Abstract—Compiler research and development has treated computation as the primary driver of performance improvements in C/C++ programs, leaving memory optimizations as a secondary consideration. Developers are currently handed the arduous task of describing both the semantics and layout of their data in memory, either manually or via libraries, prematurely lowering high-level data collections to a low-level view of memory for the compiler. Thus, the compiler can only glean conservative information about the memory in a program, e.g., alias analysis, and is further hampered by heavy memory optimizations. This paper proposes the Memory Object Intermediate Representation (MEMOIR), a language-agnostic SSA form for sequential and associative data collections, objects, and the fields contained therein. At the core of MEMOIR is a decoupling of the memory used to store data from that used to logically organize data. Through its SSA form, MEMOIR compilers can perform element-level analysis on data collections, enabling static analysis on the state of a collection or object at any given program point. To illustrate the power of this analysis, we perform dead element elimination, resulting in a 26.6% speedup on mcf from SPECINT 2017. With the degree of freedom to mutate memory layout, our MEMOIR compiler performs field elision and dead field elimination, reducing peak memory usage of mcf by 20.8%.

I. INTRODUCTION

Imperative programming languages require developers to describe their programs via direct updates to the program state. Some of these languages, namely C, give developers direct access to memory, making the ceiling for manual memory optimizations nearly unlimited. Using this degree of freedom, developers have been able to build operating systems [1], optimizing compilers [2,3], and interpreters [4,5].

However this manual control comes with the caveat that all memory optimizations must be created manually. This spawned mostly out of necessity, as compilers of the time were almost solely translation units, taking C as a portable assembly language and translating it to the target machine code. As such, developers were required to prematurely optimize [6] memory, before the compiler could perform meaningful optimizations.

For projects where performance is a primary goal, manual memory optimizations are prevalent throughout the source code. Anytime a developer wants to change a data structure, they must consider the implications of that change on existing memory optimizations. A daunting task, as memory optimizations are performed by careful consideration of both the data structure definition and its multitude of allocations. However this leaves compilers with lacking degrees of freedom, as these decisions are fixed before compilation.

As a result, production compiler optimizations either focus on scalar values or are limited in their applicability when memory is involved. Modern compilers seek to perform more aggressive transformations, such as automatic vectorization and parallelization [7–20], to fully utilize modern, multi-core processors. Such transformations require precise information about data and control dependencies in the program [21–23]. For programs operating on scalars, these dependencies can be easily analyzed with SSA forms [24,25]. However these techniques are severely limited when dealing with applications operating on complex data structures holding increasingly large amounts of data that must be stored in memory.

At present, only fixed-length arrays and objects have SSA forms [26,27]. Compilers, therefore, must rely on pointer analyses for data flow information about memory objects. This information can be improved by field-sensitive [28] and type-based [29] analyses, however common manual memory optimizations create spurious dependencies and ambiguity that the compiler cannot resolve. An example of this is allocation reuse, wherein a memory location is used to represent multiple objects over the execution of the program. This optimization is common for vectors, which may use the same memory location for different elements throughout its lifetime. This aggregates the disjoint lifetimes of individual elements into a single, long-lived lifetime. Through such premature optimizations, the compiler cannot distinguish between dependencies injected by the developer and those logically necessary.

The problems facing modern compilers are the culmination of ambiguous memory behavior and lacking degrees of freedom for dependency breaking transformations. To remedy this, the compiler requires unambiguous memory operations via strong guarantees about the type, allocation, and usage of memory within the program. Memory behavior must be presented to the compiler in a form that can be easily analyzed and meaningfully transformed.

This paper proposes the Memory Object Intermediate Representation (MEMOIR). MEMOIR provides the compiler with an SSA representation for sequential and associative data collections. Additionally it defines a representation for objects and their fields. By decoupling the representation of memory used to store data from the memory used to logically organize data, MEMOIR grants powerful guarantees for transformation and enables data flow analysis for elements of collections and fields of objects via def-use chains. MEMOIR also grants the degrees of freedom necessary to change the memory layout and broader memory organization of a program. By providing an IR that is amenable to both analysis and transformation, MEMOIR compilers can emit performant code without placing the burden of memory optimization on developers.
This paper makes the following contributions:

- Makes the empirical observation that the majority of memory used within well-established C/C++ benchmark suites has a higher-level data structure (§III).
- Proposes MEMOIR, an SSA IR for associative and sequential collections, objects and their fields (§IV).
- Presents dead element elimination, redundant indirection elimination, and field elision, novel transformations that rely on element-level analyses (§V).
- Implements algorithms to generate and lower MEMOIR without incurring spurious copies (§VI).
- Evaluates the impact of MEMOIR compiler transformations on well-established C/C++ benchmarks (§VII).

Our MEMOIR compiler implementation is publicly available at https://github.com/arcana-lab/memoir.

II. BACKGROUND

Compilers rely on data flow analyses to perform optimizations, such as dead code elimination and common sub-expression elimination. The SSA form [24] was introduced to provide a sparse representation for the data flow of variables throughout a program via def-use chains. This representation enables compilers to perform data flow analyses more efficiently and succinctly [30]. As such, SSA has become the de facto standard for modern, optimizing compiler intermediate representations (IR) such as LLVM [3] and GCC [2].

The Jalapeño compiler [31] sought to represent a subset of memory objects in Java programs with an SSA form. Their work made use of Array SSA form [26] to capture the semantics of fixed-size arrays and the Extended Array SSA form [27] to represent the fields of statically-, strongly-typed objects. This enabled the compiler to perform parallelization of array processing applications and load-store propagation for accesses to object fields by analyzing elements of these simple data collections at any given program point. While this work enabled compilers to more easily analyze simple array and object structures in a program, it does not provide a general representation for data structures that are sequential or associative. These data structures have operational semantics that cannot be represented as simple arrays or individual objects. Without abstracting these collections into a general form, the compiler is unable to provide a unified analysis.

In dealing with a similar problem, tensor compilers have provided an inspiring generalization for higher-dimensional array structures. Efficient memory layout of tensors is heavily tied to the structure of the data stored therein, commonly classified as either sparse or dense. Sparse tensors are commonly stored as tree structures to reduce memory usage, limiting the ability for compilers to perform optimizations because of the linked data structure. To remedy this, the TACO compiler [32–34] provides a general representation for tensor operations, abstracting away the storage structure as either sparse or dense. The compiler can thus perform unified analysis of tensors while still reasoning about the structure of the stored data when generating code, illustrating the power of generalized representations while still maintaining rich semantics.

III. MOTIVATIONS

At present, TACO and Jalapeño’s representations do not generalize to common data structures. These include, linked lists, key-value stores, and memory allocations that grow and shrink throughout program execution. To illustrate the prevalence of these data structures in modern programs, we inspected the memory usage and accesses of C/C++ benchmarks in the SPECspeed 2017 Integer suite (SPECFINT 2017).

