



RESEARCH ARTICLE

# Roaring bitmaps: Implementation of an optimized software library

Daniel Lemire<sup>1</sup>  | Owen Kaser<sup>2</sup>  | Nathan Kurz<sup>3</sup> | Luca Deri<sup>4</sup> | Chris O'Hara<sup>5</sup> | François Saint-Jacques<sup>6</sup> | Gregory Ssi-Yan-Kai<sup>7</sup>

<sup>1</sup>TELUQ, Université du Québec, Montreal, QC, Canada

<sup>2</sup>Faculty of Computer Science, University of New Brunswick, Saint John, NB, Canada

<sup>3</sup>Orinda, CA, USA

<sup>4</sup>Istituto di Informatica e Telematica/Consiglio Nazionale delle Ricerche, Pisa, Italy

<sup>5</sup>Kissmetrics, San Francisco Bay Area, CA, USA

<sup>6</sup>AdGear Technologies, Inc., Montreal, QC, Canada

<sup>7</sup>42 Quai Georges Gorse, Boulogne-Billancourt, France

## Correspondence

Daniel Lemire, TELUQ, Université du Québec, 5800 Saint-Denis, Office 1105, Montreal, QC H2S 3L5, Canada.  
Email: lemire@gmail.com

## Funding information

Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: 261437

## Summary

Compressed bitmap indexes are used in systems such as Git or Oracle to accelerate queries. They represent sets and often support operations such as unions, intersections, differences, and symmetric differences. Several important systems such as Elasticsearch, Apache Spark, Netflix's Atlas, LinkedIn's Pivot, Metamarkets' Druid, Pilosa, Apache Hive, Apache Tez, Microsoft Visual Studio Team Services, and Apache Kylin rely on a specific type of compressed bitmap index called Roaring. We present an optimized software library written in C implementing Roaring bitmaps: CRoaring. It benefits from several algorithms designed for the single-instruction-multiple-data instructions available on commodity processors. In particular, we present vectorized algorithms to compute the intersection, union, difference, and symmetric difference between arrays. We benchmark the library against a wide range of competitive alternatives, identifying weaknesses and strengths in our software. Our work is available under a liberal open-source license.

## KEYWORDS

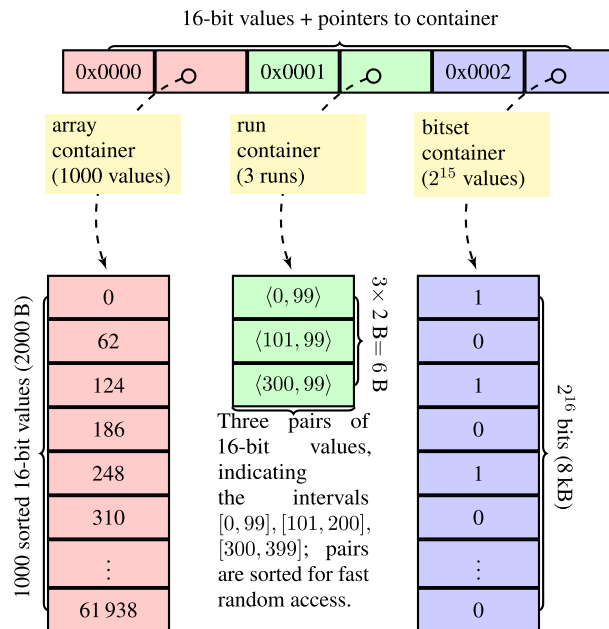
bitmap indexes, database indexes, Jaccard index, SIMD instructions, vectorization

## 1 | INTRODUCTION

Contemporary computing hardware offers performance opportunities through improved parallelism, by having more cores and better single-instruction-multiple-data (SIMD) instructions. Meanwhile, software indexes often determine the performance of big data applications. Efficient indexes not only improve latency and throughput but also reduce energy usage.<sup>1</sup>

Indexes are often made of sets of numerical identifiers (stored as integers). For instance, inverted indexes map query terms to document identifiers in search engines, and conventional database indexes map column values to record identifiers. We often need efficient computation of the intersection ( $A \cap B$ ), the union ( $A \cup B$ ), the difference ( $A \setminus B$ ), or the symmetric difference ( $(A \setminus B) \cup (B \setminus A)$ ) of these sets.

The bitmap (or bitset) is a time-honored strategy to represent sets of integers concisely. Given a universe of  $n$  possible integers, we use a vector of  $n$  bits to represent any one set. On a 64-bit processor,  $\lceil n/64 \rceil$  inexpensive bitwise operations suffice to compute set operations between 2 bitmaps:



**FIGURE 1** Roaring bitmap containing the first 1000 multiples of 62, all integers in the intervals  $[2^{16}, 2^{16} + 100)$ ,  $[2^{16} + 101, 2^{16} + 201)$ , and  $[2^{16} + 300, 2^{16} + 400)$  and all even integers in  $[2 \times 2^{16}, 3 \times 2^{16})$  [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

- the intersection corresponds to the bitwise AND;
- the union corresponds to the bitwise OR;
- the difference corresponds to the bitwise ANDNOT;
- the symmetric difference corresponds to the bitwise XOR.

Unfortunately, when the range of possible values ( $n$ ) is too wide, bitmaps can be too large to be practical. For example, it might be impractical to represent the set  $\{1, 2^{31}\}$  using a bitset. For this reason, we often use compressed bitmaps.

Although there are many ways to compress bitmaps (see Section 2.1), several systems rely on an approach called Roaring including Elasticsearch,<sup>2</sup> Metamarkets' Druid,<sup>3</sup> eBay's Apache Kylin,<sup>4</sup> Netflix's Atlas,<sup>5</sup> LinkedIn's Pinot,<sup>6</sup> Pilosa,<sup>7</sup> Apache Hive, Apache Tez, Microsoft Visual Studio Team Services,<sup>8</sup> and Apache Spark.<sup>9,10</sup> In turn, these systems are in widespread use, eg, Elasticsearch provides the search capabilities of Wikipedia.<sup>11</sup> Additionally, Roaring is used in machine learning,<sup>12</sup> in data visualization,<sup>13</sup> in natural language processing,<sup>14</sup> in Resource Description Framework databases,<sup>15</sup> and in geographical information systems.<sup>16</sup> Wang et al<sup>17</sup> recommended Roaring bitmaps as a superior alternative to other compressed bitmaps.

Roaring partitions the space  $[0, 2^{32})$  into *chunks* consisting of ranges of  $2^{16}$  integers ( $[0, 2^{16})$ ,  $[2^{16}, 2^{17})$ , ...).<sup>18,19,\*</sup> For a value in the set, its least significant 16 bits are stored in a container corresponding to its chunk (as determined by its most significant 16 bits), using 1 of 3 possible container types:

- bitset containers made of  $2^{16}$  bits or 8 kB;
- array containers made of up to 4096 sorted 16-bit integers;
- run containers made of a series of sorted  $\langle s, l \rangle$  pairs indicating that all integers in the range  $[s, s + l]$  are present.<sup>†</sup>

At a high level, we can view a Roaring bitmap as a list of 16-bit numbers (corresponding to the most significant 2 B of the values present in the set), each of which is coupled with a reference to a container holding another set of 16-bit numbers corresponding to the least significant 2 B of the elements sharing the same prefix (see Figure 1).

We dynamically pick the container type to minimize memory usage. For example, when intersecting 2 bitset containers, we determine whether the result is an array or a bitset container on-the-fly. As we add or remove values, a container's type might change. No bitset container may store fewer than 4097 distinct values; no array container may store more than

\*There are 64-bit extensions to Roaring.<sup>20</sup>

†The Roaring implementation<sup>2</sup> found in Elasticsearch lacks run containers but has “negated” array containers, where all but the missing values are stored.

**TABLE 1** Our Roaring optimizations and how they relate to prior work

Containers	Optimization	Section	Prior Work
bitset→array	converting bitsets into arrays	Section 3.1	
bitset+array	setting, flipping, or resetting the bits of a bitset at indexes specified by an array, with and without cardinality tracking	Section 3.2	
bitset	computing the cardinality using a vectorized Harley-Seal algorithm	Section 4.1.1	Muła et al <sup>24</sup>
bitset+bitset	computing AND/OR/XOR/ANDNOT between 2 bitsets with cardinality using a vectorized Harley-Seal algorithm	Section 4.1.2	Muła et al <sup>24</sup>
array+array	computing the intersection between 2 arrays using a vectorized algorithm	Section 4.2	Lemire et al <sup>22</sup> Schlegel et al <sup>23</sup>
array+array	computing the union between 2 arrays using a vectorized algorithm	Section 4.3	
array+array	computing the difference between 2 arrays using a vectorized algorithm	Section 4.4	
array+array	computing the symmetric difference between 2 arrays using a vectorized algorithm	Section 4.5	

4096 distinct values. If a run container has more than 4096 distinct values, then it must have no more than 2047 runs; otherwise, the number of runs must be less than half the number of distinct values.

Roaring offers logarithmic-time random access: to check for the presence of a 32-bit integer, we seek the container corresponding to the 16 most significant bits using a binary search. If this prefix is not in the list, we know that the integer is not present. If a bitmap container is found, we check the corresponding bit; if an array or run container is found, we use a binary search.

The early implementations of Roaring are in Java.<sup>18,21</sup> Although Java offers high performance, it also abstracts away the hardware, making it more difficult for the programmer to tune the software to the microarchitecture. As an experiment, we built an optimized Roaring bitmap library in C (CRoaring). Based on this work, we make 2 main contributions.

- We present several nontrivial algorithmic optimizations (see Table 1). In particular, we show that a collection of algorithms exploiting SIMD instructions can enhance the performance of a data structure like Roaring in some cases, above and beyond what state-of-the-art optimizing compilers can achieve. To our knowledge, it is the first work to report on the benefits of advanced SIMD-based algorithms for compressed bitmaps.

Although the approach we use to compute array intersections using SIMD instructions in Section 4.2 is not new,<sup>22,23</sup> our work on the computation of the union (Section 4.3), difference (Section 4.4), and symmetric difference (Section 4.4) of arrays using SIMD instructions might be novel and of general interest.

- We benchmark our C library against a wide range of alternatives in C and C++. Our results provide guidance as to the strengths and weaknesses of our implementation.

We focus primarily on our novel implementation and the lessons we learned: we refer to earlier work for details regarding the high-level algorithmic design of Roaring bitmaps.<sup>18,19</sup> Because our library is freely available under a liberal open-source license, we hope that our work will be used to accelerate information systems.

## 2 | INTEGER-SET DATA STRUCTURES

The simplest way to represent a set of integers is as a sorted array, leading to an easy implementation. Querying for the presence of a given value can be done in logarithmic time using a binary search. Efficient set operations are already supported in standard libraries (eg, in C++ through the Standard Template Library or STL). We can compute the intersection, union, difference, and symmetric difference between 2 sorted arrays in linear time:  $O(n_1 + n_2)$  where  $n_1$  and  $n_2$  are the cardinalities of the 2 arrays. The intersection and difference can also be computed in time  $O(n_1 \log n_2)$ , which is advantageous when 1 array is small.

Another conventional approach is to implement a set of integers using a hash set—effectively a hash table without any values attached to the keys. Hash sets are supported in the standard libraries of several popular languages (eg, in C++ through `unordered_set`). Checking for the presence of a value in a hash set can be done in expected constant time, giving other hash-set operations favorable computational complexities. For example, adding a set of size  $n_2$  to an existing

set of size  $n_1$  can be done in expected linear time  $O(n_2)$ —an optimal complexity. The intersection between 2 sets can be computed in expected  $O(\min(n_1, n_2))$  time. However, compared to a sorted array, a hash set is likely to use more memory. Moreover, and maybe more critically, accessing data in the hash set involves repeated random accesses to memory, instead of the more efficient sequential access made possible by a sorted array.

## 2.1 | Compressed bitsets

A bitset (or bitmap) has both the benefits of the hash set (constant-time random access) and of a sorted array (good locality), but suffers from impractical memory usage when the universe size is too large compared to the cardinality of the sets. Hence, we use compressed bitmaps. Although there are alternatives,<sup>25</sup> the most popular bitmap compression techniques are based on the word-aligned run-length encoding compression model inherited from Oracle (byte-aligned bitmap compression<sup>26</sup>): WAH,<sup>27</sup> Concise,<sup>28</sup> EWAH<sup>29</sup> (used by the Git<sup>30</sup> version control system), COMPAX,<sup>31</sup> VLC,<sup>32</sup> and VAL-WAH,<sup>33</sup> among others.<sup>34–36</sup> On a  $W$ -bit system, the  $r$  bits of the bitset are partitioned into sequences of  $W'$  consecutive bits, where  $W' \approx W$  depends on the technique used; for EWAH,  $W' = W$ ; for WAH and many others,  $W' = W - 1$ . When such a sequence of  $W'$  consecutive bits contains only 1s or only 0s, we call it a *fill* word; otherwise, we call it a *dirty* word. For example, using  $W' = 8$ , the uncompressed bitmap 0000000001010000 contains 2 words: a fill word (00000000) and a dirty word (01010000). Techniques such as BBC, WAH, or EWAH use special marker words to compress long sequences of identical fill words. When accessing these formats, it may be necessary to read every compressed word to determine whether it indicates a sequence of fill words or a dirty word. The EWAH format supports a limited form of skipping because its marker words give not only the length of the sequences of fill words but also the number of consecutive dirty words. EWAH was found to have superior performance to WAH and Concise<sup>33</sup> in an extensive comparison. A major limitation of formats like BBC, WAH, Concise, or EWAH is that random access is slow. That is, to check whether a value is present in a set can take linear time  $O(n)$ , where  $n$  is the compressed size of the bitmap. When intersecting a small set with a more voluminous one, these formats may have suboptimal performance.

Guzun and Canahuate<sup>37</sup> compared Concise, WAH, EWAH, and Roaring (without run containers) with a hybrid system combining EWAH and uncompressed bitsets. In many of their tests (but not all), Roaring was superior to Concise, WAH, and EWAH. Similarly, Wang et al<sup>17</sup> compared a wide range of bitmap compression techniques. They concluded their comparison with a strong recommendation in favor of Roaring: “Use Roaring for bitmap compression whenever possible. Do not use other bitmap compression methods such as BBC, WAH, EWAH, PLWAH, CONCISE, VALWAH, and SBH.”

## 2.2 | BitMagic

Mixing container types in one data structure is not unique to Roaring. O’Neil and O’Neil’s RIDBit is an external-memory B-tree of bitsets and lists.<sup>38,39</sup> Culpepper and Moffat’s HYB+M2 mixes compressed arrays with bitmaps.<sup>40</sup> Lemire et al<sup>22</sup> decomposed HYB+M2 into chunks of bitsets and arrays that fit in the CPU cache.

The BitMagic library is probably the most closely related data structure<sup>41</sup> with a publicly available implementation. Like Roaring, it is a 2-level data structure similar to RIDBit.<sup>38,39</sup> There are, however, a few differences between Roaring and BitMagic.

