelerating `LIMIT` queries, though it could extend to a wider range of query types, especially when partition metadata can be used to produce aggregates on scans without iterating over the data. Our formalization of pruning functions could strengthen such systems, while our spatial extension could support efficient pruning of spatial predicates, and our reduction-handling techniques could improve filtered reductions over partitioned data.

Compilers for Irregular Traversal Patterns. Multiple domain-specific languages for other domains (and classes of data structures) have used data structure properties in the compilation of asymptotically-efficient code, e.g., sparse tensor algebra [11, 12, 41] with extensions to more general sparse array processing [32, 44, 67, 78], sparse grids [35] with extensions to meshes [86], and graphs [89]. We provide a similar framework for tree data structures and pruning traversals. Though the graphics community has explored languages for collision detection [8] and ray tracing [57, 58], each relies on hand-written traversal routines. In contrast, our higher-level representation (Section 3) broadens support for more general spatial queries.

Parallelizing Traversals. Significant work has explored parallelizing recursive programs and tree traversals outside of domain-specific languages [10, 23, 42, 63, 68, 72, 74]. These works are largely orthogonal to ours, as they seek to efficiently parallelize code patterns such as those generated by our compiler.

11 Conclusion and Future Work

We describe the first technique for compiling high-level queries into pruning tree traversals. Our technique is enabled by the key insight that hand-written tree traversals are based on necessary and sufficient conditions as functions of node metadata, which imply or disprove query predicates without iterating all data in the subtree. This technique is further enabled by an efficient derivation

procedure for generating these conditions based on standard symbolic interval analysis with a novel extension to spatial operators. Our results demonstrate that accelerated filters and joins do not need to rely on bespoke data-structure-specific code, but can instead be generated by a compiler. This suggests a path toward query engines that treat tree-based acceleration as a reusable, derivable optimization rather than a small set of inflexible primitives. We envision many directions for future work, including:

DBMS Integration. Integrating our techniques into a database management system (DBMS) requires both implementing physical relational operators for tree queries (e.g., a generalized Index Scan) and developing cost models to drive operator selection. Such cost models must consider both query predicates and data distributions to assess pruning potential. For example, a DBMS observing frequent long-running queries that filter or aggregate on a column could automatically build bounds or reduction annotations for that column, respectively. However, this may only be profitable given certain data distributions on the column, as shown in Figure 11. Determining when a profile warrants index construction requires accurate modeling of compute gains against the memory overhead of storing the index, which we believe requires further research on cost models.

Tree Design. Different choices of stored metadata induce different asymptotics of tree traversals, but also impact memory usage; synthesizing a tree design (i.e., the choice of metadata) automatically for a query or set of queries would enable automatic optimization of queries.

Space Partitioning Trees. Space partitioning trees such as k-d trees [7] (as opposed to bounding volume hierarchies) offer weaker invariants for non-point data: their implicit bounding volumes only promise overlap with primitives beneath a node, and geometry often needs to be duplicated in the tree [19]. Nevertheless, they offer benefits for some classes of queries (e.g., in-order all-hit ray tracing). Developing predicate analysis for overlap volumes and deriving deduplication techniques could enable further query acceleration.

Scheduling. The ray tracing community has performed extensive research on improving the performance of tree traversals on parallel hardware (e.g., packet tracing [82], techniques for improving coherence on GPUs [2], and wavefront traversals [1]). Supporting these optimizations for a larger class of queries via program transformations could be useful.

Construction Algorithms. It is well-known that the construction of a tree on geometry has a significant impact on query performance [3, 5, 27, 37, 43, 47, 56]. Exploring the design space of construction time versus tree quality is important work that could be made easier by compiler techniques.

Together, these directions broaden the scope of pruning-based optimization and move us toward compiler-driven methods for accelerated query processing. We hope this work serves as a foundation for research in tree traversal design and derivation.

Data-Availability Statement

Performance results were generated with a publicly available artifact [66] containing all benchmarking code and scripts, as well as instructions for reproducibility. The Bonsai compiler is also available at https://github.com/rootjalex/bonsai. Benchmarking results may very based on the hardware used.

Acknowledgments

We thank our reviewers for their valuable feedback. We also thank Amanda Liu, Bala Vinaithirthan, Ben Driscoll, Bobby Yan, Brennan Shacklett, Devanshu Ladsaria, Genghan Zhang, Ishita Gupta, James Dong, Katherine Mohr, Liza Pertseva, Marco Siracusa, Matt Pharr, Nestan Tsiskaridze, Olivia

Hsu, Rohan Sawhney, Rohan Yadav, Rubens Lacouture, Sai Gautham Ravipati, Scott Kovach, and Zander Majercik for their helpful feedback on drafts of this work. Alexander and Christophe were supported by the Qualcomm Innovation Fellowship during part of this work; Alexander was also supported by the NSF Graduate Research Fellowship and took part in this work while interning at Adobe Research under Andrew.

