

# Fast Succinct Retrieval and Approximate Membership using Ribbon

Peter C. Dillinger  
Facebook  
peterd@fb.com

Lorenz Hübschle-Schneider  
Karlsruhe Institute of Technology  
huebschle@kit.edu

Peter Sanders  
Karlsruhe Institute of Technology  
sanders@kit.edu

Stefan Walzer  
Cologne University  
walzer@cs.uni-koeln.de

July 20, 2021

## Abstract

A *retrieval* data structure for a static function  $f : S \rightarrow \{0, 1\}^r$  supports queries that return  $f(x)$  for any  $x \in S$ . Retrieval data structures can be used to implement a static approximate membership query data structure (AMQ) (i.e., a Bloom filter alternative) with false positive rate  $2^{-r}$ . The information-theoretic lower bound for both tasks is  $r|S|$  bits. While *succinct* theoretical constructions using  $(1 + o(1))r|S|$  bits were known, these could not achieve very small overheads in practice because they have an unfavorable space–time tradeoff hidden in the asymptotic costs or because small overheads would only be reached for physically impossible input sizes. With *bumped ribbon retrieval (BuRR)*, we present the first practical succinct retrieval data structure. In an extensive experimental evaluation BuRR achieves space overheads well below 1% while being faster than most previously used retrieval data structures (typically with space overheads at least an order of magnitude larger) and faster than classical Bloom filters (with space overhead  $\geq 44\%$ ). This efficiency, including favorable constants, stems from a combination of simplicity, word parallelism, and high locality.

We additionally describe *homogeneous ribbon filter AMQs*, which are even simpler and faster at the price of slightly larger space overhead.

# 1 Introduction

A *retrieval data structure* (sometimes called “static function”) represents a function  $f : S \rightarrow \{0, 1\}^r$  for a set  $S \subseteq \mathcal{U}$  of  $n$  keys from a universe  $\mathcal{U}$  and  $r \in \mathbb{N}$ . A query for  $x \in S$  must return  $f(x)$ , but a query for  $x \in \mathcal{U} \setminus S$  may return any value from  $\{0, 1\}^r$ . An analogy that even predates computers is the classical 2-column table with the left column for keys and the right column for associated information. What retrieval data structures achieve is to store “almost” only the right column and still allow finding the correct information for a given key. As a toy example, the left column may contain names of movies and the right column might store ratings of these movies (say 1–4 stars). One can then build a database storing all the movie ratings using close to 2 bits per table entry.

More formally, the information-theoretic minimum amount of space needed by a retrieval data structure is  $nr$  bits in the general case. In this sense a retrieval data structure using  $s$  bits has (*space overhead*)  $\frac{s}{nr} - 1$ . It is by now a classical and widely used result that retrieval data structures can be built with small constant overhead (e.g., about 20%), linear construction time, constant query time, and constant factors that make all this practical, e.g. [10]. Of course an important question is whether the overhead can be brought close to 0 to obtain a *succinct* retrieval data structure using  $(1 + o(1))nr$  bits. Several results in this direction have been obtained (see below) but they all involve a steep space–performance tradeoff. Our key contribution is a simple succinct retrieval data structure where the practical running time penalty for achieving very small overhead ( $< 1\%$ ) is small – in some situations our data structures are even faster than the previously best constructions with 23% overhead.

**Applications.** Retrieval data structures are an important basic tool for building compressed data structures. Perhaps the most widely used application is storing  $r$ -bit fingerprints of the keys which allows implementing an *approximate membership query data structure* (AMQ, aka *Bloom filter replacement*) that supports membership queries with *false positive rate*  $\varphi = 2^{-r}$ . A membership query for key  $x$  will simply compare the fingerprint of  $x$  with the stored value for  $x$ . The values will be the same if  $x$  is in the set. Otherwise, they are the same only with probability  $2^{-r}$ .

Another application uses a result from cuckoo hashing [25, 26, 36], namely that given four hash functions  $h_1, h_2, h_3, h_4 : S \rightarrow [1.024|S|]$  there exists, with high probability, a choice function  $f : S \rightarrow [4]$  such that  $x \mapsto h_{f(x)}(x)$  is injective.<sup>1</sup> A 2-bit retrieval data structure for  $f$  therefore gives rise to a perfect hash function [10], see also [13]. Retrieval data structures can also be used to directly store compact names of objects, e.g., in column-oriented databases [43]. This takes more space than perfect hashing but allows to encode the ordering of the keys into the names.

In several of these applications, retrieval data structures occupy a considerable fraction of RAM in large server farms. Even small reductions (say 10%) in their space consumption thus translate into considerable cost savings. Whether or not these space savings should be pursued at the price of increased access costs depends on the number of queries per second. The lower the access frequency, the more worthwhile it is to occasionally spend increased access costs for a permanently lowered memory budget. Sophisticated implementations use multiple variants of compressed data structures at once based on known access frequencies of different parts of the database [42]. Thus, the entire set of Pareto-optimal variants with respect to the space–access cost tradeoff is relevant for applications.

We remark that once we can do 1-bit retrieval with low overhead, we can use that to store data with prefix-free **variable-bit-length encoding** (e.g. Huffman or Golomb codes). We can store the  $k$ -th bit of element  $x$  as data to be retrieved for the input tuple  $(x, k)$ . This can be further improved by storing  $R$  1-bit retrieval data structures where  $R$  is the largest number of bits needed for representing an input [31, 4, 28]. By interleaving these data structures, one can make queries almost as fast as the case of fixed  $r$ .

---

<sup>1</sup>In this paper,  $[k]$  stands for  $\{0, \dots, k - 1\}$ , and  $a..b$  stands for  $\{a, \dots, b\}$ .

**Linear Algebra Based Retrieval Data Structures.** Due to these important applications, there has been considerable interest in finding retrieval data structures with very small space overhead. A simple, elegant and highly successful basic approach uses linear algebra over the finite field  $\mathbb{Z}_2 = \{0, 1\}$  [16, 27, 1, 45, 13, 10, 18, 20]. Refer to Section 1.4 for a discussion of alternative and complementary techniques. The idea is to store a table  $Z \in \{0, 1\}^{m \times r}$  with  $m \geq n$  entries of  $r$  bits each and to define  $f(x)$  as the bit-wise xor of a set of table entries whose positions are determined by a hash function  $h$ . This can be viewed as the matrix product  $\vec{h}(x)Z$  where  $\vec{h}(x)$  is the characteristic (row)-vector of  $h(x)$ . For given  $h$ , the main task in building the data structure is to find the right table entries such that  $h(x)Z = f(x)$  holds for every key  $x$ . This is equivalent to solving a system of linear equations  $AZ = \mathbf{b}$  where  $A = (\vec{h}(x))_{x \in S} \in \{0, 1\}^{n \times m}$  and  $\mathbf{b} = (f(x))_{x \in S} \in \{0, 1\}^{n \times r}$ . Note that rows in the constraint matrix  $A$  correspond to elements of the input set  $S$ . In the following, we will thus switch between the terms “row” and “element” depending on which one is more natural in the given context. Going back to our view of retrieval as omitting the key column of a table, the train of thought is this: A natural idea would be to have a hash function point to a location where the key’s information is stored while the key itself need not be stored. This fails because of hash collisions. We therefore allow the information for each key to be dispersed over several locations. While this gets rid of the collision problem, it is not obvious that this idea is sound otherwise.

An encouraging observation is that even for  $m = n$ , the system  $AZ = \mathbf{b}$  is solvable with constant probability if the rows of  $A$  are chosen uniformly at random [14, 45]. With linear query time and cubic construction time, we can thus achieve optimal space consumption using the linear algebra approach. For a practically useful approach, however, we want  $A$  to be sparse so that queries can be answered in (near) constant time and we want a (near) linear time algorithm for solving  $AZ = \mathbf{b}$ . This is possible if  $m > n$ .

For  $m > 1.22n$ , the hash function  $h$  can choose random 3-element subsets. In that case  $AZ = \mathbf{b}$  is solvable with high probability. Moreover, a simple greedy algorithm (*hypergraph peeling*) can solve  $AZ = \mathbf{b}$  in linear time [41, 32]. This linear-time peeling approach can be boosted to achieve arbitrarily small overheads by choosing more and more nonzeros per row in appropriate nonuniform ways [37, 50]. Concretely, the *spatial coupling* approach [50] chooses  $k$  random nonzeros in a window of possible positions, achieving space overhead  $\approx e^{-k}$ . However, these approaches face a non-favorable space-performance tradeoff – a query will require  $k$  random memory accesses that usually incur close to  $k$  cache faults for large inputs. Similarly, the peeling algorithm has poor locality. Moreover, the coupling approach has lower-order overhead terms that make it less attractive for small  $n$ .

Higher locality is possible by choosing a larger number of nonzeros in a narrow range. Specifically, the SGAUSS approach [20] chooses  $w$  random bits  $c(x) \in \{0, 1\}^w$  and a random starting position  $s(x) \in [m - w - 1]$ , i.e.,  $\vec{h}(x) = 0^{s(x)-1}c(x)0^{m-s(x)-w+1}$ . For reasons explained later, we call  $w$  the *ribbon width*. For  $m = (1 + \varepsilon)n$  some ribbon width  $w = \mathcal{O}(\log(n)/\varepsilon)$  suffices to make the system  $AZ = \mathbf{b}$  solvable with high probability. After sorting the rows by  $s(x)$ , the solution can be found in time  $\mathcal{O}(n/\varepsilon^2)$  using Gaussian elimination [20] and bit-parallel row operations; see also Figure 1 (a).

**Global Assumptions.** Our model of computation is a word RAM with word size  $\Omega(\log n)$ . Since memory access costs often determine the practical performance of retrieval data structures, we also take the external memory model [47] into account. The analysis assumes that hash functions behave like random functions.<sup>2</sup>

## 1.1 Contribution

We advance the linear algebra approach to the point where space overhead is almost eliminated while keeping or improving the running times of previous constructions.

---

<sup>2</sup>This is a standard assumption in many papers and can also be justified by standard constructions [17].

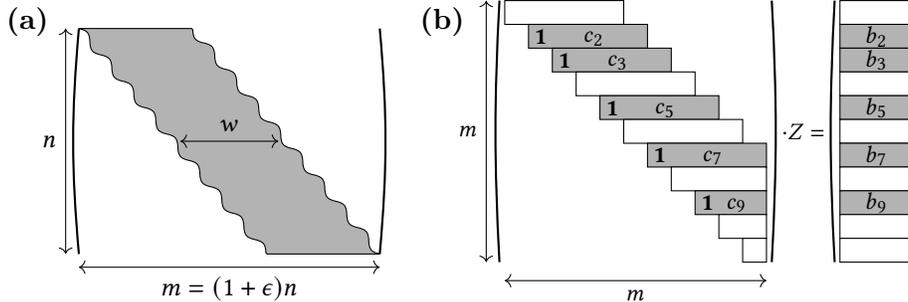


Figure 1: **(a)** Typical shape of the random matrix  $A$  with rows  $(\vec{h}(x))_{x \in S}$  sorted by starting positions. The shaded “ribbon” region contains random bits. Gaussian elimination never causes any fill-in outside of the ribbon.

**(b)** Shape of the linear system  $M$  central to Boolean banding on the fly.

**Ribbon solving.** Our first contribution is a simple algorithm we could not resist to also call *ribbon* as in *Rapid Incremental Boolean Banding ON the fly*. It maintains a system of linear equations in row-echelon form as shown in Figure 1 (b). It does so *on-the-fly*, i.e. while equations arrive one by one in arbitrary order. For each index  $i$  of a column there may be at most one equation that has its leftmost one in column  $i$ . When an equation with row vector  $a$  arrives and its slot is already taken by a row  $a'$ , then ribbon performs the row operation  $a \leftarrow a + a'$  and continues with the modified row. An invariant is that rows have all their nonzeros in a range of size  $w$  and which allows to process rows with a small number of bit-parallel word operations. This insertion process is *incremental* in that insertions do not modify existing rows. This improves performance and allows to cheaply roll back the most recent insertions which will be exploited below. More details are given in Section 2.

**Standard Ribbon.** When employing no further tricks, we obtain *standard ribbon retrieval*, which is essentially the same data structure as in [20] except with a different solver that is faster by noticeable constant factors. A problem is that  $w$  has to become impractically large when  $n$  is large and  $\epsilon$  is small. For example, in our experiments the smallest overhead we could achieve for  $n = 10^6$  and (already quite expensive)  $w = 128$  were around 5.8%. To some degree this can be mitigated by sharding techniques [49], but in this paper we pursue a more ambitious route.

**BuRR.** Our main contribution is *bumped ribbon retrieval (BuRR)* which reduces the required ribbon width to a constant that only depends on the targeted space efficiency. BuRR is based on several crucial observations. First of all, the ribbon solving approach succeeds to insert most rows (representing elements of  $S$ ) even when  $w$  is small. Thus, by eliminating those rows/elements that cause a linear dependency, we obtain a compact retrieval data structure for the non-bumped elements. The bumped elements are handled by a fallback data structure which, by recursion, can be a BuRR data structure again. We show that only  $\mathcal{O}(\frac{n \log w}{w})$  elements are bumped. Thus, after a constant number of layers, a final retrieval data structure can handle the few remaining elements.

This basic *bumped retrieval* approach is adopted from the updateable retrieval data structure *filtered retrieval (FiRe)* [43]. However, even to shrink the input size by a moderate constant factor, FiRe needs a constant number of bits per element (around 4). This leads to very high space overhead for small  $r$ . A crucial observation for BuRR is that there is no need to provide per-element bumping information. Rather, it suffices to bump sets of rows whose starting position  $s(x)$  is in a specified range. The reason is that linear dependencies in  $A$  are largely unrelated to the actual bit patterns  $c(x)$  but mostly caused by fluctuations in the number of elements mapped to different parts of the matrix  $A$ . Thus, by selectively bumping ranges of starting positions in overloaded parts of the system, we can obtain a solvable system. Furthermore, our analysis shows that we can drastically limit the spectrum of possible bumping ranges.

The scheme we have eventually adopted subdivides the possible starting positions into *buckets* of width  $b = \mathcal{O}(w^2/\log w)$  and allows to bump a single initial range of each bucket. In its basic setting this requires  $\mathcal{O}(\log w)$  bits of metadata per bucket or  $\mathcal{O}((\log(w)/w)^2)$  bits per element. A further reduction to  $\mathcal{O}(\log(w)/w^2)$  bits per element is possible by only allowing a constant number of possible range sizes. Our analysis is for the case that range sizes come from  $\{0, t = 3w/8, b\}$ . Since, moreover, the range size  $b$  (bump entire bucket) is only rarely needed, we can encode this in such a way that only slightly more than one bit of metadata per bucket suffices.

Besides metadata, space overhead results from the  $m - n + n_b$  excess slots of the table where  $n_b$  is the number of bumped elements. Trying out possible values of  $\varepsilon = \frac{m-n}{n} > 0$  one sees that the overhead due to excess elements is always  $\Omega(1/w)$  and will thus dominate the overhead due to metadata. However, we show that by choosing  $\varepsilon < 0$ , i.e., by *overloading* the table, we can almost completely eliminate excess table entries so that the minuscule amount of metadata becomes the dominant remaining overhead. Thus, overloading is the final crucial ingredient of BuRR and leads to the following result:

**Theorem 1.** *On a word RAM with word size  $\Omega(\log n)$ , an  $r$ -bit BuRR data structure with ribbon width  $w = \mathcal{O}(\log n)$  and  $r = \mathcal{O}(w)$  has expected construction time of  $\mathcal{O}(nw)$ , space overhead  $\mathcal{O}(\frac{\log w}{rw^2})$ , and query time  $\mathcal{O}(1 + \frac{rw}{\log n})$ .*

Note that this implies constant query time if  $rw = \mathcal{O}(\log n)$  and linear construction time if  $w \in \mathcal{O}(1)$ . For wider ribbons, construction time is slightly superlinear. However, in practice this does not necessarily mean that BuRR is slower than other approaches with asymptotically better bounds as the factor  $w$  involves operations with very high locality. An analysis in the external memory model reveals that BuRR construction is possible with a single scan of the input and integer sorting of  $n$  objects of size  $\mathcal{O}(\log n)$  bits, see Section 7.3. Many other approaches have a sorting volume that grows as the output size approaches the information-theoretic lower bound or construction can be externalized only indirectly via sharding.

**Homogeneous Ribbon Filters.** For the application of Ribbon to AMQs, we can also compute a *uniformly random* solution of the *homogeneous* equation system  $AZ = 0$ , i.e., we compute a retrieval data structure that will retrieve  $0^r$  for all elements of  $S$  but is unlikely to produce zero for other inputs. Since  $AZ = 0$  is always solvable, there is no need for bumping. The crux is that the false positive rate is no longer  $2^{-r}$  but higher. In Section 5 we show that with table size  $m = (1 + \varepsilon)n$  and  $\varepsilon = \Omega(\frac{\max(r, \log w)}{w})$  the difference is negligible. Homogeneous Ribbon filters are simpler and faster than BuRR but have higher space overhead. Our experiments indicate that together, BuRR and homogeneous ribbon filters cover a large part of the best tradeoffs for AMQs.

**Analysis outline.** To get an intuition for the relevant linear systems, it is useful to consider two simplifications. First, assume that  $\vec{h}(x)$  contains a block of uniformly random *real* numbers from  $[0, 1]$  rather than random bits. Secondly, assume that we sort the rows by starting position and use Gaussian elimination rather than ribbon to produce a row echelon form. In Figure 2 (a) we show such a matrix and illustrate with  $\times$ -marks where the pivots would be placed and in yellow the entries that are eliminated (with one row operation each); both with probability 1, i.e. barring coincidences where a row operation eliminates more than one entry. The  $\times$ -marks trace a diagonal through the matrix except that the green column and the red row are skipped because the end of the (gray) area of non-zeroes is reached. “Column failures” correspond to unused space and “row failures” correspond to failed insertions. This view remains largely intact when handling *Boolean* equations in *arbitrary* order except that the *ribbon diagonal* which we introduce as an analogue to the trace of pivot positions has a more abstract meaning and probabilistically suffers from row and column failures depending on its *distance* to the ribbon border.

The idea of standard ribbon is to give the gray ribbon area an expected slope of less than 1 such row failures are unlikely. BuRR, as illustrated in Figure 2 (b) largely avoids both failure types by using a slope bigger than 1 but removing ranges of rows in strategic positions. Homogeneous ribbon filters,

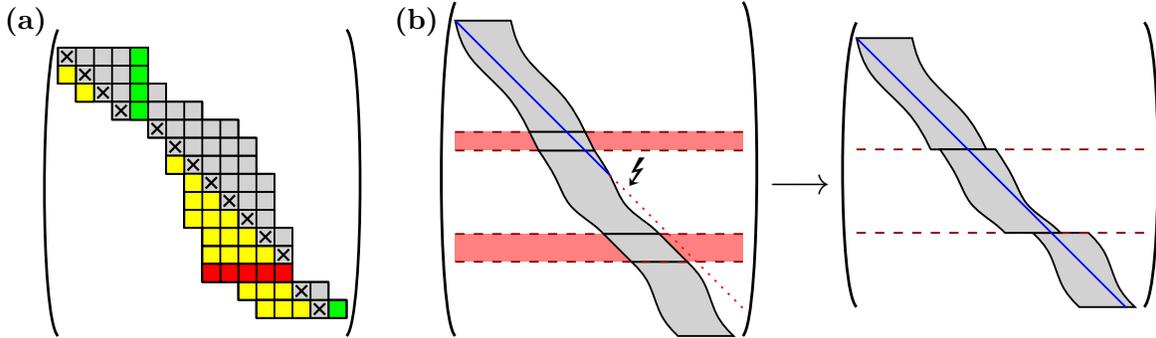


Figure 2: **(a)** The simplified ribbon diagonal (made up of  $\times$ -marks) passing through  $A$ . **(b)** The idea of BuRR: When starting with an “overloaded” linear system and removing sets of rows strategically, we can often ensure that the ribbon diagonal does not collide with the ribbon border (except possibly in the beginning and the end).

despite being the simplest approach, have the most subtle analysis as both failure types are allowed to occur. While row failures cannot cause outright construction failure, they are linked to a compromised false positive rate in a non-trivial way.

Our proofs involve mostly simple techniques as would be used in the analysis of linear probing, which is unsurprising given that [20] has already established a connection to Robin Hood hashing. We also profit from queueing theory via results we import from [20].

**Implementation.** There is a large design space of possible implementations of BuRR which we describe in Section 7. Here, we outline a simple variant that achieves very good performance in practice and is a generalization of the theoretically analyzed approach. We first scan  $S$  and generate a sequence of pairs  $\mathbf{S} = \langle (\text{MHC}(x), f(x)) : x \in S \rangle$ , where  $\text{MHC}(x)$  is an  $\mathcal{O}(\log n)$  bit *master hash code* uniquely identifying element  $x$ . All further hash values are subsequently derived from the MHCs. To build a layer of the BuRR data structure,  $\mathbf{S}$  is then sorted according to the bucket addressed by the starting position  $s(x)$  of the matrix rows represented by  $\mathbf{S}$ . We use a fast in-place integer sorter for this purpose [2]. Then buckets are processed one after the other from *left to right*. Within a bucket, elements are inserted into the row-echelon form from *right to left*; see also Figure 3. If all elements of a bucket were successfully inserted, no elements of the bucket are bumped. Otherwise, suppose the first failed insertion occurs at the  $k$ -th column of the bucket and the next-largest bumping threshold representable by metadata is  $t \geq k$ . Then we roll back the insertion of all elements  $x$  in the current bucket with starting column  $\leq t$  and store  $t$  as a bumping threshold for the current bucket. The reason for building the buckets backwards is that this reveals the appropriate bumping decision for this bucket without need for an explicit balancing algorithm for finding bumping thresholds. We profit from the fact that the total number of performed row operations does not depend on the insertion order (see Section 3). Thus, even though some insertions might need close to  $b = \Theta(w^2/\log w)$  row operations, the overall cost per element remains  $\mathcal{O}(w)$ .

Our implementation supports several variants for metadata. One good compromise between space and speed stores 2 bits per bucket encoding the threshold values  $\{0, \ell, u, b\}$ . The special case  $\ell = u$  is used in our analysis. Another slightly more compact variant (“1<sup>+</sup>-bit”) stores one bit encoding threshold values from the set  $\{0, t\}$  and additionally stores a hash table of exceptions for thresholds  $> t$ .

When all buckets are processed, we perform back-substitution for the row-echelon form representing the non-bumped elements to compute the resulting table  $Z$  for the current layer. At least for small  $r$ , *interleaved representation* of  $Z$  works well, where blocks of size  $w \times r$  of  $Z$  are stored column-wise. A query for  $x$  can then retrieve one bit of  $f(x)$  at a time by applying a population count instruction to pieces of rows retrieved from at most two of these blocks. This is particularly advantageous for negative

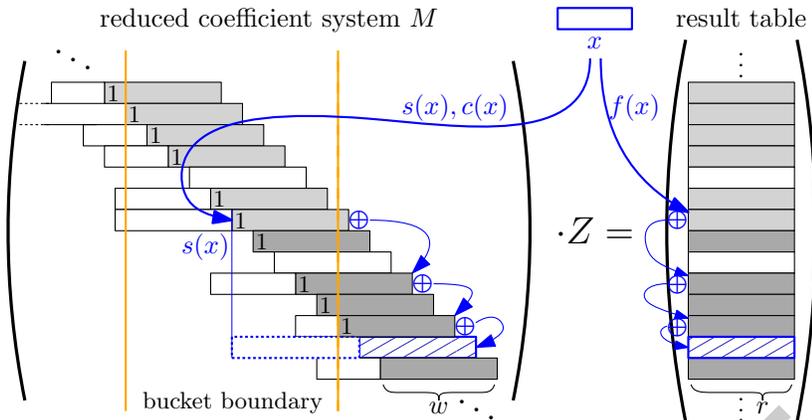


Figure 3: Insertion of a key  $x$  into the system. Already filled rows are marked light gray for keys belonging to the previous bucket, and dark gray for those inserted so far from the current bucket. White boxes preceding stored coefficients mark coefficients that were reduced to 0 and show the keys' starting positions. The reduced coefficients and function value of  $x$  are eventually stored in the blue-shaded slot. Two empty slots remain in the bucket.

queries to AMQs, where only two bits need to be retrieved on average.

The MHCs of bumped elements are collected as the input for the next layer. At a choosable layer (4 in analysis and experiments), the remaining elements are stored in a retrieval data structure which does not require bumping.