- In BitMagic, special pointer values are used to indicate a full (containing  $2^{16}$  values) or an empty container; Roaring does not need to mark empty containers (they are omitted), and it can use a run container to represent a full container efficiently.
- Roaring relies on effective heuristics to generate a memory-efficient container. For example, when computing the union between 2 array containers, we guess whether the output is likely to be more efficiently represented as a bitset container, as opposed to an array (see the work of Lemire et al<sup>19</sup> for a detailed discussion). BitMagic does not attempt to optimize the resulting containers, eg, the intersection of 2 bitset containers is a bitset container, even when another container type could be more economical.

Roaring keeps a cardinality counter updated for all its bitset containers. BitMagic keeps track of the set cardinality, but not at the level of the bitset containers. Hence, whereas deleting a value in a bitset container in Roaring might force a conversion into an array container, BitMagic cannot know that a conversion is necessary without help: the user must manually call the `optimize` method.

When aggregating bitmaps, it is not always useful to maintain the cardinality of the intermediate result: for this purpose, Roaring allows specific *lazy* operations. The BitMagic library, instead, uses a flag to indicate whether the

precomputed cardinality is valid and systematically invalidates it following any aggregation operation. Thus, by default, BitMagic operations can be faster because the cardinality of the bitset containers is not tracked, but the resulting bitmaps may not be compressed as well as they could be: bitset containers might be used to store few values.

Intuitively, one might think that tracking the cardinality of a bitset as we operate on it—as we do in Roaring by default—could be expensive. However, as we show in Sections 3 and 4, careful optimization can make the overhead of tracking the cardinality small on current processors.

- Roaring uses a key-container array, with 1 entry per nonempty container; BitMagic's top-level array has  $\lceil n/2^{16} \rceil$  entries to represent a set of values in  $[0, n)$ . Each entry is made of a pointer to a container. Such a flat layout might give BitMagic a speed advantage over Roaring in some cases. For example, when doing random-access queries, Roaring relies on a moderately expensive binary search at the top level, whereas BitMagic has direct access. Nevertheless, BitMagic's simpler layout has higher memory usage when many of its containers are empty.
- In addition to full and empty containers, BitMagic supports 2 kinds of containers: bitsets and “gap” containers. They are equivalent to Roaring's bitset and run containers, but BitMagic has no equivalent to Roaring's array containers. Moreover, while both CRoaring and BitMagic use dynamic arrays as containers, CRoaring attempts to use as little memory as possible, whereas BitMagic always uses one of a handful of possible run-container sizes (ie, in 2-byte words: 128, 256, 512, or 1280). This difference suggests that Roaring may offer better compression ratios for some data sets. However, it may be easier to provide a custom memory allocator for BitMagic.
- When appropriate, Roaring uses an intersection approach based on binary search between array containers instead of systematically relying on a linear-time merge like RIDBit.<sup>38,39</sup> That is, we use galloping intersection (also known as exponential intersection) to aggregate 2 sorted arrays of sizes  $n_1$  and  $n_2$  in linearithmic time ( $O(\min(n_1, n_2) \log \max(n_1, n_2))$ ).<sup>42</sup> BitMagic does not use array containers.
- Unlike the current Roaring implementations, the BitMagic library uses *tagged pointers* to distinguish the container types: it uses the least significant bit of the address value as a container-type marker bit. In our C implementation, Roaring uses an array of bytes, using 1 byte per container, to indicate the container type.
- BitMagic includes algorithms for fast random sampling unlike Roaring.

Overall, BitMagic is simpler than Roaring, but we expect that it can sometimes use more memory.

BitMagic includes the following SIMD-based optimizations on x64 processors with support for the SSE2 and SSE4 instruction sets.

- It uses manually optimized SIMD instructions to compute the AND, OR, XOR, and ANDNOT operations between 2 bitset containers.
- It uses a mix of SIMD instructions and the dedicated population-count instruction (`popcnt`) for its optimized functions that compute only the cardinality of the result from AND, OR, XOR, and ANDNOT operations between 2 bitset containers.
- It uses SIMD instructions to negate a bitset quickly.

Naturally, many of our Roaring optimizations (see Section 4) would be applicable to BitMagic and similar formats.

### 3 | FASTER ARRAY-BITSET OPERATIONS WITH BIT-MANIPULATION INSTRUCTIONS

Like most commodity processors, Intel and AMD processors benefit from *bit-manipulation instructions*.<sup>43</sup> Optimizing compilers often use them, but not always in an optimal manner.

#### 3.1 | Converting bitsets into arrays

Two useful bit-manipulation instructions are `blsi`, which sets all but the least significant 1-bit to zero (ie,  $x \& -x$  in C), and `tzcnt`, which counts the number of trailing zeroes (largest  $i$  such that  $x/2^i$  is an integer). Using the corresponding Intel intrinsics (`_blsi_u64` and `_mm_tzcnti_64`), we can extract the locations of all 1-bits in a 64-bit word ( $w$ ) to an array (`out`) efficiently.

```
pos = 0
while (w != 0) {
    uint64_t temp = _blsi_u64(w);
    out[pos++] = _mm_tzcnti_64(w);
    w ^= temp;
}
```



Such code is useful when we need to convert a bitset container into an array container. We can ensure that only a handful of instructions are needed per bit set in the bitset container.

### 3.2 | Array-bitset aggregates

Several other such instructions are useful when implementing Roaring. For example, the `bt` (bittest) instruction returns the value of a bit, whereas other bit-manipulation instructions can set (`bts`), clear (`btr`), and flip bits (`btc`) as they query their value. On a Pentium 4, these instructions had a throughput of 1 instruction every 2 cycles and a latency of 6 cycles. However, starting with the Intel Sandy Bridge microarchitecture (2011), they became much faster with a throughput of 2 instructions every cycle and a latency of 1 cycle<sup>44</sup> for data in registers.

Roaring has bitset containers, which are implemented as arrays of 64-bit words. Some of the most common operations on bitsets involve getting, setting, flipping, or clearing the value of a single bit in one of the words. Sometimes, it is also necessary to determine whether the value of the bit was changed.

We are interested in the following scenario: given a bitset and an array of 16-bit values, we wish to set (or clear or flip) all bits at the indexes corresponding to the 16-bit values. For example, if the array is made of the values {1, 3, 7, 96, 130}, then we might want to set the bits at indexes 1, 3, 7, 96, 130 to 1. This operation is equivalent to computing the union between a bitset and an array container.

- If we do not care about the cardinality and merely want to set the bit, we can use the simple C expression `(w[pos >> 6] |= UINT64_C(1) << (pos & 63))`. Compilers can translate this expression into 2 shifts (one to identify the word index and one to shift the single set bit) and a bitwise OR.

On a recent x64 processor (see Section 5.2) and using a recent popular C compiler (see Section 5.1), we estimate that we can set 1 bit every  $\approx 3.6$  cycles, taking into account the time required to read the positions of the bits to be set from an array.

We can do better by using a single shift (to identify the word index) followed by the instruction `bts` (“bit test and set”). Indeed, if the position of the bit in the bitset is given by the `pos` variable, then we shift `pos` right by 6 bits and store the result in `offset`. With the help of `offset`, we fetch the affected word from memory into a variable `load`. We call `bts` with `pos` and `load`. This sets the corresponding bit in `load`, so we can store it back in memory at index `offset`. Thus, we generate a total of 4 instructions (1 load, 1 shift, 1 `bts`, and 1 store), assuming that the position of the bit to set is already in a register. We find that we are able to set bits about twice as fast (1 bit every  $\approx 1.7$  cycles). We benefit from the fact that processors are superscalar, so that they can execute more than one instruction per cycle.

- We could set bits and track the cardinality but we expect branch predictions to be sometimes difficult, and branch mispredictions can be expensive (eg, over 10 cycles of penalty per misprediction). Using the following C code, we set bits and track the cardinality while avoiding branches.

```
uint64_t old_w = words[pos >> 6];
uint64_t new_w = old_w | (UINT64_C(1) << (pos & 63));
cardinality += (old_w ^ new_w) >> (pos & 63);
words[pos >> 6] = new_w;
```

This code first retrieves the word `old_w` at index `pos/64` (or equivalently `pos >> 6`) in an array of 64-bit words `words`. We then set the bit at index `pos % 64` (or equivalently `pos & 63`) to 1 and call the result `new_w`. We can write back `new_w` to the array `words` and use the bitwise XOR between `old_w` and (`new_w`) to determine whether the bit value was changed. We get about  $\approx 3.6$  cycles per bit set, or about the same speed we get when we use a similar C code without tracking the cardinality.

We can also achieve the same result, set the bit, and adjust the cardinality counter as follows: use a shift (to identify the word index), the instruction `bts`, then an instruction like `sbb` (“integer subtraction with borrow”) to modify the cardinality counter according to the carry flag eventually set by `bts`. Indeed, if the position of the bit in the bitset is given by the `pos` variable, then we can proceed as follows.

- We shift `pos` right by 6 bits (eg, `shrx`) and store the result in `offset`.
- We load the 64-bit word at index `offset` and store it in `load`.
- We call `bts` with `pos` and `load`. This sets the corresponding bit in `load`, so we can store it back in memory at index `offset`.
- The call to `bts` stored the value of the bit in the carry (CF) flag. If the bit was set, then the carry flag has a value of 1; otherwise, it has a value of 0. We want to increment our counter when the carry flag has a value of 1 and leave it unchanged when it has a value of 0. Thus, we can call the subtract with borrow (`sbb`) instruction, passing our

cardinality counter as well as the parameter  $-1$ . It will add  $-1$  to the carry flag (getting 0 and  $-1$ ) and then subtract this value from the counter, effectively adding 0 or 1 to it.

In total, we need 5 instructions (1 load, 1 shift, 1 `bts`, 1 store, and 1 `sbb`), which is one more instruction than if we did not have to track the cardinality.

With this approach, we find that we can set a bit while maintaining a cardinality counter every  $\approx 1.8$  cycles on our processor. It is only 0.1 cycles per bit set slower than if we do not track the cardinality.

According to our experiments, you can set a bit or set a bit while maintaining the cardinality, with almost the same throughput—even when all the data are in the L1 processor cache. A similar conclusion holds for bit flipping or bit clearing. Thus, it is inexpensive to track the cardinality of bitset containers as we modify individual bits on recent x64 processors.

We also find that if we use bit-manipulation instructions and hand-tuned assembly, we can roughly double the speed at which we can change bits in a bitset, compared to a compiler-generated machine code with all optimization flags activated. It is true not only when setting bits but also when clearing or flipping them—in which cases we need to use the `btr` and `btc` instructions instead. Informally, we tested various C compilers and could not find any that could perform nearly as well as the hand-tuned assembly code. This may, of course, change in the future.

## 4 | VECTORIZED PROCESSING

Modern commodity processors use parallelism to accelerate processing. SIMD instructions offer a particular form of processor parallelism<sup>45</sup> that proves advantageous for processing large volumes of data.<sup>46</sup> Whereas regular instructions operate on a single machine word (eg, 64 bits), SIMD instructions operate on large registers (eg, 256 bits) that can be used to represent “vectors” containing several distinct values. For example, a single SIMD instruction can add sixteen 16-bit integers in one 32 B vector register to the corresponding 16-bit integers in another vector register using a single cycle. Moreover, modern processors are capable of superscalar execution of certain vector instructions, allowing the execution of several SIMD instructions per cycle. For example, an Intel Skylake processor is capable of executing 2 vector additions every cycle.<sup>44</sup>

SIMD instructions are ideally suited for operations between bitset containers. When computing the intersection, union, difference, or symmetric difference between 2 (uncompressed) bitsets, we only have to load the data in registers, apply a logical operation between 2 words (AND, OR, ANDNOT, or XOR), and, optionally, save the result to memory. All of these operations have corresponding SIMD instructions. Therefore, instead of working over 64-bit words, we work over larger words (eg, 256 bits), dividing the number of instructions (eg,  $\div 4$ ) and giving significantly faster processing.

Historically, SIMD instructions have gotten wider and increasingly powerful. The Pentium 4 was limited to 128-bit instructions. Later, x64 processors introduced increasingly powerful 128-bit instruction sets: SSSE3, SSE4, SSE4.1, and SSE4.2. Today, we find rich support for 256-bit vector registers in the AVX2 instruction set available in recent x64 processors from Intel (starting with the Haswell microarchitecture, 2013) and AMD (starting with the Excavator microarchitecture, 2015). Upcoming Intel commodity processors will support 512-bit SIMD instructions (AVX-512). In many cases, the benefits of SIMD instructions increase as they get wider, which means that SIMD instructions tend to become more compelling with time.

Compilers can automatically translate C or C++ code into the appropriate SIMD instructions, even when the code does not take into account vectorization (scalar code). However, for greater gains, it is also possible to design algorithms and code that take into account the vector instructions. In these instances, we can use the SIMD intrinsics available in C and C++ to call SIMD instructions without having to use assembly code (see Table 2). These intrinsics are supported across several compilers (eg, LLVM's Clang, GNU GCC, Intel's compiler, and Visual Studio).

### 4.1 | Vectorized population count over bitsets

We can trivially vectorize operations between bitsets. Indeed, it suffices to compute bitwise operations over vectors instead of machine words. By aggressively unrolling the resulting loop, we can produce a highly efficient code. Optimizing compilers can often automatically vectorize such code. It is more difficult, however, to also compute the cardinality of the result efficiently. Ideally, we would like to vectorize simultaneously the operations between bitsets and the computation of the cardinality of the result.

**TABLE 2** Some relevant AVX and AVX2 instructions with reciprocal throughput<sup>a</sup> in CPU cycles on recent (Skylake) Intel processors

Instruction	C Intrinsic	Description	Rec. Throughput <sup>44</sup>
vpand	<code>_mm256_and_si256</code>	256-bit AND	0.33
vpor	<code>_mm256_or_si256</code>	256-bit OR	0.33
vpxor	<code>_mm256_xor_si256</code>	256-bit XOR	0.33
vpsadbw	<code>_mm256_sad_epu8</code>	sum of the absolute differences of the byte values to the low 16 bits of each 64-bit word	1
vpaddq	<code>_mm256_add_epi64</code>	add 64-bit integers	0.5
vpshufb	<code>_mm256_shuffle_epi8</code>	<i>shuffle</i> bytes within 128-bit lanes	1

<sup>a</sup>The reciprocal throughput is the number of processor clocks it takes for an instruction to execute. A reciprocal throughput of 0.5 means that we execute 2 such instructions per clock.