References

[1] Timo Aila and Tero Karras. 2010. Architecture considerations for tracing incoherent rays. In *Proceedings of the Conference on High Performance Graphics* (Saarbrucken, Germany) *(HPG '10)*. Eurographics Association, Goslar, DEU, 113–122.

[2] Timo Aila and Samuli Laine. 2009. Understanding the efficiency of ray traversal on GPUs. In *Proceedings of the Conference on High Performance Graphics 2009* (New Orleans, Louisiana) *(HPG '09)*. Association for Computing Machinery, New York, NY, USA, 145–149. doi:10.1145/1572769.1572792

[3] Ciprian Apetrei. 2014. Fast and Simple Agglomerative LBVH Construction. In *Computer Graphics and Visual Computing (CGVC)*, Rita Borgo and Wen Tang (Eds.). The Eurographics Association, 41–44. doi:10.2312/cgvc.20141206

[4] Josh Barnes and Piet Hut. 1986. A hierarchical O(N log N) force-calculation algorithm. *Nature* 324, 6096 (1986), 446–449. doi:10.1038/324446a0

[5] Carsten Benthin, Daniel Meister, Joshua Barczak, Rohan Mehalwal, John Tsakok, and Andrew Kensler. 2024. H-PLOC: Hierarchical Parallel Locally-Ordered Clustering for Bounding Volume Hierarchy Construction. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 7, 3 (2024), 1–14.

[6] Carsten Benthin, Ingo Wald, Sven Woop, and Attila T. Áfra. 2018. Compressed-leaf bounding volume hierarchies. In *Proceedings of the Conference on High-Performance Graphics* (Vancouver, British Columbia, Canada) *(HPG '18)*. Association for Computing Machinery, New York, NY, USA, Article 6, 4 pages. doi:10.1145/3231578.3231581

[7] Jon Louis Bentley. 1975. Multidimensional binary search trees used for associative searching. *Commun. ACM* 18, 9 (Sept. 1975), 509–517. doi:10.1145/361002.361007

[8] Gilbert Louis Bernstein. 2019. *Designing Languages for Parallel Portability of Physical Simulations, Using Relational Algebraic Abstractions.* Ph.D. Dissertation. Stanford University.

[9] Robert D. Blumofe and Charles E. Leiserson. 1999. Scheduling multithreaded computations by work stealing. *J. ACM* 46, 5 (Sept. 1999), 720–748. doi:10.1145/324133.324234

[10] Yanju Chen, Junrui Liu, Yu Feng, and Rastislav Bodik. 2022. Tree traversal synthesis using domain-specific symbolic compilation. In *Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems* (Lausanne, Switzerland) *(ASPLOS '22)*. Association for Computing Machinery, New York, NY, USA, 1030–1042. doi:10.1145/3503222.3507751

[11] Stephen Chou and Saman Amarasinghe. 2022. Compilation of Dynamic Sparse Tensor Algebra. *Proc. ACM Program. Lang.* 6, OOPSLA2, Article 175 (Oct. 2022), 30 pages. doi:10.1145/3563338

[12] Stephen Chou, Fredrik Kjolstad, and Saman Amarasinghe. 2018. Format Abstraction for Sparse Tensor Algebra Compilers. *Proc. ACM Program. Lang.* 2, OOPSLA, Article 123 (Oct. 2018), 30 pages. doi:10.1145/3276493

[13] E. F. Codd. 1970. A relational model of data for large shared data banks. *Commun. ACM* 13, 6 (June 1970), 377–387. doi:10.1145/362384.362685

[14] Douglas Comer. 1979. Ubiquitous B-Tree. *ACM Comput. Surv.* 11, 2 (June 1979), 121–137. doi:10.1145/356770.356776

[15] Duncan Coutts, Roman Leshchinskiy, and Don Stewart. 2007. Stream fusion: from lists to streams to nothing at all. In *Proceedings of the 12th ACM SIGPLAN International Conference on Functional Programming* (Freiburg, Germany) *(ICFP '07)*. Association for Computing Machinery, New York, NY, USA, 315–326. doi:10.1145/1291151.1291199

[16] C. J. Date. 1989. *A guide to the SQL standard (2nd ed.).* Addison-Wesley Longman Publishing Co., Inc., USA.

[17] Isil Dillig, Thomas Dillig, Boyang Li, and Ken McMillan. 2013. Inductive invariant generation via abductive inference. In *Proceedings of the 2013 ACM SIGPLAN International Conference on Object Oriented Programming Systems Languages & Applications* (Indianapolis, Indiana, USA) *(OOPSLA '13)*. Association for Computing Machinery, New York, NY, USA, 443–456. doi:10.1145/2509136.2509511

[18] Max J. Egenhofer and John Herring. 1990. A Mathematical Framework for the Definition of Topological Relationships. In *Proceedings of the Fourth International Symposium on Spatial Data Handling*. International Geographical Union, Zurich, Switzerland, 803–813. https://web.archive.org/web/20100614161335/http://www.spatial.maine.edu/~max/MJEJRH-SDH1990.pdf Accessed: 2025-05-01.