**Further results.** We have several further results around variants of BuRR that we summarize here.

Perhaps most interesting is **bump-once ribbon retrieval (Bu<sup>1</sup>RR)** that improves the worst-case query time by guaranteeing that each element can be retrieved from one out of two layers – its *primary layer* or the next one. The primary layer of the elements is now distributed over all layers (except for the last). When building a layer, the elements bumped from the previous layer are inserted into the row-echelon form first. The layer sizes have to be chosen in such a way that no bumping is needed for these elements with high probability. Only then the elements with the current layer as their primary layer are inserted – now allowing bumping. See Section 7.4 for details.

For building large retrieval data structures, **parallel construction** is important. Doing this directly is difficult for ribbon retrieval since there is no efficient way to parallelize back-substitutions. However, we can partition the equation system into parts that can be solved independently by bumping  $w$  consecutive columns. Note that this can be done in a way transparent to the query algorithm by using the bumping mechanism that is present anyway. See Section 7.3 for details.

For large  $r$ , we accelerate queries by working with **sparse bit patterns** that set only a small fraction of the  $w$  bits in the window used for BuRR. In some sense, we are covering here the middle ground between Ribbon and spatial coupling [50]. Experiments indicate that setting 8 out of 64 bits indeed speeds up queries for  $r \in \{8, 16\}$  at the price of increased (but still small) overhead. Analysis and further exploration of this middle ground may be an interesting area for future work.

## 1.2 Experimental Evaluation

We performed extensive experiments to evaluate our ribbon-based data structured and competitors. We summarize our findings here with details provided in Section 8.

**Ribbon is fastest for overhead < 44%.** Two preliminary remarks are in order: Firstly, since every retrieval data structure can be used as a filter but not vice versa, our experiments are for filters, which admits a larger number of competitors. Secondly, to reduce complexity (for now), our speed ranking combines construction time per element and query costs (for each element, one negative query, one positive query, and one with a 50% chance of being positive).

In Figure 4 we show the tradeoff between space overhead and computation cost for a range of

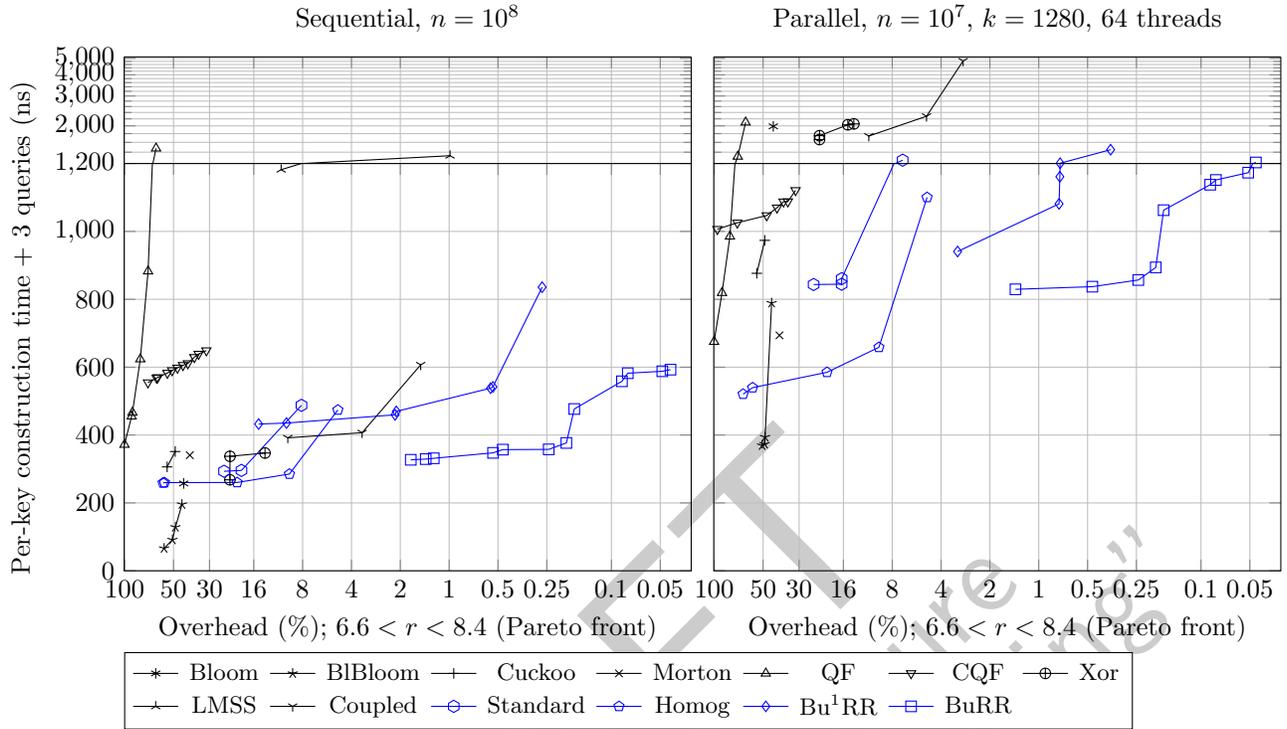


Figure 4: Performance–overhead trade-off for measured false-positive rate in 0.003–0.01 (i.e.,  $r \approx 8$ ), for different AMQs and inputs. Ribbon-based data structures are in blue. For each category of approaches, only points are shown that are not dominated by variants in the same category. Sequential benchmarks use a single filter of size  $n$  while the parallel benchmark uses 1280 filters of size  $n$  and utilizes 64 cores. Logarithmic vertical axis above 1200 ns.

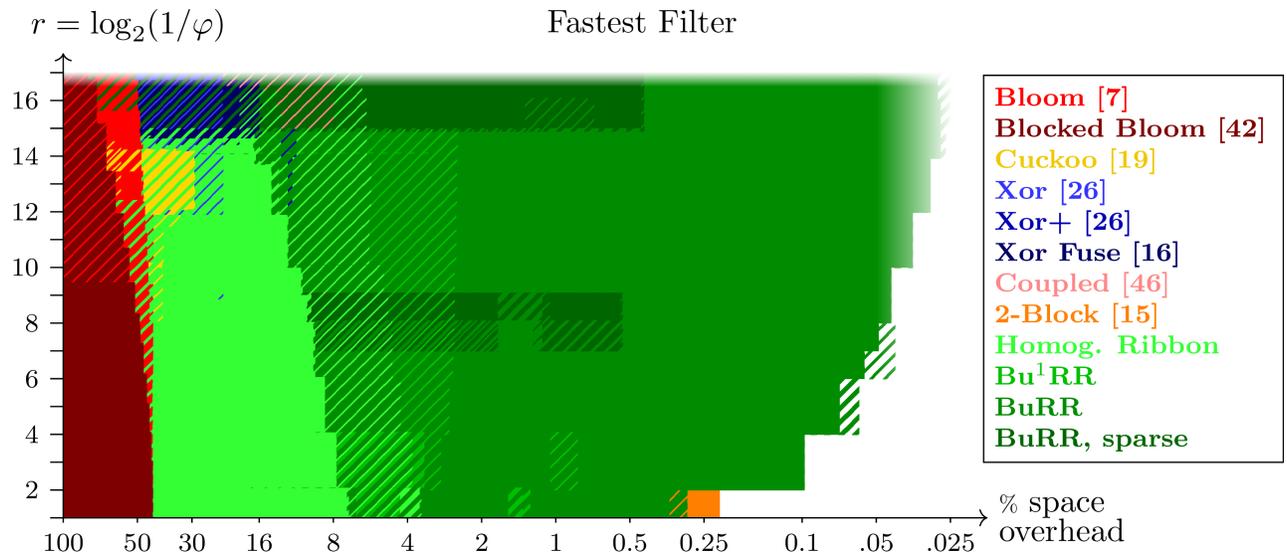


Figure 5: Fastest AMQ category for different choices of overhead and false-positive rate  $\varphi = 2^{-r}$ . Shaded regions indicates a dependency on the input type. Ranking metric: construction time per key plus time for three queries, of which one is positive, one negative, and one mixed (50% chance of either).

AMQs for false positive rate  $\varphi \approx 2^{-8}$  (i.e.,  $r = 8$  for BuRR) and large inputs.<sup>3</sup> Of most interest are *Pareto-optimal* configurations, i.e., those not dominated by other configurations with respect to both space and time. In a parallel workload on the right of Figure 4 all cores access many AMQs randomly.

Only three AMQs have Pareto-optimal configurations for this case: BuRR for space overhead below 5% (actually achieving between 1.4% and 0.2% for a narrow time range of 830–890 ns), homogeneous ribbon for space overhead below 44% (actually achieving between 20% and 10% for a narrow time range 580–660 ns), and *blocked Bloom filters* [46] with time around 400 ns at the price of space overhead of around 50%. All other tried AMQs are dominated by homogeneous ribbon and BuRR. Somewhat surprisingly, this even includes plain Bloom filters [7] which are slow because they incur several cache faults for each insertion and positive query. Since plain Bloom filters are extensively used in practice (often in cases where a static interface suffices), we conclude that homogeneous ribbon and BuRR are fast enough for a wide range of applications, opening the way for substantial space savings in those settings. BuRR is at least twice as fast as all tried retrieval data structures.<sup>4</sup> The filter data structures that support counting and deletion (Cuckoo filters [24] and the related Morton filters [11] as well as the quotient filters QF [39] and CQF [5]) are slower than the best static AMQs.

The situation changes slightly when going to a sequential workload with large inputs as shown on the left of Figure 4.<sup>5</sup> Blocked Bloom and BuRR are still the best filters for large and small overhead, respectively. But now homogeneous ribbon and (variants of) the hypergraph peeling based Xor filters [30, 19] share the middle-ground of the Pareto curve between them. Also, plain Bloom filters are almost dominated by Xor filters with half the overhead. The reason is that modern CPUs can handle several main memory accesses in parallel. This is very helpful for Bloom and Xor, whose queries do little else than computing the logical (x)or of a small number of randomly chosen memory cells. Nevertheless, the faster variants of BuRR are only moderately slower than Bloom and Xor filters while having at least an order of magnitude smaller overheads.

**Further Results.** Other claims supported by our data are:

- **Good ribbon widths are  $w = 32$  and  $w = 64$ .** Ribbon widths as small as  $w = 16$  can achieve small overhead but at least on 64-bit processors,  $w \in \{32, 64\}$  seems most sensible. The case  $w = 32$  is only 15–20% faster than  $w = 64$  while the latter has about four times less overhead. Thus the case  $w = 64$  seems the most favorable one. This confirms that the linear dependence of the construction time on  $w$  is to some extent hidden behind the cache faults which are similar for both values of  $w$  (this is in line with our analysis in the external memory model).
- **Bu<sup>1</sup>RR is slower than BuRR** by about 20%, which may be a reasonable price for better worst-case query time in some real-time applications.<sup>6</sup>
- **The 1<sup>+</sup>-bit variant of BuRR is smaller but slower** than the variant with 2-bit metadata per bucket, as expected, though not by a large margin.
- **Smaller inputs and smaller  $r$  change little.** For inputs that fit into cache, the Pareto curve is still dominated by blocked Bloom, homogeneous ribbon, and BuRR, but the performance penalty for achieving low overhead increases. For  $r = 1$  we have data for additional competitors. GOV [28],

<sup>3</sup>Small deviations of parameters are necessary because not all filters support arbitrary parameter choices. Also note that different filters have different functionality: (Blocked) Bloom allows dynamic insertion, Cuckoo, Morton and Quotient additionally allow deletion and counting. Xor [10, 19, 30], Coupled [50], LMSS [37] and all Ribbon variants are static retrieval data structures.

<sup>4</sup>FiRe [43] is likely to be faster but has two orders of magnitude higher overhead; see Section 8 for more details.

<sup>5</sup>Arguably, on modern machines a *sequential* workload that leaves most of the area of a processor idle is highly inefficient and thus unrealistic. However, this benchmark approximates a parallel workload that has high locality and thus puts little load on the memory subsystem. In other words, Figure 4 shows two extremes in a spectrum of conceivable workloads.

<sup>6</sup>Part of the performance difference might be due to implementation details; see Section 7.4.

which relies on structured Gaussian elimination, is several times slower than BuRR and exhibits an unfavorable time-overhead tradeoff. 2-block [18] uses two small dense blocks of nonzeros and can achieve very small overhead at the cost of prohibitively expensive construction.

- **For large  $r$ , Xor filters and Cuckoo filters come into play.** Figure 5 shows the fastest AMQ depending on overhead and false positive rate  $\varphi = 2^{-r}$  up to  $r = 16$ . While blocked Bloom, homogeneous ribbon, and BuRR cover most of the area, they lose ground for large  $r$  because their running time depends on  $r$ . Here Xor filters and Cuckoo filters make an appearance.
- **Bloom filters and Ribbon filters are fast for *negative queries*** where, on average, only two bits need to be retrieved to prove that an element is not in the set. This improves the relative standing of plain Bloom filters on large and parallel workloads with mostly negative queries.
- **Xor filters [30] and Coupled [50] have fast queries** since they can exploit parallelism in memory accesses. They suffer, however, from slow construction on large sequential inputs due to poor locality, and exhibit poor query performance when accessed from many threads in parallel. For small  $n$ , large  $r$ , and overhead between 8% and 20% Coupled becomes the fastest AMQ.

### 1.3 Related Work on Retrieval – Baustelle!

It has long been known that some matrices with random entries are likely to have full rank, even when sparse [14] and density thresholds for random  $k$ -XORSAT formulas to be solvable – either at all or with a linear time peeling algorithm – have been determined [41, 23, 15].

Building on such knowledge, the retrieval problem and its first solution was identified by Botelho, Pagh and Ziviani [9, 8, 10] in the context of perfect hashing. In our terminology, their rows  $\vec{h}(x)$  contain 3 random 1-entries per key which makes  $AZ = b$  solvable with peeling, provided  $m > 1.22n$ .

Several works develop the idea from [10]. In [27, 28] only  $m > 1.089n$  is needed in principle (or  $m > 1.0238n$  for  $|\vec{h}(x)| = 4$ ) but a Gaussian solver has to be used, or more recently [50] where  $\vec{h}(x)$  has  $k$  ones with suitably correlated positions and  $m \approx (1 + e^{-k})n$  is used while still allowing a peeling solver. With some squinting, a class of linear erasure correcting codes from [37] can be interpreted as a retrieval data structure of a similar vein, where  $|\vec{h}(x)| \in \{5, \dots, k\}$  is random with expectation  $\mathcal{O}(\log k)$ .

Two recent approaches also based on sparse matrix solving are [18, 20] where  $\vec{h}(x)$  contains two blocks or one block of random bits. Our ribbon approach builds on the latter.

A very different idea originated in [45], before retrieval was invented and intended as a Bloom filter replacement, relying heavily on lookup tables.

Generalizations of the retrieval problem where  $f : S \rightarrow \{0, 1\}^{\leq r}$  associates keys with information of varying length or where  $f$  is assumed to have low entropy have been considered and reduced, with good efficiency, to the standard case [31, 4, 28].

**Comparison to Ribbon.** In Table 1 we list asymptotic performance guarantees of ribbon and other approaches. It is not hard to see that our approach opens up new trade-offs between running times and overhead, for instance when  $r$  is constant and constant query time is desired.

Realistically, however, once a small constant overhead of, say,  $\varepsilon = 0.05$  is attained, additional space savings have limited relevance and for constant  $\varepsilon$  several approaches are tied with BuRR. A meaningful comparison cannot neglect constant factors, cache efficiency, branch complexity, etc, which the theoretical analyses summarised in this table systematically neglect. We believe that the case for ribbon is best made by our extensive experiments in Section 8.

We end this section with a discussion of two techniques that look good on paper and give reasons why we choose not to use them.

Table 1: Performance of various approaches for retrieval. Running times are for a word RAM with word size  $\Theta(\log n)$ . The parameters  $k \in \mathbb{N}$  and  $\alpha, \varepsilon > 0$  are constants with respect to  $n$ , and  $r, w = \mathcal{O}(\log n)$ .

	Year	$t_{\text{construct}}$	$t_{\text{query}}$	multiplicative overhead	shard size	Solver
	[37]	2001	$\mathcal{O}(n \log k)$	$\mathcal{O}(\log k)^\dagger$	$\frac{1}{k}$	peeling
	[45]	2009	$\mathcal{O}(n)$	$\mathcal{O}(1)$	$\mathcal{O}(\frac{\log \log n}{(\log n)^{1/2}})$	$\sqrt{\log n}$ lookup table
	[10]	2013	$\mathcal{O}(n)$	$\mathcal{O}(1)$	0.2218	– peeling
Standard Ribbon →	[43]	2014	$\mathcal{O}(n)$	$\mathcal{O}(1)$	$\Omega(1/r)$	$\mathcal{O}(1)$ sorting/sharding
	[27]	2016	$\mathcal{O}(n^{1+2\alpha})$	$\mathcal{O}(1)$	$0.024 + \mathcal{O}(\frac{\log n}{n^\alpha})$	$n^\alpha$ structured Gauss
	[20]	2019	$\mathcal{O}(n/\varepsilon^2)$	$\mathcal{O}(r/\varepsilon)$	$\varepsilon$	– Gauss
	[20]	2019	$\mathcal{O}(n/\varepsilon)$	$\mathcal{O}(r)$	$\varepsilon + \mathcal{O}(\frac{\log n}{n^\varepsilon})$	$n^\varepsilon$ Gauss
	[18]	2019	$\mathcal{O}(n^{1+2\alpha})$	$\mathcal{O}(r)$	$\mathcal{O}(\frac{\log n}{n^\alpha})$	$n^\alpha$ structured Gauss
	[50]	2021	$\mathcal{O}(nk)$	$\mathcal{O}(k)$	$(1 + o_k(1))e^{-k}$	– peeling
		BuRR	$\mathcal{O}(nw)$	$\mathcal{O}(1 + \frac{rw}{\log n})$	$\mathcal{O}(\frac{\log w}{rw^2})$	–
	( $\leftrightarrow$ with $w = \Theta(\log n)$ ):	$\mathcal{O}(n \log n)$	$\mathcal{O}(r)$	$\mathcal{O}(\frac{\log \log n}{r \log^2 n})$	–	on-the-fly Gauss

† Expected query time. Worst case query time is  $\mathcal{O}(D)$ .

## 1.4 More Related Work

We now discuss related and alternative approaches to achieve fast succinct retrieval data structures. Some more details on methods used in the experiments are also discussed in Section 8.

**Shards.** A widely used technique in hashing-based data structures is to use a splitting hash function to first divide the input set into many much smaller sets (shards, buckets, chunks, bins, ...) that can then be handled separately [28, 3, 18, 20, 1, 45]. This incurs only linear time overhead during preprocessing and constant time overhead during a query, and allows us to limit the impact of superlinear cost of further processing to the size of the shard. Even to Ribbon, this could be used in multiple ways. For example, by statically splitting the table into pieces of size  $n^\varepsilon$  for standard ribbon, one can achieve space overhead  $\varepsilon + \mathcal{O}(n^{-\varepsilon})$ , preprocessing time  $\mathcal{O}(n/\varepsilon)$ , and query time  $\mathcal{O}(1)$  [20]. However, for  $\varepsilon < 0.01$  the use of asymptotics becomes questionable here. For example, for  $n \leq 2^{64}$ ,  $n^{0.01}$  is smaller than 1.6 which is not a sensible shard size considering per-shard space overheads of  $\Omega(\log n)$  bits. Before arriving at the current form of BuRR, we designed several variants based on sharding but never achieved better overhead than  $\Omega(1/w)$ . The current overhead of  $\mathcal{O}(\log w/w^2)$  comes from using the splitting technique in a “soft” way – elements are assigned to buckets for the purpose of defining bumping information but the ribbon solver may implicitly allocate them in subsequent buckets.

**Table lookup.** The first asymptotically efficient succinct retrieval data structure we are aware of [45] uses two levels of sharding to obtain very small shards of size  $\mathcal{O}(\sqrt{\log n})$  with small asymptotic overhead. It then uses dense random matrices per shard to obtain per-shard retrieval data structures. This can be done in constant time per shard by tabulating the solutions of all possible matrices. This leads to a multiplicative overhead of  $\mathcal{O}(\log \log n / \sqrt{\log n})$ . However, note that for  $n \leq 2^{64}$ ,  $\log \log n / \sqrt{\log n} \geq 3/4$  so that we do not get a meaningful bound for realistic input sizes.

Belazzougui and Venturini [4] use slightly larger buckets of size  $\mathcal{O}((1 + \log \log(n)/r) \log \log(n) / \log n)$ . Using carefully designed random lookup tables they show that linear construction time, constant lookup time, and overhead  $\mathcal{O}((\log \log n)^2 / \log n)$  is possible. This even applies to result sets with arbitrary distributions. Although their overhead is asymptotically smaller than the  $\mathcal{O}(\log \log n / \sqrt{\log n})$  above, the numerical test for  $n \leq 2^{64}$  and  $r = 1$  yields 0.657 which is also not useful for practical input sizes.

Recall that the overhead of BuRR is  $\mathcal{O}(\log(w)/(rw^2))$  which can be made arbitrarily small independent of  $n$ . Even if we set  $w = \log n$  for better comparability, we get  $\mathcal{O}(\log \log(n)/(r \log^2 n))$  which is

asymptotically better (and a numerical test gives 0.4% even for a realistic  $n = 2^{32}$ ).

In general, lookup tables are often problematic for compressed data structures in practice – they cause additional space overhead and cache faults. Even if the table is small and fits into cache, this may yield efficient benchmarks but can still cause cache faults in practical workloads where the data structure is only a small part in a large software system with a large working set.

**Cascaded bumping.** Hash tables consisting of multiple shrinking levels are also used in *multilevel adaptive hashing* [12] and *filter hashing* [25]. However, these do not maintain bumping information and thus have to access all levels.

## 2 Ribbon Insertions

In this section we briefly review the SGAUSS construction for retrieval from [20]. We enhance it with a new solver called **R**apid **I**ncremental **B**oolean **B**anding **O**N the fly (*Ribbon*), which is the basis of all ribbon variants considered later.

**The SGAUSS construction.** For a parameter  $w \in \mathbb{N}$  that we call the *ribbon width*, the vector  $\vec{h}(x) \in \{0, 1\}^m$  is given by a random *starting position*  $s(x) \in [m - w - 1]$  and a random *coefficient vector*  $c(x) \in \{0, 1\}^w$  as  $\vec{h}(x) = 0^{s(x)-1}c(x)0^{m-s(x)-w+1}$ . Note that even though  $m$ -bit vectors like  $\vec{h}(x)$  are used to simplify mathematical discussion, such vectors can be represented using  $\log(m) + w$  bits. The matrix  $A$  with rows  $(\vec{h}(x))_{x \in S}$  sorted by  $s(x)$  has all of its 1-entries in a “ribbon” of width  $w$  that randomly passes through the matrix from the top left to the bottom right, as in Figure 1 (a). The authors show:

**Theorem 2** ([20, Thm 2]). *For any constant  $0 < \varepsilon < \frac{1}{2}$  and  $\frac{n}{m} = 1 - \varepsilon$  there is a suitable choice for  $w = \Theta(\frac{\log n}{\varepsilon})$  such that with high probability the linear system  $(h(x) \cdot Z = f(x))_{x \in S}$  is solvable for any  $r \in \mathbb{N}$  and any  $f : S \rightarrow \{0, 1\}^r$ . Moreover, after sorting  $(\vec{h}(x))_{x \in S}$  by  $s(x)$ , Gaussian elimination can compute a solution  $Z$  in expected time  $\mathcal{O}(n/\varepsilon^2)$ .*

**Boolean banding on the fly.** For ribbon retrieval we use the same hash function  $\vec{h}$  as in SGAUSS except that we force coefficient vectors  $c(x)$  to start with 1, which slightly improves presentation and prevents construction failures caused by single keys with  $c(x) = 0^w$ . The main difference lies in how we solve the linear system. The *insertion phase* maintains a system  $M$  of linear equations in row echelon form using on-the-fly Gaussian elimination [6]. This system is of the form shown in Figure 1 (b) and has  $m$  rows that we also call *slots*. The  $i$ -th slot contains a  $w$ -bit vector  $c_i \in \{0, 1\}^w$  and  $b_i \in \{0, 1\}^r$ . Logically, the  $i$ -th slot is either empty ( $c_i = 0^w$ ) or specifies a linear equation  $c_i \cdot Z_{[i, i+w]} = b_i$  where  $c_i$  starts with a 1. With  $Z_{[i, i+w]} \in \{0, 1\}^{w \times r}$  we refer to rows  $i, \dots, i + w - 1$  of  $Z \in \{0, 1\}^{m \times r}$ . We ensure  $c_i \cdot Z_{[i, i+w]}$  is well-defined even when  $i + w - 1 > m$  with the invariant that  $c_i$  never selects “out of bounds” rows of  $Z$ .