#### 4.1.1 | Vectorized Harley-Seal population count

Commodity processors have dedicated instructions to count the 1-bits in a word (the “population count”): `popcnt` for x64 processors and `cnt` for the 64-bit ARM architecture. On recent Intel processors, `popcnt` has a throughput of 1 instruction per cycle for both 32-bit and 64-bit registers.<sup>44</sup>

Maybe surprisingly, Muła et al found that it is possible to do better.<sup>24</sup> We can use a vectorized approach based on the circuit-based Harley-Seal algorithm.<sup>47</sup> It is inspired by a carry-save adder circuit: given 3 bit values ( $a, b, c \in \{0, 1\}$ ), we can sum them up into the 2-bit value  $(a \text{ XOR } b) \text{ XOR } c$  (least significant) and  $(a \text{ AND } b) \text{ OR } ((a \text{ XOR } b) \text{ AND } c)$  (most significant). We can sum up 3 individual bit values to a 2-bit counter using 5 logical operations, and we can generalize this approach to 256-bit vectors. Starting with 3 input vectors, we can generate 2 new output vectors, 1 containing the least significant bits and 1 containing the most significant (or *carry*) bits with 5 bitwise logical operations (2 XORS, 2 ANDs, and 1 OR):

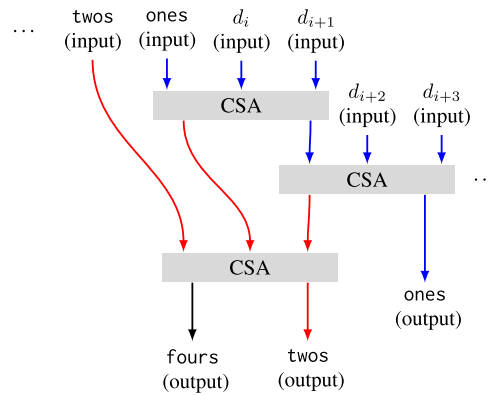
```
void CSA(__m256i *h, __m256i *l, __m256i a,
        __m256i b, __m256i c) {
    // starts with a,b,c, sets H,L, u is temp.
    __m256i u = _mm256_xor_si256(a, b);
    *h = _mm256_or_si256(_mm256_and_si256(a, b),
        _mm256_and_si256(u, c));
    *l = _mm256_xor_si256(u, c);
}
```

From such an adder function, we can derive an efficient population-count function, effectively composing 3-input adders into a circuit with 16 inputs, which encodes the population count of its 16 inputs as a 5-bit binary number. For instance, Knuth<sup>48</sup> presents a construction requiring only 63 logical operations in the case of a 16-input circuit. Although using slightly more logical operations (75), the iterated Harley-Seal approach goes further: it increments the existing contents of a 5-bit binary number by the population count.

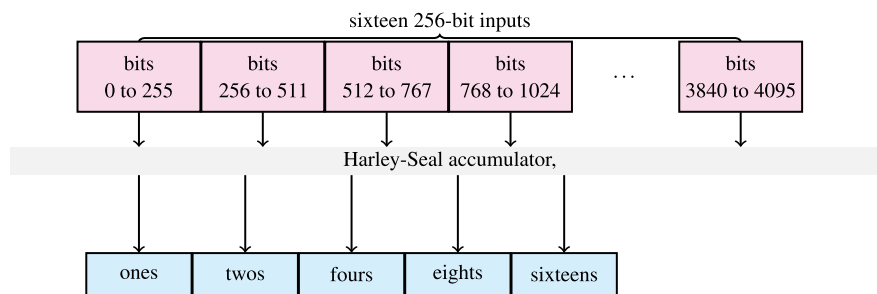
In an SIMD setting, our Harley-Seal circuit takes sixteen 256-bit vectors and updates five 256-bit vectors serving as a bit-sliced accumulator<sup>49</sup>: in other words, if one counts the number of 1s among the least significant bits of the 16 inputs, the population count (0-16) affects the least significant bits of the 5 accumulator vectors. Similarly, the population count of the second least significant bits of the 16 inputs affects the second least significant bits of the accumulator vectors. Equivalently, the 5 accumulator vectors can be viewed as providing 256 different 5-bit accumulators. The accumulator vectors are named *ones*, *twos*, *fours*, *eights*, and *sixteens*, reflecting their use in the accumulator: if the least significant bits of the accumulator vectors are all initially zero and the population count of the least significant bits of the 16 input vectors is 9, then our Harley-Seal approach will result in the least significant bits of *eights* and *ones* being set, whereas the least significant bits of *sixteens*, *fours*, and *twos* would be clear. (The accumulator went from storing 0 to storing 9.) If the accumulator updated a second time with the same inputs, then the least significant bits of *sixteens* and *twos* (only) would be set.

Figure 2 (from the work of Muła et al<sup>24</sup>) illustrates a simplified case with only *ones*, *twos*, and *fours*. Figure 3 shows the high-level process: we load 16 vectors from the source bitset, we compute the bitwise operations, and we obtain the result in *ones*, *twos*, *fours*, *eights*, and *sixteens*. This processes a 4 kb block of an input bitset using approximately 75 AVX operations, not considering any data-movement instructions (see Figure 4). We can compute the total value of all 256 accumulators as  $16 \times \text{count}(\text{sixteens}) + 8 \times \text{count}(\text{eights}) + 4 \times \text{count}(\text{fours}) + 2 \times \text{count}(\text{twos}) + \text{count}(\text{ones})$  (13 more operations), where  $\text{count}(b)$  denotes the population count of  $b$ .





**FIGURE 2** Illustration of the Harley-Seal algorithm aggregating 4 new inputs ( $d_i, d_{i+1}, d_{i+2}, d_{i+3}$ ) to inputs ones and twos, producing ones, twos, and fours<sup>24</sup> [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 3** Vectorized population count for a 4kb bitset, folding sixteen 256-bit inputs into five 256-bit outputs [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

```
CSA(&twosA, &ones, ones, A[i], A[i + 1]);
CSA(&twosB, &ones, ones, A[i + 2], A[i + 3]);
CSA(&foursA, &twos, twos, twosA, twosB);
CSA(&twosA, &ones, ones, A[i + 4], A[i + 5]);
CSA(&twosB, &ones, ones, A[i + 6], A[i + 7]);
CSA(&foursB, &twos, twos, twosA, twosB);
CSA(&eightsA, &fours, fours, foursA, foursB);
CSA(&twosA, &ones, ones, A[i + 8], A[i + 9]);
CSA(&twosB, &ones, ones, A[i + 10], A[i + 11]);
CSA(&foursA, &twos, twos, twosA, twosB);
CSA(&twosA, &ones, ones, A[i + 12], A[i + 13]);
CSA(&twosB, &ones, ones, A[i + 14], A[i + 15]);
CSA(&foursB, &twos, twos, twosA, twosB);
CSA(&eightsB, &fours, fours, foursA, foursB);
CSA(&sixteens, &eights, eights, eightsA, eightsB);
```

**FIGURE 4** Accumulator circuit over 16 inputs ( $A[i+0], \dots, A[i+15]$ )

Iterated 16 times, we can process the entire input bitset. Moreover, it is only necessary to read the total value of all 256 accumulators after the last iteration. However, we need to ensure that, at the beginning of each iteration, none of our 5-bit accumulators ever begins with a count above 15; otherwise, an accumulator might need to store a value exceeding 31. Therefore, at the end of each block, we count the 1-bits in each of the four 64-bit words of *sixteens* and add the result to a vector counter *c* before zeroing *sixteens*. To count the 1-bits in *sixteens* quickly, we use vector registers as lookup tables mapping 4-bit values (integers in  $[0, 16)$ ) to their corresponding population counts, and the `vpshufb` can effectively look up 32 byte values at once. Thus, we can divide each byte of *sixteens* into its low nibble and high nibble and look up the population count of each. To get a population count for the 64-bit words, we use the instruction `vpsadbw` (with the intrinsic `_mm256_sad_epu8`): this instruction adds the absolute values of the differences between byte values within 64-bit subwords. We illustrate the code in Figure 5: the vectors `lookuppos` and `lookupneg` store the positives and negatives of the population counts with a canceling offset of 4 so that their subtraction is the sum of the

```

__m256i popcount256(__m256i v) {
__m256i lookuppos = _mm256_setr_epi8(
    4 + 0, 4 + 1, 4 + 1, 4 + 2,
    4 + 1, 4 + 2, 4 + 2, 4 + 3,
    4 + 1, 4 + 2, 4 + 2, 4 + 3,
    4 + 2, 4 + 3, 4 + 3, 4 + 4,
    4 + 0, 4 + 1, 4 + 1, 4 + 2,
    4 + 1, 4 + 2, 4 + 2, 4 + 3,
    4 + 1, 4 + 2, 4 + 2, 4 + 3,
    4 + 2, 4 + 3, 4 + 3, 4 + 4);
__m256i lookupneg = _mm256_setr_epi8(
    4 - 0, 4 - 1, 4 - 1, 4 - 2,
    4 - 1, 4 - 2, 4 - 2, 4 - 3,
    4 - 1, 4 - 2, 4 - 2, 4 - 3,
    4 - 2, 4 - 3, 4 - 3, 4 - 4,
    4 - 0, 4 - 1, 4 - 1, 4 - 2,
    4 - 1, 4 - 2, 4 - 2, 4 - 3,
    4 - 1, 4 - 2, 4 - 2, 4 - 3,
    4 - 2, 4 - 3, 4 - 3, 4 - 4);
__m256i low_mask = _mm256_set1_epi8(0x0f);
__m256i lo = _mm256_and_si256(v, low_mask);
__m256i hi = _mm256_and_si256(_mm256_srli_epi16(v, 4),
    low_mask);
__m256i popcnt1 = _mm256_shuffle_epi8(lookuppos, lo);
__m256i popcnt2 = _mm256_shuffle_epi8(lookupneg, hi);
return _mm256_sad_epu8(popcnt1, popcnt2);
}

```

**FIGURE 5** A C function using AVX2 intrinsics to compute the 4 population counts of the four 64-bit words in a 256-bit vector [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

population counts. After the 16th iteration, the population count of the entire input bitset is computed as  $16 \times c + 8 \times \text{count}(\text{eights}) + 4 \times \text{count}(\text{fours}) + 2 \times \text{count}(\text{twos}) + \text{count}(\text{ones})$ .

By contrast, in the approach using the dedicated population-count instruction, the compiler generates 1 load, 1 `popcnt`, and 1 add per input 64-bit word. Current x64 processors decompose complex machine instructions into low-level instructions called  $\mu\text{ops}$ . The `popcnt` approach generates 3  $\mu\text{ops}$  per word. For the SIMD approach, we process sixteen 256-bit vectors using 98  $\mu\text{ops}$  including 16 loads, 32 bitwise ANDs (`v pand`), 15 bitwise ORs (`v por`), and 30 bitwise XORs (`v pxor`), or about 1.5  $\mu\text{ops}$  to process each 64-bit word. That is, the vectorized approach generates half the number of  $\mu\text{ops}$ . We analyzed our implementations with the Intel Architecture Code Analyzer (IACA)<sup>50</sup> for our target x64 microarchitecture (Section 5.2). Unsurprisingly, IACA predicts a throughput of one word per cycle with the dedicated population-count function and a throughput of slightly more than two words per cycle with the vectorized approach. Microbenchmarking agrees with IACA and shows that our approach is roughly twice as fast as relying exclusively on dedicated population-count instructions.

#### 4.1.2 | Vectorized operations over bitsets with population counts

To aggregate 2 bitsets (eg, with the intersection/AND, union/OR, etc) while maintaining the cardinality, we found that an efficient approach is to load 2 machine words from each bitset, compute the bitwise operation, apply the population-count instruction on the 2 resulting words, and write them out before processing the next words. To benefit from vectorization, we could compute the bitwise logical operations using SIMD instructions and then reload the result from memory and apply the population-count instructions. The interplay between SIMD instructions and a population-count function operating over 64-bit words would not be ideal.

Although an obvious application of the vectorized population count is the efficient computation of the cardinality of a bitset container, we also put it to good use whenever we need to compute the intersection, union, difference, or symmetric difference between 2 bitset containers, while, at the same time, gathering the cardinality of the result. We can also use it to compute the cardinality of the intersection, union, difference, or symmetric difference, without materializing them. As reported by Muła et al,<sup>24</sup> in these cases, the benefits of the vectorized approach can exceed a factor of 2. Indeed, if we use a vectorized approach to the population count, we can count the bits directly on the vector registers, without additional data transfer.

#### 4.2 | Vectorized intersections between arrays

Because array containers represent integers values as sorted arrays of 16-bit integers, we can put to good use an algorithm based on a vectorized string comparison function present in recent x64 processors<sup>22,23</sup> (SSE 4.1's `pcmpistrm`). While

intended for string comparisons, we can repurpose the instruction to compute intersections. Each input array is divided into blocks of eight 16-bit integers. There are 2 block pointers (one for each array) initialized to the first blocks. We use the `pcompstrm` instruction to conduct an all-against-all comparison of 16-bit integers between these 2 blocks, and we extract matching integers from a mask produced by the instruction. The number of 1s in the resulting mask indicates the number of matching integers. To avoid the cost of comparing pairs of blocks that cannot match, a merge-like process considers the maximum element within each block to find those block pairs to which `pcompstrm` should be applied. That is, whenever the current maximum value in the block of 1 array is smaller than or equal to the maximum of the other block, we load the next block. In particular, when the maxima of the 2 blocks are equal, we load a new block in each array (see Algorithm 1).

---

**Algorithm 1** Generic block-based intersection algorithm.

---

**Require:** two nonempty arrays made of a multiple of  $K$  distinct values in sorted order

- 1: load a block  $B_1$  of  $K$  elements from the first array
  - 2: load a block  $B_2$  of  $K$  elements from the second array
  - 3: **while** true **do**
  - 4:   write the intersection between the two blocks  $B_1 \cap B_2$  to the output
  - 5:   **if**  $\max B_1 = \max B_2$  **then**
  - 6:     load a new block  $B_1$  of  $K$  elements from the first array, or terminate if exhausted
  - 7:     load a new block  $B_2$  of  $K$  elements from the second array, or terminate
  - 8:   **else if**  $\max B_1 < \max B_2$  **true then**
  - 9:     load a new block  $B_1$  of  $K$  elements from the first array, or terminate if exhausted
  - 10: **else if**  $\max B_2 < \max B_1$  **true then**
  - 11:   load a new block  $B_2$  of  $K$  elements from the second array, or terminate if exhausted
- 

We illustrate a simplified function in Figure 6. The `_mm_cmpistrm` intrinsic compares all 16-bit integers from one 128-bit vector with all 16-bit integers from another vector, returning a mask. From this mask, we compute the number of matching values (using the `_mm_popcnt_u32` intrinsic to call the `popcnt` instruction). Moreover, we can shuffle the values in one of the vectors (`vA`) using a shuffling intrinsic (`_mm_shuffle_epi8`) and a table of precomputed shuffling masks (`mask16`). We could design an equivalent AVX2 function<sup>22</sup> using 256-bit vectors, but AVX2 lacks the equivalent of the `_mm_cmpistrm` intrinsic.

Special-case processing is required when our arrays do not have lengths divisible by 8: we need to finish the algorithm with a small scalar intersection algorithm. Special processing is also required if an array starts with zero, because the `pcompstrm` instruction is designed for string values and it processes null values as special (string ending) values.