[19] M. Y. Eltabakh, Walid G. Aref, and Mourad Ouzzani. 2007. Duplicate Elimination in Space-partitioning Tree Indexes . In *Scientific and Statistical Database Management, International Conference on*. IEEE Computer Society, Los Alamitos, CA, USA, 18. doi:10.1109/SSDBM.2007.10

[20] U. Emre, A. Kanak, and S. Steinberg. 2025. High-Performance Elliptical Cone Tracing. *Computer Graphics Forum* 44, 7 (2025), e70230. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.70230 doi:10.1111/cgf.70230
[21] Christer Ericson. 2004. *Real-Time Collision Detection.* CRC Press, Inc., USA.
[22] Peng Fan, Wei Wang, Ruofeng Tong, Hailong Li, and Min Tang. 2024. gDist: Efficient Distance Computation between 3D Meshes on GPU. In *SIGGRAPH Asia 2024 Conference Papers* (Tokyo, Japan) *(SA '24)*. Association for Computing Machinery, New York, NY, USA, Article 71, 11 pages. doi:10.1145/3680528.3687619
[23] Matteo Frigo, Charles E. Leiserson, and Keith H. Randall. 1998. The implementation of the Cilk-5 multithreaded language. In *Proceedings of the ACM SIGPLAN 1998 Conference on Programming Language Design and Implementation* (Montreal, Quebec, Canada) *(PLDI '98)*. Association for Computing Machinery, New York, NY, USA, 212–223. doi:10.1145/277650.277725
[24] Abdullah Gani, Aisha Siddiqa, Shahaboddin Shamshirband, and Fariza Hanum. 2016. A survey on indexing techniques for big data: taxonomy and performance evaluation. *Knowledge and Information Systems* 46, 2 (2016), 241–284. doi:10.1007/s10115-015-0830-y
[25] Goetz Graefe. 1993. Query evaluation techniques for large databases. *ACM Comput. Surv.* 25, 2 (June 1993), 73–169. doi:10.1145/152610.152611
[26] Goetz Graefe. 2009. Fast Loads and Fast Queries. In *Data Warehousing and Knowledge Discovery*, Torben Bach Pedersen, Mukesh K. Mohania, and A. Min Tjoa (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 111–124.
[27] Yan Gu, Yong He, Kayvon Fatahalian, and Guy Blelloch. 2013. Efficient BVH construction via approximate agglomerative clustering. In *Proceedings of the 5th High-Performance Graphics Conference* (Anaheim, California) *(HPG '13)*. Association for Computing Machinery, New York, NY, USA, 81–88. doi:10.1145/2492045.2492054
[28] Gaël Guennebaud, Benoît Jacob, et al. 2010. Eigen v3. http://eigen.tuxfamily.org.
[29] Antonin Guttman. 1984. R-trees: a dynamic index structure for spatial searching. In *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data* (Boston, Massachusetts) *(SIGMOD '84)*. Association for Computing Machinery, New York, NY, USA, 47–57. doi:10.1145/602259.602266
[30] Christophe Gyurgyik, Alexander J Root, and Fredrik Kjolstad. 2026. Decoupling Data Layouts from Bounding Volume Hierarchies. *Proceedings of the ACM on Programming Languages* 10, PLDI, Article 175 (June 2026). doi:10.1145/3808253
[31] Joseph M. Hellerstein, Jeffrey F. Naughton, and Avi Pfeffer. 1998. *Generalized search trees for database systems.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 101–112.
[32] Rawn Henry, Olivia Hsu, Rohan Yadav, Stephen Chou, Kunle Olukotun, Saman Amarasinghe, and Fredrik Kjolstad. 2021. Compilation of Sparse Array Programming Models. *Proc. ACM Program. Lang.* 5, OOPSLA, Article 128 (Oct. 2021), 29 pages. doi:10.1145/3485505
[33] Richard D. Hipp and The SQLite Development Team. 2025. SQLite. Available at https://www.sqlite.org/. Accessed: October 19, 2025.
[34] Michael P Howard, Antonia Statt, Felix Madutsa, Thomas M Truskett, and Athanassios Z Panagiotopoulos. 2019. Quantized bounding volume hierarchies for neighbor search in molecular simulations on graphics processing units. *Computational Materials Science* 164 (2019), 139–146.
[35] Yuanming Hu, Tzu-Mao Li, Luke Anderson, Jonathan Ragan-Kelley, and Frédo Durand. 2019. Taichi: a language for high-performance computation on spatially sparse data structures. *ACM Trans. Graph.* 38, 6, Article 201 (Nov. 2019), 16 pages. doi:10.1145/3355089.3356506
[36] Mioara Joldes, Jean-Michel Muller, and Valentina Popescu. 2017. Tight and Rigorous Error Bounds for Basic Building Blocks of Double-Word Arithmetic. *ACM Trans. Math. Softw.* 44, 2, Article 15res (oct 2017), 27 pages. doi:10.1145/3121432
[37] Tero Karras. 2012. Maximizing parallelism in the construction of BVHs, octrees, and k-d trees. In *Proceedings of the Fourth ACM SIGGRAPH / Eurographics Conference on High-Performance Graphics* (Paris, France) *(EGGH-HPG'12)*. Eurographics Association, Goslar, DEU, 33–37.
[38] Timothy L. Kay and James T. Kajiya. 1986. Ray tracing complex scenes. *SIGGRAPH Comput. Graph.* 20, 4 (Aug. 1986), 269–278. doi:10.1145/15886.15916
[39] Matthew J. Keeter. 2020. Massively parallel rendering of complex closed-form implicit surfaces. *ACM Trans. Graph.* 39, 4, Article 141 (Aug. 2020), 10 pages. doi:10.1145/3386569.3392429
[40] Zuhair Khayyat, William Lucia, Meghna Singh, Mourad Ouzzani, Paolo Papotti, Jorge-Arnulfo Quiané-Ruiz, Nan Tang, and Panos Kalnis. 2015. Lightning fast and space efficient inequality joins. *Proc. VLDB Endow.* 8, 13 (Sept. 2015), 2074–2085. doi:10.14778/2831360.2831362
[41] Fredrik Kjolstad, Shoaib Kamil, Stephen Chou, David Lugato, and Saman Amarasinghe. 2017. The Tensor Algebra Compiler. *Proc. ACM Program. Lang.* 1, OOPSLA, Article 77 (Oct. 2017), 29 pages. doi:10.1145/3133901
[42] Chaitanya Koparkar, Mike Rainey, Michael Vollmer, Milind Kulkarni, and Ryan R. Newton. 2021. Efficient tree-traversals: reconciling parallelism and dense data representations. *Proc. ACM Program. Lang.* 5, ICFP, Article 91 (Aug. 2021), 29 pages. doi:10.1145/3473596