We consider the equations  $\vec{h}(x) \cdot Z = f(x)$  for  $x \in S$  one by one, in arbitrary order, and try to integrate each into  $M$  using Algorithm 1, which we explain now.

A key’s equation may be modified several times before it can be added to  $M$ , but a loop invariant is that its form is

$$c \cdot Z_{[i, i+w]} = b \text{ for some } i \in [m], c \in 1 \circ \{0, 1\}^{w-1}, b \in \{0, 1\}^r. \quad (1)$$

The initial equation  $\vec{h}(x) \cdot Z = f(x)$  of key  $x \in S$  has this form with  $i = s(x)$ ,  $c = c(x)$  and  $b = f(x)$ .

**Case 1:** In the simplest case, slot  $i$  of  $M$  is empty and we can store Equation (1) in it.

**Case 2:** Otherwise slot  $i$  of  $M$  is occupied by an equation  $c_i \cdot Z_{[i, i+w]} = b_i$ . We perform a *row operation* to obtain the new equation

$$c' \cdot Z_{[i, i+w]} = b' \text{ with } c' = c \oplus c_i \text{ and } b' = b \oplus b_i, \quad (2)$$

---

**Algorithm 1:** Adding a key’s equation to the linear system  $M$ .

---

```

1  $(i, c, b) \leftarrow (s(x), c(x), f(x))$ 
2 loop
3   if  $M.c[i] = 0$  then // slot  $i$  of  $M$  is empty
4      $(M.c[i], M.b[i]) \leftarrow (c, b)$ 
5     return SUCCESS
6    $(c, b) \leftarrow (c \oplus M.c[i], b \oplus M.b[i])$ 
7   if  $c = 0$  then
8     if  $b = 0$  then return REDUNDANT
9     else return FAILURE
10   $j \leftarrow \text{findFirstSet}(c)$  // a.k.a. BitScanForward
11   $i \leftarrow i + j$ 
12   $c \leftarrow c \gg j$  // logical shift last toward first

```

---

which, in the presence of the equation in slot  $i$  of  $M$ , puts the same constraint on  $Z$  as Equation (1). Both  $c$  and  $c_i$  start with 1, so  $c'$  starts with 0. We consider the following sub-cases.

**Case 2.1:**  $c' = 0^w$  and  $b' = 0^r$ . The equation is void and can be ignored. This happens when the original equation of  $x$  is implied by equations previously added to  $M$ .

**Case 2.2:**  $c' = 0^w$  and  $b' \neq 0^r$ . The equation is unsatisfiable. This happens when the key’s original equation is inconsistent with equations previously added to  $M$ .

**Case 2.3:**  $c'$  starts with  $j > 0$  zeroes followed by a 1. Then Equation (2) can be rewritten as  $c'' \cdot Z_{[i', i'+w)} = b'$  where  $i' = i + j$  and  $c''$  is obtained from  $c'$  by discarding the  $j$  leading zeroes of  $c$  and appending  $j$  trailing zeroes.<sup>7</sup>

Termination is guaranteed since  $i$  increases with each loop iteration.

**“On-the-fly” and “incremental.”** The insertion phase of Ribbon is *on-the-fly* [6], i.e. maintains a row echelon form as keys arrive. This allows us to determine the longest prefix  $(x_1, \dots, x_n)$  of a sequence  $S = (x_1, x_2, x_3, \dots)$  of keys for which construction succeeds: Simply insert keys until the first failure. We say the insertion phase is *incremental* since an insertion may lead to a new row in  $M$  but does not modify existing rows. This allows us to easily undo the most recent successful insertions by clearing the slots of  $M$  that were filled last. These properties are not shared by SGAUSS and will be exploited by BuRR in Section 6.

### 3 Analysis of Ribbon Insertions

Given a set  $S$  of  $n$  keys we wish to analyze the process of inserting these keys into the system  $M$  using Algorithm 1. In particular, we are interested in the number of successful and failed insertions, the set of occupied slots in  $M$  and the total running time. Recall that  $A \in \{0, 1\}^{n \times m}$  contains the rows  $\vec{h}(x)$  for  $x \in S$  sorted by  $s(x)$ , see Figure 1 (a). Our analysis considers the *ribbon diagonal*, which is a line passing through  $A$ . We begin with an instructive simplification.

#### 3.1 A Warm Up: The Simplified Ribbon Diagonal

We make the following two assumptions:

---

<sup>7</sup>Note that in the bit-shift of Algorithm 1 the roles of “leading” and “trailing” may seem reversed because the least-significant “first” bit of a word is conventionally thought of as the “right-most” bit.

- (M1) Keys are inserted in the order they appear in  $A$  (sorted by  $s(x)$ ). This ensures that the insertion of each key  $x \in S$  fails or succeeds within the first  $w$  steps because no 1-entries can exist in  $M$  beyond column  $s(x) + w - 1$ .
- (M2) Inserting  $x \in S$  fills the first free slot  $i \in [s(x), s(x) + w - 1]$  unless all of these slots are occupied, in which case the insertion fails. This ignores the role of  $c(x)$ .

Figure 2 visualizes the process with an  $\times$  in position  $(j, i)$  if the insertion of the  $j$ -th key fills slot  $i$  of  $M$ . These points approximately trace a diagonal line from top left to bottom right and we call it the *simplified ribbon diagonal*  $d_{\text{simp}}$ . We make the following observations:

- (O1) If  $d_{\text{simp}}$  were to cross the bottom border of the ribbon, it skips a column (shown in green). Column  $i$  is skipped if and only if slot  $i$  of  $M$  remains empty.
- (O2) If  $d$  were to cross the right border of the ribbon, it skips a row (shown in red). Row  $j$  is skipped if and only if the  $j$ -th key is not inserted successfully.
- (O3) The area enclosed between  $d$  and the left border of the ribbon (shown in yellow) is an upper bound on the number of row operations performed during successful insertions.

### 3.2 The Ribbon Diagonal

A formal analysis can salvage much of the intuition from the simplified model. First, we show that (M1), though not (M2), can be made without loss of generality. For an adjusted definition of the ribbon diagonal, we then prove probabilistic versions of (O1), (O2) and (O3). The following notation will be useful.

- $S_i = \{x \in S \mid s(x) \leq i\}$  and  $s_i = |S_i|$ , for  $i \in [m]$ .
- $S' \subseteq S$  is the set of keys not inserted successfully. Moreover,  $S'_i = S_i \cap S'$  and  $s'_i = |S'_i|$ .
- $r_i$ , for  $i \in [m]$ , is the rank of the first  $i$  columns of  $A$ .
- $P_M$  is the set of slots of  $M$  that end up being filled.

**On (M1): The order of keys is irrelevant.** Since  $M$  arises from  $A$  by row operations, which do not affect ranks of sets of columns, we conclude that  $r_i$  is the rank of the first  $i$  columns of  $M$ , regardless of the order in which keys are handled. From the form of  $M$  (see Figure 1 (b)) it is clear that  $i \in P_M \Leftrightarrow r_i = r_{i-1} + 1$ . Therefore, the set  $P_M$  and thus the number  $n - |P_M| = |S'|$  of failed insertions is also invariant.

Assuming all insertions are successful, the number of row operations performed for key  $x$  is at most the distance of  $s(x)$  to the slot  $i(x) \in P_M$  that is filled. An invariant upper bound  $\Delta$  on the number of row operations, which are the dominating contribution to construction time, is then

$$\Delta := \sum_{x \in S} (i(x) - s(x)) = \sum_{i \in P_M} i - \sum_{x \in S} s(x). \quad (3)$$

Except for the time related to failed insertions, which we have to bound separately, we can derive everything we want from  $S$  and the invariants  $P_M, |S'|$ . We can therefore assume (M1).

**Definition and properties of  $d$ .** Given **(M1)**, we formally define the *ribbon diagonal*  $d$  as the following set of matrix positions in  $A$ .

$$d = \{(d_i, i) \mid i \in [m]\} \text{ where } d_i = r_i + s'_{i-w+1}.$$

It is useful to imagine the “default case” to be  $r_i = r_{i-1} + 1$  and  $s'_i = s'_{i-1}$ . We then have  $d_i = d_{i-1} + 1$  and the ribbon diagonal indeed moves diagonally down and to the right. An empty slot  $i \notin P_M$  correspond to a right-shift (due to  $r_i = r_{i-1}$ ) and a failed insertion of a key with  $s(x) = i - w + 1$  correspond to a down-shift (due to  $s'_{i-w+1} > s'_{i-w}$ ).

Let us first check that  $d$  is actually within the ribbon. More precisely:

**Lemma 3.** *For any  $i \in [m]$ ,  $d_i$  is not below the bottom ribbon border  $s_i$  and at most one position above the top ribbon border  $s_{i-w} + 1$ .*

*Proof.* The first claim holds because

$$d_i = r_i + s'_{i-w+1} \leq r_i + s'_i = |P_M \cap [1, i]| + s'_i \leq s_i$$

where the last step uses that each key in  $S_i$  can fail to be inserted or fill a slot in  $M$ , but not both. The latter is true because

$$d_i = |P_M \cap [1, i]| + s'_{i-w+1} \geq (s_{i-w+1} - s'_{i-w+1}) + s'_{i-w+1} = s_{i-w+1} \geq s_{i-w}.$$

where the first “ $\geq$ ” uses that the first  $s_{i-w+1}$  rows cause  $s_{i-w+1} - s'_{i-w+1}$  slots with index at most  $i$  to be filled.  $\square$

The first part of Lemma 3 ensures that the height  $h_i := s_i - d_i$  of the ribbon diagonal above the bottom ribbon border is non-negative. It plays a central role in the precise versions of **(O1)** to **(O3)** we prove next. The main adjustment we have to make is that  $d$  is probabilistically repelled when *close* to the ribbon border, while  $d_{\text{simp}}$  only responds to outright collisions.

**Lemma 4** (Precise version of **(O1)**). *We have  $\Pr[i \notin P_M \mid h_{i-1} = k] \leq 2^{-k}$  for any  $k \in \mathbb{N}_0$ .*

*Proof.* A useful alternative way to think about Algorithm 1 uses language from linear probing: A key  $x$  probes slots  $s(x), s(x) + 1, \dots, s(x) + w - 1$  one by one. When probing an empty slot,  $x$  is inserted into that slot with probability  $\frac{1}{2}$ , otherwise it keeps probing.<sup>8</sup> Now consider slot  $i$ . Of the  $s_{i-1}$  keys with starting position at most  $i - 1$ , precisely  $r_{i-1}$  are successfully inserted to slots in  $[1, i - 1]$  and  $s'_{i-w}$  insertions fail without probing slot  $i$ . Therefore  $s_{i-1} - s'_{i-w} - r_{i-1} = s_{i-1} - d_{i-1} = h_{i-1}$  keys probe slot  $i$ . So conditioned on  $h_{i-1} = k$ , slot  $i$  remains empty with probability at most  $2^{-k}$ .  $\square$

**Lemma 5** (Precise version of **(O2)**). *Let  $x$  be a key with  $s(x) = i$ .*

**(a)** *Let  $i' \in [i, i + w]$  be the column of  $A$  where the ribbon diagonal passes the row of  $x$ . Assume  $i' - i = w - k$ , i.e.  $i'$  is  $k$  positions left of the right ribbon border. Conditioned on this,  $\Pr[x \in S'] \leq 2^{-k}$ .*

**(b)** *A simple variant of this claim is: If  $h_i \leq w - k$  for some  $k \in \mathbb{N}$  then  $\Pr[x \in S'] \leq 2^{-k}$ .*

*Proof.* **(a)** Let  $i = s(x)$ . We may assume that  $x$  is the last key with starting position  $i$  as this can only increase  $\Pr[x \in S']$ . This means  $x$  corresponds to row  $s_i$  and hence  $d_{i'} \geq s_i$ . Of the  $s_i - 1$  keys that

<sup>8</sup>This uses that for any key  $x$  and  $i \in [s(x) + 1, s(x) + w - 1]$  the random coefficient  $a_i$  that  $x$  has for slot  $i$  remains fully random until slot  $i$  is reached, since the bits that are added to  $a_i$  during row operations are uncorrelated with  $a_i$ .

are handled before  $x$ , exactly  $r_{i-1}$  are inserted before slot  $i$  and at least  $s'_{i-1}$  were not inserted successfully. The number of slots in  $[i, m]$  that are occupied when  $x$  is handled is therefore at most

$$\begin{aligned} s_i - 1 - r_{i-1} - s'_{i-1} &\leq d_{i'} - 1 - r_{i-1} - s'_{i-1} = r_{i'} + s_{i'-w+1} - 1 - r_{i-1} - s'_{i-1} \\ &\leq r_{i'} - r_{i-1} - 1 \leq i' - i = w - k. \end{aligned}$$

This means at least  $k$  slots within  $[i, i + w - 1]$  are empty. The probability that  $x$  cannot be inserted despite probing these  $k$  slots is  $2^{-k}$ .

(b) The assumption gives an alternative way to derive the same intermediate step:

$$s_i - 1 - r_{i-1} - s'_{i-1} \leq s_i - r_i - s'_{i-w+1} = s_i - d_i = h_i \leq w - k. \quad \square$$

**Lemma 6** (Precise version of (O3)). *Let  $\text{op}_+$  be the number of row operations performed during successful insertions. We have  $\text{op}_+ \leq n(w - 1)$  (trivially) as well as  $\text{op}_+ \leq \sum_{i \in [m]} h_i$ .*

*Proof.* First assume that all insertions succeed and consider Equation (3). Since  $i(x) - s(x) \leq w - 1$  holds for all  $x \in S$  the trivial bound  $\text{op}_+ \leq \Delta \leq n(w - 1)$  follows. Now consider the right hand side of Equation (3). The sum  $\sum_{x \in S} (s(x) - 1)$  can be interpreted as the area in  $A$  (i.e. the number of matrix positions) left of the ribbon. Moreover we have

$$\sum_{i \in P_M} (i - 1) = \sum_{i \in [m]} |P_M \cap [i + 1, m]| = \sum_{i \in [m]} n - r_i = \sum_{i \in [m]} n - d_i.$$

which is the area below the ribbon diagonal. This makes  $\Delta$  the area enclosed between the ribbon diagonal and the lower ribbon border. A column-wise computation of this area yields  $\text{op}_+ \leq \Delta = \sum_{i \in [m]} h_i$  as desired.

Contrary to our initial assumption, there may be keys that fail to be inserted. But our bounds remain valid in the presence of such keys: The number  $\text{op}_+$  only counts operations made for successfully inserted keys and hence does not change. Our bounds  $n(w - 1)$  and  $\sum_{i \in [m]} h_i$  are easily seen to increase by  $w - 1$  and  $w$ , respectively, for each additional “failing” key we take into account.  $\square$

## 4 Analysis of Standard Ribbon Retrieval

By *standard ribbon* we mean the original design from [20], except that we use our improved solver. We sketch an implementation in Algorithm 2 and recall the broad strokes of the analysis from [20] which will help us to analyze homogeneous ribbon filters in Section 5.

Given  $n \in \mathbb{N}$  keys we allocate a system  $M$  of size  $m = n/(1 - \varepsilon) + w - 1$  and try to insert all keys using Algorithm 1. If any insertion fails, the entire construction is restarted with new hash functions. Otherwise, we obtain a solution  $Z$  to  $M$  in the *back substitution phase*. The rows of  $Z$  are obtained from bottom to top. If slot  $i$  of  $M$  contains an equation then this equation uniquely determines row  $i$  of  $Z$  in terms of later rows of  $Z$ . If slot  $i$  of  $M$  is empty, then row  $i$  of  $Z$  can be initialized arbitrarily.

The expected “slope” of the ribbon is  $1 - \varepsilon$ , giving us reason to hope that the ribbon diagonal will stick to the left ribbon border making failures unlikely.

**Lemma 7.** *The heights  $h_i := s_i - d_i$  for  $i \in [m]$  satisfy:*

(a)  $\mathbb{E}[h_i] \leq \mathcal{O}(1/\varepsilon)$

(b)  $\forall k \in \mathbb{N} : \Pr[h_i > k] = \exp(-\Omega(\varepsilon k)).$

---

**Algorithm 2:** The construction algorithm of standard ribbon.

---

**Input:**  $f : S \rightarrow \{0, 1\}^r$  for some  $S \subseteq \mathcal{U}$  of size  $n$ .  
**Parameters:**  $w \in \mathbb{N}$ ,  $\varepsilon > 0$ .

- 1  $m \leftarrow n/(1 - \varepsilon) + w - 1$ ; allocate system  $M$  of size  $m$
- 2 pick hash functions  $s : \mathcal{U} \rightarrow [m - w + 1]$ ,  $c : \mathcal{U} \rightarrow \{0, 1\}^w$
- 3 **for**  $x \in S$  **do**
- 4  $\text{ret} \leftarrow \text{insert}(x)$  // using Algorithm 1
- 5 **if**  $\text{ret} = \text{FAILURE}$  **then**
- 6  $\text{restart}$
- 7  $Z \leftarrow 0^{m \times r}$
- 8 **for**  $i = m$  **down to** 1 **do**
- 9  $Z_i \leftarrow M.c[i] \cdot Z_{i..i+w-1}$  // back substitution
- 10 **return**  $(s, c, Z)$

---

*Proof idea from [20].* By definition of  $h_i$ ,  $d_i$  and  $r_i$  we have

$$\begin{aligned} h_i - h_{i-1} &= (s_i - s_{i-1}) - (r_i - r_{i-1}) - (s'_{i-w+1} - s'_{i-w}) \\ &\leq (s_i - s_{i-1}) - (r_i - r_{i-1}) = (s_i - s_{i-1}) - \mathbb{1}_{i \in P_M}. \end{aligned}$$

The number  $s_i - s_{i-1}$  of keys with starting position  $i$  has distribution  $\text{Bin}(n, \frac{1}{m-w+1})$  which is approximately  $\text{Po}(1 - \varepsilon)$ . By Lemma 4 we have  $\Pr[i \notin P_M] \leq 2^{-h_{i-1}}$ . Roughly speaking this means that  $\Pr[\mathbb{1}_{i \in P_M} \neq 1]$  is negligible as soon as  $h_{i-1}$  rises to a value large enough to threaten the upper bounds we intend to prove. A coupling argument then allows us to upper bound  $h_i$  in terms of a so-called M/D/1 queue. In every time step  $\text{Po}(1 - \varepsilon)$  customers arrive and 1 customer can be serviced. The stated bounds on (a) expectation and (b) tails of  $h_i$  stem from the literature on such queues.

We remark that the term  $s'_{i-w+1} - s'_{i-w}$  that we ignored relates to failed insertions. It translates to customers abandoning the queue after waiting for  $w$  time steps without being serviced.  $\square$

By choosing  $w = \Omega(\frac{\log n}{\varepsilon})$  it follows from Lemma 7 (b) that  $h_i \leq w/2$  for all  $i \in [m]$  whp. Lemma 5 (b) then ensures that all keys can be inserted successfully whp. Combining Lemma 6 with Lemma 7 (a) shows that the expected number of row operations during construction is  $\mathcal{O}(n/\varepsilon)$ .

## 5 Analysis of Homogeneous Ribbon Filters

In this section we give a precise description and analysis of homogeneous ribbon filters, which are even simpler than filters based on standard ribbon but unsuitable for retrieval.

Recall the approach for constructing a filter by picking hash functions  $\vec{h} : \mathcal{U} \rightarrow \{0, 1\}^m$ ,  $f : \mathcal{U} \rightarrow \{0, 1\}^r$  and finding  $Z \in \{0, 1\}^{m \times r}$  such that all  $x \in S$  satisfy  $\vec{h}(x) \cdot Z = f(x)$ , while most  $x \in \mathcal{U} \setminus S$  do not. We now dispose of the fingerprint function  $f$ , effectively setting  $f(x) = 0$  for all  $x \in \mathcal{U}$ . A filter is then given by a solution  $Z$  to the *homogeneous* system  $(\vec{h}(x) \cdot Z = 0^r)_{x \in S}$ . The FP rate for  $Z$  is  $\varphi_Z = \Pr_{a \sim H}[a \cdot Z = 0^r]$  where  $H$  is the distribution of  $\vec{h}(x)$  for  $x \in \mathcal{U}$ . An immediate issue with the idea is that  $Z = 0^{m \times r}$  is a valid solution but gives  $\varphi_Z = 1$ . We therefore pick  $Z$  *uniformly at random* from all solutions. If  $\vec{h}$  has the form  $\vec{h}(x) = 0^{s(x)-1} c(x) 0^{m-s(x)-w+1}$  from standard ribbon retrieval, we call the resulting filter a *homogeneous ribbon filter*. The full construction is shown in Algorithm 3. Two notable simplifications compared to Algorithm 2 are that no function  $f$  is needed and that a restart is

never required. Note, however, that free variables must now be sampled uniformly at random<sup>9</sup> during back substitution.

---

**Algorithm 3:** The construction algorithm of homogeneous ribbon filters.

---

**Input:**  $S \subseteq \mathcal{U}$  of size  $n$ .  
**Parameters:**  $r \in \mathbb{N}, w \in \mathbb{N}, \varepsilon > 0$ .

- 1  $m \leftarrow n/(1 - \varepsilon) + w - 1$ ; allocate system  $M$  of size  $m$
- 2 pick hash functions  $s : \mathcal{U} \rightarrow [m - w + 1]$ ,  $c : \mathcal{U} \rightarrow \{0, 1\}^w$
- 3 sort  $S$  approximately by  $s(x)$  (see Lemma 9)
- 4 **for**  $x \in S$  **do**
- 5    $\lfloor$  insert( $x$ ) // using Algorithm 1 with  $f \equiv 0$ . Cannot fail!
- 6  $Z \leftarrow 0^{m \times r}$
- 7 **for**  $i = m$  **down to** 1 **do**
- 8   **if**  $M.c[i] = 0$  **then** // slot unused?
- 9      $\lfloor$  sample  $Z_i \sim U(\{0, 1\}^r)$  // randomly initialize free variable
- 10   **else**
- 11      $\lfloor$   $Z_i \leftarrow M.c[i] \cdot Z_{i..i+w-1}$  // back substitution
- 12 **return**  $(s, c, Z)$

---

The overall FP rate is  $\varphi = \mathbb{E}[\varphi_Z]$  where  $Z$  depends on the randomness in  $(\vec{h}(x))_{x \in S}$  and the free variables. A complication is that  $\varphi = 2^{-r}$  no longer holds, instead there is a gap  $\varphi - 2^{-r} > 0$ . We show that this gap is negligible under two conditions. Firstly, the filter must be *underloaded*, with  $\varepsilon \approx \frac{m-n}{m} > 0$ , which leads to a memory overhead of  $\mathcal{O}(\varepsilon)$ . Secondly, the ribbon width  $w$  must satisfy  $w = \Omega(r/\varepsilon)$ . The good news is that there is no dependence of  $w$  on  $n$  (such as  $w = \Omega(\frac{\log n}{\varepsilon})$  required in standard ribbon) and that no sharding or bumping is required. More precisely, we prove the following.

**Theorem 8.** *Let  $r \in \mathbb{N}$  and  $\varepsilon \in (0, \frac{1}{2}]$ . There is  $w \in \mathbb{N}$  with  $\frac{w}{\max(r, \log w)} = \mathcal{O}(1/\varepsilon)$  such that the homogeneous ribbon filter has false positive rate  $\varphi \approx 2^{-r}$  and space overhead  $\mathcal{O}(\varepsilon)$ . On a word RAM with word size  $\geq w$  expected construction time is  $\mathcal{O}(n/\varepsilon)$  and query time is  $\mathcal{O}(r)$ .*

Note that when targeting  $w = \Theta(\log n)$  we can achieve an overhead of  $\varepsilon = \mathcal{O}(\frac{\max(r, \log \log n)}{\log n})$ .

### 5.1 Proof of Theorem 8

The easier part is to prove the running time bounds. The query time of  $\mathcal{O}(r)$  is, in fact, obvious. For the construction time, we reuse results for standard ribbon. Though insertions cannot fail, the set of *redundant* keys, i.e. keys for which Algorithm 1 returns “REDUNDANT” rather than “SUCCESS” now demands attention.