Evidently, we can easily modify this function if we only wish to get the cardinality. For the most part, we merely skip the memory store instructions.

```
int32_t intersect(uint16_t *A, size_t lengthA,
                 uint16_t *B, size_t lengthB,
                 uint16_t *out) {
    size_t count = 0; // size of intersection
    size_t i = 0, j = 0;
    int vectorlength = sizeof(__m128i) / sizeof(uint16_t);
    __m128i vA, vB, rv, sm16;
    while (i < lengthA) && (j < lengthB) {
        vA = _mm_loadu_si128((__m128i *)&A[i]);
        vB = _mm_loadu_si128((__m128i *)&B[j]);
        rv = _mm_cmpistrm(vB, vA,
                        _SIDD_UWORD_OPS | _SIDD_CMP_EQUAL_ANY | _SIDD_BIT_MASK);
        int r = _mm_extract_epi32(rv, 0);
        sm16 = _mm_load_si128(mask16 + r);
        __m128i p = _mm_shuffle_epi8(vA, sm16);
        _mm_storeu_si128((__m128i *)&(out + count), p);
        count += _mm_popcnt_u32(r);
        uint16_t a_max = A[i + vectorlength - 1];
        uint16_t b_max = B[j + vectorlength - 1];
        if (a_max <= b_max) i += vectorlength;
        if (b_max <= a_max) j += vectorlength;
    }
    return count;
}
```

**FIGURE 6** Optimized intersection function between 2 sorted 16-bit arrays using SSE4.1 intrinsics [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

When the `pcompstrm` instruction is available, as it is on all recent x64 processors, it is advantageous to replace a conventional list intersection function by such a vectorized approach. However, we still rely on a galloping intersection when 1 of the 2 input arrays is much smaller. For example, the vectorized approach is obviously not helpful to intersect an array made of 1000 values with one made of 2 values. There are vectorized versions of the galloping intersection<sup>22</sup> that could further improve our performance in these cases, but we leave such an optimization for future work.

### 4.3 | Vectorized unions between arrays

To compute the union between 2 array containers, we adapt an approach originally developed for merge sort using SIMD instructions.<sup>51</sup> Given 2 sorted vectors, we can generate 2 sorted output vectors that are the result of a “merger” between the 2 input vectors, by using a sorting network<sup>52,53</sup> whose branch-free implementation uses SIMD minimum and maximum instructions. We can compute such a merger with 8 minimum instructions and 8 maximum instructions, given 2 vectors made of eight 16-bit integers (see Figure 7).

We can put this fast merge function to good use by initially loading 1 vector from each sorted array and merging them, so that the small values go into 1 vector ( $B_1$ ) and the large values go into the other ( $B_2$ ). Vector  $B_1$  can be queued for output. We then advance a pointer in 1 of the 2 arrays, choosing the array with a smallest new value, and then load a new vector. We merge this new vector with  $B_2$ . We repeat this process; see Algorithm 2. We terminate with a scalar merge because the arrays’ lengths might not be a multiple of the vector size.

---

**Algorithm 2** Generic block-based union algorithm.

---

**Require:** two nonempty arrays made of a multiple of  $K$  distinct values in sorted order

---

- 1: load a block  $B_1$  of  $K$  elements from the first array
  - 2: load a block  $B_2$  of  $K$  elements from the second array
  - 3: Merge blocks  $B_1$  and  $B_2$  so that the smallest elements are in block  $B_1$  and the largest are in block  $B_2$ . Output the elements from  $B_1$ , after removing the duplicates
  - 4: **while** there is more data in both lists **do**
  - 5:   **if** the next value loaded from the first array would be smaller than the next value in the second array **then**
  - 6:     load a new block  $B_1$  of  $K$  elements from the first array
  - 7:   **else**
  - 8:     load a new block  $B_1$  of  $K$  elements from the second array
  - 9:   Merge blocks  $B_1$  and  $B_2$  so that the smallest elements are in block  $B_1$  and the largest are in block  $B_2$
  - 10:   output the elements from  $B_1$ , after removing the duplicates while taking into account the last value written to output
  - 11: **while** there is more data in one list **do**
  - 12:   load a new block  $B_1$  of  $K$  elements from the remaining array
  - 13:   Merge blocks  $B_1$  and  $B_2$  so that the smallest elements are in block  $B_1$  and the largest are in block  $B_2$ ,
  - 14:   output the elements from  $B_1$ , after removing the duplicates while taking into account the last value written to output
  - 15: Output the elements from  $B_2$ , after removing the duplicates
- 

The queued vector might include duplicate values. To “deduplicate” them, we use a single vectorized comparison between each new vector with a version of itself offset back by 1 element (taking care to handle the final previously stored value). For the mask resulting from this comparison, we can quickly extract the integers that need to be written, using a lookup table and an appropriate permutation (see Figure 8).

Our vectorized approach is implemented to minimize mispredicted branches. We use 128-bit SIMD vectors instead of 256-bit AVX vectors. Larger vectors would further reduce the number of branches, but informal experiments indicate that the gains are modest. A downside of using larger vectors is that more of the processing occurs using scalar code, when the array lengths are not divisible by the vector lengths.

### 4.4 | Vectorized differences between arrays

Unlike union and intersection, the difference operator is an asymmetric operation:  $A \setminus B$  differs from  $B \setminus A$ . We remove the elements of the second array from the first array.

```

void sse_merge(const __m128i *vInput1,
               const __m128i *vInput2, // input 1 & 2
               __m128i *vecMin, __m128i *vecMax) {
    vecTmp = _mm_min_epu16(*vInput1, *vInput2);
    *vecMax = _mm_max_epu16(*vInput1, *vInput2);
    vecTmp = _mm_alignr_epi8(vecTmp, vecTmp, 2);
    *vecMin = _mm_min_epu16(vecTmp, *vecMax);
    *vecMax = _mm_max_epu16(vecTmp, *vecMax);
    vecTmp = _mm_alignr_epi8(*vecMin, *vecMin, 2);
    *vecMin = _mm_min_epu16(vecTmp, *vecMax);
    *vecMax = _mm_max_epu16(vecTmp, *vecMax);
    vecTmp = _mm_alignr_epi8(*vecMin, *vecMin, 2);
    *vecMin = _mm_min_epu16(vecTmp, *vecMax);
    *vecMax = _mm_max_epu16(vecTmp, *vecMax);
    vecTmp = _mm_alignr_epi8(*vecMin, *vecMin, 2);
    *vecMin = _mm_min_epu16(vecTmp, *vecMax);
    *vecMax = _mm_max_epu16(vecTmp, *vecMax);
    vecTmp = _mm_alignr_epi8(*vecMin, *vecMin, 2);
    *vecMin = _mm_min_epu16(vecTmp, *vecMax);
    *vecMax = _mm_max_epu16(vecTmp, *vecMax);
    vecTmp = _mm_alignr_epi8(*vecMin, *vecMin, 2);
    *vecMin = _mm_min_epu16(vecTmp, *vecMax);
    *vecMax = _mm_max_epu16(vecTmp, *vecMax);
    vecTmp = _mm_alignr_epi8(*vecMin, *vecMin, 2);
    *vecMin = _mm_min_epu16(vecTmp, *vecMax);
    *vecMax = _mm_max_epu16(vecTmp, *vecMax);
    *vecMin = _mm_alignr_epi8(*vecMin, *vecMin, 2);
}

```

**FIGURE 7** Fast merge function using Intel single-instruction–multiple-data intrinsics, takes 2 sorted arrays of eight 16-bit integers, and produces 2 vectors (vecMin, vecMax) containing the 16 integer inputs in a sorted sequence, with the 8 smallest integers in vecMin and 8 largest in vecMax [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

```

// returns how many unique values are found in n
// the last integer of "o" is last written value
int store_unique(__m128i o, __m128i n, uint16_t *output) {
    // new vector v starting with last value of "o"
    // and containing the 7 first values of "n"
    __m128i v = _mm_alignr_epi8(n, o, 16 - 2);
    // compare v and n and create a mask
    int M = _mm_movemask_epi8(
        _mm_packs_epi16(_mm_cmpeq_epi16(v, n), _mm_setzero_si128()));
    // "number" represents the number of unique values
    int number = 8 - _mm_popcnt_u32(M);
    // load a "shuffling" mask
    __m128i key = _mm_loadu_si128(uniqshuf + M);
    // the first "number" values of "val" are the unique values
    __m128i val = _mm_shuffle_epi8(n, key);
    // we store the result to "output"
    _mm_storeu_si128((__m128i *)output, val);
    return number;
}

```

**FIGURE 8** Fast function that takes a vector containing sorted values with duplicates and stores the values with the duplicates removed, returning how many values were written. We provide a second vector containing the last value written as its last component [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

If we can compute the intersection quickly between blocks of data (eg, with the `_mm_cmpistrm` intrinsic), we can also quickly obtain the difference. In our vectorized intersection algorithm (Section 4.2), we identify the elements that are part of the intersection using a bitset. In a difference algorithm, picking blocks of values from the first array, we can successively remove the elements that are present in the second array, by iterating through its data in blocks. The intersection, expressed as a bitset, marks the elements for deletion, and we can conveniently accumulate these indications by taking the bitwise OR of the bitsets. Indeed, suppose that we want to compute  $A \setminus B$  where  $A = \{1, 3, 5, 7\}$  and  $B = \{1, 2, 3, 4, 6, 7, 8, 9\}$  using blocks of 4 integers:

- We compute the intersection between the first (and only) block from  $A$  given by  $\{1, 3, 5, 7\}$  and the first block from  $B$  given by  $\{1, 2, 3, 4\}$ . We find that 1,3 are part of the intersection, but not 5,7. We can represent this result with the bitset 0b0011 (or 0x3).



- We advance to the next block in the second array, loading {6, 7, 8, 9}. The intersection between the 2 blocks is  $\{1, 3, 5, 7\} \cap \{6, 7, 8, 9\} = \{7\}$ , which we can represent with the bitset 0b1000.
- We can compute the bitwise OR between the 2 bitsets 0b0011 and 0b1000 to get 0b1011. It indicates that all but the third value in the block {1, 3, 5, 7} must be removed. Hence, we output {3}.

Algorithm 3 describes the difference algorithm at a high level.

As with our other algorithms, we need to finish the computation using scalar code because the inputs may not be divisible by the vector size. Moreover, if we only need to compute the size of the difference, we can use the same algorithm, but we simply do not store the result, only tracking the cardinality.

---

**Algorithm 3** Generic block-based difference algorithm.

---

**Require:** two nonempty arrays made of a multiple of  $K$  distinct values in sorted order

- 1: load a block  $B_1$  of  $K$  elements from the first array
  - 2: load a block  $B_2$  of  $K$  elements from the second array
  - 3: **while** true **do**
  - 4:   mark for deletion the elements in  $B_1 \cap B_2$
  - 5:   **if**  $\max B_1 = \max B_2$  **then**
  - 6:     write  $B_1$  to the output after removing the elements marked for deletion
  - 7:     load a new block  $B_1$  of  $K$  elements from the first array, or terminate if exhausted
  - 8:     load a new block  $B_2$  of  $K$  elements from the second array, or break the loop if exhausted
  - 9:   **else if**  $\max B_1 < \max B_2$  **then**
  - 10:     write  $B_1$  to the output after removing the elements marked for deletion
  - 11:     load a new block  $B_1$  of  $K$  elements from the first array, or terminate if exhausted
  - 12:   **else if**  $\max B_2 < \max B_1$  **then**
  - 13:     load a new block  $B_2$  of  $K$  elements from the second array, or break the loop if exhausted
  - 14: write  $B_1$  after removing the elements marked for deletion as well as all unprocessed data from the first array to the output
- 

## 4.5 | Vectorized symmetric differences between arrays

The symmetric difference can be computed using a block-based algorithm similar to the union algorithm (Algorithm 2). When computing the union, we remove duplicates, keeping just one instance of a value repeated twice. For instance, given {1, 1, 2, 2, 3}, we want to output {1, 2, 3}. When computing the symmetric difference, we need to entirely remove a value if it is repeated. Thus, given {1, 1, 2, 2, 3}, we want to get the set {3}. We can use a fast vectorized function for this purpose (see Figure 9), and it is similar to the corresponding deduplication function used for the vectorized union (Figure 8).

In the union, we can immediately write all values after removing duplicates. With the symmetric difference, the largest value in the current block should not be written immediately. Indeed, we need to account for the fact that the next block might start with this largest value and that it may become necessary to omit it. Thus, we never write the current largest value of the current block when computing the symmetric difference. Instead, it is (potentially) added when the next block is processed—note that `val` in Figure 9 is computed from `v2` (which has this value), rather than from `n`.

```
int store_xor(__m128i o, __m128i n, uint16_t *output) {
    __m128i v1 = _mm_alignr_epi8(n, o, 16 - 4);
    __m128i v2 = _mm_alignr_epi8(n, o, 16 - 2);
    __m128i e1 = _mm_cmpeq_epi16(v1, v2);
    __m128i er = _mm_cmpeq_epi16(v2, n);
    __m128i erl = _mm_or_si128(e1, er);
    int M = _mm_movemask_epi8(
        _mm_packs_epi16(erl, _mm_setzero_si128()));
    int nv = 8 - _mm_popcnt_u32(M);
    __m128i key = _mm_dddqu_si128(uniqshuf + M);
    __m128i val = _mm_shuffle_epi8(v2, key);
    _mm_storeu_si128((__m128i *)output, val);
    return nv;
}
```

**FIGURE 9** Fast function taking a vector containing sorted values with duplicates and entirely removing the duplicated values, returning how many values were written. We provide a second vector containing the last value written as its last component [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

Of course, when the algorithm terminates, we must write the largest value encountered if it were not repeated. In any case, we need to terminate the algorithm with scalar code given that the size of our input lists may not be divisible by the vector size.

## 5 | EXPERIMENTS

We present experiments on realistic data sets. Our benchmarks directly carry out operations on collections of in-memory sets. These original sets are immutable: they are not modified during the benchmarks. For a validation of Roaring in a database system with external memory processing, see, for example, the work of Chambi et al.,<sup>3</sup> where they showed that switching to Roaring from Concise improved system-level benchmarks.

### 5.1 | Software

We use the GNU GCC compiler (version 5.4) on a Linux server. All codes were compiled with full optimization (`-O3 -march=native`) as it provided the best performance.

We published our C code as the CRoaring library; it is available on GitHub under an Apache license<sup>‡</sup>. The CRoaring library is amalgamated into a single source file (`roaring.c`) and a single header file (`roaring.h`).<sup>54</sup> The code is portable, having been tested on several operating systems (Linux, Windows, and macOS), on several processors (ARM and x64), and on several compilers (GNU GCC, LLVM's Clang, Intel, and Microsoft Visual Studio). We offer interfaces to the library in several other languages (eg, C#, C++, Rust, Go, and Python); it is also available as a module of the popular open-source in-memory database Redis.<sup>55</sup> When instantiating the original Roaring bitmaps, we call the functions `roaring_bitmap_run_optimize` and `roaring_bitmap_shrink_to_fit` to minimize memory usage. We use version 0.2.33 of the CRoaring library in our experiments.

To assess the library and compare it with alternatives, we prepared benchmarking software in C and C++, also freely available on GitHub.<sup>§</sup> We consider the following alternatives.

- We use the C++ BitMagic library<sup>¶</sup> (version 3.7.1, published in July 2016). During compilation, we enable the SIMD optimizations offered by the library by defining the appropriate flags (`BMSSE2OPT` and `BMSSE42OPT`). We instantiate bitmaps using the default C++ template (`bm::bvector<>`), since there is no other documented alternative. We follow the documented approach to optimizing memory usage, calling `set_new_blocks_strat(bm::BM_GAP)` followed by `optimize()` after creating the original bitmaps. We do not call `optimize()` during the timed benchmarks as it may affect negatively the timings.

During the optimization procedure, BitMagic allocates a temporary 8kB memory block that fails to get deallocated. This adds an unnecessary 8kB to the memory usage of each bitmap—an undesirable effect. Because this memory allocation is part of the private code of the library, we cannot manually deallocate it without modifying the library. However, as a workaround, we copy each of the original bitmaps after calling `optimize()` using the copy constructor: this new copy does not have such a useless 8kB memory block.