[43] C. Lauterbach, M. Garland, S. Sengupta, D. Luebke, and D. Manocha. 2009. Fast BVH Construction on GPUs. *Computer Graphics Forum* 28, 2 (2009), 375–384. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2009.01377.x doi:10.1111/j.1467-8659.2009.01377.x

[44] Peiming Liu, Alexander J Root, Anlun Xu, Yinying Li, Fredrik Kjolstad, and Aart J.C. Bik. 2024. Compiler Support for Sparse Tensor Convolutions. *Proc. ACM Program. Lang.* 8, OOPSLA2, Article 281 (Oct. 2024), 29 pages. doi:10.1145/3689721

[45] Ahmed R. Mahmood, Sri Punni, and Walid G. Aref. 2019. Spatio-temporal access methods: a survey (2010 - 2017). *Geoinformatica* 23, 1 (Jan. 2019), 1–36. doi:10.1007/s10707-018-0329-2

[46] Morgan McGuire. 2017. Computer Graphics Archive. https://casual-effects.com/data `https://casual-effects.com/data`.

[47] Daniel Meister and Jiří Bittner. 2017. Parallel locally-ordered clustering for bounding volume hierarchy construction. *IEEE transactions on visualization and computer graphics* 24, 3 (2017), 1345–1353.

[48] Daniel Meister, Shinji Ogaki, Carsten Benthin, Michael J. Doyle, Michael Guthe, and Jiří Bittner. 2021. A Survey on Bounding Volume Hierarchies for Ray Tracing. *Computer Graphics Forum* 40, 2 (2021), 683–712. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.142662 doi:10.1111/cgf.142662

[49] Don P. Mitchell. 1991. Three Applications of Interval Analysis in Computer Graphics. In *Frontiers of Rendering, Course Notes*. ACM SIGGRAPH. Course No. 14, SIGGRAPH '91.

[50] Mohamed F. Mokbel, Thanaa M. Ghanem, and Walid G. Aref. 2003. Spatio-temporal Access Methods. *IEEE Data Engineering Bulletin* 26, 2 (2003), 40–49.