![Figure 1: Classification of SPECINT 2017 heap memory usage.](image)

Figure 1 shows the breakdown of heap memory usage, as reported by Valgrind’s dynamic heap analysis tool [35]. These benchmarks use a negligible amount of stack memory. We manually classify collections into one of six classes. Sequential collections are contiguous in index space, e.g., arrays, vectors and linked lists. Associative collections store a relation between key-value pairs and have an index space composed of the keys stored therein. Objects are fixed-length data structures composed of heterogeneously-typed fields. Trees and graphs represent any tree or graph-based data structure, making no distinction between their storage formats. Unstructured collections have no well-defined structure or are dictated by an externally defined memory layout, such as a file. Of note, the existing work on SSA forms described in §II does not fully cover sequences as Array SSA does not provide a representation for the reallocation of sequential collections; associative arrays, trees and graphs do not have any SSA representation; and unstructured data has no representation at present, aside from file readers represented as streams [36].

As illustrated above, modern applications utilize complex data structures. However, these data structures entail increasingly complex memory dependencies in the IR due to ambiguities in their representation. These memory dependencies result in overly-conservative analyses, limiting existing techniques and constraining the compiler’s optimization space. This paper focuses on three sources of ambiguity in modern compiler representations. These are linked data structures, stateful data accesses, and premature memory optimizations.
Linked data structures are the source of the undecideability of memory alias analysis [37]. These data structures hinder static analyses, as they are unable to determine with certainty where in the list a value came from. However, these data structures do possess an underlying index space that the compiler can use to disambiguate accesses. By representing linked lists as a sequence of values, similarly to an array, the index space can be exploited by the compiler. This comes from decoupling the representation of the memory used to store the data from the memory used to logically organize the data.

Stateful data accesses create data dependencies between unrelated memory operations. Listing 1 illustrates a piece of code that is unoptimizable by language-agnostic IR optimizations. Even though the key 0 is statically known to be unique from 1, neither clang++, g++, nor icpc can propagate the constant 10 to the return statement: to the compiler every write to the map may update the state of the entire collection. In the code example, this is caused by a premature lowering of std::unordered_map to its implementation in the standard library.

Listing 1: Example of stateful data access in C++.
```c++
1 int work (std::unordered_map<int, int> &map) {
2    map[0] = 10;
3    map[1] = 11;
4    return map[0];
5 }
```

Memory optimizations performed by developers, such as over-provisioning of arrays, cause memory analyses to fail when details about element-level liveness is needed. This result from memory locations with disjoint lifetimes being aggregated into the same, long lived lifetime. As a result, the compiler cannot disambiguate accesses between unique objects, creating unresolvable false dependencies.

To remedy these shortcomings of the compiler we propose MEMOIR, an SSA representation for associative and sequential data collections, objects and the fields contained therein. MEMOIR provides a representation for the most commonly used memory classes in SPECINT 2017: objects, sequences and associative arrays. MEMOIR’s SSA form enables the analyses needed by modern optimizations, which require a detailed understanding of the memory locations being accessed throughout the program. To guarantee this SSA form, we propose the use of the MEMOIR type system, which enforces static-, strong-typing for SSA collection variables.

IV. SSA FORM FOR COLLECTIONS AND OBJECTS

MEMOIR is an intermediate representation with data collections as first-class citizens. To enable element-level analysis and transformation on generalized data collections, MEMOIR provides three main properties for collections.

- Collections have a static single assignment (SSA).
- Collections are value types.
- Collections and their elements have static, strong types.

These properties make MEMOIR data collections immutable. MEMOIR consists of named variables for collections, instructions for construction, access, and data flow of collections and objects, and a type system to represent object layout.

MEMOIR instructions allow the creation of new collections from existing ones with unambiguous operations to add, remove or redefine elements. The syntax for MEMOIR collections, objects and types is shown in Figure 2.

A. Collections

We define a collection to be an index-value mapping. Indices and values stored within a collection have static, strong types. We refer to an individual index-value pair in the collection as an element. We refer to the set of indices present in the index-value mapping as the index space of a collection. In this paper we explore two variants of collections with different constraints on the index space, sequential and associative, explained in §IV-C and §IV-D, respectively. These variants harken back to the concepts of position- and value-dependent containers introduced by prior work [38].

B. Operations on SSA Collections

Reading elements of a collection is performed with the READ operator, of the form \( v = \text{READ}(c, i) \), where \( c \) is the collection being accessed and \( i \) is the index being read from. The value held at that index is stored in the variable \( v \). Reading from an uninitialized element is considered undefined behavior. Reading from an index not in the index space of the collection is similarly undefined behavior.

The USE\( \phi \)'s as introduced in prior work [27], links accesses to the same collection in control flow order. This allows sparse data flow analyses to associate a lattice variable with each access, disaggregating use information from the definition. As USE\( \phi \)'s are not needed for every analysis, and increase program size by an additional instruction per-read, they can be constructed and destructed on demand via copy-folding [30].

Redefinition of elements in MEMOIR is accomplished by the write operation. Write operations are of the form \( c_1 = \text{WRITE}(c_0, i, v) \), where \( c_1 \) is a copy of the input collection \( c_0 \) with the exception of \( c_1(i) = v \). With the write operation, a fixed-size collection is put in SSA form. Write operations in MEMOIR are similar to functional updates or DEF\( \phi \)'s, as introduced in prior work [26,27].

Insertion and Removal of elements in an SSA form is provided by the insert and remove operations. These operations are the only way changes to the index space of collections can be represented. Insert operations are of the form \( c_1 = \text{INSERT}(c_0, i, v) \), where \( c_1 \) is a copy of \( c_0 \) with index \( i \) inserted and \( c'[i] = v \) if \( v \) is defined. Remove operations, of the form \( c_1 = \text{REMOVE}(c_0, i) \), where \( c_1 \) is a copy of \( c_0 \) except for the element at index \( i \) being removed. The semantics of these operations depends on the type of collection being operated on, the details of which are expanded upon in §IV-C and §IV-D.

Copying Elements of a Collection is performed by the copy operation, of the form \( c_1 = \text{COPY}(c_0) \), which creates a new collection, \( c_1 \), with the same index-value mapping of \( c_0 \). Additional semantics for sequences are explained in §IV-C.

The Size of a Collection can be queried with the size operation, of the form \( n = \text{size}(c_0) \). The result of this query is the number of index-value pairs in the collection.
C. Sequences

A sequence is a collection of data organized with a contiguous index space. This index space is defined as the linear order of \( \{ n \in \mathbb{N} \mid 0 \leq n < l \} \), where \( l \) is the sequence’s length. SSA sequence variables are denoted as \( \text{id} \). For example, \( S_0 : \text{Seq} \langle \text{i32} \rangle \) is a sequence with elements of 32-bit signed integers. More information about element types is included in §IV-E.

A new sequence with \( n \) elements of type \( T \) can be created via a \text{new Seq}<T>(n)\) instruction; \( n \) need not be statically known, but the length of the sequence is fixed upon allocation. Elements are uninitialized upon allocation. For operations on sequences, the \text{end} symbol is used as syntactic sugar for the \text{size} of the sequence being accessed.