This extra copy to improve BitMagic's memory usage makes the creation of the bitmaps slightly slower, but it does not harm BitMagic's timings in our benchmarks.

- We use the C++ EWAHBoolArray library for comparison with the EWAH format (version 0.5.15).<sup>#</sup> The code from this library is part of the Git version-control tool.<sup>30</sup> After creating the original bitmaps, we call their `trim` method to minimize memory usage. The `trim` method is not called during the timed benchmarks.
- For comparison against uncompressed bitsets, we developed a portable C library, `cbitsset`<sup>||</sup> (version 0.1.6). The library is written using portable C, without processor-specific optimizations. It uses 64-bit words. To represent a set of integers in  $[0, n)$ , it attempts to use as little as  $\lceil n/64 \rceil \times 8$  bytes, plus some overhead.
- For comparison against the WAH and Concise formats,<sup>28</sup> we use the Concise C++ library<sup>\*\*</sup> (version 0.1.8). Since Concise and WAH are similar formats, we can support both of them using the same C++ code without sacrificing

<sup>‡</sup><https://github.com/RoaringBitmap/CRoaring>

<sup>§</sup><https://github.com/RoaringBitmap/CBitmapCompetition>

<sup>¶</sup><https://sourceforge.net/projects/bmagic/>

<sup>#</sup><https://github.com/lemire/EWAHBoolArray>

<sup>||</sup><https://github.com/lemire/cbitset>

<sup>\*\*</sup><https://github.com/lemire/Concise>

performance with a C++ template. After creating the original bitmaps, we call their `compact` method to minimize memory usage, but never during the timed benchmarks.

The EWAH, WAH, and Concise formats can work with words of various sizes (eg, 32 bits and 64 bits). When using 64 bits, we sometimes gain some speed at the expense of a significant increase in memory usage.<sup>19</sup> For simplicity, we only consider 32-bit words, thus maximizing compression.

We also include the standard C++ library bundled with our compiler, which provides `std::vector` and `std::unordered_set`.

- We use `std::unordered_set` as the basis for testing hash-set performance. The `std::unordered_set` template is hardly the sole way to implement a hash set in C++, but since it is part of the standard library, we consider it a good default implementation. To reduce memory usage, we call the `rehash` method on our `std::unordered_set` instances after constructing them, passing the number of items as an argument. Implementing set operations for `std::unordered_set` is straightforward. For example, to compute an intersection between 2 `std::unordered_set`, we pick the smallest, iterate over its content, and check each value in the largest one. Our implementations have good computational complexity. Thus, computing the intersection between 2 hash sets of sizes  $n_1$  and  $n_2$  has complexity  $O(\min(n_1, n_2))$ , the union and symmetric differences have complexity  $O(n_1 + n_2)$ , and the difference has complexity  $O(n_1)$ .
- To minimize the memory usage of our original `std::vector` instances, we call their `shrink_to_fit` method after constructing them, but never during the timed benchmarks. To implement set operations over the `std::vector` data structure, we keep the values sorted and make use of the functions provided by STL (`std::set_intersection`, `std::set_union`, `std::set_difference`, `std::set_symmetric_difference`, ...). These functions take iterators pointing at the inputs and output, so we can use them to, for example, measure the size of the output without necessarily materializing it. To locate a value within an `std::vector`, we use `std::binary_search`, a logarithmic-time function.

Of course, when comparing different software libraries, we expect performance differences due to specific implementation issues. Although this is unavoidable, we alleviate the problem by ensuring that all of our codes and all of the libraries' codes are freely available, thus easing reproducibility and further work.

## 5.2 | Hardware

We use a Linux server with an Intel i7-6700 processor (Skylake microarchitecture, 3.4 GHz, 32 kB of L1 data cache, 256 kB of L2 cache per core, and 8 MB of L3 cache). The server has 32 GB of RAM (DDR4 2133, double-channel). Our benchmarks do not access disk and are single threaded. The processor always runs at its highest clock speed, and Turbo Boost is disabled. Times are collected using the `rdtsc` instruction, as described in Intel's documentation.<sup>56</sup> Prior to benchmarking, we checked that all results could be reproduced reliably within a margin of error of 5% or better.

## 5.3 | Data

We focus on the cases where the integer values span a wide range of values and where sets contain many integers. Thus, we make use of 8 data sets from earlier work.<sup>19,57</sup> Our data sets are derived from bitmap indexes built on real-world tables. Some are taken as is, whereas others are generated from tables that have been lexicographically sorted prior to indexing.<sup>29</sup> We have 2 sets of bitmaps from each data source (eg, CENSUS1881 and CENSUS1881<sup>sort</sup>). For each data set, we randomly chose 200 sets, each of which contains the record ids of rows matching predicate  $A = v$ , for some column  $A$  and column-value  $v$ . We present the basic characteristics of these sets in Table 3. All of the data sets are publicly available as part of our CRoaring software library. The universe size of a data set is the smallest value  $n$  such that all sets are contained in  $[0, n)$ . The density is the cardinality divided by the universe size.

We take our data sets as reasonable examples that can serve to illustrate the benefits (as well as the lack of benefit) of our Roaring optimizations. We stress that no single data structure (bitset, array, hash set, and compressed bitset) can be best for all data sets and all applications. The choice depends on the data characteristics and is nontrivial.<sup>58</sup>

We want to benchmark computational efficiency to evaluate our optimization. Thus, we want to avoid disk access entirely. Hence, our data sets easily fit in memory by design. We further validate our results with a large in-memory data set in Appendix B.

**TABLE 3** Characteristics of our realistic data sets

	Universe Size	Average Cardi- nality Per Set	Average Density
CENSUSINC	199 523	34 610.1	0.17
CENSUSINC <sup>sort</sup>	199 523	30 464.3	0.15
CENSUS1881	4 277 806	5019.3	0.001
CENSUS1881 <sup>sort</sup>	4 277 735	3404.0	0.0008
WEATHER	1 015 367	64 353.1	0.063
WEATHER <sup>sort</sup>	1 015 367	80 540.5	0.079
WIKILEAKS	1 353 179	1376.8	0.001
WIKILEAKS <sup>sort</sup>	1 353 133	1440.1	0.001

In our benchmarks, the original collection of sets is immutable. For example, if we compute the intersection between 2 sets, we generate a new data structure with the result. This reflects a common usage scenario, eg, Druid<sup>3</sup> relies on collections of immutable and persistent compressed bitmaps to accelerate queries.

## 5.4 | Memory usage

Data structures that fail to fit in memory must be retrieved from slower storage media. Thus, everything else being equal, we want our data structures to use as little memory as possible.

Exactly computing the memory usage of complex data structures in C and C++ can be difficult; the same C code running with different libraries or a different operating system might use memory differently. Thus, we need to proceed with care.

For the `std::vector` implementation, we know from first principles that the memory usage (when ignoring a small amount of constant overhead) is 32 bits per entry. For the hash sets (`std::unordered_set`), we use a custom memory allocator that keeps track of allocation requests and record the provided byte counts, assuming, for simplicity, that memory allocations carry no overhead. For continuous storage, this estimate should be accurate, but for a hash set (where there are numerous allocations), we underestimate the true memory usage. Because our custom memory allocation might add an unnecessary computational overhead, we run the timing tests without it.

We proceed similarly with Roaring. Because we implemented Roaring using the C language, we instrument the `malloc/free` calls so that they register the memory allocated and freed. We also run our experiments with and without this instrumentation, to make sure that we can record the speed without undue overhead. The CRoaring library supports a compact and portable serialization format. For most of our data sets, the in-memory size and the compact serialized size are nearly identical. There are only 3 data sets where there were noteworthy differences between in-memory and serialized sizes: CENSUS1881<sup>sort</sup> (2.77 bits per value vs 2.16 bits per value), WIKILEAKS (2.58 bits per value vs 1.63 bits per value), and WIKILEAKS<sup>sort</sup> (7.04 bits per value vs 5.89 bits per value). Not coincidentally, these are our data sets with the fewest values per set. The difference can be explained by the overhead introduced by each new container. It suggests that the memory layout in the CRoaring library could be improved in some cases, by reducing the memory overhead of the containers.

For uncompressed bitsets, EWAH, WAH, and Concise, we rely on the memory usage in bytes reported by the corresponding software libraries. Because they store all their data in a single array, we can be confident that this approach provides a reliable count.

For the BitMagic library, we rely on the numbers provided by the `calc_stat` method, retrieving the `memory_used` attribute. An examination of the source code reveals that it should provide a fair estimate of the memory usage.

We present the estimated memory usage in Table 4.

- The memory usage of the uncompressed bitset approach can be excessive (up to 888 bits per 32-bit value). However, it is less than 32 bits per value in 4 data sets: CENSUSINC, CENSUSINC<sup>sort</sup>, WEATHER, and WEATHER<sup>sort</sup>.
- The chosen hash-set implementation is clearly not economical memory-wise, using 195 bits per 32-bit integer. The underlying data structure is a hash table composed of an array of buckets. Because the maximum load factor is 1.0 by default, we have at least as many buckets as we have values. Each bucket is a singly linked list starting with a 64-bit pointer to the first value node. Each value node uses 128 bits: 64 bits for a pointer to the next value node and 64 bits to store the actual integer value. Although we use 32-bit integers as values, the implementation pads the node size to

**TABLE 4** Memory usage in bits per value. For each data set, the best result is in bold. The hash set is implemented using STL's (`unordered_set`)

	Bitset	Vector	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	5.66	32.0	195	4.46	<b>2.60</b>	3.29	3.37	2.94
CENSUSINC <sup>sort</sup>	6.01	32.0	195	1.89	0.600	0.640	0.650	<b>0.550</b>
CENSUS1881	524	32.0	195	46.3	<b>15.1</b>	33.8	34.3	25.6
CENSUS1881 <sup>sort</sup>	888	32.0	195	16.0	<b>2.16</b>	2.91	2.95	2.48
WEATHER	15.3	32.0	195	8.79	<b>5.38</b>	6.67	6.82	5.88
WEATHER <sup>sort</sup>	11.4	32.0	195	0.960	<b>0.340</b>	0.540	0.540	0.430
WIKILEAKS	796	32.0	195	29.8	<b>5.89</b>	10.8	10.9	10.2
WIKILEAKS <sup>sort</sup>	648	32.0	195	24.0	<b>1.63</b>	2.63	2.67	2.23

128 bits for memory alignment. Thus, at least 192 bits ( $64 + 128$ ) are necessary to store each value, which explains the 195 bits encountered in practice in our data sets.

- The memory usage of EWAH, WAH, and Concise is similar to one another (within 20%). Concise is slightly more economical in most cases. Roaring uses less memory than Concise in all data sets except CENSUSINC<sup>sort</sup>, where Concise uses 10% less memory. In 2 cases (CENSUS1881 and WIKILEAKS), Roaring uses much less memory than Concise (30%-40%).
- In some cases, BitMagic uses a surprising amount of memory (up to 46.3 bits per 32-bit integer). Something similar was observed by Pieterse et al.<sup>59</sup> who wrote that “although it [BitMagic] maintained acceptable performance times, its memory usage is exorbitant.” The BitMagic library does not appear to be optimized for low memory usage.

## 5.5 | Sequential access

It is useful to be able to iterate through all values of a set. All of the data structures under consideration allow for efficient in-order access, except for the hash set. The fastest data structure for this problem is the sorted array. Setting it aside, we compare the other software implementations in Table 5. In this benchmark, we simply go through all values, as a means to compute the total cardinality. WAH and Concise have similar performance, with WAH being marginally faster. EWAH has also similar performance, being slightly faster for some data sets and slightly slower for others. Depending on the data set, the worst performance is either with the bitset or the hash set. Indeed, iterating through a bitset can be slow when it is memory inefficient.

The hash set, BitMagic, EWAH, WAH, and Concise implementation rely on an iterator approach, as is typical in C++. Our benchmark code takes the form of a simple for loop.

```
for(auto j = b.begin(); j != b.end(); ++j)
    count++;
```

The bitset and CRoaring implementations favor an approach that is reminiscent of the functional for-each construct, as implemented by STL's `std::for_each` and Java's `forEach`. They take a function as a parameter, calling the function

**TABLE 5** Time needed to iterate through all values, checking the cardinality in the process. We report the number of CPU cycles used per value

	Bitset	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	7.53	28.5	20.4	<b>5.87</b>	13.1	9.23	9.34
CENSUSINC <sup>sort</sup>	6.79	12.6	6.86	<b>5.33</b>	6.03	7.63	7.85
CENSUS1881	42.7	67.5	70.9	<b>5.16</b>	29.4	29.4	31.7
CENSUS1881 <sup>sort</sup>	41.4	20.4	11.0	<b>6.32</b>	9.10	10.0	10.1
WEATHER	10.0	40.7	31.7	<b>6.43</b>	15.9	12.4	12.8
WEATHER <sup>sort</sup>	6.73	12.9	6.79	<b>5.25</b>	5.87	7.46	7.68
WIKILEAKS	46.2	24.6	12.2	<b>9.61</b>	18.4	16.1	16.6
WIKILEAKS <sup>sort</sup>	33.2	13.1	7.96	<b>5.86</b>	7.95	8.95	9.19



**TABLE 6** Performance comparisons for random-access checks (membership queries). We report the number of CPU cycles used on average per query

	Bitset	Vector	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	<b>3.74</b>	415	49.2	33.6	63.6	3260	19 300	19 100
CENSUSINC <sup>sort</sup>	<b>3.70</b>	547	46.7	49.7	67.4	947	3690	3800
CENSUS1881	<b>2.36</b>	174	38.2	11.6	17.4	8820	26 700	30 000
CENSUS1881 <sup>sort</sup>	<b>2.51</b>	159	36.6	11.0	18.3	561	1810	1880
WEATHER	<b>3.54</b>	564	53.0	27.4	76.8	13 700	65 200	69 000
WEATHER <sup>sort</sup>	<b>3.47</b>	661	48.8	42.0	53.7	2270	7490	7330
WIKILEAKS	<b>3.68</b>	172	45.4	20.9	27.7	1030	2410	2350
WIKILEAKS <sup>sort</sup>	<b>2.79</b>	152	43.2	14.8	20.1	248	531	493

for each element of the set, continuing as long as the function returns true. The expected function signature has an optional void-pointer parameter.

```
typedef bool (*roaring_iterator)(uint32_t value, void *param);
```

In our benchmark, the void pointer is used to pass a variable that gets incremented with each value encountered.

## 5.6 | Membership queries

All of the software libraries under consideration support membership queries: given an integer  $x$ , we can ask whether it belongs to the set. Recall that for the `std::vector` tests, we use the `std::binary_search` function provided by the standard library.

To test the performance of random access, for each data set, we compute the size of the universe ( $[0, n)$ ) and check for the presence of the integers  $\lfloor n/4 \rfloor$ ,  $\lfloor n/2 \rfloor$ , and  $\lfloor 3n/4 \rfloor$  in each set of the data set. We present the results in Table 6.

The bitset consistently has the best performance, using less than a handful of CPU cycles per query. The hash set and BitMagic have the next best random-access performance, followed by Roaring, followed by the arrays (`vector`). The arrays have much worse performance than Roaring, sometimes by an order of magnitude.