[51] Chandrakana Nandi, Max Willsey, Amy Zhu, Yisu Remy Wang, Brett Saiki, Adam Anderson, Adriana Schulz, Dan Grossman, and Zachary Tatlock. 2021. Rewrite rule inference using equality saturation. *Proc. ACM Program. Lang.* 5, OOPSLA, Article 119 (Oct. 2021), 28 pages. doi:10.1145/3485496

[52] Thomas Neumann. 2011. Efficiently compiling efficient query plans for modern hardware. *Proc. VLDB Endow.* 4, 9 (June 2011), 539–550. doi:10.14778/2002938.2002940

[53] Julie L. Newcomb, Andrew Adams, Steven Johnson, Rastislav Bodik, and Shoaib Kamil. 2020. Verifying and improving Halide's term rewriting system with program synthesis. *Proc. ACM Program. Lang.* 4, OOPSLA, Article 166 (Nov. 2020), 28 pages. doi:10.1145/3428234

[54] Long-Van Nguyen-Dinh, Walid G. Aref, and Mohamed F. Mokbel. 2010. Spatio-Temporal Access Methods: Part 2 (2003 - 2010). *IEEE Data Eng. Bull.* 33, 2 (2010), 46–55. http://sites.computer.org/debull/A10june/Aref.pdf

[55] Jia Pan, Sachin Chitta, and Dinesh Manocha. 2012. FCL: A general purpose library for collision and proximity queries. In *2012 IEEE International Conference on Robotics and Automation*. 3859–3866. doi:10.1109/ICRA.2012.6225337

[56] J. Pantaleoni and D. Luebke. 2010. HLBVH: hierarchical LBVH construction for real-time ray tracing of dynamic geometry. In *Proceedings of the Conference on High Performance Graphics* (Saarbrucken, Germany) *(HPG '10)*. Eurographics Association, Goslar, DEU, 87–95.

[57] Arsène Pérard-Gayot, Richard Membarth, Roland Leißa, Sebastian Hack, and Philipp Slusallek. 2019. Rodent: Generating Renderers without Writing a Generator. *ACM Trans. Graph.* 38, 4, Article 40 (jul 2019), 12 pages. doi:10.1145/3306346.3322955

[58] Arsène Pérard-Gayot, Martin Weier, Richard Membarth, Philipp Slusallek, Roland Leißa, and Sebastian Hack. 2017. RaTrace: Simple and Efficient Abstractions for BVH Ray Traversal Algorithms. In *Proceedings of the 16th ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences* (Vancouver, BC, Canada) *(GPCE 2017)*. Association for Computing Machinery, New York, NY, USA, 157–168. doi:10.1145/3136040.3136044

[59] Matt Pharr, Wenzel Jakob, and Greg Humphreys. 2023. *Physically Based Rendering: From Theory to Implementation* (4 ed.). MIT Press. https://mitpress.mit.edu/9780262048026/physically-based-rendering/

[60] Mark Raasveldt and Hannes Mühleisen. 2019. DuckDB: an Embeddable Analytical Database. In *Proceedings of the 2019 International Conference on Management of Data* (Amsterdam, Netherlands) *(SIGMOD '19)*. Association for Computing Machinery, New York, NY, USA, 1981–1984. doi:10.1145/3299869.3320212

[61] Jonathan Ragan-Kelley, Andrew Adams, Sylvain Paris, Marc Levoy, Saman Amarasinghe, and Frédo Durand. 2012. Decoupling algorithms from schedules for easy optimization of image processing pipelines. *ACM Trans. Graph.* 31, 4, Article 32 (jul 2012), 12 pages. doi:10.1145/2185520.2185528

[62] Jonathan Ragan-Kelley, Connelly Barnes, Andrew Adams, Sylvain Paris, Frédo Durand, and Saman Amarasinghe. 2013. Halide: a language and compiler for optimizing parallelism, locality, and recomputation in image processing pipelines. *SIGPLAN Not.* 48, 6 (June 2013), 519–530. doi:10.1145/2499370.2462176

[63] Bin Ren, Shruthi Balakrishna, Youngjoon Jo, Sriram Krishnamoorthy, Kunal Agrawal, and Milind Kulkarni. 2019. Extracting SIMD Parallelism from Recursive Task-Parallel Programs. *ACM Trans. Parallel Comput.* 6, 4, Article 24 (Dec. 2019), 37 pages. doi:10.1145/3365663

[64] Andrew Reynolds, Haniel Barbosa, Daniel Larraz, and Cesare Tinelli. 2020. Scalable Algorithms for Abduction via Enumerative Syntax-Guided Synthesis. In *Automated Reasoning: 10th International Joint Conference, IJCAR 2020, Paris, France, July 1–4, 2020, Proceedings, Part I* (Paris, France). Springer-Verlag, Berlin, Heidelberg, 141–160. doi:10.1007/978-3-030-51074-9_9

[65] Alexander J Root, Maaz Bin Safeer Ahmad, Dillon Sharlet, Andrew Adams, Shoaib Kamil, and Jonathan Ragan-Kelley. 2024. Fast Instruction Selection for Fast Digital Signal Processing. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 4* (Vancouver, BC, Canada) *(ASPLOS '23)*. Association for Computing Machinery, New York, NY, USA, 125–137. doi:10.1145/3623278.3624768