On account of their sequential index space, sequences have additional semantics for the \text{remove}, \text{copy} and \text{swap} operations. A visual summary of these differences is shown in Figure 3 The first is the ability to operate on \text{ranges}, as the index space is contiguous. This is done by specifying a second index for the argument, which is the end of the range. We will use the shorthand \( S[a:b] \) to represent the range of a sequence from \( a \) to \( b \) (exclusive) when explaining the semantics of such operations. For example, \( S_1 = \text{REMOVE}(S_0, i, j) \) creates \( S_1 \), where \( S_1[0:i] = S_0[0:i] \) and \( S_1[i:end] = S_0[j:end] \).

Additionally, the \text{insert} operation can insert elements from a sequence at a given index. For example, \( S_2 = \text{INSERT}(S_1, i, \phi) \), which creates \( S_2 \), where \( S_2[0:i] = S_1[0:i] \) and \( S_2[i+\text{size}(S_1)-1:j] = \phi \). If it is, then \( S_2 \) = \text{INSERT}(S_1, i, j) \) creates \( S_3 \), a copy of \( S_0 \) except for the ranges \( i:j \) and \( k:k+j-1 \) swapped. Similarly, \( S_3,S_2 = \text{SWAP}(S_1,i,j,S_0,k) \) swaps ranges \( i:j \) and \( k:k+j-1 \) between sequences \( S_1 \) and \( S_0 \).

D. Associative Arrays

An associative array is a mapping from keys to values. For example, \( A_0 : \text{Assoc}\langle \text{f32} \rangle \) is an associative array of 32-bit floating point keys to boolean values.

In addition to the collection operations defined in §IV-B, the \text{has} operator, denoted \( \text{has}(A_0, k) \), checks if the given key \( k \) is contained in the associative array \( A_0 \). If it is, then a true value is written to the boolean variable \( \text{has} \), otherwise false is written. Additionally, the keys operator \( S_1 = \text{keys}(A_0) \), creates a sequence \( S_1 \) containing the keys of \( A_0 \); there are no guarantees on the order of keys in \( S_1 \).

The identity equality is used for key comparisons on primitive types. MEMOIR uses shallow equality for reference types, where keys alias iff they reference the same object. For object types MEMOIR checks equality for each element and field.

E. Field Arrays, Objects and Types

Objects flow throughout a program via object references, denoted as \( \text{id} \). Individual objects have explicit creation and deletion sites with the \text{new} and \text{delete} operators, respectively. Semantically, nested objects are stored as unique references within read-only elements of the collection, these references can be read into a variable via the \text{read} operator.

Fields of objects are accessed via field arrays, borrowed from the concept of heap arrays [27]. Field arrays are an associative array, mapping an object reference to a field value. Each field of an object type-defined below has a unique field array and are instantiated with the object type definition. By construction, a field array \text{cannot alias} with any other field.
of the object. This representation for fields allows MEMOIR compilers to easily track all accesses of a given field, even when its owning object is an element of a collection. It also decouples the access of fields from the layout of fields within the object, enabling myriad degrees of freedom to MEMOIR compilers. Among these is the ability to reorder, add or delete fields within an object.

To specify the layout of objects and provide a mechanism for such transformations, MEMOIR includes the object type: an ordered list of individually addressable, typed data fields. For example, \( T \rightarrow \{ a : \text{i32}, b : \text{f32} \} \) defines a new type \( T \) containing two fields. The first having unique identifier \( a \) and is of 32-bit signed integer type. The second being \( b \) of 32-bit floating point type. The field arrays for them being \( F_{T, a} : \phi_T \rightarrow a \) and \( F_{T, b} : \phi_T \rightarrow b \), respectively.

MEMOIR includes support for unsigned and signed integers (e.g., \( \text{i32} \)), booleans, floating point values, and C-style pointers. C-style pointers are included to support operations that require access to locations within conventional memory allocations, such as memory-mapped regions. Reference types are also supported, being nullable references to an object of a given object type. For example, \( \text{seq} \subset \phi_T \) is a sequence type containing elements that are references to structs of type \( \tau_T \).

Object types can contain nested object types, but may not be recursively defined. This ensures that object types have a finite, and statically-known, size in memory and a finite depth equality function when used as keys of associative arrays.

V. Empowering the Compiler

By providing a more comprehensive representation for collections, MEMOIR enables automatic transformations on both the computation performed on data and its memory layout. In this section we will explore this optimization space.

We will operate on a constrained form of LLVM [3], where irreducible loops [39] are not permitted. The \( \mu \)-operation [40] is used to represent a \( \phi \) in the context of a loop: the first operand is the initial value and the second is used by later iterations. The ARG\( \phi \) and RET\( \phi \) functions are used for interprocedural data flow. ARG\( \phi \)s are attached to each parameter of a function, mapping a parameter to its incoming argument from each possible call site. RET\( \phi \)s are attached to each call function and map live-out variables from each possible return statement. During partial compilation, externally visible functions have an unknown operand from an unknown call site added to their ARG\( \phi \)s, similarly, RET\( \phi \)s of indirect calls have an unknown incoming value from an unknown function.

For analysis, we define an expression tree, representing an expression and the expressions that compose it, as well as a range, describing a contiguous subspace of a sequence via expression trees. We then define an analysis for the range of live elements in a sequence at each program point.

Def. 1. An expression tree is a tree where every internal node is an operator and every leaf is either a variable or a constant. We will define a partial ordering of expression trees, \( t_1 \subseteq t_2 \) iff \( t_2 \) contains \( t_1 \) as a subtree.

Def. 2. A range of a sequence is a contiguous subspace of its index space \( [\ell : u] \), where \( \ell \) and \( u \) are expression trees.

Def. 3. A range lattice \( L \) is a lattice comprising \( n \) lattice points \( \{ L_1, ..., L_n \} \). Each lattice point \( L_i \) maps to a range \( [\ell_i : u_i] \). These lattice points are partially ordered with the subset relation \( \subseteq \) with \( L_i \subseteq L_j \) iff \( \ell_i \subseteq \ell_j \) and \( u_i \subseteq u_j \).

Def. 4. The disjunctive merge operator \( \land \) unites the ranges of two lattice points, defined as:

\[
L_i \land L_j = [\ell_i : u_i] \land [\ell_j : u_j] = [\min(\ell_i, \ell_j) : \max(u_i, u_j)]
\]

Def. 5. The conjunctive merge operator \( \lor \) intersects the ranges of two lattice points, defined as:

\[
L_i \lor L_j = [\ell_i : u_i] \lor [\ell_j : u_j] = [\max(\ell_i, \ell_j) : \min(u_i, u_j)]
\]

Live Range Analysis computes ranges for the live elements of sequence variables in a program. This uses a similar approach as the range analysis described in [41], extending it to be context-sensitive and modifying it to operate on sequence variables. The analysis, detailed in Algorithm 1, operates on a constraints graph derived from a constraints system, the rules for which are defined in Table 1, being a backwards propagation of liveness information.