Roaring is inferior to BitMagic in these tests even though they implement a similar 2-level data structure. Two design elements can explain this inferior performance.

- Roaring uses a binary search that is used to first locate (if it exists) a corresponding container, before searching the container. BitMagic does not require such a binary search because, given values in the range  $[0, n)$ , it materializes  $\lceil n/2^{16} \rceil$  containers in a flat array, with containers marked as empty or full as needed. Thus, BitMagic can locate the corresponding array directly, without any branching.
- Roaring can store up to 4096 values in array containers, whereas BitMagic never stores more than 1280 values in its array containers. Hence, where Roaring uses large array containers, BitMagic use bitset containers. It is faster to look up the value of a bit within a bitset container than to do a binary search within an array container having more than 1280 values.

Both of these elements also contribute to BitMagic's higher memory usage.

Both the bitset and the hash set have expected constant-time random-access complexity. However, the bitset is an order of magnitude faster in this instance.

EWAH, WAH, and Concise have terrible random-access performance as expected, with EWAH faring slightly better. Indeed, unlike our other data structures, these suffer from a linear-time random-access complexity, which means that the time required to check the value of a bit is proportional to the compressed size of the bitmap. Although Concise usually compresses better than either EWAH or WAH, we see that it can also have a slightly worse random access performance.

## 5.7 | Intersections, unions, differences, and symmetric differences

Recall that our data sets are made of 200 sets each. We consider the 199 intersections, unions, differences, and symmetric differences between successive sets. Our functions leave the inputs unchanged, but they materialize the result as a new

set, and so, the timings include the construction of a new data structure. After each computation, we check the cardinality of the result against a precomputed result, thus helping to prevent the compiler from optimizing the computation (see Table 7).

The relative rankings of the different implementations are similar, whether we consider intersections, unions, differences, or symmetric differences. Intersections run faster, whereas unions and symmetric differences take longer.

What is immediately apparent is that for the CENSUSINC, CENSUSINC<sup>sort</sup>, and WEATHER data sets, an uncompressed bitset is faster. However, Roaring comes in second, having sometimes half the speed. However, Roaring uses 10× less memory than an uncompressed bitset on CENSUSINC<sup>sort</sup>. The performance of an uncompressed bitset is much worse on several other data sets such as CENSUS1881, CENSUS1881<sup>sort</sup>, WIKILEAKS<sup>sort</sup>, and WIKILEAKS.

**TABLE 7** Performance comparisons for intersections, unions, differences, and symmetric differences of integer sets. We report the number of CPU cycles used, divided by the number of input values. For each data set, the best result is in bold

<b>(A) Two-by-two intersections: given 200 sets, we compute 199 intersections and check the cardinality of each result</b>								
	Bitset	Vector	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	<b>0.130</b>	7.84	55.3	0.380	0.220	1.41	1.80	2.15
CENSUSINC <sup>sort</sup>	<b>0.150</b>	4.52	28.5	0.250	0.160	0.280	0.450	0.560
CENSUS1881	19.5	3.28	0.480	1.07	<b>0.090</b>	3.74	14.5	20.9
CENSUS1881 <sup>sort</sup>	33.0	3.96	1.41	0.830	<b>0.140</b>	0.500	1.62	1.94
WEATHER	0.420	7.36	35.5	0.590	<b>0.380</b>	2.64	3.58	4.53
WEATHER <sup>sort</sup>	0.330	4.36	15.1	0.160	<b>0.080</b>	0.190	0.340	0.410
WIKILEAKS	26.4	4.15	14.9	3.06	<b>1.43</b>	2.86	6.10	6.38
WIKILEAKS <sup>sort</sup>	19.6	4.15	26.5	1.92	<b>0.580</b>	0.730	1.28	1.43
<b>(B) Two-by-two unions: given 200 sets, we compute 199 unions and check the cardinality of each result</b>								
	Bitset	Vector	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	<b>0.130</b>	7.69	210	0.350	0.270	2.02	2.11	2.44
CENSUSINC <sup>sort</sup>	<b>0.140</b>	5.45	173	0.270	0.300	0.560	0.650	0.760
CENSUS1881	13.6	6.10	265	3.68	<b>1.13</b>	28.2	29.3	33.1
CENSUS1881 <sup>sort</sup>	23.7	6.18	201	2.47	<b>0.950</b>	2.50	3.03	3.38
WEATHER	<b>0.400</b>	7.70	250	0.600	0.570	4.46	4.59	5.41
WEATHER <sup>sort</sup>	0.310	6.13	181	0.200	<b>0.170</b>	0.480	0.560	0.620
WIKILEAKS	20.6	6.95	229	5.74	<b>3.87</b>	9.23	10.9	11.3
WIKILEAKS <sup>sort</sup>	18.4	7.15	182	4.20	<b>2.16</b>	2.37	2.57	2.63
<b>(C) Two-by-two differences: given 200 sets, we compute 199 differences and check the cardinality of each result</b>								
	Bitset	Vector	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	<b>0.120</b>	8.38	177	0.340	0.280	1.72	1.98	2.34
CENSUSINC <sup>sort</sup>	<b>0.120</b>	5.54	134	0.260	0.240	0.420	0.540	0.650
CENSUS1881	11.1	5.03	236	2.56	<b>0.610</b>	15.8	21.9	26.7
CENSUS1881 <sup>sort</sup>	19.4	5.76	171	2.94	<b>0.570</b>	1.46	2.31	2.59
WEATHER	<b>0.390</b>	8.32	248	0.550	0.440	3.60	4.28	5.10
WEATHER <sup>sort</sup>	0.290	5.78	149	0.210	<b>0.150</b>	0.340	0.450	0.510
WIKILEAKS	18.4	6.05	215	6.00	<b>2.65</b>	5.98	8.25	8.44
WIKILEAKS <sup>sort</sup>	15.2	5.74	161	4.43	<b>1.27</b>	1.59	1.93	1.99
<b>(D) Two-by-two symmetric differences: given 200 sets, we compute 199 symmetric differences and check the cardinality</b>								
	Bitset	Vector	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	<b>0.120</b>	8.79	266	0.350	0.370	1.96	2.04	2.38
CENSUSINC <sup>sort</sup>	<b>0.130</b>	5.99	215	0.280	0.350	0.620	0.660	0.790
CENSUS1881	13.2	6.38	369	3.61	<b>1.13</b>	27.5	28.9	33.3
CENSUS1881 <sup>sort</sup>	22.5	6.41	243	2.56	<b>0.950</b>	2.47	2.97	3.38
WEATHER	<b>0.400</b>	8.72	359	0.590	0.760	4.37	4.50	5.41
WEATHER <sup>sort</sup>	0.310	6.23	237	<b>0.220</b>	<b>0.220</b>	0.510	0.560	0.630
WIKILEAKS	20.0	7.41	301	5.60	<b>3.92</b>	8.96	10.7	11.4
WIKILEAKS <sup>sort</sup>	17.7	6.67	228	4.24	<b>1.95</b>	2.33	2.51	2.66

**TABLE 8** Given 200 sets, we compute a single union and report the number of CPU cycles used, divided by the number of input values

	Bitset	Vector	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	0.090	43.4	98.4	0.160	<b>0.050</b>	0.500	1.83	2.25
CENSUSINC <sup>sort</sup>	<b>0.100</b>	37.2	60.2	0.190	0.110	0.160	0.470	0.590
CENSUS1881	9.85	542	1010	<b>2.18</b>	2.56	276	388	402
CENSUS1881 <sup>sort</sup>	16.2	619	353	4.58	<b>3.26</b>	172	198	225
WEATHER	0.350	94.1	237	0.280	<b>0.160</b>	2.35	4.92	6.02
WEATHER <sup>sort</sup>	0.250	69.1	82.3	0.210	<b>0.040</b>	0.320	0.580	0.690
WIKILEAKS	15.4	515	413	11.9	<b>3.01</b>	328	415	450
WIKILEAKS <sup>sort</sup>	14.5	496	274	8.92	<b>1.94</b>	116	133	149

Excluding the uncompressed bitset, the fastest data structure is generally Roaring. On the CENSUS1881 data set, Roaring is an order of magnitude faster than any other approach at computing intersections, and generally much faster.

In most instances, the hash set offers the worst performance. This may seem surprising as hash sets have good computational complexity. However, hash sets have also poor locality, ie, nearby values are not necessarily located in nearby cache lines.

## 5.8 | Wide unions

In applications, it is not uncommon that we need to compute the union of many sets at once. Thus, we benchmark the computation of the union of all 200 sets within a data set. Although various optimizations of this problem are possible, we proceed sequentially, taking the first set, computing the union with the second set, and so forth. This simple approach uses little memory and provides good performance. In earlier work, this sequential approach was compared to other strategies in some detail.<sup>19</sup>

Although we leave the inputs unchanged, the bitset, BitMagic, and Roaring implementations can do some of the operations in-place. For example, we can copy the first bitset and then compute the bitwise OR operation in-place with the following bitsets. Roaring has a convenience function to compute the union of several bitmaps at once (`roaring_bitmap_or_many`).

We present our result in Table 8. Roaring is generally faster, except for 2 cases.

- On CENSUSINC<sup>sort</sup>, the uncompressed bitset is slightly faster (by about 20%).
- On CENSUS1881, BitMagic is slightly faster than Roaring (by about 10%).

In several instances, Roaring is several times faster than any other alternative.

## 5.9 | Fast counts

Sometimes, we do not wish to materialize the result of an intersection or of a union, we merely want to compute the cardinality of the result as fast as possible. For example, one might want to compute the Jaccard index (Tanimoto similarity) between 2 sets ( $|A \cap B| / |A \cup B|$ ) or the cosine similarity ( $|A \cap B| / \sqrt{|A||B|}$ ).

Given a Roaring data structure, we can compute the total cardinality efficiently. For all bitset and array containers, we can simply look up the cardinality, whereas a simple sum over the length of all runs can determine the cardinality of a run container. The same is true with other data structures such as BitMagic, the `vector`, and so forth.

With this in mind, we can reduce the problem of computing the resulting cardinalities of various operations to the problem of computing the size of the intersection.

- The size of the union is  $|A| + |B| - |A \cap B|$ .
- The size of the difference is  $|A| - |A \cap B|$ .
- The size of the symmetric difference is  $|A| + |B| - 2|A \cap B|$ .

We present the time required to compute the size of the intersection in Table 9A. Compared to Table 7A, where we present the time necessary to compute the full intersection and measure its size, these new timings can be twice as small or better. Unsurprisingly, it is efficient to avoid materializing the intersection.

**TABLE 9** Performance comparisons for computing the sizes of the intersections, unions, differences, and symmetric differences of integer sets. We report the number of CPU cycles used, divided by the number of input values. For each data set, the best result is in bold

<b>(A) Two-by-two intersections: given 200 sets, we compute the sizes of 199 intersections, without materializing the intersections</b>								
	Bitset	Vector	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	<b>0.090</b>	4.85	13.3	0.140	0.130	0.800	0.980	1.25
CENSUSINC <sup>sort</sup>	<b>0.090</b>	2.08	6.76	0.120	0.110	0.190	0.260	0.340
CENSUS1881	6.31	1.45	0.460	0.090	<b>0.080</b>	3.71	9.04	12.1
CENSUS1881 <sup>sort</sup>	11.4	1.70	1.05	0.120	<b>0.060</b>	0.440	1.02	1.27
WEATHER	0.270	4.54	17.5	<b>0.210</b>	0.260	1.64	2.09	2.75
WEATHER <sup>sort</sup>	0.190	1.97	7.19	0.090	<b>0.060</b>	0.150	0.220	0.260
WIKILEAKS	11.7	2.12	12.8	1.57	<b>0.870</b>	2.46	3.85	4.11
WIKILEAKS <sup>sort</sup>	8.71	1.69	10.4	0.510	<b>0.260</b>	0.540	0.760	0.830
<b>(B) Two-by-two unions: given 200 sets, we compute the sizes of 199 unions, without materializing the unions</b>								
	Bitset	Vector	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	<b>0.070</b>	4.62	13.4	0.370	0.130	1.11	0.990	1.32
CENSUSINC <sup>sort</sup>	<b>0.070</b>	1.87	6.76	0.200	0.120	0.290	0.280	0.380
CENSUS1881	7.62	1.27	0.220	1.18	<b>0.100</b>	12.1	11.6	16.8
CENSUS1881 <sup>sort</sup>	12.6	1.57	0.840	0.520	<b>0.140</b>	1.06	1.28	1.65
WEATHER	<b>0.230</b>	4.36	17.6	0.570	0.270	2.50	2.10	2.92
WEATHER <sup>sort</sup>	0.180	1.75	7.20	0.140	<b>0.060</b>	0.230	0.240	0.300
WIKILEAKS	11.8	2.37	10.6	2.22	<b>1.33</b>	4.08	4.46	4.89
WIKILEAKS <sup>sort</sup>	10.5	1.57	9.40	1.16	<b>0.490</b>	1.00	1.02	1.16
<b>(C) Two-by-two differences: given 200 sets, we compute the sizes of 199 differences, without materializing the differences</b>								
	Bitset	Vector	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	<b>0.090</b>	5.21	47.4	0.280	0.130	0.950	1.00	1.26
CENSUSINC <sup>sort</sup>	<b>0.100</b>	2.28	29.5	0.190	0.120	0.230	0.280	0.350
CENSUS1881	8.26	1.63	44.6	1.60	<b>0.080</b>	8.14	10.8	14.5
CENSUS1881 <sup>sort</sup>	14.2	1.99	21.5	3.19	<b>0.090</b>	0.720	1.04	1.28
WEATHER	0.280	5.11	61.4	0.450	<b>0.270</b>	2.05	2.14	2.77
WEATHER <sup>sort</sup>	0.210	2.19	32.3	0.170	<b>0.060</b>	0.190	0.230	0.270
WIKILEAKS	13.5	2.53	40.3	5.27	<b>1.09</b>	3.28	4.15	4.29
WIKILEAKS <sup>sort</sup>	11.0	1.88	25.8	3.64	<b>0.370</b>	0.790	0.950	1.03
<b>(D) Two-by-two symmetric differences: given 200 sets, we compute the sizes of 199 symmetric differences, without materializing the symmetric differences</b>								
	Bitset	Vector	Hash Set	BitMagic	Roaring	EWAH	WAH	Concise
CENSUSINC	<b>0.090</b>	4.80	13.3	0.390	0.130	1.13	0.980	1.30
CENSUSINC <sup>sort</sup>	<b>0.100</b>	2.10	6.77	0.220	0.120	0.310	0.280	0.380
CENSUS1881	9.79	1.46	0.470	1.11	<b>0.080</b>	12.0	11.6	16.9
CENSUS1881 <sup>sort</sup>	16.3	1.78	0.800	0.520	<b>0.130</b>	1.07	1.27	1.66
WEATHER	0.290	4.51	17.6	0.600	<b>0.270</b>	2.52	2.09	2.90
WEATHER <sup>sort</sup>	0.220	1.97	7.19	0.140	<b>0.060</b>	0.240	0.240	0.300
WIKILEAKS	14.8	2.23	11.3	2.21	<b>1.29</b>	4.09	4.47	4.93
WIKILEAKS <sup>sort</sup>	12.8	1.70	8.77	1.14	<b>0.470</b>	0.990	1.02	1.16

The results are otherwise somewhat similar: except for a few data sets, Roaring is best. However, we observe that BitMagic is much closer in performance to Roaring, even surpassing it slightly in one data set, than in Table 7A. As reported in Section 2.2, the BitMagic library includes optimizations specific to the computation of the cardinality.