**Lemma 9.** *Consider the setting of Theorem 8.*

- (a) *The fraction of keys that lead to redundant insertions is  $\exp(-\Omega(\varepsilon w))$ .*
- (b) *The expected number of row additions during construction is  $\mathcal{O}(n/\varepsilon)$ .*

*Proof.* (a) Any key  $x \in S$  with a starting position  $i = s(x)$  for which  $h_i \leq w/2$  is, by Lemma 5 (b), inserted successfully with probability at least  $1 - 2^{-w/2}$ . By Lemma 7 the expected fraction of

---

<sup>9</sup>Our implementation uses trivial pseudo-random assignments instead: a free variable in row  $i$  is assigned  $pi \bmod 2^r$  for some fixed large odd number  $p$ .

positions to which this argument does not apply is  $\exp(-\Omega(\varepsilon w))$ . From this it is not hard to see that the expected fraction of *keys* to which this argument does not apply is also  $\exp(-\Omega(\varepsilon w))$ . The fraction of keys not inserted successfully is therefore  $\mathcal{O}(2^{-w/2}) + \exp(-\Omega(\varepsilon w)) = \exp(-\Omega(\varepsilon w))$ .

- (b) Combining Lemma 6 with Lemma 7 (a) shows that the expected number of row operations during successful insertions is  $\mathcal{O}(n/\varepsilon)$ . Redundant keys are somewhat annoying to deal with. They are the reason we partially sort the key set in Algorithm 3. If keys are sorted into buckets of  $b$  consecutive starting positions each and buckets handled from left to right, then no attempted insertion can take longer than  $b + w$  steps. Thus,  $b \leq \exp(\Omega(\varepsilon w))$  ensures that redundant insertions contribute  $\mathcal{O}(n)$  to expected total running time.  $\square$

To get a grip on the false positive rate, we start with the following simple observation.

**Lemma 10.** *Let  $p$  be the probability that for  $y \in \mathcal{U} \setminus S$  the vector  $\vec{h}(y)$  is in the span of  $(\vec{h}(x))_{x \in S}$ . The false positive rate of the homogeneous ribbon filter is*

$$\varphi = p + (1 - p)2^{-r}.$$

*Proof.* First assume there exists  $S' \subseteq S$  with  $\vec{h}(y) = \sum_{x \in S'} \vec{h}(x)$  which happens with probability  $p$ . In that case

$$\vec{h}(y) \cdot Z = \left( \sum_{x \in S'} \vec{h}(x) \right) \cdot Z = \sum_{x \in S'} (\vec{h}(x) \cdot Z) = 0$$

and  $y$  is a false positive. Otherwise, i.e. with probability  $1 - p$ , an attempt to add  $\vec{h}(y) \cdot Z = 0$  to  $M$  after all equations for  $S$  were added would have resulted in a (non-redundant) insertion in some row  $i$ . During back substitution, only one choice for the  $i$ -th row of  $Z$  satisfies  $\vec{h}(y) \cdot Z = 0$ . Since the  $i$ -th row was initialized randomly we have  $\Pr[\vec{h}(y) \cdot Z = 0 \mid \vec{h}(y) \notin \text{span}((\vec{h}(x))_{x \in S})] = 2^{-r}$ .  $\square$

We now derive an asymptotic bound on  $p$  in terms of large  $w$  and small  $\varepsilon$ .

**Lemma 11.** *There exists a constant  $C$  such that whenever  $C \frac{\log w}{w} \leq \varepsilon \leq \frac{1}{2}$  we have  $p = \exp(-\Omega(\varepsilon w))$ .*

*Proof.* We may imagine that  $S \subseteq \mathcal{U}$  and  $y \in \mathcal{U} \setminus S$  are obtained from a set  $S^+ \subseteq \mathcal{U}$  of size  $n + 1$  by picking  $y \in S^+$  at random and setting  $S = S^+ \setminus \{y\}$ . Then  $p$  is simply the expected fraction of keys in  $S^+$  that are contained in some *dependent set*, i.e. in some  $S' \subseteq S^+$  with  $\sum_{x \in S'} \vec{h}(x) = 0^m$ . Clearly,  $x$  is contained in a dependent set if and only if it is contained in a *minimal* dependent set. Such a set  $S'$  always “touches” a consecutive set of positions, i.e.  $\text{pos}(S') := \bigcup_{x \in S'} [s(x), s(x) + w - 1]$  is an interval.

We call an interval  $I \subseteq [m]$  *long* if  $|I| \geq w^2$  and *short* otherwise. We call it *overloaded* if  $S_I := \{x \in S^+ \mid s(x) \in I\}$  has size  $|S_I| \geq |I| \cdot (1 - \varepsilon/2)$ . Finally, we call a position  $i \in [m]$  *bad* if one of the following is the case:

- (B1)  $i$  is contained in a long overloaded interval.
- (B2)  $i \in \text{pos}(S')$  for a minimal dependent set  $S'$  with long non-overloaded interval  $\text{pos}(S')$ .
- (B3)  $i \in \text{pos}(S')$  for a minimal dependent set  $S'$  with short interval  $\text{pos}(S')$ .

We shall now establish the following

**Claim:**  $\forall i \in [m] : \Pr[i \text{ is bad}] = \exp(-\Omega(\varepsilon w))$ .

For each  $i \in [m]$  the contributions from each of the badness conditions (B1, B2, B3) can be bounded separately. In all cases we use our assumption  $\varepsilon \geq C \frac{\log w}{w}$ . It ensures that  $\exp(-\Omega(\varepsilon w))$  is at most  $\exp(-\Omega(\log w)) = w^{-\Omega(1)}$  and can “absorb” factors of  $w$  in the sense that by adapting the constant hidden in  $\Omega$  we have  $w \exp(-\Omega(\varepsilon w)) = \exp(-\Omega(\varepsilon w))$ .

(B1) Let  $I \subseteq [m]$  be any interval and  $X_1, \dots, X_{n+1}$  indicate which of the keys in  $S^+$  have a starting position within  $I$ . For  $n \gg w$  and  $X := \sum_{j \in [n+1]} X_j$  we have

$$\mu := \mathbb{E}[X] \leq \frac{(n+1)|I|}{m-w+1} \approx \frac{n|I|}{m-w+1} = (1-\varepsilon)|I|.$$

Using a Chernoff bound (Lemma 15 (a)), the probability for  $I$  to be overloaded is (for  $n \gg w$ )

$$\Pr[X \geq (1-\varepsilon/2)|I|] \leq \Pr[X \geq \underbrace{(1+\varepsilon/2)}_{\delta} \underbrace{(1-\varepsilon)|I|}_{\geq \mu}] \stackrel{\text{Lem.15}}{\leq} \exp\left(\frac{-\varepsilon^2(1-\varepsilon)|I|}{12}\right). \quad (4)$$

The probability for  $i \in [m]$  to be contained in a long overloaded interval is bounded by the sum of Equation (4) over all lengths  $|I| \geq w^2$  and all  $|I|$  offsets that  $I$  can have relative to  $i$ . The result is of order  $\exp(-\Omega(\varepsilon^2 w^2))$  and hence small enough.

(B2) Consider a long interval  $I$  that is not overloaded, i.e.  $|I| \geq w^2$  and  $|S_I| \leq (1-\varepsilon/2)|I|$ . There are at most  $2^{|S_I|}$  sets  $S'$  of keys with  $\text{pos}(S') = I$  and each is a dependent set with probability  $2^{-|I|}$  because each of the  $|I|$  positions of  $I$  that  $S'$  touches imposes one parity condition.

A union bound on the probability for  $I$  to support at least one dependent set is therefore  $2^{-|I|} \cdot 2^{|S_I|} = 2^{-\frac{\varepsilon}{2}|I|} = \exp(-\Omega(\varepsilon|I|))$ .

Similar as in (B1) for  $i \in [m]$  we can sum this probability over all admissible lengths  $|I| \geq w^2$  and all offsets that  $i$  can have in  $I$  to bound the probability that  $i$  is bad due to (B2).

(B3) Let  $S_{\text{red}} \subseteq S^+$  be the set of redundant keys. By Lemma 9 we have  $\mathbb{E}[|S_{\text{red}}|] = n \cdot \exp(-\Omega(\varepsilon w))$ .

Now if  $i$  is bad due to (B3) then  $i \in \text{pos}(S')$  for some minimal dependent set  $S'$  with short  $\text{pos}(S')$ . At least one key from  $S'$  is redundant (regardless of the insertion order). In particular,  $i$  is within short distance ( $< w^2$ ) of the starting position of a redundant key  $x$ . Therefore at most  $|S_{\text{red}}| \cdot 2w^2$  positions are bad due to (B3), which is an  $\exp(-\Omega(\varepsilon w))$ -fraction of all positions as desired.

Simple tail bounds on the number of keys with the same starting position suffice to show the following variant of the claim:

$$\textbf{Claim': } \forall x \in S^+ : \Pr[s(x) \text{ is bad}] = \exp(-\Omega(\varepsilon w)).$$

Now assume that the key  $y \in S^+$  we singled out is contained in a minimal dependent set  $S'$ . It follows that all of  $\text{pos}(S')$  would be bad. Indeed, either  $\text{pos}(S')$  is a short interval ( $\rightarrow$  B3) or it is long. If it is long, then it is overloaded ( $\rightarrow$  B1) or not overloaded ( $\rightarrow$  B2). In any case  $s(y) \in \text{pos}(S')$  would be bad.

Therefore, the probability  $p$  for  $y \in S^+$  to be contained in a dependent set is at most the probability for  $s(y)$  to be bad. This is upper-bounded by  $\exp(-\Omega(\varepsilon w))$  using Claim'.  $\square$

We are now ready to prove Theorem 8.

*Proof of Theorem 8.* We already dealt with running times in Lemma 9.

The constraint  $\frac{w}{\max(r, \log w)} = \mathcal{O}(1/\varepsilon)$  leaves us room to assume  $\varepsilon w > Cr$  and  $\varepsilon w > C \log w$  for a constant  $C$  of our choosing. Concerning the false positive rate we obtain

$$p \stackrel{\text{Lem 11}}{\leq} \exp(-\varepsilon w) \leq \exp(-2 \log(w) - r) \leq \frac{1}{w^2} e^{-r} \leq \varepsilon^2 2^{-r}$$

and hence  $\varphi \stackrel{\text{Lem 10}}{=} p + (1-p)2^{-r} \leq p + 2^{-r} \leq \varepsilon^2 2^{-r} + 2^{-r} = (1 + \varepsilon^2)2^{-r}$ .

which is close to  $2^{-r}$  as desired. Concerning the space overhead, recall its definition as  $\frac{\text{SPACE}}{\text{OPT}} - 1$  where  $\text{SPACE}$  is the space usage of the filter and  $\text{OPT} = -\log_2(\varphi)n$  is the information-theoretic lower bound for filters that achieve false positive rate  $\varphi$ . We have:

$$\begin{aligned} \text{OPT} &= -\log_2(\varphi)n \geq -\log_2((1 + \varepsilon^2)2^{-r})n = (r - \log_2(1 + \varepsilon^2))n \geq (r - \varepsilon^2)n \\ \text{and } \text{SPACE} &= rm = r(m - w + 1) + \mathcal{O}(rw) = \frac{rn}{1 - \varepsilon} + \mathcal{O}(wr) \\ \text{which yields } \frac{\text{SPACE}}{\text{OPT}} &= \frac{r}{(1 + \varepsilon)(r - \varepsilon^2)} + \mathcal{O}\left(\frac{w}{n}\right) \leq \frac{1}{(1 + \varepsilon)(1 - \varepsilon^2)} + \mathcal{O}\left(\frac{w}{n}\right) \leq 1 + 3\varepsilon, \end{aligned}$$

where the last step uses  $\varepsilon \leq \frac{1}{2}$ . □

## 6 Analysis of Bumped Ribbon Retrieval (BuRR)

We now single out one variant of BuRR and analyze it fully, thereby proving Theorem 1. The analysis could undoubtedly be extended to cover other variants of BuRR (see Section 7), but in the interest of a cleaner presentation we will not do so.

Recall the idea illustrated in Figure 2 (b): We use  $m < n$ , making the data structure *overloaded*. This ensures that the ribbon diagonal  $d$  rarely hits the bottom ribbon border and (O1)/Lemma 4 suggests that almost all slots in  $M$  can be utilized. An immediate problem is that  $d$  would necessarily hit the right ribbon border in at least  $n - m$  places, causing at least  $n - m$  insertions to fail. We deal with this by removing contiguous ranges of keys in strategic positions such that without the corresponding rows,  $d$  never hits the right ribbon border. A small amount of “metadata” indicates the ranges of removed keys. These keys are *bumped* to a fallback retrieval data structure. Many variants of this approach are possible, see Section 7.

### 6.1 Proof of Theorem 1

Consider Algorithm 4. In what follows,  $C$  refers to a constant from Lemma 15. For  $n$  keys,  $M$  is given  $m = n/(1 + \frac{C \log w}{8w}) + w - 1$  rows<sup>10</sup>. The  $m - w + 1$  possible starting positions are partitioned into *buckets* of size  $b = \frac{w^2}{C \log w}$ . The first  $\frac{3}{8}w$  slots of a bucket are called its *head*, and the larger rest is called its *tail*. Keys implicitly belong to (the head or tail of) a bucket according to their starting position. For each bucket the algorithm has three choices:

1. No keys belonging to the bucket are bumped.
2. The keys belonging to the head of the bucket are bumped.
3. All keys of the bucket are bumped.

These choices are made greedily as follows. Buckets are handled from left to right. For each bucket, we first try to insert all keys belonging to the bucket’s tail. If at least one insertion fails, then the successful insertions are undone and the entire bucket is bumped, i.e. Option 3 is used. Otherwise, we also try to insert the keys belonging to the bucket’s head. If at least one insertion fails, all insertions of head keys are undone and we choose Option 2, otherwise, we choose Option 1. The main ingredient in the analysis of this algorithm is the following lemma, proved later in this section.

**Lemma 12.** *The expected fraction of empty slots in  $M$  is  $\mathcal{O}(w^{-3})$ .*

<sup>10</sup>We ignore rounding issues for a clearer presentation and assume that  $w$  is large. This causes a certain disconnect to practical application where concrete values like  $w = 32$  are used.

---

**Algorithm 4:** The construction algorithm of BuRR as analyzed in Section 6.

---

**Input:**  $f : S \rightarrow \{0, 1\}^r$  for some  $S \subseteq \mathcal{U}$  of size  $n$ .  
**Parameters:**  $w \in \mathbb{N}$ .

- 1  $m \leftarrow n / (1 + \frac{C \log w}{8w}) + w - 1$ ; allocate system  $M$  of size  $m$
- 2 pick hash functions  $s : \mathcal{U} \rightarrow [m - w + 1]$ ,  $c : \mathcal{U} \rightarrow \{0, 1\}^w$
- 3  $b \leftarrow \frac{w^2}{C \log w}$ , #buckets  $\leftarrow \frac{m-w+1}{b}$  // bucket size & number of buckets
- 4 **for**  $j \in [\text{\#buckets}]$  **do** // partition
  - 5  $B_j \leftarrow \{x \in S \mid \lceil s(x)/b \rceil = j\}$
  - 6  $H_j \leftarrow \{x \in B_j \mid s(x) - (j-1)b \leq \frac{3}{8}w\}$  // head
  - 7  $T_j = B_j \setminus H_j$  // tail
- 8  $S_{\text{bumped}} \leftarrow \emptyset$
- 9 **for**  $j \in [\text{\#buckets}]$  **do**
  - 10 // insertAll(X): attempt Algorithm 1 for all  $x \in X$ , roll back on failure
  - 11 **if** insertAll( $T_j$ ) **then**
    - 12 **if** insertAll( $H_j$ ) **then**
      - 13 | meta[ $j$ ]  $\leftarrow$  BUMPNOTHING
      - 14 **else**
      - 15 |  $S_{\text{bumped}} \leftarrow S_{\text{bumped}} \cup H_j$
      - 16 | meta[ $j$ ]  $\leftarrow$  BUMPHEAD
    - 17 **else**
    - 18 |  $S_{\text{bumped}} \leftarrow S_{\text{bumped}} \cup B_j$
    - 19 | meta[ $j$ ]  $\leftarrow$  BUMPALL
- 20  $Z \leftarrow 0^{m \times r}$
- 21 **for**  $i = m$  **down to** 1 **do**
  - 22 |  $Z_i \leftarrow M.c[i] \cdot Z_{i,i+w-1}$  // back substitution
- 23  $D_{\text{bumped}} \leftarrow \text{construct}(S_{\text{bumped}})$  // recursive, unless base case reached
- 24 **return**  $D = (s, c, Z, \text{meta}, D_{\text{bumped}})$

---

If the fraction of empty slots is significantly higher than expected, we simply restart the construction with new hash functions until satisfactory (this is not reflected in Algorithm 4). After back substitution, we obtain a solution vector of  $mr$  bits. Additionally, we need to store the choices we made, which takes  $\lceil \log_2 3 \rceil = 2$  bits of metadata per bucket. Given that  $|P_M| = m \cdot (1 - \mathcal{O}(w^{-3}))$  keys are taken care of, this suggests a space overhead of

$$\begin{aligned} \varepsilon &= \frac{\text{SPACE}}{\text{OPT}} - 1 \leq \frac{mr + 2\frac{m}{b}}{|P_M| \cdot r} - 1 \leq \frac{1 + 2\frac{1}{rb}}{1 - \mathcal{O}(w^{-3})} - 1 \\ &= (1 + \frac{2}{rb})(1 + \mathcal{O}(w^{-3})) - 1 = \mathcal{O}(\frac{1}{rb}) + \mathcal{O}(w^{-3}) = \mathcal{O}(\frac{\log w}{rw^2}) + \mathcal{O}(w^{-3}) = \mathcal{O}(\frac{\log w}{rw^2}). \end{aligned}$$

The last step uses the assumption  $r = \mathcal{O}(w)$ . The trivial bound in Lemma 6 implies that  $\mathcal{O}(bw)$  row operations are performed during the *successful* insertions in a bucket. There can be at most one failed insertion for each bucket which takes  $\mathcal{O}(b)$  row operations since insertions cannot extend past the next (still empty) bucket. Since  $w = \mathcal{O}(\log n)$  bits fit into a word of a word RAM, these row operations take  $\mathcal{O}(\frac{m}{b} \cdot (bw + b)) = \mathcal{O}(nw)$  time in total.

A query of a non-bumped key involves computing the product of the  $w$ -bit vector  $c(x)$  and a block  $Z(x)$  of  $w \times r$  bits from the solution matrix  $Z \in \{0, 1\}^{m \times r}$ . The  $wr$  bit operations can be carried out in  $\mathcal{O}(1 + \frac{wr}{\log n})$  steps on a word RAM with word size  $\Omega(\log n)$ . A complication is that if  $w, r \in \omega(1) \cap o(\log n)$

then we are forced us to handle several rows of  $Z(x)$  in parallel (xor-ing a  $c(x)$ -controlled selection) or several columns of  $Z(x)$  in parallel (bitwise AND with  $c(x)$  and popcount). Numbers “much bigger than 1 and much smaller than  $\log n$ ” are a somewhat academic concern, so we believe an academic resolution (not reflected in our implementation<sup>11</sup>) is sufficient: We resort to the standard techniques of tabulating the results of a suitable set of vector matrix products. Back substitution has the same complexity as  $n$  queries and therefore takes  $\mathcal{O}(n(1 + \frac{wr}{\log n})) = \mathcal{O}(nw)$  time.

To complete the construction, we still have to deal with the  $n - |P_M| = n - m(1 - \mathcal{O}(w^{-3})) = \mathcal{O}(\frac{n \log w}{w})$  bumped keys. A query can easily identify from the metadata whether a key is bumped, so all we need is another retrieval data structure that is consulted in this case. We can recursively use bumped ribbon retrieval again. However, to avoid compromising worst-case query time we only do this for four levels. Let  $S^{(4)}$  be the set of keys bumped four times. We have  $|S^{(4)}| = \mathcal{O}(n \frac{\log^4 w}{w^4}) = \mathcal{O}(n \frac{\log w}{rw^2})$  and we can afford to store  $S^{(4)}$  using a retrieval data structure with constant overhead, linear construction time and  $\mathcal{O}(1)$  worst-case query time, e.g., using minimum perfect hash functions [10].<sup>12</sup>

## 6.2 Proof of Lemma 12

We give an induction-like argument showing that “most” buckets satisfy two properties:

**(P1)** All slots of the bucket are filled.

**(P2)** The height  $h_i := s_i - d_i$  of the ribbon diagonal over the lower ribbon border at the last position  $i$  of the bucket satisfies  $h_i \geq \frac{w}{4}$ .<sup>13</sup>

**Claim 13.** *If (P2) holds for a bucket  $B_0$  then (P1) and (P2) hold for the following bucket  $B_1$  with probability  $1 - w^{-3}$ .*

*Proof.* Let  $i_0$  and  $i_1$  be the last positions of buckets  $B_0$  and  $B_1$ , respectively. By **(P2)** for  $B_0$  we have  $h_{i_0} \geq \frac{w}{4}$ .

**Case 1:**  $h_{i_0} < \frac{5}{8}w$ . We claim that with probability  $1 - \mathcal{O}(w^{-3})$  all keys belonging to  $B_1$  (head and tail) can be inserted and **(P1)** and **(P2)** are fulfilled afterwards. The situation is illustrated in Figure 6 on the left.

The dashed black lines show the expected trajectories of the ribbon borders for bucket  $B_1$ . The lower expected border travels in a straight line from  $(s_{i_0}, i_0)$  to the point that is  $b = i_1 - i_0$  positions to the right and  $\mathbb{E}[|\{x \in S \mid s(x) \in B_1\}|] = \frac{n}{m-w+1}b = (1 + \frac{C \log w}{8w})b = b + \frac{w}{8}$  positions below. The actual position of the border randomly fluctuates around the expectation. At each point the (vertical) deviation exceeds  $\frac{w}{8}$  with probability at most  $\mathcal{O}(w^{-5})$  by Lemma 15 **(b)**. A union bound shows that it is at most  $\frac{w}{8}$  everywhere in the bucket and thus within the region shaded red with probability at least  $1 - \mathcal{O}(w^{-3})$ . The region shaded yellow represents a “safety distance” of another  $\frac{w}{8}$  that we wish to keep from the ribbon border. Finally, the blue line is a perfect diagonal starting from  $(d_{i_0}, i_0)$ , which we claim the ribbon diagonal also follows. Due to  $s_{i_0} - d_{i_0} = h_{i_0} \geq \frac{w}{4}$  the diagonal does not intersect the lower yellow region. To see that it does not intersect the right yellow region, note that it passes through position  $(s_{i_0}, i_0 + h_{i_0})$  which, by this case’s assumption is at least  $\frac{3}{8}w$  positions left of the right ribbon border. This is sufficient to compensate for the width of the red region ( $\frac{1}{8}w$ ), the width of the yellow region ( $\frac{1}{8}w$ ) as well as the difference in slope due to overloading which accounts for a relative vertical shift of another  $\frac{1}{8}w$ .

<sup>11</sup>Though AVX512 instructions such as VPOPCNTDQ may benefit a corresponding niche.

<sup>12</sup>Our implementation is optimized for  $w = \Omega(\log n)$  and can simply use ribbon retrieval with an appropriate  $\varepsilon > 0$ .

<sup>13</sup>The key set underlying the definitions of  $s_i$  and  $d_i$  excludes the bumped keys.

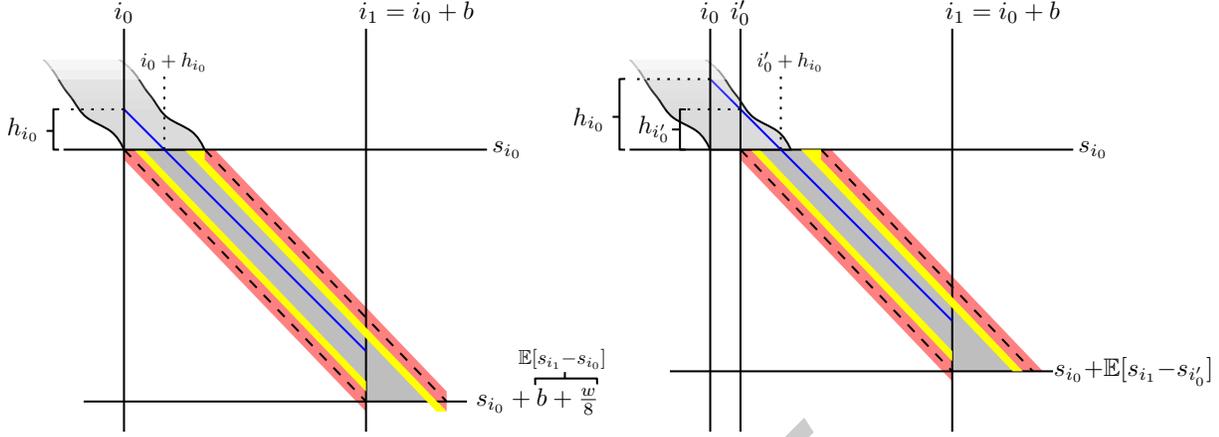


Figure 6: Situation in Cases 1 (left) of 2 (right) of the proof of Claim 13.