Def. 6. A constraints graph, \( G = (N, E, L, C) \), where: \( N \) is a set of vertices where \( n_a \) is a vertex; \( E \) is a set of edges where \( n_a \rightarrow n_b \) is an edge; \( L : E \rightarrow S \) is a function from edges to constraints; \( C : E \rightarrow c \) is a partial function from edges to call site \( c \), representing context-sensitive edges.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_2 = \phi(S_1, S_0) )</td>
<td>( S_2 \subseteq S_1, S_2 \subseteq S_0 )</td>
</tr>
<tr>
<td>( ... = \text{READ}(S_0, i) )</td>
<td>( i + [0 : 1] \subseteq S_0 )</td>
</tr>
<tr>
<td>( S_1 = \text{USE}\phi(S_0) )</td>
<td>( S_1 \subseteq S_0 )</td>
</tr>
<tr>
<td>( S_1 = \text{WRITE}(S_0, i, v) )</td>
<td>( S_1 \subseteq S_0 )</td>
</tr>
<tr>
<td>( S_1 = \text{INSERT}(S_0, i, v) )</td>
<td>( S_1 \land [0 : i] \subseteq S_0, S_1 \land [i + 1 : \text{end}] - 1 \subseteq S_0 )</td>
</tr>
<tr>
<td>( S_2 = \text{INSERT}(S_1, i, S_0) )</td>
<td>( S_2 \land [i : i +</td>
</tr>
<tr>
<td>( S_1 = \text{REMOVE}(S_0, i, j) )</td>
<td>( S_1 \land [0 : i] \subseteq S_0, S_1 \land [j : \text{end}] \subseteq S_0 )</td>
</tr>
<tr>
<td>( S_1 = \text{COPY}(S_0, i, j) )</td>
<td>( S_1 \land [0 : i] \subseteq S_0, S_1 \land [i : j] - i \subseteq S_0 )</td>
</tr>
<tr>
<td>( S_1, S_2 = \text{SWAP}(S_1, i, j, k) )</td>
<td>( S_1 \land [0 : i] \subseteq S_0, S_1 \land [i : j] - i \subseteq S_0 )</td>
</tr>
<tr>
<td>( S_1 = \text{ARG}\phi(S_1, S_0) )</td>
<td>( S_2 \subseteq S_1, S_0 \subseteq S_0 )</td>
</tr>
<tr>
<td>( S_2 = \text{RET}\phi(S_1, S_0) )</td>
<td>([%a : %b] \subseteq S_1, [%a : %b] \subseteq S_0 )</td>
</tr>
</tbody>
</table>

**TABLE I**: Constraint rules for sequences. Where \( |S_i| = \text{size}(S_i) \), and \( \%a, \%b \) are new parameters of the function.
Dead Element Elimination Using the live slice range analysis, we describe the dead element elimination transformation in Algorithm 2. This transforms sequence access and construction to only operate on the live slice range, eliminating dead code. The materialization function (defined below) is used to perform available expression analysis [44] on an expression tree, constructing operations. Following dead element elimination, constant propagation and folding are applied, simplifying the if-else regions along with a conventional sink pass to move computation into its newly conditioned execution.

Def. 7. The materialization function $M(e, p)$ is a partial function that analyzes an expression tree $e$ at a given program point $p$. This function constructs the necessary operations and returns the resultant value $v$, which is either a variable or a constant. $M(e, p) = e$ iff $e$ is a constant, a parameter of the function containing $p$, or a variable dominating $p$. $M(e, p) = g$ iff $g$ is a variable which dominates $p$ and has the same global value number [45] as $e$. $M(e, p) = \mathit{op}(M(e_1, p), \ldots, M(e_n, p))$ iff $e$ is an expression tree such that $e_i \subseteq e$ for all $i = 1, \ldots, n$, $\mathit{op}$ is an operation with no side effects and $M(e_i, p)$ are defined for all $i$. Otherwise, $M(e, p)$ is undefined.
Listing 3: Abridged quick sort implementation in MEMOIR.

```c
fn qsort(S1: Seq<\tau>, %lo: index, %hi: index) %a = ARG(S1, %lo, %hi)
    \{ ... Lines 2-6 unchanged ... \}
for %i in %lo to %hi // partition
    %p = (%hi - %lo) / 2 // pivot position
    %q = SWAP(S1, %p, %lo) // move pivot
    %pv = SWAP(S1, %lo) // pivot value
    if %i < %lo or (%hi - %lo) == 1 return
    else if %p = %lo + 1 to %hi // partition
        %i = μ(%lo, %q) // temp. pivot pos.
        %s = μ(%s, %i)
    \{ ... Lines 12-13 unchanged ... \}
    if %s < %lo or %hi < %s // partition
        \{ ... Lines 2-6 unchanged ... \}
        if %a <= %lo < %b and %a <= %q < %b
            Sdeel = SWAP(S1, %i, %j)
            \{ ... Lines 12-13 unchanged ... \}
            if %a <= %i < %b
                %jv = READ(S1, %j)
                Sdeel = WRITE(S1, %i, %jv)
            else if %a <= %j < %b
                %iv = READ(S1, %i)
                Sdeel = WRITE(S1, %j, %iv)
            S = φ(Sdeel, Sdeel, Sdeel, S1)
        \{ ... Lines 15-19 unchanged ... \}
```

Listing 4: Optimized quick sort implementation in MEMOIR.

```c
fn qsort(S1: Seq<\tau>, %lo: index, %hi: index)
    %a = ARG(S1, %lo, %hi)
    \{ ... Lines 2-6 unchanged ... \}
for %i in %lo to %hi // partition
    %p = (%hi - %lo) / 2 // pivot position
    %q = SWAP(S1, %p, %lo) // move pivot
    %pv = SWAP(S1, %lo) // pivot value
    if %i < %lo or (%hi - %lo) == 1 return
    else if %p = %lo + 1 to %hi // partition
        %i = μ(%lo, %q) // temp. pivot pos.
        %s = μ(%s, %i)
    \{ ... Lines 12-13 unchanged ... \}
    if %s < %lo or %hi < %s // partition
        \{ ... Lines 2-6 unchanged ... \}
        if %a <= %lo < %b and %a <= %q < %b
            Sdeel = SWAP(S1, %i, %j)
            \{ ... Lines 12-13 unchanged ... \}
            if %a <= %i < %b
                %jv = READ(S1, %j)
                Sdeel = WRITE(S1, %i, %jv)
            else if %a <= %j < %b
                %iv = READ(S1, %i)
                Sdeel = WRITE(S1, %j, %iv)
            S = φ(Sdeel, Sdeel, Sdeel, S1)
        \{ ... Lines 15-19 unchanged ... \}
```

Field Elision is a novel optimization enabled by the element-level analysis and decoupled field semantics of MEMOIR. Field elision converts a field of an object into a key-value pair stored in an associative array. This reduces the memory usage of possibly unused fields and increases the spatial locality of definitely used fields. Data structure splicing [46] pursues a similar goal by migrating fields that are not accessed with their co-located fields frequently to their own object. However, this entails additional pointer fields to locate the migrated fields, increasing the size of objects. Field elision avoids this by introducing a collection instead of pointer fields.