[66] Alexander J Root, Gyurgyik Christophe, Goel Purvi, Kayvon Fatahalian, Jonathan Ragan-Kelley, Adams Andrew, and Fredrik Berg Kjolstad. 2026. *Artifact for PLDI 2026 Paper: Bonsai: Compiling Queries into Work-Efficient Tree Traversals*. doi:10.5281/zenodo.19091467

[67] Alexander J Root, Bobby Yan, Peiming Liu, Christophe Gyurgyik, Aart J.C. Bik, and Fredrik Kjolstad. 2024. Compilation of Shape Operators on Sparse Arrays. *Proc. ACM Program. Lang.* 8, OOPSLA2, Article 312 (Oct. 2024), 27 pages. doi:10.1145/3689752

[68] Laith Sakka, Kirshanthan Sundararajah, Ryan R. Newton, and Milind Kulkarni. 2019. Sound, fine-grained traversal fusion for heterogeneous trees. In *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation* (Phoenix, AZ, USA) *(PLDI 2019)*. Association for Computing Machinery, New York, NY, USA, 830–844. doi:10.1145/3314221.3314626

[69] Hanan Samet. 2005. *Foundations of Multidimensional and Metric Data Structures (The Morgan Kaufmann Series in Computer Graphics and Geometric Modeling)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[70] Rohan Sawhney. 2021. *FCPW: Fastest Closest Points in the West*.

[71] Rohan Sawhney and Keenan Crane. 2020. Monte Carlo geometry processing: a grid-free approach to PDE-based methods on volumetric domains. *ACM Trans. Graph.* 39, 4, Article 123 (Aug. 2020), 18 pages. doi:10.1145/3386569.3392374

[72] Tao B. Schardl, William S. Moses, and Charles E. Leiserson. 2019. Tapir: Embedding Recursive Fork-join Parallelism into LLVM's Intermediate Representation. *ACM Trans. Parallel Comput.* 6, 4, Article 19 (Dec. 2019), 33 pages. doi:10.1145/3365655

[73] P. Griffiths Selinger, M. M. Astrahan, D. D. Chamberlin, R. A. Lorie, and T. G. Price. 1979. Access path selection in a relational database management system. In *Proceedings of the 1979 ACM SIGMOD International Conference on Management of Data* (Boston, Massachusetts) *(SIGMOD '79)*. Association for Computing Machinery, New York, NY, USA, 23–34. doi:10.1145/582095.582099

[74] Vidush Singhal, Laith Sakka, Kirshanthan Sundararajah, Ryan Newton, and Milind Kulkarni. 2024. Orchard: Heterogeneous Parallelism and Fine-grained Fusion for Complex Tree Traversals. *ACM Trans. Archit. Code Optim.* 21, 2, Article 41 (May 2024), 25 pages. doi:10.1145/3652605

[75] John M. Snyder. 1992. Interval analysis for computer graphics. *SIGGRAPH Comput. Graph.* 26, 2 (July 1992), 121–130. doi:10.1145/142920.134024

[76] Sivaprasad Sudhir, Wenbo Tao, Nikolay Laptev, Cyrille Habis, Michael Cafarella, and Samuel Madden. 2023. Pando: Enhanced Data Skipping with Logical Data Partitioning. *Proc. VLDB Endow.* 16, 9 (May 2023), 2316–2329. doi:10.14778/3598581.3598601

[77] Liwen Sun, Michael J. Franklin, Sanjay Krishnan, and Reynold S. Xin. 2014. Fine-grained partitioning for aggressive data skipping. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data* (Snowbird, Utah, USA) *(SIGMOD '14)*. Association for Computing Machinery, New York, NY, USA, 1115–1126. doi:10.1145/2588555.2610515

[78] Shiv Sundram, Muhammad Usman Tariq, and Fredrik Kjolstad. 2024. Compiling Recurrences over Dense and Sparse Arrays. *Proc. ACM Program. Lang.* 8, OOPSLA1, Article 103 (April 2024), 26 pages. doi:10.1145/3649820

[79] Ruby Y. Tahboub, Grégory M. Essertel, and Tiark Rompf. 2018. How to Architect a Query Compiler, Revisited. In *Proceedings of the 2018 International Conference on Management of Data* (Houston, TX, USA) *(SIGMOD '18)*. Association for Computing Machinery, New York, NY, USA, 307–322. doi:10.1145/3183713.3196893

[80] Thibault Tricard. 2024. Interval Shading: using Mesh Shaders to generate shading intervals for volume rendering. *Proc. ACM Comput. Graph. Interact. Tech.* 7, 3, Article 43 (Aug. 2024), 11 pages. doi:10.1145/3675380