Now our arguments nicely interlock to show that the ribbon diagonal  $\{(d_i, i) \mid i \in B_1\}$  follows this designated path: As long as  $d$  stays away from the yellow region, it remains  $\frac{w}{8}$  positions away from the lower and right ribbon borders so each slot remains empty with probability at most  $2^{-w/8}$  by Lemma 4 and each insertion fails with probability at most  $2^{-w/8}$  by Lemma 5. Conversely, as long as no insertion fails and no slot remains empty,  $d$  proceeds along a diagonal path.

Two small caveats concern the area outside of the rectangle. We do not know the right ribbon border above row  $s_{i_0}$ ; however, those rows correspond to keys from the previous bucket and would have been bumped if their insertions failed. We also do not know the lower border to the right of  $i_1$ ; however here Lemma 5 (b) helps: We may use that the vertical distance of the ribbon diagonal to the top ribbon border is at most  $\frac{w}{8}$  to conclude that the keys with the last  $w - 1$  starting positions are also inserted successfully with probability  $2^{-w/8}$ .

This establishes **(P1)** with probability  $1 - \mathcal{O}(w^{-3})$ . Then **(P2)** follows easily: The extreme case is when both  $h_{i_0} = \frac{w}{4}$  and  $|\{x \in S \mid s(x) \in B_1\}| = b$  take the minimum permitted values. In that case we have  $h_{i_1} = \frac{w}{4}$ , so in general  $h_{i_1} \geq \frac{w}{4}$  follows.

**Case 2:  $h_{i_0} \geq \frac{5}{8}w$ .** We claim that all keys belonging to the tail of  $B_1$  can be inserted and that afterwards **(P1)** and **(P2)** are fulfilled with probability  $1 - \mathcal{O}(w^{-3})$ . In case the keys in the head of  $B_1$  can also be successfully inserted this cannot hurt **(P1)** or **(P2)** because the number of filled slots and the height could only increase due to the additional keys.<sup>14</sup>

The situation is illustrated in Figure 6 on the right. We only consider the keys in the tail of  $B_1$  which starts at position  $i'_0 + 1$  where  $i'_0 = i_0 + \frac{3}{8}w$ . Note that for  $i \in [i_0, i'_0]$  the ribbon diagonal  $(d_i, i)$  follows an ideal diagonal trajectory with probability  $1 - \mathcal{O}(2^{-\Omega(w)})$  since keys from  $B_0$  are successfully inserted and the distance to the bottom border is at least  $\frac{1}{4}w$ . This implies that all slots in the head of  $B_1$  are filled by keys from  $B_0$  and  $h_{i'_0} = h_{i_0} - \frac{3}{8}w$ . Since  $h_{i_0} \in [\frac{5}{8}w, w]$  we have  $h_{i'_0} \in [\frac{1}{4}w, \frac{5}{8}w]$ , which allows us to reason as in Case 1 to show that all slots in the tail of bucket  $B_1$  are filled and  $h_{i_1} \geq \frac{w}{4}$  with probability  $1 - \mathcal{O}(w^{-3})$ .  $\square$

**Handling failures and the first bucket.** The following claim is needed to deal with the rare cases where Claim 13 does not apply.

<sup>14</sup>Note that our analysis suggests that  $B_1$  is already full after the tail-keys are inserted, which means that the head keys can only be inserted if they all “overflow” into the next bucket.

**Claim 14.** If (P2) does not hold for a bucket  $B_0$  then with probability  $1 - \mathcal{O}(w^{-3})$  all keys of the next bucket  $B_1$  (head and tail) are successfully inserted.

*Proof.* The ribbon diagonal  $d$  starts at a height  $h_{i_0} < \frac{w}{4}$ , which is lower than desired, and might hit the lower ribbon border within  $B_1$ . However,  $d$  avoids the right ribbon border, because (recycling ideas from Case 1 of Claim 13)  $d$  would have to pierce the diagonal starting at the desired height  $\frac{w}{4}$  first and would afterwards stay on that diagonal with probability  $1 - \mathcal{O}(w^{-3})$ . This allows for some slots in  $B_1$  to remain empty but implies all keys are successfully inserted with probability  $1 - \mathcal{O}(w^{-3})$ .  $\square$

We now classify the buckets. On the one hand, there are *bad buckets* for which one of the considered events with probability  $1 - \mathcal{O}(w^{-3})$  fails to occur. On the other hand there are *good buckets* and *recovery buckets* to which Claim 13 and Claim 14 apply, respectively, and for which the corresponding high probability events occur. A *recovery sequence* is a maximal contiguous sequence of recovery buckets. Since such a sequence cannot be preceded by a good bucket, the number of recovery sequences is at most the number of bad buckets plus 1 (the first bucket is always a recovery bucket or bad). Only bad buckets contain fewer than  $b$  keys, so a recovery sequence of  $k$  buckets contains at least  $kb$  keys, all of which are inserted successfully by Claim 14. At most  $w - 1$  of these insertions fill slots after the sequence so there are at most  $w - 1$  empty slots within a recovery sequence. With  $x$  denoting the number of bad buckets, the number  $m - |P_M|$  of empty slots in total is

$$m - |P_M| \leq xb + (x + 1)(w - 1) + w - 1$$

where the last  $w - 1$  accounts for slots  $[m - w + 1, m]$  that do not belong to any bucket. The dominating term is  $xb$  so using  $\mathbb{E}[x] = \mathcal{O}(\frac{m}{b}w^{-3})$  we obtain  $\mathbb{E}[\frac{m - |P_M|}{m}] = \mathcal{O}(w^{-3})$ , which completes the proof of Lemma 12.

## 7 The Design Space of BuRR

There is a large design space for implementing BuRR. We outline it in some breadth because there was a fruitful back-and-forth between different approaches and their analysis, i.e., different approaches gradually improved the final theoretical results while insights gained by the analysis helped to navigate to simple and efficient design points. The description of the design space helps to explain some of the gained insights and might also show directions for future improvements of BuRR. To also accommodate more “hasty” readers, we nevertheless put emphasis on the the main variant of BuRR analyzed Section 6 and also move some details to appendices. We first introduce a simple generic approach and discuss concrete instantiations and refinements in separate subsections. In Appendix B, we describe further details.

As all layers of BuRR function in the same way, we need only explain the construction process of a single layer. The BuRR construction process makes the ribbon retrieval approach of Section 2 more dynamic by *bumping* ranges of keys when insertion of a row into the linear system fails by causing a linear dependence. Bumping is effected by subdividing the table for the current layer into *buckets* of size  $b$ . More concretely, bucket  $B$  contains metadata for keys  $x$  with  $s(x) \in Bb + 1..(B + 1)b$ . We also say that  $x$  is *allocated* to bucket  $B$  even though retrieving  $x$  can also involve subsequent buckets. In Section 6, we showed that it basically suffices to adaptively remove a fraction of the keys from buckets with high *load* to make the equation system solvable, i.e., to make all remaining keys retrievable from the current layer. The structure of the linear system easily absorbs most variance within and between buckets but bigger fluctuations are more efficiently handled with bumping.

**Construction.** We first sort the keys by the buckets they are allocated to. For simplicity, we set each key’s leading coefficient  $\vec{h}(x)[s(x)]$  to 1. As the starting positions are distributed randomly, this is not an issue. The ribbon solving approach is adapted to build the row-echelon form bucket by bucket from left to right (see Appendix B.4 for a discussion of variants). Consider now the solution process for a bucket  $B$ . Within  $B$ , we place the keys in some order that depends on the concrete strategy; see Appendix B.1. One useful order is from right to left based on  $s(x)$ . We store metadata that indicates one or several *groups* of keys allocated to the bucket that are bumped. These groups correspond to ranges in the placement order, not necessarily from  $s(x)$ . Which groups can be represented depends on the metadata compression strategy, which we discuss in Sections 7.1 and B.2. For example, in the right-to-left order mentioned above, it makes sense to bump keys  $x$  with  $s(x) \in Bb + 1..Bb + t$  for some threshold  $t$ , i.e., a leftmost range of slots in a bucket. This makes sense because that part of the bucket is already occupied by elements placed during construction of the previous bucket (see also Figure 3).

If placement fails for one key in a group, all keys in that group are bumped together. Such transactional grouping of insertions is possible by recording offsets of rows inserted for the current group, and clearing those rows if reverting is needed. This implies that we need to record which rows were used by keys of the current group so that we can revert their insertion if needed.

Keys bumped during the construction of a layer are recorded and passed into the construction process of the next layer. Note that additional or independent hashing erases any relationships among keys that led to bumping. In the last layer, we allocate enough space to ensure that no keys are bumped, as in the standard ribbon approach.

**Querying.** At query time, if we try to retrieve a key  $x$  from a bucket  $B$ , we check whether  $x$ ’s position in the insertion order indicates that  $x$  is in a bumped group. If not, we can retrieve  $x$  from the current layer, otherwise we have to go on to query the next layer.

**Overloading.** The tuning parameter  $\varepsilon = 1 - m/n$  is very important for space efficiency. While other linear algebra based retrieval data structures need  $\varepsilon > 0$  to work, a decisive feature of BuRR is that negative  $\varepsilon$  almost eliminates empty table slots by avoiding underloaded ranges of columns.

We discuss further aspects of the design space of BuRR in additional subsections. By default, bucket construction is *greedy*, i.e., proceeds as far as possible. Appendix B.3 presents a *cautious* variant that might have advantages in some cases. Appendix B.4 justifies our choice to construct buckets from left to right. Section 7.2 discusses how more sparse bit patterns can improve performance. Construction can be parallelized by bumping ranges of  $w$  consecutive table slots. This separates the equation system into independent blocks; see Section 7.3. In that section, we also explain external memory construction that with high probability needs only a single scan of the input and otherwise scanning and integer sorting of simple objects. The computed table and metadata can be represented in various forms that we discuss in Appendix B.5. In particular, *interleaved storage* allows to efficiently retrieve  $f(x)$  one bit at a time, which is advantageous when using BuRR as an AMQ. We can also reduce cache faults by storing metadata together with the actual table entries.

A very interesting variant of BuRR is *bump-once ribbon retrieval* ( $Bu^1RR$ ) described in Section 7.4 that guarantees that each key can be found in one of two layers.

## 7.1 Threshold-Based Bumping

Recall that BuRR places the keys one bucket  $B$  at a time and within  $B$  according to some ordering – say from right to left defined by a position in  $1..b$ . A very simple way to represent metadata is to store a single threshold  $j_B$  that remembers (or approximates) the first time this insertion process fails for  $B$  ( $j_B = 0$  if insertion does not fail). During a query, when retrieving a key  $x$  that has position  $j$  in the insertion process,  $x$  is bumped if  $j \geq j_B$ . We need  $\log b$  bits if we conflate  $b - 1$  and  $b$  onto  $b - 1$  and bump the entire bucket. It turns out that for small  $r$  (few retrieved bits), the space overhead for this

threshold dominates the overhead for empty slots. Thus, we now discuss several ways to further reduce this space.

**2-bit Thresholds and  $c$ -bit Thresholds.** metadata implies that we have to choose between four possible thresholds values. It makes sense to have support for threshold values for  $j_B = 0$  (no bumping needed in this bucket) and  $j_B = b$  (bump entire bucket). The latter is a convenient way to ensure correctness in a nonprobabilistic way, thus obviating restarts with a fresh hash function. The former threshold makes sense because the effect of obviously bumping some keys from *every* bucket could also be achieved by choosing a larger value of  $\varepsilon$ . This leaves two threshold values  $\ell, u$ , as tuning parameters. The experiments in Section 8 use linear formulas of the form  $\lceil (c_1 - c_2\varepsilon)b \rceil$  for  $\ell$  and  $u$  with the values of  $c_1$  and  $c_2$  dependent on  $w$ , but it also turns out that  $u = 2\ell$  is a good choice so that the cases 0,  $\ell$ , and  $2\ell$  can be decoded with a simple linear formula and only the case  $j_B = b$  needs special case treatment. Moreover, this approach can be generalized to  $c$ -bit thresholds, where we use  $2^c - 1$  equally spaced thresholds starting at 0 plus the threshold  $b$ .

**1<sup>+</sup>-Bit Thresholds.** The experiments performed for 2-bit thresholds indicated that actually choosing  $\ell = u$  performs quite well. Indeed, the analysis in Section 6 takes up this scheme. Moreover, both experiments and theory show that the threshold  $b$  (bump entire bucket) occurs only rarely. Hence, we considered compression schemes that store only a single bit most of the time, using some additional storage to store larger bumping thresholds. We slightly deviate from the theoretical setting by allowing arbitrary larger thresholds in order to reduce the space incurred by empty buckets. Thus, we considered a variant where the threshold values 0 (bump nothing),  $t$  (bump something), and also values  $t + 1..b$  (bump a lot) are possible but where the latter (rare) cases are stored in a separate small hash table  $H^+$  whose keys are bucket indices.

Compared to 2-bit thresholds, we get a space-time trade-off, however, because accessing the exception table  $H^+$  costs additional time (even if it is very small and will often fit into the cache). Thus, a further refinement of 1<sup>+</sup>-bit thresholds is to partition the buckets into ranges of size  $b^+$  and to store one bit for each such range to indicate whether any bucket in that range has an entry in  $H^+$ .

## 7.2 Sparse Bit Patterns

A query to a BuRR data structure as described so far needs to process about  $rw/2$  bits of data to produce  $r$  bits of output. Despite the opportunity for word parallelism, this can be a considerable cost. It turns out that BuRR also works with significantly more sparse bit patterns. We can split  $w$  bits into  $k$  groups and set just one randomly chosen bit per group. The downside of sparse patterns is that they incur more linear dependencies. This will induce more bumped keys and possibly more empty slots. Our experiments indicate that the compromise is quite interesting. For example for  $w = 64$  and  $k = 8$  we can reduce the expected number of 1-bits by a factor of four and observe faster queries for large  $r$  at the cost of a slight increase of space overhead. Figure 5 indicates that our implementation of sparse coefficient BuRR is indeed a good choice for  $r \in \{8, 16\}$ .

## 7.3 Parallelization and Memory Hierarchies

As a static data structure, queries are trivial to parallelize on a shared-memory system. Parallelizing construction can use (hard) sharding, i.e., subdividing the data structure into pieces that can be built independently. An interesting property of BuRR is that sharding can be done transparently in a way that does not affect query implementation and thus avoids performance overheads for queries. To subdivide the data structure, we set the bumping information in such a way that each piece is delimited by empty ranges of at least  $w$  columns in the constraint matrix. This has the effect that the equation system can be solved independently in parallel for ranges of columns separated by such gaps. With the parametrization

we choose for BuRR, the fraction of additionally bumped keys will be negligible as long as  $n \gg pw^2$  where  $p$  is the number of threads.

For large BuRRs that fill a significant part of main memory, space consumption during construction can be a major limitation. Once more, sharding is a simple way out – by constructing the data structure one shard at a time, the temporary space overhead for construction is only proportional to the size of a shard rather than to the size of the entire data structure.

An alternative is to consider a “proper” external memory algorithm. The input is a file  $F$  of  $n$  key–value pairs. The output is a file containing the layers of the BuRR data structure. A difficulty here is that keys could be much larger than  $\log n$  bits and that random accesses to keys is much more expensive than just scanning or sorting. We therefore outline an algorithm that, with high probability, reads the key file only once. The trick is to first substitute keys by *master hash codes* (MHCs) with  $c \log n$  bits for a constant  $c > 2$ . Using the well-known calculations for the birthday problem, this size suffices that the MHCs are unique with high probability. If a collision should occur, the entire construction is restarted.<sup>15</sup> Otherwise, the keys are never accessed again – all further hash bits are obtained by applying secondary hash functions to the MHC.<sup>16</sup>

Construction then begins by scanning the input file  $F$  and producing a stream  $F_1$  of pairs (MHC, function value)<sup>17</sup>. Construction at layer  $i$  amounts to reading a stream  $F_i$  of MHC–value pairs. This stream  $F_i$  is then sorted by a bucket ID that is derived from the MHCs. A collision is detected when two pairs with the same MHC show up. These are easy to detect since they will be sorted to the same bucket and the same column within that bucket. Constructing the row-echelon form (REM) then amounts to simultaneously scanning  $F_i$ , the REM and the right-hand side. At no time during this process do we need to keep more than two buckets worth of data in main memory. Bumped MHC–value pairs are output as a stream  $F_{i+1}$  that is the input for construction of the next layer. Back-substitution amounts to a backward scan of the REM and the right-hand side – producing the table for layer  $i$  as an output.

Overall, the I/O complexity is the I/Os for scanning  $n$  keys plus sorting  $\mathcal{O}(n)$  items consisting of a constant number of machine words ( $\mathcal{O}(\log n)$  bits). The fact that there are multiple layers contributes only a small constant factor to the sorting term since these layers are shrinking geometrically with a rather large shrinking factor.

#### 7.4 Bu<sup>1</sup>RR: Accessing Only Two Layers in the Worst Case

BuRR as described so far has worst-case constant access time when the the number of layers is constant (our analysis and experiments use four layers). However, for real-time applications, the resulting worst case might be a limitation. Here, we describe how the worst-case number of accessed layers can be limited to two. The idea is quite simple: Rather than mapping all keys to the first layer, we map the keys to all layers using a distribution still to be determined. We now guarantee that a key originally mapped to layer  $i$  is either retrieved there or from layer  $i + 1$ . A query for key  $x$  now proceeds as follows. First the primary layer  $i(x)$  for that key is determined. Then the bumping information of layer  $i$  is queried to find out whether  $i$  is bumped. If not,  $x$  is retrieved from layer  $i$ , otherwise it is retrieved from layer  $i + 1$  *without* querying the bumping information for layer  $i + 1$ .

For constructing layer  $i$ , the input consists of keys  $E'_i$  bumped from layer  $i - 1$  ( $E'_1 = \emptyset$ ) and keys  $E_i$  having layer  $i$  as their primary layer ( $E_i = \emptyset$  for the last layer). First, the bumped keys  $E'_i$  are inserted

<sup>15</sup>For use in AMQs, restarts are not needed since duplicate MHCs lead to identical equations that will be ignored as redundant by the ribbon solver.

<sup>16</sup>We use a (very fast) linear congruential mapping [34, 22] that, with some care, (multiplier congruent 1 modulo 4 and an odd summand) even defines a bijective mapping [34]. We also tried linear combinations of two parts of the MHC [33] which did not work well however for a 64 bit MHC.

<sup>17</sup>When used as an AMQ, the function value need not be stored since it can be derived from the MHC.

bucket by bucket. In this phase, a failure to insert a key causes construction to fail. Then the keys  $E_i$  are processed bucket by bucket as before, recording bumped keys in  $E'_{i+1}$ .

The size of layer  $i$  can be chosen as  $(1 + \varepsilon)(|E'_i| + |E_i|)$  to achieve the same overloading effect as in basic BuRR. An exception is the last layer  $i^*$  where we choose the size large enough so that construction succeeds with high probability. Note that if  $|E_i|$  shrinks geometrically, we can choose  $i^*$  such that  $|E_{i^*}|$  is negligible.

Except in the last layer, construction can only fail due to keys already bumped, which is a small fraction even for practical  $w$ . In the unlikely<sup>18</sup> case of construction failing, construction of that layer can be retried with more space. Tests suggest that increasing space by a factor of  $\frac{w+1}{w}$  has similar or better construction success probability than a fresh hash function.

A simpler Bu<sup>1</sup>RR construction has layers of predetermined sizes, all in one linear system. Construction reliability and/or space efficiency are reduced slightly because of reduced adaptability. For moderate  $n \approx w^3$ , layers of uniform size can work well, especially if the last layer is of variable size.

Our implementation of Bu<sup>1</sup>RR, tested to scale to billions of keys, uses layers with sizes shrinking by a factor of two, each with a power of two number of buckets. To a first approximation, the primary layer  $i(x)$  of a key is simply the number of leading zeros in an appropriate hash value, up to the maximum layer. To consistently saturate construction with expected overload of  $\alpha = -\varepsilon$ , this is modified with a bias for the first layer. A portion ( $\alpha$ ) of values with a leading 0 (not first layer) are changed to a leading 1 (re-assigned to first layer), so  $|E_0| \approx (1 + \alpha)2^{-1}m$  and other  $|E_i| \approx (1 - \alpha)2^{-(i+1)}m$ . With bumped entries,  $|E'_i| \approx \alpha 2^{-i}m$ , expected overload is consistent through the layers:  $|E_i| + |E'_i| \approx (1 + \alpha)2^{-(i+1)}m$ .

## 8 Experiments

**Implementation Details.** We implemented BuRR in C++ using template parameters that allow us to navigate a large part of the design space mapped in Section 7. We use sequential construction using 64-bit master-hash-codes (MHCs) so that the input keys themselves are hashed only once. Linear congruential mapping is used to derive more pseudo-random bits from the MHC. When not otherwise mentioned, our default configuration is BuRR with left-to-right processing of buckets, aggressive right-to-left insertion within a bucket, threshold-based bumping, interleaved storage of the solution  $Z$ , and separately stored metadata. The data structure has four layers, the last of which uses  $w' := \min(w, 64)$  and  $\varepsilon \geq 0$ , where  $\varepsilon$  is increased in increments of 0.05 until no elements are bumped. For 1<sup>+</sup>-bit, we choose  $t := \lceil -2\varepsilon b + \sqrt{b/(1 + \varepsilon)}/2 \rceil$  and  $\varepsilon := -2/3 \cdot w/(4b + w)$ . For 2-bit, parameter tuning showed that  $\ell := \lceil (0.13 - \varepsilon/2)b \rceil$ ,  $u := \lceil (0.3 - \varepsilon/2)b \rceil$ , and  $\varepsilon := -3/w$  work well for  $w = 32$ ; for  $w \geq 64$ , we use  $\ell = \lceil (0.09 - 3\varepsilon/4)b \rceil$ ,  $u = \lceil (0.22 - 1.3\varepsilon)b \rceil$ , and  $\varepsilon := -4/w$ .

In addition, there is a prototypical implementation of Bu<sup>1</sup>RR from [22]; see Section 7.4. Both BuRR and Bu<sup>1</sup>RR build on the same software for ribbon solving from [22]. For validation we extend the experimental setup used for Cuckoo and Xor filters [29], with our code available at [https://github.com/lorenzhs/fastfilter\\_cpp](https://github.com/lorenzhs/fastfilter_cpp) and <https://github.com/lorenzhs/BuRR>.

**Experimental Setup.** All experiments were run on a machine with an AMD EPYC 7702 processor with 64 cores, a base clock speed of 2.0 GHz, and a maximum boost clock speed of 3.35GHz. The machine is equipped with 1 TiB of DDR4-3200 main memory and runs Ubuntu 20.04. We use the `clang++` compiler in version 11.0 with optimization flags `-O3 -march=native`. During sequential experiments, only a single core was active at any time to minimize interference.

We looked at different input sizes  $n \in \{10^6, 10^7, 10^8\}$ . Like most studies in this area, we first look at a **sequential** workload on a powerful processor with a considerable number of cores. However, this seems

<sup>18</sup>We do not have a complete analysis of this case yet but believe that our analysis in Section 6 can be adapted to show that the construction process will succeed with high probability for  $w = \Omega(\log n)$ .