If a field is deemed a candidate for elision via affinity analysis [47,48], the transformation is applied. To apply this for a given candidate \(T.a\), with field array \(F_t.a: \delta T \to U\), construct \(A_{t.a} = \new Assoc<T,U>\) at the beginning of the program’s entry function. For each reference to \(F_t.a\) replace it with \(A_{t.a}\). If \(F_t.a\) is used by an \(\text{ARG}\), create a new parameter in the corresponding function, passing \(A_{t.a}\) as an argument, replacing all uses of \(F_t.a\) in the function with the new parameter. Finally, remove field \(a\) from the definition of \(\tau\).

Redundant Indirection Elimination (RIE) simplifies indirect accesses (e.g., \(a[b[i]]\)) [49] to associative arrays when the index is derived from constant data (e.g., elements of \(b\) are constant). When detected, the keys of an associative array can be replaced with the indices of a sequence or the keys of another associative array. This transforms the collection's type, removing an access to the index collection (e.g., \(b[i]\)).

The analysis is as follows: for an associative array, \(A_0 = \new Assoc<T,U>\), let \(R\) be the set of variables in the def-use chain of \(A_0\), let \(R\) be the set of these variables. If any variable in \(R\) may, but not must, reference \(A_0\) (e.g., control divergence), RIE is not applicable. Otherwise, the transformation checks all accesses \(\varepsilon[k]\), where \(r \in R\). If all keys \(k\) are of the form \(k = \READ(c_i, i)\), where \(c\) is a sequence or associative array and all instances of \(c\) must reference the same collection, then:

1) If \(c\) is a sequence, construct \(c' = \new Seq<U>(\text{size}(c))\).
2) If \(c\) is an assoc. array, construct \(c' = \new Assoc<V,U>\), where \(V\) is the key type of \(c\).
3) \(\forall r \in R\), replace accesses \(\varepsilon[k]\) with \(c'[i]\), where \(k=c[i]\).

Dead Field Elimination utilizes a simple data flow analysis provided by MEMOIR’s element-level analysis to eliminate fields of objects that are trivially dead. If a field array is never read from, i.e., it is never redefined with a \(\text{USE}\phi\) and is never passed into an unknown function during partial compilation, it is deemed dead. If so, all writes to the field array and all variables in its def-use chain are removed and the field is eliminated from the type definition.

VI. A MEMOIR COMPILER

We will now describe our implementation of a MEMOIR compiler, the pipeline of which is laid out in Figure 4. First, we introduce the \(\text{LIB}\) library, a C/C++ programming interface for developers to grant the compiler with guarantees and degrees of freedom necessary to construct and optimize a MEMOIR program. Second, we present the SSA construction algorithm, converting the mutable collections in MUT into immutable SSA collections in MEMOIR. Finally, we discuss the lowering algorithms employed to generate LLVM IR, namely SSA destruction and heap/stack selection.
The **Mut Library.** To instantiate a MEMOIR program, the compiler relies on guarantees for allocation, access and typing of data collections, objects and elements. Similar to past language extensions [50,51], we use a library-compiler codesign to achieve this. To this end, we present the Mut library, consisting of sequences, associative arrays and objects. The Mut library contains explicit operators to directly mutate these collections—outlined in Figure 5—mirroring those available in the standard C++ library. Collections and objects in Mut have the same type properties as those laid out for Memoir in §IV, namely that they have strong, static types and mutable value semantics [52]. By employing the Mut library in their program, developers provide the guarantees and degrees of freedom necessary to construct and optimize a Memoir program. Additionally, by providing an extension rather than a replacement for C/C++, existing programs can be incrementally ported to benefit from Memoir.

**SSA Construction.** We now present the algorithm for constructing the SSA form of Memoir. We will only cover construction of SSA collections; construction of field arrays for object accesses is performed in the same manner as prior work [27]. For collections, a conventional \( \phi \) insertion using the dominance frontier is used to insert \( \phi \)-functions for Mut operations. A depth-first traversal of the CFG dominator tree is performed, applying the rewrite rules in Figure 5 are to Mut operations. Reaching definitions are updated accordingly—\( ReachDef(v') = ReachDef(v) \) and \( ReachDef(v) = v' \) for each \( v,v' \) pair in the rewrite rule.

---

**Fig. 5:** Mut operations (left) and their mapping to Memoir operations (right). Useφ’s are only constructed on demand.

---

**Algorithm 3: Algorithm to destruct the SSA form.**

```
in: f a Memoir function
for B ∈ preorder DFS of f’s CFG dominator tree do
  I = Def(v) in storage order of B do
    if v = new Seq(n) then UPDATE(v,v[0:n])
    else if v = keys(a) then UPDATE(v,v[0:end])
    else if v = ϕ(S0, ..., Sn) then
      construct S' = ϕ(S'0, ..., S'n).
      t0 = φ(t0, ..., t0), and
      t0 = φ(t0, ..., t0), where S0 = S'[f0, t0]...
      UPDATE(v,S0[f0:t0])
    else if v = Useφ(c0) then replace uses of v with c0
    else if v = Operands(I) is not dead after this use then
      replace uses of S0 with COPY(S0)
    if v = Write(c0, i, ...) then
      replace uses of v with c
    if v = Remove(c, ...) then
      construct remove(c, ...)
      replace uses of v with c
    if v = COPY(S0, a, b) then UPDATE(v,S[a:b])
    if v = Insert(S0, i, S1) then
      construct insert(S0, i, S1)
      replace uses of v with S0
    if v = Swap(S0, i, j, k) then
      construct swap(S0, i, j, k)
      replace uses of v with S0
    if v,w = Swap(S0, i, j, k) then
      construct swap(S0, i, j, k)
      replace uses of v with S0 and uses of w with S1
  end

  function UPDATE(S, S'[a:b])
    foreach u of S do
      if U=S[c:d] then replace S[c:d] with S'[a+c:a+d]
    else replace S with S'[a:b]
  end

  function COPY(c, Collection)
    if c = S0[l:j] then return COPY(S0, i, j)
    else return COPY(c)
  end
```

**SSA Destruction.** After SSA construction and optimizations are applied, Memoir collections are lowered to LLVM IR. SSA destruction, shown in Algorithm 3, coalesces collection variables, replacing SSA operations with ones that operate directly on their memory representation. Just as in scalar SSA, great care must be taken in destructing Memoir, as a naïve approach could drastically increase the number of allocations and copies in the program. Our Memoir compiler employs an SSA destruction algorithm with a focus on avoiding such spurious copies. These operations act directly on collections and views of sequences, which represent a contiguous subset of their index space, denoted \( S[<f,t>] \). A variable being dead or alive is in reference to the program point following the instruction under inspection.