In Table 9, we also present results for the computation of the sizes of the unions, differences, and symmetric differences of integer sets. We can compare these timings with Table 7, where we compute the complete results.

**TABLE 10** CPU cycles used per input element before and after our single-instruction-multiple-data (SIMD)-based optimizations (Section 4)

		CENSUSINC	WEATHER
2-by-2 intersections	scalar code	0.450	0.990
	SIMD code	0.220	0.380
	scalar/SIMD	2.0	2.6
2-by-2 unions	scalar code	0.410	0.880
	SIMD code	0.270	0.570
	scalar/SIMD	1.5	1.5
2-by-2 difference	scalar code	0.520	1.10
	SIMD code	0.280	0.440
	scalar/SIMD	1.9	2.5
2-by-2 sym. diff.	scalar code	0.540	1.06
	SIMD code	0.370	0.760
	scalar/SIMD	1.5	1.4
2-by-2 intersection counts	scalar code	0.360	0.890
	SIMD code	0.130	0.260
	scalar/SIMD	2.8	3.4
wide union	scalar code	0.060	0.250
	SIMD code	0.050	0.160
	scalar/SIMD	1.2	1.6

## 5.10 | Effect of our optimizations

When compiling scalar C code, the compiler and its standard library use advanced (eg, SIMD) instructions. However, we are interested in the effect of our specific SIMD optimizations (see Sections 3 and 4). Within the CRoaring library, we can deactivate these optimizations at compile time, falling back on portable C code. The compiler still uses advanced instructions, but without our direct assistance. In Table 10, we present a comparison, using 2 data sets where the optimizations were especially helpful. In some cases, we double the speed, or better. In these data sets, the functions to compute the size (or count) of the intersection, union, difference, and symmetric differences were all similarly helped. In other data sets, our optimizations had smaller benefits or no benefit at all (though they did no harm).

We find that these results stress the value of our optimizations for several reasons.

- As already stated, even when our optimizations are disabled, the compiler still relies on advanced (SIMD) instructions.
- We are attempting to optimize already efficient code: without our optimizations, CRoaring is already faster than most alternatives in our benchmarks.
- Our optimizations focus on only some specific cases (see Section 4). The bulk of the code base remains the same when we deactivate our optimizations. For this reason, we should not expect them to help in all or even most benchmarks. For example, none of the operations involving run containers have received specific optimizations.

## 6 | CONCLUSION

No single approach can be best in all cases, but Roaring offers good all-around performance in our tests as was already reported.<sup>17,19</sup> Our advanced SIMD-based optimizations further improved already good speeds in several cases. Other similar data structures (like BitMagic) could make use of our optimizations.

We have identified 2 relative weaknesses of our implementation that should be the subject of future work.

- A limitation of our implementation is the lack of advanced optimizations for all operations between a run container and another container type, including another run container. We have determined, by profiling, that these operations are frequently a significant cost (see Appendix A), so they may offer an avenue for further performance gains. In particular, it may prove interesting to attempt to vectorize operations over run containers.
- When comparing BitMagic and Roaring, 2 relatively similar designs, we found that BitMagic often has better random-access performance (membership queries), eg, by up to about a factor of 3. This can be explained by the



more compact layout used by Roaring. Careful optimizations might be able to improve Roaring's performance on these queries.<sup>60</sup>

Our investigations have focused on recent x64 microarchitectures. Our CRoaring library can run on a wide range of processors. In the future, we might investigate performance issues on other important architectures such as ARM.

Several features of CRoaring will be reviewed in the future. For example, CRoaring supports copy-on-write at the container level, which means that we can copy Roaring bitmaps, in part or in full, quickly. Given that we have determined that memory allocations and data copies can sometimes be a significant cost (see Appendix A), it seems that this might be helpful. However, the copy-on-write capability adds implementation complexity. Moreover, CRoaring is designed to allow memory-file mapping, thus potentially saving extraneous memory allocations when bitmaps are retrieved from external memory. We can potentially combine copy-on-write containers with memory-file mapping.

Bitmaps can do more than just intersections, differences, symmetric differences, and unions; they can support advanced queries such as top- $k$ .<sup>61</sup> Further work should expand on our experimental validation.

The BitMagic library is similar to CRoaring. However, one distinguishing feature of BitMagic is the ability to produce highly compressed serialized outputs for storage on disk. Future work could consider adding this aspect to CRoaring. Moreover, CRoaring could—like BitMagic—use tagged pointers and a customized memory allocator for reduced memory usage and better speed.

One possible application for CRoaring would be as a drop-in replacement for an existing Java library through the Java Native Interface. Future work could assess the potential gains and downsides from such a substitution (see Appendix C).

## ACKNOWLEDGEMENTS

We thank the author of the BitMagic library (A. Kuznetsov) for reviewing our manuscript and helping us tune the BitMagic benchmark. This work was supported by the Natural Sciences and Engineering Research Council of Canada under grant 261437.

## ORCID

Daniel Lemire  <http://orcid.org/0000-0003-3306-6922>

Owen Kaser  <http://orcid.org/0000-0001-6204-5085>

## REFERENCES

1. Graefe G. Database servers tailored to improve energy efficiency. Paper presented at: Proceedings of the 2008 EDBT Workshop on Software Engineering for Tailor-Made Data Management, ACM; 2008; Nantes, France.
2. Grand A. Frame of reference and roaring bitmaps. 2015; <https://www.elastic.co/blog/frame-of-reference-and-roaring-bitmaps>. Last checked [April 2017].
3. Chambi S, Lemire D, Godin R, Boukhalfa K, Allen CR, Yang F. Optimizing druid with roaring bitmaps. Paper presented at: Proceedings of the 20th International Database Engineering & Applications Symposium (IDEAS'16), ACM; 2016; New York, NY.
4. Apache. Kylin: extreme OLAP engine for big data. 2017; <http://kylin.apache.org/>. Last checked [April 2017].
5. Harrington B, Rapoport R. Introducing Atlas: Netflix's primary telemetry platform. 2014; <https://medium.com/netflix-techblog/introducing-atlas-netflixs-primary-telemetry-platform-bd31f4d8ed9a>. Last checked [April 2017].
6. Naga PN. Real-time analytics at massive scale with Pinot. 2017; <https://engineering.linkedin.com/analytics/real-time-analytics-massive-scale-pinot>. Last checked [April 2017].
7. Germond C. Pilosa launches breakthrough open source software to dramatically accelerate data queries. 2017; <https://www.pilosa.com/press/pilosa-launches-breakthrough-open-source-software/>. Last checked [April 2017].
8. Wu C. VSTS now uses roaring bitmaps. Private communication; 2017.
9. Zaharia M, Chowdhury M, Franklin MJ, Shenker S, Stoica I. Spark: cluster computing with working sets. Paper presented at: Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing (HotCloud'10), USENIX Association; 2010; Berkeley, CA.
10. Interlandi M, Shah K, Tetali SD, et al. Titian: data provenance support in spark. Paper presented at: Proceedings of the VLDB Endowment; 2015; Kohala Coast, Hawaii. <https://doi.org/10.14778/2850583.2850595>
11. Wikimedia. Cirrussearch. GitHub Engineering Blog. 2017; <https://www.mediawiki.org/wiki/Help:CirrusSearch>. Last checked [April 2017].
12. Abuzaid F, Bradley JK, Liang FT, et al. Yggdrasil: an optimized system for training deep decision trees at scale. *Adv Neural Inf Process Syst*. 2016;3817-3825.
13. Siddiqui T, Kim A, Lee J, Karahalios K, Parameswaran A. Effortless data exploration with zenvisage: an expressive and interactive visual analytics system. Paper presented at: Proceedings of the VLDB Endowment; 2016; New Delhi, India.

14. van Cranenburgh A, Scha R, Bod R. Data-oriented parsing with discontinuous constituents and function tags. *J Language Modelling*. 2016;4(1):57-111. <https://doi.org/10.15398/jlm.v4i1.100>
15. Fokou G, Jean S, Hadjali A, Baron M. Cooperative techniques for SPARQL query relaxation in RDF databases. Paper presented at: Proceedings of the 12th European Semantic Web Conference on The Semantic Web. Latest Advances and New Domains-Volume 9088, Springer-Verlag; 2015; New York, NY. <https://doi.org/10.1007/978-3-319-18818-815>
16. Krogh B, Jensen CS, Torp K. Efficient in-memory indexing of network-constrained trajectories. Paper presented at: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS'16), ACM; 2016; New York, NY. <https://doi.org/10.1145/2996913.2996972>
17. Wang J, Lin C, Papakonstantinou Y, Swanson S. An experimental study of bitmap compression vs. inverted list compression. Paper presented at: Proceedings of the 2017 ACM SIGMOD International Conference on Management of Data (SIGMOD'17), ACM; 2017; New York, NY.
18. Chambi S, Lemire D, Kaser O, Godin R. Better bitmap performance with roaring bitmaps. *Softw Prac Exper*. 2016;46(5):709-719. <https://doi.org/10.1002/spe.2325>
19. Lemire D, Ssi-Yan-Kai G, Kaser O. Consistently faster and smaller compressed bitmaps with roaring. *Softw Pract Exper*. 2016;46(11):1547-1569.
20. Chambi S, Lemire D, Godin R. Nouveaux modèles d'index bitmap compressés à 64 bits. *Revue des Nouvelles Technologies de l'Information*. 2016;RNTI-B-12:1-16.
21. Grand A. LUCENE-5983: RoaringDocIdSet. 2014; <https://issues.apache.org/jira/browse/LUCENE-5983>. Last checked [April 2017].
22. Lemire D, Boytsov L, Kurz N. SIMD compression and the intersection of sorted integers. *Softw Pract Exper*. 2016;46(6):723-749. <https://doi.org/10.1002/spe.2326>
23. Schlegel B, Willhalm T, Lehner W. Fast sorted-set intersection using SIMD instructions. Paper presented at: Proceedings of the 2nd International Workshop on Accelerating Data Management Systems Using Modern Processor and Storage Architectures (ADMS'11); 2011; Seattle, WA.
24. Muła W, Kurz N, Lemire D. Faster population counts using AVX2 instructions. 2017; <https://arxiv.org/abs/1611.07612>. Last checked [April 2017].
25. Navarro G, Provel E. Fast, small, simple rank/select on bitmaps. Paper presented at: Klasing R, ed. *Experimental Algorithms, Lecture Notes in Computer Science*. Vol. 7276. Berlin, Germany: Springer; 2012:295-306. [https://doi.org/10.1007/978-3-642-30850-5\\_26](https://doi.org/10.1007/978-3-642-30850-5_26)
26. Antoshenkov G. *Byte-Aligned Bitmap Compression*. DCC'95. Washington, DC: IEEE Computer Society; 1995:476.
27. Wu K, Stockinger K, Shoshani A. *Breaking the Curse of Cardinality on Bitmap Indexes*, SSDBM'08. Berlin, Germany: Springer; 2008:348-365.
28. Colantonio A, Di Pietro R. Concise: compressed 'n' composable integer set. *Inf Process Lett*. 2010;110(16):644-650. <https://doi.org/10.1016/j.ipl.2010.05.018>
29. Lemire D, Kaser O, Aouiche K. Sorting improves word-aligned bitmap indexes. *Data Knowl Eng*. 2010;69(1):3-28. <https://doi.org/10.1016/j.datak.2009.08.006>
30. Martí V. Counting objects. GitHub Engineering Blog. 2015; <http://githubengineering.com/counting-objects/>. Last checked [April 2017].
31. Fusco F, Stoecklin MP, Vlachos M. NET-FLI: on-the-fly compression, archiving and indexing of streaming network traffic. *Proc. of the VLDB Endowment*. 2010;3(2):1382-1393. <https://doi.org/10.14778/1920841.1921011>
32. Corrales F, Chiu D, Sawin J. *Variable Length Compression for Bitmap Indices*. DEXA'11. Berlin, Germany: Springer-Verlag; 2011:381-395.
33. Guzun G, Canahuate G, Chiu D, Sawin J. *A Tunable Compression Framework for Bitmap Indices*. ICDE'14. Chicago, IL: IEEE; 2014:484-495.
34. Chang J, Chen Z, Zheng W, et al. SPLWAH: a bitmap index compression scheme for searching in archival internet traffic. Paper presented at: 2015 IEEE International Conference on Communications (ICC); 2015; London, UK. <https://doi.org/10.1109/ICC.2015.7249457>
35. Wu Y, Chen Z, Cao J, et al. CAMP: a new bitmap index for data retrieval in traffic archival. *IEEE Commun Lett*. 2016;20(6):1128-1131.
36. Wu Y, Chen Z, Wen Y, Zheng W, Cao J. COMBAT: a new bitmap index coding algorithm for big data. *Tsinghua Sci Technol*. 2016;21(2):136-145. <https://doi.org/10.1109/TST.2016.7442497>
37. Guzun G, Canahuate G. Hybrid query optimization for hard-to-compress bit-vectors. *VLDB J*. 2016;25(3):339-354. <https://doi.org/10.1007/s00778-015-0419-9>
38. O'Neil E, O'Neil P, Wu K. Bitmap index design choices and their performance implications. Paper presented at: Proceedings of the 11th International Database Engineering and Applications Symposium (IDEAS'07), IEEE Computer Society; 2007; Washington, DC. <https://doi.org/10.1109/IDEAS.2007.19>
39. Rinfret D, O'Neil P, O'Neil E. Bit-sliced index arithmetic. Paper presented at: Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data (SIGMOD'01), ACM; 2001; New York, NY. <https://doi.org/10.1145/375663.375669>
40. Culpepper JS, Moffat A. Efficient set intersection for inverted indexing. *ACM Trans Inf Syst*. 2010;29(1):1:1-1:25. <https://doi.org/10.1145/1877766.1877767>
41. Kuznetsov A. The BitMagic C++ library. 2016; <https://github.com/tlk00/BitMagic>. Last checked [August 2017].
42. Bentley JL, Yao ACC. An almost optimal algorithm for unbounded searching. *Inf Process Lett*. 1976;5(3):82-87.
43. Wikipedia. Bit Manipulation Instruction Sets. 2017; [https://en.wikipedia.org/wiki/Bit\\_Manipulation\\_Instruction\\_Sets](https://en.wikipedia.org/wiki/Bit_Manipulation_Instruction_Sets). Last checked [April 2017].
44. Fog A. Instruction tables: lists of instruction latencies, throughputs and micro-operation breakdowns for Intel, AMD and VIA CPUs. Technical Report. Ballerup, Denmark: Copenhagen University College of Engineering; 2016. [http://www.agner.org/optimize/instruction\\_tables.pdf](http://www.agner.org/optimize/instruction_tables.pdf). Last checked [April 2017].