[81] Edgar Velázquez-Armendáriz, Shuang Zhao, Miloš Hašan, Bruce Walter, and Kavita Bala. 2009. Automatic bounding of programmable shaders for efficient global illumination. *ACM Trans. Graph.* 28, 5 (Dec. 2009), 1–9. doi:10.1145/1618452.1618488

[82] Ingo Wald, Philipp Slusallek, Carsten Benthin, and Markus Wagner. 2001. Interactive Rendering with Coherent Ray Tracing. *Computer Graphics Forum* 20, 3 (2001), 153–165. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8659.00508 doi:10.1111/1467-8659.00508

[83] 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

[84] Sven Woop, Carsten Benthin, Ingo Wald, Gregory S. Johnson, and Eric Tabellion. 2014. Exploiting local orientation similarity for efficient ray traversal of hair and fur. In *Proceedings of High Performance Graphics* (Lyon, France) *(HPG '14)*. Eurographics Association, Goslar, DEU, 41–49.

[85] Yichen Yang, Phitchaya Phothilimthana, Yisu Wang, Max Willsey, Sudip Roy, and Jacques Pienaar. 2021. Equality Saturation for Tensor Graph Superoptimization. In *Proceedings of Machine Learning and Systems*, A. Smola, A. Dimakis, and I. Stoica (Eds.), Vol. 3. 255–268. https://proceedings.mlsys.org/paper_files/paper/2021/file/cc427d934a7f6c0663e5923f49eba531-Paper.pdf

[86] Chang Yu, Yi Xu, Ye Kuang, Yuanming Hu, and Tiantian Liu. 2022. MeshTaichi: A Compiler for Efficient Mesh-Based Operations. *ACM Trans. Graph.* 41, 6, Article 252 (Nov. 2022), 17 pages. doi:10.1145/3550454.3555430

[87] Yihong Zhang, Dan Suciu, Yisu Remy Wang, and Max Willsey. 2025. Database Theory in Action: Search-Based Program Optimization. In *28th International Conference on Database Theory (ICDT 2025) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 328)*, Sudeepa Roy and Ahmet Kara (Eds.). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 34:1–34:6. doi:10.4230/LIPIcs.ICDT.2025.34

[88] Yihong Zhang, Yisu Remy Wang, Max Willsey, and Zachary Tatlock. 2022. Relational E-Matching. *Proc. ACM Program. Lang.* 6, POPL, Article 35 (jan 2022), 22 pages. doi:10.1145/3498696

[89] Yunming Zhang, Mengjiao Yang, Riyadh Baghdadi, Shoaib Kamil, Julian Shun, and Saman Amarasinghe. 2018. GraphIt: a high-performance graph DSL. *Proc. ACM Program. Lang.* 2, OOPSLA, Article 121 (Oct. 2018), 30 pages. doi:10.1145/3276491

[90] Andreas Zimmerer, Damien Dam, Jan Kossmann, Juliane Waack, Ismail Oukid, and Andreas Kipf. 2025. Pruning in Snowflake: Working Smarter, Not Harder. In *Companion of the 2025 International Conference on Management of Data* (Berlin, Germany) *(SIGMOD/PODS '25)*. Association for Computing Machinery, New York, NY, USA, 757–770. doi:10.1145/3722212.3724447

A Scalar Symbolic Interval Analysis

Interval analysis is a recursive bottom-up technique. For symbolic interval analysis, the compiler builds an AST that represents the lower bound and upper bound of a symbolic expression, where varying parameters are replaced with their interval (or volumes) and uniform parameters are singular-valued intervals. To support additional operations, one needs to define the semantics of how an operator operates on an interval instead of a scalar value; this is generally done by reasoning about the monotonicity of an operation. Section 6.2 describes how such reasoning can be done for boolean combinators and comparison operations; here, we document how intervals can be computed numerically, allowing for analyzing predicates containing computation.

Numerical operations. Addition is monotonically increasing in both arguments; thus, the upper bound is the sum of the upper bound of its arguments. Subtraction, however, is monotonically increasing in its first argument, but monotonically decreasing in the second argument. These properties result in the following bounds:

$$\lceil x + y \rceil \mapsto \lceil x \rceil + \lceil y \rceil \qquad \lceil x - y \rceil \mapsto \lceil x \rceil - \lfloor y \rfloor$$

$$\lfloor x + y \rfloor \mapsto \lfloor x \rfloor + \lfloor y \rfloor \qquad \lfloor x - y \rfloor \mapsto \lfloor x \rfloor - \lceil y \rceil$$

Multiplication is non-linear, and therefore requires evaluating all interval end-points[4]:

$$\text{let } S = \{\lceil x \rceil * \lceil y \rceil, \lceil x \rceil * \lfloor y \rfloor, \lfloor x \rfloor * \lceil y \rceil, \lfloor x \rfloor * \lfloor y \rfloor\} \text{ in}$$

$$\lceil x * y \rceil \mapsto \max(S) \qquad \lfloor x * y \rfloor \mapsto \min(S)$$