**Collection Lowering.** Finally, new operators are lowered to either a heap or stack allocation using the implementation in the standard library, e.g., std::vector or std::map. If an escape analysis computed on a new operator indicates that the collection or object is dead at all exit points of its containing function, it will be allocated on the stack; otherwise it is allocated on the heap.
VII. Evaluation

To evaluate MEMOIR’s construction, destruction and optimizations we have implemented a MEMOIR compiler on the production-quality LLVM 9.0.0 compiler infrastructure, allowing us to utilize the open-source NOELLE [22,53] compiler framework for its loop-level analysis. The developer effort of our compiler implementation is shown in Table II. We have also manually converted the hot collections, which spread throughout much of the codebase, from SPECINT 2017’s mcf_s, deepsjeng_s and LLVM’s opt to use collections and objects from the MUT library. We evaluate the compilation time and effectiveness of our MEMOIR compiler (§VII-B) and the effectiveness of MEMOIR optimizations (§VII-C). We also analyze LLVM passes that MEMOIR could improve (§VII-D).

A. Experimental Setup

Our evaluation was performed on a server with two Intel Xeon Gold 6258R (Cascade) processors running at 2.70GHz. Each processor has 28 cores with 2-way hyper-threading, 32KiB 8-way L1d and 1MiB 16-way L2, backed by 38.5MiB 11-way LL, all having a 64B line size. They are supported by 512GiB of DDR4 main memory running at 2933MT/s. The OS used is Red Hat Enterprise Linux 8.7.

We evaluate against four production-quality compilers: GCC 8.5.0 (GCC), ICC 18.0.1 (ICC), LLVM 9.0.0 (LLVM9), and LLVM 14.0.6 (LLVM14). All performance results are gathered from 10 executions per configuration, all values are the median of those executions relative to LLVM9.

B. Compilation

We evaluated the effectiveness of our compiler at avoiding spurious copies and providing compilation time on par with modern optimizations. Table III shows the breakdown of our MEMOIR compiler’s performance on mcf and deepsjeng from SPECINT 2017 as well as LLVM’s opt middle-end compiler. Compilation time for solely performing SSA construction and destruction (MEMOIR O0) is on par with that of clang -00. Compiling with all MEMOIR optimizations and opt -03 causes a reasonable increase in compilation time and could be improved with further engineering. We also show that no spurious copies are introduced.

TABLE II: MEMOIR passes require reasonable developer effort.

<table>
<thead>
<tr>
<th>MEMOIR</th>
<th>LLVM SLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEE</td>
<td>1211</td>
</tr>
<tr>
<td>DFE</td>
<td>267</td>
</tr>
<tr>
<td>FE</td>
<td>580</td>
</tr>
<tr>
<td>RIE</td>
<td>461</td>
</tr>
</tbody>
</table>

TABLE III: Our MEMOIR compiler has reasonable compilation time with no spurious copies added by SSA construction.

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Compile Time (ms)</th>
<th># Collections</th>
<th>Source</th>
<th>SSA</th>
<th>Binary</th>
</tr>
</thead>
<tbody>
<tr>
<td>mcf</td>
<td>70.6</td>
<td>776.4</td>
<td>20.9</td>
<td>663.2</td>
<td>5</td>
</tr>
<tr>
<td>deepsjeng</td>
<td>246.0</td>
<td>1867.6</td>
<td>34.8</td>
<td>852.8</td>
<td>2</td>
</tr>
<tr>
<td>LLVM opt</td>
<td>225.9</td>
<td>668.4</td>
<td>52.0</td>
<td>414.7</td>
<td>8</td>
</tr>
</tbody>
</table>

C. Performance and Memory Usage Impact of Optimizations

We evaluated the impact of MEMOIR optimizations on mcf_s and deepsjeng_s from SPECINT 2017. To evaluate the impact of the optimizations described in §V we utilize a separate compilation pipeline. The aforementioned programs are manually transformed, following the algorithms described in §V. These programs are then compiled using clang 9.0.0, with optimization level o3. By transforming these optimizations manually instead of automatically, we are able to isolate their impact without additional noise caused by different collection implementations from the baseline, which may stress different regions of the program.

We evaluate multiple permutations of MEMOIR compiler transformations: Dead Element Elimination (DEE), Dead Field Elimination (DFE), Field Elision (FE), and Redundant Indirection Elimination (RIE). The ALL configuration has all of the above transformations applied. Evaluation of LLVM’s opt is not included as the MEMOIR optimizations explored in this paper were not applicable. Figures 6 and 7 show the execution time and maximum resident set size (max RSS), respectively, with the ALL configuration. We see that MEMOIR optimizations are able to reduce the memory usage of both mcf and deepsjeng, while only mcf sees a reduction in execution time. For deepsjeng, only field elision and key folding were applicable, eliding a 16-bit field from the hottest data structure. This allowed for better packing of the struct in memory, reducing the memory usage by 16.6% but entailing a 5.1% increase in execution time due to cache performance.

All MEMOIR optimizations were applicable to mcf, we will now explore their individual impact and interplay. DEE results in a 26.6% speedup over LLVM9 by reducing the computational complexity of mcf’s quick sort on a sequence of length n = K + B from O(nlog(n)) to O(n + Blog(B)). FE, in the absence of other optimizations, causes a 10.4% increase in execution time and a 3.3% increase in max RSS, where a single pointer field is elided. The resulting associative array is lowered to a hashable, whose key-value store causes increases memory usage and expansion of the table causes the increased execution time. When combined with RIE, the positive impact of FE can be seen with a 10.4% decrease in max RSS with
only a 1.3% increase in execution time, as the associative array is converted into a single sequence. This removes the storage of the key, only needing to store the value.

When combined with DFE the memory size of the object is shrunk to 56 bytes, allowing fields of more than one object to be stored on the same cache line. Because accesses to these objects are filtered by the value of the first field, filtered iterations (i.e., where the rest of the object is not read), can be processed along with the previous iteration, achieving a 4.7% speedup over the baseline. This combination also greatly reduces the memory usage of the benchmark, reducing the max RSS by 20.8%, approximately 844MiB. When ALL optimizations are applied an additional 2.1% speedup is seen over solely DEE and max RSS reduction remains 20.8%.

D. Pass Analysis

We evaluated optimizations in the LLVM 16.0.3\(^2\) compiler pipeline to locate areas that traditional compiler techniques could be improved by MEMOIR, the goal being to find an upper-bound for such benefits. Results were gathered via manually inserted counters, from an invocation of `opt -O3` on the whole-program bitcode generated by gclang.

Global Value Numbering is restricted by LLVM’s opaque memory locations, which prevent assignment of memory and pointer operations to existing congruence classes, introducing a large number of memory-related value numbers, as seen in Figure 10. With element-level information, these memory locations can be mapped to a smaller set of congruence classes.

Sink is constrained by memory operations, which create barriers that instructions cannot be safely moved across. Figure 11 illustrates how common of an occurrence this is, as many attempts to sink an instruction fail because of instructions that may write or reference the same memory location. With unambiguous representations for what memory operations are being performed and which elements are being operated on, MEMOIR could enable additional, safe, code motion.