45. Flynn MJ. Some computer organizations and their effectiveness. *IEEE Trans Comput.* 1972;C-21(9):948-960. <https://doi.org/10.1109/TC.1972.5009071>
46. Koblenz G, Ishizaki K, Inoue H. Bringing Apache Spark closer to SIMD and GPU. 2017; <http://www.spark.tc/simd-and-gpu/>. Last checked [August 2017].
47. Warren HS Jr. The quest for an accelerated population count. In: Wilson G, Oram A, eds. *Beautiful Code: Leading Programmers Explain How They Think*. Sebastopol, CA: O'Reilly Media; 2007:147-160.
48. Knuth DE. *Combinatorial Algorithms, Part 1, The Art of Computer Programming*. Vol. 4A. Boston, MA: Addison-Wesley; 2011.
49. O'Neil P, Quass D. Improved query performance with variant indexes. Paper presented at: Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data (SIGMOD'97); 1997; Tucson, AZ.
50. Intel Corporation. Intel IACA tool: A Static Code Analyser. 2012; <https://software.intel.com/en-us/articles/intel-architecture-code-analyzer>. Last checked [April 2017].
51. Inoue H, Taura K. SIMD- and cache-friendly algorithm for sorting an array of structures. *Proc. of the VLDB Endowment*. 2015;8(11):1274-1285. <https://doi.org/10.14778/2809974.2809988>
52. Batchier KE. Sorting networks and their applications. Paper presented at: Proceedings of the Spring Joint Computer Conference (AFIPS'68), ACM; 1968; New York, NY.
53. Knuth DE. *The Art of Computer Programming, Volume 3: (2nd Ed.) Sorting and Searching*. Redwood City, CA: Addison Wesley Longman Publishing Co, Inc; 1998.
54. Hipp R. The SQLite amalgamation. 2015; <https://www.sqlite.org/amalgamation.html>. Last checked [April 2017].
55. Sanfilippo S. Redis modules. 2017; <https://redis.io/modules>. Last checked [August 2017].
56. Paoloni G. *How to Benchmark Code Execution Times on Intel IA-32 and IA-64 Instruction Set Architectures*. Santa Clara, CA: Intel Corporation; 2010.
57. Lemire D, Kaser O, Gutarra E. Reordering rows for better compression: beyond the lexicographical order. *ACM Trans Database Syst.* 2012;37(3). <https://doi.org/10.1145/2338626.2338633>. Article 20.
58. Damme P, Habich D, Hildebrandt J, Lehner W. Insights into the comparative evaluation of lightweight data compression algorithms. Paper presented at: Proceedings of the 20th International Conference on Extending Database Technology (EDBT'17); 2017; Konstanz, Germany.
59. Pieterse V, Kourie DG, Cleophas L, Watson BW. Performance of C++ bit-vector implementations. Paper presented at: Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT'10), ACM; 2010; New York, NY. <https://doi.org/10.1145/1899503.1899530>
60. Khuong PV, Morin P. Array layouts for comparison-based searching. 2017; <https://arxiv.org/abs/1509.05053>. Last checked [April 2017].
61. Kaser O, Lemire D. Compressed bitmap indexes: beyond unions and intersections. *Softw Pract Exper.* 2016;46(2):167-198. <https://doi.org/10.1002/spe.2289>
62. Anh VN, Moffat A. Index compression using 64-bit words. *Softw Pract Exper.* 2010;40(2):131-147. <https://doi.org/10.1002/spe.v40:2>

**How to cite this article:** Lemire D, Kaser O, Kurz N, et al. Roaring bitmaps: Implementation of an optimized software library. *Softw Pract Exper.* 2018;48:867-895. <https://doi.org/10.1002/spe.2560>

## APPENDIX A: PROFILING OF ROARING BITMAP COMPUTATIONS

As is typical, given a data set and an operation, the software spends most of its time in a few functions spanning few lines of code. Using Linux's `perf` tool, we can identify these hot functions using our test platform and data sets (see Section 5). As part of our publicly available benchmark package, we include a script to automatically run a profiling benchmark. A detailed analysis would be difficult, as there are too many details to take into account. However, we can at least identify the most expensive operation. We also report the percentage of the total time spent in the corresponding operation. This percentage should be viewed as a rough approximation. When the percentage is relatively low (eg, 20%), then it is an indication that there is no clear bottleneck and that many functions contribute to the total running time. Although we collected detailed data on all operations and all data sets, we only present the results for some key operations, for simplicity (see Table A1). The various benchmarks and data sets are presented in our experimental section (Section 5).

In several instances, the same operation is identified as most significant both with regular code and with optimizations (as per Sections 3 and 4), but the percentage is lower with optimizations. This reflects the fact that the optimizations were successful at reducing the running time. In cases where the function does not benefit from optimizations, the percentage is mostly unchanged, as expected. It is the case when run containers are involved, as operations over these containers have not benefited from particular optimizations.

**TABLE A1** Most expensive operations for each benchmark and data set as determined by software profiling

<b>2-By-2 Intersection</b>	<b>Regular Code</b>	<b>With Optimizations</b>
CENSUSINC	array-array (56%)	array-array (19%)
CENSUSINC <sup>sort</sup>	array-run (42%)	array-run (45%)
CENSUS1881	galloping array-array (38%)	galloping array-array (39%)
CENSUS1881 <sup>sort</sup>	high-level (26%)	high-level (26%)
WEATHER	array-array (72%)	array-array (35%)
WEATHER <sup>sort</sup>	array-run (30%)	array-run (31%)
WIKILEAKS	run-run (38%)	run-run (38%)
WIKILEAKS <sup>sort</sup>	run-run (19%)	run-run (20%)
<b>2-By-2 Union</b>	<b>Regular Code</b>	<b>With Optimizations</b>
CENSUSINC	array-array (47%)	array-array (20%)
CENSUSINC <sup>sort</sup>	array-run (44%)	array-run (46%)
CENSUS1881	memcpy (20%)	memcpy (19%)
CENSUS1881 <sup>sort</sup>	memory allocation (20%)	memory allocation (20%)
WEATHER	array-array (49%)	bitset→array conversion (23%)
WEATHER <sup>sort</sup>	array-run (26%)	array-run (25%)
WIKILEAKS	run-run (20%)	run-run (20%)
WIKILEAKS <sup>sort</sup>	memory allocation (15%)	memory allocation (13%)
<b>2-By-2 Diff.</b>	<b>Regular Code</b>	<b>With Optimizations</b>
CENSUSINC	array-array (52%)	bitset→array conversion (18%)
CENSUSINC <sup>sort</sup>	array-run (25%)	array-run (27%)
CENSUS1881	array-array (19%)	memcpy (18%)
CENSUS1881 <sup>sort</sup>	memory allocation (20%)	memory allocation (20%)
WEATHER	array-array (26%)	array-array (25%)
WEATHER <sup>sort</sup>	run-run (22%)	run-run (22%)
WIKILEAKS	run-run (24%)	run-run (22%)
WIKILEAKS <sup>sort</sup>	memory allocation (14%)	memory allocation (16%)
<b>2-By-2 Sym. Diff.</b>	<b>Regular Code</b>	<b>With Optimizations</b>
CENSUSINC	array-array (52%)	bitset-array (24%)
CENSUSINC <sup>sort</sup>	array-array (21%)	run-run (18%)
CENSUS1881	memcpy (21%)	memcpy (21%)
CENSUS1881 <sup>sort</sup>	memory allocation (20%)	memory allocation (20%)
WEATHER	array-array (72%)	array-array (40%)
WEATHER <sup>sort</sup>	run-run (22%)	run-run (21%)
WIKILEAKS	run-run (14%)	run-run (14%)
WIKILEAKS <sup>sort</sup>	memory allocation (15%)	memory allocation (15%)
<b>Single-Union</b>	<b>Regular Code</b>	<b>With Optimizations</b>
CENSUSINC	bitset-array (56%)	bitset-array (52%)
CENSUSINC <sup>sort</sup>	run-bitset (62%)	run-bitset (70%)
CENSUS1881	bitset-array (81%)	bitset-array (36%)
CENSUS1881 <sup>sort</sup>	bitset→array conversion (61%)	bitset→array conversion (63%)
WEATHER	bitset-array (80%)	bitset-array (78%)
WEATHER <sup>sort</sup>	run-bitset (66%)	run-bitset (68%)
WIKILEAKS	run-bitset (64%)	run-bitset (66%)
WIKILEAKS <sup>sort</sup>	run-bitset (25%), bitset→array (25%)	run-bitset (26%), bitset→array (26%)
<b>Intersection-Count</b>	<b>Regular Code</b>	<b>With Optimizations</b>
CENSUSINC	array-array (71%)	array-array (29%)
CENSUSINC <sup>sort</sup>	array-run (55%)	array-run (62%)
CENSUS1881	galloping array-array (54%)	galloping array-array (62%)
CENSUS1881 <sup>sort</sup>	high-level (46%)	high-level (47%)
WEATHER	array-array (81%)	array-array (45%)
WEATHER <sup>sort</sup>	run-run (42%)	run-run (46%)
WIKILEAKS	run-run (66%)	run-run (66%)
WIKILEAKS <sup>sort</sup>	run-run (47%)	run-run (52%)

In some instances, eg, for the union between bitmaps in the CENSUS1881 data set, the memcpy function as well as accompanying memory allocations become a significant cost because we clone many containers. In other cases (eg, CENSUS1881<sup>sort</sup> and WIKILEAKS<sup>sort</sup>), memory allocation is significant.

**TABLE B1** Memory usage in bits per value as well as timings in cycles per value for various operations for the large ClusterData in-memory data set. For each metric, the best reported result is in bold. For the vector, sequential access is practically free

	Bitset	Vector	BitMagic	Roaring	EWAH	WAH	Concise
bits/value	100	32.0	38.5	<b>13.5</b>	33.4	33.7	22.6
seq. access	28.9	—	46.3	<b>6.87</b>	25.5	27.3	27.6
membership	<b>4.92</b>	1910	82.0	381	19 900 000	58 600 000	48 000 000
intersections	25.4	11.2	9.11	<b>2.58</b>	18.0	24.5	27.8
unions	25.4	14.5	9.77	<b>4.97</b>	33.5	34.9	37.5
differences	25.3	14.4	8.05	<b>2.48</b>	26.2	31.4	33.0
sym. differences	25.3	16.6	9.91	<b>4.09</b>	32.6	34.8	37.8
wide unions	5.01	457	3.92	<b>3.87</b>	53.0	68.7	69.5
intersection counts	2.27	8.23	5.16	<b>1.37</b>	14.0	16.2	18.9
union counts	2.14	8.92	8.89	<b>1.35</b>	17.7	16.5	19.3
diff. counts	2.28	9.27	7.24	<b>1.37</b>	16.5	16.8	18.9
sym. diff. counts	2.27	8.29	9.17	<b>1.35</b>	17.8	16.6	19.2

When intersecting bitmaps in the CENSUS1881<sup>sort</sup> data set, much of the time is spent accessing and comparing the high-level 16-bit keys, without access to the containers.

## APPENDIX B: LARGE IN-MEMORY DATA SET

To validate our results on a large data set, we generated 100 arrays of 10 million distinct integers in the range  $[0, 10^9]$  using the ClusterData distribution from Anh and Moffat.<sup>62</sup> This distribution leaves relatively small gaps between successive integers, with occasional large gaps. As with our main experiments, all our source code and scripts are available online. We ran these experiments using the same hardware and software as in Section 5, using a separate script that is included with our benchmarking software.

In total, we generated 1 billion integers; thus, our benchmarks could use gigabytes of memory (RAM). We excluded the hash-set implementation (`std::unordered_set`) from these tests because it failed to complete all of them without exceeding the memory available on our benchmarking machine.

We present our comparison in Table B1. Compared with our main experiments (on smaller data sets), the most obvious difference is that the membership queries are significantly worse for the run-length encoded schemes (EWAH, WAH, and Concise), which is not surprising: that membership queries are resolved by scanning the data from the beginning of the set each time in these cases. To a lesser extent, Roaring also suffers from a worse performance for membership queries when compared to the bitset or to BitMagic.

We compare Roaring without our manual SIMD-based optimizations with the scalar code in Table B2. Our SIMD-based optimizations are worthwhile, sometimes multiplying the performance by a factor of 5.

## APPENDIX C: JAVA VERSUS C

In earlier work, we used Java to benchmark Roaring against several other set data structures, including 32-bit EWAH.<sup>19</sup> There are many differences between Java and C that make comparisons difficult with our current results in C. For example, Java uses a garbage collector whereas C does not. Nevertheless, we can run the SIMD-optimized C and Java versions side by side on the same machine and report the results. We make our dual-language benchmark freely available to ensure reproducibility ([https://github.com/lemire/java\\_vs\\_c\\_bitmap\\_benchmark](https://github.com/lemire/java_vs_c_bitmap_benchmark)). For simplicity, we focus on a single data set (CENSUS1881). On our testing hardware (Section 5.2), we use Oracle's JDK 1.8.0\_101 for Linux (x64), the JavaEWAH library (version 1.0.6), and the RoaringBitmap library (version 0.6.51). We use Oracle's JMH benchmarking system to record the average wall-clock time in Java after many warm-up trials. Our configuration differs in many ways from our earlier work; hence, results are not directly comparable.<sup>19</sup>

Because the Java and C benchmarks are written using entirely different software and because the Java timings rely on an extensive framework (JMH), we cannot be certain that the Java-to-C comparison is entirely fair, especially for small timings. For example, the Java benchmark may not account for the memory allocation in the same manner as



**TABLE B2** CPU cycles used per input element before and after our single-instruction-multiple-data (SIMD)-based optimizations (Section 4)

		ClusterData
2-by-2 intersections	scalar code	8.11
	SIMD code	2.58
	scalar/SIMD	3.1
2-by-2 unions	scalar code	9.34
	SIMD code	4.97
	scalar/SIMD	1.9
2-by-2 difference	scalar code	8.33
	SIMD code	2.48
	scalar/SIMD	3.4
2-by-2 sym. diff.	scalar code	9.19
	SIMD code	4.09
	scalar/SIMD	2.2
2-by-2 intersection counts	scalar code	6.82
	SIMD code	1.37
	scalar/SIMD	5.0
wide union	scalar code	4.79
	SIMD code	3.87
	scalar/SIMD	1.2

**TABLE C1** Performance comparisons for intersections between Java and C/C++ implementations. We report timings in microseconds for executing the whole benchmark (eg. compute all pairwise intersections). A ratio greater than 1 indicates that Roaring is faster

	C/C++			Java			Java/C Ratio	
	EWAH	Roaring	Ratio	EWAH	Roaring	Ratio	EWAH	Roaring
2-by-2 intersections	1900	42	45	2480	43	58	1.3	1.0
2-by-2 unions	12 800	605	21	28 500	750	38	2.2	1.2
wide union	77 600	677	114	107 000	1240	86	1.4	1.8

the C timings. To help improve fairness, we added a few “warm-up” passes in C/C++, just like the warm-up trials generated in Java by JMH: it helped performance slightly. The small beneficial effect is probably related to caching and branch prediction.

We report timings (in microseconds) and give the speed ratio between 32-bit EWAH and Roaring in Table C1. The relative benefits of Roaring over EWAH in these tests are sometimes more important in C/C++ and sometimes more important in Java. The Java code is slower by up to a factor of 2, but the difference between Java and C can also be negligible. The performance differences between Roaring and EWAH in these tests dominate the differences that are due to the programming languages and implementations. That is, the Java version of Roaring is several times faster than the C++ version of EWAH.