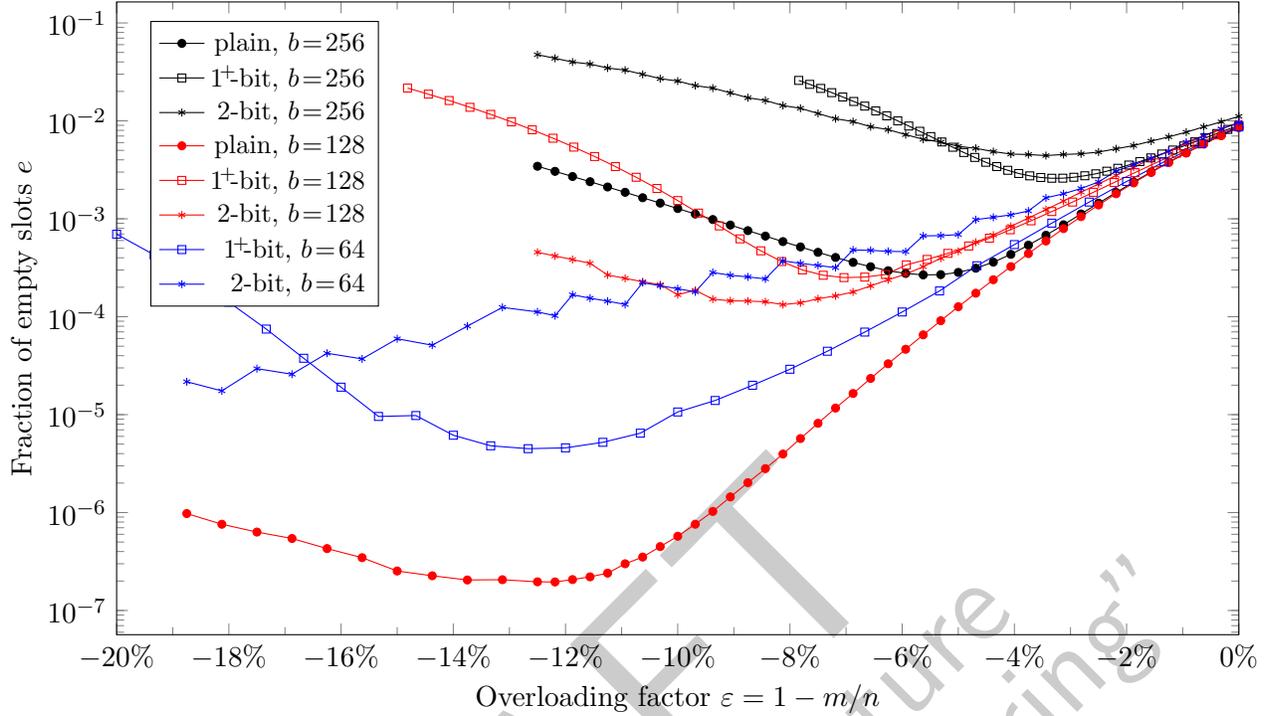


Figure 7: Fraction of empty slots for various configurations of bumped ribbon retrieval with  $w = 64$ , depending on the overloading factor  $\varepsilon$ .

unrealistic since in most applications, one would not let most cores lay bare but use them. Unless these cores have a workload with very high locality this would have a considerable effect on the performance of the AMQs. We therefore also look at a scenario that might be the opposite extreme to a sequential unloaded setting. We run the benchmarks on all available hardware threads in **parallel**. Construction builds many AMQs of size  $n$  in parallel. Queries select AMQs randomly. This emulates a large AMQ that is parallelized using sharding and puts very high load on the memory system.

### Space Overhead of BuRR

Figure 7 plots the fraction  $e$  of empty slots of BuRR for  $w = 64$  and several combinations of bucket size  $b$  and different threshold compression schemes. Similar plots are given in the appendix in Figure 10 for  $w = 32$ ,  $w = 128$ , and for  $w = 64$  with sparse coefficients. Note that (for an infinite number of layers), the overhead is about  $o = e + \mu/(rb(1 - e))$  where  $r$  is the number of retrieved bits and  $\mu$  is the number of metadata bits per bucket. Hence, at least when  $\mu$  is constant, the overhead is a monotonic function in  $e$  and thus minimizing  $e$  also minimizes overhead.

We see that for small  $|\varepsilon|$ ,  $e$  decreases exponentially. For sufficiently small  $b$ ,  $e$  can get almost arbitrarily small. For fixed  $b > w$ ,  $e$  eventually reaches a local minimum because with threshold-based compression, a large overload enforces large thresholds ( $> w$ ) and thus empty regions of buckets. Which actual configuration to choose depends primarily on  $r$ . Roughly, for larger  $r$ , more and more metadata bits (i.e., small  $b$ , higher resolution of threshold values) can be invested to reduce  $e$ . For fixed  $b$  and threshold compression scheme, one can choose  $\varepsilon$  to minimize  $e$ . One can often choose a larger  $\varepsilon$  to get slightly better performance due to less bumping with little impact on  $o$ . Perhaps the most delicate tuning parameters are the thresholds to use for 2-bit and 1+-bit compression (see Section 7.1). Indeed, in Figure 7 1+-bit compression has lower  $e$  than 2-bit compression for  $b = 64$  but higher  $e$  for  $b = 128$ . We expect that 2-bit compression could always achieve smaller  $e$  than 1+-bit compression, but we have not

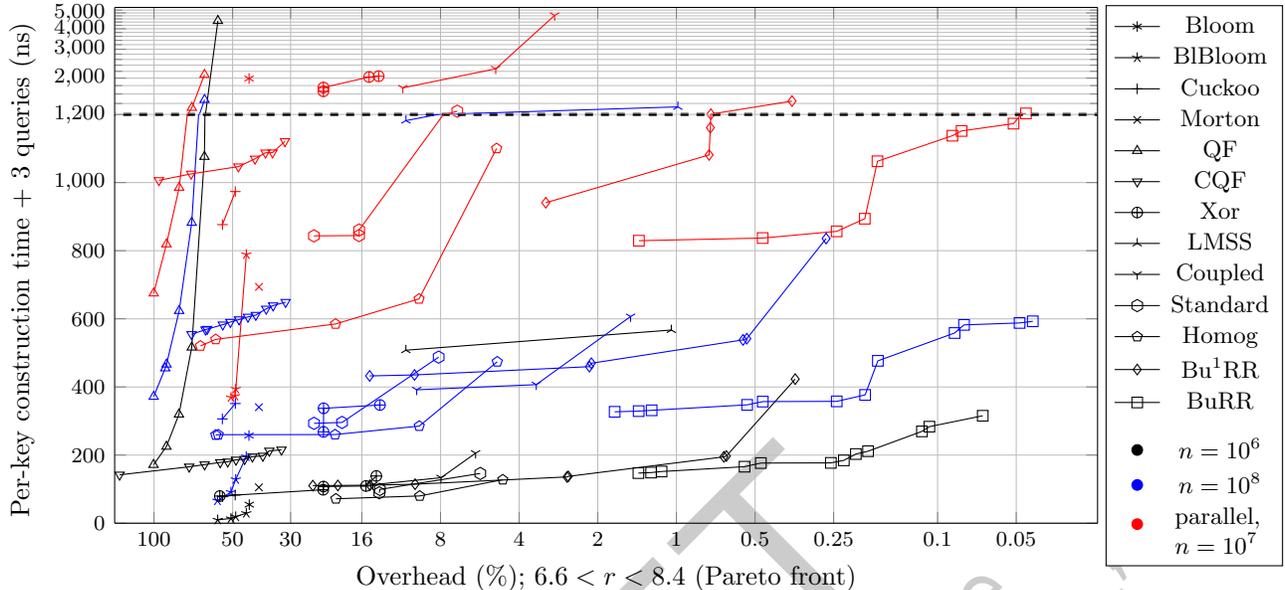


Figure 8: Performance–overhead trade-off for false-positive rate  $2^{-r}$ , approximately 0.3% to 1%, for different AMQs and inputs. For each category of approaches, only points are shown that are not dominated by variants in the same category. Sequential benchmarks use a single AMQ of size  $n$  while the parallel benchmark uses 1280 AMQs and 64 cores. Logarithmic vertical axis above 1200 ns.

found choices for the threshold values that always ensure this. Table 4 in Appendix C summarizes key parameters of some selected BuRR configurations.

In all following experiments, we use  $b = 2^{\lfloor w^2/(2 \log_2 w) \rfloor}$  for uncompressed and 2-bit compressed thresholds, and  $b = 2^{\lfloor w^2/(4 \log_2 w) \rfloor}$  when using 1<sup>+</sup>-bit threshold compression.

### Performance of BuRR Variants

We have performed experiments with numerous configurations of BuRR. See Table 2 in Appendix C for a small sample; we will publish a complete list online. The scatter plot in Figure 8 summarizes the performance–overhead trade-off for  $r \approx 8$ . Plots for different values of  $r$  and for construction and query times separately are in Appendix C (Figures 11 to 15).

A small **ribbon width** of  $w = 16$  is feasible but does not pay off with respect to performance because its high bumping rates drive up the expected number of layers accessed. Choosing  $w = 32$  yields the best performance in many configurations but the penalty for going to  $w = 64$  is very small while reducing overheads. In contrast,  $w = 128$  has a large performance penalty for little additional gain – overheads far below 1% are already possible with  $w = 64$ . Thus, on a 64-bit machine,  $w = 64$  seems the logical choice in most situations.

With respect to performance, **1<sup>+</sup>-bit** compression is slightly slower than **2-bit** compression or uncompressed thresholds but not by a large margin. However, **1<sup>+</sup>-bit** achieves the lowest overheads. **Interleaved** table representation (see Appendix B.5) is mostly faster than **contiguous** representation. This might change for sufficiently large  $r$  and use of advanced bit-manipulation or SIMD instructions. Nevertheless, **sparse coefficients** with 8 out of 64 bits using contiguous representation achieve significantly better query performance than the best dense variant with comparable or better overhead when contiguous storage is efficiently addressable, i.e.,  $r$  is a multiple of 8.

**Bu<sup>1</sup>RR** is around 20% slower than BuRR and also somewhat inferior with respect to the achieved overheads. This may in part be due to less extensive parameter tuning. When worst-case guarantees

matter, Bu<sup>1</sup>RR thus might be a good choice.

In a parallel scaling experiment with  $10^{10}$  keys, construction and queries both scaled well, shown in Figure 16 in Appendix C. Constructing many AMQs in parallel achieves speedups of 65 – 71 depending on the configuration when using all 64 cores plus hyperthreads of our machine (50 without hyperthreading). Individual query times are around 15 % higher than sequentially when using all cores, and 50 % higher when using all hyperthreads. This approximately matches the speedups for construction.

## Comparison with Other Retrieval and AMQ Data Structures

To compare BuRR with other approaches, we performed an extensive case study of different **AMQs** and retrieval data structures. To compare space overheads, we compare  $r$ -bit retrieval data structures to AMQs with false positive rate  $2^{-r}$ . Our benchmark extends the framework used to evaluate Cuckoo and Xor filters [29], with our modified version available at [https://github.com/lorenzhs/fastfilter\\_cpp](https://github.com/lorenzhs/fastfilter_cpp). In addition to adding our implementations of standard ribbon, homogeneous ribbon, BuRR, and Bu<sup>1</sup>RR, we integrated existing implementations of Quotient Filters [38, 39] and LMSS, Coupled, GOV, 2-block, and BPZ retrieval [48, 49]. We extended the implementations of LMSS, Coupled, and BPZ to support the cases of  $r = 8$  and  $r = 16$  in addition to one-bit retrieval.<sup>19</sup>

We also added a parallel version of the benchmark where each thread constructs a number of AMQs independently, but queries access all of them randomly. In the Cuckoo filter implementation, we replaced calls to `rand()` with a `std::mt19937` Mersenne Twister to eliminate a parallel bottleneck. The implementation of LMSS cannot be run simultaneously in multiple threads of the same process and was excluded from the parallel benchmark.

Both the sequential and the parallel benchmark use three query workloads: positive queries, negative queries, and a mixed set containing equal numbers of positive and negative queries. We report many results in the form of construction time per key plus the time for one query from each of the three sets, measured by dividing the running time for construction plus  $n$  queries of each type by the input size  $n$ . This metric is a reasonable tradeoff between construction and queries; we provide figures for the individual components in Appendix C (Figures 13 to 15).

Once more, the scatter plot in Figure 8 summarizes the performance–overhead trade-off for  $r \approx 8$ ; other values of  $r$  are covered in Appendix C (Figures 11 and 12). In addition, Figure 5 gives an overview of the fastest approach for different values of  $r$  and overhead. We now discuss different approaches progressing from high to low space overhead.

**Bloom Filters Variants:** Plain Bloom filters [7] set  $k \sim \log(1/\varphi)$  random bits in a table for each element in order to achieve false-positive rate  $\varphi$ . They are the most well-known and perhaps most widely used AMQ. However, they have an inherent space overhead of at least 44 % compared to the information-theoretic optimum. Moreover, for large inputs they cause almost  $k$  cache faults for each insertion and positive query. **Blocked Bloom filters** [46] are faster because they set all of the  $k$  bits in the same cache block. The downside is that this increases the false-positive rate, in particular for large  $k$ . This can be slightly reduced using **two blocks**.

**Cuckoo Filters** [24] store a random fingerprint for each key (similar to retrieval-based AMQs). However, to allow good space efficiency, several positions need to be possible. This introduces an intrinsic space overhead of a few bits per element that is further increased by some empty slots that are required to allow fast insertions. The latter overhead is reduced in **Morton filters** [11] which can be viewed as a compressed representation of cuckoo filters. In our measurements, cuckoo and Morton filters are the most space efficient dynamic AMQ for small  $\varphi$ , but are otherwise outperformed by other constructions.

**Quotient Filters (QF)** [5] can be viewed as a compressed representation of a Bloom filter with a single hash function. QFs support not only insertions but also deletions and counting. Similar to cuckoo

---

<sup>19</sup>This is easy for peeling-based approaches, but far more work would be required to do the same for GOV and 2-block.

filters, they incur an overhead of a few bits per element (2–3 depending on the implementation) plus a multiplicative overhead due to empty entries. Search time grows fairly quickly as the multiplicative overhead approaches zero. **Counting Quotient Filters (CQF)** [44] mitigate this dependence at the cost of less locality for low fill degrees. Overall, quotient filters are good if their rich set of operations is needed but cannot compete with alternatives (e.g., blocked Bloom filters and cuckoo filters) for dynamic AMQs without deletions. Compared to static AMQs, they have comparable or slower speed than BuRR but two orders of magnitude higher overhead.

**Xor filter/retrieval.** **Xor Filters** and **Xor+ Filters** [30] are a recent implementation of peeling-based retrieval that can reduce overhead to 23 % and 14 %, respectively. An earlier implementation [10] is 35–90 % slower to construct and also has slightly slower queries. **Fuse** filters [19], a variant of Xor, achieve higher loads and are fast to construct if  $n$  is not too large, but construction becomes slow for large  $n$  and parallel queries are less efficient than for plain Xor filters. In the plots we only show the Pareto-optimal variants in each case and call all of them Xor in the following and in Section 1. Among the tried retrieval data structures, Xor has the fastest queries for sequential settings but is otherwise dominated by BuRR, which has one to two orders of magnitude smaller overheads.

**Low overhead peeling-based retrieval.** By using several hash functions in a nonuniform way, peeling-based retrieval data structures can in principle achieve arbitrarily small overheads and thus could be viewed as the primary competitors of ribbon-based data structures. **LMSS** [37] was originally proposed as an error-correcting code but can also be used for retrieval. However, in the experiments it is clearly dominated by BuRR. The more recent **Coupled** peeling approach [50] can achieve Pareto-optimal query times for large sequential inputs but is otherwise dominated by ribbon-based data structures. Coupled has faster construction times than LMSS but in that respect is several times slower than BuRR for large  $n$  and in our parallel benchmark, even when it is allowed an order of magnitude more overhead. Nevertheless, when disregarding ribbon-based data structures, Coupled comes closest to a practical retrieval data structure with very low overhead. For small inputs ( $n = 10^6$ ),  $r \approx 16$  and overhead between 8 and 15 %, it is even the fastest AMQ in our benchmark (see Figure 5). Perhaps for large  $r$  and by engineering faster construction algorithms, Coupled could become more competitive in the future.

**Standard Ribbon** can achieve overhead around 10 %. However, it often fails for large  $n$ , requiring larger space overhead when used without sharding on top of it. For AMQs this can be elegantly remedied using **homogeneous ribbon filters**. Thus, in the heatmap, homogeneous ribbon occupies the area between (blocked) Bloom filters and BuRR and its variants. However, the performance advantage over BuRR in parallel and large settings is not very large (typically 20 %).

**BuRR** and its variants take the entire right part of the heatmap. Compared to the best competitor – homogeneous ribbon filters – overhead drops from around 10 % to well below 1 % at a moderate performance penalty. In particular, due to BuRR’s high locality, performance is even better than for successful competitors like Xor, Cuckoo, or Bloom filters.

**2-block** [18] can be viewed as a generalization of ribbon-based approaches that use two rather than one block of nonzeros in each row of the constraint matrix. Unfortunately this makes the equation system much more difficult to solve. This implies very expensive construction even when aggressively using the sharding trick. In our experiments, an implementation by Walzer [49] for  $r = 1$  achieves smaller overhead than BuRR with  $w = 128$  at the price of an order of magnitude larger construction time. It is however likely that a BuRR implementation able to handle  $w = 256$  would dominate 2-block.

**Techniques not tried.** There are a few interesting retrieval data structures for which we had no available implementation. **FiRe** [43] is likely to be the fastest existing retrieval data structure and also supports updates to function values as well as a limited form of insertions. FiRe maps elements to buckets fitting into a cache line. Per-bucket metadata is used to uniquely map some elements to data slots available in

the bucket while bumping the remaining elements. This requires a constant number of metadata bits per input element (around 4) and thus implies an overhead two orders of magnitude larger than BuRR. Additionally, the only known implementation of FiRe is closed source and owned by SAP, and was not available to us.

We are not aware of implementations of the lookup-table based approaches [45, 4] and do not view them as practically promising for the reasons discussed in Section 1.4.

## 9 Conclusion

We present a new static retrieval data structure which achieves excellent trade-offs between high space efficiency and fast running times. Like [20], the construction involves solving a system of linear equations over the two-element field, where each row contains a block of random bits in a random position. We augment this setup with a load balancing idea: The construction algorithm is given less space than required to store all keys but is permitted to *bump* sets of keys in overloaded regions to a fallback data structure. There are two major benefits: Firstly, space utilization is significantly improved, and secondly, the width  $w$  of blocks need no longer scale with  $\mathcal{O}(\log n)$ , which improves running times. We give a self-contained analysis in terms of an easy-to-grasp visual concept (the ribbon diagonal) that we believe will readily extend to other ribbon variants (see also Section 4).

In an extensive experimental evaluation, we achieve overheads around 1% at running times with which previously, only overheads of more than 10% were achievable. Overheads below 0.1% are possible at modest additional cost.

## Acknowledgements

We thank Martin Dietzfelbinger for early contributions to this line of research. We thank others at Facebook for their supporting roles, including Jay Zhuang, Siying Dong, Shrikanth Shankar, Affan Dar, and Ramkumar Vadivelu.

## References

- [1] Martin Aumüller, Martin Dietzfelbinger, and Michael Rink. Experimental variations of a theoretically good retrieval data structure. In *Proc. 17th ESA*, pages 742–751, 2009. doi:10.1007/978-3-642-04128-0\_66.
- [2] Michael Axtmann, Sascha Witt, Daniel Ferizovic, and Peter Sanders. Engineering in-place (shared-memory) sorting algorithms. *CoRR*, 2020. arXiv:2009.13569.
- [3] Djamel Belazzougui, Paolo Boldi, Giuseppe Ottaviano, Rossano Venturini, and Sebastiano Vigna. Cache-oblivious peeling of random hypergraphs. In *Proc. DCC*, pages 352–361, 2014. doi:10.1109/DCC.2014.48.
- [4] Djamel Belazzougui and Rossano Venturini. Compressed static functions with applications. In Sanjeev Khanna, editor, *Proc. 24th SODA*, pages 229–240. SIAM, 2013. doi:10.1137/1.9781611973105.17.
- [5] Michael A. Bender, Martin Farach-Colton, Rob Johnson, Russell Kraner, Bradley C. Kuszmaul, Dzejla Medjedovic, Pablo Montes, Pradeep Shetty, Richard P. Spillane, and Erez Zadok. Don't thrash: How to cache your hash on flash. *Proc. VLDB Endow.*, 5(11):1627–1637, 2012. doi:10.14778/2350229.2350275.

- [6] Valerio Bioglio, Marco Grangetto, Rossano Gaeta, and Matteo Sereno. On the fly gaussian elimination for LT codes. *Communications Letters, IEEE*, 13(12):953 – 955, 12 2009. doi:10.1109/LCOMM.2009.12.091824.
- [7] Burton H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Commun. ACM*, 13(7):422–426, 1970. doi:10.1145/362686.362692.
- [8] Fabiano Cupertino Botelho. *Near-Optimal Space Perfect Hashing Algorithms*. PhD thesis, Federal University of Minas Gerais, 2008. URL: <http://cmph.sourceforge.net/papers/thesis.pdf>.
- [9] Fabiano Cupertino Botelho, Rasmus Pagh, and Nivio Ziviani. Simple and space-efficient minimal perfect hash functions. In *Proc. 10th WADS*, pages 139–150, 2007. doi:10.1007/978-3-540-73951-7\_13.
- [10] Fabiano Cupertino Botelho, Rasmus Pagh, and Nivio Ziviani. Practical perfect hashing in nearly optimal space. *Inf. Syst.*, 38(1):108–131, 2013. doi:10.1016/j.is.2012.06.002.
- [11] Alex D. Breslow and Nuwan Jayasena. Morton filters: fast, compressed sparse cuckoo filters. *VLDB J.*, 29(2-3):731–754, 2020. doi:10.1007/s00778-019-00561-0.
- [12] Andrei Z. Broder and Anna R. Karlin. Multilevel adaptive hashing. In David S. Johnson, editor, *Proc. 1st SODA*, pages 43–53. SIAM, 1990. URL: <http://dl.acm.org/citation.cfm?id=320176.320181>.
- [13] Bernard Chazelle, Joe Kilian, Ronitt Rubinfeld, and Ayellet Tal. The Bloomier filter: An efficient data structure for static support lookup tables. In *Proc. 15th SODA*, pages 30–39. SIAM, 2004. URL: <http://dl.acm.org/citation.cfm?id=982792.982797>.
- [14] Colin Cooper. On the rank of random matrices. *Random Structures & Algorithms*, 16(2):209–232, 2000. doi:10.1002/(SICI)1098-2418(200003)16:2<209::AID-RSA6>3.0.CO;2-1.
- [15] Martin Dietzfelbinger, Andreas Goerdt, Michael Mitzenmacher, Andrea Montanari, Rasmus Pagh, and Michael Rink. Tight thresholds for cuckoo hashing via XORSAT. In *Proc. 37th ICALP (1)*, pages 213–225, 2010. doi:10.1007/978-3-642-14165-2\_19.
- [16] Martin Dietzfelbinger and Rasmus Pagh. Succinct data structures for retrieval and approximate membership (extended abstract). In *Proc. 35th ICALP (1)*, pages 385–396, 2008. doi:10.1007/978-3-540-70575-8\_32.
- [17] Martin Dietzfelbinger and Michael Rink. Applications of a splitting trick. In *Proc. 36th ICALP (1)*, pages 354–365, 2009. doi:10.1007/978-3-642-02927-1\_30.
- [18] Martin Dietzfelbinger and Stefan Walzer. Constant-time retrieval with  $O(\log m)$  extra bits. In *Proc. 36th STACS*, pages 24:1–24:16, 2019. doi:10.4230/LIPIcs.STACS.2019.24.
- [19] Martin Dietzfelbinger and Stefan Walzer. Dense peelable random uniform hypergraphs. In *Proc. 27th ESA*, pages 38:1–38:16, 2019. doi:10.4230/LIPIcs.ESA.2019.38.
- [20] Martin Dietzfelbinger and Stefan Walzer. Efficient Gauss elimination for near-quadratic matrices with one short random block per row, with applications. In *Proc. 27th ESA*, pages 39:1–39:18, 2019. doi:10.4230/LIPIcs.ESA.2019.39.
- [21] Peter C. Dillinger. RocksDB FastLocalBloom, 2019. URL: [https://github.com/facebook/rocksdb/blob/master/util/bloom\\_impl.h](https://github.com/facebook/rocksdb/blob/master/util/bloom_impl.h).