Constant Folding is blocked by an inability to propagate constants across memory barriers, the breakdown of which is seen in Figure 12. The element-level analysis provided by MEMOIR allows constants to be propagated along a collection’s DEF-USE chain, an optimization performed in [27] for fields. Prior work has also proposed conditional constant propagation algorithms for Array SSA [54], which could be repurposed by MEMOIR compilers.

VIII. RELATED WORK

A. Compiler Intermediate Representations

Array SSA [26] provides a first class representation for arrays, but fails to generalize to sequential collections that grow and shrink throughout execution. Extended Array SSA [27] provides a representation for accessing fields of objects with static, strong types, introducing what this paper calls field arrays. MEMOIR generalizes both of these representations, making it amenable to techniques that use them [31,54–57].

Memory SSA [58] and similar works [59,60] provide a sparse representation of points-to information, enabling flow analyses on memory blocks. However, they do not grant degrees of freedom to the compiler as they do not provide a representation for the structure of data in memory.

MLIR [61] provides a compiler infrastructure for progressive lowering from higher- to lower-level representations. MLIR’s memref dialect represents memory objects, but is purely a wrapper for the LLVM memory representation, providing no additional guarantees as the underlying memory is mutable. MLIR seems a useful substrate for MEMOIR compilers, but the concepts introduced by MEMOIR are
missing, making its introduction a matter of engineering. Additionally, there is currently no MLIR-based compiler capable of meaningfully handling the complexity of SPEC benchmarks.

SDFG’s [62] represent the program as a stateless, acyclic graph of data containers and computation, wrapped by stateful nodes. This representation does not provide meaningful abstractions for their containers, suffering from the memory ambiguities shown in §III, especially stateful data accesses which require each access be wrapped in a stateful node. Similarly, HPVM [63] provides a dataflow graph representation of the program, with explicit data movement between nodes. These nodes may have side effects, which are modeled with the same low level memory representation as LLVM, making HPVM subject to the same memory aliasing problems.

B. SSA Destruction and Register Allocation Techniques

SSA destruction of data collections in MEMOIR bears a strong resemblance to the register allocation problem, known to be NP-complete [64]. We present an algorithm with a primary focus on avoiding spurious copies, simply preserving the user-defined allocation. Adapting strategies from scalar variable allocation [65–68], could enable MEMOIR compilers to optimize the allocations. Advanced escape analyses could be used to improve the stack allocation algorithm [69–72].

C. Precise Reasoning with Collections and Structured Heaps

There is a growing body of work related to the static analysis of data collections. Dillig et al. [38] proposed a framework for analyzing the contents of containers—analogous to element-level analysis in this paper—for the purpose of verification. Unifying their framework with MEMOIR provides an exciting opportunity for sound collection lowering to implementations outside of the standard library. More recently, structured heaps [73] have been proposed as a means of enabling static analysis for programs using dynamic memory, to which MEMOIR bears a strong, albeit higher-level, resemblance.

D. Programming Languages Amenable to MEMOIR

At its core, MEMOIR proposes collections as value types. In this paper, we implement a library in C/C++ to provide this functionality, however many languages exist which provide the guarantees needed for a MEMOIR compiler. Languages with mutable value semantics [52], which degrades references to second-class citizens, are amenable to SSA construction, as they are analogous to our MUT library. Such languages include Swift’s struct types [74] and Hylo [75].

Languages with single-ownership, i.e., “borrowing”, which guarantee that only one mutable reference will exist at a time can be used to construct a MEMOIR program. An example of this is Rust [76], which is steadily entering the programming zeitgeist. Similarly to Rust, newer languages such as Mojo [77] and Vale [78] have similar ownership models. Of note, use ϕ’s cannot be constructed for these languages, as multiple immutable references may exist at once. While the aforementioned languages are promising directions of future work, the lack of accepted benchmark suites implemented in them, unlike C/C++, was deemed too large a barrier to adoption in our research at present.

Collection-oriented languages [79–82] have existed for many years now. APL [79] and SETL [82] serve as prime examples of their philosophy, focusing on arrays and sets, respectively, as prime concepts of the language. As such they are interesting source languages for compilation and, furthermore, their implementations provide a wealth of resources on optimizing collection-oriented programs [83]. A recent example of these concepts being exploited outside of their original languages is parallel block-delayed sequences [84], which implements loop-fusion techniques on sequences as a library for Parallel ML and C++. Investigating the extent these optimizations could be performed statically with MEMOIR provides an interesting starting point for this line of research.

IX. Conclusion

This paper introduced MEMOIR, a compiler intermediate representation in an SSA form for data collections and objects stored in memory. The core of MEMOIR comes from a decoupling of the memory used to store data from the memory used to logically organize data. Through this decoupling, the compiler is granted a generalized representation of both sequential and associative collections with well-defined operational semantics that cover common operations performed on them. This paper also introduced a prototype MEMOIR compiler, which is capable of performing element-level analysis on collections thanks to its SSA form and unambiguous operations. Using this analysis, the compiler is able to perform novel memory optimizations that must be applied manually today, including efficient layout of fields, selection of heap or stack allocation and copy elision. Furthermore, MEMOIR enables traditional, scalar analyses and transformations be applied to elements of collections. As an example, we generalized live variable analysis to be live range analysis and used it to perform dead element elimination, a generalization of dead variable elimination. With additional work on MEMOIR compilers, we believe that developers can be liberated from the burden of performing low-level, manual memory optimizations. Additionally, the avoidance of premature lowering and optimization of data collections can remove barriers to optimizations, improving the applicability of traditional compiler optimizations.

Acknowledgements

We thank members of the ARCANA Lab for their support and feedback on this work. We also thank the anonymous reviewers for insightful comments and invaluable suggestions that made this work stronger. This material is based upon work supported by the U.S. Department of Energy under the contract number DE-SC0022268. This material is also based upon work supported by the National Science Foundation under Grants NSF-2119069, NSF-2107042, NSF-2028851, NSF-1908488. All opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF.
A. Abstract

Our artifact includes source files for the prototype MEMOIR compiler described and evaluated in the paper. In addition to this it includes source-code patches for SPEC CPU2017 and LLVM opt. When applied, these patches port the programs to LLVM in the paper, i.e., 100s-1000s.

We have provided a Dockerfile to generate an image with all dependencies necessary to evaluate the artifact. Our artifact requires a network connection to pull down external dependencies and benchmark suite.

B. Artifact Check-list (meta-information)

- Algorithm: SSA Construction, SSA Destruction, Dead Element Elimination, Dead Field Elimination, Field Elision, Redundant Indirection Elimination
- Program: LLVM; Optional: SPEC2017
- Compilation: LLVM9; Optional: LLVM14, GCC, ICC
- Transformations: MEMOIR optimizations and porting implemented as ed scripts.
- Binary: None
- Data set: SPEC CPU2017 Integer (not included)
- Run-time environment: Linux (RHEL 8.7)
- Hardware: x64
- Run-time state: Yes
- Execution: Sole-user
- Metrics: Execution time, Maximum resident set size, Compilation time, Number of collections, Significant lines of code
- Output: Tables 2 and 3, Figures 5, 6, 7 and 8.
- Experiments: Performance and Compiler Evaluation
- How much disk space required (approximately)?: Generated artifact is 11GiB
- How much time is needed to prepare workflow (approximately)?: 30 minutes
- How much time is needed to complete experiments (approximately)?: 45 minutes (1 run, with minimal configurations enabled), 4 hours (1 run, with all configurations enabled)
- Publicly available?: Yes
- Code licenses (if publicly available)?: MIT License
- Data licenses (if publicly available)?: N/A
- Workflow framework used?: Unix Makefiles, Bash, customization described below.
- Archived?: Yes, https://doi.org/10.5281/zenodo.1018391

C. Description

1) How delivered: The artifact is available on Zenodo, at https://zenodo.org/records/10201049.