Thus far, no numerical operations have been type-specific.[5] However, the bounds of division are quite different for floating-point versus integer types. Computing the bounds of a floating-point division is a well-studied but complex problem [36], and we do not attempt to tersely express accurate symbolic bounds for floating-point division here. Integer division and modulo can be reasoned about somewhat more easily, but require a significant amount of control flow to handle reasoning about the sign of each operand. For brevity, we do not include the upper and lower bounds of these operators, but do walk through the (symbolic) casework needed for the upper bound[6] of integer (Euclidean) division below.

$$\lceil x/y \rceil \mapsto \begin{cases} \lceil x \rceil / \lfloor y \rfloor & \lceil x \rceil > 0 \wedge \lfloor y \rfloor > 0 \\ \lceil x \rceil / \lceil y \rceil & \lceil x \rceil < 0 \wedge \lfloor y \rfloor > 0 \\ \lfloor x \rfloor / \lceil y \rceil & \lfloor x \rfloor < 0 \wedge \lceil y \rceil < 0 \\ \lfloor x \rfloor / \lfloor y \rfloor & \lfloor x \rfloor > 0 \wedge \lceil y \rceil < 0 \\ \max(-\lfloor x \rfloor, \lceil x \rceil) & \text{otherwise} \end{cases}$$

In order, these cases correspond to: x can be positive and y must be positive; x must be negative and y must be positive; x can be negative and y must be negative; x must be positive and y must be negative; and lastly, y can be positive or negative. In each of these cases, if the result must be negative, the lowest-magnitude negative result is computed, and if the result can be positive, the highest-magnitude positive result is computed.

Monotonic Functions. The bounds of monotonic functions, such as `min`, `max`, `ceil`, `floor`, `exp`, `sqrt`, or `ln`, are simply the function applied to the corresponding bounds of its argument(s).

[4]Inlining all multiplications blows up the AST; `let` statements avoid this.
[5]A practical concern here is integer overflow, under which these rules are incorrect. This is most easily avoided by defining numerical operations to saturate.
[6]The lower bound does not exist when the interval of the denominator includes zero.

Conditionals. Boolean bounds can also be used to bound conditional statements. We denote the ternary conditional operator (AKA if-then-else) as `ite`, and provide the following expression for the upper bound:

$$\lceil \texttt{ite}(a, x, y) \rceil \mapsto \max(\texttt{ite}(\lfloor a \rfloor, \lceil x \rceil, \lceil y \rceil), \texttt{ite}(\lceil a \rceil, \lceil x \rceil, \lceil y \rceil))$$

Logically, if the condition *a must be true* ($\lfloor a \rfloor = \lceil a \rceil = true$), then the upper bound of `ite` is just the upper bound of x, $\lceil x \rceil$. If a must be false ($\lfloor a \rfloor = \lceil a \rceil = false$), then the upper bound is just $\lceil y \rceil$. Otherwise, a is unbounded, and the upper bound is just the maximum of the two possible upper bounds. The same reasoning is applied to produce the expression for the lower bound:

$$\lfloor \texttt{ite}(a, x, y) \rfloor \mapsto \min(\texttt{ite}(\lfloor a \rfloor, \lfloor x \rfloor, \lfloor y \rfloor), \texttt{ite}(\lceil a \rceil, \lfloor x \rfloor, \lfloor y \rfloor))$$

Non-monotonic Functions. There are many interesting non-monotonic functions as well, which require slightly more complex reasoning to bound their values. Piecewise-monotonic (monotonic on certain intervals) functions can be bounded relatively easily. For example, consider the `trunc` function from the SQL standard [16], which accepts a floating-point value and an integer number representing the number of digits past 0 to round to. If the rounding integer is uniform (constant), this function is monotonic in the floating-point argument, but if not (e.g., the rounding integer is a varying piece of indexed data), then we require more control-flow to be generated:

$$\lceil \texttt{trunc}(x, y) \rceil \mapsto \texttt{ite}(\lceil x \rceil > 0, \texttt{trunc}(\lceil x \rceil, \lceil y \rceil), \texttt{trunc}(\lceil x \rceil, \lfloor y \rfloor))$$

If x can be positive, the upper bound of $\texttt{trunc}(x, y)$ is the upper bound of x with the most precision (with $\lceil y \rceil$ decimals), but if x is strictly negative, then the upper bound is the upper bound of x with the *least* precision (with $\lfloor y \rfloor$ decimals). Similar reasoning can be applied to the lower bound of `trunc`, as well as other functions like `pow` and `round`.

Received 2025-11-07; accepted 2026-04-03