- [22] Peter C. Dillinger and Stefan Walzer. Ribbon filter: practically smaller than Bloom and Xor. *CoRR*, 2021. arXiv:2103.02515.
- [23] Olivier Dubois and Jacques Mandler. The 3-XORSAT threshold. In *Proc. 43rd FOCS*, pages 769–778, 2002. doi:10.1109/SFCS.2002.1182002.
- [24] Bin Fan, David G. Andersen, and Michael Kaminsky. Cuckoo filter: Better than Bloom. *login.*, 38(4), 2013. URL: <https://www.usenix.org/publications/login/august-2013-volume-38-number-4/cuckoo-filter-better-bloom>.
- [25] Dimitris Fotakis, Rasmus Pagh, Peter Sanders, and Paul G. Spirakis. Space efficient hash tables with worst case constant access time. *Theory Comput. Syst.*, 38(2):229–248, 2005. doi:10.1007/s00224-004-1195-x.
- [26] Nikolaos Fountoulakis and Konstantinos Panagiotou. Sharp load thresholds for cuckoo hashing. *Random Struct. Algorithms*, 41(3):306–333, 2012. doi:10.1002/rsa.20426.
- [27] Marco Genuzio, Giuseppe Ottaviano, and Sebastiano Vigna. Fast scalable construction of (minimal perfect hash) functions. In *Proc. 15th SEA*, pages 339–352, 2016. doi:10.1007/978-3-319-38851-9\_23.
- [28] Marco Genuzio, Giuseppe Ottaviano, and Sebastiano Vigna. Fast scalable construction of ([compressed] static — minimal perfect hash) functions. *Information and Computation*, 2020. doi:10.1016/j.ic.2020.104517.
- [29] Thomas Mueller Graf and Daniel Lemire. fastfilter\_cpp, 2019. URL: [https://github.com/FastFilter/fastfilter\\_cpp](https://github.com/FastFilter/fastfilter_cpp).
- [30] Thomas Mueller Graf and Daniel Lemire. Xor filters: Faster and smaller than Bloom and cuckoo filters. *ACM J. Exp. Algorithmics*, 25:1–16, 2020. doi:10.1145/3376122.
- [31] Jóhannes B. Hreinsson, Morten Krøyer, and Rasmus Pagh. Storing a compressed function with constant time access. In *Proc. 17th ESA*, pages 730–741, 2009. doi:10.1007/978-3-642-04128-0\_65.
- [32] Svante Janson and Malwina J. Luczak. A simple solution to the  $k$ -core problem. *Random Struct. Algorithms*, 30(1-2):50–62, 2007. doi:10.1002/rsa.20147.
- [33] Adam Kirsch and Michael Mitzenmacher. Less hashing, same performance: Building a better Bloom filter. *Random Struct. Algorithms*, 33(2):187–218, 2008. doi:10.1002/rsa.20208.
- [34] D. E. Knuth. *The Art of Computer Programming — Seminumerical Algorithms*, volume 2. Addison Wesley, 2nd edition, 1981.
- [35] Harald Lang, Thomas Neumann, Alfons Kemper, and Peter A. Boncz. Performance-optimal filtering: Bloom overtakes cuckoo at high-throughput. *Proc. VLDB Endow.*, 12(5):502–515, 2019. doi:10.14778/3303753.3303757.
- [36] Marc Lelarge. A new approach to the orientation of random hypergraphs. In *Proc. 23rd SODA*, pages 251–264. SIAM, 2012. doi:10.1137/1.9781611973099.23.
- [37] Michael Luby, Michael Mitzenmacher, Mohammad Amin Shokrollahi, and Daniel A. Spielman. Efficient erasure correcting codes. *IEEE Trans. Inf. Theory*, 47(2):569–584, 2001. doi:10.1109/18.910575.

- [38] Tobias Maier. lpqfilter, 2020. URL: <https://github.com/TooBiased/lpqfilter>.
- [39] Tobias Maier, Peter Sanders, and Robert Williger. Concurrent expandable AMQs on the basis of quotient filters. In *Proc. 18th SEA*, pages 15:1–15:13, 2020. doi:10.4230/LIPIcs.SEA.2020.15.
- [40] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press, New York, NY, USA, 2nd edition, 2017.
- [41] Michael Molloy. Cores in random hypergraphs and Boolean formulas. *Random Struct. Algorithms*, 27(1):124–135, 2005. doi:10.1002/rsa.20061.
- [42] Ingo Müller, Cornelius Ratsch, and Franz Färber. Adaptive string dictionary compression in in-memory column-store database systems. In *Proc. 17th EDBT*, pages 283–294, 2014. doi:10.5441/002/edbt.2014.27.
- [43] Ingo Müller, Peter Sanders, Robert Schulze, and Wei Zhou. Retrieval and perfect hashing using fingerprinting. In *Proc. 14th SEA*, pages 138–149, 2014. doi:10.1007/978-3-319-07959-2\_12.
- [44] Prashant Pandey, Michael A. Bender, Rob Johnson, and Rob Patro. A general-purpose counting filter: Making every bit count. In *Proc. 2017 SIGMOD*, pages 775–787. ACM, 2017. doi:10.1145/3035918.3035963.
- [45] Ely Porat. An optimal Bloom filter replacement based on matrix solving. In *Proc. 4th CSR*, pages 263–273, 2009. doi:10.1007/978-3-642-03351-3\_25.
- [46] Felix Putze, Peter Sanders, and Johannes Singler. Cache-, hash-, and space-efficient Bloom filters. *ACM Journal of Experimental Algorithmics*, 14, 2009. doi:10.1145/1498698.1594230.
- [47] J. S. Vitter and E. A. M. Shriver. Algorithms for parallel memory, I: Two level memories. *Algorithmica*, 12(2/3):110–147, 1994. doi:10.1007/BF01185207.
- [48] Stefan Walzer. Experimental comparison of retrieval data structures, 2020. URL: <https://github.com/sekti/retrieval-test>.
- [49] Stefan Walzer. *Random Hypergraphs for Hashing-Based Data Structures*. PhD thesis, Technische Universität Ilmenau, 2020. URL: [https://www.db-thueringen.de/receive/dbt\\_mods\\_00047127](https://www.db-thueringen.de/receive/dbt_mods_00047127).
- [50] Stefan Walzer. Peeling close to the orientability threshold: Spatial coupling in hashing-based data structures. In *Proc. 32nd SODA*, pages 2194–2211. SIAM, 2021. doi:10.1137/1.9781611976465.131.

## A Chernoff Bounds

**Lemma 15.** Let  $(X_j)_{j \in [N]}$  be i.i.d. indicator random variables,  $X := \sum_{j \in [N]} X_j$  and  $\mu := \mathbb{E}X$ .

(a) For  $\delta \in [0, 1]$  we have  $\Pr[|X - \mu| \geq \delta\mu] \leq 2 \exp(-\delta^2\mu/3)$ .

(b) There exists  $C > 0$  such that for any  $w \in \mathbb{N}$  and  $\mu \leq \frac{2w^2}{C \log w}$  we have  $\Pr[|X - \mu| \geq \frac{w}{8}] = \mathcal{O}(w^{-5})$ .

*Proof.* (a) This combines standard Chernoff bounds on the probability of  $\{X \geq (1 + \delta)\mu\}$  and  $\{X \leq (1 - \delta)\mu\}$  as found for instance in [40, Chapter 4].

(b) We set  $\delta = \frac{w}{8\mu}$  and apply (a). This gives

$$\Pr[|X - \mu| \geq \frac{w}{8}] \leq 2 \exp(-\delta^2 \mu / 3) = 2 \exp(-\frac{w^2}{192\mu}) \leq 2 \exp(-\frac{C \log w}{384}) = 2w^{-C/384}.$$

Choosing  $C = 1920$  achieves the desired bound.<sup>20</sup> □

## B More on the Design Space of BuRR

### B.1 Different Insertion Orders Within a Bucket

**Right-to-left.** Assuming that we process buckets from left to right (see also Appendix B.4), a simple and useful insertion order within a bucket is from right to left, i.e., by decreasing values of  $s(x)$ . This ordering takes into account that the leftmost part of the bucket is already mostly filled with keys spilled over from the previous bucket whereas the next bucket is still empty. Thus, proceeding from the right to the left, placement is initially easy. With overloaded buckets (negative  $\varepsilon$ ), placement gets more and more difficult and gradually needs to place more and more keys in the next bucket. The analysis in Section 6 suggests that linear dependencies mostly occur in two ways. Either, when the overload of this bucket (or some subrange of it) is too large, or when it runs into the elements placed into the left part of the bucket when the previous bucket was constructed. We make the former event unlikely by choosing appropriate values of  $b$  and  $\varepsilon$ . When the latter event happens, we can bump a range of keys allocated to the leftmost range of the bucket. This bumping scheme is discussed in Section 7.1. The right-to-left ordering then helps us to find the right threshold value. An illustration is shown in Figure 3.

**(Quasi)random order.** The above simple insertion order is limited to situations where the overload per bucket is less than  $w$  most of the time. Otherwise, placement will often fail early, bumping many keys that could actually be placed because their  $s$  falls into a range of slots that remain empty. We can achieve more flexibility in choosing  $\varepsilon$  by spreading keys more uniformly over the bucket during the insertion process. We tried several such approaches. Bu<sup>1</sup>RR uses additional hash bits to make the ordering for bumping independent of the position within the bucket. Another interesting variant is most easy to explain when  $b$  is a power of two. We use  $c = \log b$  bits of hash function value to define the position within a bucket. However, rather than directly using the value as a column number, we use its *mirror image*, i.e., a value  $h_{c-1}h_{c-2} \dots h_1h_0$  addresses column  $h_0h_1 \dots h_{c-2}h_{c-1}$  of a bucket. We also tried a tabulated random permutation, which according to early experiments, works slightly worse than the mirror permutation.

### B.2 Metadata for Bumping Multiple Groups

A disadvantage of threshold-based bumping is that a single failed placement of a key implies that all subsequently placed keys allocated to that bucket must be bumped. This penalty can be mitigated by subdividing the positions in the insertion order into multiple groups that can be bumped separately, e.g., by storing a single bit that indicates whether a group is bumped. Choosing *groups of uniform (expected) size* is simple and fast. It works well when highly compressed metadata is not of primary importance, e.g., when  $r$  is large.

Better compression can be achieved by choosing *groups of variable size*. Bu<sup>1</sup>RRs use groups whose sizes are a geometric progression with a factor about  $\sqrt{2}$  between subsequent group sizes. For example, to cover a bucket of size  $b = 1024$  using 8 bits of metadata, one could use groups of expected sizes 28, 40, 57, 80, 113, 160, 226, and 320. Note that smaller groups cannot and need not be bumped since the main effect of bumping “too much” is that fewer keys spill over to the next bucket which can be rectified there.

---

<sup>20</sup>We do not attempt to optimise  $C$  here. In practice much smaller values of  $C$  are sufficient, see Section 8

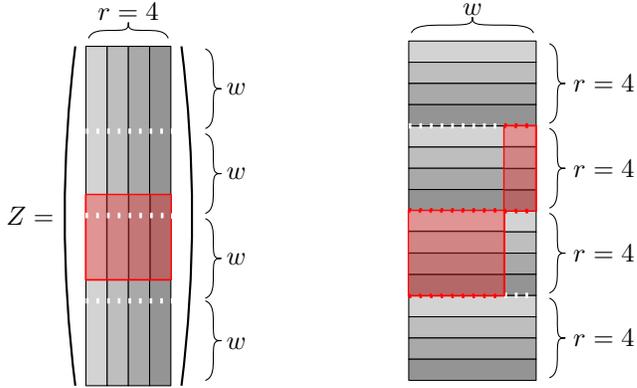


Figure 9: A solution matrix (left) and  $w$ -bit interleaved column-major layout (right) of that matrix. The red shaded region shows the bits used in a query crossing a boundary.

This is also the underlying reason why highly compressed representations of the threshold metadata from Section 7.1 are sufficient.

### B.3 Aggressive versus Cautious Bumping

By default, the generic BuRR solving approach described in Section 7 is greedy, i.e., it tries to place as many keys as possible into the current bucket. This will usually spill over close to  $w$  keys into the next bucket. This increases the likelihood that construction in that bucket fails early. It might be a better approach to be more cautious and try to avoid this situation. For example, when after processing a column  $j$ , more than  $\alpha w$  keys are already placed in the next bucket then all further keys are bumped from the current bucket. More sophisticated balancing approaches are conceivable.

For example, we use a form of cautious bumping for our implementation of Bu<sup>1</sup>RR (see Section 7.4). After placing keys bumped from the previous layer, when processing bucket  $i$ , we first try to place keys in its largest group. Then we try to place the keys in the second largest group in bucket  $i - 1$ , the third largest group in bucket  $i - 2$ , etc. This can be viewed as driving a “wedge” through the buckets.

### B.4 Different Global Insertion Orders

Many of our implemented variants of BuRR process buckets from left to right and, within a bucket, place keys  $x$  from right to left with respect to  $s(x)$ . We also tried the dual approach – traversing the buckets from right to left and then inserting from left to right within a bucket. This behaves identically with respect to space efficiency but leads to far more row operations and much higher construction times. The straightforward ordering from left to right both between and inside buckets does not work well with aggressive bumping – placement frequently fails early. We expect that it may turn out to be a natural order for a cautious bumping strategy. Many other insertion orders can be considered. However, the global left-to-right order has the advantage that spilling keys to the next bucket is cheap since it is still empty. Thus, other orders might have higher insertion time.

### B.5 Table Representation

**Interleaved Versus Contiguous Storage.** *Contiguous* storage is the “obvious” representation of the table by  $m$  slots of  $r$  contiguous bits each. *Interleaved* means that the table is organized as  $rm/w$  words of  $w$  bits each where word  $i$  represents bit  $i \bmod r$  of  $w$  subsequent table entries [22]. This organization allows the extraction of one retrieved bit from two adjoining machine words using population-count instructions. Interleaved representation is advantageous for uses of BuRR as an AMQ data structure since a negative query only has to extract two bits in expectation. Moreover, the implementation directly works for any value of  $r$ .

The contiguous representation, despite its conceptual simplicity, is more difficult to implement, in particular when  $r$  is not a power of two. On the other hand, we expect it to be more efficient when all  $r$  bits need to be retrieved, in particular when  $r$  is large and when the implementation makes careful use of available bit-manipulation<sup>21</sup> and SIMD instructions. This is particularly true when the sparse bit patterns from Section 7.2 are used.

**Embedded Versus Separate Metadata.** The “obvious” way to represent metadata is as a separate array with one entry for each bucket (and a separate hash table for the  $1^+$ -bit representation). On the other hand, if the metadata cannot be assumed to be resident in cache, it is more cache-efficient to store one bucket completely in one cache line, holding both its table entries and meta-information. Then, assuming  $b \geq w$ , querying the data structure accesses only one cache line plus possibly the next cache line when the accessed part of the table extends there. In preliminary experiments with variants of this approach, we observed performance improvements of up to 10% in some cases. We believe that, depending on the implementation and the use case, the difference could also be bigger but have not investigated this further since there are too many disadvantages to this approach: In particular, in the most space-efficient configurations of BuRR, buckets can be bigger than a cache line and the metadata will often fit into cache anyway. Furthermore, when the memory system is not too contended, metadata and primary table can be accessed in parallel, thus eliminating the involved overhead. Finally, embedded metadata is more complicated to implement.

## C Further Experimental Data

---

<sup>21</sup>For example, the BMI2 bit manipulation operations PDEP and PEXT in newer x86 processors look useful.

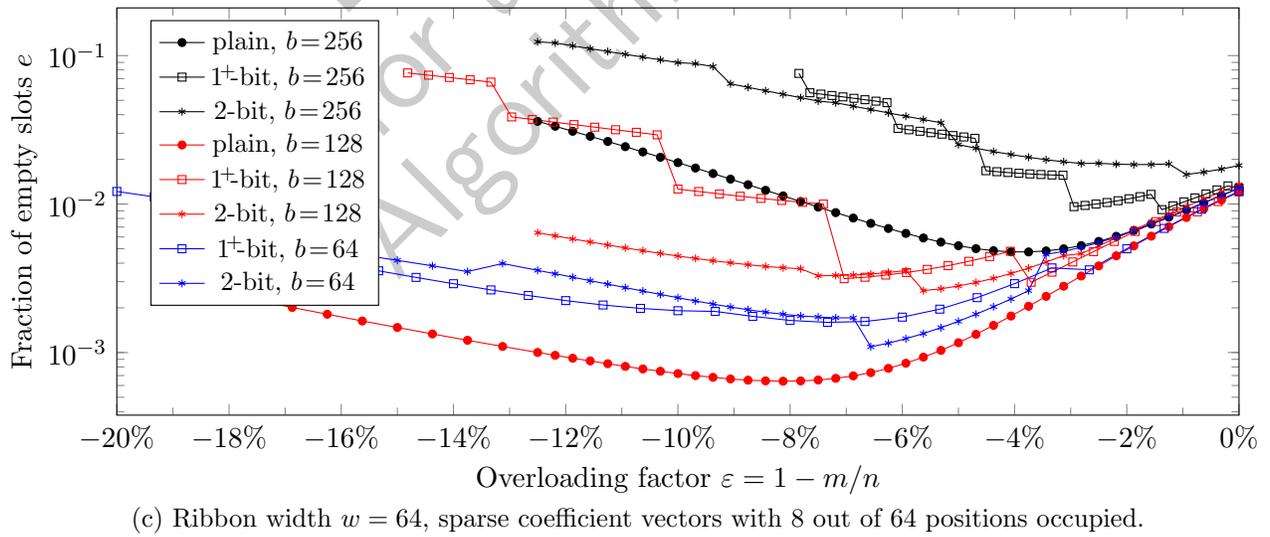
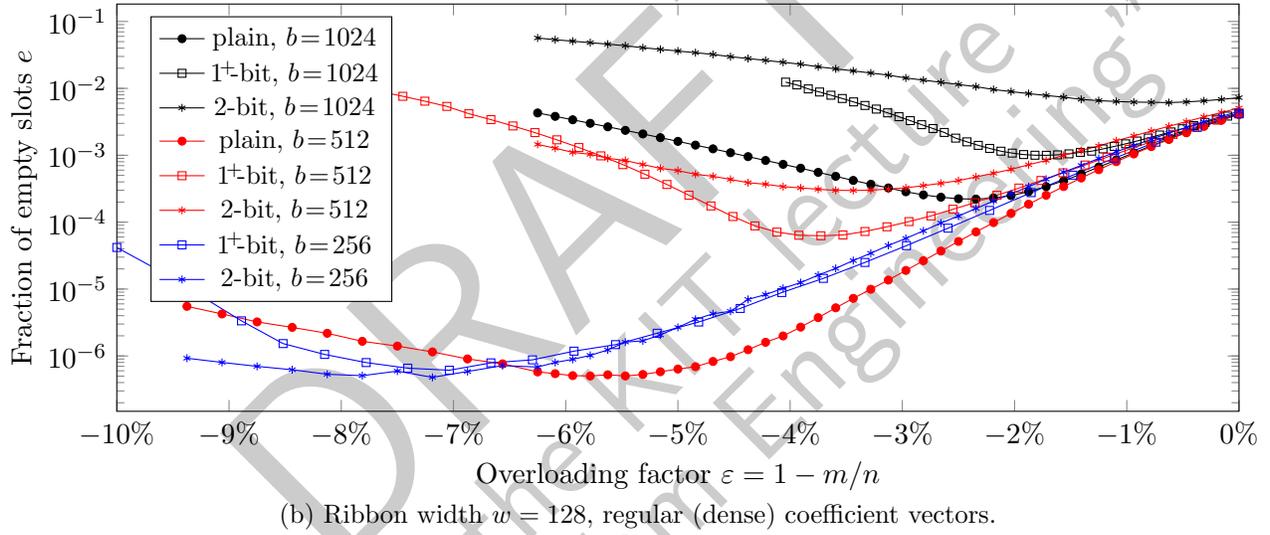
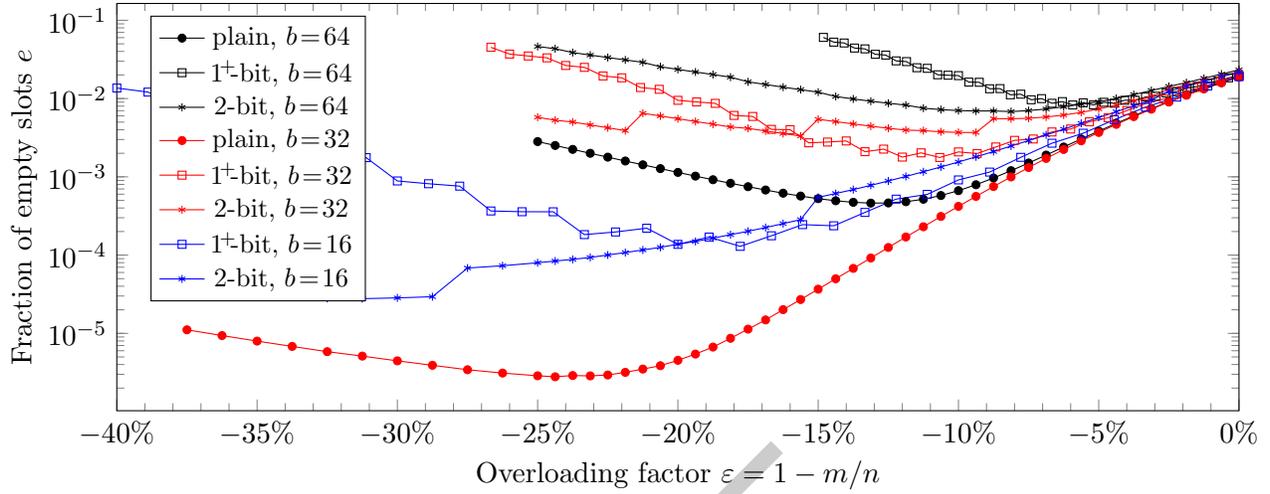


Figure 10: Fraction of empty slots for various configurations of bumped ribbon retrieval, depending on the overloading factor  $\varepsilon$ .

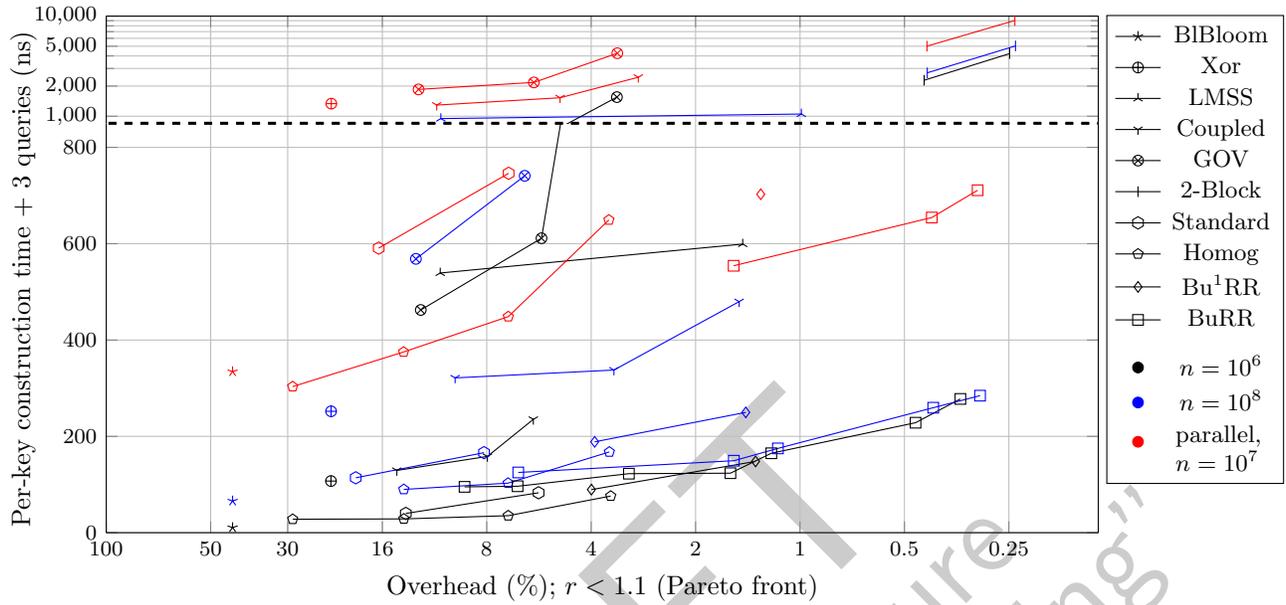


Figure 11: Performance–overhead trade-off for false-positive rate  $> 46\%$  for different AMQs and different inputs. This large false-positive rate is the only one for which we have implementations for GOV [28] and 2-block [18]. Note that the vertical axis switches to a logarithmic scale above 900 ns.