2) Hardware dependencies: An x64 processor. This artifact has been tested on both Intel and AMD x64 machines.

3) Software dependencies: Running the artifact requires LLVM 9.0.0, Julia 1.9.4, scc, and gcclang. It optionally depends on LLVM 14.0.6, GCC 8.5.0, and ICC 18.0.1.

All of the above dependencies are handled when building the docker image. This artifact has been tested with both docker 24.0.5 and Podman 4.6.1.

4) Data sets: For performance evaluation, the SPEC CPU2017 Integer benchmark suite is required. The LLVM opt data set is automatically downloaded from the network when setting up the artifact.

D. Installation

We have provided a Dockerfile to handle dependencies. If evaluating the artifact with SPEC2017, the easiest solution is to copy your SPEC2017.tar.gz into the artifact directory, which will be transferred to your docker container. If you choose to use docker, run the following two commands from within the installed artifact directory:

```
docker build -t cgo24-artifact .
docker run -it cgo24-artifact
```

After installation, run make config within the artifact directory to configure the artifact. Details about the configuration are included in §G.

E. Experiment workflow

After installation and configuration of the artifact, run make all. The flow of this experiment after executing make all is as follows:

1) MEMOIR is compiled (make memoir-setup).
2) All benchmarks (enabled by the configuration) are compiled (make benchmark-setup benchmarks).
3) All benchmarks are executed, and statistics about their execution time and memory usage are gathered via /usr/bin/time (make performance).
4) Figures 5, 6, 7 and 8 are constructed (make figures).
5) Statistics about the MEMOIR compiler are gathered and Tables 2 and 3 are recreated (make figures).

F. Evaluation and expected result

As the output of the experiment flow, the figures/ directory will be populated with the tables and figures from the paper. Note that we have made various improvements to our compiler in preparation for publicizing our artifact, as such, the exact data in tables 2 and 3 may differ from those in the paper. The takeaway from the evaluation remains the same and is detailed below. Figures 5, 6, 7 and 8 are plotted relative to LLVM9.

Table 2 (figures/table_2.txt) evaluates our prototype MEMOIR compiler. The first portion evaluates the compilation time for the MEMOIR compiler and LLVM compiler using no optimizations (O0) and all optimizations (O3). Expected: the compilation time is within an order of magnitude of the corresponding LLVM compilation time. The second portion evaluates the number of collections in the source program, MEMOIR program (SSA), and the resulting binary. Expected: the compiler does not introduce any new collections, i.e., the number of collections in the binary is less than or equal to the number of collections in the source program.

Table 3 (figures/table_3.txt) evaluates the development effort of our prototype MEMOIR compiler. The SLOC for LLVM passes is not included. Expected: the SLOC of MEMOIR compiler passes is on the same order as those shown for LLVM in the paper, i.e., 100s-1000s.
Figure 5 (figures/figure_5.pdf) evaluates the impact of MEMOIR optimizations on execution time. **Expected:** The execution time with MEMOIR optimizations enabled is reduced by \(\sim 26.6\%\) compared to LLVM9 (allowing for differences in machines). **Expected:** No alternative compiler (if enabled) is able to achieve the same order of magnitude execution time reduction as MEMOIR.

Figure 6 (figures/figure_6.pdf) evaluates the impact of MEMOIR optimizations on memory usage (max RSS). **Expected:** The maximum resident set size with MEMOIR optimizations enabled is reduced by \(\sim 20.8\%\). **Expected:** No alternative compiler (if enabled) is able to achieve the same order of magnitude max rss reduction as MEMOIR.

Figure 7 (figures/figure_7.pdf) evaluate the impact of each MEMOIR optimization, in isolation and concert, on execution time. **Expected:** The execution time reduction with all MEMOIR optimizations enabled is approximately equal to the reduction of DEE and DFE-FE-RIE combined.

Figure 8 (figures/figure_8.pdf) evaluate the impact of each MEMOIR optimization, in isolation and concert, on memory usage (max RSS). **Expected:** The maximum resident set size reduction with all MEMOIR optimizations enabled is approximately equal to the reduction of DEE and DFE-FE-RIE combined, and DFE-FE-RIE is approximately equal to DFE and FE-RIE combined.

G. Experiment customization

We provide various knobs for any evaluator to customize how they would like to run the experiment. An evaluator can customize their configuration by running `make config` in the top-level directory of the artifact. This configuration can be manually set in the `Makefile.config` file. The configuration options are as follows:

- **SPEC2017** can be enabled/disabled by providing or not providing a path to a SPEC2017 tarball with the `SPEC_TARBALL` environment variable to "y".
  
  **Note:** without a SPEC2017 installation, no performance evaluation can be performed, disabling figures 5, 6, 7 and 8. A single run of SPEC2017 without the following configurations enabled is expected to take \(\sim 45\) minutes.

- **Sweeping Compilers** can be enabled/disabled by setting the `SWEEEP_COMPILERS` environment variable to "y". When enabled, figures 5, 6, 7 and 8 will include performance evaluation for GCC and LLVM14. Since ICC requires a license, you must provide a path to your ICC installation when prompted.

  **Note:** enabling this option will increase evaluation time by \(\sim 1.5\) hours per run.

- **Sweeping Optimizations** can be enabled/disabled by setting the `SWEEEP_OPTIMIZATIONS` environment variable to "y". When enabled, the ablation study in figures 7 and 8 will be generated. Without enabling this option, figures 7 and 8 will not be generated.

  **Note:** enabling this option will increase evaluation time by \(\sim 2\) hours per run.

- **Number of Runs** can be set with the `NUM_RUNS` environment variable. This will run each configuration of each benchmark `NUM_RUNS` times. By default, this value is set to three.

  **Note:** This will increase the evaluation time by a multiplication of `NUM_RUNS`.

H. Notes

For more detailed information about the artifact and how to extend it to be used on new benchmarks and compiler configurations, please read the `README.md` in the artifact.

I. Methodology

Submission, reviewing and badging methodology:

- [http://cTuning.org/ae/submission-20190109.html](http://cTuning.org/ae/submission-20190109.html)
- [http://cTuning.org/ae/reviewing-20190109.html](http://cTuning.org/ae/reviewing-20190109.html)
- [https://www.acm.org/publications/policies/artifact-review-badging](https://www.acm.org/publications/policies/artifact-review-badging)

REFERENCES