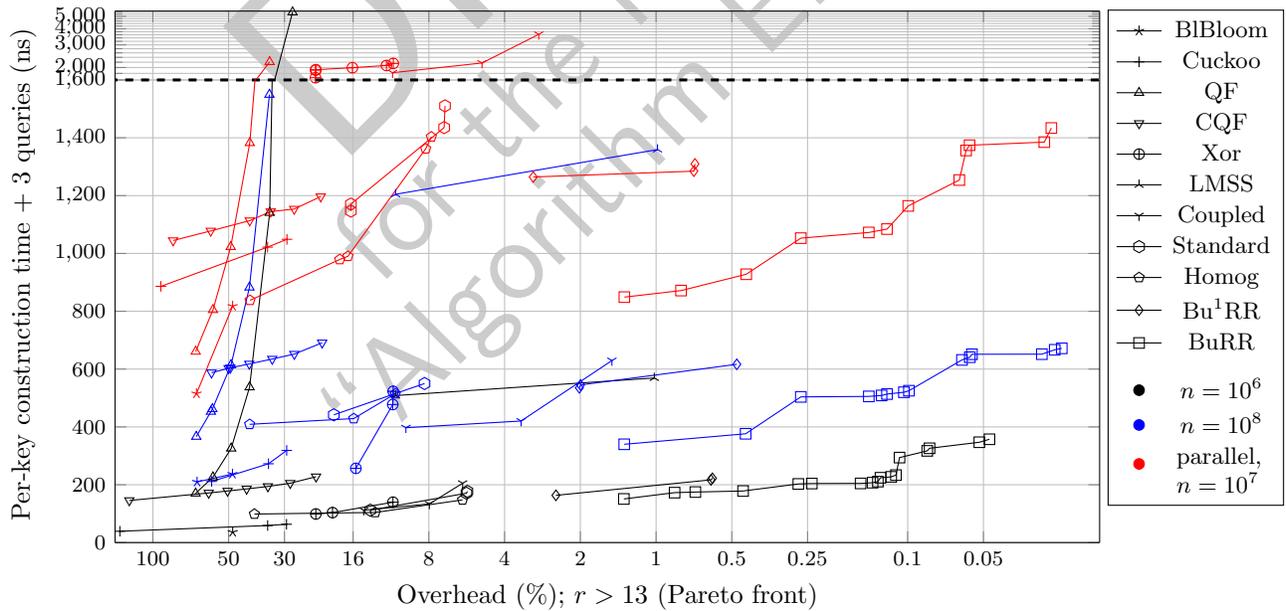


Figure 12: Performance–overhead trade-off for false-positive rate  $< 2^{-13} \approx 0.01\%$  for different AMQs and different inputs. Logarithmic vertical axis above 1600 ns.

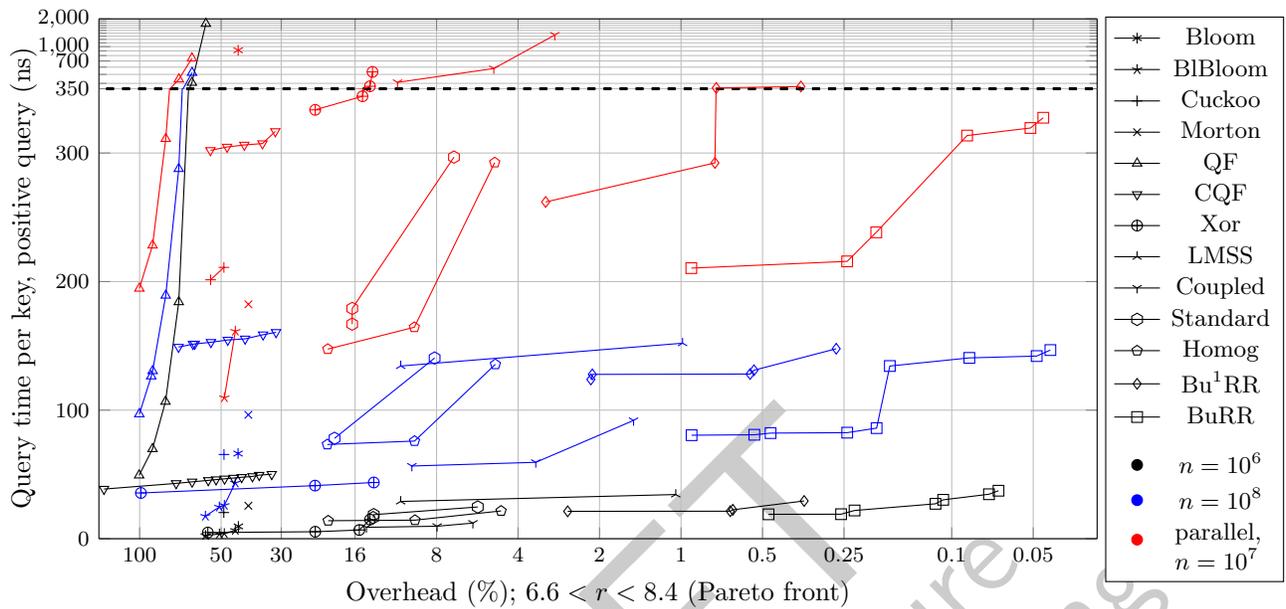


Figure 13: Query time–overhead trade-off for positive queries, false-positive rate between 0.3% and 1% for different AMQs and different inputs. Note that Xor filters have excellent query time sequentially where random fetches can be performed in parallel but are far from optimal in the parallel setting where the total number of memory accesses matters most. Logarithmic vertical axis above 350 ns.

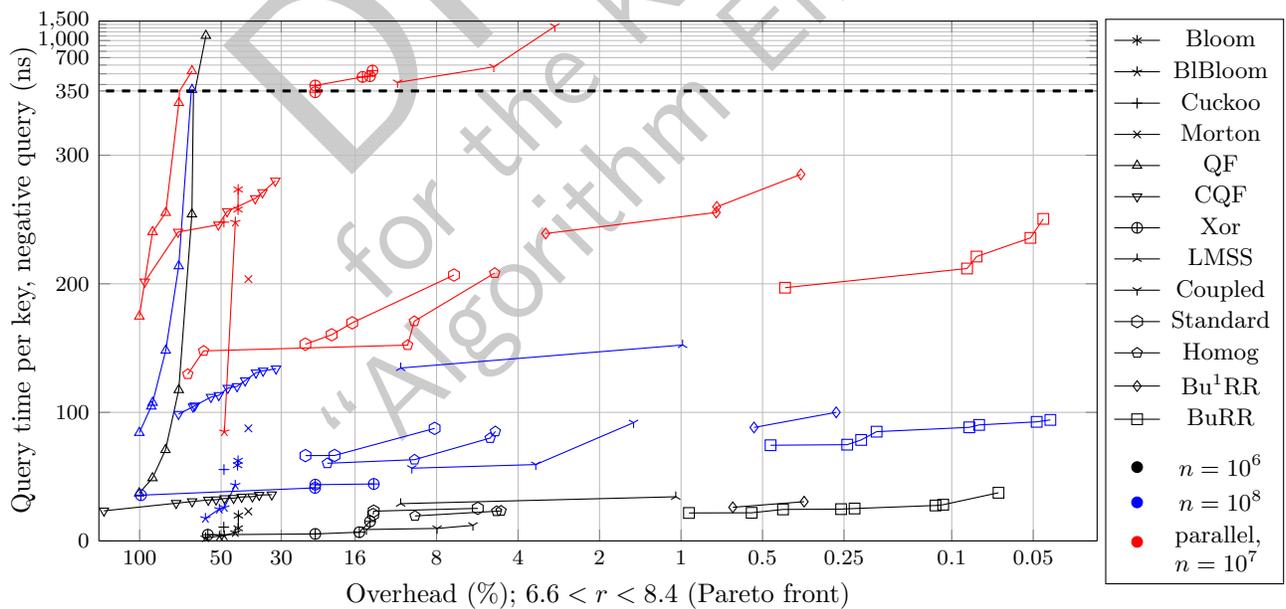


Figure 14: Query time–overhead trade-off for negative queries, false-positive rate between 0.3% and 1% for different AMQs and different inputs. Again, Xor filters perform well sequentially but suffer in the parallel case. Logarithmic vertical axis above 350 ns.

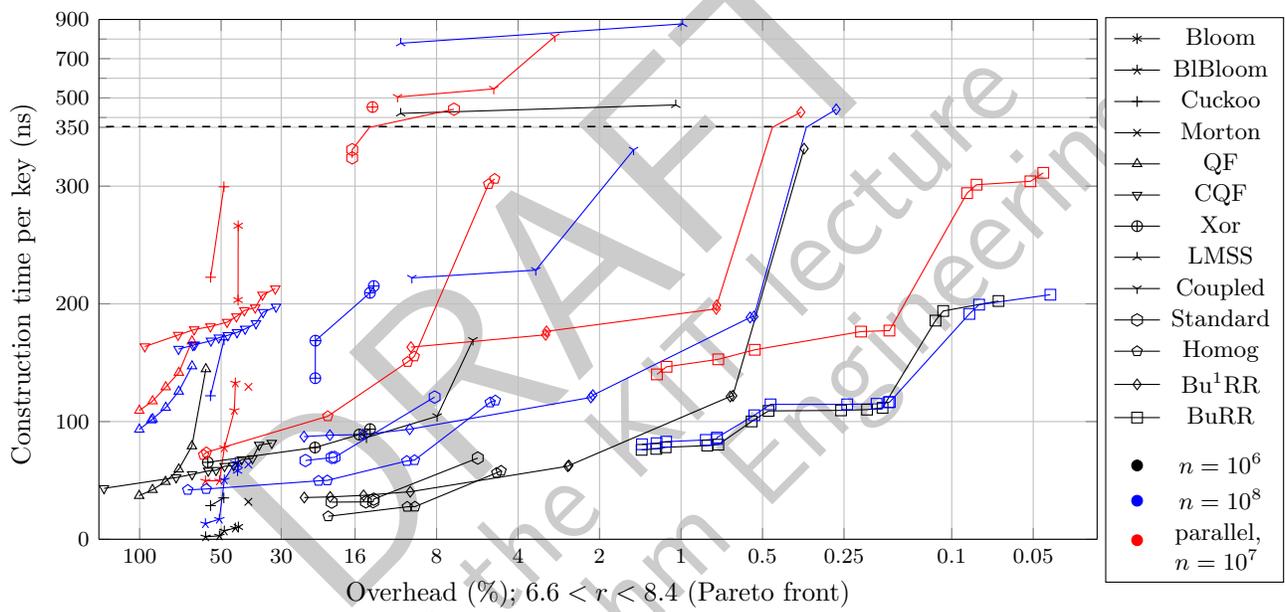


Figure 15: Construction time–overhead trade-off for false-positive rate between 0.3% and 1% for different AMQs and different inputs. Compressed vertical axis above 350 ns.

Table 2: Experimental performance comparisons. Overhead, construction and query times (positive and negative queries) for various AMQs. Tested configurations:  $n = 10^6$  keys,  $n = 10^8$  keys, both sequential, as well as 1280 AMQs with  $n = 10^7$  keys each (total:  $1.28 \times 10^{10}$  keys), constructed and queried in parallel using 64 threads, with each query operating on a randomly chosen AMQ.

Configuration	Space	ns/key, $n = 10^6$			ns/key, $n = 10^8$			parallel, $n = 10^7$		
	ovr %	con	pos	neg	con	pos	neg	con	pos	neg
↓ False positive rate around 1 %, ribbons using $r = 7$ ↓										
Blocked Bloom [35]	52.0	3	3	3	17	24	24	50	116	101
Blocked Bloom [21]	49.8	7	4	4	50	26	26	78	109	85
Blocked Bloom [21] $k = 2$	45.0	9	7	7	57	43	43	110	149	194
Cuckoo12 [24] †	46.3	29	10	7	118	56	52	239	162	282
Cuckoo12 [24]	40.4	35	12	7	166	58	51	288	180	271
Morton [11]	40.6	32	25	22	64	96	87	130	182	203
Xor [30] $r = 7$ ‡	23.0	91	8	8	169	56	56	644	333	348
Xor [30] $r = 8$	23.0	91	5	5	159	41	41	586	386	392
Xor+ [30] $r = 8$	14.4	94	14	15	209	86	85	853	372	475
XorFuse [19] $r = 8$	16;14	89	6	6	215	43	44	453	532	534
LMSS [37] $D = 12, c = 0.91, r = 8$	11.1	421	28	28	779	134	134	not tested		
LMSS [37] $D = 150, c = 0.99, r = 8$	1.0	464	34	34	877	152	152	not tested		
Coupled [50] $k = 4, r = 8$	8;4	104	9	9	229	59	59	546	579	578
Coupled [50] $k = 7, r = 8$	6;2	169	12	12	331	91	91	813	1337	1342
Quotient Filter [5]	81.9	69	432	272	114	225	169	133	385	308
Counting Quotient Filter [44]	67;55	60	45	31	172	153	113	183	307	252
Standard Ribbon $w = 64$	14;20	32	16	20	70	78	66	324	234	194
Standard Ribbon $w = 128$	6;8	69	24	25	121	140	87	464	296	206
Homog. Ribbon $w = 16$	52.2	19	14	20	42	73	61	69	148	128
Homog. Ribbon $w = 32$	20.7	20	13	19	50	73	60	105	147	168
Homog. Ribbon $w = 64$	9.9	28	14	19	67	75	63	155	164	170
Homog. Ribbon $w = 128$	4.9	58	21	23	118	135	85	306	292	208
<b>Bu<sup>1</sup>RR</b> $w = 32$ †	10.3	40	21	26	94	125	88	163	275	286
<b>Bu<sup>1</sup>RR</b> $w = 32$	2.4	62	21	26	121	123	88	174	261	239
<b>BuRR</b> plain $w = 32$	1.4	76	20	26	81	82	79	151	247	210
<b>BuRR</b> 2-bit $w = 32$	1.3	77	19	26	82	82	80	152	233	245
<b>BuRR</b> 1 <sup>±</sup> -bit $w = 32$	0.82	80	29	34	84	88	88	158	240	231
<b>Bu<sup>1</sup>RR</b> $w = 64$	0.62	121	21	26	188	128	90	197	292	300
<b>BuRR</b> plain $w = 64$	0.48	109	18	24	115	82	74	182	229	213
<b>BuRR</b> 2-bit $w = 64$	0.25	110	19	24	115	82	74	190	215	215
<b>BuRR</b> 1 <sup>±</sup> -bit $w = 64$	0.21	110	27	33	115	86	84	189	238	228
<b>Bu<sup>1</sup>RR</b> $w = 128$	0.31	332	29	30	442	147	100	427	369	285
<b>BuRR</b> plain $w = 128$	0.18	209	27	27	214	141	88	319	317	226
<b>BuRR</b> 2-bit $w = 128$	0.10	186	27	27	191	140	88	294	313	211
<b>BuRR</b> 1 <sup>±</sup> -bit $w = 128$	0.06	208	34	37	214	142	92	304	319	235

† Larger space allocated to improve construction time.

‡ Potentially unfavorable bit alignment.

; Standard Ribbon, XorFuse, Coupled, and CQF space overhead depend on  $n$ .

Table 3: Experimental performance comparisons (continued from Table 2).

Configuration	Space	ns/key, $n = 10^6$			ns/key, $n = 10^8$			parallel, $n = 10^7$		
	ovr %	con	pos	neg	con	pos	neg	con	pos	neg
↓ False positive rate around <b>10 %</b> , ribbons using $r = 3$ ↓										
Xor [30] $r = 3$ ‡	23.0	91	6	6	169	51	51	633	292	270
Standard Ribbon $w = 64$	14;20	27	9	9	65	54	51	326	118	138
Standard Ribbon $w = 128$	6;8	62	14	14	114	67	67	467	139	191
Homog. Ribbon $w = 16$	34.6	16	9	9	40	58	59	67	141	140
Homog. Ribbon $w = 32$	16.1	17	8	8	49	45	45	101	153	151
Homog. Ribbon $w = 64$	8.0	24	8	8	65	45	45	158	132	145
Homog. Ribbon $w = 128$	4.0	57	13	13	119	65	65	324	162	193
<b>Bu<sup>1</sup>RR</b> $w = 32$	2.8	60	14	14	119	67	67	175	193	212
<b>BuRR</b> plain $w = 32$	3.2	73	14	21	78	64	64	145	180	170
<b>BuRR</b> 2-bit $w = 32$	2.5	74	14	21	79	64	65	147	227	214
<b>BuRR</b> 1 <sup>+</sup> -bit $w = 32$	1.6	77	23	30	82	67	74	150	202	215
<b>Bu<sup>1</sup>RR</b> $w = 64$	0.8	118	13	13	187	66	67	195	192	220
<b>BuRR</b> plain $w = 64$	1.1	104	13	20	110	63	62	177	160	172
<b>BuRR</b> 2-bit $w = 64$	0.6	106	13	21	111	64	64	188	174	181
<b>BuRR</b> 1 <sup>+</sup> -bit $w = 64$	0.4	106	22	29	111	66	71	189	196	221
↓ False positive rate around $2^{-11} \approx 0.05\%$ , ribbons using $r = 11$ ↓										
Cuckoo16 [24]	30.1	31	11	7	156	56	50	309	180	294
Cuckoo16 [24] †	35.7	28	10	7	119	56	44	243	188	308
CuckooSemiSort	26.6	64	15	14	259	79	79	376	264	326
Morton [11]	36.8	38	40	35	69	167	156	127	314	305
Xor [30] $r = 12$	23.0	89	8	7	163	57	57	632	415	435
Xor+ [30] $r = 11$ ‡	12.8	98	16	16	215	101	99	759	470	494
Quotient Filter [5]	93.6	71	485	304	109	235	175	138	402	324
Standard Ribbon, $w = 64$	14;20	38	23	21	76	143	69	342	294	205
Standard Ribbon, $w = 128$	6;8	71	33	26	124	158	91	442	336	243
Homog. Ribbon, $w = 32$	28.5	24	18	20	53	82	63	109	230	179
Homog. Ribbon, $w = 64$	12.1	32	18	20	70	86	65	165	218	195
Homog. Ribbon, $w = 128$	6.5	58	30	25	115	155	89	281	333	305
<b>Bu<sup>1</sup>RR</b> , $w = 32$	2.4	66	28	27	125	152	92	178	405	277
<b>BuRR</b> plain $w = 32$	0.94	81	28	26	86	145	83	149	327	229
<b>BuRR</b> 2-bit $w = 32$	0.95	82	27	26	87	145	81	150	328	217
<b>BuRR</b> 1 <sup>+</sup> -bit $w = 32$	0.61	85	35	35	90	148	92	158	352	220
<b>Bu<sup>1</sup>RR</b> , $w = 64$	0.57	128	28	27	196	152	95	203	400	296
<b>BuRR</b> plain $w = 64$	0.32	116	27	25	121	146	79	175	334	250
<b>BuRR</b> 2-bit $w = 64$	0.17	118	27	25	123	146	77	164	328	209
<b>BuRR</b> 1 <sup>+</sup> -bit $w = 64$	0.14	115	33	34	120	147	90	189	353	239
<b>BuRR</b> plain $w = 128$	0.13	214	35	28	220	159	92	305	364	244
<b>BuRR</b> 2-bit $w = 128$	0.08	202	35	28	208	159	92	297	358	226
<b>BuRR</b> 1 <sup>+</sup> -bit $w = 128$	0.05	214	42	38	219	160	96	317	389	294

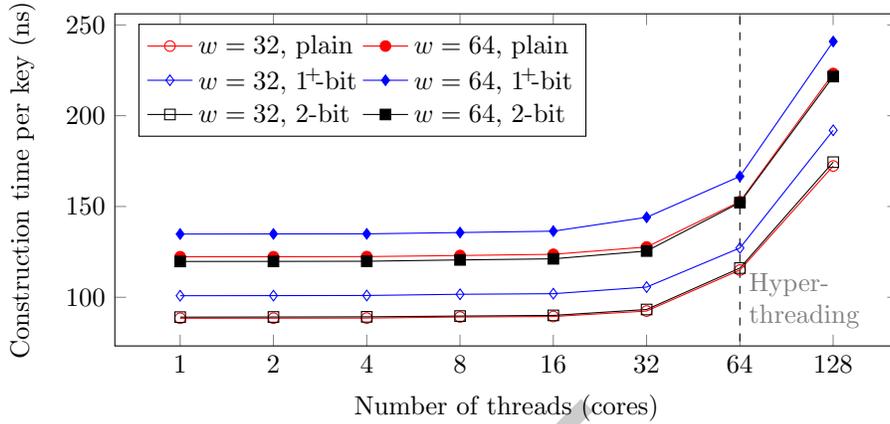
† Larger space allocated to improve construction time.

‡ Potentially unfavorable bit alignment.

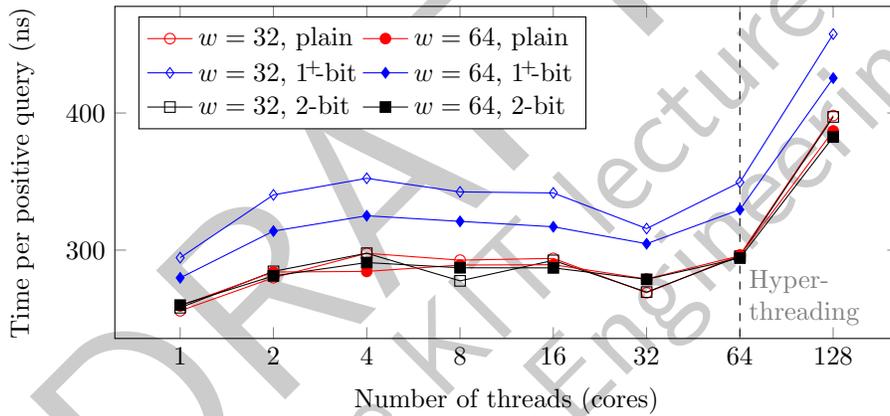
; Standard Ribbon space overhead depends on  $n$ .

Table 4: Selected BuRR configurations for various  $r$ . Sparse coefficient vectors used for rows with threshold compression mode marked <sup>s</sup>.

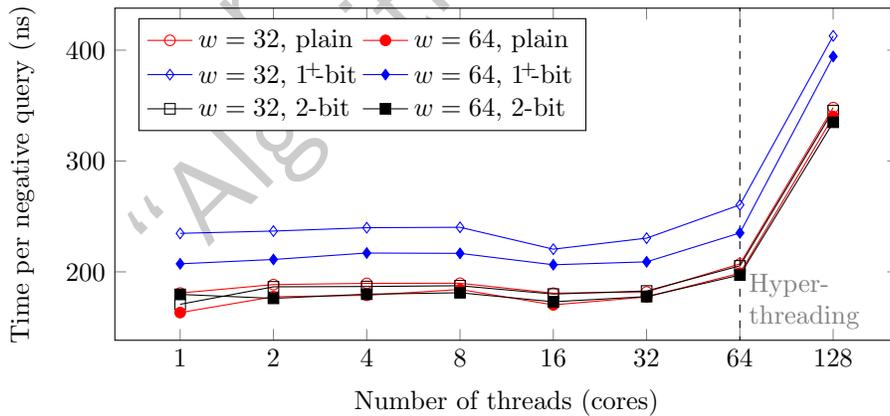
$r$	$w$	$b$	thresh mode	overloading factor $\varepsilon$	empty slots (%)	metabits /bucket	estimated overhead (%)
1	128	512	1 <sup>+</sup> -bit	-0.0470588	0.017477	1.3194	0.275219
1	64	256	2-bit	-0.034375	0.442091	2	1.226810
1	64	128	2-bit <sup>s</sup>	-0.05625	0.261241	2	1.827833
1	32	64	2-bit	-0.08125	0.681598	2	3.828044
2	128	512	1 <sup>+</sup> -bit	-0.0470588	0.017477	1.3194	0.146348
2	64	128	1 <sup>+</sup> -bit	-0.0888889	0.062375	1.2511	0.551393
2	64	128	2-bit	-0.08125	0.013288	2	0.794642
2	64	128	2-bit <sup>s</sup>	-0.05625	0.261241	2	1.044537
2	32	64	2-bit	-0.08125	0.681598	2	2.254821
4	128	512	1 <sup>+</sup> -bit	-0.0431373	0.008797	1.4402	0.079125
4	64	128	2-bit	-0.08125	0.013288	2	0.403965
4	64	128	2-bit <sup>s</sup>	-0.05625	0.261241	2	0.652889
4	32	64	2-bit	-0.08125	0.681598	2	1.468210
8	128	512	1 <sup>+</sup> -bit	-0.0431373	0.008797	1.4402	0.043961
8	64	128	2-bit	-0.08125	0.013288	2	0.208626
8	64	128	2-bit <sup>s</sup>	-0.05625	0.261241	2	0.457065
8	32	64	2-bit	-0.08125	0.681598	2	1.074904
16	128	512	1 <sup>+</sup> -bit	-0.0411765	0.007102	1.5117	0.025557
16	64	128	2-bit	-0.08125	0.013288	2	0.110957
16	64	64	2-bit <sup>s</sup>	-0.065625	0.109362	2	0.304888
16	32	64	plain	-0.13125	0.046154	6	0.632362



(a) Construction of 10 000 filters with  $10^6$  keys each ( $w = 32$ ) or 5 000 filters with  $2 \cdot 10^6$  keys each ( $w = 64$ ), for a total of  $10^{10}$  keys; strong scaling.



(b) Scaling behavior of positive queries on the filters from (a). Each query accesses a randomly chosen filter. Tested with  $10^8$  queries per thread.



(c) Scaling behavior of negative queries on the filters from (a). Each query accesses a randomly chosen filter. Tested with  $10^8$  queries per thread.

Figure 16: Scaling experiments for parallel construction and querying of BuRR.